Climate Time Series Analysis

Classical Statistical and Bootstrap Methods

Atmospheric and Oceanographic Sciences Library

42

Manfred Mudelsee

Springer

Climate Time Series Analysis

ATMOSPHERIC AND OCEANOGRAPHIC SCIENCES LIBRARY

VOLUME 42

For other titles published in this series, go to
www.springer.com/series/5669

Manfred Mudelsee

Climate Time Series Analysis

Classical Statistical and Bootstrap Methods

Springer

Dr. Manfred Mudelsee
Climate Risk Analysis
Schneiderberg 26
30167 Hannover
Germany

Alfred Wegener Institute for Polar and Marine Research
Bussestrasse 24
27570 Bremerhaven
Germany
mudelsee@mudelsee.com

ISSN 1383-8601
ISBN 978-90-481-9481-0 e-ISBN 978-90-481-9482-7
DOI 10.1007/978-90-481-9482-7
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010930656

Cover illustration: Wave@2009 JupiterImages Corporation

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my parents,

Anna-Luise Mudelsee,
née Widmann

and

Richard Mudelsee

Preface

Climate is a paradigm of a complex system. Analysing climate data is an exciting challenge. Analysis connects the two other fields where climate scientists work, measurements and models. Climate time series analysis uses statistical methods to learn about the time evolution of climate. The most important word in this book is "estimation." We wish to know the truth about the climate evolution but have only a limited amount of data (a time series) influenced by various sources of error (noise). We cannot expect our guess (estimate), based on data, to equal the truth. However, we can determine the typical size of that deviation (error bar). Related concepts are confidence intervals or bias. Error bars help to critically assess estimation results, they prevent us from making overstatements, they guide us on our way to enhance the knowledge about the climate. Estimates without error bars are useless.

The complexity of the climate system and the nature of the measurement or modelling act may introduce (1) non-normal distributional shape, (2) serial dependence, (3) uneven spacing and (4) timescale uncertainties. These difficulties prohibit in many cases the classical statistical approach to derive error bars by means of calculating the theoretical distribution of the estimates. Therefore we turn to the bootstrap approach, which generates artificial resamples of the time series in the computer, repeats for each resample the estimation (yielding the replication) and calculates the error bars from the distribution of the replications. The typical number of replications is 2000. This computing-intensive approach yields likely more realistic error bars.

Still, there is theoretical work to be done: how to best preserve the shape and serial dependence in the bootstrap resamples, how to estimate with smallest error bars. Uneven spacing in time series analysis has not been the preferred study object of statisticians. Timescale uncertainties and their effect on error bars (widening, but how much?) is almost completely unexplored. This book adapts existing and introduces new bootstrap algorithms for handling such problems.

We test our methods by means of Monte Carlo experiments. When the true parameter values are known, it is possible to generate random samples and calculate bootstrap error bars and confidence intervals and check whether, for example, a 95% confidence interval for the estimated parameter does indeed contain in 95% of the Monte Carlo runs the known parameter. The number of Monte Carlo runs is typically 47,500. The computational burden increases to 2000 $\times$ 47,500. To create of this book required relatively powerful computers. In Chapter 9, we look on what may become possible when quantum computers exist.

Chapter 1 introduces you to climate time series and their statistical properties. Chapter 2 gives stochastic models of serial dependence or persistence, which are needed in Chapter 3, where bootstrap resampling, the determination of error bars and the construction of confidence intervals is explained. This concludes Part I on fundamental concepts. Chapters 4, 5 and 6 employ the concepts in the univariate setting (Part II), where the sample consists of only one time series. Chapters 7 and 8 deal with the bivariate setting (Part III).

Each of the chapters has a section "Background material," which contains supplementary material from statistics and climatology. You find also reported "stories"—comments, discussions and replies on certain papers in a scientific journal. Such exchanges, as also the "discussion" parts in read statistical papers, provide insight into the production of science—often more intimate than what polished journal articles reveal. The chapters have also a section entitled "Technical issues," where you find, besides information about numerical algorithms, listed software with internet links.

Intuition and creativity is needed for developing statistical estimation techniques for complex problems. Therefore I praise occasionally the artistic scientist, not at least in response to papers that make derogative remarks on that capacity. On the other hand, the artist in us must not forget to look for previous work on the same subject done in other disciplines and to scrutinize the own development by means of objective methods, such as Monte Carlo tests.

Regarding the notation, I have tried to find a route between convention on the one hand and consistency on the other. However, the most important symbols, including t for sampled time, x for a sampled climate variable, n for data size and $\{t(i), x(i)\}_{i=1}^{n}$ for a time series sample, possess their role throughout the book. I take this opportunity to introduce the counterpart of the time series sample, the stochastic process, $\{T(i), X(i)\}_{i=1}^{n}$. I hope that not only statisticians find that traditional distinction (Fisher 1922) between sample (i.e., numbers) and process (i.e., random variables) useful. Regarding the reference list, this notes only the first of the places of a publisher and it gives, in square brackets, additional information. This is not done consistently (e.g., the doi is given mostly to identify more recent papers published by the American Geophysical Union). The author list may be more aptly denoted as "first-author list."

The URL for this book is `http://www.manfredmudelsee.com/book`. It has the links to the sites of the software (including own products) and the data. It has also, inevitably, an errata section. As the person responsible for the content, I offer my apologies in advance of the discovered errors, and I thank you for informing me. My email address is `mudelsee@mudelsee.com`.

Sincere thanks go to my academic teachers, Augusto Mangini and Karl Stattegger, and the hosts of my subsequent stays, Howell Tong and Qiwei Yao, Gerd Tetzlaff, Maureen Raymo and Gerrit Lohmann. They and the colleagues at the respective institutions (Institute of Environmental Physics at the University of Heidelberg, Germany; today's Institute of Geosciences at the University of Kiel, Germany; today's School of Mathematics, Statistics and Actuarial Science at the University of Kent, Canterbury, UK; Institute of Meteorology at the University of Leipzig, Germany; Department of Earth Sciences at Boston University, USA; Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany) helped me to shape my thinking and flourish in the field of climate time series analysis.

The above and following had an influence, gratefully acknowledged, on this book via discussing with me or supplying data, knowledge or literature: Mersku Alkio, Susana Barbosa, Rasmus Benestad, André Berger, Wolfgang Berger (whom I owe the term "ramp"), Mark Besonen, Matthias Bigler, Michael Börngen, Armin Bunde, Steven Burns, Dragos Chirila (who went through the whole manuscript), Ann Cowling, Michel Cruzifix, Anthony Davison (who went through Chapters 1, 2, 3, 4, 5 and 6 of the manuscript), Cees Diks, Reik Donner, Heinz Engel, Dominik Fleitmann, Imola Fodor, Eigil Friis-Christensen, Martin Girardin, the late Clive Granger, Uwe Grünewald, Peter Hall, Gerald Haug, Jonathan Hosking, Daniela Jacob, Malaak Kallache (who went through Chapter 6), Vit Klemeš, Demetris Koutsoyiannis, Thomas Laepple, Peter Laut, Martin Losch (who went through Chapter 9), Werner Metz, Alberto Montanari, Eric Moulines, Alfred Musekiwa, Germán Prieto, Stefan Rahmstorf, Regine Röthlisberger, Henning Rust, Michael Sarnthein, Denis Scholz, Michael Schulz, Walter Schwarzacher, Martin Trauth, Dietmar Wagenbach, Heinz Wanner, Eric Wolff, Peili Wu and Carl Wunsch.

The computing centres from following institutions provided computing time: Alfred Wegener Institute and University of Leipzig. Following institutions gave data: British Antarctic Survey, Cambridge, UK; Global Runoff Data Centre, Koblenz, Germany; National Oceanic and Atmospheric Administration, Washington, DC, USA. Libraries from following research institutes and universities helped with literature: Alfred Wegener Institute, Boston University, University of Massachusetts Boston, Cambridge, Halle, Hannover, Harvard, Heidelberg, Kassel, Leipzig, Massachusetts Institute of Technology, Michigan State University and Yale. Following institutions funded own research that contributed to this book: British Antarctic Survey, Deutsche Forschungsgemeinschaft, European Commission, Niedersächsisches Ministerium für Wissenschaft und Kultur and Risk Prediction Initiative.

Rajiv Monsurate helped adapting the Latex style file.

Last, but not least, I thank the editors at Springer as well as former Kluwer for their patience over the past six years: Chris Bendall, Robert Doe, Gert-Jan Geraeds, Kevin Hamilton, Lawrence Mysak and Christian Witschel.

Hannover, Germany
December 2009

Manfred Mudelsee

Acknowledgements

Copyright permissions are gratefully acknowledged for reproducing Figs. 1.14 and 2.8 (American Geophysical Union, Washington, DC) and the photograph of the author (Silke Storjohann, Hamburg).

Contents

List of Algorithms

List of Figures

List of Tables

Part I

Fundamental Concepts

Chapter 1

Introduction

Superiority of quantitative methods over qualitative

—Popper

"Weather is important but hard to predict"—lay people and scientists alike will agree. The complexity of that system limits the knowledge about it and therefore its predictability even over a few days. It is complex because many variables within the Earth's atmosphere, such as temperature, barometric pressure, wind velocity, humidity, clouds and precipitation, are interacting, and they do so nonlinearly. Extending the view to longer timescales, that is, the climate system in its original sense (the World Meteorological Organization defines a timescale boundary between weather and climate of 30 years), and also to larger spatial and further processual scales considered to influence climate (Earth's surface, cryosphere, Sun, etc.), does not reduce complexity. This book loosely adopts the term "climate" to refer to this extended view, which shall also include "paleoclimate" as the climate within the geologic past.

Man observes nature and climate to learn, or extract information, and to predict. Since the climate system is complex and not all variables can be observed at arbitrary spatial and temporal range and resolution, our knowledge is, and shall be, restricted and uncertainty is introduced. In such a situation, we need the statistical language to acquire quantitative information. For that, we take the axiomatic approach by assuming that to an uncertain event ("it rains tomorrow" or "before 20,000 years the tropics were more than 5°C colder than at present") a probability (real number between 0 and 1) can be assigned (Kolmogoroff 1933). Statistics then deciphers/infers events and probabilities from data. This is an

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_1,

assumption like others in the business: three-dimensional space, time arrow and causality, mathematical axioms (Kant 1781; Polanyi 1958; Kandel 2006). The book also follows the optimistic path of Popper (1935): small and accurately known ranges of uncertainty about the climate system enable more precise climate hypotheses to be tested, leading to enhanced knowledge and scientific progress. Also if one shares Kuhn's (1970) view, paradigm shifts in climatology have better success chances if they are substantiated by more accurate knowledge. It is the aim of this book to provide methods for obtaining accurate information from complex time series data.

Climate evolves in time, and a stochastic process (a time-dependent random variable representing a climate variable with not exactly known value) and time series (the observed or sampled process) are central to statistical climate analysis. We shall use a wide definition of trend and decompose a stochastic process, X, as follows:

$$X(T) = X_{\text{trend}}(T) + X_{\text{out}}(T) + S(T) \cdot X_{\text{noise}}(T), \tag{1.1}$$

where T is continuous time, $X_{\text{trend}}(T)$ is the trend process, $X_{\text{out}}(T)$ is the outlier process, $S(T)$ is a variability function scaling $X_{\text{noise}}(T)$, the noise process. The trend is seen to include all systematic or deterministic, long-term processes such as a linear increase, a step change or a seasonal signal. The trend is described by parameters, for example, the rate of an increase. Outliers are events with an extremely large absolute value and are usually rare. The noise process is assumed to be weakly stationary with zero mean and autocorrelation. Giving $X_{\text{noise}}(T)$ standard deviation unity enables introduction of $S(T)$ to honour climate's definition as not only the mean but also the variability of the state of the atmosphere and other compartments (Brückner 1890; Hann 1901; Köppen 1923). A version of Eq. (1.1) is written for discrete time, $T(i)$, as

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \tag{1.2}$$

using the abbreviation $X(i) \equiv X(T(i))$, etc. However, for unevenly spaced $T(i)$ this is a problematic step because of a possibly non-unique relation between $X_{\text{noise}}(T)$ and $X_{\text{noise}}(i)$, see Section 2.1.2.1. The observed, discrete time series from process $X(i)$ is the set of size n of paired values $t(i)$ and $x(i)$, compactly written as $\{t(i), x(i)\}_{i=1}^{n}$. To restate, the aim of this book is to provide methods for obtaining quantitative estimates of parameters of $X_{\text{trend}}(T)$, $X_{\text{out}}(T)$, $S(T)$ and $X_{\text{noise}}(T)$ using the observed time series data $\{t(i), x(i)\}_{i=1}^{n}$.

A problem in climate analysis is that the observation process superimposes on the climatic process. $X_{\text{noise}}(T)$ may show not only climatic but also measurement noise. Outliers can be produced by power loss in the recording instrument. Non-climatic trends result, for example, from changing the recording situation. An example is temperature measurements made in a town that are influenced by urbanization (meaning an increased heat-storage capacity). However, measurement noise can in principle be reduced by using better instruments, and outliers and trends owing to the observation system can be removed from the data—climatologists denote such observation trend free data as homogeneous.

A further problem in real-world climatology is that also the time values have to be estimated, by dating (Section 1.1). Dating errors are expected to add to the noise and make the result more uncertain.

Consider a second climate variable, $Y(T)$, composed as $X(T)$ in Eq. (1.1) of trend, outliers, variability and noise. The interesting new point is the dependence between $X(T)$ and $Y(T)$. Take as example the relation between concentration of CO_2 in the atmosphere and the global surface temperature. In analogy to univariate X, we write $\{X(T), Y(T)\}$, $\{T(i), X(i), Y(i)\}$ and $\{t(i), x(i), y(i)\}_{i=1}^{n}$ for such bivariate processes and time series. This book describes methods only for uni- and bivariate time series. Possible extensions to higher dimensions are mentioned in Chapter 9.

$\{t(i), x(i), y(i)\}_{i=1}^{n}$ need not result from the natural climate system but may also be the output from a mathematical climate model. Such models attempt to rebuild the climate system by connecting climate variables with governing mathematical–physical equations. Owing to the limited performance of computers and the uncertain knowledge about climatic processes, climate models are necessarily limited in spatial, processual and temporal resolution (McAvaney et al. 2001; Randall et al. 2007). On the other hand, climate models offer the possibility to perform and repeat climate experiments (say, studying the influence of doubled concentrations of CO_2 in the atmosphere on precipitation in dependence on different model implementations of the role of clouds).

1.1 Climate archives, variables and dating

Climate archives "contain" the time series. The measured variables $(x(i), y(i))$ either are of direct interest, as in case of precipitation and temperature, or they bear indirect information (indicator or proxy variables). The estimated times $(t(i))$, in geosciences often called timescale, are obtained either by direct, absolute dating methods or indirectly by comparison with another, dated time series. Table 1.1 gives an overview about climate archives and absolute dating methods. Table 1.2 informs about climate variables and their proxies studied in this book. More details are provided in Figs. 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 and

Table 1.1. Main types of climate archives, covered time ranges and absolute dating methods.

Climate archive	*Time range*									
	Holocene				Pleistocene		Pl.			Geological epoch
	10^0	10^1	10^2	10^3	10^4	10^5	10^6	10^7	10^9	Years B.P.
Marine sediment core										
Ice core										
Lake sediment core										
Speleothem										
Tree-rings										
Documents										
Direct measurements										
Climate model										
										Dating method
										K/Ar
										Dosimeter
										U/Th
										^{14}C
										^{210}Pb
										Layer counting

Dark shading means "frequently used," *light shading* means "occasionally used." Pl., Pliocene; B.P., "before the present." Background material (Section 1.6) gives details and references on geological epochs (also before Pliocene), archives and dating.

1.10, where some of the time series analysed in this book are presented, and in the background material (Section 1.6).

1.2 Noise and statistical distribution

The noise, $X_{\text{noise}}(T)$, has been written in Eq. (1.1) as a zero-mean and unit-standard deviation process, leaving freedom as regards its other second and higher-order statistical moments, which define its distributional shape and also its spectral and persistence properties (next section). The probability density function (PDF), $f(x)$, defines

$$\text{prob}\left(a \le X_{\text{noise}}(T) \le a+\delta\right)\big|_{\delta\to 0} = \int_a^{a+\delta} f(x)dx, \tag{1.3}$$

putting our incomplete knowledge in quantitative form.

For analysing, by means of explorative tools, the shape of $f(x)$ using time series data $\{t(i), x(i)\}_{i=1}^{n}$, it is important to estimate and remove the trend from the data. An unremoved trend would deliver a false, broadened picture of $f(x)$. Trend removal has been done for constructing Fig. 1.11, which shows histograms as estimates of the distributions of $X_{\text{noise}}(T)$ for various climate time series. The estimation of trends is

Table 1.2. Climate archives and variables studied in this book (selection).

Climate archive	*Location*	*Time range* (a)	*Proxy variable*	*Resolution* (a)	*Climate variable*
Marine sediment core	Eastern equatorial Pacific	10^6	$\delta^{18}O$, benthic foraminifera	10^3	Ice volume, bottom water temperature
Ice core	Antarctica	10^5	CO_2, air bubbles	10^3	CO_2, atmosphere
			δD, ice	10^2	Air temperature
	Greenland	10^5	SO_4 content, ice	10^0	Volcanic activity
			Ca content, ice	10^0	Aeolian dust, wind
			Dust content, ice	10^0	Aeolian dust, wind
			Conductivity,[a] ice	10^0	Soluble material, wind
			Na content, ice	10^0	Seasalt, wind
Tree-rings	Worldwide	10^4	$\Delta^{14}C$, wood	10^0	Solar irradiance, ocean circulation
Lake sediment core	Boston area	10^3	Varve thickness	10^0	Wind[b]
Speleothem	Southern Oman	10^4	$\delta^{18}O$, carbonate	10^1	Monsoon rainfall
Documents	Weikinn source texts	10^3		10^0	Floods, river Elbe
Climate model	Hadley Centre, HadCM3	10^2		10^0	River runoff
Direct measurements	Siberia, North Atlantic	10^2		10^{-1}	Surface temperature

Time range refers to the length of a record, resolution to the order of the average time spacing (see Section 1.4). "Proxy variable" denotes what was actually measured on which material. "Climate variable" refers to the climatic variations recorded by the variations in the proxy variable. The ability of a proxy variable to indicate a climate variable depends on the characteristic timescales (between resolution and time range). For example, $\delta^{18}O$ variations in benthic foraminifera over timescales of only a few decades do not record ice-volume variations (which are slower). The Weikinn source texts are given by Weikinn (1958, 1960, 1961, 1963, 2000, 2002).

[a] Electrical conductivity of the melted water.

[b] Extremely thick varves (graded beds) indicate extremely high wind speed (hurricane).

one of the primary tasks in climate time series analysis and described in Chapter 4. In Fig. 1.11, outliers, sitting at the tail of the distribution, are tentatively marked. The variability, $S(T)$, has only been normalized in those panels in Fig. 1.11 where it is not time-constant.

As the histogram estimates of the PDFs reveal, some distributions (Fig. 1.11b, i, j) exhibit a fairly symmetrical shape, resembling a Gaussian

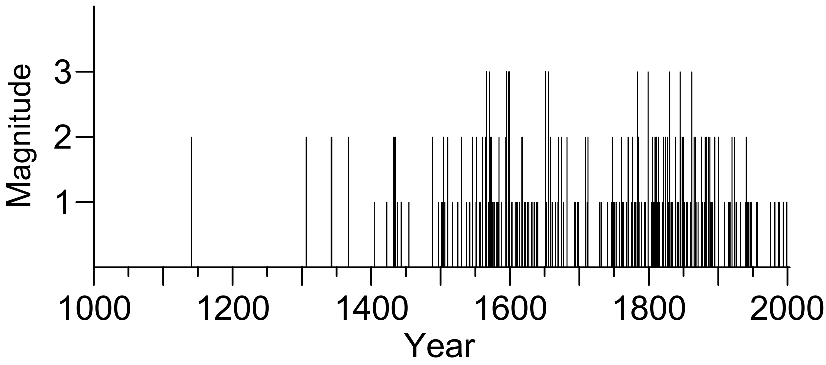


Figure 1.1. Documentary data: floods of the river Elbe during winter over the past 1000 years. x, the flood magnitude, is in three classes (1, minor; 2, strong; 3, exceptionally strong). Hydrological winter is from November to April. Data for $t \leq 1850$ were extracted from Curt Weikinn's compilation (Weikinn 1958, 1960, 1961, 1963, 2000, 2002) of source texts on hydrography in Europe; accuracy of flood dates is ~ 1 month. Data for $t > 1850$ were inferred from daily measurements of water stage and runoff (volume per time interval) at Elbe station Dresden (Global Runoff Data Centre, Koblenz, Germany) via a calibration of magnitude versus water stage/runoff (Mudelsee et al. 2003). Because floods can last up to several weeks, only the peaks in stage/runoff were used to ensure independence of the data. Total number of points is 211. Data sparseness for $t \lessapprox 1500$ is likely caused by document loss (inhomogeneity). One climatological question associated with the data is whether floods occur at a constant rate or there is instead a trend. (Data from Mudelsee et al. 2003.)

(Fig. 3.1). Other PDFs (Fig. 1.11c–h, k), however, have more or less strongly right-skewed shape. Possibly Fig. 1.11d (Vostok δD) reflects a bimodal distribution.

Table 1.3 informs about the size of the variability, $S(T)$, in relation to the uncertainty associated with the pure measurement for the time series analysed here. $S(T)$ reflects the variability of the climate around its trend (Eq. 1.1), the limited proxy quality when no directly measured variables are available and, finally, measurement error. As is evident from the data shown, the measurement error is often comparably small in climatology. It is in many studies that use proxy variables one of the major tasks to quantify the proxy error. For example, if δ^{18}O in shells of benthic foraminifera from deep-sea sediment cores is used as proxy for global ice volume, bottom-water temperature fluctuations make up nearly 1/3 of $S(T)$, see Table 1.3.

A relation proxy variable–climate variable established under laboratory conditions is not perfect but shows errors, quantifiable through regression (Chapter 8). Assuming that such a relation holds true also in

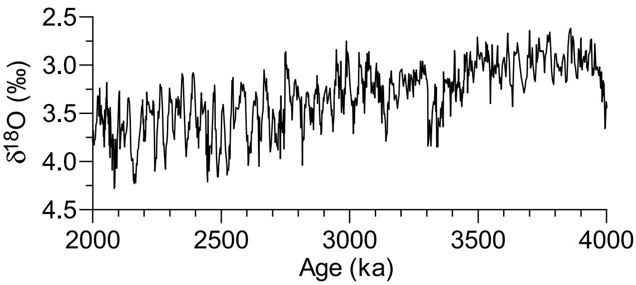


Figure 1.2. Marine sediment core data: δ^{18}O record from Ocean Drilling Program (ODP) site 846 (eastern equatorial Pacific) within 2–4 Ma. The core was drilled from a ship through $\sim$ 3300 m water into the ocean floor, it has a length of $\sim$ 460 m and a diameter of $\sim$ 35 cm. The oxygen isotope record (Shackleton et al. 1995b) was measured on the calcareous shells of benthic foraminifera, mainly *C. wuellerstorfi* and *Uvigerina* spp., using a mass spectrometer. Values are given in delta notation: $\delta^{18}\mathrm{O} = \left[(^{18}\mathrm{O}/^{16}\mathrm{O})_{\mathrm{sample}}/(^{18}\mathrm{O}/^{16}\mathrm{O})_{\mathrm{PDB}} - 1\right] \cdot 1000‰$, where $(^{18}\mathrm{O}/^{16}\mathrm{O})$ is the number ratio of oxygen isotopes ^{18}O and ^{16}O and PDB is "Pee Dee Belemnite" standard. A value of 0.64‰ was added to all δ^{18}O values from *C. wuellerstorfi* to correct for a species-dependent offset (Shackleton and Hall 1984). The depth scale was transformed into a timescale in several steps (Shackleton et al. 1995a). First, biostratigraphic positions, that is, core depths documenting first or last appearances of marine organisms, provided a rough time frame. (Unlike many other marine sediment cores, site ODP 846 lacks a magnetostratigraphy, that is, recorded events of reversals of the Earth's magnetic field, which might had improved the temporal accuracy at this step.) Second, a proxy record of sediment density was measured using a gamma-ray attenuation porosity evaluation (GRAPE) tool. Third, the ODP 846 GRAPE record was tuned (Section 1.6) to the combined GRAPE record from ODP sites 849, 850 and 851. This stacked GRAPE record had in turn been previously tuned to the time series of solar insolation at 65°N (Berger and Loutre 1991), calculated using standard procedures from astronomy. The reason behind this cross-tuning procedure is the observation (Hays et al. 1976) that Earth's climatic variations in the order of tens of thousands to several hundreds of thousands of years are influenced by solar insolation variations. Since the sedimentation rate in the geographic region of site ODP 846 varies with climate (Shackleton et al. 1995a), one cannot assume a constant accumulation of the marine archive. Hence, the dates of sediment samples between the biostratigraphic fixpoints cannot be obtained by interpolation and the GRAPE density records had to be used to obtain an absolute timescale by tuning. Note that time runs "in paleoclimatic manner" from the right to the left. In the same fashion, the δ^{18}O scale is inverted such that glacial conditions (large ice volume, low bottom water temperature or large δ^{18}O values) are indicated by the bottom part and interglacial conditions by the top part of the plot. The number of data points, n, within the shown interval is 821, the average spacing is $\sim$ 2.4 a. A comparison between absolutely dated and tuned magnetostratigraphic timescales for the Pliocene to early Pleistocene (Mudelsee 2005) suggests an average age deviation of $\sim$ 25 ka; this value can also serve to indicate the magnitude of the absolute error of the ODP 846 timescale. The record indicates variations in global ice volume and regional bottom water temperature. One task is to quantify the long-term δ^{18}O increase, which reflects the glaciation of the northern hemisphere in the Pliocene.

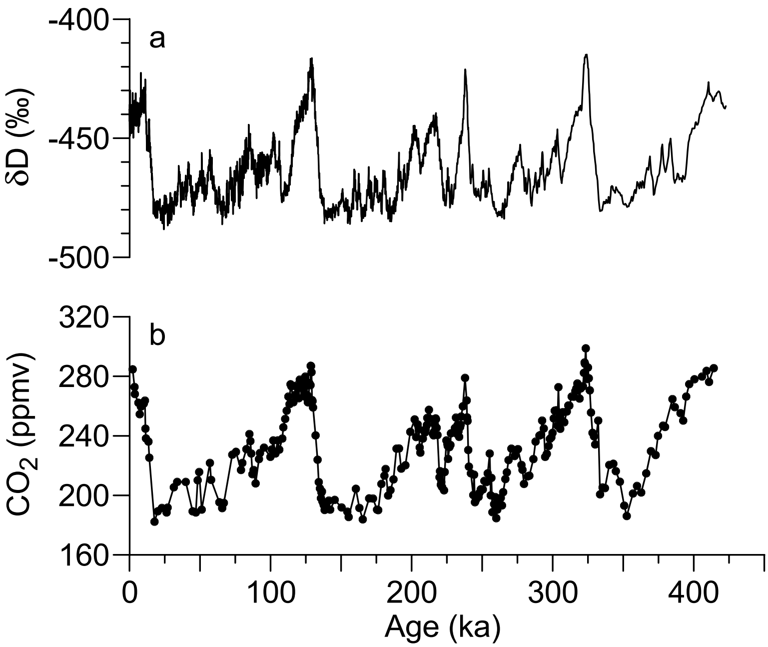


Figure 1.3. Ice core data: deuterium and CO_2 records from the Vostok station (Antarctica) over the past 420,000 years. The core was drilled into the ice (diameter: 12 cm, length: 3623 m) and recovered in segments. The deuterium record (**a**) was measured on ice material using a mass spectrometer. Values are given in delta notation: $\delta\mathrm{D} = [(\mathrm{D/H})_{\mathrm{sample}}/(\mathrm{D/H})_{\mathrm{SMOW}} - 1] \cdot 1000‰$, where (D/H) is the number of D particles over the number of H particles and SMOW is "Standard Mean Ocean Water" standard. Total number of points, n, is 3311. The CO_2 record (**b**) was measured on air bubbles enclosed in the ice. Values are given as "parts per million by volume," n is 283. In **b,** values (*dots*) are connected by lines; in **a,** only lines are shown. The present-day CO_2 concentration ($\sim$ 389 ppmv) is not recorded in **b.** The construction of the timescale (named GT4) was achieved using a model of the ice accumulation and flow. Besides glaciological constraints, it further assumed that the points at 110 and 390 ka correspond to dated stages in the marine isotope record. Construction of the CO_2 timescale required additional modelling because in the ice core, air bubbles are younger in age than ice at the same depth. One climatological question associated with the data is whether variations in CO_2 (the values in air bubbles presenting the atmospheric value accurately) lead over or lag behind those of deuterium (which indicate temperature variations, low δD meaning low temperature). (Data from Petit et al. 1999.)

the geologic past increases the proxy error. Spatially extending the range for which a variable is thought to be representative is a further source of error. This is the case, for example, when variations in air temperature in the inversion height above Antarctic station Vostok are used to repre-

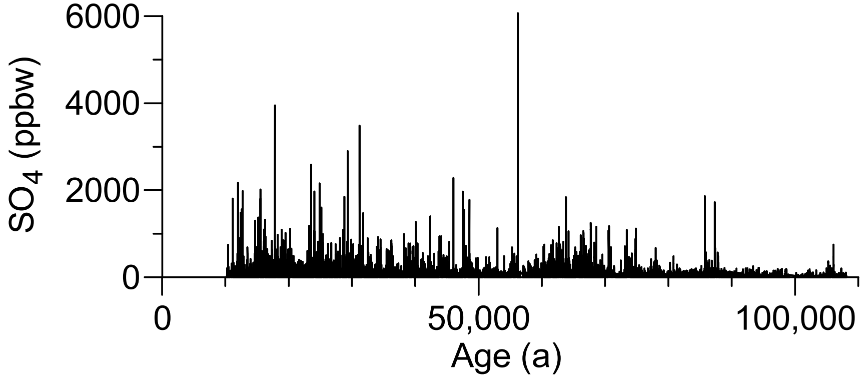


Figure 1.4. Ice core data: sulfate record from the NGRIP core (Greenland) over the interval from ~ 10 to ~ 110 ka. The sulfate content was determined by continuously melting the ice core along its axis and measuring SO_4 of the melt water by means of a photometer (continuous flow analysis, CFA; see Röthlisberger et al. (2000) and Bigler et al. (2002)); ppbw, parts per billion by weight. Meltspeed and signal dispersion limit the length resolution to ~ 1 cm over the measured record length (1530 m). In the young part of the record ($t \leq 105$ ka), the NGRIP timescale was obtained by tuning to the ss09sea timescale of the Greenland GRIP ice core (Johnsen et al. 2001) using the records of ice isotopes (North Greenland Ice Core Project members 2004), electrical conductivity and dielectric properties. In the old part, the NGRIP timescale was obtained by tuning to the GT4 timescale of the Vostok ice core (Fig. 1.3) using the records of δ^{18}O and methane concentration. (An absolutely dated alternative to the GRIP ss09sea timescale was published by Shackleton et al. (2004).) The sulfate record was finally averaged to 1-year resolution. Using the Ca and Na records, proxies for mineral dust and seasalt content, respectively, it is possible to remove peaks in the sulfate record from dust and salt input—the remaining peaks in the "excess" SO_4 record, shown here, likely reflect the input from volcanic eruptions via the atmosphere. The record therefore bears the possibility to reconstruct volcanic activity throughout the last glacial period. (Data from Bigler M 2004, personal communication.)

sent those of the total southern hemisphere. However, such uncertainties are often unavoidable when general statements about the climate system are sought. All individual noise influences on a climate variable (natural variability, proxy and measurement noise) seem to produce a process $X_{\text{noise}}(T)$ with a PDF that is better described by a product than a sum of individual PDFs and that likely has a right-skewed shape, such as the lognormal distribution (Aitchison and Brown 1957).

1.3 Persistence

The other property of $X_{\text{noise}}(T)$ besides distributional shape regards serial dependence. The autocovariance, $E\left[X_{\text{noise}}(T_1) \cdot X_{\text{noise}}(T_2)\right]$ for

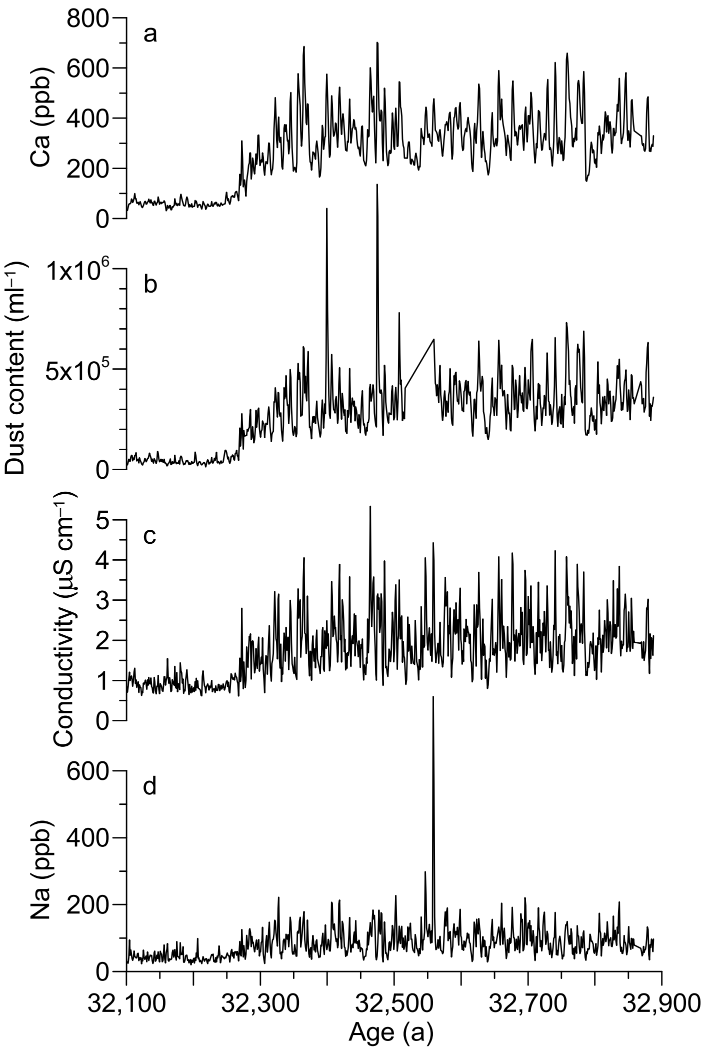


Figure 1.5. Ice core data: Ca concentration (**a**), dust content (**b**), electrical conductivity (**c**) and Na concentration (**d**) from the NGRIP core (Greenland) during the onset of Dansgaard–Oeschger (D–O) event 5. The four variables were measured using CFA on the melted water (Fig. 1.4). ppb, parts per billion; ml^{-1}, number of particles per ml; $S\,m^{-1}$, SI unit for electrical conductivity. A data gap (hiatus) exists at around 32,550 a in the dust-content record. Records were "downsampled" to annual resolution. The Ca record indicates variations of mineral dust transported to the atmosphere over Greenland, the dust content indicates atmospheric dust load, electrical conductivity is a proxy for input of soluble material (integrating various environmental signals) and Na is a proxy for seasalt. One climatological question is whether the changes in all four variables happened simultaneously at the onset of D–O event 5. D–O events are short-term warmings during the last glacial period. (Data from Röthlisberger R 2004, personal communication.)

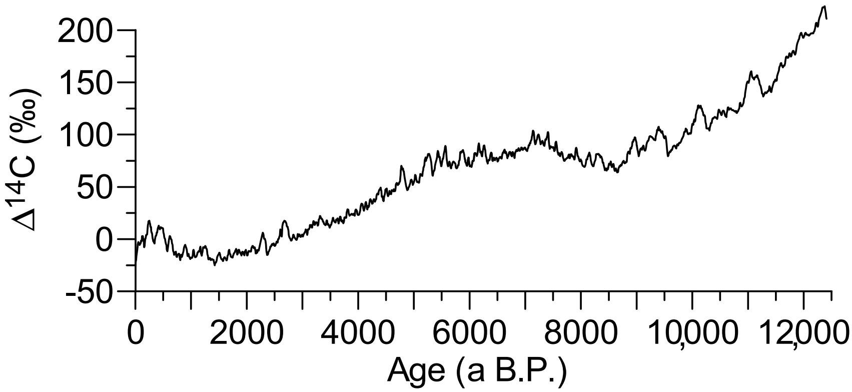


Figure 1.6. Tree-ring data: record of atmospheric radiocarbon content over the past 12,410 years. The tree-ring radiocarbon equilibrates with atmospheric radiocarbon via the photosynthetic cycle. The ^{14}C radioactivity was measured by counting the β particles on CO_2 produced by combusting the wood material. Original sampling resolution was yearly (individual tree-rings) and lower; data shown are 5-year averages ($n = 2483$). The values are presented in delta notation (Fig. 1.3) with the oxalic acid standard of the National Bureau of Standards, for conventional reasons "Δ" is used instead of "δ." The timescale (given as years before present (B.P.) where "present" is, as in "radiocarbon terminology," the year 1950) is based on a counted tree-ring chronology, established by matching radiocarbon patterns from individual trees. Since the age spans of the trees overlap, it is possible to go back in time as far as shown (and beyond). Since the radiocarbon data act as a proxy for solar activity (high Δ^{14}C means low solar irradiance), it is possible to analyse Sun–climate connections by studying correlations between Δ^{14}C and climate proxy records. (Data from Reimer et al. 2004.)

$T_1 \neq T_2$, is here of interest; higher-order moments are neglected. Lag-1 scatterplots ($x(i-1)$ versus $x(i)$) of the climate time series, using detrended $\{t(i), x(i)\}_{i=1}^{n}$ as realizations of the noise process, explore the autocovariance structure (Fig. 1.12). It is evident that all examples exhibit a more or less pronounced orientation of the points along the 1:1 line. This indicates positive serial dependence, or "memory," also called persistence in the atmospheric sciences. The reason for that memory effect is twofold. First, it is characteristic for many types of climatic fluctuations (Wilks 1995). Second, it can be induced by the sampling of the data. A record sampled at high resolution has often stronger persistence than when sampled at low resolution (see next section).

The lag-1 scatterplots (Fig. 1.12) reflect also the right-skewed shape of many of the distributions (more spreading towards right-up) and let some outliers appear.

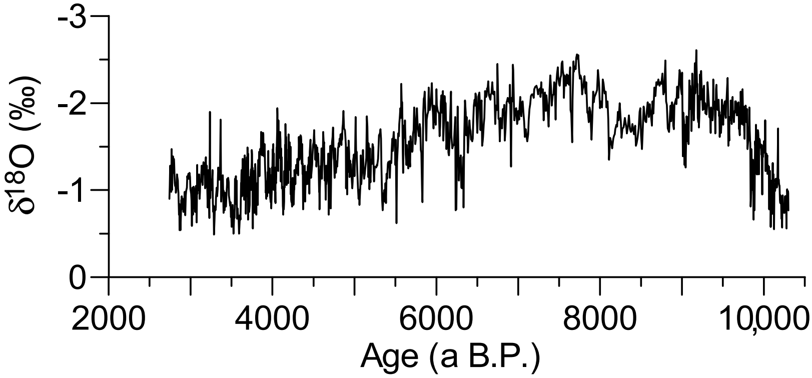


Figure 1.7. Speleothem data: oxygen isotope record from stalagmite Q5 from southern Oman over the past 10,300 years. Along the growth axis of the nearly 1 m long stalagmite, every $\sim$ 0.7 mm about 5 mg material ($CaCO_3$) was drilled, yielding $n = 1345$ samples. The carbonate powder was analysed with an automatic preparation system linked to a mass spectrometer to determine the δ^{18}O values. (The (^{18}O/^{16}O) ratio is given relative to the Vienna Pee Dee Belemnite (VPDB) standard analogously to the description in Fig. 1.3.) The timescale (years before 1950) is based on 18 U/Th mass-spectrometric ages, obtained on separated and purified material. Dates for samples between absolutely dated positions were obtained by linear interpolation. Time runs from right to left. The δ^{18}O scale is inverted "in paleoclimatic manner" so that the transition from the last glacial to the present Holocene interglacial at around 10 ka is "upwards." Note that growth of stalagmite Q5 ceased at $\sim$ 2740 a B.P. Climatological questions associated with the data are whether the transition to the Holocene occurred synchronously with climatic transitions in other locations and whether there exist solar influences on the variations in monsoon rainfall (indicated by δ^{18}O variations, low δ^{18}O reflecting strong monsoon). (Data from Fleitmann et al. 2003.)

1.4 Spacing

Archives other than documentary collections or climate models require measurements on the archive material. Material-size requirements lead in many cases to a constant length interval, $L(i)$, from which material for one measurement is taken, and also the length spacing, $l(i)$, between the measurement mid-points on the length axis is often constant (Fig. 1.13). Dating transfers from length into the time domain with the "sample duration," $D(i)$, and the temporal spacing, $d(i) = t(i) - t(i-1)$, here in this book briefly denoted as "spacing." The spacing is frequently nonconstant: archives normally accumulate not at a constant rate. They might also be subject to postdepositional length distortions such as compressing in the case of ice cores. Archives that allow pre-sampling (visual)

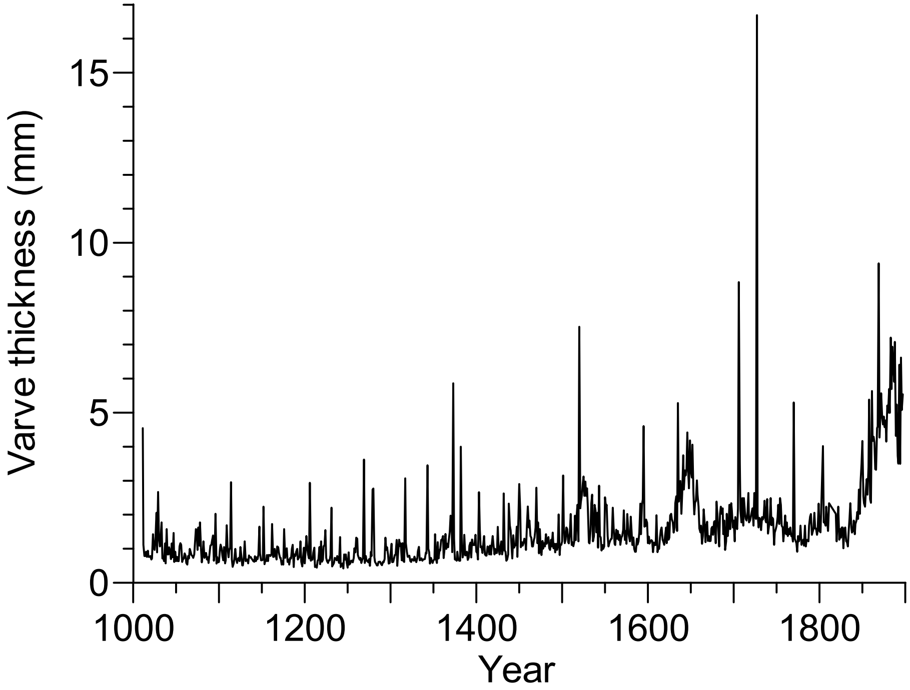


Figure 1.8. Lake sediment core data: varve thickness record from Lower Mystic Lake (Boston area) over the past 1000 years. Multiple overlapping cores were retrieved from the lake, split and photographed in the laboratory. The sediments consist of varves of alternating siliciclastic (bright) and biogenic (dark) layers. The total combined length of the records is about 2 m. Sediment blocks extracted from cores were embedded in epoxy resin to produce petrographic thin sections and X-ray densitometry slabs. A master, composite sequence of stratigraphy was constructed from high-resolution imagery of observations made via petrographic microscopy, back scattered electron microscopy and X-ray densitometry (Besonen 2006). Age control from varve counting was confirmed by means of radiocarbon dating on terrestrial macrofossils. In addition to varve thickness, Besonen (2006) determined the dates of graded beds based on visual examination of the petrographic thin sections and X-ray imagery. A thick varve and a graded bed can be jointly used as a proxy for hurricane activity in the area of the lake. Hurricane-strength precipitation saturates the watershed, results in erosive overland flow that entrains sediment and carries it into the lake where it is deposited as a graded bed. This is enhanced by hurricane-strength winds that disturb vegetation and uproot trees, exposing loose sediment (Besonen 2006). The proxy information was verified by means of pollen data and documentary information (available from about 1630 to the present). The time series ($n = 877$) covers the interval from A.D. 1011 to 1897, minor hiatuses are present (1720–1721, 1803, 1812–1818), also a major above the depth corresponding to 1897. The record bears information on hurricane activity in the Boston area over the past 1000 years. (Data from Besonen et al. 2008.)

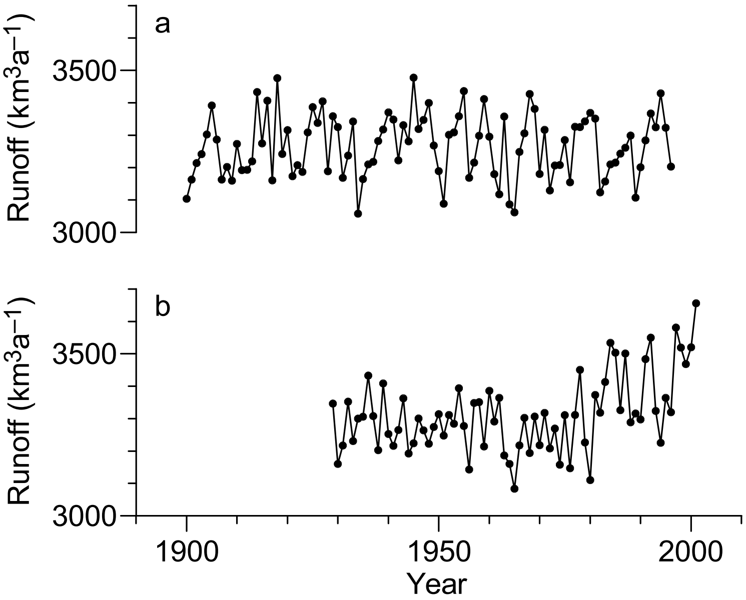


Figure 1.9. Climate model data: runoff from Arctic rivers. **a** Natural forcing only; **b** combined anthropogenic and natural forcing. In a climate model, the physical equations for energy, momentum and mass conservation are numerically solved in time steps over a spatial grid. HadCM3 (Gordon et al. 2000) is a coupled Atmosphere–Ocean General Circulation Model (AOGCM) for the global domain, run by the Hadley Centre for Climate Prediction and Research, Bracknell, United Kingdom. The atmospheric component has a horizontal grid spacing of 2.5° in latitude by 3.75° in longitude and 19 vertical levels. The oceanic component has 20 vertical levels on a 1.25° by 1.25° grid. The time step used for integrating the differential equations representing the atmospheric component was 30 min, for the oceanic component one hour. The total interval simulated (∼ 140 years) was longer than the data shown (**a** 1900–1996; **b** 1929–2001). Plotted are annual-mean ensemble averages, for which the model year starts on 1 December. The averages were constructed from four ensemble runs, that is, runs with identical forcings but different initial conditions. The initial conditions used were taken from states separated by 100 years in a HadCM3 run, in which external forcings where set to have no year-to-year variations ("control run"). Unlike previous models, HadCM3 does not require flux adjustments of heat and water at the air–sea interface to maintain a stable climate without drift behaviour (Johns et al. 1997; Stott et al. 2000). This makes the results obtained with HadCM3 more reliable than previous results. The natural forcing included changes in the amount of stratospheric aerosols stemming from volcanic eruptions and variations in solar irradiation. The anthropogenic forcing included changes in atmospheric concentrations of CO_2, methane, sulfate aerosols and ozone. The river runoff records were generated (Wu et al. 2005) by summing the precipitation contributions from affected grid cells north of 65°N. Model simulations can be used to analyse past and forecast future climate changes. Questions associated with the data are those after the size and the timing of changes in runoff as a result of an intensified hydrological cycle caused by anthropogenically induced warming. (Data from Wu et al. 2005.)

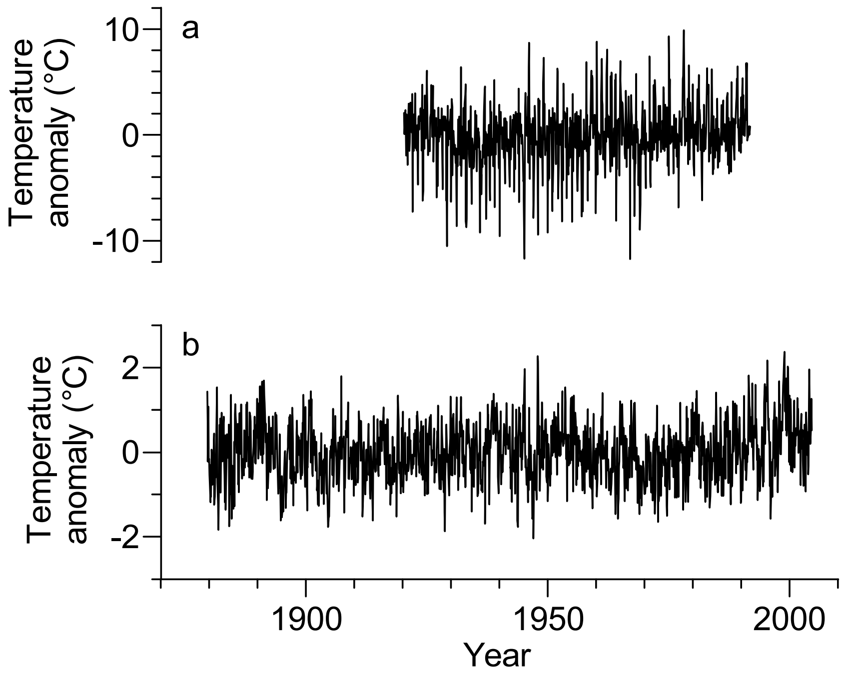


Figure 1.10. Measured data: surface air temperature records from Siberia (**a**) and North Atlantic (**b**). Data are monthly temperature anomalies with respect to the 1961–1990 means from a gridded, global set. Siberia is presented by the grid cell 50–55°N, 90–95°E, effectively reflecting station Krasnojarsk; the North Atlantic by 35–40°N, 25–30°W. Shown are the gap-free time intervals (**a** May 1920 to November 1991, $n = 859$; **b** July 1879 to July 2004, $n = 1501$). The annual cycles were removed by subtracting the monthly averages. (Raw data from Jones and Moberg 2003.)

detection of time-equidistant sampling points, such as tree-rings, varves (that is, annually laminated sediments) or speleothems (Fig. 1.14), appear to be the exception rather than the rule. That mixture of deterministic and stochastic influences on the spacing, is pictured in Fig. 1.15. The Elbe floods (Fig. 1.1) are an example where $d(i)$ (or equivalently $t(i)$) is the major research object, not $x(i)$, see Chapter 6.

The nonzero sample duration, $D(i)$, imposed by material requirements, can be subject to extension to $D'(i)$ by diffusion-like processes in the archive (Fig. 1.13). Besides physical diffusion of material, for example in ice cores, bioturbation in sedimentary archives (mixing by activities of worms and other animals in the upper (young) layer) can play a role. The other data archives studied here (Table 1.1) likely have no diffusion effects.

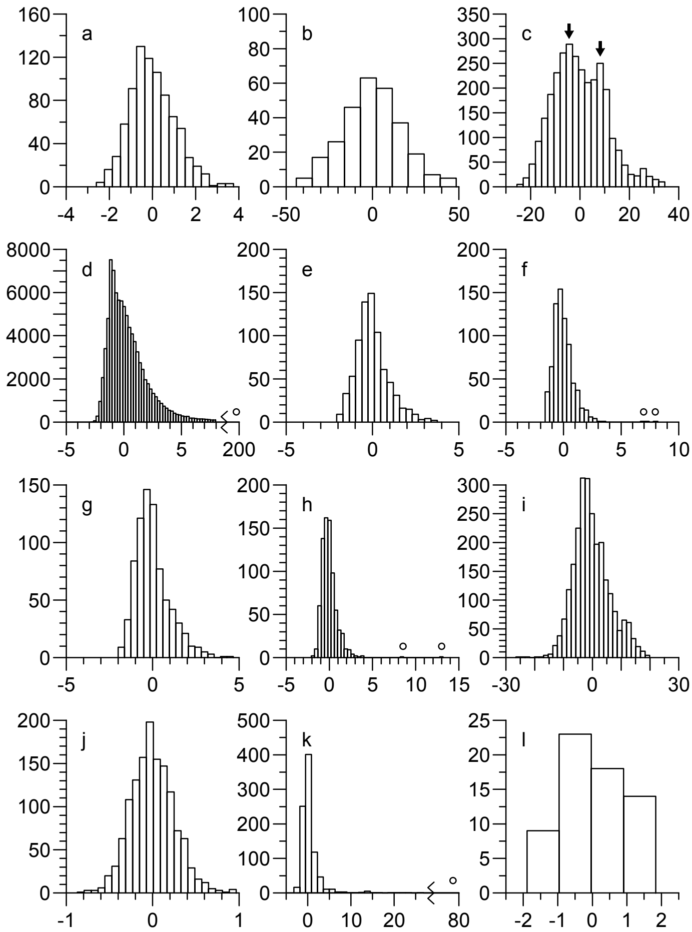


Figure 1.11. Statistical noise distributions of selected climate time series. **a** ODP 846 δ^{18}O; **b** Vostok CO_2; **c** Vostok δD; **d** NGRIP SO_4; **e** NGRIP Ca; **f** NGRIP dust content; **g** NGRIP electrical conductivity; **h** NGRIP Na; **i** tree-ring Δ^{14}C; **j** Q5 δ^{18}O; **k** Lower Mystic Lake varve thickness; **l** HadCM3 runoff. The distributions are estimated with histograms. Data and units are given in Figs. 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8 and 1.9. In **a** and **e–h,** the trend component was estimated (and removed prior to histogram calculation) using a ramp regression model (Figs. 4.6 and 4.7); in **b** and **c** using a harmonic filter (Section 5.2.4.3); in **d** and **k** using the running median (Figs. 4.16 and 4.17); in **i** using nonparametric regression (Fig. 4.14); in **j** using a combination of a ramp model in the early and a sinusoidal in the late part (Fig. 4.18); and in **l** using the break regression model (Fig. 4.12). Outliers are tentatively marked with *open circles* (note broken axes in **d, k**). In **c,** the modes of the suspected bimodal distribution are marked with *arrows.* In **a, e–h** and **j,** time-dependent variability, $S(T)$, was estimated using a ramp regression model (Chapter 4); in **d** and **k** using the running MAD (Figs. 4.16 and 4.17); and in **l** using a linear model. Normalizing (dividing by $S(T)$) for those time series was carried out prior to histogram calculation. The other time series assume constant $S(T)$, values are given in Table 1.3.

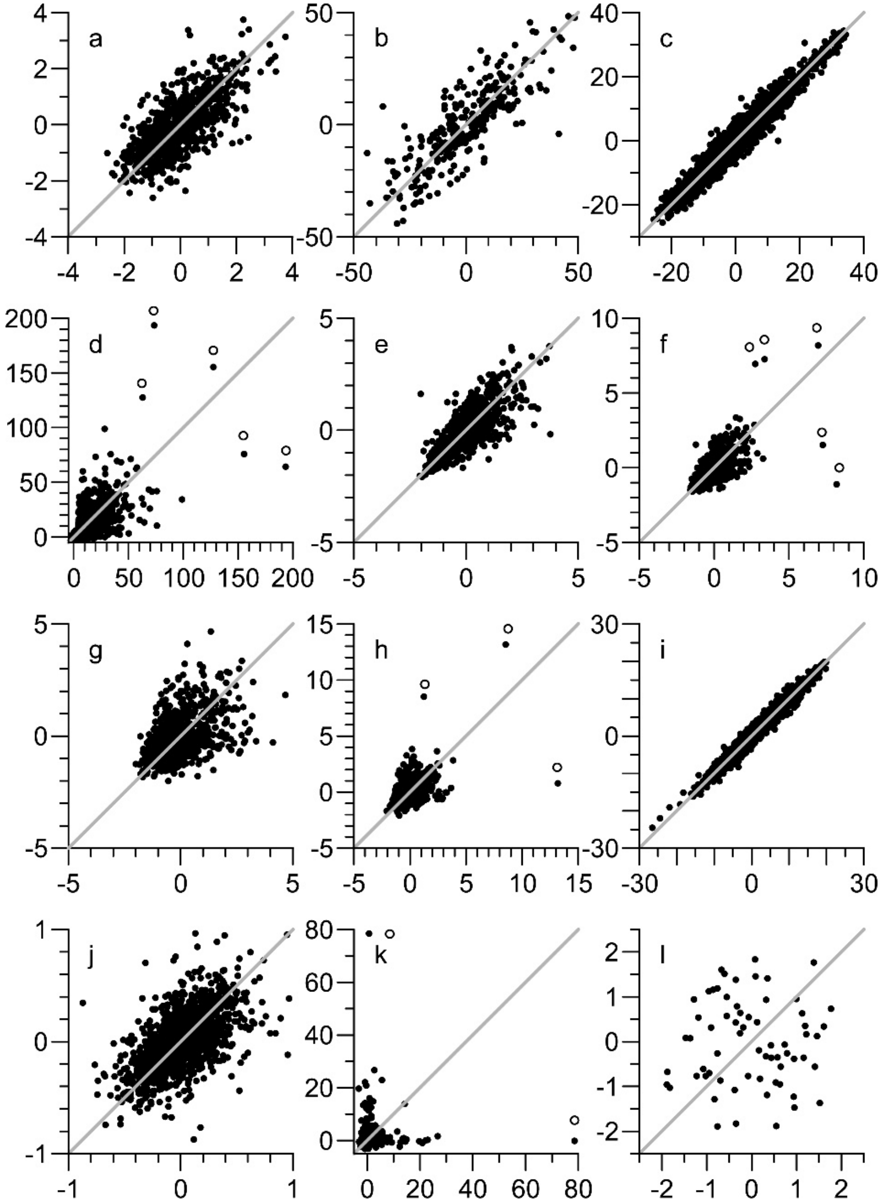


Figure 1.12. Persistence of noise in selected climate time series. **a** ODP 846 δ^{18}O; **b** Vostok CO_2; **c** Vostok δD; **d** NGRIP SO_4; **e** NGRIP Ca; **f** NGRIP dust content; **g** NGRIP electrical conductivity; **h** NGRIP Na; **i** tree-ring Δ^{14}C; **j** Q5 δ^{18}O; **k** Lower Mystic Lake varve thickness; **l** HadCM3 runoff. Noise data are shown as lag-1 scatterplots (in each panel, $x(i-1)$ is plotted on the ordinate against $x(i)$ on the abscissa as *points*), together with 1:1 lines (*grey*). Data and units are given in Figs. 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8 and 1.9. Detrending and $S(T)$ normalization prior to analysis was carried out as in Fig. 1.11. Note that in **d,** all points are shown (unlike as in Fig. 1.11d). Outliers are tentatively marked with *open circles*.

Table 1.3. Measurement and proxy errors in selected climate time series (Table 1.2).

Archive	*Variable*	*Error range*		
		Total, S(T)	*Measurement*	*Proxy*
Marine core	δ^{18}O	0.2–0.3‰[a]	0.06‰[b]	$\sim 1/3$[c]
Ice core	CO_2 content	17.5 ppmv[a]	2–3 ppmv[d]	Small[e]
	δD	10.5‰[a]	$\leq$ 1‰[d]	7‰[f]
	SO_4 content	40.5 ppbw[g]	10%[h]	Unknown[i]
	Ca content	43 ppb[j]	10%[h]	Unknown[i]
	Dust content	$0.56 \cdot 10^5$ ml^{-1}[j]	10%[h]	Unknown[i]
	Conductivity	0.37 µS cm^{-1}[j]	10%[h]	Unknown[i]
	Na	28 ppb[j]	10%[h]	Unknown[i]
Tree-rings	Δ^{14}C	6.2‰[a]	$\sim$ 2‰[k]	Small[l]
Speleothem	δ^{18}O	0.25‰[a]	0.08‰[m]	Unknown[n]
Lake core	Varve thickness	0.33 mm[g]	0.1 mm[o]	NA[p]
Climate model	River runoff	93 km^{3}a^{-1}[q]	0	NA
Direct measure-	Temperature	0.69°C[r]	0.03°C[s]	0
ment		2.97°C[t]	0.03°C[s]	0

Measurement errors were usually determined using repeated measurements. Proxy errors refer to the climate variables in Table 1.2 unless otherwise noted. NA, not applicable.
[a] Standard deviation of detrended $\{t(i), x(i)\}_{i=1}^{n}$ (Fig. 1.11).
[b] Shackleton et al. (1995b).
[c] As ice-volume indicator, relative error. This error comes from other variations than of ice volume: mainly of bottom water temperature and to a lesser degree of salinity (Mudelsee and Raymo 2005).
[d] Petit et al. (1999).
[e] Raynaud et al. (1993).
[f] As air-temperature indicator; own determination of $MS_E^{1/2}$ (Eq. 4.8) after Jouzel et al. (2007: Fig. S4 therein).
[g] Average MAD value (Figs. 4.16 and 4.17), divided by 0.6745 (a standard normal distribution has an MAD of $\sim$ 0.6745).
[h] Relative error (Röthlisberger et al. 2000).
[i] Trace substances are part of a complex system, involving variations at the source, during transport (wind) and at deposition; they represent a more local or regional climate signal.
[j] Time-average of $\widehat{S}(i)$ (Fig. 4.7).
[k] Reimer et al. (2004).
[l] Δ^{14}C in tree-rings on yearly to decadal resolution has a (small) proxy error as atmospheric Δ^{14}C indicator because the wood formation is not constant (the major portion is formed in spring and early summer) and because tree-ring thickness varies (Stuiver et al. 1998). Δ^{14}C variations are a good proxy of solar activity variations because other influences (variations in ocean circulation, changes in the intensity of the Earth's magnetic field) are weak on Holocene timescales (Solanki et al. 2004).
[m] Fleitmann et al. (2003).
[n] Unknown on longer timescales (Table 1.2) because observed monsoon rainfall time series (Parthasarathy et al. 1994) are too short (150 a) to permit comparison.
[o] Time-average; depends on varve distinctiveness and human component (Besonen MR 2010, personal communication).
[p] Only information about hurricane existence sought, not about hurricane strength.
[q] Time-average of $\widehat{S}(i)$ (Fig. 4.12).
[r] North Atlantic, time-average.
[s] Upper limit (Tetzlaff G 2006, personal communication).
[t] Siberia, time-average.

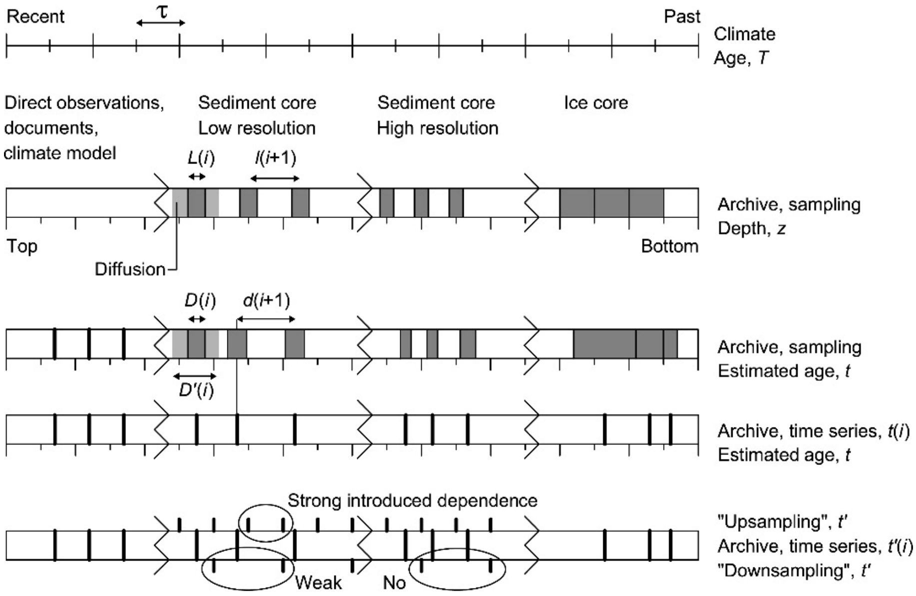


Figure 1.13. Sampling of time series from climate archives. The archive, documenting climate over a time span, is sampled (depth domain), dated (time domain) and possibly interpolated to an evenly spaced time grid. τ denotes a typical timescale of climatic fluctuations, $X_{\text{noise}}(T)$. $L(i)$, length over which material is sampled (*dark shading*); $l(i)$, length spacing between mid-points of $L(i)$; $D(i)$, time-domain analogue of $L(i)$; $d(i)$, time-domain analogue of $l(i)$, denoted as "spacing." *Light shading* indicates effects of a diffusion-like process, that is, extension of $D(i)$ to $D'(i)$. Diffusion need not act symmetrically. *Thick vertical lines* indicate $t(i)$. Terms "sediment core", "ice core", etc. denote here the sampling type rather than a specific archive (for example, a speleothem is often sampled like a "sediment core"). In case of ice cores, $t(i)$ often is not the average time but the time at the upper end of the sample. Real ice cores contain two sub-archives, ice material and enclosed air bubbles, with different age–depth relations (Chapter 8). Interpolation to a fine grid ("upsampling") introduces a strong additional dependence in addition to climatic dependence; "downsampling" introduces weak or no additional dependence. High-resolution time series ($d(i)$ small) have the advantage that this effect is weaker than for low-resolution records. (Note that our usage of "grid" is not restricted to two dimensions.)

The sampled time series $\{t(i), x(i)\}_{i=1}^{n}$ carries information about observed climatic variations up to an upper bound equal to the record length and down to a lower bound of

$$\max\left(\tau, D'(i), \bar{d}\,\right), \tag{1.4}$$

where $\bar{d}$ is the average of $d(i)$. Whereas the upper bound is obvious, the lower bound is explained as follows. The "persistence time," τ, of the climatic noise measures the decay of the autocorrelation function ("memory loss") of $X_{\text{noise}}(T)$, see Chapter 2. Deterministic influences acting on shorter timescales are by definition (Eq. 1.1) not part of the description. Information within interval $D'(i)$ is lost by the sampling process and eventual diffusion. Information theory shows that for evenly spaced time series ($d(i) = d = \text{const.}$) the lower limit is $2\ d$ (or one over Nyquist frequency). The factor 2 is omitted in Eq. (1.4) because for uneven spacing the bound may be lower than for even spacing (Chapter 5).

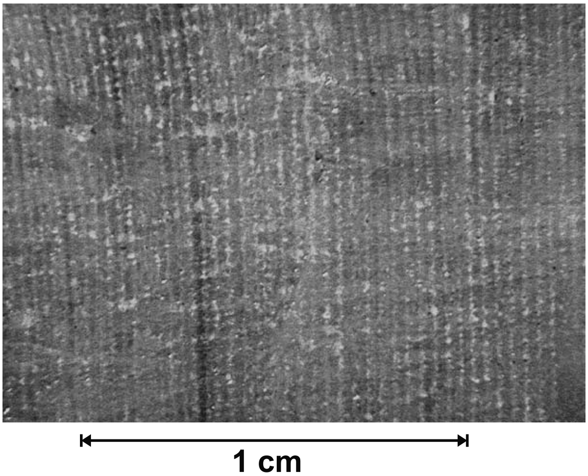


Figure 1.14. Plain-light photomicrograph from a polished section of stalagmite S3 from southern Oman. U/Th dating of samples and the seasonally varying monsoon precipitation pattern in the geographic region suggest that the laminae are annual. Dark (bright) layers reflect a higher (lower) density of pores and fluid inclusions (Fleitmann 2001). The stalagmite covers the period from approximately A.D. 1215 to 1996. Annual layer thickness and oxygen isotopic (δ^{18}O) composition, measured on the stalagmite, record variations in the intensity of Indian monsoonal rainfall. (From Burns et al. (2002), with permission from the publisher.)

Interpolation of the unevenly spaced time series $\{t(i), x(i)\}_{i=1}^{n}$ is in climatology usually done to obtain an evenly spaced series $\{t'(i), x'(i)\}_{i=1}^{n'}$. This series can then be analysed with sophisticated statistical methods for which currently only implementations exist that require even spacing. This advantage, however, is accompanied by following disadvantages. First, additional serial dependence can be introduced, depending mainly on n'. If $n' > n$ ("upsampling"), that effect is strong (Fig. 1.13).

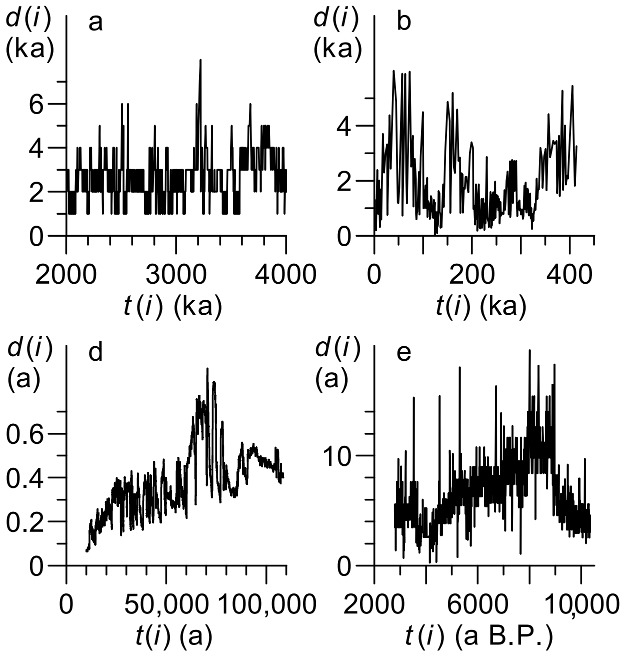


Figure 1.15. Spacing of selected climate time series. **a** ODP 846 δ^{18}O; **b** Vostok CO_2; **c** Vostok δD; **d** NGRIP SO_4; **e** Q5 δ^{18}O. Data are given in Figs. 1.2, 1.3, 1.4 and 1.7. In **d,** $d(i)$ is shown for the $D(i) = 0.5$ cm data; the time series with $t(i) = 1\,$a (Fig. 1.4) is obtained from the 0.5-cm data using "downsampling." The ice core data (**b–d**) reflect to some degree the effects of ice compaction, that means, $d(i)$ increases with $t(i)$. The Q5 speleothem spacing time series (**e**) suggests visually a strong negative correlation with the speleothem δ^{18}O series (Fig. 1.7). This is explained as follows. Low (high) δ^{18}O means strong (weak) Indian monsoonal rainfall, this in turn faster (slower) movement of the rainwater through the soil, weaker (stronger) uptake of soil-CO_2, lower (higher) pH of the water, reduced (enhanced) solution of soil-carbonate, less (more) material for calcite precipitation, small (large) annual stalagmite layers and, finally, a higher (lower) temporal spacing because the depth spacing is nearly constant (Fig. 1.7). Note that at places with other soil properties, the relation δ^{18}O–spacing may be different (Burns et al. 2002). The values of the average spacing, $\bar{d}$, and the coefficient of variation of spacing, CV_d, which is defined as the standard deviation of the spacing divided by $\bar{d}$, are as follows. **a** $\bar{d} = 2.40\,$a, $CV_d = 0.41$; **b** $\bar{d} = 1.46\,$a, $CV_d = 0.82$; **c** $\bar{d} = 0.13\,$a, $CV_d = 0.85$; **d** $\bar{d} = 0.32\,$a, $CV_d = 0.47$; **e** $\bar{d} = 5.62\,$a, $CV_d = 0.49$.

If $n' \approx n$ it is weaker, and only for $n' < n$ ("downsampling") it may be absent. That means, interpolation does not allow to go in resolution below the limit set by Eq. (1.4). Second, depending on the type of interpolation method (linear, cubic spline, etc.), $x'(i)$ may show serious deviations from $x(i)$ in terms of variability or noise properties. That is, interpolation takes us a step further away from the observed process.

Achieving insight into shorter-term climatic processes through sampling an archive is best done by increasing the resolution. Reducing $d(i)$ might require reducing $D(i)$ by employing a measurement method that consumes less material. However, the restriction imposed by diffusion processes and climatic persistence still applies (Eq. 1.4). "Overlapped sampling," $d(i) < D(i)$, is no means to obtain higher resolved information than with $d(i) \geq D(i)$.

1.5 Aim and structure of this book

We have certain hypotheses about time-dependent climate processes, about $X_{\text{trend}}(T)$, $X_{\text{out}}(T)$, $S(T)$ and $X_{\text{noise}}(T)$, which we would like to test. Alternatively, we wish to estimate parameters of climate processes. For that purpose, we use certain methods that take uncertainty into account, that means, statistical methods. Smaller error bars or narrower confidence intervals for the results obtained with the methods, guarantee better testability or more accurate knowledge. To construct confidence intervals, in principle, two approaches exist: classical and bootstrap. The classical approach makes substantial assumptions, such as normally distributed data, no serial dependence, and, often, even time spacing, whereas the bootstrap approach does not make such. Since the preceding sections showed that the assumptions made by the classical approach may be violated when applied to climate time series analysis, the bootstrap may yield more reliable results.

That does not imply that all results obtained on climate time series using classical methods are invalid. However, caution as regards their statistical accuracy is appropriate. The reasons why the classical approach was used are obvious. First, the bootstrap was invented late (Efron 1979), but it soon became accepted in the statistical community and recognized/accepted in the natural sciences (Casella 2003). Bootstrap methods for time series (serially dependence) lag one decade behind in their development. Second, there has been an increase in computing power, which made bootstrap calculations feasible.

This book presents the bootstrap approach adapted to a number of statistical analysis methods that have been found useful for analysing climate time series at least by the author. Linear and nonlinear regression (Chapter 4), spectral analysis (Chapter 5) and extreme value time series analysis (Chapter 6) are explained in case of univariate climate time series analysis (Part II). Correlation (Chapter 7) as well as lagged and other variants of regression (Chapter 8) come from the field of bivariate time series (Part III). Application of each method is illustrated with one or more climate time series, several of which already presented. A section ("Background material") reports alternative techniques and provides a look at the literature that is intended to serve climatologists who wish to learn about the statistical basics of the method, as well as statisticians who wish to learn about the relevant climatology encountered. While both lists cannot be exhaustive, this is more the aim for the also given literature where the bootstrap approach to a statistical method has been used in climatology and related fields as, for example, ecology. A further section ("Technical issues") informs about details such as numerical accuracy and software implementations, it gives also internet references where the computer programs implementing the method can be obtained.

Some topics are not covered in this book. Extension to tri- and higher dimensional multivariate time series seems to be straightforward. Methods from dynamical systems theory, attempting to describe climate as a low-dimensional chaotic system, are likely too demanding in terms of data size (Section 1.6). Also other methods that require even spacing are not dealt with but briefly reviewed in Section 1.6.

However, before starting with adaptions of the bootstrap approach to statistical methods in climatology we need to review bootstrap methodology for time series in some detail, which is done in Chapter 3. Necessary statistical concepts such as confidence intervals or standard errors are also explained. One bootstrap variant ("parametric bootstrap") employed in this book assumes a statistical model of the climatological persistence (Chapter 2). These chapters complete Part I.

Sceptics among the readers may ask whether or not the bootstrap approach brings indeed more reliable results than the classical approach. Therefore you will find throughout the book comparisons between both approaches. These are based on Monte Carlo simulations, that means, artificial time series with pre-defined attributes, for which the true result is known a priori. In the same way, different bootstrap variants are also compared with each other. Finally, the (adverse) effects of interpolation are also explored by means of Monte Carlo simulations.

The final part (IV) of the book is an outlook on future directions in climate time series analysis with the bootstrap. Chapter 9 outlines climate archive modelling to take into account timescale uncertainties and includes "normal" extensions to novel estimation problems and higher dimensions. We also look on paradigm changes that may result from a strong increase in computing power in the future and influence the way how we model the climate and analyse climate time series.

1.6 Background material

The **prologue** is a translation from Popper (1935: p. 78 therein). Other relevant books on quantification and philosophy of science are predominantly written by physicists: Einstein (1949), Heisenberg (1969), Lakatos and Musgrave (1970), von Weizsäcker (1985) and Sokal and Bricmont (1998).

As **statistics texts,** accessible to non-statisticians, describing the various roads to probability and estimation, may serve Priestley (1981: Chapters 1–3 therein), Fine (1983), Davison (2003) and Wasserman (2004). The Bayesian road (Lindley 1965; Spall 1988; Bernardo and Smith 1994; Bernardo et al. 2003) seems not to be so well followed in geosciences, but this may change in future. Davis (1986) is a text book written by a geologist; Wilks (2006) and von Storch and Zwiers (1999) were written by climatologists. The latter three contain parts on time series analysis. As text books on time series analysis, accessible to non-statisticians, the following can be used: Priestley (1981), Diggle (1990), Brockwell and Davis (1996) and Shumway and Stoffer (2006); the latter work includes software examples in the R computing environment. A further book on time series analysis is by Anderson (1971). The only book devoted to time series analysis of unevenly spaced data seems to be Parzen (1984); an early review is by Marquardt and Acuff (1982); there is a thesis (Martin 1998) from the field of signal processing. We finally mention the Encyclopedia of statistical sciences (Kotz et al. 1982a,b, 1983a,b, 1985a,b, 1986, 1988a,b, 1989, 1997, 1998, 1999).

Climatology text books: The reports by Working Group I of the Intergovernmental Panel on Climate Change (IPCC–WG I) (Houghton et al. 2001; Solomon et al. 2007) are useful on weather (that is, meteorology) and short-term climate. Paleoclimate, covering longer-term processes in, say, the Holocene (last $\sim$ 10,000 years) and before, is described by Crowley and North (1991), Bradley (1999) and Cronin (2010). We finally mention the Encyclopedia of Atmospheric Sciences (Holton et al. 2003), the Encyclopedia of Earth System Science (Nierenberg 1992), the Encyclopedia of Geology (Selley et al. 2005), the Glossary of Geology (Neuendorf et al. 2005), the Handbook of Hydrology (Maidment 1993) and the Encyclopedia of Ocean Sciences (Steele et al. 2001).

The form of **decomposition** in Eq. (1.1) of a process into trend, outliers, variability and noise is non-standard. Outliers are often considered as gross errors in the data that only have to be removed. However, in climatology, outliers may bear information on extreme events and can also be the object of analysis (Chapter 6). The notion of systematic behaviour of a trend leaves space for interpretation of what can be included. Certainly worth so are nonlinear trends to account for climatic changes (Chapter 4). Also incorporated are harmonic signals like the daily or annual cycle, which can be recorded in climate archives. Since the focus here in this book is on longer-term processes, we omit to write an own seasonal signal into Eq. (1.1); such an approach is common in econometrics (Box et al. 1994). Other, longer-term cyclic influences on climate are also astronomical in origin, such as variations in solar activity or Milankovitch variations in Earth orbital parameters. However, since their imprint in the climate system as regards amplitude, phase and frequency, is not precisely known (and also sometimes debated), these signals are investigated in this book by analysing the spectral properties of the noise process (Chapter 5).

Detailed accounts of **climate archives** give the following. Usage of marine sediment cores is a standard method (has been applied over decades), see Kennett (1982), Seibold and Berger (1993) and the series of reports on and results of scientific drilling into the ocean floor (Deep Sea Drilling Project 1969–1986; Ocean Drilling Program 1986–2004, 1988–2007). Ice cores (Oeschger and Langway 1989; Hammer et al. 1997) and lake sediment cores (Negendank and Zolitschka 1993; Zolitschka 1999) are likewise regularly employed. Usefulness of speleothems (Baker et al. 1993; Gillieson 1996; Daoxian and Cheng 2002; Fairchild et al. 2007) is recognized since the 1990s. Dendroclimatology has a long tradition (Douglass 1919, 1928, 1936; Schweingruber 1988). Analysis of documentary climate data is described by Pfister (1999), Brázdil et al. (2005) and Glaser (2001). Construction and use of climate models is a growing field, see McGuffie and Henderson-Sellers (1997) or Randall et al. (2007). From this book's data analysis view, climate modelling is similar to probing and measuring a natural climate archive.

An upper limit to the **time range** over which climate can be studied is set by the age of Earth (~ 4.6 Ga). The course of the evolution of Earth, including its climate, division and subdivision into different geological epochs, is described by Stanley (1989). A geological timescale refers to a chronology of events (first or last appearance of species, reversals of Earth's magnetic field, climatic, etc.) which is updated as new data and new datings become available. Currently used are: Gradstein et al. (2004) covering the whole time range, Cande and Kent (1992, 1995) going back before the Cenozoic (last ~ 65 Ma) into the late Cretaceous, Berggren et al. (1995b) for the Cenozoic and Berggren et al. (1995a) for the last 6 Ma. (Note the various meanings of "timescale" in geosciences.)

Absolute dating methods almost entirely use one of the many clocks provided by natural radioactive elements. A comprehensive treatise is Geyh and Schleicher (1990), see also Walker (2005). K/Ar dating (Dalrymple and Lanphere 1969) utilizes the decay of ^{40}K. The potassium isotope has a half-life, $T_{1/2}$, of 1.266 Ga (Section 8.7), it decays into ^{40}Ar with a chance of $\sim 11\%$ and ^{40}Ca ($\sim 89\%$). One measures ^{40}K and also the amount of ^{40}Ar that accumulated in a sample since argon was removed by a process whose age is to be determined. Such a zeroing process can be a volcanic eruption, which produced the rock sample. The natural decay chains in uranium and thorium provide a wealth of clocks, running on a wide range of timescales (Ivanovich and Harmon 1992). U/Th dating utilizes the decays of ^{234}U to ^{230}Th ($T_{1/2} \approx 245\,\mathrm{ka}$) and ^{230}Th to ^{226}Ra ($T_{1/2} \approx 76\,\mathrm{ka}$). Since speleothems contain essentially no thorium at the time of formation, dating means measuring the amount of accumulated thorium since that time. ^{210}Pb dating (Appleby and Oldfield 1992) takes the decay chain of ^{210}Pb ($T_{1/2} \approx 22.3\,\mathrm{a}$) to ^{206}Pb. Radiocarbon dating (Taylor 1987) employs the decay of ^{14}C to ^{14}N ($T_{1/2} \approx 5730\,\mathrm{a}$). $T_{1/2}$ determines the limits for a reliable age determination. For ages below, say, $\sim 0.1 \cdot T_{1/2}$ and above $\sim 10 \cdot T_{1/2}$, the uncertainties introduced at the determination of the amounts of parent or daughter products become likely too large. Using modern mass spectrometers, this range can be somewhat widened. Besides measurement uncertainties and those owing to imperfectly known half-lifes, another error source is bias that occurs when assumptions, such as complete zeroing or absent sample contamination, are violated. In fact, eliminating measurement bias is often the major task in absolute dating. Using an archive as a dosimeter for dating (Table 1.1) means to measure the dose (effect) a sample has received over time exposed to a dose-rate (effect per time interval). One example is electron-spin-resonance dating, where the effect consists in the number of trapped electrons (for example in carbonate material in a sediment core) and the dose-rate is from natural radioactivity (Grün 1989); the other is cosmic-ray-exposure dating, where one of the effects used regards the number of ^{10}Be atoms transported to an archive from the atmosphere, where cosmic rays had produced them (Gosse and Phillips 2001). Another absolute dating method is counting of yearly layers, either of tree-rings or growth layers in a stalagmite (Fig. 1.14). The assumption that layers present a constant time interval is crucial. Documentary data contain together with the variable usually also the date (which is susceptible to reporting errors).

Relative dating methods rely on an assumed relation between the measured series in the depth domain, $\{z(i), x(i)\}_{i=1}^{n_X}$, and another, dated time series, $\{t_Y(j), y(j)\}_{j=1}^{n_Y}$. If the relation between X and Y is simple (linear, no lag), $t_Y(j)$ can be projected onto $t_X(i)$ rather easily. If it is more complex, a mathematical model may have to be used. Climatologists denote that procedure as correlation or "tuning." As illustration we note that besides the GT4 timescale for the Vostok ice core (Fig. 1.3), two tuned timescales exist. One uses as $x(i)$ Vostok $\delta^{18}O$ in air bubbles and as $y(i)$ the precession of Earth's orbit (Shackleton 2000); the other uses as $x(i)$ Vostok methane content in air and as $y(i)$ mid-July insolation at 30°N (Ruddiman and Raymo 2003). One critical point with relative dating is how well the assumed relation holds. Bayesian approaches to timescale construction were developed by Agrinier et al. (1999) for a geomagnetic polarity record from the Cretaceous–Cenozoic and by Blaauw and Christen (2005) for a Holocene archive in form of a peat-bog core. Section 4.4 gives more details and references on the approaches.

Most before mentioned textbooks on climate and climate archives contain also information on **proxy variables** and how well those indicate climate. Other sources are Broecker and Peng (1982) and Henderson (2002). $\delta^{18}O$ in shells of marine living foraminifera (Fig. 1.2) was in the beginning seen as a "paleothermometer" (Emiliani 1955) until Shackleton (1967) showed that the major recorded climate variable is global ice volume, although he partly withdraw later from this position (Shackleton 2000). The main idea is that polar ice is isotopically light (low $\delta^{18}O$) and that during an interglacial (warm) more of that is as water in the ocean, where foraminifera build their calcareous, $\delta^{18}O$-light shells. Stacks of ice volume records, such as that from the "Spectral Analysis, Mapping, and Prediction" (SPECMAP) project (Imbrie et al. 1984), going back nearly 800 ka, and that of Shackleton et al. (1995b), extending into the Miocene (before $\sim$ 5.2 Ma), were produced and a nomenclature (Prell et al. 1986) of marine isotope stages (MISs) erected. A recently constructed Plio- to Pleistocene $\delta^{18}O$ stack is by Lisiecki and Raymo (2005). Atmospheric CO_2 is rather accurately reflected by CO_2 in air bubbles from Antarctic ice cores (Fig. 1.3), mainly because CO_2 mixes well in the atmosphere (Raynaud et al. 1993). The currently longest record comes from the European Project for Ice Coring in Antarctica (EPICA), Dome C site, the core covering the past $\sim$ 800 ka (Section 8.6.1). For earlier times, other proxies for atmospheric CO_2 have to be used, such as the size and spatial density of stomata in fossil leaves (Kürschner et al. 1996), resulting in significantly larger proxy errors. δD variations in polar ice (Fig. 1.3) reflect variations in air temperature as this variable determines how enriched the precipitation becomes during its net transport from the mid-latitudes to the poles (Rayleigh destillation) (Dansgaard and Oeschger 1989). As regards the various proxy variables from the NGRIP ice core (Figs. 1.4 and 1.5), see the captions and references given therein. Radiocarbon (Fig. 1.6) is produced in the upper atmosphere via reactions with cosmogenic neutrons; the cosmic-ray flux is modulated by the Sun's activity through the solar wind. Another influence that can be seen using Δ^{14}C is variations in the exchange between the oceanic carbon storage and the atmosphere, see Beer et al. (1994) and Cini Castagnoli and Provenzale (1997). Pollen records and their proxy quality are explained by Moore et al. (1991) and Traverse (2007). The proxy quality of δ^{18}O in speleothems from the Arabian peninsula as indicator of monsoon rainfall is largely based on Rayleigh destillation processes (Fleitmann et al. 2004, 2007a).

Ergodicity. Detrended and normalized $x(i)$ were used for analysing the distributional shape for the process $X_{\text{noise}}(T)$ (Fig. 1.11). That is, instead of an ensemble of different realizations at a particular time, one realization was taken at different times. A process for which this replacement gives same results is called ergodic. Since in climatological practice no repeated experiment can be carried out, except with climate models, ergodicity has to be added to the set of made assumptions in this book.

Density estimation. The histograms in Fig. 1.11 were constructed using a bin width equal to $3.49\, s_{n-1}\, n^{-1/3}$ (Scott 1979), where s_{n-1} is the sample standard deviation. More elaborated approaches to density estimations use kernel functions (Silverman 1986; Simonoff 1996; Wasserman 2006). Applications of density estimation to climatology have been made occasional. They include analyses of the Pleistocene ice age (Matteucci 1990; Mudelsee and Stattegger 1997) and of the recent planetary-scale atmospheric circulation (Hansen and Sutera 1986). Standard references on statistical properties of distributions are Johnson et al. (1994, 1995) on continuous univariate and Kotz et al. (2000) on continuous multivariate distributions. Random variables that are composed of products or ratios of other random variables have since long successfully defied analytical derivation of their PDF. Only very simple forms, like $Z = X^2 + Y^2$ with Gaussian X and Y, which has a chi-squared density (right-skewed), can be solved. See Haldane (1942) or Lomnicki (1967) for other cases.

Bioturbation in deep-sea sediments acts as a low-pass filter (Eq. 1.4) (Goreau 1980; Dalfes et al. 1984; Pestiaux and Berger 1984). However, since the accumulated sediment passes the bioturbation zone (the upper few tens of cm of sediment) unidirectionally, signal processing techniques, termed "deconvolution," have been successfully developed to use that information to improve the construction of the timescale (Schiffelbein 1984, 1985; Trauth 1998). An example demonstrating what effects have to anticipated when sampling natural climate archives such as sediment cores is given by Thomson et al. (1995), who found offsets of ~ 1.1 ka between ages of large ($> 150\,\mu$m diameter) foraminifera and fine bulk carbonate at same depth in a core. The most likely explanation is a size-dependent bioturbation that preferentially transports fine material downwards because that is cheaper in terms of energy.

Inhomogeneities in time series owing to systematic changes in the observation system (i.e., the archive) may arise in manifold ways. It is evidently of importance to detect and correct for these effects. A simple case is a sudden change, such as when the time at which daily temperature is recorded, is shifted. This type can be detected using methods (Basseville and Nikiforov 1993) that search for an abrupt change in the mean, $X_{\text{trend}}(T)$. Inhomogeneities in the form of gradual changes in mean, or variability, may be analysed using regression techniques (Chapter 4). Quality assessment of climate data deals predominantly with types and sizes of inhomogeneities (Peterson et al. 1998a,b). Inhomogeneities in the form of periodic changes of the observation system can influence the estimated spectral properties (Chapter 5).

Physics' **nonlinear dynamical systems theory** has developed time series analysis techniques (Abarbanel et al. 1993; Kantz and Schreiber 1997; Diks 1999; Chan and Tong 2001; Tsonis and Elsner 2007; Donner and Barbosa 2008) that can be applied to study, for example, the question whether the climatic variability sampled by $\{t(i), x(i)\}_{i=1}^{n}$ is the product of low-dimensional chaos. A positive answer would have serious consequences for the construction of climate models because only a handful of independent climate variables had to be incorporated; and also the degree of climate predictability would be precisely known (Lyapunov exponents). Although it was meteorology that boosted development of dynamical systems theory by constructing a simplified atmosphere model (Lorenz 1963), we will not pursue related time series analysis methods for two reasons. First, for most applications in climatology the data sizes are not sufficient to allow reasonably accurate conclusions. For example, Nicolis and Nicolis (1984) analysed one late Pleistocene (last ~ 900 ka) δ^{18}O time series (cf. Fig. 1.2) and found a "climatic attractor" with dimensionality ~ 3.1, meaning that four variables could explain the ice age. Grassberger (1986), and later Ruelle (1990), convincingly refuted that claim, which was based on a data size of a few hundred instead of several thousand necessary (Eckmann and Ruelle 1992). Later, Mudelsee and Stattegger (1994) analysed the longest Plio-/Pleistocene δ^{18}O records then available. They found no low-dimensional attractor and could only conclude that at least five variables are acting. Since one assumption for such analyses is that the proxy quality of the measured variable (δ^{18}O, indicating ice volume) holds over all timescales sampled, the limits owing to the sampling process (Eq. 1.4) and the proxy quality (Table 1.2) effectively prohibit exploration of low-dimensional climatic chaos—not to mention the amount of measurements required. Lorenz (1991) considered that decoupled climatic subsystems with low dimensionality could be found. Second, nonlinear dynamical systems methods reconstruct the physical phase space by the method of delay-time coordinates (Packard et al. 1980). Instead of using multivariate time series $\{t(i), x(i), y(i), z(i), \ldots\}_{i=1}^{n}$ (forming the data matrix), this method takes $\{t(i), x(i), x(i+L), x(i+2L), \ldots\}_{i=1}^{n'}$, with $n' < n$ and L (integer) appropriately selected. The delay-time method requires equidistance. For many climate time series encountered in practice, this would mean interpolation, which this book does not advocate (Section 1.4).

Even time spacing is also required for current implementations of two other analysis techniques. The first, Singular Spectrum Analysis or SSA (Broomhead and King 1986), also uses delay-time coordinates explained in the preceding paragraph to reconstruct the data matrix from one univariate time series. The eigenvectors associated with the largest eigenvalues yield the SSA decomposition of the time series into trend and other more variable portions. There exists a successful approach based on computer simulations to assess the significance of eigenvalues in the presence of persistence, which has been applied to observed equidistant temperature time series (Allen and Smith 1994). Again, because for many real-world paleoclimatic time series interpolation would have to be performed, we do not include SSA here. Note that similar to SSA is Principal Component Analysis (PCA), also termed Empirical Orthogonal Function (EOF) analysis, which does the same as SSA on multivariate time series. PCA is a standard method to search for patterns in high-dimensional meteorological time series such as pressure and temperature fields (Preisendorfer 1988; von Storch and Zwiers 1999). The second time series analysis method that requires even spacing and is often applied in climatology, is wavelet analysis, which composes a time series using "wave packets," localized in time and frequency. Percival and Walden (2000) is a textbook accessible to non-statisticians. Applications to climatology include Fligge et al. (1999), who analyse sunspot time series (Fig. 2.12), and Torrence and Compo (1998), who analyse time series of the El Niño–Southern Oscillation (ENSO) climatic mode. (El Niño is defined by sea-surface temperature anomalies in the eastern tropical Pacific, while the Southern Oscillation Index is a measure of the atmospheric circulation response in the Pacific–Indian Ocean region.) It might well be possible to develop adaptions of phase-space reconstruction and nonlinear dynamical systems analysis, SSA, PCA and wavelet analysis to explore unevenly spaced time series directly, circumventing adverse effects of interpolation—at the moment, such adaptions seem not to be available (but see Section 5.3 as regards wavelets).

Chapter 2

Persistence Models

Climatic noise often exhibits persistence (Section 1.3). Chapter 3 presents bootstrap methods as resampling techniques aimed at providing realistic confidence intervals or error bars for the various estimation problems treated in the subsequent chapters. The bootstrap works with artificially produced (by means of a random number generator) resamples of the noise process. Accurate bootstrap results need therefore the resamples to preserve the persistence of $X_{\text{noise}}(i)$. To achieve this requires a model of the noise process or a quantification of the size of the dependence. Model fits to the noise data inform about the "memory" of the climate fluctuations, the span of the persistence. The fitted models and their estimated parameters can then be directly used for the bootstrap resampling procedure.

It turns out that for climate time series with discrete times and uneven spacing, the class of persistence models with a unique correspondence to continuous-time models is rather limited. This "embedding" is necessary because it guarantees that our persistence description has a foundation on physics. The first-order autoregressive or AR(1) process has this desirable property.

2.1 First-order autoregressive model

The AR(1) process is a simple persistence model, where a realization of the noise process, $X_{\text{noise}}(i)$, depends on just the value at one time step earlier, $X_{\text{noise}}(i-1)$. We analyse even and uneven spacing separately.

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_2,

2.1.1 Even spacing

In Eq. (1.2) we let the time increase with constant spacing $d(i) = d > 0$ and write the discrete-time Gaussian AR(1) noise model,

$$\begin{aligned} X_{\text{noise}}(1) &= \mathcal{E}_{\text{N}(0,\,1)}(1), \\ X_{\text{noise}}(i) &= a \cdot X_{\text{noise}}(i-1) + \mathcal{E}_{\text{N}(0,\,1-a^2)}(i), \qquad i = 2, \ldots, n. \end{aligned} \tag{2.1}$$

Herein, $-1 < a < 1$ is a constant and $\mathcal{E}_{\text{N}(\mu,\,\sigma^2)}(\cdot)$ is a Gaussian random process with mean μ, variance σ^2 and no serial dependence, that means, $E\left[\mathcal{E}_{\text{N}(\mu,\,\sigma^2)}(i) \cdot \mathcal{E}_{\text{N}(\mu,\,\sigma^2)}(j)\right] = 0$ for $i \neq j$. It readily follows that $X_{\text{noise}}(i)$ has zero mean and unity variance, as assumed in our decomposition (Eq. 1.2). Figure 2.1 shows a realization of an AR(1) process.

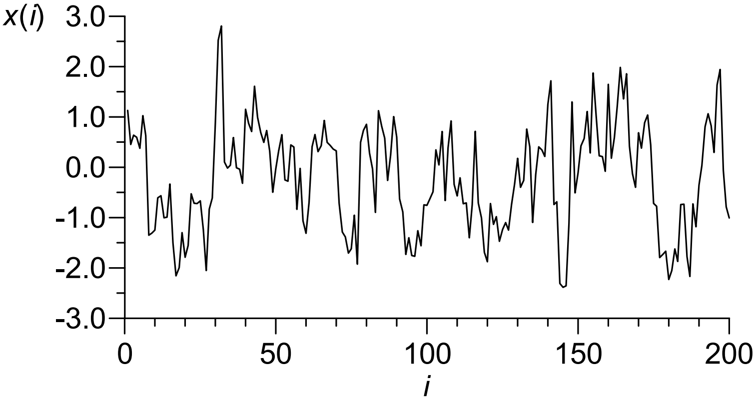


Figure 2.1. Realization of an AR(1) process (Eq. 2.1); $n = 200$ and $a = 0.7$.

The autocorrelation function,

$$\begin{aligned} \rho(h) &= \frac{E\Big[\big\{X_{\text{noise}}(i+h) - E\big[X_{\text{noise}}(i+h)\big]\big\} \cdot \big\{X_{\text{noise}}(i) - E\big[X_{\text{noise}}(i)\big]\big\}\Big]}{\Big\{VAR\big[X_{\text{noise}}(i+h)\big] \cdot VAR\big[X_{\text{noise}}(i)\big]\Big\}^{1/2}} \\ &= E\Big[X_{\text{noise}}(i+h) \cdot X_{\text{noise}}(i)\Big], \end{aligned} \tag{2.2}$$

where h is the time lag, E is the expectation operator and VAR is the variance operator, is given by (Priestley 1981: Section 3.5 therein)

$$\rho(h) = a^{|h|}, \qquad h = 0, \pm 1, \pm 2, \ldots. \tag{2.3}$$

For $a > 0$, this behaviour may be referred to as "exponentially decreasing memory" (Fig. 2.2).

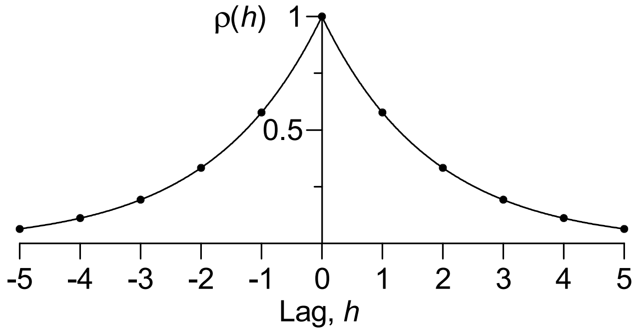


Figure 2.2. Autocorrelation function of the AR(1) process, $a > 0$. In the case of even spacing (Section 2.1.1) $\rho(h)$ is given by $a^{|h|} = \exp\left[-|h| \cdot d/\tau\right]$, in the case of uneven spacing (Section 2.1.2) by $\exp\left[-|T(i+h) - T(i)|/\tau\right]$. In both cases, the decrease is exponential with decay constant τ.

Note that the assumptions in Eq. (1.2), namely $E\left[X_{\text{noise}}(i)\right] = 0$ and $VAR\left[X_{\text{noise}}(i)\right] = 1$, required the formulation of the AR(1) model as in Eq. (2.1), which is non-standard. See Section 2.6 for the standard formulation.

Persistence estimation for the AR(1) model means estimation of the autocorrelation parameter, a. To illustrate autocorrelation estimation, assume that from the time series data, $\{x(i)\}_{i=1}^{n}$, the outliers have been removed and the trend and variability properties (Eq. 1.2) determined and used (as in Fig. 1.11) to extract $\{x_{\text{noise}}(i)\}_{i=1}^{n}$, realizations of the noise process. An estimator of the autocorrelation parameter, that means, a recipe how to calculate a from $\{x_{\text{noise}}(i)\}_{i=1}^{n}$, is given by

$$\widehat{a} = \sum_{i=2}^{n} x_{\text{noise}}(i) \cdot x_{\text{noise}}(i-1) \Bigg/ \sum_{i=2}^{n} x_{\text{noise}}(i)^2. \tag{2.4}$$

(Chapter 3 introduces estimators and the "hat notation.") Note that estimator $\widehat{a}$ is biased, that means, if $\{X_{\text{noise}}(i)\}$ is an AR(1) process with parameter a, then $E\left(\widehat{a}\right) \neq a$. Only approximation formulas exist for the bias in general autocorrelation estimation. Such formulas can be used for bias correction. Similarly, also the estimation variance, $VAR\left(\widehat{a}\right)$, is only approximately known. In general, bias and variance decrease with n. The background material (Section 2.6) gives various bias and variance formulas, informs about bias correction and lists other autocorrelation estimators.

The suitability of the AR(1) model can be assessed using the estimation residuals,

$$\epsilon(i) = x_{\text{noise}}(i) - \widehat{a} \cdot x_{\text{noise}}(i-1), \qquad i = 2, \ldots, n. \tag{2.5}$$

As realizations of a standard normal random process, the residuals should not exhibit patterns in the lag-1 scatterplot (Fig. 1.12).

2.1.1.1 Effective data size

Persistence ($a > 0$) means a reduced information content of a time series compared to a situation without positive serial dependence. In a statistical estimation, more data have then to be available to achieve a confidence interval (Chapter 3) of same width. An effective data size, n', can be defined for estimators of parameters of processes with persistence via the estimation variance. Consider the mean estimator, $\bar{X} = \sum_{i=1}^{n} X(i)/n$, and the AR(1) process Eq. (2.1) for two cases: $a > 0$ and $a = 0$. Then

$$VAR\left(\bar{X}\right) = VAR[X(i)]\,/n'_{\mu} \qquad\qquad (a > 0)$$

(the index refers to mean estimation) is set equal to

$$VAR\left(\bar{X}\right) = VAR[X(i)]\,/n \qquad\qquad (a = 0).$$

Bayley and Hammersley (1946) show that

$$n'_{\mu} = n\left[1 + 2\sum_{i=1}^{n-1}(1 - i/n)\,\rho(i)\right]^{-1}, \tag{2.6}$$

which can for the AR(1) process with the autocorrelation given in Eq. (2.3) be readily solved using the geometric series as well as the arithmetic-geometric series:

$$n'_{\mu} = n\left\{1 + \frac{2}{n}\frac{1}{1-a}\left[a\left(n - \frac{1}{1-a}\right) - a^n\left(1 - \frac{1}{1-a}\right)\right]\right\}^{-1}. \tag{2.7}$$

von Storch and Zwiers (1999: Section 17.1 therein) define a related quantity, the decorrelation time as

$$\tau_D = \lim_{n\to\infty} \frac{n}{n'_{\mu}}. \tag{2.8}$$

An AR(1) process thus has $\tau_D = (1 + a)/(1 - a)$.

Even for moderate values of n ($\gtrapprox 50$) and a ($\lessapprox 0.5$), the influence of persistence on n'_{μ} can be considerable (Section 2.6). Eq. (2.7) is valid only for the mean estimator. Because the definition of n' depends of the type of estimation (von Storch and Zwiers 1999), such formulas have limited practical relevance. Section 2.6 gives n' for variance and correlation estimation.

2.1.2 Uneven spacing

In Eq. (1.2), we let the time increase with an uneven spacing $d(i) > 0$ and write the discrete-time Gaussian AR(1) noise model,

$$\begin{aligned} X_{\text{noise}}(1) &= \mathcal{E}_{\text{N}(0,\,1)}(1), \\ X_{\text{noise}}(i) &= \exp\left\{-\left[T(i) - T(i-1)\right]/\tau\right\} \cdot X_{\text{noise}}(i-1) \\ &\quad + \mathcal{E}_{\text{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau\})}(i), \qquad i = 2, \ldots, n. \end{aligned} \tag{2.9}$$

The "loss of memory" increases with the time difference scaled by the persistence time, τ (Fig. 2.2). The random innovation, $\mathcal{E}(\cdot)$, is now heteroscedastic instead of homoscedastic as in the case of even spacing. It follows that this noise model for uneven spacing has zero mean, unity variance and autocorrelation

$$E\left[X_{\text{noise}}(i+h) \cdot X_{\text{noise}}(i)\right] = \exp\left[-|T(i+h) - T(i)|/\tau\right]. \tag{2.10}$$

Estimation of the persistence time using noise data $\{t(i), x_{\text{noise}}(i)\}_{i=1}^{n}$ is more complex than in the case of even spacing. A least-squares estimation uses the sum of squares,

$$S(\widetilde{\tau}) = \sum_{i=2}^{n} \left[x_{\text{noise}}(i) - \exp\left\{-\left[t(i) - t(i-1)\right]/\widetilde{\tau}\right\} \cdot x_{\text{noise}}(i-1)\right]^2, \tag{2.11}$$

and takes the minimizer as τ estimator, $\widehat{\tau} = \operatorname{argmin}\left[S(\widetilde{\tau})\right]$. The minimization has to be carried out numerically (Section 2.7). In the case of equidistance, $t(i) - t(i-1) = d \;\forall i$, the least-squares estimator corresponds to the estimator given in Eq. (2.4), with $\widehat{a} = \exp\left(-d/\widehat{\tau}\right)$.

The bias in the estimation of τ for unevenly spaced data seems to defy an analytical derivation. Figure 2.3 shows the bias studied by means of Monte Carlo simulations. The simulations demonstrate that the bias is similar to the value in a situation with even spacing.

Suitability of the AR(1) model for uneven spacing can be assessed using the residuals,

$$\begin{aligned} \epsilon(i) = x_{\text{noise}}(i) - \exp\left\{-\left[t(i) - t(i-1)\right]/\widehat{\tau}\right\} \cdot x_{\text{noise}}(i-1), \\ i = 2, \ldots, n. \end{aligned} \tag{2.12}$$

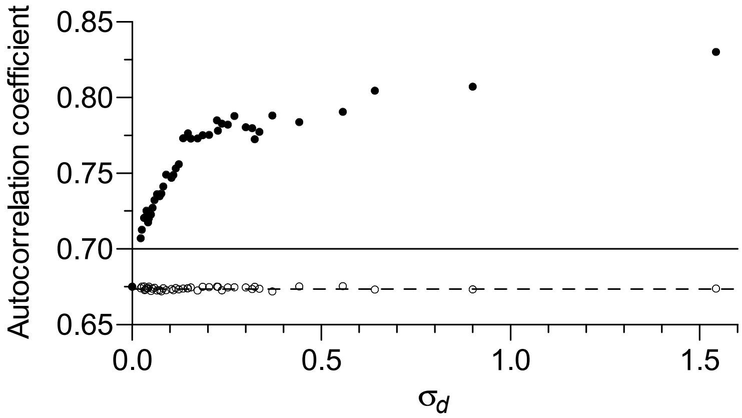


Figure 2.3. Monte Carlo study of the bias in the autocorrelation estimation of an AR(1) process, known mean, uneven spacing. Time series were generated after Eq. (2.9) with $n = 50$ and $\tau = -1/\ln(0.7) \approx 2.804$ by means of a random number generator (Section 2.7). The start was set to $T(1) = t(1) = 1$; the spacing, $d(i)$, was drawn from a gamma distribution with a pre-defined order parameter (Section 2.7) and subsequently scaled such that $t(n) = 50$ or $\bar{d} = 1$. The "equivalent autocorrelation coefficient" is $\bar{a} = \exp(-\bar{d}/\tau) = 0.7$. The standard deviation of the spacing, σ_d, was used as a measure of the unevenness. For each time grid (σ_d fixed), a number ($n_{\mathrm{sim}} =$ 10,000) of time series were generated and $\widehat{\tau}$ determined after Eq. (2.11). Shown (*open symbols*) is the average of the quantity $\exp(-\bar{d}/\widehat{\tau})$ over the simulations. Also shown (*filled symbols*) is the average of the estimator $\widehat{a}$ (Eq. 2.4) applied to the linearly interpolated, equidistant time series (same start and end, same data size). (The standard error ($\sim 1/\sqrt{n_{\mathrm{sim}}}$) of the estimation averages is smaller than the symbol size.) The true autocorrelation value (*solid line*) is underestimated by the $\widehat{\tau}$ estimator. This negative bias is excellently described by the bias approximation (*dashed line*) of White (1961) from the case of even spacing. The interpolation, on the other hand, leads to serious overestimation. This effect is owing to the serial dependence introduced by the interpolation (Fig. 1.13), it increases with σ_d. No formulas to correct for this bias exist.

2.1.2.1 Embedding in continuous time

In continuous time, the AR(1) noise model is given in "differential notation" by

$$dX_{\mathrm{noise}}(T) = a \cdot X_{\mathrm{noise}}(T)\, dT + dW(T), \tag{2.13}$$

where a is the autocorrelation parameter, dT is a time increment and $dW(T)$ is an innovation term of the Wiener process (also called Brownian motion), $W(T)$. As the discrete-time model, the continuous-time AR(1) model has an exponentially decaying autocorrelation function (Priestley 1981).

Let us consider the continuous-time noise model to be sampled at discrete times, which may be unevenly spaced, resulting in the discrete-time model, $\{T(i), X_{\text{noise}}(i)\}_{i=1}^{n}$. Robinson (1977) showed that for $a > 0$, this resulting model equals the discrete-time AR(1) model given in Eq. (2.9). The discrete-time AR(1) model is said to be embedded in continuous time, it determines uniquely the underlying continuous-time AR(1) model given in Eq. (2.13). This is an important property of the AR(1) model because the embedding allows a foundation on physics, which works in continuous time (differential equations).

2.2 Second-order autoregressive model

We assume even spacing and write the discrete-time Gaussian second-order autoregressive or AR(2) noise model,

$$X_{\text{noise}}(i) = a_1 \cdot X_{\text{noise}}(i-1) + a_2 \cdot X_{\text{noise}}(i-2) + \mathcal{E}(i), \tag{2.14}$$

where $\mathcal{E}(i)$ is a stationary purely random process (no serial dependence). This AR(2) process is not strictly stationary but, conditional on a_1 and a_2, only asymptotically stationary (its moments approach saturation with $i \to \infty$); see Section 2.6. The regions of asymptotic stationarity in the a_1–a_2 plane are shown in Fig. 2.4. The behaviour of the autocorrelation function of the AR(2) process depends on where a_1 and a_2 lie. For $a_2 \geq -a_1^2/4$, $a_1 > 0$ and $a_2 < 0$, $\rho(h)$ decays smoothly to zero and for $a_2 \geq -a_1^2/4$, $a_1 < 0$ and $a_2 < 0$, $\rho(h)$ alternates its sign as it decays. In the connection with spectral analysis (Chapter 5), the case $a_2 < -a_1^2/4$ is interesting, because then $\rho(h)$ shows besides a decay a quasi-cyclical behaviour with a period of

$$T = 2\pi/\theta, \tag{2.15}$$

where $\cos(\theta) = a_1/(2\sqrt{-a_2})$. Figure 2.5 shows a realization of the process in Eq. (2.14).

Commonly used estimators of the AR(2) model include the Yule–Walker estimators,

$$\begin{aligned}\widehat{a_1} &= \widehat{\rho}(1) \cdot [1 - \widehat{\rho}(2)] \,/\left[1 - \widehat{\rho}(1)^2\right], \\ \widehat{a_2} &= \left[\widehat{\rho}(2) - \widehat{\rho}(1)^2\right] /\left[1 - \widehat{\rho}(1)^2\right],\end{aligned} \tag{2.16}$$

where

$$\widehat{\rho}(h) = \widehat{R}(h) \Big/ \widehat{R}(0), \qquad h = 1, 2, \tag{2.17}$$

is the autocorrelation estimator and

$$\widehat{R}(h) = \sum_{i=h+1}^{n} x_{\text{noise}}(i) \cdot x_{\text{noise}}(i-h), \qquad h = 0, 1, 2, \tag{2.18}$$

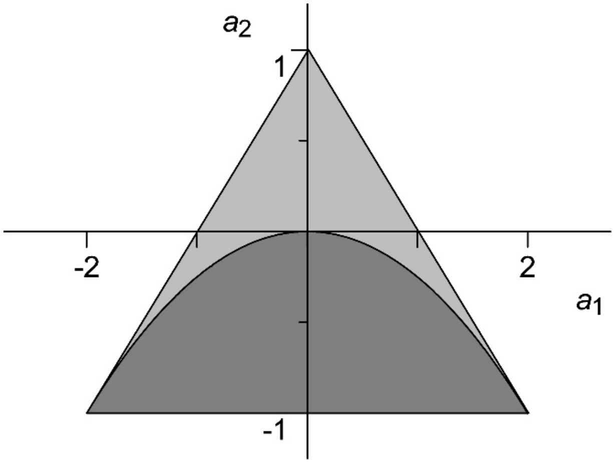


Figure 2.4. Regions of asymptotic stationarity for the AR(2) process (Eq. 2.14) (*shaded*). The region for complex roots (*dark shaded*), which allows quasi-periodic behaviour, lies below the parabolic, $a_2 < -a_1^2/4$.

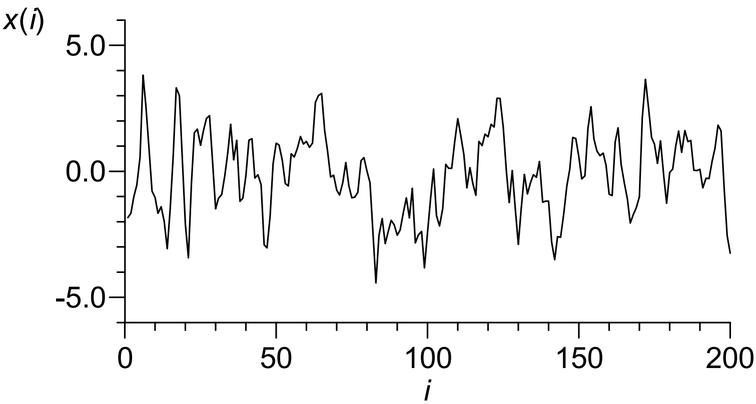


Figure 2.5. Realization of an AR(2) process (Eq. 2.14); $n = 200$, $a_1 = 1.0$, $a_2 = -0.4$ and $\mathcal{E}(i) = \mathcal{E}_{\mathrm{N}(0,\,1)}(i)$. The first 5000 data were discarded to approach asymptotic stationarity. The graph exhibits a quasi-cyclical behaviour with an approximate period of 9.5 time units.

is the autocovariance estimator. Estimation bias occurs also in case of the AR(2) model, approximations were given by Tjøstheim and Paulsen (1983).

2.3 Mixed autoregressive moving average model

We assume even spacing and write the discrete-time Gaussian mixed autoregressive moving average or ARMA(p, q) noise model,

$$\begin{aligned} X_{\text{noise}}(i) = a_1 \cdot X_{\text{noise}}(i-1) + \cdots + a_p \cdot X_{\text{noise}}(i-p) \\ + b_0 \cdot \mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(i) + \cdots + b_q \cdot \mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(i-q), \\ i = \max(p, q) + 1, \ldots, n. \end{aligned} \tag{2.19}$$

Note that the model is given only for a subset of $i = 1, \ldots, n$. Similar conditions as in the preceding sections may be formulated for $\{a_1, \ldots, a_p; b_0, \ldots, b_q\}$ to ensure stationarity. The AR(1) model (Section 2.1), AR(2) model (Section 2.2), the autoregressive model of general order (AR(p))—these are special cases of the ARMA(p, q) model ($q = 0$). The moving average process of general order (MA(q)) arises from the ARMA(p, q) model with $p = 0$; it is not considered further in this book. Estimation techniques for the ARMA(p, q) model are mentioned in the background material (Section 2.6).

As regards the context of this book, the problem with the discrete-time ARMA(p, q) model under uneven time spacing is that no embedding in a continuous-time process can be proven. Indeed, already for a discrete-time, real-valued Gaussian AR(1) process with $a < 0$ it was shown (Chan and Tong 1987) that no embedding in a continuous-time, real-valued Gaussian AR(1) process exists. No embedding of $X_{\text{noise}}(i)$ means no foundation on physics. Suppose, for example, that physical laws governing the climate system to be analysed yield an ARMA(p_1, q_1) continuous-time noise model. Even with a "perfect" estimation (estimation bias and variance both zero), it would then not be possible to determine the model parameters $\{a_1, \ldots, a_{p_1}; b_0, \ldots, b_{q_1}\}$ uniquely from an unevenly spaced sample time series $\{t(i), x_{\text{noise}}(i)\}_{i=1}^{n}$. For climate time series, a perfect time estimation would also be required, which is not usually possible. A further complication arises from "model aliasing." Bartlett (1946) showed that an evenly sampled continuous-time AR(p) process becomes a discrete-time ARMA$(p, p-1)$ process (the "alias"), which has implications already for the AR(2) model (Section 2.2). However, for a certain type of uneven spacing, namely "missing observations" (Fig. 1.15e), where $\{t(i)\}_{i=1}^{n}$ is a subset of $\{t(j)\}_{j=1}^{m}$ with $d(j) = \text{const.}$ and $m - n \ll m$, the embedding problem vanishes and estimation techniques exist (Section 2.6).

The majority of sampled climate time series, at least within this book, exhibits uneven, irregular spacing (Fig. 1.15), for which only the simple AR(1) model ensures the embedding property. Fortunately, this is no serious limitation as climatic theory shows that climatic noise is to a first order of approximation well described by the AR(1) process (Section 2.5).

2.4 Other models

The discrete-time Gaussian ARMA(p, q) process (Eq. 2.19) composes $X_{\text{noise}}(i)$ as a linear combination of past $X_{\text{noise}}(j), j < i$, and innovations, $\mathcal{E}_{\text{N}(0,\,\sigma^2)}(j)$. We briefly review other processes that can be seen as extensions of the ARMA(p, q) process. These processes might provide somewhat more realistic models for $X_{\text{noise}}(i)$. However, usage of many of these models seems to be restricted to evenly spaced time series (perhaps with missing values) because of the embedding problem (Section 2.1.2.1) and lack of statistical theory.

2.4.1 Long-memory processes

The AR(1) process has an exponentially ("fast") decaying autocorrelation function (Fig. 2.2). Also the ARMA(p, q) process has a similar bound (Brockwell and Davis 1991),

$$|\rho(h)| \leq C\, r^{|h|}, \qquad h = 0, \pm 1, \pm 2, \ldots, \tag{2.20}$$

where $C > 0$ and $0 < r < 1$. A long-memory process is a stationary process (loosely speaking, with time-constant statistical properties such as mean and standard deviation) for which

$$\rho(h) \rightarrow C\, h^{2H-1} \qquad \text{as } h \rightarrow \infty, \tag{2.21}$$

where $C \neq 0$ and $H < 0.5$. This decrease is slower than in the case of ARMA(p, q), hence it is said to exhibit long-range serial dependence or long memory.

Examples of long-memory processes are

1. fractional Gaussian noise and

2. fractional autoregressive integrated moving average models, denoted as ARFIMA(p, δ, q).

The relation between the ARFIMA(p, δ, q) and ARMA(p, q) models is as follows. δ defines (Section 2.6) a fractional difference operator, $(1 - B)^{\delta}$, where $|\delta| < 0.5$ and B is the backshift operator. The backshift operator shifts one step back in time, for example, $B\, X_{\text{noise}}(i) = X_{\text{noise}}(i - 1)$. The ARFIMA$(p, \delta, q)$ model is then an ARMA(p, q) model (Eq. 2.19), where $X_{\text{noise}}(j)$ is replaced by $(1 - B)^{\delta}\, X_{\text{noise}}(j)$. For the trivial case $\delta = 0$, the ARFIMA(p, δ, q) model reduces to the ARMA(p, q) model, which already shows the embedding problem (Section 2.3). (One can define a nonstationary ARIMA model by allowing $\delta = 1, 2, \ldots$.) The ARFIMA(p, δ, q) model has $H = \delta$ (Brockwell and Davis 1991). Although continuous-time ARFIMA(p, δ, q) models have been developed (Comte and Renault 1996), the embedding for $\delta \neq 0$ seems not yet to have been analysed.

In the special case of the ARFIMA($0, \delta, 0$) model it has been shown (Hwang 2000) that for uneven spacing the estimation of δ is biased, with the bias depending on the spacing. It appears that in the general case the theory of long-memory processes for unevenly spaced data is not well developed enough to be applied in the context of the present book. Section 2.6 gives more details on estimation of long-memory models for evenly spaced data, while Sections 2.5.2 and 2.5.3 present examples where long- and short-range models are fitted to climate series.

2.4.2 Nonlinear and non-Gaussian models

Stationary nonlinear models allow a richer structure to be given to the noise process, $X_{\text{noise}}(i)$. Of particular interest for climatology is the class of threshold autoregressive models (Tong and Lim 1980). Let the real line $\mathbf{R}$ be partitioned into l non-overlapping, closed segments, $\mathbf{R} = \mathbf{R_1} \cup \mathbf{R_2} \cup \cdots \cup \mathbf{R}_{\boldsymbol{l}}$. The discrete-time Gaussian self-exciting threshold autoregressive process of order $(l; k, \ldots, k)$ or SETAR$(l; k, \ldots, k)$ process, where k is repeated l times, is given by

$$X_{\text{noise}}(i) = a_0^{(m)} + \sum_{j=1}^{k} a_j^{(m)} \cdot X_{\text{noise}}(i-j) + \mathcal{E}_{\mathrm{N}(0,\, \sigma^{2(m)})}(i), \qquad (2.22)$$

conditional on $X_{\text{noise}}(i-j) \in \mathbf{R}_{\boldsymbol{m}}; m = 1, 2, \ldots, l$. As an example, Fig. 2.6 shows a realization of the SETAR$(2; 1, 1)$ process,

$$X_{\text{noise}}(i) = \begin{cases} +2.0 + 0.8 X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,1)}(i) & \text{if } X_{\text{noise}}(i-1) \le 0, \\ -1.0 + 0.4 X_{\text{noise}}(i-1) + \mathcal{E}_{\mathrm{N}(0,\,2)}(i) & \text{if } X_{\text{noise}}(i-1) > 0, \end{cases} \qquad (2.23)$$

which may be a model of random fluctuations between two climate regimes with different mean values and persistence times. Also quasi-cyclical behaviour can be reproduced by threshold models. In practical applications the number of regimes, l, is usually limited to a few. Estimation is carried out iteratively: guess of l, maximum likelihood estimation of parameters, calculation of a goodness-of-fit measure such as AIC or a normalized version (Tong and Yeung 1991), analysis of residuals and autocorrelation functions, improved guess of l, etc. Continuous-time threshold autoregressive models have been formulated (Tong and Yeung 1991) but it seems that the embedding problem for unevenly spaced

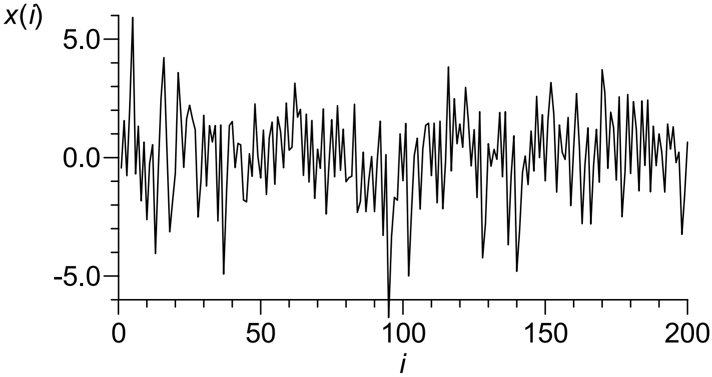


Figure 2.6. Realization of a SETAR(2; 1, 1) process (Eq. 2.23); $n = 200$. (The first 5000 data were discarded.)

time series has not yet been analysed. This would mainly concern the SETAR(2; 1, 1) case.

Many more types of nonlinear persistence models can be perceived (Section 2.6). It may for some climate data even be useful to consider in Eq. (1.2) the process $S(i) \cdot X_{\text{noise}}(i)$ as belonging to the class of stochastic volatility models, for which $S(i)$ depends on past $X_{\text{noise}}(i-j)$, $j > 0$, and/or past $S(i-j)$. This process of time-varying variability could model "burst" phenomena such as earthquakes and serve also as a formulation of the outlier process in Eq. (1.2). One common problem, however, with complex, nonlinear time series models, is the embedding of the discrete-time process in continuous time (Section 2.1.2.1). We have to concede that complex, unevenly observed climatic processes may not be accessible to a meaningful parametric estimation.

Also non-Gaussian random innovation terms can be used to construct ARMA(p, q) models. In such cases, however, formulas for the estimation bias are hardly available. One possibility is to introduce a transformation,

$$X_{\text{noise}}(i) = f\left(X'_{\text{noise}}(i)\right), \tag{2.24}$$

where $X'_{\text{noise}}(i)$ is a Gaussian process, and to infer f from a probability density estimation (Section 1.6) using $\{x_{\text{noise}}(i)\}_{i=1}^{n}$.

2.5 Climate theory

A dynamical view of the climate system gives motivation that climatic persistence may to a first order of approximation be written as an AR(1) process. This was shown by an influential paper entitled "Stochastic climate models" (Hasselmann 1976).

The dynamical view seems to be challenged by a series of papers claiming a “universal power law” in temperature records. This law indicates long memory, not short as the AR(1) models suggests. Long memory of temperature fluctuations over timescales from days to decades should seriously impact the development of climate theory. It would further have enormous practical consequences as weather forecasts for intervals considerably longer than what is currently feasible (a few days) would become principally possible.

The re-analysis of some crucial data here (Section 2.5.2) makes it hard to accept the “universal power law.” It should nevertheless be kept in mind that the AR(1) model need not be a good higher-order description of $X(T)$. However, after allowing nonlinear trends and outlier processes (Eq. 1.2), the AR(1) model is likely not a bad candidate to describe $X_{\text{noise}}(T)$.

2.5.1 Stochastic climate models

The derivation of the AR(1) model of climate persistence is based on three assumptions.

> **Assumption 1** is timescale separability. The climate system as a whole (p. 3) is composed of a slowly varying component (“climate” in original sense), representing oceans, biosphere and cryosphere, and a fast varying component (“weather”), representing the atmosphere.

The differential equations governing the climate evolution may then be written as

$$\frac{dX(T)}{dT} = F(X(T), Y(T)), \qquad \text{timescale } \tau_X, \tag{2.25}$$

$$\frac{dY(T)}{dT} = G(X(T), Y(T)), \qquad \text{timescale } \tau_Y, \tag{2.26}$$

where $\tau_X \gg \tau_Y$, X and Y are the slowly and fast varying components (vectors of possibly high dimension), respectively, and F and G are some nonlinear system operators. τ_Y is of the order of a few days, τ_X of several months to years and more (Hasselmann 1976).

It was previously thought that the influence of Y on X, of weather on climate, could be accounted for by simply averaging,

$$\frac{dX(T)}{dT} \simeq F^*(X(T)), \qquad \text{timescale } \tau_X, \tag{2.27}$$

where the modified climate system operator, F^*, is the time average of $F(X, Y)$.

Since the work of Hasselmann (1976) it is accepted that the weather noise cannot so easily be cancelled out. Consider $0 \leq T \leq \tau_X$. Then

$$\begin{aligned} \frac{dX(T)}{dT} &\simeq F(X(0), Y(T)), \\ &= W(T), \end{aligned} \tag{2.28}$$

where W is a stochastic (Wiener) process. Discretization yields

$$X(T+1) = X(T) + \mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(T). \tag{2.29}$$

Here we have made

Assumption 2. The unknown weather components $Y(T)$ add up to yield after the central limit theorem (Priestley 1981: Section 2.14 therein) a Gaussian purely random process $\mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(T)$.

Now let $T > \tau_X$. Then the time-dependence of $F(X(T), Y(T))$ has to be taken into account. Since the climate system trajectories have to be bounded, we must invoke a negative feedback mechanism. The simplest model for that is given by

Assumption 3. The negative feedback is proportional to the climate variable, $X(T)$, yielding $F(X(T), Y(T)) = -\beta \cdot X(T) + W(T)$.

Assumption 3 makes Eq. (2.29), which is a nonstationary random walk process (Section 2.6), to a stationary AR(1) process,

$$X(T+1) = a \cdot X(T) + \mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(T), \tag{2.30}$$

with $0 \leq a = 1 - \beta < 1$. (Strictly speaking, this is an "asymptotically stationary" AR(1) process, see background material.) This explanation of Hasselmann's (1976) derivation of climate's AR(1) model is from von Storch and Zwiers (1999: Section 10.4.3 therein). Another account of the 1976 paper and its influence on climatology is given by Arnold (2001).

The suitability of the AR(1) noise model depends on how well Assumptions 1, 2 and 3 are fulfilled. Assumption 1 (timescale separability) is generally thought to be well fulfilled. The root cause is that the atmosphere has a much smaller density and heat capacity than most of the rest of the climate system, allowing weather processes to run faster. One caveat may be that some climatically relevant biological processes, such as algae growth (Lovelock and Kump 1994), may act also on short timescales. The validity of Assumption 2 is difficult to prove. It might well be that some weather influences do not add but rather multiply with each other, producing non-Gaussian distributional shape (Section 1.2; Sura et al. (2005)). But Assumption 2 can be relaxed to recognize this, leading to non-Gaussian AR(1) models. Assumption 3 is certainly not exactly fulfilled but may be a good first-order approximation. More sophisticated feedback mechanisms would lead to higher-order ARMA(p, q) or nonlinear models. An interesting case would be a nonlinear dynamical climate system with several local attractors and the occurrence probability within the attracting regions depending on the weather noise (Hasselmann 1999; Arnold 2001). The present knowledge about feedback processes is, however, too limited to permit the theoretical derivation of the precise model form. A further point is that external climate forcing mechanisms (e.g., volcanic eruptions) have to be included for achieving a full set of climate equations. The size of such forcings is currently not well understood (Section 8.4).

The dynamical equations in this section are for the evolution of high-dimensional climate variables, $X(T)$, and not just for the noise part of one variable, $X_{\mathrm{noise}}(T)$. In Eq. (1.1), we composed a climate variable of trend, outliers and noise. By allowing nonlinear trends and outlier processes, effects of violations of the assumptions made above are reduced. This lends credence to the AR(1) noise model.

2.5.2 Long memory of temperature fluctuations?

Peng et al. (1994) introduced Detrended Fluctuation Analysis (DFA) to measure persistence in DNA sequences. Peng et al. (1995) elaborated DFA in more detail and applied it to heartbeat time series. Koscielny-Bunde et al. (1996) introduced DFA to climate time series analysis and found a "universal power law governing atmospheric variability."

DFA uses a time series $\{t(i), x(i)\}_{i=1}^{n}$ with constant spacing $d > 0$ to calculate the so-called profile,

$$y(i) = \sum_{j=1}^{i} x(j), \qquad i = 1, \ldots, n. \tag{2.31}$$

The profile is divided into non-overlapping, contiguous segments of length l (multiple of d), discarding the $\mathrm{mod}(n, l/d)$ last points. The $y(i)$ series is detrended by segment-wise fitting and subtracting polynomials of order 0, 1, 2, etc. Most commonly used are mean and linear detrending. Koscielny-Bunde et al. (1998a,b) studied the influence of different detrending types, also other than polynomial detrending. The fluctuation function, $F(l)$, is the standard deviation of detrended $y(i)$ within a segment, averaged over all segments. $F(l)$ is usually plotted on a double-logarithmic scale because power laws, $F(l) \propto l^{\alpha}$, appear in such plots as a straight line, with slope α.

For a polynomial of order 0 and $x(i)$ from the Gaussian AR(1) model in Eq. (2.1), it readily follows that $F(l) \propto l^{1/2}$. For data with long-range dependence (Eq. 2.21), the power law $F(l) \propto l^{H+1/2}$ results, that means, $\alpha = H + 1/2$ (Talkner and Weber 2000). Thus, DFA can be seen as a method to estimate long-range dependence.

Koscielny-Bunde et al. (1996, 1998a,b) analysed daily temperature series covering typically the past 100 a using DFA and found $\alpha \approx 0.65$ for many records. The claimed universal long-range power-law dependence would have serious theoretical and practical consequences. Govindan et al. (2002) went further and analysed temperature output from a number of AOGCMs, the currently most sophisticated mathematical tools for climate simulation. Since the AOGCMs did not produce a single, universal value but rather a scatter of α values, it was concluded that the models were not able to provide realistic climate forecasts. In particular, the predicted size of global warming would be overestimated. Later, Fraedrich and Blender (2003) used DFA to analyse monthly temperature series. They disputed the existence of the universal power law and suggested the following: $\alpha = 1$ for oceanic data, $\alpha = 0.5$ for inner continental data and $\alpha = 0.65$ for data from the transition regions. This led to an exchange of arguments (Bunde et al. 2004; Fraedrich and Blender 2004), where in particular the results from the temperature record from Siberian station Krasnojarsk were assessed controversially. It appears that in reply to a criticism by Ritson (2004), the originators partly stepped back (Vyushin et al. 2004) from the claimed universality to a position with two memory laws, one for the ocean, the other for the continents.

Here we re-analyse the Krasnojarsk temperature record and also one series of North Atlantic air temperatures to assess whether the power-law exponents are similar or not. It is also asked whether the power laws are actually good models of the serial dependence in the temperature data and whether simple AR(1) models are not already sufficient. A simulation study helps to quantify the uncertainty of the result coming from sampling variations. We use the same gridded raw data set as Fraedrich and Blender (2003); also the technique for removing the annual cycle, the orders of the DFA polynomials (0, 1) and the scaling range of l = 1–15 a for α determination are identical to what these authors employed. Two points may limit comparability of results. First, we restrict ourselves to those time intervals where the monthly series have no gaps. Fraedrich and Blender (2003) took longer series that start in 1900, without explaining how they adapted their methodology to the case of missing data. Second, Fraedrich and Blender (2003) gave the coordinates 30°W, 50°N for the North Atlantic, but the raw data for the grid cells around this point start clearly later than 1900. We take a grid cell from somewhat more south that starts earlier.

The results of the DFA are as follows (Fig. 2.7). The $F(l)$ curves show increases that resemble on first sight a power law. The α estimates in Table 2.1 were determined by fitting the power-law regression model to the $F(l)$ points inside the selected scaling range 1–15 a. Note that fits of a linear regression model to logarithmically transformed l and F values would likely lead to a biased estimation, because Gaussian additive (measurement) noise of $X(i)$ and $F(i)$ would be lost by the transformation (Section 2.6). The resulting values exhibit a variation with the DFA detrending type that is considerably larger than the standard errors for α, indicating systematic estimation errors. The α estimations via DFA deviate also considerably from the values obtained via ARFIMA fits (Table 2.1).

Table 2.1. Result of DFA study (Fig. 2.7), estimated power-law exponents α.

Station (*grid point*)	*DFA*		ARFIMA$(1, \delta, 0)$
	Mean detrending	*Linear detrending*	
Krasnojarsk (50–55°N, 90–95°E)	0.65 ± 0.02	0.79 ± 0.02	0.61
North Atlantic (35–40°N, 25–30°W)	0.58 ± 0.01	0.68 ± 0.01	0.73

DFA errors are standard errors from unweighted least-squares regression (see Chapter 4). ARFIMA models were fitted using Whittle's approximate maximum likelihood technique (Beran 1994: Chapter 5 therein) and α determined via the relation $\alpha = \delta + 1/2$.

A close inspection of the $F(l)$ curves (Fig. 2.7) reveals marked deviations from straight line in the double-logarithmic plots, especially for larger l. Such a behaviour might be referred to as "crossover" (Peng et al. 1995) and different scaling regions with different α values could be further investigated. Philosophy of science, however, says that this is a problematic step because it violates the principle of parsimony. A simpler explanation of the $F(l)$ curves is that the class of power-law models is not ideally suited to describe the data.

The DFA study therefore explores also how the simplest persistence model, AR(1), is suited to describe the data. The AR(1) model (Eq. 2.1) was fitted using the estimator Eq. (2.4) and bias correction (Eq. 2.45), yielding $\widehat{a}' = 0.23$ (Siberia) and $\widehat{a}' = 0.44$ (North Atlantic). For

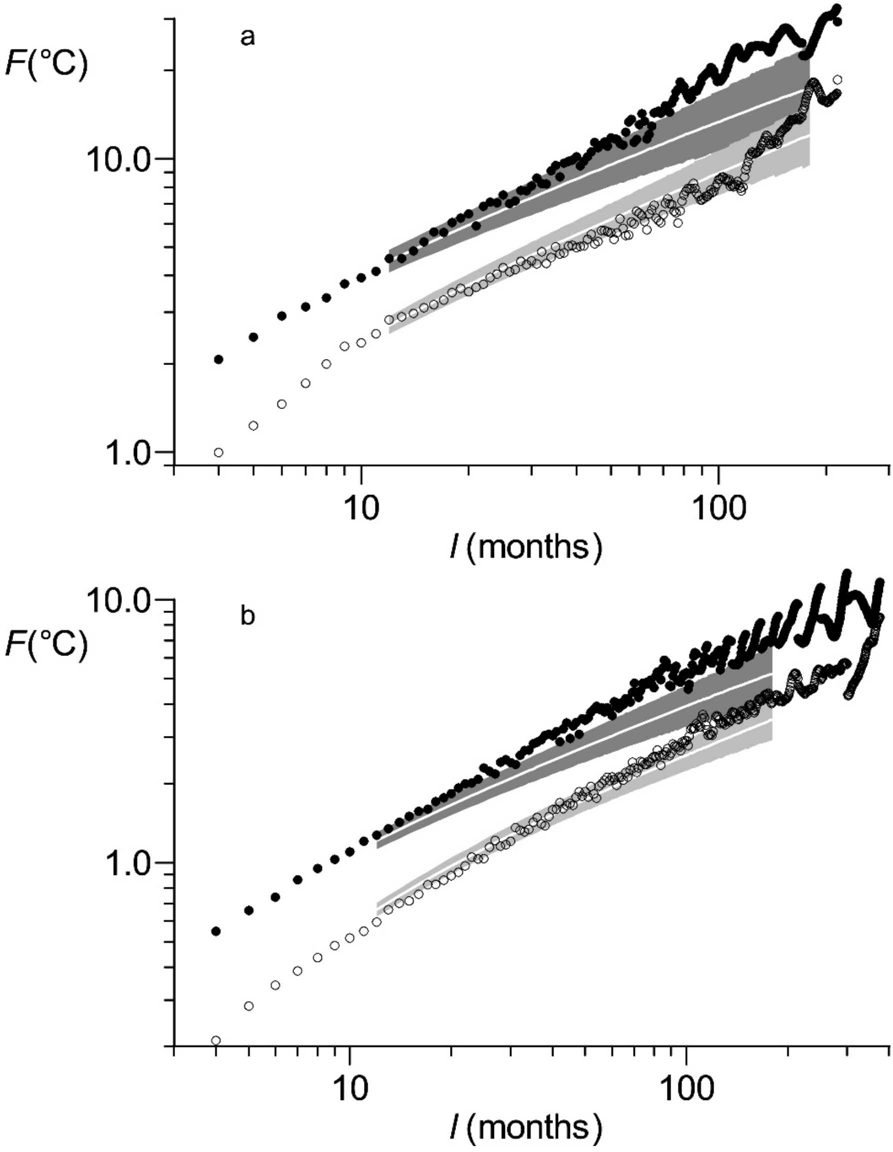


Figure 2.7. Detrended Fluctuation Analysis for temperature records (Fig. 1.10) from Siberia (**a**) and North Atlantic (**b**). Shown as *filled* (*open*) *symbols* are fluctuation functions, F, against segment sizes, l [months] $= 4, \ldots, n/4$, for the mean-detrended (linearly detrended) DFA variants. Also shown as *shaded areas* are the 90% confidence bands from simulation experiments based on AR(1) model fits (*dark*, mean-detrended; *light*, linearly detrended); the median (50%) simulation results are drawn as *white lines*. Simulation results are plotted for the range $l = 1$–$15\,\mathrm{a}$, for which power laws and the related question after the suitability of short- and long-memory models for the data are discussed in the main text.

both cases, $n_{\mathrm{sim}} = 10{,}000$ AR(1) time series were generated (identical means, variances and autocorrelations as the data) and two DFA variants applied (mean detrending and linear detrending). The central 90% of simulated $F(l)$ at each point l in the scaling region are shown as shaded area (Fig. 2.7). This is a percentile confidence band, that is, a set of percentile confidence intervals (Chapter 3) for F over a range of l values. The confidence bands contain large portions of the $F(l)$ curves from the data. This indicates that DFA is not an ideal method to discriminate between power-law and AR(1) models. The median simulation result illustrates this point, where the AR(1) model produces an almost perfect straight line, which could be mis-interpreted as power-law behaviour (Fig. 2.7). However, systematic deviations exist between AR(1) and power-law models in the DFA plots for larger l. These could indicate some significant long-memory behaviour. This finding is supported by the ARFIMA fits, which have lower AICC values (Eq. 2.46) than simple AR(1) fits.

As regards the dispute about the universality of the power law in temperature series, on basis of the AR(1) and ARFIMA estimations (Table 2.1), we conclude that the oceanic data have a stronger memory than the land data. Because the difficulties associated with DFA in interpreting the double-logarithmic plots and selecting the suitable detrending method, ARFIMA models with their elaborated estimation techniques (Beran 1994) are to be preferred for quantifying long memory. This could also be the reason why DFA is almost completely absent from the statistical literature.

Although the evidence for long-memory dependence in temperature time series seems yet not strong, more records should be analysed with ARFIMA estimation for achieving a better overview. However, analysing aggregated spatial averages, such as northern hemisphere temperature (Rybski et al. 2006), is likely unsuited for this purpose because the aggregation of short-memory AR(1) processes with distributed autocorrelation parameters yields a long-memory process (Granger 1980): the long memory may be a spurious effect of the aggregation. This has been noted also by Mills (2007). Aggregation is likely not a problem for the data analysed here, which come from one station (Siberia) and less than or equal to four stations (North Atlantic), respectively. Hemispheric averages may, however, result from processing several thousand individual records (Chapter 8).

2.5.3 Long memory of river runoff

Hurst found in an influential paper (Hurst 1951) evidence for long memory in runoff records from the Nile. The long-memory property of runoff time series has subsequently been confirmed for a number of rivers; see, for example, Mandelbrot and Wallis (1969), Hosking (1984), Mesa and Poveda (1993), Montanari et al. (1997), Montanari (2003), Pelletier and Turcotte (1997), Koutsoyiannis (2002), Bunde et al. (2005) and Koscielny-Bunde et al. (2006). Hurst's finding inspired the development of the theory of long-memory processes (Section 2.4.1) and of their estimation (Hosking 1984). Up to now, no widely accepted physical explanation of the "Hurst phenomenon" of long memory, that means on the basis of the physical–hydrological system properties, has been found (Koutsoyiannis 2005a,b).

The paper by Mudelsee (2007) presents an explanation, which suggests that a river network aggregates short-memory precipitation and converts it into long-memory runoff. River basins (Fig. 2.8) form a network of tributaries, confluences and reservoirs that has been geometrically described as a fractal object (Rodriguez-Iturbe and Rinaldo 1997). Consider a single, hydrologically homogeneous area A_j, that is, a reservoir with a linear input–output rule described by a dimensionless positive constant k_j. If the input to the reservoir, given by precipitation minus evaporation, is a purely random process, then it has been shown (Klemeš 1978) that the output, $X_j(i)$, is an AR(1) process (Section 2.1) with autocorrelation parameter $a_j = 1/(k_j + 1)$. This further implies that the runoff at a point in a river is not from a single reservoir but a cascade (Klemeš 1974) of reservoirs, one feeding the next (Fig. 2.8):

$$Q(i) = \sum_{j=1}^{m} X_j(i), \qquad i = 1, \ldots, n, \tag{2.32}$$

where $X_j(i)$ are mutually independent AR(1) processes with autocorrelation parameter a_j. It has been shown previously that if the $X_j(i)$ have identical means (zero), identical standard deviations (unity) and the a_j are either beta-distributed (Granger 1980) or uniformly distributed (Linden 1999), then for $m \to \infty$ the aggregated process $X_j(i)$ is a long-memory process. Monte Carlo simulations (Mudelsee 2007) reveal that the estimated long-memory parameter δ increases with m and that saturation of δ sets in already for $m \approx 100$. This leads to the suggestion (Mudelsee 2007) that long memory in river runoff results from spatial aggregation of many short-memory reservoir contributions.

To test the aggregation hypothesis, Mudelsee (2007) studied the long-memory parameter δ of fitted ARFIMA$(1, \delta, 0)$ models in dependence on the basin size, A. The idea of the $\delta(A)$ estimation is that with increasing A also the number m of contributions $X_j(i)$ grows. Thereby should also δ increase, from zero ($m = 1$) to a saturation level below

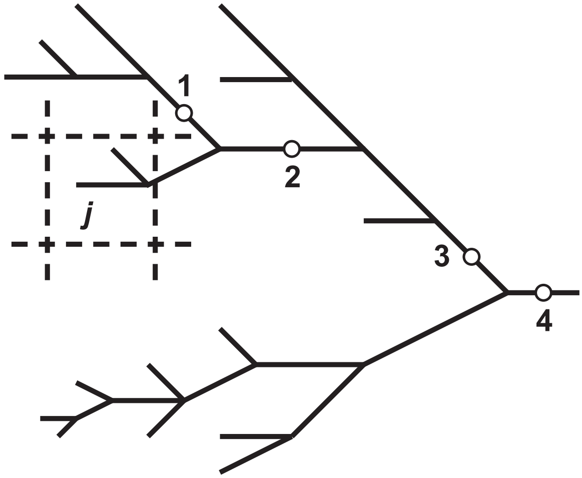


Figure 2.8. River network. Runoff at a point (e.g., 4) is the spatial aggregation of runoff from upstream (to the left in the picture). Shown is also a hypothetical spatial unit j with area A_j. The basin size, A, for a point is given by the sum of the areas A_j upstream. (From Mudelsee (2007), with permission from the publisher.)

0.5 (m large). The fact that the distribution of the a_j for real rivers is difficult to derive empirically or analytically, can be ignored at this low level of sophistication of the hypothesis. The resulting $\delta(A)$ curve for one of the longest available runoff records, from the river Weser (Germany) (Fig. 2.9), basically confirms the aggregation explanation of the Hurst phenomenon.

Mudelsee (2007) estimated $\delta(A)$ curves also for other rivers, finding similar $\delta(A)$ increases (Elbe, Rhine, Colorado and Nile) but also in one case (Mississippi) a $\delta(A)$ decrease. This paper discusses the validity of the various assumptions made by the aggregation hypothesis. One particular criticism is that the linear input–output release rule may be violated for very large reservoirs. Another obstacle is the requirement of very long time series (above, say, 70 years) for obtaining sufficient accuracy. A major criticism is that the aggregation of AR(1) processes is not an ARFIMA process (Linden 1999), and that the result (Fig. 2.9) may therefore be affected by estimation bias. This paper (Mudelsee 2007) further finds little evidence for long memory in precipitation records from the same regions as the river basins. It thus appears appropriate to reserve the concept of the "Hurst phenomenon" for hydrological time series, and not for climate time series in general.

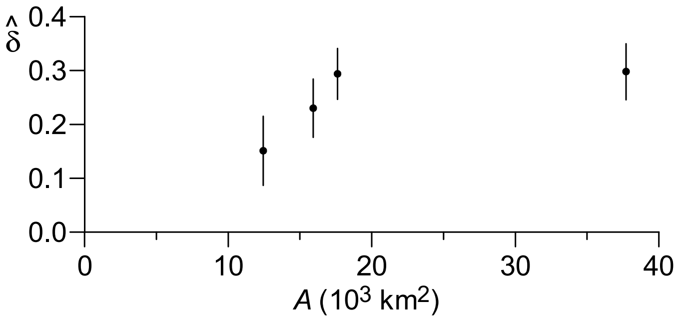


Figure 2.9. Long-memory parameter in dependence on basin size, river Weser. The δ estimates (*dots*) are shown with bootstrap standard errors (Doornik and Ooms 2003). The time series are monthly runoff values from January 1857 to April 2002. δ was estimated using an ARFIMA$(1, \delta, 0)$ model and maximum likelihood (Doornik and Ooms 2003). Prior to the estimations, the data were logarithmically transformed, the annual cycle removed by subtracting the day-wise long-term averages and the linear trends removed. This ARFIMA model had for all four river stations better AIC values than the ARFIMA$(0, \delta, 0)$ model. (After Mudelsee 2007.)

2.6 Background material

Textbooks on the theory and estimation of ARMA(p, q) processes were written by Priestley (1981), Brockwell and Davis (1991, 1996), Box et al. (1994) and Chatfield (2004). Nonlinear time series models are covered by Priestley (1988), Tong (1990) and Fan and Yao (2003). The latter two books have the notable aim to bridge the gap between statistics and nonlinear dynamics, see also Tong (1992, 1995). Long-memory processes are the topic of Beran (1994, 1997), Doukhan et al. (2003) and Robinson (2003). ARFIMA(p, δ, q) processes are reviewed by Beran (1998) and fractional Gaussian noise processes by Mandelbrot (1983). Tables of series and other formulas can be found in the books by Abramowitz and Stegun (1965) and Gradshteyn and Ryzhik (2000).

The **AR(1) model standard formulation** is (Priestley 1981: Section 3.5.2 therein)

$$X_{\text{noise}}(i) = a \cdot X_{\text{noise}}(i-1) + \mathcal{E}(i), \tag{2.33}$$

where $\mathcal{E}(i)$ is a stationary purely random process with mean μ_ϵ and standard deviation σ_ϵ. In the general case this noise model is not stationary. For $\mu_\epsilon \neq 0$ and $|a| < 1$, $E\left[X_{\text{noise}}(i)\right]$ is not constant but approaches with time a saturation value of $\mu_\epsilon/(1-a)$. For $\mu_\epsilon = 0$ and $|a| < 1$, $VAR\left[X_{\text{noise}}(i)\right]$ is not constant but approaches with time a saturation value of $\sigma_\epsilon^2/(1-a^2)$. The case $\mu_\epsilon = 0$ and $|a| = 1$ results in a random walk process (p. 60). The standard formulation in Eq. (2.33) describes an "asymptotically stationary" process, whereas Eq. (2.1) describes a strictly stationary process. In practical applications such as random number generation (Section 2.7), the standard model (Eq. 2.33) can be used if the transient sequence of numbers (say, the first 5000) is discarded.

The **covariance** between two random variables, X and Y, is

$$COV\left[X,Y\right] = E\Big[\big(X - E\left[X\right]\big) \cdot \big(Y - E\left[Y\right]\big)\Big]. \tag{2.34}$$

A special case is $COV\left[X,X\right] = VAR\left[X\right]$.

The **effective data size** is reduced for a persistent process. This is shown (Fig. 2.10a) for the case of mean estimation of an AR(1) process, where even for moderate values ($n \gtrapprox 50$ and $a \lessapprox 0.5$), n'_μ is considerably smaller than n. The data size reduction is quantified by Eq. (2.7). A simplified version based on the decorrelation time (Eq. 2.8) underestimates n'_μ by less than 5% for $n \gtrapprox 50$ and $a \lessapprox 0.5$, as shown by Fig. 2.10b. Even for moderate values of n ($\gtrapprox 50$) and a ($\lessapprox 0.5$), the influence of persistence on n'_μ can be considerable. Eq. (2.3) is valid only for the mean estimator. The effective data size for variance estimation of an AR(1) process,

$$n'_{\sigma^2} = n \left[1 + 2 \sum_{i=1}^{n-1} \left(1 - i/n\right) \rho(i)^2\right]^{-1} \tag{2.35}$$

(Bayley and Hammersley 1946), is given by

$$n'_{\sigma^2} = n \left\{1 + \frac{2}{n}\,\frac{1}{1-a^2}\left[a^2\left(n - \frac{1}{1-a^2}\right) - a^{2n}\left(1 - \frac{1}{1-a^2}\right)\right]\right\}^{-1}. \tag{2.36}$$

Likewise, the effective data size for correlation estimation between two processes $X(i)$ and $Y(i)$ with autocorrelation functions $\rho_X(i)$ and $\rho_Y(i)$,

$$n'_{\rho} = n \left[1 + 2 \sum_{i=1}^{n-1} \left(1 - i/n\right) \rho_X(i)\, \rho_Y(i)\right]^{-1} \tag{2.37}$$

(von Storch and Zwiers 1999), is in the case of two AR(1) processes with persistence parameters a_X and a_Y given by

$$\begin{aligned} n'_{\rho} = n \Bigg\{1 + \frac{2}{n}\,\frac{1}{1-a_X a_Y}\Bigg[& a_X a_Y\left(n - \frac{1}{1-a_X a_Y}\right) \\ & -\left(a_X a_Y\right)^n\left(1 - \frac{1}{1-a_X a_Y}\right)\Bigg]\Bigg\}^{-1}. \end{aligned} \tag{2.38}$$

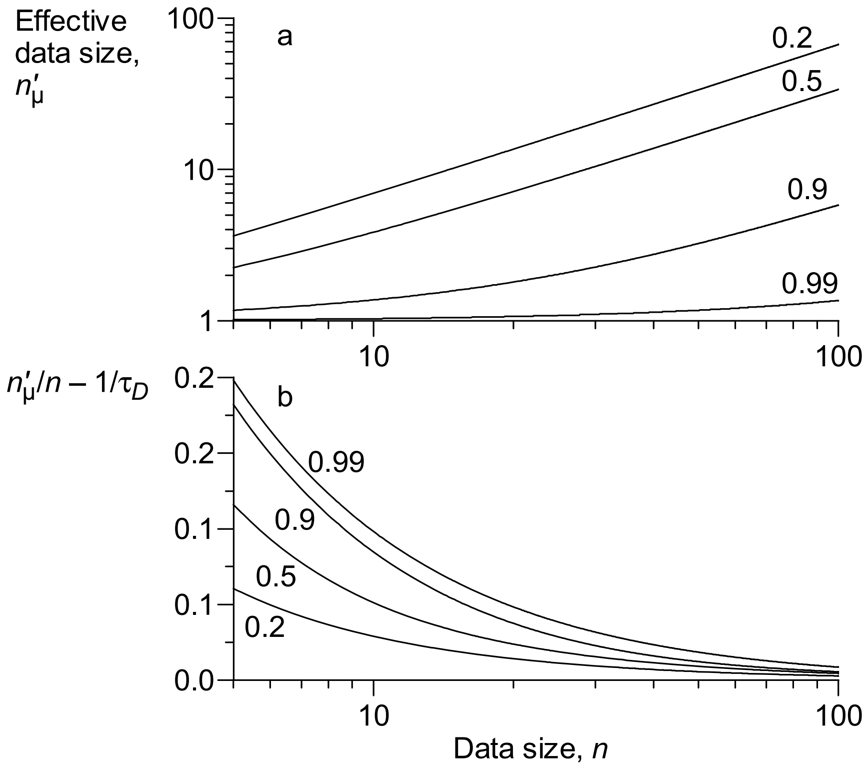


Figure 2.10. Effective data size, mean estimation of an AR(1) process. **a** Dependence of n'_μ on n after Eq. (2.7), for various persistence values a (0.2, 0.5, 0.9 and 0.99). **b** Comparison of the exact expression (Eq. 2.7) with a simplified version based on the decorrelation time (Eq. 2.8).

Early papers in climatology on effective data size and the influence of persistence on estimation variance include Matalas and Langbein (1962), Leith (1973), Laurmann and Gates (1977), Thiébaux and Zwiers (1984), Trenberth (1984a,b) and Zwiers and von Storch (1995).

Various **approximate bias** (and variance) formulas have been published for estimators of the autocorrelation parameter a in evenly spaced AR(1) models (Eqs. 2.1 and 2.33). Marriott and Pope (1954) analysed $\widehat{a}$ (Eq. 2.4) and gave

$$E\left(\widehat{a}\right) \simeq \left(1 - 2/n\right) a. \tag{2.39}$$

White (1961) gave an approximation of higher order in terms of powers of $(1/n)$:

$$E\left(\widehat{a}\right) \simeq \left(1 - 2/n + 4/n^2 - 2/n^3\right)\, a + \left(2/n^2\right)\, a^3 + \left(2/n^2\right)\, a^5. \tag{2.40}$$

One drawback of these approximations is that they are not accurate for large a. For $a \to 1$, also $\widehat{a} \to 1$ (Eq. 2.4) and the bias, $E(\widehat{a}) - a$, approaches zero. This behaviour is not contained in Eqs. (2.39) or (2.40). It is, however, contained in the bias formula of Mudelsee (2001a):

$$E(\widehat{a}) \simeq [1 - 2/(n-1)]\; a + \left[2/(n-1)^2\right] \left(a - a^{2n-1}\right) / \left(1 - a^2\right). \quad (2.41)$$

Mudelsee (2001a) showed that this approximation is more accurate than Eq. (2.40) for $a \gtrapprox 0.88$. The estimation variance of $\widehat{a}$ is to a low approximation order (Bartlett 1946)

$$VAR(\widehat{a}) \simeq \left(1 - a^2\right)/n \quad (2.42)$$

and to a higher order (White 1961)

$$VAR(\widehat{a}) \simeq \left(1/n - 1/n^2 + 5/n^3\right) - \left(1/n - 9/n^2 + 53/n^3\right) a^2 - \left(12/n^3\right) a^4. \quad (2.43)$$

Higher-order approximations of the first four moments of $\widehat{a}$ are given by Shenton and Johnson (1965). From a practical point of view, it is more realistic to assume that the mean of $X_{\text{noise}}(i)$ is unknown and has to be subtracted. In case of the AR(1) process with unknown mean, the analogue of the estimator in Eq. (2.4) is

$$\widehat{a} = \sum_{i=2}^{n} [x_{\text{noise}}(i) - \bar{x}_{\text{noise}}] \cdot [x_{\text{noise}}(i-1) - \bar{x}_{\text{noise}}] \Bigg/ \sum_{i=2}^{n} [x_{\text{noise}}(i) - \bar{x}_{\text{noise}}]^2, \quad (2.44)$$

where $\bar{x}_{\text{noise}} = \sum_{i=1}^{n} x_{\text{noise}}(i)/n$ is the sample mean. The approximate expectation of this estimator is (Kendall 1954)

$$E(\widehat{a}) \simeq a - (1 + 3a)/(n-1). \quad (2.45)$$

Monte Carlo simulations (Fig. 2.11) indicate that this approximation can be used for bias correction in situations with uneven spacing and moderate autocorrelation (a less than, say, 0.9). The bias-corrected autocorrelation coefficient, $\widehat{a}'$, is obtained from $\widehat{a}$ by inserting $\widehat{a}'$ for a on the right-hand side and $\widehat{a}$ for $E(\widehat{a})$ on the left-hand side in one of the equations describing the bias, say Eq. (2.45), and solving this equation for $\widehat{a}'$. Approximations for the bias of least-squares and Yule–Walker estimators of AR(p) processes with known/unknown mean have been given for $p \leq 6$ by Shaman and Stine (1988). Sample mean subtraction is a special case of detrending. It follows that if we obtain $\{x_{\text{noise}}(i)\}_{i=1}^{n}$ from the data $\{x(i)\}_{i=1}^{n}$ by removing an estimated trend function, $\{x_{\text{trend}}(i)\}_{i=1}^{n}$ (Eq. 1.2), in principle we have to replace $\bar{x}_{\text{noise}}$ in Eq. (2.44) by $x_{\text{trend}}(i)$. For trends more complex than a constant function, bias properties of such estimators seem, however, to be analytically untractable.

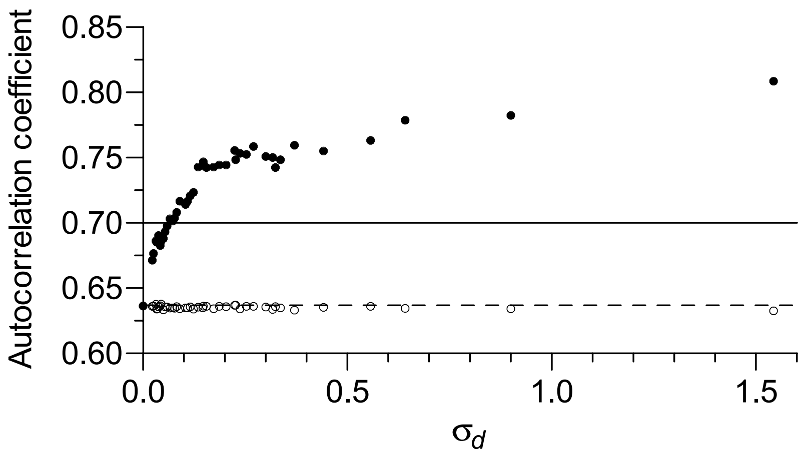


Figure 2.11. Monte Carlo study of the bias in the autocorrelation estimation of an AR(1) process, unknown mean, uneven spacing. Identical Monte Carlo parameters and time series properties were used as in Fig. 2.3. The estimators are also the same, with the exception that the sample mean was removed from the time series prior to the estimations. The negative bias approximation (*dashed line*) is from the case of even spacing (Kendall 1954). See Fig. 2.3 for further explanation.

Another AR(1) parameter estimator was introduced (Houseman 2005), based on an estimation function (which has zero expectation at the true parameter value). This author gave the estimation function robustness with respect to outliers (Chapter 3), included a linear regression term (Chapter 4), that is, performed a joint estimation, and presented an application to unevenly spaced water monitoring time series from Boston Harbor.

ARMA($\boldsymbol{p, q}$) estimation. A least-squares estimation (Brockwell and Davis 1991: Chapter 8 therein) can be used to fit even-spacing ARMA(p, q) models to data $\{x_{\text{noise}}(i)\}_{i=1}^{n}$. Besides least squares, statistical practice normally uses the maximum likelihood principle, which means to search for the ARMA(p, q) parameters $\{a_1, \ldots, a_p; b_0, \ldots, b_q\}$ that maximize the likelihood that the fitted model had produced the data (Brockwell and Davis 1991: Chapter 8 therein). This may be numerically difficult but is no fundamental restriction. Another important point is model identification, that means, selection of p and q. Guidance for that gives the sample autocorrelation function (Eq. 2.17), which can be compared with the autocorrelation function of the model candidate; analogously used are the partial autocorrelation functions (Brockwell and Davis 1991: Chapter 9 therein). Some quantitative measures of goodness of fit exist, such as Akaike's (1973) information criterion (AIC). A bias-corrected version of the AIC, referred to as the AICC (Hurvich and Tsai 1989), is calculated from the maximized likelihood, L, plus a penalty term for the number of parameters,

$$\mathrm{AICC} = -2\ln(L) + 2(p+q+1)n/(n-p-q-2). \tag{2.46}$$

The penalty is a mathematical expression of Ockham's razor: it is easier to fit a model with more parameters, or alternatively: it is preferable to explain the data using a model with fewer parameters. The context of this book, uneven spacing and the embedding problem for ARMA(p, q) processes, forces us to take a rather parsimonious position by adopting the AR(1) model ($p = 1, q = 0$). It could therefore be that the parametric AR(1) model misses some properties, representable in higher-order continuous-time models, of the observed noise process. A milder case is when the spacing arises from equidistance with some missing observations (Fig. 1.15). Then a possible solution is to use a state-space representation, Kalman filtering and an adaption of the likelihood function. This approach has been pioneered by Jones (1981, 1985, 1986) and Jones and Tryon (1987).

A **maximum likelihood estimate for the AR(1) model** with uneven spacing is given by Robinson (1977). For even and uneven spacing, it is useful to visually check the fit residuals (predictions for $x_{\mathrm{noise}}(i)$ by the fitted model minus the data $x_{\mathrm{noise}}(i)$), which are for a proper fit realizations of a purely random process. For higher-order ARMA(p, q) models, it seems that no one embarked on the derivation of analytical approximations of the estimation bias.

ARFIMA(p, δ, q) models were introduced by Granger and Joyeux (1980) and Hosking (1981). For $|\delta| < 0.5$, the fractional difference operator $(1 - B)^\delta$ is defined by

$$\begin{aligned}(1-B)^\delta &= \sum_{k=0}^{\infty} \binom{\delta}{k} (-B)^k \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(\delta+1)}{\Gamma(k+1)\,\Gamma(\delta-k+1)} (-B)^k ,\end{aligned} \tag{2.47}$$

where B is the backshift operator and $\Gamma(\cdot)$ is the gamma function. Maximum likelihood and other estimation techniques are described by Beran (1994, 1997).

Double-logarithmic transformations followed by linear regression are generally not suited to estimate power-law models when the original data have Gaussian error distributions. Although this has since long been known in various disciplines (Rützel 1976; Freund and Minton 1979; Miller 1984; Jansson 1985; Mudelsee and Stattegger 1994), the transformation is still frequently encountered in various applied sciences today. The low power of the double-logarithmic plot to discriminate between scaling and no scaling in noisy data has been criticized by Tsonis and Elsner (1995). These authors suggest a test for scaling, namely to plot the slope with bootstrap confidence interval (Chapter 3) against the scale and look for a plateau (constant slope). However, for too small data sizes no plateau behaviour might be found, despite the existence of scaling. Maraun et al. (2004) showed in a Monte Carlo study analysing artificial time series with DFA that well over 100,000 data points may be required. This result together with the other criticisms make DFA irrelevant for climate time series analysis.

A **random walk** process arises from the case $a = 1$ in Eq. (2.33). For this process, $VAR\left[X_{\text{noise}}(i)\right]$ increases linearly with time, that means, the random walk is not stationary. In climatology, where the variables are within certain bounds, the random walk has to be modified to serve as a noise model. In that manner, it has been applied to short-term temperature fluctuations (Kärner 2002). In case of Pleistocene timescales (Table 1.1), Wunsch (2003) suggested a random walk for explaining the 100-ka ice-age cycle. He put bounds to the ice-volume variable $X_{\text{noise}}(i)$; when the system attempts to leave the permitted range, it is thrown back. Mostly other fields than climatology, such as econometrics, apply tests of the hypothesis "$a = 1$" for the autoregressive process, or its generalization ("unit-roots tests") for the ARMA(p, q) process (Fuller 1996: Chapter 10 therein). Several bootstrap hypothesis tests (Chapter 3) for unit-roots were examined by means of Monte Carlo simulations (Palm et al. 2008). A climatological application is the paper by Stern and Kaufmann (2000), where unit-roots were identified in tests of hemispheric temperature records, circa 1855–1995. Because these tests generally have poor power (loosely speaking, detection probability; see Eq. (3.41)) (Chatfield 2004: Section 13.4 therein), and because of the nonstationarity, we will not consider further random walk models for $X_{\text{noise}}(i)$.

Further **examples** of time-series models fitted to climate data are the following. It is fair to say that the vast majority of such papers used the AR(1) process with $a > 0$ as a model of $X_{\text{noise}}(i)$. One classic, from meteorology, is Gilman et al. (1963). As an example from late Pleistocene climatology, we cite Roe and Steig (2004), who characterize Arctic and Antarctic climate by means of the AR(1) persistence time. Since this book adopts the AR(1) noise model, we will encounter various applications of it in the following chapters. It may be noted that

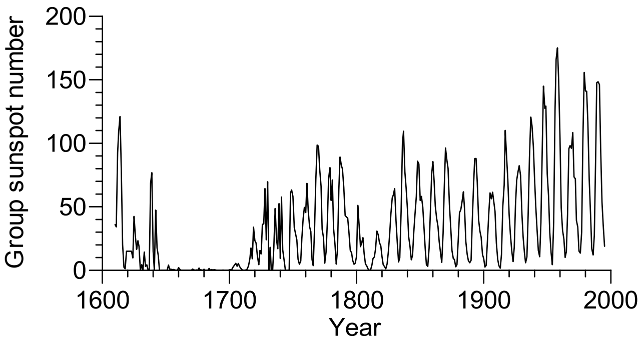


Figure 2.12. Group sunspot number, 1610–1995. Sunspots are dark spots on the Sun's surface, visible from the Earth with a telescope. They present regions of reduced temperature. Satellite measurements, available since 1980, show that the solar activity correlates positively with the number of sunspots (Willson and Hudson 1988), that means, the solar constant is no constant. The group sunspot number is a way of counting the sunspots as groups and thought to give a more accurate picture over the previous centuries than using the individual sunspot data (Hoyt and Schatten 1998). The long lasting minimum during approximately 1645–1715 is the Maunder Minimum. Beer et al. (1998) demonstrate that this was not a period without solar activity variations. (Data have $d(i) = 1\,\mathrm{a}$ and are from Hoyt and Schatten (1998).)

in hydrology cases of $a < 0$ (anti-persistence or blue noise) are found (Milly and Wetherald 2002). Annual layer thickness in ice cores may also exhibit blue noise on very short (annual) timescales (Fisher et al. 1985): if the true thickness of a layer is, say, larger than measured, then the true thickness of a neighboured layer is likely smaller than measured (since the overall thickness is constrained). Yule (1927) fitted an AR(2) model to sunspot data, 1749–1924. These data (Fig. 2.12), which exhibit quasi-periodic behaviour with period $\sim 11\,\mathrm{a}$, have since this pioneering work been the hobbyhorse of time series analysts. Tong and Lim (1980) took a SETAR$(2; 4, 12)$ process, which reproduces the sunspot cycles' asymmetry (rise and descent). Jones (1981) fitted various ARMA(p, q) models, and Priestley (1981: Section 11 therein) compared the fits of AR(p), ARMA(p, q) and threshold autoregressive models to the sunspot data. Seleshi et al. (1994) fitted a high-order autoregressive model to the sunspot data but found that AR(p) or ARMA(p, q) models gave no satisfactory fit to the rainfall series, 1900–1991, from Addis Ababa. This stimulated their search for a transformation of the rainfall data which could produce a better relation with solar activity variations. Matyasovszky (2001) fitted an AR(4) and threshold autoregressive models to the longest record of monthly instrumental observations, the central England temperature time series (Manley 1974), which starts in January 1659. He found four regimes, one of which has a limit cycle of about 2 a period and might correspond to the meteorological phenomenon of the quasi-biennial oscillation, which refers to zonal wind-speed variations in the tropical stratosphere. Stedinger and Crainiceanu (2001) considered ARMA(p, q) models for the logarithmically transformed maximum annual runoff of the river Missouri from 1898 to 1998. Koen and Lombard (1993) applied ARMA(p, q) modelling to astronomical time series. Stattegger (1986) used ARMA(p, q) processes to describe variations in the composition of heavy metals in sedimentary deposits of rivers in Austria. Newton et al. (1991) fitted ARMA(p, q) models to the Pleistocene SPECMAP δ^{18}O curve. Giese et al. (1999) searched for suitable ARMA(p, q) and ARIMA models for δ^{18}O variations in a Pliocene–Pleistocene deep-sea sediment core without success. Stephenson et al. (2000) analysed the wintertime North Atlantic Oscillation (NAO) index from 1864 to 1998, that is, the difference between standardized December–March mean sea-level pressures measured at Lisbon and Iceland. (Other versions exist, which use Azores instead of Lisbon.) The NAO index is used as a measure to summarize the mean westerly atmospheric flow over the North Atlantic region, which in turn influences the weather in Europe (Hurrell 1995). It was found (Stephenson et al. 2000) that an ARFIMA(1, 0.13, 0) model describes the data as good as an AR(10) model and better than the asymptotic stationary AR(1) or random walk models. The more parsimonious long-memory model was preferred to make NAO forecasts. Divine et al. (2008) presented a model of piecewise AR(1) processes and a maximum likelihood technique for estimating the autocorrelation parameters and the change-point times. This model constitutes an interesting augmentation of the SETAR($l; k, \ldots, k$) model. Divine et al. (2008) applied their method to detect changes in records of the NAO, ENSO and ice core δ^{18}O. Kallache et al. (2005) fitted ARFIMA(p, δ, q) models to runoff records from small rivers in southern Germany and found via an AICC variant that the parsimonious AR(1) model, contained in the ARFIMA model, is often a suitable description but also that frequently a nonzero long-term parameter δ is required. These examples illustrate that there is an interplay between what is put into the noise process and what is put into the trend. Taking more complex noise processes does reduce the need to consider complex trend functions. This book is devoted to using simpler noise processes and more complex trend functions. Besides the necessity to keep the noise simple because of the embedding problem (Section 2.3), this position allows to quantify trend parameters, that means for example, climate changes, using regressions (Chapter 4).

2.7 Technical issues

The **gamma distribution** in its standard form with order parameter a has following PDF:

$$f(x) = x^{a-1} \exp(-x) / \Gamma(a), \qquad x \geq 0. \tag{2.48}$$

Ahrens and Dieter (1974) devised an algorithm for generating gamma random variables. See Johnson et al. (1994: Chapter 17 therein) for more details on the gamma distribution.

The **gamma function** is defined by

$$\Gamma(z) = \int_0^\infty y^{z-1} \exp(-y) dy. \tag{2.49}$$

When z is an integer, $\Gamma(z+1) = z!$, otherwise approximations have to be used for calculation. Lanczos (1964) devised an algorithm for approximating $\ln[\Gamma(z)]$. Ratios of gamma functions can then be numerically advantageously evaluated as $\Gamma(a) / \Gamma(b) = \exp\{\ln[\Gamma(a)] - \ln[\Gamma(b)]\}$.

The **beta distribution** in its standard form with parameters $p > 0$ and $q > 0$ has following PDF:

$$f(x) = x^{p-1}(1-x)^{q-1} / B(p,q), \qquad 0 \leq x \leq 1. \tag{2.50}$$

The beta function, B, is given by $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$. See Johnson et al. (1995: Chapter 25 therein) for more details on the beta distribution.

TAUEST is a FORTRAN 77 program that estimates the persistence time of an AR(1) process and uneven spacing. The minimization of the least-squares sum (Eq. 2.11) is done using Brent's search (Press et al. 1992). The software includes residual analysis, bias correction and construction of a bootstrap percentile confidence interval (which is described in Chapter 3). TAUEST is described by Mudelsee (2002). The software is available at the web site for this book.

Jones (1981) gave FORTRAN 77 subroutines to fit data to a continuous-time autoregressive process. This program uses a state space representation and a maximum likelihood principle.

ITSM 2000 is a Windows package for the estimation of ARMA, ARFIMA and other models under even spacing. It includes tools for transformations, regression, forecasting, smoothing and spectral estimation. A version for a limited data size is included in the book by Brockwell and Davis (1996), a full version can be obtained from B & D Enterprises, Inc. (`pjbrock@stat.colostate.edu`, email from 16 April 2004).

STAR is a DOS program for fitting threshold autoregressive models to evenly spaced time series. The book by Tong (1990) informs that the software could be obtained from Microstar Software (Canterbury, UK).

S-Plus routines for fitting long-memory models to data are listed in the monograph by Beran (1994). This book contains also an S-Plus routine for the simulation of ARFIMA processes. Hosking (1984) gives another algorithm for ARFIMA simulation.

Ox is a computer language, for which a package (Doornik and Ooms 2001) for maximum likelihood fitting of ARFIMA models is available (`http://www.doornik.com`, 18 December 2005).

Random number generators are required to perform Monte Carlo experiments such as that from Fig. 2.3. Also bootstrap resampling (Chapter 3) uses such tools. Almost exclusively employed are pseudorandom numbers, which are generated by mathematical algorithms. One simple form is the multiplicative congruential generator, $Z_i = A\, Z_{i-1} \pmod{M}$, $i \geq 1$, where A, M and Z_i (pseudorandom numbers) are integers. Z_1 is primed (seeded). Z_i can be mapped onto the interval $[0; 1]$ to produce a uniform distribution. The uniform serves also as basis to generate other types of distribution. For example, the Gaussian arises from the uniform by the transformation given by Box and Muller (1958). The success of Monte Carlo experiments and bootstrap resampling depends critically on whether the generator "supplies sequences of numbers from which arbitrarily selected nonoverlapping subsequences appear to behave like statistically independent sequences and where the variation in an arbitrarily chosen subsequence of length k (≥ 1) resembles that of a sample drawn from the uniform distribution on the k-dimensional unit hypercube" (Fishman 1996: p. 587 therein). Uniform filling of the hypercube can be assessed by inspecting whether regular patters are absent in two- or three-dimensional hyperplanes. Park and Miller (1988) show that good random number generators "are hard to find." They give also a multiplicative congruential generator with $A = 16{,}807$ and $M = 2{,}147{,}483{,}647$, which may serve as minimal standard. Schrage (1979) lists a FORTRAN 77 code of this generator type. Other generators are described by Press et al. (1992: Chapter 7 therein), Fishman (1996: Chapter 7 therein) and Knuth (2001: Chapter 3 therein); the latter book (Section 3.4.2, Algorithm P therein) contains a recipe for producing random permutations.

Chapter 3

Bootstrap Confidence Intervals

In statistical analysis of climate time series, our aim (Chapter 1) is to estimate parameters of $X_{\text{trend}}(T)$, $X_{\text{out}}(T)$, $S(T)$ and $X_{\text{noise}}(T)$. Denote in general such a parameter as θ. An estimator, $\widehat{\theta}$, is a recipe how to calculate θ from a set of data. The data, discretely sampled time series $\{t(i), x(i)\}_{i=1}^{n}$, are influenced by measurement and proxy errors of $x(i)$, outliers, dating errors of $t(i)$ and climatic noise. Therefore, $\widehat{\theta}$ cannot be expected to equal θ. The accuracy of $\widehat{\theta}$, how close it comes to θ, is described by statistical terms such as standard error, bias, mean squared error and confidence interval (CI). These are introduced in Section 3.1.

With the exploration of new archives or innovations in proxy, measurement and dating techniques, new $\widehat{\theta}$ values, denoted as estimates, become available and eventually join or replace previous estimates. A telling example from geochronology is where θ is the time before present when the Earth's magnetic field changed from reversed polarity during the Matuyama epoch to normal polarity during the Brunhes epoch, at the beginning of the late Pleistocene. Estimates published over the past decades include 690 ka (Cox 1969) and 730 ka (Mankinen and Dalrymple 1979), both based on K/Ar dating; and 790 ka (Johnson 1982) and 780 ka (Shackleton et al. 1990), both based on astronomical tuning. The currently accepted value is 779 ka with a standard error of 2 ka (Singer and Pringle 1996), written as 779 ± 2 ka, based on ^{40}Ar/^{39}Ar dating (a high-precision variant of K/Ar dating). An example with a much greater uncertainty regards the case where θ is the radiative forcing (change in net vertical irradiance at the tropopause) of changes in atmospheric concentrations of mineral dust, where even the sign of θ is uncertain (Penner et al. 2001; Forster et al. 2007). It is evident that the

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_3,

growth of climatological knowledge depends critically on estimates of θ that are accompanied by error bars or other measures of their accuracy.

Bootstrap resampling (Sections 3.2 and 3.3) is an approach to construct error bars and CIs. The idea is to draw random resamples from the data and calculate error bars and CIs from repeated estimations on the resamples. For climate time series, the bootstrap is potentially superior to the classical approach, which relies partly on unrealistic assumptions regarding distributional shape, persistence and spacing (Chapter 1). However, the bootstrap, developed originally for data without serial dependence, has to be adapted before applying it to time series. Two classes of adaptions exist for taking persistence into account. First, nonparametric bootstrap methods resample sequences, or blocks, of the data. They preserve the dependence structure over the length of a block. Second, the parametric bootstrap adopts a dependence model. As such, the AR(1) model (Chapter 2) is our favorite.

It turns out that both bootstrap resampling types have the potential to yield acceptably accurate CIs for estimated climate parameters. A problem for the block bootstrap arises from uneven time spacing. Another difficult point is to find optimal block lengths. This could make the parametric bootstrap superior within the context of this book, especially for small data sizes (less than, say, 50). The block bootstrap, however, is important when the deviations from AR(1) persistence seem to be strong. Various CI types are investigated. We prefer a version (so-called BCa interval) that automatically corrects for estimation bias and scale effects. Computing-intensive calibration techniques can further increase the accuracy.

3.1 Error bars and confidence intervals

Let θ be the parameter of interest of the climatic process $\{X(T)\}$ and $\widehat{\theta}$ be the estimator. Extension to a set of parameters is straightforward. Any meaningful construction lets the estimator be a function of the process, $\widehat{\theta} = g\left(\{X(T)\}\right)$. That means, $\widehat{\theta}$ is a random variable with statistical properties. The standard deviation of $\widehat{\theta}$, denoted as standard error, is

$$\mathrm{se}_{\widehat{\theta}} = \left[VAR\left(\widehat{\theta}\right)\right]^{1/2}. \tag{3.1}$$

The bias of $\widehat{\theta}$ is

$$\mathrm{bias}_{\widehat{\theta}} = E\left(\widehat{\theta}\right) - \theta. \tag{3.2}$$

$\mathrm{bias}_{\widehat{\theta}} > 0$ ($\mathrm{bias}_{\widehat{\theta}} < 0$) means a systematic overestimation (underestimation). $\mathrm{se}_{\widehat{\theta}}$ and $\mathrm{bias}_{\widehat{\theta}}$ are illustrated in Fig. 3.1. Desirable estimators have small $\mathrm{se}_{\widehat{\theta}}$ and small $\mathrm{bias}_{\widehat{\theta}}$. In many estimations, a trade-off problem

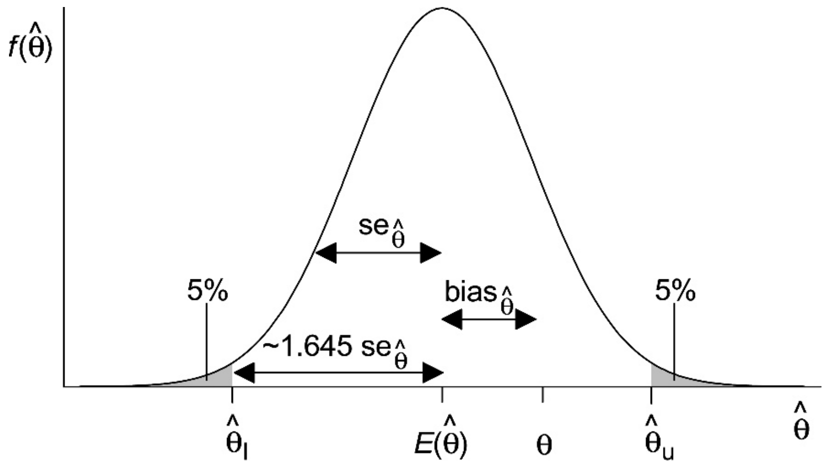


Figure 3.1. Standard error ($\mathrm{se}_{\widehat{\theta}}$), bias ($\mathrm{bias}_{\widehat{\theta}}$) and equi-tailed confidence interval ($\mathrm{CI}_{\widehat{\theta},1-2\alpha} = [\widehat{\theta}_{\mathrm{l}}; \widehat{\theta}_{\mathrm{u}}]$) for a Gaussian distributed estimator, $\widehat{\theta}$. The true parameter value is θ; the confidence level is $1 - 2\alpha = 90\%$.

between $\mathrm{se}_{\widehat{\theta}}$ and $\mathrm{bias}_{\widehat{\theta}}$ occurs. A convenient measure is the root mean squared error,

$$\begin{aligned} \mathrm{RMSE}_{\widehat{\theta}} &= \left\{ E\left[\left(\widehat{\theta} - \theta \right)^2 \right] \right\}^{1/2} \\ &= \left(\mathrm{se}_{\widehat{\theta}}^{\,2} + \mathrm{bias}_{\widehat{\theta}}^{\,2} \right)^{1/2}. \end{aligned} \tag{3.3}$$

The coefficient of variation is

$$\mathrm{CV}_{\widehat{\theta}} = \mathrm{se}_{\widehat{\theta}} \Big/ \left| E\left(\widehat{\theta} \right) \right|. \tag{3.4}$$

While $\widehat{\theta}$ is a best guess of θ or a point estimate, a CI is an interval estimate that informs how good a guess is (Fig. 3.1). The CI for θ is

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}_{\mathrm{l}}; \widehat{\theta}_{\mathrm{u}} \right], \tag{3.5}$$

where $0 \le 1 - 2\alpha \le 1$ is a prescribed value, denoted as confidence level. The practical examples in his book consider 90% ($\alpha = 0.05$) or 95% ($\alpha = 0.025$) CIs, which are reasonable choices for climatological problems. $\widehat{\theta}_{\mathrm{l}}$ is the lower, $\widehat{\theta}_{\mathrm{u}}$ the upper endpoint of the CI. $\widehat{\theta}_{\mathrm{l}}$ and $\widehat{\theta}_{\mathrm{u}}$ are random variables and have statistical properties such as standard error or bias. The properties of interest for CIs are the coverages,

$$\gamma_{\mathrm{l}} = \mathrm{prob}\left(\theta \leq \widehat{\theta}_{\mathrm{l}}\right), \tag{3.6}$$

$$\gamma_{\mathrm{u}} = \mathrm{prob}\left(\theta \geq \widehat{\theta}_{\mathrm{u}}\right) \tag{3.7}$$

and

$$\gamma = \mathrm{prob}\left(\widehat{\theta}_{\mathrm{l}} < \theta < \widehat{\theta}_{\mathrm{u}}\right) = 1 - \gamma_{\mathrm{l}} - \gamma_{\mathrm{u}}. \tag{3.8}$$

Exact CIs have coverages, γ, equal to the nominal value $1 - 2\alpha$. Construction of exact CIs requires knowledge of the distribution of $\widehat{\theta}$, which can be achieved only for simple problems. In more complex situations, only approximate CIs can be constructed (Section 3.1.3). As regards the division of the nominal coverage between the CI endpoints, this book adopts a practical approach and considers only equi-tailed CIs, where nominally $\gamma_{\mathrm{l}} = \gamma_{\mathrm{u}} = \alpha$. As a second CI property besides coverage, we consider interval length, $\widehat{\theta}_{\mathrm{u}} - \widehat{\theta}_{\mathrm{l}}$, which is ideally small.

Preceding paragraphs considered estimators on the process level. In practice, on the sample level, we plug in the data $\{t(i), x(i)\}_{i=1}^{n}$ for $\{T(i), X(i)\}_{i=1}^{n}$. Following the usual convention, we denote also the estimator on the sample level as $\widehat{\theta}$. An example is the autocorrelation estimator (Eq. 2.4).

3.1.1 Theoretical example: mean estimation of Gaussian white noise

Let the process $\{X(i)\}_{i=1}^{n}$ be given by

$$X(i) = \mathcal{E}_{\mathrm{N}(\mu,\, \sigma^2)}(i), \qquad i = 1, \ldots, n, \tag{3.9}$$

which is called a Gaussian purely random process or Gaussian white noise. There is no serial dependence, and the times $T(i)$ are not of interest. Consider as estimator $\widehat{\theta}$ of the mean, μ, the sample mean, written on process level as

$$\widehat{\mu} = \bar{X} = \sum_{i=1}^{n} X(i)/n. \tag{3.10}$$

Let also σ be unknown and estimated by the sample standard deviation, $\widehat{\sigma} = S_{n-1}$, given in the next example (Eq. 3.19). The properties of $\bar{X}$ readily follow as

$$\mathrm{se}_{\bar{X}} = \sigma \cdot n^{-1/2}, \tag{3.11}$$

$$\mathrm{bias}_{\bar{X}} = 0, \tag{3.12}$$

$$\mathrm{RMSE}_{\bar{X}} = \mathrm{se}_{\bar{X}} \tag{3.13}$$

and

$$\mathrm{CV}_{\bar{X}} = \sigma \cdot n^{-1/2} \cdot \mu^{-1}. \tag{3.14}$$

An exact CI of level $1-2\alpha$ can be constructed by means of the Student's t distribution of $\bar{X}$ (von Storch and Zwiers 1999):

$$\mathrm{CI}_{\bar{X},1-2\alpha} = \left[\bar{X} + t_{n-1}(\alpha) \cdot S_{n-1} \cdot n^{-1/2}; \bar{X} + t_{n-1}(1-\alpha) \cdot S_{n-1} \cdot n^{-1/2}\right]. \tag{3.15}$$

$t_\nu(\beta)$ is the percentage point at β of the t distribution function with ν degrees of freedom (Section 3.9).

On the sample level, we write the estimated sample mean,

$$\widehat{\mu} = \bar{x} = \sum_{i=1}^{n} x(i)/n, \tag{3.16}$$

the estimated standard error,

$$\widehat{\mathrm{se}}_{\bar{x}} = \left\{ \sum_{i=1}^{n} \left[x(i) - \bar{x}\right]^2 / n^2 \right\}^{1/2}, \tag{3.17}$$

and the confidence interval,

$$\mathrm{CI}_{\bar{x},1-2\alpha} = \left[\bar{x} + t_{n-1}(\alpha) \cdot s_{n-1} \cdot n^{-1/2}; \bar{x} + t_{n-1}(1-\alpha) \cdot s_{n-1} \cdot n^{-1/2}\right], \tag{3.18}$$

where s_{n-1} is given by Eq. (3.25).

The performance of the CI in Eq. (3.18) for Gaussian white noise is analysed by means of a Monte Carlo simulation experiment. The CI performs excellent in coverage (Table 3.1), as expected from its exactness. The second CI property, length, decreases with data size. It can be further compared with CI lengths for other location measures.

3.1.2 Theoretical example: standard deviation estimation of Gaussian white noise

Consider the Gaussian white-noise process (Eq. 3.9) with unknown mean, and as estimator of σ the sample standard deviation, written on process level as

$$\widehat{\sigma} = S_{n-1} = \left\{ \sum_{i=1}^{n} \left[X(i) - \bar{X}\right]^2 / (n-1) \right\}^{1/2}. \tag{3.19}$$

Table 3.1. Monte Carlo experiment, mean estimation of a Gaussian purely random process. $n_{\text{sim}} = 4{,}750{,}000$ random samples of $\{X(i)\}_{i=1}^{n}$ were generated after Eq. (3.9) with $\mu = 1.0$, $\sigma = 2.0$ and various n values. An exact confidence interval $\text{CI}_{\bar{x},1-2\alpha}$ was constructed for each simulation after Eq. (3.18) with $\alpha = 0.025$. Average CI length, empirical $\text{RMSE}_{\bar{X}}$ and empirical coverage were determined subsequently. The entries are rounded.

n	$\text{RMSE}_{\bar{x}}^{\text{a}}$	*Nominal*[b]	⟨ *CI length* ⟩[c]	*Nominal*[d]	$\gamma_{\bar{x}}^{\text{e}}$	*Nominal*
10	0.6327	0.6325	2.7832	2.7832	0.9499	0.9500
20	0.4474	0.4472	1.8476	1.8476	0.9498	0.9500
50	0.2828	0.2828	1.1310	1.1310	0.9501	0.9500
100	0.2000	0.2000	0.7916	0.7917	0.9499	0.9500
200	0.1415	0.1414	0.5570	0.5571	0.9499	0.9500
500	0.0894	0.0894	0.3513	0.3513	0.9500	0.9500
1000	0.0633	0.0632	0.2482	0.2482	0.9499	0.9500

[a] Empirical $\text{RMSE}_{\bar{X}}$, given by $\left[\sum_{i=1}^{n_{\text{sim}}} (\bar{x} - \mu)^2 / n_{\text{sim}}\right]^{1/2}$.
[b] $\sigma \cdot n^{-1/2}$.
[c] Average value over n_{sim} simulations.
[d] $2 \cdot t_{n-1}(1-\alpha) \cdot \sigma \cdot c \cdot n^{-1/2}$, where c is given by Eq. (3.24).
[e] Empirical coverage, given by the number of simulations where $\text{CI}_{\bar{x},1-2\alpha}$ contains μ, divided by n_{sim}. Standard error of $\gamma_{\bar{x}}$ is (Efron and Tibshirani 1993) nominally $[2\alpha(1-2\alpha)/n_{\text{sim}}]^{1/2} = 0.0001$.

The properties of S_{n-1} are as follows:

$$\text{se}_{S_{n-1}} = \sigma \cdot \left(1 - c^2\right)^{1/2}, \tag{3.20}$$

$$\text{bias}_{S_{n-1}} = \sigma \cdot (c - 1), \tag{3.21}$$

$$\text{RMSE}_{S_{n-1}} = \sigma \cdot [2(1-c)]^{1/2} \tag{3.22}$$

and

$$\text{CV}_{S_{n-1}} = \left(1/c^2 - 1\right)^{1/2}, \tag{3.23}$$

where

$$c = [2/(n-1)]^{1/2} \cdot \Gamma(n/2) \,/\, \Gamma((n-1)/2). \tag{3.24}$$

On the sample level, we write

$$\widehat{\sigma} = s_{n-1} = \left\{ \sum_{i=1}^{n} [x(i) - \bar{x}]^2 / (n-1) \right\}^{1/2} \tag{3.25}$$

Table 3.2. Monte Carlo experiment, standard deviation estimation of a Gaussian purely random process. $n_{\mathrm{sim}} = 4{,}750{,}000$ random samples of $\{X(i)\}_{i=1}^{n}$ were generated after Eq. (3.9) with $\mu = 1.0$, $\sigma = 2.0$ and various n values. An exact confidence interval $\mathrm{CI}_{s_{n-1},1-2\alpha}$ was constructed for each simulation after Eq. (3.26) with $\alpha = 0.025$. Average CI length, empirical $\mathrm{RMSE}_{S_{n-1}}$ and empirical coverage were determined subsequently.

n	$\mathrm{RMSE}^{\mathrm{a}}_{s_{n-1}}$	*Nominal*[b]	$\langle$ *CI length* $\rangle$[c]	*Nominal*[d]	$\gamma^{\mathrm{e}}_{s_{n-1}}$	*Nominal*
10	0.4677	0.4677	2.2133	2.2133	0.9500	0.9500
20	0.3232	0.3233	1.3818	1.3819	0.9500	0.9500
50	0.2018	0.2018	0.8174	0.8174	0.9499	0.9500
100	0.1421	0.1420	0.5659	0.5659	0.9499	0.9500
200	0.1002	0.1002	0.3960	0.3960	0.9500	0.9500
500	0.0633	0.0633	0.2489	0.2489	0.9500	0.9500
1000	0.0447	0.0447	0.1757	0.1757	0.9501	0.9500

[a] Empirical $\mathrm{RMSE}_{S_{n-1}}$, given by $\left[\sum_{i=1}^{n_{\mathrm{sim}}} (s_{n-1} - \sigma)^2 / n_{\mathrm{sim}}\right]^{1/2}$.
[b] $\sigma \cdot [2(1-c)]^{1/2}$.
[c] Average value over n_{sim} simulations.
[d] $\left[\left(\chi^2_{n-1}(1-\alpha)\right)^{-1/2} - \left(\chi^2_{n-1}(\alpha)\right)^{-1/2}\right] \cdot \sigma \cdot c \cdot (n-1)^{1/2}$.
[e] Empirical coverage, given by the number of simulations where $\mathrm{CI}_{s_{n-1},1-2\alpha}$ contains σ, divided by n_{sim}. Standard error of $\gamma_{s_{n-1}}$ is nominally $[2\alpha(1-2\alpha)/n_{\mathrm{sim}}]^{1/2} = 0.0001$.

and use the chi-squared distribution of S^2_{n-1} (von Storch and Zwiers 1999) to find

$$\mathrm{CI}_{s_{n-1},1-2\alpha} = \left[s_{n-1} \left[(n-1) / \chi^2_{n-1}(\alpha) \right]^{1/2} ; \right.$$
$$\left. s_{n-1} \left[(n-1) / \chi^2_{n-1}(1-\alpha) \right]^{1/2} \right], \quad (3.26)$$

where $\chi^2_{\nu}(\beta)$ is the percentage point at β of the chi-squared distribution function with ν degrees of freedom (Section 3.9).

The performance of the CI in Eq. (3.26) for Gaussian white noise is analysed by means of a Monte Carlo simulation experiment. The CI performs excellent in coverage (Table 3.2), as expected from its exactness. The CI property length can be compared with CI lengths for other measures of spread or variation.

3.1.3 Real world

The two theoretical examples (Sections 3.1.1 and 3.1.2) presented convenient settings. $X(i)$ was normally distributed and persistence was absent, for which reasons the spacing was not relevant. The simple estimators $\widehat{\mu}$ and $\widehat{\sigma}$ could then be applied for mean and standard deviation estimation, which allowed to deduce their distributions as Student's t and

Table 3.3. Monte Carlo experiment, mean and median estimation of a lognormal purely random process. $n_{\mathrm{sim}} = 4{,}750{,}000$ random samples of $\{X(i)\}_{i=1}^{n}$ were generated after $X(i) = \exp\left[\mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(i)\right]$, $i = 1, \ldots, n$, with $\mu = 1.0$, $\sigma = 1.0$ and various n values. The density function is skewed (Fig. 3.2). Analysed as estimators of the centre of location of the distribution were the sample mean (Eq. 3.16) and the sample median, $\widehat{m}$ (see background material, Section 3.8). $\mathrm{CI}_{\bar{x},1-2\alpha}$ was constructed after Eq. (3.18) with $\alpha = 0.025$.

n	$\mathrm{RMSE}_{\widehat{m}}$	$\mathrm{RMSE}_{\bar{x}}$	$\gamma_{\bar{x}}^{\mathrm{a}}$	*Nominal*	C^{b}
10	1.1647	1.8575	0.8392	0.9500	−0.1108
20	0.7893	1.3140	0.8670	0.9500	−0.0830
50	0.4884	0.8309	0.8991	0.9500	−0.0509
100	0.3430	0.5880	0.9170	0.9500	−0.0330
200	0.2418	0.4155	0.9296	0.9500	−0.0204
500	0.1526	0.2627	0.9399	0.9500	−0.0101
1000	0.1078	0.1858	0.9442	0.9500	−0.0058

[a] Standard error of $\gamma_{\bar{x}}$ is nominally 0.0001.
[b] Empirical coverage error of $\mathrm{CI}_{\bar{x},1-2\alpha}$, given by $\gamma_{\bar{x}}$ minus nominal value.

chi-squared, respectively. Finally, exact CIs were obtained using the percentage points of the distributions of the estimators.

In the real climatological world, however, such simple assumptions regarding distributional shape, persistence and spacing cannot be expected to be fulfilled (Chapter 1). In the practical setting, further questions than just after mean and standard deviation are asked, leading to more complex parameters, θ. The major part of the rest of this book is devoted to such problems. Also the estimators of those parameters have commonly more complex distributions, $f(\widehat{\theta})$.

Example 3 (Table 3.3) goes a small step from the theoretical in the direction of the real world. This case illustrates the effects of violations of the distributional assumption. Example 3 assumes that $X(i)$ are Gaussian distributed, although the prescribed true distribution is lognormal. This leads to a Student's t CI with an empirical coverage that deviates from the nominal value by several standard errors (Table 3.3). The difference is the coverage error (see next paragraph), its absolute value decreases with the data size. This CI is not exact but only approximate. Table 3.4 summarizes theoretical and practical settings.

Coverage error, C, is defined by means of a single-sided CI endpoint (Efron and Tibshirani 1993), for example,

$$C = \gamma_l - \alpha. \tag{3.27}$$

If C decreases with sample size as $\mathcal{O}\left(n^{-1/2}\right)$, that means, if C is composed of terms of powers of $1/n$ that are greater than or equal to $1/2$,

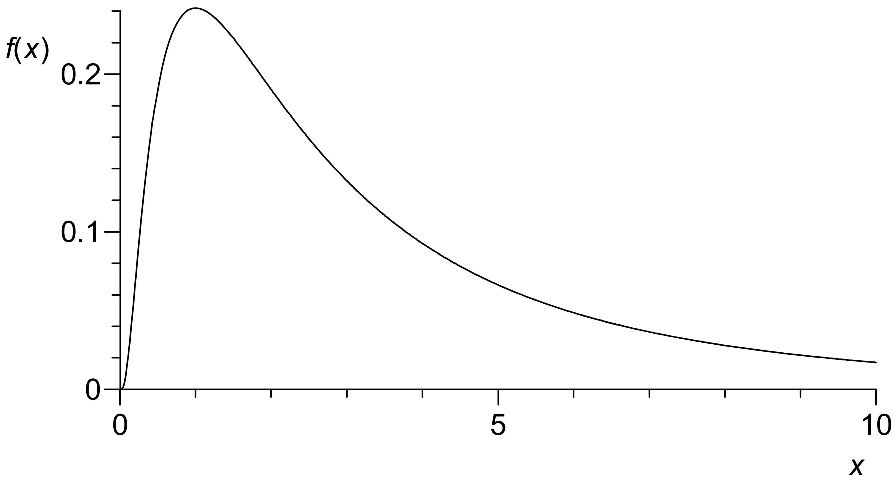


Figure 3.2. Lognormal density function from Example 3 (Table 3.3), with $\mu = 1.0$ and $\sigma = 1.0$. The expression for $f(x)$ is given by Eq. (3.61).

then the CI is called first-order accurate; if C is of $\mathcal{O}\left(n^{-1}\right)$, then the CI is called second-order accurate; and so forth. The same CI accuracy applies also to two-sided CIs. Desirable approximate CIs have a high-order accuracy. Coverage accuracy is the major criterion employed in this book for assessing the quality of a CI. As a second property we consider interval length, $\widehat{\theta}_{\mathrm{u}} - \widehat{\theta}_{\mathrm{l}}$, which is ideally small. Related to CI accuracy is CI correctness (Efron and Tibshirani 1993: Section 22.2 therein), which refers to the difference between an exact CI endpoint (which has $C = 0$) and an approximate CI endpoint, expanded in terms of powers of n.

For practical situations it is conceivable that different estimators, $\widehat{\theta}_1$ and $\widehat{\theta}_2$, of the same parameter, θ, exist. Consider for example parameter estimation of the AR(p) model, for which Priestley (1981: Section 5.4.1 therein) gives four sets of estimators, namely exact likelihood, least squares, approximate least squares and Yule–Walker. Each estimator has its own properties such as standard error, bias, RMSE, CI length or CI coverage accuracy.

An important attribute of an estimator is robustness, which means that the $\widehat{\theta}$ properties depend only weakly on made assumptions (shape, persistence and spacing). Robust estimators perform better (e.g., have smaller RMSE or higher coverage accuracy) than non-robust in non-ideal situations. Example 3 shows that the sample median as an estimator of the centre of location of a distribution is more robust (with regard to $\mathrm{RMSE}_{\widehat{\theta}}$) than the mean. In essence, because of the complexity of the setting in the real world and the dependence on the situation and the aims of the analysis, there is no general rule how to construct best an estimator. It has something of an art, which is not meant negatively. In this light, the growth of climatological knowledge does not only depend on more and better data but also on improved methods to analyse them.

Table 3.4 shows also how real-world climatological estimation problems may be tackled. The classical approach comes from theory and aims to extend the applicability by introducing countermeasures. Regarding distributional shape, a measure may be to estimate the shape of the noise data (Section 1.6). Then one looks up and applies the estimator for the parameter θ that performs for this particular shape best in terms of a user-specified property, say RMSE. The CI follows from the estimator's distribution. The problem is that only for simple shapes and parameters, knowledge is available that would allow this procedure. (In this regard, the lognormal without is clearly simpler than the lognormal with shift parameter (Section 3.8).) Transformations of the data, such that the noise part has a simple shape, can also be tried, but then the problem is that the systematic part of the model (Eq. 1.2) can take intractable forms, see Atkinson and Cox (1988) on this dilemma. (The double-logarithmic transformation described in Section 2.6 was in the converse direction. It produced a simpler systematic part and a more complex noise part.)

Regarding persistence, the effective data size, n', can be used instead of n for CI calculation. The problem here is that n' depends on the persistence model and on which estimator is used (Chapter 2). One may take n'_μ (Eq. 2.7), n'_{σ^2} (Eq. 2.36) or n'_ρ (Eq. 2.38) for the AR(1) process and hope that deviations to the problem at hand are small. Regarding spacing, it is fair to say that the classical approach mostly ignores unevenness because its influence on n' and the distribution of $\widehat{\theta}$ can in the general case not be deduced. As a result, the classical approach often contents itself with approximate normality, that is, with $f(\widehat{\theta})$ approaching normal shape as $n \to \infty$. For many theoretical estimations, approximate normality can be proven. However, the point is that in practice n is limited and it is mostly unknown how accurate the normal approximation of the CI is.

3.2 Bootstrap principle

Table 3.4 lists also the bootstrap approach to solve practical estimation problems. These tasks include constructing CIs for estimators more complex than the mean, and this in the presence of non-normal distribu-

Table 3.4. Estimation settings (theoretical and practical) and approaches (classical and bootstrap) to solve practical problems.[a]

	Distributional shape	*Persistence*	*Spacing*	*Estimator*, $\widehat{\theta}$	*Distribution of* $\widehat{\theta}$, $f(\widehat{\theta})$	*Confidence interval*, $\mathrm{CI}_{\widehat{\theta},1-2\alpha}$
Setting						
Theoretical[b]	Known, normal	No (yes)	Not relevant (even)	Tractable	Deducible	Exact
Example 1	Normal	No	Not relevant	$\widehat{\mu}$ (Eq. 3.10)	t distribution	Exact
Example 2	Normal	No	Not relevant	$\widehat{\sigma}$ (Eq. 3.19)	χ^2 distribution	Exact
Practical	Non-normal	Yes	Uneven	More complex than $\widehat{\mu}$ or $\widehat{\sigma}$	Often not deducible	Exact only if $f(\widehat{\theta})$ deducible
Example 3	Lognormal	No	Not relevant	$\widehat{\mu}$ (Eq. 3.10), $\widehat{m}$	Ignored	Student's t approximation
Approach						
Classical	Find shape, apply suitable $\widehat{\theta}$ or transform x	Effective data size	Ignore		Assume normality	Approximate, based on assumptions
Bootstrap	Not very relevant[c]	Block bootstrap or parametric	Not relevant[d]		Not very relevant[c]	Approximate, based on fewer assumptions

[a] Indicated are the main lines of settings and approaches. Exceptions exist; for example, theory deals also with uneven spacing (Parzen 1984).
[b] Theory must necessarily impose restrictions to shape, persistence and spacing to obtain tractable problems.
[c] Distributional properties can influence bootstrap CI accuracy (see text), but this is a minor effect.[a]
[d] Restriction: Parametric persistence models more complex than AR(1) are not considered for uneven spacing because of the embedding problem (Section 2.3).

tions, persistence and uneven spacing. The main idea of the bootstrap is to use the data to mimic the unknown distribution function, which is now replaced by the empirical distribution function (Eq. 3.43). Mimicking the data generating process is achieved by drawing random samples from the data set. The simplest form is the ordinary bootstrap, that means, drawing one by one with replacement. Preserving the persistence properties of time series data requires adaptions of the ordinary bootstrap, which are explained in Section 3.3. Re-applying the estimation procedure to the new random samples, called resamples, yields new estimates, called replications. Section 3.4 explains CI construction using the replications. Figure 3.3 shows the bootstrap principle and the workflow. It gives also a simple bootstrap CI variant (bootstrap normal CI).

The bootstrap means that numerical simulation replaces theoretical derivation of the distribution of an estimator. This can be an improvement, especially if the complexity of the problem defies obtaining an exact theoretical result. However, also the bootstrap is not free of assumptions. The main requirement is that the properties distributional shape and persistence are preserved by the bootstrap resampling. There is also "simulation noise," but this can be made arbitrarily small by using a large number of resamples, B. Assumptions made at CI construction add to the fact that in complex situations, bootstrap CIs, like classical CIs, are not exact but approximate. In complex cases, for small sample size, non-smooth functionals such as the median and without underlying theory, even the bootstrap may fail to yield acceptable results (LePage and Billard 1992). However, bootstrap CIs seem to be more flexible and require less strict assumptions than classical CIs (Table 3.4). A word on usage of "simulation:" henceforth we reserve this for Monte Carlo experiments, where statistical methods are tested by means of artificial data from models with pre-defined properties. The bootstrap procedure, on the other hand, is referred to as "resampling."

3.3 Bootstrap resampling

The ordinary bootstrap, resampling one by one with replacement, is a nonparametric method because it can virtually be applied to data from any continuous PDF without involvement of distributional parameters. By resampling one by one, the serial dependence in $\{X(i)\}_{i=1}^{n}$ is lost. For the analysis of time series, the ordinary bootstrap has therefore to be adapted to take serial dependence into account. This can be done nonparametrically, by resampling block by block of data. Alternatively,

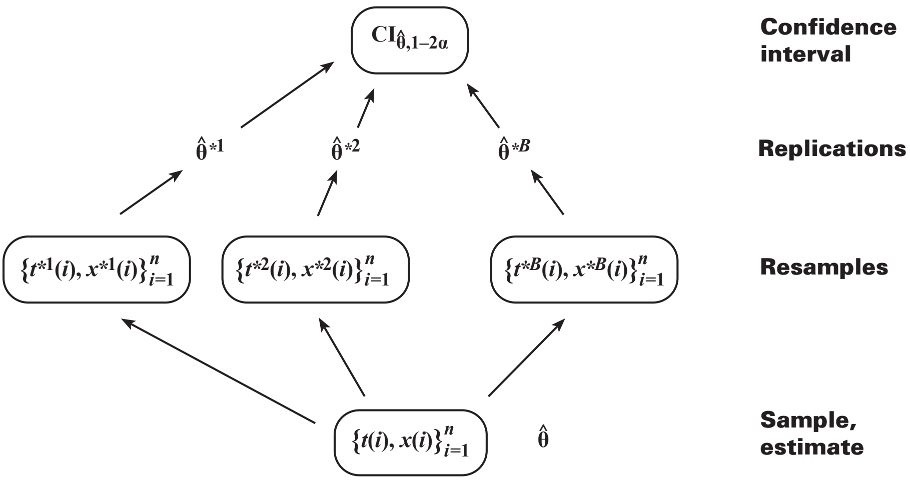


Figure 3.3. Bootstrap principle for constructing confidence intervals. Given is a sample of data and an estimate of a parameter of interest. Using bootstrap resampling (Section 3.3) new data sets—resamples—are formed. The resamples ideally preserve fully the statistical properties of the process that generated the data. For convenience of presentation, we assume that this process (Eq. 1.2) consists only of the noise part; the following chapters analyse bootstrap resampling where the model has also a systematic part. In the simple case where $t(i)$ are perfectly known and also persistence is absent, $t^*(i) = t(i), i = 1, \ldots, n$, and $\{x^*(i)\}_{i=1}^{n}$ is obtained by drawing randomly, one by one and with replacement, n elements from the set of sample values, $\{x(i)\}_{i=1}^{n}$. The resamples are marked with an asterisk and numbered with an index, $b = 1, \ldots, B$. The number of resamples, B, is typically a few thousand. The estimator is applied to each of the resamples, yielding B new estimates—the replications. The set of replications $\{\widehat{\theta}^{*b}\}_{b=1}^{B}$ is then used for CI construction. Several methods exist for that purpose (Section 3.4), which can, for example, correct for estimation bias. In the simple case of normal bootstrap confidence intervals, henceforth denoted briefly as normal CIs, $\mathrm{CI}_{\widehat{\theta},1-2\alpha} = [\widehat{\theta} + z(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}; \widehat{\theta} - z(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}]$, where $\widehat{\mathrm{se}}_{\widehat{\theta}^*}$ is the sample standard error of the replications, denoted as estimated bootstrap standard error, and $z(\alpha)$ is the percentage point of the normal distribution (Section 3.9).

persistence can be modelled. The preferred model in the case of climate time series is the AR(1) process (Chapter 2).

For convenience of presentation, this chapter omits the effects of errors in the timescale, $t(i)$, that means, it sets $t^*(i) = t(i), i = 1, \ldots, n$, or briefly $\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$. Bootstrap adaptions for solving estimation problems associated with an uncertain timescale, which are relevant for climatology, seem not to have been developed yet in the statistical literature. The subsequent chapters present some possible bootstrap adaptions. These are steps into new territory.

3.3.1 Nonparametric: moving block bootstrap

The moving block bootstrap algorithm, denoted as MBB, divides the time series values $\{x(i)\}_{i=1}^{n}$ into sequences or blocks of l consecutive points (Algorithm 3.1). The blocks may overlap, their number is $n-l+1$. MBB draws randomly a block and inserts the contained values as the first l resample values, $\{x^*(i)\}_{i=1}^{l}$. The next randomly drawn block yields $\{x^*(i)\}_{i=l+1}^{2l}$, and so forth. When the last point, $x^*(n)$, has been inserted, the algorithm stops; remaining block values are discarded. The resampled times are unchanged (Algorithm 3.1). One indexes the first resample as $\left\{t^{*1}(i), x^{*1}(i)\right\}_{i=1}^{n}$ and repeats MBB until B resamples exist.

A possible adaption of the MBB to uneven spacing is introduced in Section 3.3.1.2. Other nonparametric bootstrap algorithms are described briefly in the background material (Section 3.8).

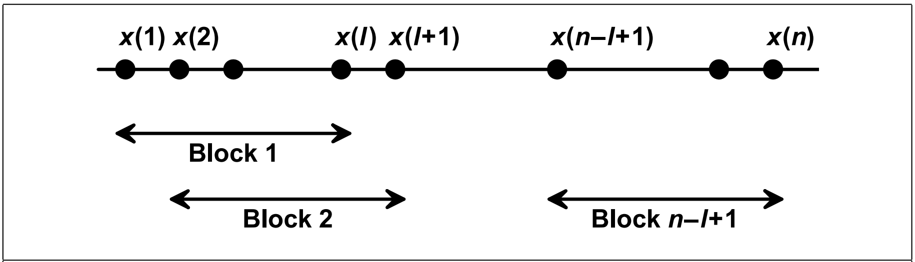
3.3.1.1 Block length selection

Selection of the block length, l, is a crucial step because it determines properties like bootstrap standard error or bootstrap CI coverage accuracy. Berkowitz and Kilian (2000: p. 20 therein) describe the trade-off problem involved as follows: "As the block size becomes too small, the [MBB] destroys the time dependency of the data and its average accuracy will decline. As the block size becomes too large, there are few blocks and [resamples] will tend to look alike. As a result, the average accuracy of the [MBB] also will decline. This suggests that there exists an optimal block size [l_{opt}] which maximizes accuracy."

A simple block length selector can be derived from Sherman et al. (1998), who adapted a formula from Carlstein (1986), to the MBB:

$$l_{\text{opt}} = NINT\left\{\left[6^{1/2} \cdot \widehat{\bar{a}} \Big/ \left(1 - \widehat{\bar{a}}^2\right)\right]^{2/3} \cdot n^{1/3}\right\}, \qquad (3.28)$$

where $NINT(\cdot)$ is the nearest integer function and $\widehat{\bar{a}} = \exp(-\bar{d}/\widehat{\tau})$ is the estimated "equivalent autocorrelation coefficient" (Fig. 2.3) of an AR(1) process fitted to the data with uneven spacing. (If $\widehat{\bar{a}} \to 0$ and $\widehat{\bar{a}} \to 1$, then take $l_{\text{opt}} = 1$ and $l_{\text{opt}} = n-1$, respectively.) In the case of even spacing, $\widehat{\bar{a}}$ can be taken from Eq. (2.4). Instead of $\widehat{\bar{a}}$, also a bias-corrected version, $\widehat{\bar{a}}'$, can be used, see Section 2.6. Employing this block length selector for real-world problems is evidently a simplification because it was developed for normal shape, AR(1) persistence, even spacing and bootstrap standard error estimation. Hall et al. (1995a) show that for bootstrap CI estimation, l_{opt} should increase at a slower rate with n. On the other hand, in practice some simplification is inevitable, and the formula might yield acceptable results. This can be assessed by



Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Resampled times unchanged	$\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$
Step 3	Blocks j (see above)	$\{x(i)\}_{i=j}^{j+l-1}, \quad j = 1, \ldots, n-l+1$
Step 4	Set counter	$c = 1$
Start resampling		
Step 5	Draw random block j^*	$j^* \in \{1, \ldots, n-l+1\}$
Step 6	Insert block data	$\{x^*(i)\}_{i=c}^{c+l-1} = \{x(i)\}_{i=j^*}^{j^*+l-1}$
	If $x^*(n)$ has been inserted	Stop inserting and exit
Step 7	Increase counter	$c \rightarrow c + l$
Step 8	Go to Step 5	
End resampling		

Algorithm 3.1. Moving block bootstrap algorithm (MBB). Note: An equation like $\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$ is used to denote $t^*(i) = t(i), i = 1, \ldots, n$.

means of Monte Carlo simulations of real-world conditions, as is done in subsequent parts of this book.

Bühlmann and Künsch (1999) presented a fully data-driven block length selector (Algorithm 3.2). They showed the equivalence of l_{opt} selection and smoothing in spectral estimation (Chapter 5).

Berkowitz and Kilian (2000) presented a brute-force block length selector:

1 Approximate the data generating process by a parametric model (e.g., ARMA).

2 Generate Monte Carlo samples from this fitted model.

Step 1 Calculate $\{Y(i)\}_{i=1}^{n} = \left\{\widehat{\mathrm{IF}}\,(X(i))\right\}_{i=1}^{n}$, where

$$\widehat{\mathrm{IF}}\,(X(j)) = n \cdot \left(\widehat{\theta} - \widehat{\theta}_{(j)}\right)$$

Step 2 Calculate $\widehat{R}(h) = n^{-1} \sum_{i=1}^{n-|h|} Y(i) \cdot Y(i+|h|), \qquad h = -n+1, \ldots, n-1$

Step 3 Calculate iteratively:

$$b_0 = n^{-1},$$

$$b_k = n^{-1/3} \left[\left(\sum_{h=-n+1}^{n-1} \widehat{R}(h)^2 \right) \times \left(6 \sum_{h=-n+1}^{n-1} w_{\mathrm{SC}}(h \cdot b_{k-1} \cdot n^{4/21})^2 \cdot h^2 \cdot \widehat{R}(h)^2 \right)^{-1} \right]^{1/3},$$

$$k = 1, 2, 3, 4,$$

$$\widehat{b} = n^{-1/3} \cdot (2/3)^{1/3} \left[\left(\sum_{h=-n+1}^{n-1} w_{\mathrm{TH}}(h \cdot b_4 \cdot n^{4/21}) \cdot \widehat{R}(h) \right) \times \left(\sum_{h=-n+1}^{n-1} w_{\mathrm{SC}}(h \cdot b_4 \cdot n^{4/21}) \cdot |h| \cdot \widehat{R}(h) \right)^{-1} \right]^{2/3}$$

Step 4 Set $l_{\mathrm{opt}} = NINT(\widehat{b}^{-1})$

Algorithm 3.2. Block length selector after Bühlmann and Künsch (1999). Notes: $\widehat{\mathrm{IF}}\,(X(i))$ is the estimated influence function (Efron and Tibshirani 1993: Section 21.3 therein). $\widehat{\theta}_{(j)}$ is the delete-one, jackknife value of $\widehat{\theta}$, that is, the $\widehat{\theta}$ value calculated from the data with the jth point removed, see Section 3.4.4. w_{SC} is the split-cosine window; $w_{\mathrm{SC}}(z) = 1$ for $|z| \leq 0.8$, $w_{\mathrm{SC}}(z) = [1 + \cos(5(z - 0.8)\pi)]/2$ for $0.8 < |z| \leq 1$ and $w_{\mathrm{SC}}(z) = 0$ for $|z| > 1$. w_{TH} is the Tukey–Hanning window; $w_{\mathrm{TH}}(z) = [1 + \cos(\pi z)]/2$ for $|z| \leq 1$ and $w_{\mathrm{TH}}(z) = 0$ for $|z| > 1$.

3 Select the parameter of interest, θ, and an estimation property of interest, say, bootstrap CI accuracy.

4 Prescribe a search grid. For example, l_{search} runs from a start to an end value with some spacing.

5 Calculate the empirical bootstrap CI coverage error (or another property) using the Monte Carlo samples and MBB with l_{search}.

6 Select l_{search} with best performance.

Other block length selectors are described briefly in the background material (Section 3.8).

3.3.1.2 Uneven spacing

Applying the MBB to unevenly spaced time series increases the estimation uncertainty because the time spacing values within the inserted block, $\{d(i)\}_{i=j^*}^{j^*+l-2}$, need not equal the spacing values at the insertion place, $\{d^*(i)\}_{i=c}^{c+l-2}$. This may reduce the ability to preserve serial dependence.

An attempt to adapt MBB to this situation could be to resample only blocks with spacing similar to the spacing at the insertion place. For example, only the β% blocks with nearest spacing could be made drawable. The unevenness in a block could be quantified by the coefficient of variation of the spacing, CV_d, similarly as was done in Fig. 2.3. In the case of equidistance, one would have $\mathrm{CV}_d = 0$ and take $\beta = 100\%$, that means, one would use MBB. It is, however, unclear which β value to take for $\mathrm{CV}_d > 0$. A second measure could be to decrease l when reducing the number of drawable blocks.

A Monte Carlo experiment (Section 3.8) tested a rather simple MBB adaption: $\beta = 50\%$ for $\mathrm{CV}_d > 0$. This was applied to mean estimation of a Gaussian AR(1) process. It turned out, however, that the accuracy of the BCa CI was lower compared to usage of the ordinary MBB under the same block length selector (Eq. 3.28). More Monte Carlo studies of β choices in dependence on CV_d and other spacing properties have to be carried out to find more accurate MBB adaptions to uneven spacing.

The practical conclusion is that for small CV_d and large deviations from AR(1) persistence, one may use MBB. On the other hand, large CV_d and minor deviations from the AR(1) model indicate to employ the parametric autoregressive bootstrap (next section). This resampling method could have a higher relevance than MBB for practical applications because the AR(1) persistence model is generally a suitable first-order approximation for weather and climate time series (Chapter 2). Such a combined approach should yield acceptable results also for small data sizes. For that purpose, we tend to prefer the ARB over the MBB resampling type on the basis of the Monte Carlo experiments of mean estimation (Tables 3.5 and 3.7). If CV_d is large and also the deviations from AR(1) dependence are large, both the MBB and the parametric autoregressive bootstrap should be tried and results compared. This difference should indicate the size of the difference of the approximate bootstrap CIs to the exact CI.

3.3.1.3 Systematic model parts and nonstationarity

For explaining the bootstrap principle (Fig. 3.3), we assumed for convenience of presentation $x(i) = x_{\mathrm{noise}}(i)$. Realistic climate processes contain more parts, like trend, outliers and variability (Eq. 1.2). The MBB can be applied to such processes by resampling from the residuals.

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Resampled times unchanged	$\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$
Step 3	Residuals (Eq. 3.29)	$r(i) = [x(i) - \widehat{x}_{\text{trend}}(i) - \widehat{x}_{\text{out}}(i)] \Big/ \widehat{S}(i)$
Step 4	Apply MBB (Algorithm 3.1) to residuals	$\{r(i)\}_{i=1}^{n}$
Step 5	Resampled residuals	$\{r^*(i)\}_{i=1}^{n}$
Step 6	Use resampled residuals to produce resamples	$x^*(i) = \widehat{x}_{\text{trend}}(i) + \widehat{x}_{\text{out}}(i) + \widehat{S}(i) \cdot r^*(i)$

Algorithm 3.3. MBB for realistic climate processes, which comprise trend, outlier and variability components.

Plugging in the estimates into the climate equation (Eq. 1.2) yields

$$r(i) = [x(i) - \widehat{x}_{\text{trend}}(i) - \widehat{x}_{\text{out}}(i)] \Big/ \widehat{S}(i), \qquad i = 1, \ldots, n, \tag{3.29}$$

where $\widehat{x}_{\text{trend}}(i)$, $\widehat{x}_{\text{out}}(i)$ and $\widehat{S}(i)$ are estimated trend, outlier and variability components, respectively. The following chapters explain such estimations. The residuals, $r(i)$, are realizations of the noise process. (Analogously, the residuals, $\epsilon(i)$, in Chapter 2 are realizations of a white-noise process.) The MBB for realistic climate processes is listed as Algorithm 3.3.

The trend, outlier and variability components allow to describe nonstationary climate processes. A further type of nonstationarity regards persistence. Consider as example ice-volume fluctuations over the past 4 Ma. In the early part (Pliocene), the persistence was weaker than in the late part (Pleistocene), when huge continental ice-sheets had been built up (Mudelsee and Raymo 2005). Such nonstationarity can be accounted for by the local block bootstrap (Paparoditis and Politis 2002), where, in the example, Pliocene resamples, $x^*(i)$, are restricted to come from the Pliocene data, $x(i)$, analogously for Pleistocene resamples. The local block bootstrap could also be applied, as an alternative to using MBB and the residuals, to produce nonparametric trend and variability estimates with CIs (Bühlmann 1998). The cited paper applies smoothing to an ozone time series from Switzerland, 1932–1996. Evidently, the size of the locality region should be chosen taking prior knowledge about the data generating process into account.

3.3.2 Parametric: autoregressive bootstrap

The autoregressive bootstrap algorithm (ARB) is the ordinary bootstrap applied to the white-noise residuals, $\epsilon(i)$. We first take the residuals, $r(i)$, from the climate equation as in Eq. (3.29). Using the persistence model for $r(i)$, the residuals $\epsilon(i)$ are then formed. $\epsilon(i)$ are treated as realizations of a white-noise process, see Eq. (2.5). We employ the AR(1) persistence model as a suitable description for climate processes (Chapter 2). Advantageously, the distributional shape need not be Gaussian. Even and uneven spacing are treated separately.

3.3.2.1 Even spacing

The ARB for even spacing is listed as Algorithm 3.4. Although the bias correction (Step 7) is only approximate (Section 2.6), this is considered an important step because ignoring bias can lead to a bad bootstrap performance (Stine 1987). Scaling, as done in Step 8 using a factor $\left[1 - (\widehat{a}')^2\right]^{-1/2}$, is non-standard. It has the computational advantage that no transient behaviour is required in Step 11. Centering (Step 9) achieves that the resample generating process has expectation zero, as the white-noise process is supposed to have. After Step 9, a further scaling with a factor $[(n-1)/(n-2)]^{1/2}$ (Stine 1987) is omitted. This factor is in the general case only approximate (Peters and Freedman 1984) and its effect is considered negligible compared with the other uncertainties. Lahiri (2003) explains the "traditional" method to generate a number of samples that is very much larger than n at Step 10 and use those at Step 11 for extracting $r^*(i)$ from the transient sequence. The advantage of the non-standard formulation (Step 8) corresponds to the advantage of strict stationarity of the non-standard formulation of the AR(1) model (Chapter 2).

3.3.2.2 Uneven spacing

The ARB for uneven spacing is listed as Algorithm 3.5. It corresponds basically to the ARB for even spacing, where the persistence parameter, a, is replaced by $\exp\{-[t(i) - t(i-1)]/\tau\}$. Bias correction for $\widehat{\tau}$ at Step 7 goes via $\widehat{\bar{a}}' = \exp(-\bar{d}/\widehat{\tau}')$.

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Resampled times unchanged	$\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$
Step 3	Estimated trend, outliers, variability	$\{\widehat{x}_{\text{trend}}(i)\}_{i=1}^{n}$, $\{\widehat{x}_{\text{out}}(i)\}_{i=1}^{n}$, $\left\{\widehat{S}(i)\right\}_{i=1}^{n}$
Step 4	Climate equation residuals (Eq. 3.29)	$\{r(i)\}_{i=1}^{n}$
Step 5	Assume $\{r(i)\}_{i=1}^{n}$ to come from AR(1) model for even spacing (Eq. 2.1)	
Step 6	Estimate AR(1) parameter (Eq. 2.4)	$\widehat{a}$
Step 7	Bias correction	$\widehat{a}'$
Step 8	White-noise residuals	$\epsilon(i) = [r(i) - \widehat{a}' \cdot r(i-1)] \times \left[1 - (\widehat{a}')^2\right]^{-1/2}$, $i = 2, \ldots, n$
Step 9	Centering	$\tilde{\epsilon}(i) = \epsilon(i) - \sum_{i=2}^{n} \epsilon(i)/(n-1)$
Step 10	Draw $\tilde{\epsilon}^*(j)$, $j = 2, \ldots, n$, with replacement from	$\{\tilde{\epsilon}(i)\}_{i=2}^{n}$
Step 11	Resampled climate residuals	$r^*(1)$ drawn from $\{r(i)\}_{i=1}^{n}$, $r^*(i) = \widehat{a}' \cdot r^*(i-1) + \left[1 - (\widehat{a}')^2\right]^{1/2} \cdot \tilde{\epsilon}^*(i)$, $i = 2, \ldots, n$
Step 12	Resampled data	$x^*(i) = \widehat{x}_{\text{trend}}(i) + \widehat{x}_{\text{out}}(i) + \widehat{S}(i) \cdot r^*(i)$, $i = 1, \ldots, n$

Algorithm 3.4. Autoregressive bootstrap algorithm (ARB), even spacing.

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Resampled times unchanged	$\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$
Step 3	Estimated trend, outliers, variability	$\{\widehat{x}_{\text{trend}}(i)\}_{i=1}^{n}$, $\{\widehat{x}_{\text{out}}(i)\}_{i=1}^{n}$, $\left\{\widehat{S}(i)\right\}_{i=1}^{n}$
Step 4	Climate equation residuals (Eq. 3.29)	$\{r(i)\}_{i=1}^{n}$
Step 5	Assume $\{r(i)\}_{i=1}^{n}$ to come from AR(1) model for uneven spacing (Eq. 2.9)	
Step 6	Estimate persistence time (Eq. 2.11)	$\widehat{\tau}$
Step 7	Bias correction	$\widehat{\tau}'$
Step 8	Abbreviation	$\widehat{a}'(i) = \exp\{-[t(i) - t(i-1)]/\widehat{\tau}'\}$, $i = 2, \ldots, n$
Step 9	White-noise residuals	$\epsilon(i) = [r(i) - \widehat{a}'(i) \cdot r(i-1)] \times \left\{1 - [\widehat{a}'(i)]^2\right\}^{-1/2}$, $i = 2, \ldots, n$
Step 10	Centering	$\tilde{\epsilon}(i) = \epsilon(i) - \sum_{i=2}^{n} \epsilon(i)/(n-1)$
Step 11	Draw $\tilde{\epsilon}^*(j)$, $j = 2, \ldots, n$, with replacement from	$\{\tilde{\epsilon}(i)\}_{i=2}^{n}$
Step 12	Resampled climate residuals	$r^*(1)$ drawn from $\{r(i)\}_{i=1}^{n}$, $r^*(i) = \widehat{a}'(i) \cdot r^*(i-1) + \left\{1 - [\widehat{a}'(i)]^2\right\}^{1/2} \times \tilde{\epsilon}^*(i)$, $i = 2, \ldots, n$
Step 13	Resampled data	$x^*(i) = \widehat{x}_{\text{trend}}(i) + \widehat{x}_{\text{out}}(i) + \widehat{S}(i) \cdot r^*(i)$, $i = 1, \ldots, n$

Algorithm 3.5. Autoregressive bootstrap algorithm (ARB), uneven spacing.

3.3.3 Parametric: surrogate data

The surrogate data approach (Algorithm 3.6), related to ARB, is a simulation rather than a resampling method. No residuals are drawn as in the ARB. Instead, climate equation residuals $\{r^*(i)\}_{i=1}^{n}$ are obtained by numerical simulation (Step 8) from the persistence model with estimated (and bias-corrected) parameters. Because also the distributional shape is specified, the surrogate data approach is bounded stronger by parametric restrictions than the ARB. Therein lies its danger: it is more prone than the ARB to systematic errors from violated assumptions.

3.4 Bootstrap confidence intervals

Estimation of θ is repeated for the resamples, $\{t^{*b}(i), x^{*b}(i)\}_{i=1}^{n}, b = 1, \ldots, B$. This yields the bootstrap replications, $\{\widehat{\theta}^{*b}\}_{b=1}^{B}$. The replications are used to construct equi-tailed $(1 - 2\alpha)$ confidence intervals, $\mathrm{CI}_{\widehat{\theta}, 1-2\alpha}$, see Fig. 3.3.

Two approaches, standard error based and percentile based, dominate theory and practice of bootstrap CI construction. The estimated bootstrap standard error is the sample standard error of the replications,

$$\widehat{\mathrm{se}}_{\widehat{\theta}^*} = \left\{ \sum_{b=1}^{B} \left[\widehat{\theta}^{*b} - \left\langle \widehat{\theta}^{*b} \right\rangle \right]^2 \Big/ (B-1) \right\}^{1/2}, \tag{3.30}$$

where $\left\langle \widehat{\theta}^{*b} \right\rangle = \sum_{b=1}^{B} \widehat{\theta}^{*b} / B$. The percentiles result from the empirical distribution function (Eq. 3.43) of the replications. The accuracy of bootstrap CIs depends critically on the similarity (in terms of standard errors or percentiles) of the distribution of the bootstrap replications and the true distribution, $f(\widehat{\theta})$. Various concepts exist for accounting for the deviations between the two distributions.

Suppressing "simulation noise" requires more resamples for percentile estimation than for bootstrap standard error estimation. This book follows the recommendation of Efron and Tibshirani (1993), and sets throughout $B = 2000$ (or 1999 for percentile CIs). For a reasonable α value such as 0.025, this means that a number of 50 replications are outside the percentile bound. An own simulation study, analysing the coefficient of variation of a CI endpoint in dependence of B, confirmed that this choice is sufficient also in a bivariate setting (Mudelsee and Alkio 2007).

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Resampled times unchanged	$\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$
Step 3	Estimated trend, outliers, variability	$\{\widehat{x}_{\text{trend}}(i)\}_{i=1}^{n}$, $\{\widehat{x}_{\text{out}}(i)\}_{i=1}^{n}$, $\left\{\widehat{S}(i)\right\}_{i=1}^{n}$
Step 4	Climate equation residuals (Eq. 3.29)	$\{r(i)\}_{i=1}^{n}$
Step 5	Assume $\{r(i)\}_{i=1}^{n}$ to come from specific model (shape, persistence)	
Step 6	Estimate model parameters	
Step 7	Bias correction	
Step 8	Simulate climate equation residuals from estimated model	$\{r^*(i)\}_{i=1}^{n}$
Step 9	Simulated data	$x^*(i) = \widehat{x}_{\text{trend}}(i) + \widehat{x}_{\text{out}}(i) + \widehat{S}(i) \cdot r^*(i)$, $i = 1, \ldots, n$

Algorithm 3.6. Surrogate data approach.

3.4.1 Normal confidence interval

The bootstrap normal confidence interval, already given in Fig. 3.3, is

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta} + z(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}; \widehat{\theta} - z(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}\right], \tag{3.31}$$

where $z(\alpha)$ is the percentage point of the normal distribution (Section 3.9).

3.4.2 Student's *t* confidence interval

The bootstrap Student's t confidence interval is

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta} + t_{\nu}(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}; \widehat{\theta} - t_{\nu}(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}\right], \tag{3.32}$$

where $t_{\nu}(\alpha)$ is the percentage point of the t distribution function with ν degrees of freedom (Section 3.9). It is in practice presumably always more accurate to prefer, as this book does, Student's t CIs over normal CIs because they recognize the reduction of degrees of freedom. (For data sizes above, say, 30, the difference becomes negligible.)

3.4.3 Percentile confidence interval

The bootstrap percentile confidence interval is

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}^*(\alpha); \widehat{\theta}^*(1-\alpha)\right], \tag{3.33}$$

that means, it is the interval between the 100αth percentage point and the $100(1-\alpha)$th percentage point of the empirical distribution of $\left\{\widehat{\theta}^{*b}\right\}_{b=1}^{B}$. Because of finite B, "simulation noise" is introduced in estimating percentile based CIs. $B = 1999$ sufficiently reduces this effect, see the introduction to this section. One takes this value instead of 2000 because then commonly used percentage points can be evaluated without interpolation (e.g., 95th percentage point $= 0.95 \cdot (1999 + 1)$th $=$ 1900th largest replication value).

3.4.4 BCa confidence interval

The bootstrap bias-corrected and accelerated (BCa) confidence interval is

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}^*(\alpha 1); \widehat{\theta}^*(\alpha 2)\right], \tag{3.34}$$

where

$$\alpha 1 = F\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z(\alpha)}{1 - \widehat{a}\left[\widehat{z}_0 + z(\alpha)\right]}\right) \tag{3.35}$$

and

$$\alpha 2 = F\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z(1-\alpha)}{1 - \widehat{a}\left[\widehat{z}_0 + z(1-\alpha)\right]}\right). \tag{3.36}$$

$F(\cdot)$ is the standard normal distribution function (Eq. 3.49). $\widehat{z}_0$, the bias correction, is computed as

$$\widehat{z}_0 = F^{-1}\left(\frac{\#\left\{\widehat{\theta}^{*b} < \widehat{\theta}\right\}}{B}\right), \tag{3.37}$$

where $\#\{\widehat{\theta}^{*b} < \widehat{\theta}\}$ means the number of replications where $\widehat{\theta}^{*b} < \widehat{\theta}$ and $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. The acceleration, $\widehat{a}$, is computed (Efron and Tibshirani 1993) as

$$\widehat{a} = \frac{\sum_{j=1}^{n}\left[\left\langle\widehat{\theta}_{(j)}\right\rangle - \widehat{\theta}_{(j)}\right]^3}{6\left\{\sum_{j=1}^{n}\left[\left\langle\widehat{\theta}_{(j)}\right\rangle - \widehat{\theta}_{(j)}\right]^2\right\}^{3/2}}, \tag{3.38}$$

where $\widehat{\theta}_{(j)}$ is the jackknife value of $\widehat{\theta}$. Consider the original sample with the jth point removed, that is, $\{t(i), x(i)\}, i = 1, \ldots, n, i \neq j$. The jackknife value is then the value of $\widehat{\theta}$ calculated using this sample of reduced size. The average, $\left\langle\widehat{\theta}_{(j)}\right\rangle$, is given by $\left[\sum_{j=1}^{n}\widehat{\theta}_{(j)}\right]\Big/ n$.

$\widehat{z}_0$ corrects for the median estimation bias; for example, if just half of the replications have $\widehat{\theta}^{*b} < \widehat{\theta}$, then $\widehat{z}_0 = 0$. The acceleration, $\widehat{a}$, takes into account scale effects, which arise when the standard error of $\widehat{\theta}$ itself depends on the true parameter value, θ.

3.5 Examples

In the first, theoretical example, we compare classical and bootstrap CIs in terms of coverage accuracy (Table 3.5). The mean of AR(1) processes with uneven spacing was estimated for two distributional shapes, normal and lognormal. The classical CI employed the effective data size for mean estimation, the bootstrap CI used the ARB algorithm and the BCa method.

The classical CI performed better for the normal than for the lognormal shape. This is because the normal assumption made at CI construction is violated in the case of the lognormal shape. With increasing data size, the lognormal approaches the normal distribution (Johnson et al. 1994: Chapter 14 therein) and the difference in performance decreases. However, this difference is still significant for $n = 1000$ in the example.

Also the bootstrap CI performed better for the normal than for the lognormal shape. This may be because persistence time estimation ($\widehat{\tau}$) and persistence time bias correction ($\widehat{\tau}'$) is less accurate for non-normally distributed data.

Table 3.5. Monte Carlo experiment, mean estimation of AR(1) noise processes with uneven spacing, normal and lognormal shape. $n_{\text{sim}} = 47{,}500$ random samples were generated from the Gaussian AR(1) process, $\{X(i)\}_{i=1}^{n}$, after Eq. (2.9) with $\tau = 1$. The samples from the lognormal AR(1) process were generated by taking $\exp[X(i)]$. The start was set to $t(1) = 1$; the time spacing, $d(i)$, was drawn from a gamma distribution (Eq. 2.48) with order parameter 16, that means, a distribution with a coefficient of variation equal to $(16)^{-1/2} = 0.25$, and subsequently scaled to $\bar{d} = 1$. Two CI types for the estimated mean were constructed, classical and bootstrap. The classical CI employed n'_μ calculated from Eq. (2.7) with $\widehat{\bar{a}}' = \exp\left(-\bar{d}/\widehat{\tau}'\right)$ plugged in for a, and the t distribution (Eq. 3.18). The bootstrap CI used the ARB (Algorithm 3.5) and the BCa method (Section 3.4.4) with $B = 1999$ and $\alpha = 0.025$.

n	$\gamma_{\bar{x}}$[a]				*Nominal*
	Distribution				
	Normal		*Lognormal*		
	CI type		*CI type*		
	Classical	*Bootstrap*	*Classical*	*Bootstrap*	
10	0.918	0.863	0.835	0.789	0.950
20	0.929	0.903	0.845	0.845	0.950
50	0.938	0.929	0.876	0.888	0.950
100	0.943	0.941	0.897	0.909	0.950
200	0.942	0.943	0.914	0.922	0.950
500	0.947	0.948	0.926	0.930	0.950
1000	0.947	0.949	0.933	0.937	0.950

[a] Standard error of $\gamma_{\bar{x}}$ is nominally 0.001.

For small sample sizes ($n \lessapprox 50$ (normal distribution) or $n \lessapprox 20$ (lognormal distribution)) the classical CI performed better than the bootstrap CI. This advantage is likely in part owing to the fact that a formula for the effective data size for mean estimation is known; it may disappear for more complex estimators, where no formula for the effective data size exists. For larger sample sizes ($n \gtrapprox 100$ (normal distribution) or $n \gtrapprox 50$ (lognormal distribution)) the bootstrap CI is as good as the classical CI (normal shape) or better (lognormal shape).

In the second, practical example, Fig. 3.4 shows the transition from a glacial (MIS 6) to the last interglacial (MIS 5) in the Vostok CO_2 record. The mean CO_2 concentration was estimated for the time intervals from 140 to 177 ka (glacial) and from 115 to 130 ka (interglacial). Student's t CIs (Section 3.4.2) were constructed using nonparametric stationary bootstrap resampling, a variant of the MBB, where the block length is

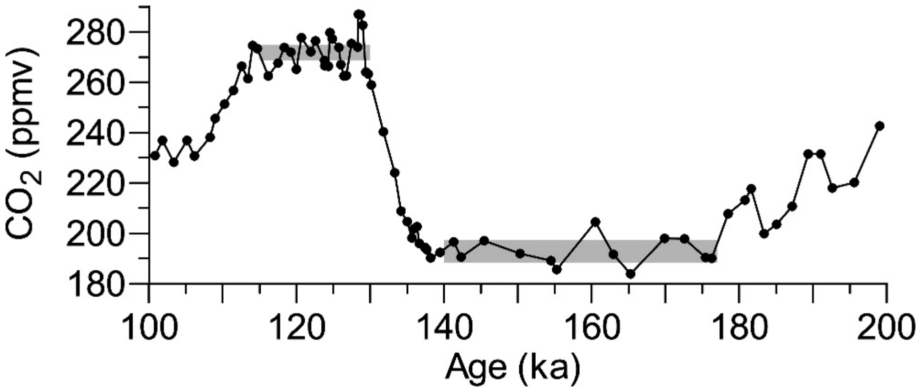


Figure 3.4. Determination of mean CO_2 levels in the Vostok record (Fig. 1.3b) during a glacial and an interglacial. The interval from 140 to 177 ka represents the glacial (MIS 6), the interval from 115 to 130 ky the interglacial (marine isotope substage 5.5). The 95% bootstrap CIs for the estimated means are shown as *shaded bars*.

not constant (Section 3.8). The number of resamples was $B = 2000$. The average block length was adjusted to $NINT\left(4 \cdot \tau / \bar{d}\right)$.

The mean glacial CO_2 level was determined as 192.8 ppmv with 95% CI [188.3 ppmv; 197.3 ppmv]; the mean interglacial CO_2 level was 271.9 ppmv with 95% CI [268.8 ppmv; 275.0 ppmv]. Because of the reduced data sizes in the intervals (glacial, $n = 13$; interglacial, $n = 24$), also the accuracies of the CIs may be reduced. The enormous glacial–interglacial amplitude in CO_2 documents the importance of this greenhouse gas for late Pleistocene climate changes, the ice age. The relation between CO_2 and temperature changes is analysed in Chapters 7 and 8.

3.6 Bootstrap hypothesis tests

By the analysis of climate time series, $\{t(i), x(i)\}_{i=1}^{n}$, we make, generally speaking, a statistical inference of properties of the climate system. One type of inference is estimation of a climate parameter, θ. In addition to a point estimate, $\widehat{\theta}$, an interval estimate, $\mathrm{CI}_{\widehat{\theta},1-2\alpha}$, helps to assess how accurate $\widehat{\theta}$ is. The bootstrap is used to construct CIs in complex situations regarding data properties shape, persistence and spacing. The second type of inference is testing a hypothesis, a statement about the climate system, using the data sample. Again, this can be a difficult task (shape, persistence, spacing), and again, the bootstrap can be a powerful tool in such a situation. Hypothesis tests are also called significance tests or statistical tests.

A hypothesis test involves the following procedure. A null hypothesis (or short: null), H_0, is formulated. H_0 is tested against an alternative hypothesis, H_1. The hypotheses H_0 and H_1 are mutually exclusive. H_0 is a simple null hypothesis if it completely specifies the data generating process. An example would be "$X(i)$ is a Gaussian white-noise process with zero mean and unit standard deviation." H_0 is a composite null hypothesis if some parameter of $X(i)$ is unspecified, for example, "Gaussian white-noise process with zero mean." Next, a test statistic, U, is calculated. Any meaningful construction lets U be a function of the data process, $U = g\left(\{T(i), X(i)\}_{i=1}^{n}\right)$. On the sample level, $u = g\left(\{t(i), x(i)\}_{i=1}^{n}\right)$. In the example H_0: "Gaussian white-noise process with $\mu = 0$" one could take $U = \bar{X} = \sum_{i=1}^{n} X(i)/n$, the sample mean. U is a random variable with a distribution function, $F_0(u)$, where the index "0" indicates that U is computed "under H_0," that is, as if H_0 were true. $F_0(u)$ is the null distribution. In the example, $F_0(u)$ would be Student's t distribution function (Section 3.9). If in the example the alternative were H_1: "$\mu > 0$," then a large, positive u value would speak against H_0 and for H_1. Using $F_0(u)$ and plugging in the data $\{t(i), x(i)\}_{i=1}^{n}$, the one-sided significance probability or one-sided P-value results as

$$\begin{aligned} P &= \operatorname{prob}\left(U \geq u \mid H_0\right) \\ &= 1 - F_0(u). \end{aligned} \tag{3.39}$$

The P-value is the probability that under H_0 a value of the test statistic greater than or equal to the observed value, u, is observed. If P is small, then H_0 is rejected and H_1 accepted, otherwise H_0 is accepted and H_1 rejected. The two-sided P-value is

$$P = \operatorname{prob}\left(|U| \geq |u| \mid H_0\right). \tag{3.40}$$

In the example, a two-sided test would be indicated for H_1: "Gaussian white-noise with $\mu \neq 0$." Besides the P-value, a second result of a statistical test is the power. In the one-sided test example:

$$\text{power} = \operatorname{prob}\left(U \geq u \mid H_1\right). \tag{3.41}$$

A type-2 error is accepting H_0, although it is a false statement and H_1 is true. The probability of a type-2 error is $\beta = 1 - \text{power}$. A type-1 error is rejecting H_0 against H_1, although H_0 is true. P, the significance probability, is therefore denoted also as type-1-error probability or false-alarm probability; u is denoted also as false-alarm level.

Although H_0 can be a composite null, it is usually more explicit than H_1. In climatological practice, the selection of H_1 should be guided by prior climatological knowledge. H_1 determines also whether a test should be designed as one- or two-sided. For example, if H_0 were "no temperature change in a climate model experiment studying the effects of doubled CO_2 concentrations, $\Delta T = 0$," then a one-sided test against H_1: "$\Delta T > 0$" would be appropriate because physics would not let one expect a temperature decrease. Because H_1 is normally rather general, it is difficult to quantify the test power. Therefore, more emphasis is put on accurate P-values. Various test statistics, $U_1, U_2, \ldots$, may be appropriate for testing H_0 against H_1. The statistic of choice has for a given data set a small type-1-error probability (small P-value) as first quality criterion. The second quality criterion is a small type-2-error probability (large power), preferably calculated for some realistic, explicit alternative. We can say that a test does not intend to prove that a hypothesis is true but rather that it does try to reject a null hypothesis. A null hypothesis becomes more "reliable" after it has been tested successfully against various realistic alternatives using various data samples, see Popper (1935). It is important that H_0 and H_1 are established independently of the data to prevent circular reasoning, see von Storch and Zwiers (1999: Section 6.4 therein). As a final general remark, it is more informative to give P-values than to report merely whether they are below certain specified significance levels, say $P < 0.1, 0.05$ or 0.01.

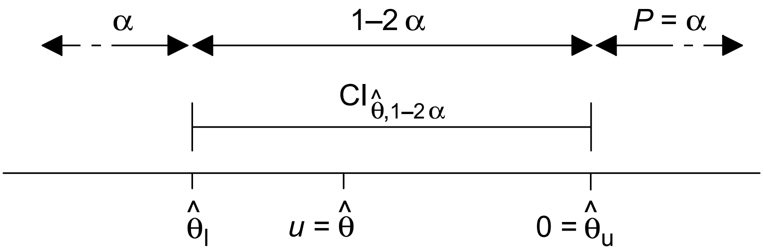


Figure 3.5. Hypothesis test and confidence interval. The parametric null hypothesis H_0: "$\theta < 0$" cannot be rejected against H_1: "$\theta \geq 0$" with a P-value of α.

When H_0 concerns a particular parameter value ($U = \theta$), a CI can be used to derive the P-value (Efron and Tibshirani 1993: Section 15.4 therein). Suppose that a test observes $u = \widehat{\theta} < 0$. Then select α such that the upper CI bound equals zero. Nominally, $\text{prob}(\theta \geq 0) = \alpha$ (Fig. 3.5). This gives a P-value of α for the test of H_0: "$\theta < 0$" against H_1: "$\theta \geq 0$." An example from a bivariate setting with data $\{x(i), y(i)\}_{i=1}^{n}$ would be the comparison of means μ_X and μ_Y. If the CI at level $1 - 2\alpha$

for the absolute value of the difference of means, $|\mu_X - \mu_Y|$, does contain zero, then H_0: "$\mu_X = \mu_Y$" cannot be rejected against H_1: "$\mu_X \neq \mu_Y$" at the level $p = 1 - 2\alpha$ in this two-sided test. A criticism to this CI method of hypothesis testing would be that the CIs are not necessarily constructed as if H_0 were true. There might be scale changes and $F_0(u)$ depend on H_0. However, the BCa CI provides a correction to this effect (Efron and Tibshirani 1993: p. 216 therein). Another option would be to construct a test statistic, U, such that $F_0(u)$ is the same for all H_0. Such a statistic is called a pivot.

Davison and Hinkley (1997: Chapter 4 therein) explain construction of hypothesis tests by approximating $F_0(u)$ with $\widehat{F}_0(u)$ obtained from bootstrap resampling or the bootstrap surrogate data approach (Section 3.3.3). The permutation test, developed in the 1930s (Edgington 1986), is the bootstrap test with the difference that no replacement is done for drawing the random samples. This book here puts more emphasis on bootstrap CIs than on bootstrap hypothesis test because CIs contain more quantitative information. We subscribe to Efron and Tibshirani's (1993: p. 218 therein) view that "hypothesis tests tend to be overused and confidence intervals underused in statistical applications."

An illustrative example is the case where θ is the anthropogenic signal proportion in the increase of the global temperature over the past 150 years. Specifically, θ can be defined as $\Delta T_{\text{with}} - \Delta T_{\text{without}}$, where ΔT_{with} is the temperature change calculated using an AOGCM and taking human activities such as fossil fuel consumption into account, and $\Delta T_{\text{without}}$ is the temperature change without the effects of human activities ("control run"). Hasselmann (1993) and Hegerl et al. (1996) developed the "fingerprint" approach to derive a powerful test statistic from the high-dimensional, gridded AOGCM output, and showed that H_0: "$\theta = 0$" can be rejected against H_1: "$\theta > 0$." One task was to quantify the natural temperature variability in the temporal and spatial domains, in order to derive the null distribution. This is difficult because the observed variability contains both natural and anthropogenic portions. It was solved using AOGCM experiments without simulated anthropogenic forcings and a surrogate data approach (Section 3.3.3), that means, several control runs with perturbed initial conditions. It is evident that an estimate, $\widehat{\theta}$, with confidence interval, $\mathrm{CI}_{\widehat{\theta},1-2\alpha}$, for the anthropogenic signal proportion would mean a step further towards quantification.

3.7 Notation

Table 3.6 summarizes the notation.

Table 3.6. Notation.

$X(T)$	Climate variable, continuous time, process level
$X_{\text{trend}}(T)$	Trend component, continuous time, process level
$X_{\text{out}}(T)$	Outlier component, continuous time, process level
$S(T)$	Variability, continuous time
$X_{\text{noise}}(T)$	Noise component, continuous time, process level
T	Continuous time
$X(i)$	Climate variable, discrete time, process level
$X_{\text{trend}}(i)$	Trend component, discrete time, process level
$X_{\text{out}}(i)$	Outlier component, discrete time, process level
$S(i)$	Variability, discrete time
$X_{\text{noise}}(i)$	Noise component, discrete time, process level
$T(i)$	Discrete time
i	Index
j	Index
$\mathcal{E}_{\mathrm{N}(\mu,\, \sigma^2)}(i)$	Gaussian noise process with mean μ and standard deviation σ, discrete time
$x(i)$	Climate variable, discrete time, sample level
$t(i)$	Discrete time, sample level
$\{t(i), x(i)\}_{i=1}^{n}$	Data or sample, discrete time series
$d(i)$	Time spacing, sample level
$\bar{d}$	Average time spacing, sample level
n	Data size
θ	(Climate) parameter
$\widehat{\theta}$	Estimator of (climate) parameter, process and sample levels, estimate
$\widehat{\theta}_1, \widehat{\theta}_2$	Other estimators
PDF	Probability density function
$f(\widehat{\theta})$	PDF of $\widehat{\theta}$
$F(\cdot)$	Probability distribution function
$F^{-1}(\cdot)$	Inverse probability distribution function
$F_{\text{emp}}(\cdot)$	Empirical distribution function
$E(\cdot)$	Expectation operator
$VAR(\cdot)$	Variance operator
$g(\cdot)$	Function
$\Gamma(\cdot)$	Gamma function
$NINT(\cdot)$	Nearest integer function
$\text{se}_{\widehat{\theta}}$	Standard error of $\widehat{\theta}$
$\text{bias}_{\widehat{\theta}}$	Bias of $\widehat{\theta}$
$\text{RMSE}_{\widehat{\theta}}$	Root mean squared error of $\widehat{\theta}$
$\text{CV}_{\widehat{\theta}}$	Coefficient of variation of $\widehat{\theta}$

Table 3.6. Notation (continued).

CI	Confidence interval
$\mathrm{CI}_{\widehat{\theta},1-2\alpha}$	Confidence interval for $\widehat{\theta}$ of level $1-2\alpha$
$\widehat{\theta}_{\mathrm{l}}$	Lower bound of CI for $\widehat{\theta}$
$\widehat{\theta}_{\mathrm{u}}$	Upper bound of CI for $\widehat{\theta}$
γ_{l}	Coverage, below lower CI bound
γ_{u}	Coverage, above upper CI bound
γ	Coverage of CI
μ	Mean
$\widehat{\mu}$	Mean estimator
$\bar{X}$	Sample mean, process level
$\bar{x}$	Sample mean, sample level
$\gamma_{\bar{x}}$	Coverage of $\mathrm{CI}_{\bar{x},1-2\alpha}$
σ	Standard deviation
$\widehat{\sigma}$	Standard deviation estimator
S_{n-1}	Sample standard deviation, process level
s_{n-1}	Sample standard deviation, sample level
$\gamma_{s_{n-1}}$	Coverage of $\mathrm{CI}_{s_{n-1},1-2\alpha}$
$z(\beta) = z_\beta$	Percentage point at β of the standard normal distribution
$t_\nu(\beta)$	Percentage point at β of the t distribution function with ν degrees of freedom
$\chi^2_\nu(\beta)$	Percentage point at β of the chi-squared distribution function with ν degrees of freedom
β	Probability
n_{sim}	Number of (Monte Carlo) simulations
c	Constant
c	Counter
C	Coverage error
$\mathcal{O}(\cdot)$	Order of
$\langle\cdot\rangle$	Average
AR(1)	Autoregressive process of order 1
AR(p)	Autoregressive process of order p
MA(q)	Moving average process of order q
ARMA(p,q)	Mixed autoregressive moving average process
n'	Effective data size
n'_μ	Effective data size for mean estimation
n'_{σ^2}	Effective data size for variance estimation
n'_ρ	Effective data size for correlation estimation

Table 3.6. Notation (continued).

a	AR(1) autocorrelation parameter (even spacing)
$\widehat{a}$	AR(1) autocorrelation parameter (even spacing) estimator
$\widehat{a}'$	AR(1) autocorrelation parameter (even spacing) estimator, bias-corrected
τ	AR(1) persistence time (uneven spacing)
$\widehat{\tau}$	AR(1) persistence time (uneven spacing) estimator
$\widehat{\tau}'$	AR(1) persistence time (uneven spacing) estimator, bias-corrected
$\bar{a}$	AR(1) equivalent autocorrelation parameter (uneven spacing)
$\widehat{\bar{a}}$	AR(1) equivalent autocorrelation parameter (uneven spacing) estimator
$\widehat{\bar{a}}'$	AR(1) equivalent autocorrelation parameter (uneven spacing) estimator, bias-corrected
$t^*, t^*(i)$	Bootstrap version of discrete time, sample level
$t^{*b}(i)$	Indexed bootstrap version of discrete time, sample level
$b = 1, \ldots, B$	Index
B	Number of bootstrap resamples
$x^*, x^*(i)$	Bootstrap version of climate variable, discrete time, sample level
$x^{*b}(i)$	Indexed bootstrap version of climate variable, discrete time, sample level
$d^*(i)$	Bootstrap version of time spacing, sample level
$\{t^*(i), x^*(i)\}_{i=1}^{n}$	Bootstrap resample
$\widehat{\theta}^*$	Bootstrap replication
$\widehat{\theta}^{*b}$	Indexed bootstrap replication
MBB	Moving block bootstrap
ARB	Autoregressive bootstrap
NBB	Non-overlapping block bootstrap
CBB	Circular block bootstrap
SB	Stationary bootstrap
MaBB	Matched-block bootstrap
TaBB	Tapered block bootstrap
l	Block length
l_{opt}	Optimal block length
l_{search}	Block length search value
$Y(i)$	Variable (l_{opt} selector after Bühlmann and Künsch (1999))
$\widehat{\text{IF}}\,(X(i))$	Estimated influence function
$\widehat{R}(h)$	Function (l_{opt} selector after Bühlmann and Künsch (1999))
$\widehat{R}(h)$	Autocovariance estimator (Chapter 2)
$\widehat{\rho}(h)$	Autocorrelation estimator (Chapter 2)
h	Lag
$b_0, b_1, b_2, b_3, b_4, \widehat{b}$	Parameters (l_{opt} selector after Bühlmann and Künsch (1999))

Table 3.6. Notation (continued).

$w_{\text{SC}}(\cdot)$	Split-cosine window
$w_{\text{TH}}(\cdot)$	Tukey–Hanning window
z	Auxiliary variable
CV_d	Coefficient of variation of the spacing
β	Percentage of drawable blocks (adaption of MBB to uneven spacing)
$\widehat{x}_{\text{trend}}(i)$	Estimated trend component, discrete time, sample level
$\widehat{x}_{\text{out}}(i)$	Estimated outlier component, discrete time, sample level
$\widehat{S}(i)$	Estimated variability, discrete time
$r(i)$	Residual of climate equation, discrete time (Eq. 1.2)
$r^*(i)$	Bootstrap version of residual of climate equation, discrete time (Eq. 1.2)
$\epsilon(i)$	White-noise residual, discrete time
$\tilde{\epsilon}(i)$	Centred white-noise residual, discrete time
$\tilde{\epsilon}^*(i)$	Bootstrap version of centred white-noise residual, discrete time
$\widehat{a}'(i)$	Abbreviation (ARB algorithm)
BCa CI	Bias-corrected and accelerated CI
ABC CI	Approximate BCa CI
$\widehat{\text{se}}_{\widehat{\theta}^*}$	Estimated bootstrap standard error
$\widehat{\theta}^*(\alpha)$	Percentage point at α of the empirical distribution of $\widehat{\theta}^*$
$\alpha 1, \alpha 2$	Other α values
$\widehat{z}_0$	Bias correction
$\widehat{a}$	Acceleration
$\#\{\}$	Number of cases
$\widehat{\theta}_{(j)}$	Jackknife value of $\widehat{\theta}$
H_0	Null hypothesis
H_1	Alternative hypothesis
U	Test statistic, process level
u	Test statistic, sample level (u is also denoted as false-alarm level)
U_1, U_2	Other test statistics, process level
$F_0(u)$	Null distribution
$\widehat{F}_0(u)$	Estimated null distribution
P	P-value, probability of a type-1 error or false-alarm probability
β	Probability of a type-2 error
M	Median
$\widehat{M}$	Sample median, process level
$\widehat{m}$	Sample median, sample level
$X'(i)$	Size-sorted $X(i)$
ϵ	Small value

Table 3.6. Notation (continued).

$\widehat{\theta}_{1}^{*b}(\lambda)$	Indexed lower bootstrap CI bound over a grid of confidence levels
λ	Variable, determines confidence level
$\widehat{p}(\lambda)$	Empirical probability (bootstrap calibration)
$y, p_0, p_1, p_2, p_3, p_4,$ q_0, q_1, q_2, q_3, q_4	Parameters ($z(\beta)$ approximation)
u, v, w	Parameters (error function approximation)
b, δ	Parameters (lognormal distribution)
p, q	Parameters (geometric distribution)
$\mathbf{Z}$	Set of whole numbers
$\mathbf{S}$	Set of numbers
AOGCM	Atmosphere–Ocean General Circulation Model
MIS	Marine isotope stage (sometimes also loosely used for marine isotope substage)
ΔT	Modelled temperature change
ΔT_{with}	Modelled temperature change, with fossil fuel consumption
$\Delta T_{\text{without}}$	Modelled temperature change, without fossil fuel consumption

3.8 Background material

We use **RMSE** instead of the mean squared error (given by $\text{RMSE}^2_{\widehat{\theta}}$). RMSE, with the same units as the data, is a handy parameter.

Standard deviation estimation for Gaussian white noise seems to have raised more interest in previous decades than today, as the discussion from 1968 in the journal The American Statistician illustrates (Cureton 1968b,a; Bolch 1968; Markowitz 1968a,b; Jarrett 1968). For example, the choice $\widehat{\sigma} = c \cdot S_{n-1}$, with c given by Eq. (3.24), yields minimal $\text{RMSE}_{\widehat{\sigma}}$ among all σ estimators for Gaussian white noise (Goodman 1953). Or, $\widehat{\sigma} = c^{-1} \cdot S_{n-1}$ yields $\text{bias}_{\widehat{\sigma}} = 0$ for Gaussian white noise, see for example Holtzman (1950). Today, it appears for practical purposes rather arbitrary whether or not to scale S_{n-1}, or whether to use $n-1$ or n. The resulting differences are likely much smaller than the effects of violations of the Gaussian assumption.

The **median** of a distribution is defined via $F(M) = 0.5$. ($F(\cdot)$ is the distribution function, see Eq. (3.49).) The sample median as estimator of M is on the process level

$$\widehat{M} = \begin{cases} X'((n+1)/2) & \text{for uneven } n, \\ 0.5 \cdot [X'(n/2) + X'(n/2+1)] & \text{for even } n, \end{cases} \tag{3.42}$$

where $X'(i)$ are the size-sorted $X(i)$. On the sample level, $\widehat{m}$ results from using $x(i)$.

A **robust estimation** procedure "performs well not only under ideal conditions, the model assumptions that have been postulated, but also under departures of from the ideal" (Bickel 1988). In the context of this book, the assumptions regard distributional shape, persistence and spacing; the performance regards an estimator and its properties such as RMSE or CI coverage accuracy. Under ideal conditions, robust estimation procedures can be less efficient (have higher $\mathrm{se}_{\widehat{\theta}}$) than non-robust procedures. For example, for Gaussianity and $n \rightarrow \infty$, $\mathrm{se}_{\widehat{m}} \rightarrow (\pi/2)^{1/2} \cdot \mathrm{se}_{\widehat{\mu}}$ (Chu 1955). Robust estimators can require sorting operations, which makes it often difficult to deduce their distribution. The term "robust" was coined by Box (1953); relevant papers on robust location estimation include Huber (1964) and Hampel (1985); for more details see Tukey (1977) or Huber (1981). Unfortunately, today's usage of "robust" in the climate research literature is rather arbitrary.

The **empirical distribution function** of a sample $\{x(i)\}_{i=1}^{n}$ is given by

$$F_{\mathrm{emp}}(x) = \frac{\text{number of values } \leq x}{n}. \tag{3.43}$$

$F_{\mathrm{emp}}(x)$ is the sample analogue of the theoretical distribution function, for example, Eq. (3.49).

Bootstrap resampling was formally introduced by Efron (1979); this article summarizes also earlier work. Singh (1981) soon recognized that the ordinary bootstrap yields inconsistent results in a setting with serial dependence. A consistent estimator, $\widehat{\theta}$, converges in probability to θ as n increases. Convergence in probability means

$$\lim_{n \rightarrow \infty} \mathrm{prob}\left(|\widehat{\theta} - \theta| > \epsilon\right) = 0 \ \forall \epsilon > 0. \tag{3.44}$$

Textbooks on bootstrap resampling include those written by Efron and Tibshirani (1993), Davison and Hinkley (1997) and Good (2005). Statistical point estimation is covered by Lehmann and Casella (1998).

The **moving block bootstrap** or MBB was introduced by Künsch (1989) and Liu and Singh (1992). The MBB resamples overlapping blocks. Carlstein (1986) had earlier suggested a method (denoted as NBB) that resamples non-overlapping blocks and does not truncate the final block. This may lead to resamples with data size less than n, that means, subsampling (see below). Hall (1985) had already considered overlapping and non-overlapping block methods in the context of spatial data. Bühlmann (1994) showed that if

1. $X(i)$ is a stationary Gaussian process with short-range dependence,
2. $\widehat{\theta}$ is a smooth function $g(\{x(i)\})$ of the data (e.g., the mean is a smooth function, but the median not) and

3. the block length, l, increases with the data size, n, within bounds, $l = \mathcal{O}\left(n^{1/2-\epsilon}\right), 0 < \epsilon < 1/2$,

then the MBB produces resamples from a process that converges to the data generating process. The MBB is then called asymptotically valid. The questions after the validity and other properties of the MBB and other bootstrap methods under relaxed assumptions (non-Gaussian processes, long-range dependence, etc.) are currently extensively studied in statistical science. For long-range dependence and the sample mean as estimator with an asymptotically Gaussian distribution, MBB can be modified to provide a valid approximation (Lahiri 1993). For long-range dependence and non-Gaussian limit distributions, MBB has to be changed to subsampling one single block (Hall et al. 1998). Block length selection is less explored for long-range dependence; intuitively, a larger length should be used than for short-range dependence. See Berkowitz and Kilian (2000), Bühlmann (2002), Politis (2003), Lahiri (2003) and references cited in these overviews.

Other **block length selectors** for the MBB and also for other non-parametric bootstrap methods have been proposed. Hall et al. (1995a) gave an iterative method based on subsamples and cross-validation. As regards the subsample size, consult Carlstein et al. (1998: p. 309 therein). Although the convergence properties in the general case are unknown, the method performed well in the Monte Carlo simulations shown. Politis and White (2004) developed a rule that selects block length as two times the smallest integer, after which the autocovariance function (Eq. 2.18) "appears negligible." However, for uneven spacing the autocovariance function is not defined and this selector not directly applicable. A related rule, based on the persistence time, τ, of the AR(1) process for uneven spacing (Section 2.1.2), would set $l = NINT\left(4 \cdot \tau / \bar{d}\right)$; Mudelsee (2003) suggested this rule for correlation estimation of bivariate, unevenly spaced time series (Chapter 7).

An **MBB adaption to uneven spacing** was analysed using a Monte Carlo experiment. The following simple rule was employed. Instead of allowing all $n - l + 1$ blocks to be drawn for insertion, only the 50% blocks closest (plus ties) in the coefficient of variation of the spacing, CV_d, were made drawable. This was applied to mean estimation of a Gaussian AR(1) process. The comparison between this MBB adaption and the ordinary MBB was made in terms of coverage accuracy and average CI length (Table 3.7). The experiment used the BCa CI and employed the block length selector after Eq. (3.28) for the MBB and its adaption. The result (Table 3.7) exhibits a reduced coverage accuracy of the MBB adaption. The following deficit outweighed the advantage of the adaption (increased similarity of CV_d between sample and resample).

Table 3.7. Monte Carlo experiment, moving block bootstrap adaption to uneven spacing. $n_{\mathrm{sim}} = 47{,}500$ random samples were generated from the Gaussian AR(1) process, $\{X(i)\}_{i=1}^{n}$, after Eq. (2.9) with $\tau = 1$. The start was set to $t(1) = 1$; the time spacing, $d(i)$, was drawn from a gamma distribution (Eq. 2.48) with order parameter 16, that means, a distribution with a coefficient of variation equal to $(16)^{-1/2} = 0.25$, and subsequently scaled to $\bar{d} = 1$. Bootstrap BCa CIs for the estimated mean were constructed with $B = 1999$ and $\alpha = 0.025$. The ordinary MBB resampling algorithm was compared with an MBB adaption to uneven spacing. The adaption made drawable only the 50% blocks closest (plus ties) in the coefficient of variation of the spacing. Both the MBB and its adaption to uneven spacing yield clearly larger coverage errors than the ARB because in that Monte Carlo experiment (Table 3.5) the prescribed AR(1) dependence matches the assumption made by the ARB (Section 3.3.2).

n	$\gamma_{\bar{x}}$ [a]		*Nominal*	⟨ *CI length* ⟩ [b]	
	Resampling method			*Resampling method*	
	MBB	*Adapted MBB*		*MBB*	*Adapted MBB*
10	0.591	0.623	0.950	0.836	0.864
20	0.799	0.788	0.950	0.915	0.890
50	0.874	0.861	0.950	0.685	0.672
100	0.901	0.888	0.950	0.510	0.505
200	0.913	0.903	0.950	0.374	0.372
500	0.929	0.920	0.950	0.244	0.244
1000	0.935	0.923	0.950	0.176	0.175

[a] Standard error of $\gamma_{\bar{x}}$ is nominally 0.001.
[b] Average value over n_{sim} simulations.

Reducing the drawable blocks to 50% reduced, in comparison with the ordinary MBB, the variation between resamples. This in turn reduced the variation between the replications (sample means of resamples). This led to narrower CIs from the adapted MBB algorithm (last two columns in Table 3.7). The CIs from the adapted MBB, finally, contained the true μ value less often than the CIs from the ordinary MBB. This means a reduced accuracy because the empirical coverages were in this case of mean estimation always less than the nominal value.

Other **nonparametric bootstrap resampling methods** than the MBB have been proposed. The circular block bootstrap (CBB) (Politis and Romano 1992a) "wraps" the data $\{x(i)\}_{i=1}^{n}$ around a circle such that $x(n)$ (Algorithm 3.1) has a successor, $x(1)$. The CBB then resamples overlapping blocks of length l from this periodic structure. That overcomes the deficit of the MBB that data near the edges, $x(1)$ or $x(n)$, have a lower probability to be resampled than data in the centre.

Also the stationary bootstrap (SB) (Politis and Romano 1994) uses the periodic structure to ensure stationarity of the resampling process. Also the SB uses overlapping blocks—however, the block length is not constant but geometrically distributed. Similar selectors as for the MBB (Section 3.3.1.1) can be used for adjusting the average block length. As regards the choice among MBB, NBB, CBB and SB, Lahiri (1999) showed that (1) overlapping blocks (MBB, CBB, SB) are better than non-overlapping blocks (NBB) in terms of RMSE of estimation of variance and related quantities like bootstrap standard error and (2) nonrandom block lengths (MBB, CBB) are, under the same criterion, at least as good as random block lengths (SB). For estimation of the distribution function and related quantities like CI points, less is known but there are indications that also here MBB and CBB perform better (Lahiri 2003: Chapter 5 therein). Some recent developments are the following. The matched-block bootstrap (MaBB) (Carlstein et al. 1998) introduces dependence between blocks to reduce bias in the bootstrap variance by imposing probability rules. One rule prefers resampling blocks such that block values at the endpoints, where the blocks are concatenated, show a higher agreement than under the MBB. The tapered block bootstrap (TaBB) (Paparoditis and Politis 2001) tapers (weights) data by means of a function before concatenating blocks. The idea is to give reduced weight to data near the block endpoints. This could make the TaBB have lower estimation bias than MBB or CBB (Paparoditis and Politis 2001). Advanced block bootstrap methods could be better than MBB for analysing equidistant climate time series, especially in the case of the MaBB, which shows good theoretical and simulation results when $X(i)$ is an AR(p) process (Carlstein et al. 1998). For uneven spacing, it could be more important to enhance MBB by matching blocks in terms of their spacing structure. This point deserves further study by means of Monte Carlo experiments. Subsampling refers to a procedure where the bootstrap resample size is less than the data size. NBB can lead to subsampling. Also the jackknife (Efron 1979), where $l = n - 1$ and one block only is resampled, is a subsampling variant. A detailed account is given by Politis et al. (1999). We finally mention the wild bootstrap, which attempts to reconstruct the distribution of a residual $r(i)$ (Eq. 3.29) by means of a two-point distribution (Wu 1986; Härdle and Marron 1991). The adaption of the wild bootstrap to nonparametric autoregression by Neumann and Kreiss (1998) has not yet been extended to uneven spacing, however.

The **autoregressive bootstrap** or ARB has been developed in the early 1980s; relevant early papers include Freedman and Peters (1984), Peters and Freedman (1984), Efron and Tibshirani (1986) and Findley (1986). Then Bose (1988) showed second-order correctness of the ARB method for estimating stationary AR(p) models—not necessarily Gaussian—and even spacing. Validity of the ARB for nonstationary AR(p) models (e.g., random walk or unit-root processes) requires subsampling, that is, drawing less than n resamples at Step 12 of the ARB (Algorithm 3.4), see Lahiri (2003: Chapter 8 therein). The ARB was extended to stationary ARMA(p, q) models with even spacing by Kreiss and Franke (1992). It seems difficult to generate theoretical knowledge about ARB performance for time series models with uneven spacing.

Other **parametric bootstrap resampling methods** than the ARB have been proposed. The sieve bootstrap (Kreiss 1992; Bühlmann 1997) assumes an AR(∞) process. Because of the high number of terms, this model is highly flexible and can approximate other persistence models than the AR(p) with $p < \infty$. Therefore the sieve bootstrap could also be called a semi-parametric bootstrap method. The deficit of this method regarding application to climate time series is that it is restricted to even spacing. The parametric bootstrap for Gaussian ARFIMA processes was shown to yield similar asymptotic coverage errors of CIs for covariance estimation as in the case of independent processes (Andrews and Lieberman 2002).

The **frequency-domain bootstrap** is explained in Chapter 5.

The **surrogate data** approach comes from dynamical systems theory in physics (Theiler et al. 1992). Contrary to the assertion in the review on surrogate time series by Schreiber and Schmitz (2000: p. 352 therein), this approach is *not* the common choice in the bootstrap literature. The same as the surrogate data approach is the so-called Monte Carlo approach (Press et al. 1992: Section 15.6 therein).

Bootstrap CIs, their construction and statistical properties are reviewed in the above mentioned textbooks and by DiCiccio and Efron (1996) and Carpenter and Bithell (2000). The challenging question "why not replace [CIs] with more informative tools?" has been raised by Hall and Martin (1996: p. 213 therein). This is based on their criticism that "the process of setting confidence intervals merely picks two points off a bootstrap histogram, ignoring much relevant information about shape and other important features." It has yet to be seen whether graphical tools such as those described by Hall and Martin (1996) will be accepted by the scientific communities. The percentile CI was proposed by Efron (1979), the BCa CI by Efron (1987). A numerical approximation to the BCa interval, called ABC interval, was introduced by DiCiccio and Efron (1992). See Section 3.9 on numerical issues concerning construction of BCa intervals. Götze and Künsch (1996) show the second-order correctness of BCa CIs for various estimators and the MBB for serially dependent processes. Hall (1988) determined theoretical coverage accuracies of various bootstrap CI types for estimators that are smooth functions of the data. Bootstrap-t CIs are formed using the standard error, $\mathrm{se}^*_{\widehat{\theta}}$, of a single bootstrap replication (Efron and Tibshirani 1993). For simple estimators like $\widehat{\mu} = \bar{X}$, plug-in estimates can be used instead of $\mathrm{se}^*_{\widehat{\theta}}$. However, for more complex estimators, no plug-in estimates are at hand. A second bootstrap loop (bootstrapping from bootstrap samples) had to be invoked, which would increase computing costs.

Bootstrap calibration can strongly increase CI coverage accuracy. Consider that a single CI point is sought, say, the lower bound, $\widehat{\theta}_{\mathrm{l}}$, for an estimate, $\widehat{\theta}$. Let the bound be calculated for each bootstrap sample, $b = 1, \ldots, B$, and over a grid of confidence levels, for example,

$$\widehat{\theta}_{\mathrm{l}}^{*b}(\lambda), \qquad \lambda = 0.01, \ldots, 0.99. \tag{3.45}$$

For each λ, compute

$$\widehat{p}(\lambda) = \frac{\#\left\{\widehat{\theta} \le \widehat{\theta}_{\mathrm{l}}^{*b}(\lambda)\right\}}{B}. \tag{3.46}$$

Finally, solve $\widehat{p}(\lambda) = \alpha$ for λ. In case a two-sided, equi-tailed CI is sought, the calibration curve $\widehat{p}(\lambda) = 1 - 2\alpha$, where

$$\widehat{p}(\lambda) = \frac{\#\left\{\widehat{\theta}_{\mathrm{l}}^{*b}(\lambda) < \widehat{\theta} < \widehat{\theta}_{\mathrm{u}}^{*b}(\lambda)\right\}}{B}, \tag{3.47}$$

is solved for λ. To calculate the CI points for a bootstrap sample requires to perform a second bootstrap–estimation loop. Analysing second-loop bootstrap methods like calibration or bootstrap-t interval construction may require enormous computing costs. Relevant papers on calibrated bootstrap CIs include Hall (1986), Loh (1987, 1991), Hall and Martin (1988), Martin (1990) and Booth and Hall (1994). Regarding the context of resampling data from serially dependent processes, Choi and Hall (2000) report that the sieve or AR(∞) bootstrap has a significantly better performance than blocking methods in CI calibration. However, the sieve bootstrap is not applicable to unevenly spaced time series. This books presents a Monte Carlo experiment on calibrated bootstrap CIs for correlation estimation (Chapter 7), with satisfying coverage performance despite the used MBB resampling.

Bootstrap hypothesis tests are detailed by Davison and Hinkley (1997: Chapter 4 therein), see also Efron and Tibshirani (1993: Chapter 15 therein) and Lehmann and Romano (2005: Chapter 15 therein).

The relation between making a test statistic pivotal and bootstrap CI calibration is described by Beran (1987, 1988). Guidelines for bootstrap hypothesis testing are provided by Hall and Wilson (1991). An extension of MBB hypothesis testing of the mean from univariate to multivariate time series has been presented by Wilks (1997). The dimensionality may be rather high, and the method may therefore be applicable to time-dependent climate fields such as gridded temperature output from a mathematical climate model. Beersma and Buishand (1999) compare variances of bivariate time series using jackknife resampling. They find significantly higher variability of future northern European precipitation amounts in the computer simulation with elevated greenhouse gas concentrations than in the simulation without (control run). Huybers and Wunsch (2005) test the hypothesis that Earth's obliquity variations influence glacial terminations during the late Pleistocene using parametric resampling of the timescale (Section 4.1.7).

Multiple hypothesis tests may be performed when analysing a hypothesis that consists of several sub-hypotheses. This situation arises in spectrum estimation (Chapter 5), where a range of frequencies is examined. The traditional method is adjusting the P-values of the individual tests to yield the desired overall P-value. A recent paper (Storey 2007: p. 347 therein) states "that one can improve the overall performance of multiple significance tests by borrowing information across all the tests when assessing the relative significance of each one, rather than calculating P-values for each test individually."

The **anthropogenic warming signal** has stimulated much work applying various types of hypothesis tests using measured and AOGCM temperature data. More details on the fingerprint approach are contained in the following papers: Hasselmann (1997), Hegerl and North (1997) and Hegerl et al. (1997). Correlation approaches to detect the anthropogenic warming signal are described by Folland et al. (1998) and Wigley et al. (2000). A recent overview is given by Barnett et al. (2005).

3.9 Technical issues

The **standard normal (Gaussian) distribution** has following PDF:

$$f(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right). \tag{3.48}$$

Figure 3.1 shows the distributional shape. The distribution function,

$$F(x) = \int_{-\infty}^{x} f(x')dx', \tag{3.49}$$

cannot be expressed in closed analytical form. We use

$$F(x) = 1 - 0.5\,\mathrm{erfcc}\left(x \Big/ \sqrt{2}\right), \tag{3.50}$$

where for $x \geq 0$ the complementary error function, erfcc, is approximated (Press et al. 1992: Section 6.2 therein) via

$$\begin{aligned}
\mathrm{erfcc}(u) \;&\approx\; v\,\exp(-w^2 - 1.26551223 + v\,(1.00002368 + v\,(0.37409196 \\
&\quad + v\,(0.09678418 + v\,(-0.18628806 + v\,(0.27886807 \\
&\quad + v\,(-1.13520398 + v\,(1.48851587 \\
&\quad + v\,(-0.82215223 + v\,0.17087277))))))))), && (3.51)\\
v \;&=\; 1/(1 + w/2), && (3.52)\\
w \;&=\; |u|. && (3.53)
\end{aligned}$$

For $x < 0$, use the symmetry, $F(-x) = 1 - F(x)$. For all x, this approximation has a relative error of less than $1.2 \cdot 10^{-7}$ (Press et al. 1992). The inverse function of $F(x)$ defines the percentage point on the x axis, $z(\beta)$, with $0 \leq \beta \leq 1$. Approximations are used for calculating $z(\beta)$; for the Monte Carlo simulation experiments in this book, the formula given by Odeh and Evans (1974) is employed:

$$z(\beta) \simeq -y - \frac{\{[(y \cdot p_4 + p_3) \cdot y + p_2] \cdot y + p_1\} \cdot y + p_0}{\{[(y \cdot q_4 + q_3) \cdot y + q_2] \cdot y + q_1\} \cdot y + q_0}, \qquad 0 < \beta < 0.5, \tag{3.54}$$

where

$$y = \left[\ln\left(\beta^{-2}\right)\right]^{1/2} \tag{3.55}$$

and

$$\begin{aligned}
p_0 &= -0.322232431088, & p_1 &= -1.0,\\
p_2 &= -0.342242088547, & p_3 &= -0.0204231210245,\\
p_4 &= -0.453642210148 \cdot 10^{-4}, & q_0 &= 0.0993484626060, \qquad (3.56)\\
q_1 &= 0.588581570495, & q_2 &= 0.531103462366,\\
q_3 &= 0.103537752850, & q_4 &= 0.38560700634 \cdot 10^{-2}.
\end{aligned}$$

If $0.5 < \beta < 1$ then $z(\beta) = -z(1 - \beta)$. This approximation produces, for example, the values $z(1 - 0.025) \approx 1.959964$ and $z(1 - 0.05) \approx 1.644854$. For $10^{-20} \leq \beta \leq 1 - 10^{-20}$, Eq. (3.54) yields an approximation that is accurate to seven decimal places (Odeh and Evans 1974). The percentage point of the standard normal distribution can be used to calculate approximate percentage points of other distributions such as Student's t and chi-squared (see following paragraphs). See the following for more details on the Gaussian distribution: Johnson et al. (1994: Chapter 13 therein) and Patel and Read (1996).

Student's t distribution with ν degrees of freedom has following PDF:

$$f(x) = \frac{\Gamma((\nu+1)/2)}{(\pi\nu)^{1/2}\,\Gamma(\nu/2)} \left(1 + x^2 / \nu\right)^{-(\nu+1)/2}, \qquad \nu = 1, 2, \ldots. \tag{3.57}$$

Approximations have to be used for calculating the percentage point, $t_\nu(\beta)$. For the Monte Carlo simulation experiments in this book, the following formula (Abramowitz and Stegun 1965: p. 949 therein) is employed:

$$\begin{aligned} t_\nu(\beta) \simeq z_\beta &+ \frac{z_\beta^3 + z_\beta}{4\nu} + \frac{5z_\beta^5 + 16z_\beta^3 + 3z_\beta}{96\nu^2} \\ &+ \frac{3z_\beta^7 + 19z_\beta^5 + 17z_\beta^3 - 15z_\beta}{384\nu^3} \\ &+ \frac{79z_\beta^9 + 776z_\beta^7 + 1482z_\beta^5 - 1920z_\beta^3 - 945z_\beta}{92{,}160\nu^4}, \end{aligned} \tag{3.58}$$

where $z_\beta = z(\beta)$ is the percentage point of the standard normal distribution. For $\nu \geq 10$ and $0.0025 \leq \beta \leq 0.9975$, this approximation has a relative accuracy of less than 0.015% (own determination using Johnson et al. (1995: Table 28.7 therein)). See Johnson et al. (1995: Chapter 28 therein) for more details on the t distribution.

The **chi-squared distribution** with ν degrees of freedom has following PDF:

$$f(x) = \exp(-x/2)x^{\nu/2-1} \Big/ \left[2^{\nu/2} \cdot \Gamma(\nu/2)\right], \qquad x \geq 0,\ \nu > 0. \tag{3.59}$$

It has mean ν and variance 2ν. Approximations are used for calculating the percentage point, $\chi^2_\nu(\beta)$. For the Monte Carlo simulation experiments in this book, the following formula (Goldstein 1973) is employed:

$$\begin{aligned} \chi^2_\nu(\beta) \simeq \nu \Bigg\{ 1 &- \frac{2}{9\nu} + \frac{4z_\beta^4 + 16z_\beta^2 - 28}{1215\nu^2} \\ &+ \frac{8z_\beta^6 + 720z_\beta^4 + 3216z_\beta^2 + 2904}{229{,}635\nu^3} \\ &+ (2/\nu)^{1/2} \left[\frac{z_\beta}{3} - \frac{z_\beta^3 - 3z_\beta}{162\nu} - \frac{3z_\beta^5 + 40z_\beta^3 + 45z_\beta}{5832\nu^2} \right. \\ &\left. + \frac{301z_\beta^7 - 1519z_\beta^5 - 32{,}769z_\beta^3 - 79{,}349z_\beta}{7{,}873{,}200\nu^3} \right] \Bigg\}^3, \end{aligned} \tag{3.60}$$

where $z_\beta = z(\beta)$ is the percentage point of the standard normal distribution. For $\nu \geq 10$ and $0.001 \leq \beta \leq 0.999$, this approximation has a relative accuracy of less than 0.05% (Zar 1978). See Johnson et al. (1994: Chapter 18 therein) for more details on the chi-squared distribution.

The **lognormal distribution** can be defined as follows. If $\ln[X(i)]$ is distributed as $\mathrm{N}(\mu, \sigma^2)$, then $X(i)$ has a lognormal distribution with parameters μ and σ (shape). It has the PDF

$$f(x) = (2\pi)^{-1/2} \cdot \sigma^{-1} \cdot x^{-1} \cdot \exp\left\{-\left[\ln(x/b)\right]^2 / \left(2\sigma^2\right)\right\}, \qquad x > 0, \tag{3.61}$$

where $b = \exp(\mu)$. The lognormal has expectation $\exp(\mu + \sigma^2/2)$ and variance $\left\{\exp(2\mu) \cdot \exp(\sigma^2) \cdot [\exp(\sigma^2) - 1]\right\}$. Other definitions with an additional shift parameter ($(X(i) - \delta)$ instead of $X(i)$) exist. See Aitchison and Brown (1957), Antle (1985), Crow and Shimizu (1988) or Johnson et al. (1994: Chapter 14 therein) for more details on the lognormal distribution.

The **geometric distribution** is a discrete distribution with

$$\mathrm{prob}\,(X = x) = p \cdot q^x, \qquad x = 0, 1, 2, \ldots, \tag{3.62}$$

where $q = 1 - p$ and $0 < p < 1$. It has expectation q/p. See Johnson et al. (1993: Chapter 5 therein) for more details on the geometric distribution.

BCa CI construction has numerical pitfalls. Regarding the bias correction, $\widehat{z}_0$, in the case of a discretely distributed, unsmooth estimator, $\widehat{\theta}$, own experiments with median estimation and $x(i) \in \mathbf{Z}$ (whole numbers) have shown that a higher CI accuracy is achieved when using instead of Eq. (3.37) the following formula:

$$\widehat{z}_0 = F^{-1}\left(\frac{\#\left\{\widehat{\theta}^{*b} < \widehat{\theta}\right\}}{B} + \frac{\#\left\{\widehat{\theta}^{*b} = \widehat{\theta}\right\}}{2B}\right). \tag{3.63}$$

Because only a finite number, B, of $\widehat{\theta}^*$ values are computed, $\widehat{\theta}^*(\alpha 1)$ and $\widehat{\theta}^*(\alpha 2)$ are calculated by interpolation. If now B is too small, the acceleration, $\widehat{a}$, too large and α too small, then $\alpha 1$ may become too small or $\alpha 2$ too large to carry out the interpolation. The choice of values for this book ($B = 2000$, $\alpha \geq 0.025$), however, prohibits this problem. See Efron and Tibshirani (1993: Section 14.7 therein) and Davison and Hinkley (1997: Section 5.3.2 therein) on the interpolation pitfall, and further Andrews and Buchinsky (2000, 2002) on the choice of B. Refer to Polansky (1999) on the finite sample bounds on coverage for percentile based CIs. As regards estimation of the acceleration, possible alternatives to Eq. (3.38) are analysed by Frangos and Schucany (1990).

The **balanced bootstrap** (Davison et al. 1986) is a bootstrap variant where over all $n \cdot B$ resampling operations, each of the values $\{x(i)\}_{i=1}^{n}$ is prescribed to be drawn equally often (B times). This can increase the accuracy of bootstrap estimates or, instead, allow to reduce B with the same accuracy as when using the "unbalanced" bootstrap with a higher number of resamples. In the case of a process without serial dependence, a simple algorithm for a balanced version of the ordinary bootstrap is as follows (Davison and Hinkley 1997: Section 9.2.1 therein). Step 1. Concatenate B copies of $\{x(i)\}_{i=1}^{n}$ into a single set $\mathbf{S}$ of size $n \cdot B$. Step 2. Permute the elements of $\mathbf{S}$ at random and call this set $\mathbf{S}^*$. Step 3. For $b = 1, \ldots, B$, take successive sets of n elements of $\mathbf{S}^*$ as balanced resamples $\{x^{*b}(i)\}_{i=1}^{n}$. In the case of serial dependence, a balanced version of the MBB would permute blocks of elements of $\mathbf{S}$. A reduced number of resamples, B, means reduced computing costs for the balanced bootstrap. How large this gain is depends on the type of estimation. The gain may not be large for quantile estimation (Davison and Hinkley 1997), which is required in BCa CI construction (Section 3.4.4).

2SAMPLES (Mudelsee and Alkio 2007) is a Fortran 90 program for performing comparisons of location measures (mean and median) and variability measures (standard deviation and MAD) between two samples. The difference measures are estimated with BCa CI. It is freely available from `http://www.mudelsee.com` (27 November 2009).

DOS and **Excel** resampling programs are freely available for download on `http://userweb.port.ac.uk/~woodm/programs.htm` (2 August 2005).

Good (2005) is a reference where routines for bootstrap resampling, BCa and bootstrap-t CI construction can be found. Also two- and multisample comparisons are included. Following languages/environments are supported: C++, EViews, Excel, GAUSS, Matlab, R, Resampling Stats, SAS, S-Plus and Stata.

A **Matlab/R** computer code for practical implementation of the block length selector of Politis and White (2004) can be downloaded from `http://econ.duke.edu/~ap172/` (29 June 2010).

Resampling Stats is a resampling software purchasable as standalone, Excel and Matlab versions from `http://www.resample.com` (2 August 2005).

Shazam is a commercial econometrics software that includes bootstrap resampling (`http://shazam.econ.ubc.ca`, 2 August 2005).

SPSS is a software package that includes bootstrap resampling and CI construction (Version 13.0: SPSS, Inc., Chicago, IL 60606, USA; IBM SPSS Statistics Version 18: `http://spss.com/software/statistics`, 5 January 2010).

Part II

Univariate Time Series

Chapter 4

Regression I

Regression is a method to estimate the trend in the climate equation (Eq. 1.1). Assume that outlier data do not exist or have already been removed by the assistance of an extreme value analysis (Chapter 6). Then the climate equation is a regression equation,

$$X(T) = X_{\text{trend}}(T) + S(T) \cdot X_{\text{noise}}(T). \tag{4.1}$$

One choice is to write $X_{\text{trend}}(T)$ as a function with parameters to be estimated. A simple example is the linear function (Section 4.1), which has two parameters, intercept and slope. A second example is the nonlinear regression model (Section 4.2). The other choice is to estimate $X_{\text{trend}}(T)$ nonparametrically, without reference to a specific model. Nonparametric regression (Section 4.3) is also called smoothing.

Trend is a property of genuine interest in climatology, it describes the mean state. This chapter deals also with quantifying $S(T)$, the variability around the trend, as second property of climate. Regression methods can be used to measure climate changes: their size and timing. For that aim, the ramp regression (Section 4.2.1) constitutes a useful parametric model of climate changes.

We compare the bootstrap with the classical approach to determine error bars and CIs for estimated regression parameters. The difficulties imposed by the data are non-Gaussian distributions, persistence and uneven spacing. We meet another difficulty, uncertain timescales. This leads to adaptions of the bootstrap (Section 4.1.7), where the resampling procedure is extended to include also the time values, $t(i)$.

The present chapter studies regression as a tool for quantifying the time-dependence of $X_{\text{trend}}(T)$, the relation between trend and time in univariate time series. A later chapter (Regression II) uses regression to

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_4,

analyse the relation in bivariate time series, between one time-dependent climate variable, $X(T)$, and another, $Y(T)$.

4.1 Linear regression

The linear regression uses a straight-line model,

$$X_{\text{trend}}(T) = \beta_0 + \beta_1 T. \tag{4.2}$$

The climate equation without outlier component is then written in discrete time as a linear regression equation,

$$X(i) = \beta_0 + \beta_1 T(i) + S(i) \cdot X_{\text{noise}}(i). \tag{4.3}$$

T is called the predictor or regressor variable, X the response variable, β_0 and β_1 the regression parameters.

4.1.1 Weighted least-squares and ordinary least-squares estimation

In a simple, theoretical setting, where the variability $S(i)$ is known and $X_{\text{noise}}(i)$ has no serial dependence, the linear regression model can be fitted to data $\{t(i), x(i)\}_{i=1}^{n}$ by minimizing the weighted sum of squares,

$$SSQW(\beta_0, \beta_1) = \sum_{i=1}^{n} \left[x(i) - \beta_0 - \beta_1 t(i)\right]^2 / S(i)^2 \,, \tag{4.4}$$

yielding the weighted least-squares (WLS) estimators

$$\widehat{\beta}_0 = \left[\sum_{i=1}^{n} x(i)/S(i)^2 - \widehat{\beta}_1 \sum_{i=1}^{n} t(i)/S(i)^2\right] \Big/ W, \tag{4.5}$$

$$\begin{aligned}\widehat{\beta}_1 = &\left\{\left[\sum_{i=1}^{n} t(i)/S(i)^2\right]\left[\sum_{i=1}^{n} x(i)/S(i)^2\right] \Big/ W - \sum_{i=1}^{n} t(i)\, x(i)/S(i)^2\right\} \\ &\times \left\{\left[\sum_{i=1}^{n} t(i)/S(i)^2\right]^2 \Big/ W - \sum_{i=1}^{n} t(i)^2/S(i)^2\right\}^{-1},\end{aligned} \tag{4.6}$$

where

$$W = \sum_{i=1}^{n} 1/S(i)^2. \tag{4.7}$$

In a practical setting, $S(i)$ is often not known and has to be replaced by $\widehat{S}(i)$. If prior knowledge indicates that $S(i)$ is constant, then one may take as estimator the square root of the residual mean square MS_E (Montgomery and Peck 1992),

$$\widehat{S}(i) = \widehat{S} = \left\{ \sum_{i=1}^{n} \left[x(i) - \widehat{\beta}_0 - \widehat{\beta}_1 t(i) \right]^2 \Big/ (n-2) \right\}^{1/2} = MS_E^{1/2}. \quad (4.8)$$

If $S(i)$ is unknown and possibly time-dependent, the following iterative estimation algorithm can be applied (Algorithm 4.1). As long as $S(i)$ is required only for weighting, this produces the correct estimators also if only the relative changes of $S(i)$, instead of the absolute values, are estimated. Analogously, if $S(i)$ is required only for weighting and known to be constant, then Eqs. (4.5) and (4.6) can be used with $S(i) = 1$, $i = 1, \ldots, n$ and $W = n$. This estimation without weighting is called ordinary least squares (OLS). For the construction of classical CIs (Section 4.1.4), however, an estimate of $S(i)$ has to be available.

Step 1	Make an initial guess, $\widehat{S}^{(0)}(i)$, of the variability.
Step 2	Estimate the regression parameters, $\widehat{\beta}_0^{(0)}$ and $\widehat{\beta}_1^{(0)}$, with the guessed variability used instead of $S(i)$ in Eqs. (4.5), (4.6) and (4.7).
Step 3	Calculate $e(i) = x(i) - \widehat{\beta}_0 - \widehat{\beta}_1 t(i), i = 1, \ldots, n$. The $e(i)$ are called the unweighted regression residuals.
Step 4	Obtain a new variability estimate, $\widehat{S}^{(1)}(i)$ from the residuals. This can be done either nonparametrically by smoothing (e.g., running standard deviation of $e(i)$) or fitting a parametric model of $S(i)$ to $e(i)$.
Step 5	Go to Step 2 with the new, improved variability estimate until regression estimates converge.

Algorithm 4.1. Linear weighted least-squares regression, unknown variability.

4.1.1.1 Example: Arctic river runoff

The climate model run with natural forcing only (Fig. 4.1a) does not exhibit a slope significantly different from zero. (See Section 4.1.4 for the determination of regression standard errors.) The run with combined anthropogenic and natural forcing (Fig. 4.1b) displays significant upwards trends in runoff. Wu et al. (2005) conjecture that there might be a change-point at around 1965, when the slope changed.

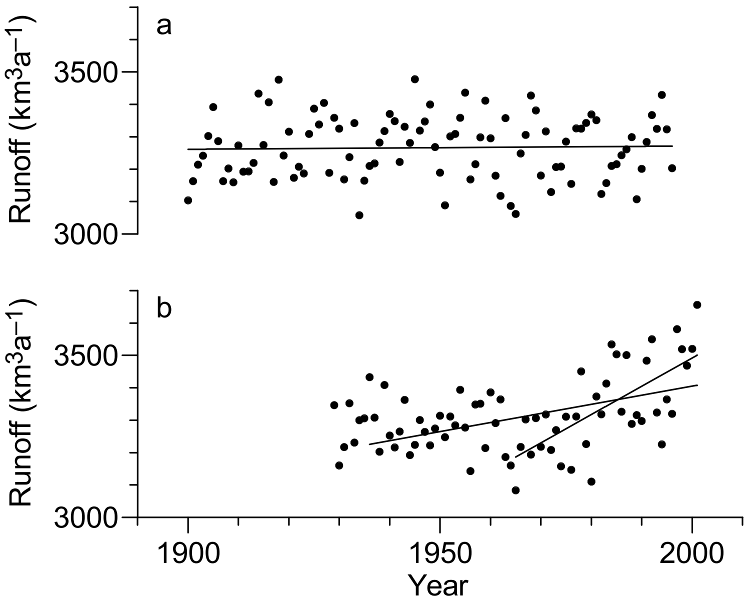


Figure 4.1. Linear regression models fitted to modelled Arctic river runoff (Fig. 1.9). **a** Natural forcing only; **b** combined anthropogenic and natural forcing. Following Wu et al. (2005), the fits (*solid lines*) were obtained by OLS regression using the data from (**a**) the whole interval 1900–1996 and (**b**) from two intervals, 1936–2001 and 1965–2001. The estimated regression parameters (Eqs. 4.5 and 4.6) and their standard errors (Eqs. 4.24 and 4.25) are as follows. **a** $\widehat{\beta}_0 = 3068 \pm 694$ km^{3}a^{-1}, $\widehat{\beta}_1 = 0.102 \pm 0.356$ km^{3}a^{-2}; **b** 1936–2001, $\widehat{\beta}_0 = -2210 \pm 1375$ km^{3}a^{-1}, $\widehat{\beta}_1 = 2.807 \pm 0.698$ km^{3}a^{-2}; **b** 1965–2001, $\widehat{\beta}_0 = -13{,}977 \pm 3226$ km^{3}a^{-1}, $\widehat{\beta}_1 = 8.734 \pm 1.627$ km^{3}a^{-2}.

4.1.2 Generalized least-squares estimation

In a practical climatological setting, $X_{\text{noise}}(i)$ often exhibits persistence. This means more structure or information content than a purely random process has. This knowledge can be used to apply the generalized least-squares (GLS) estimation, where the following sum of squares is minimized:

$$SSQG(\boldsymbol{\beta}) = (\mathbf{x} - \mathbf{T}\boldsymbol{\beta})' \, \mathbf{V}^{-1} \, (\mathbf{x} - \mathbf{T}\boldsymbol{\beta}) \, . \tag{4.9}$$

Herein,

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ (parameter vector)}, \tag{4.10}$$

$$\mathbf{x} = \begin{bmatrix} x(1) \\ \vdots \\ x(n) \end{bmatrix} \text{ (data vector)}, \tag{4.11}$$

$$\mathbf{T} = \begin{bmatrix} 1 & t(1) \\ \vdots & \vdots \\ 1 & t(n) \end{bmatrix} \text{ (time matrix)} \tag{4.12}$$

and $\mathbf{V}$ is an $n \times n$ matrix, the covariance matrix. The solution is the GLS estimator,

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{T}'\mathbf{V}^{-1}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{V}^{-1}\mathbf{x}. \tag{4.13}$$

GLS has the advantage of providing smaller standard errors of regression estimators than WLS in the presence of persistence. Analogously, in the case of time-dependent $S(i)$, the WLS estimation is preferable (Sen and Srivastava 1990) to OLS estimation. The covariance matrix has the elements

$$V(i_1, i_2) = S(i_1) \cdot S(i_2) \cdot E\left[X_{\text{noise}}(i_1) \cdot X_{\text{noise}}(i_2)\right], \tag{4.14}$$

$i_1, i_2 = 1, \ldots, n$. Climatological practice normally requires to estimate besides the variability also the persistence (Chapter 2) to obtain the $\mathbf{V}$ matrix. In the case of the AR(1) persistence model for uneven spacing (Eq. 2.9), the only unknown besides $S(i)$ required for calculating $\mathbf{V}$ is the persistence time, τ. The estimated $\mathbf{V}$ matrix has then the elements

$$\widehat{V}(i_1, i_2) = \widehat{S}(i_1) \cdot \widehat{S}(i_2) \cdot \exp\left[-|t(i_1) - t(i_2)|/\widehat{\tau}'\right], \tag{4.15}$$

$i_1, i_2 = 1, \ldots, n$, where $\widehat{\tau}'$ is the estimated, bias-corrected persistence time (Section 2.6). For even spacing, replace the exponential expression by $(\widehat{a}')^{|i_1 - i_2|}$. (In the case of persistence models more complex than AR(1), $\mathbf{V}$ is calculable and, hence, GLS applicable only for evenly spaced time series.) The autocorrelation or persistence time estimation formulas (Eqs. 2.4 and 2.11) are applied to the weighted WLS regression residuals,

$$r(i) = \left[x(i) - \widehat{\beta}_0 - \widehat{\beta}_1 t(i)\right] \Big/ \widehat{S}(i), \tag{4.16}$$

$i = 1, \ldots, n$. Detrending by a linear regression is not the same as mean subtraction, and the bias of those autocorrelation and persistence time estimators need not follow the approximations given for mean subtraction (Section 2.6), but are unknown. However, the deviations are likely negligible compared with the other uncertainties. Also in the case of unknown persistence, an iterative procedure similar to that for WLS can be applied, which is called estimated generalized least squares (EGLS) (Sen and Srivastava 1990: Section 7.3 therein). Section 4.1.4.1 gives an EGLS procedure for the case of AR(1) persistence.

4.1.3 Other estimation types

Least squares (OLS, WLS, GLS) is one type of fit criterion. Another is maximum likelihood (Section 2.6, p. 58). Further criteria result from further preferences in the regression procedure. A notable choice is robustness against the influence of outlier data, $X_{\text{out}}(i)$. This can be achieved by minimizing instead of the sum of squares (Eq. 4.4), the median of squares,

$$\widehat{m}\left\{\left[x(i)-\beta_0-\beta_1 t(i)\right]^2 / S(i)^2\right\}_{i=1}^{n} . \tag{4.17}$$

Preferably (background material) is to minimize the trimmed sum of squares,

$$SSQT(\beta_0,\beta_1)=\sum_{i=j+1}^{n-j}\left[x'(i)-\beta_0-\beta_1 t'(i)\right]/S'(i)^2 , \tag{4.18}$$

where $j = INT(\delta n)$, $INT(\cdot)$ is the integer function, $0 < \delta < 0.5$, $x'(i)$ is size-sorted $x(i)$, and $t'(i)$ and $S'(i)$ are the “slaves,” correspondingly rearranged. Trimming excludes the $2j$ most extreme terms from contributing to the estimation. Also by the minimization of the sum of absolute deviations,

$$SSQA(\beta_0,\beta_1)=\sum_{i=1}^{n}\left|x(i)-\beta_0-\beta_1 t(i)\right|/S(i) , \tag{4.19}$$

outlier values (if not already excluded by means of a prior analysis) can be given less influence on regression estimates than in least-squares minimization. Such criteria could also be preferable (in terms of, say, standard errors of estimates) to least squares when instead of $X_{\text{out}}(i)$ we considered heavy-tailed or skewed $X_{\text{noise}}(i)$ distributions.

The various criteria introduced so far and the related minimization techniques represent the computational aspect of the regression estimation problem. The second and perhaps more relevant aspect is suitability of the linear regression model. In climatology this means whether a linear increase or decrease is not too simple for describing $X_{\text{trend}}(T)$. Model suitability can be evaluated graphically via various types of plots of the regression residuals (Eq. 4.16). These realizations of the noise process should nominally not exhibit more structure than the assumed persistence model.

4.1.4 Classical confidence intervals

Assume that for a data set $\{t(i), x(i)\}_{i=1}^{n}$ the following assumptions hold:

1. $X_{\text{noise}}(i)$ is of Gaussian shape;
2. the covariance matrix $\mathbf{V}$ (Eq. 4.14), containing persistence and variability properties, is correctly estimated;
3. the linear model (Eq. 4.3) is correct.

Then CIs for the GLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of Eq. (4.13) can be constructed from their Student's t distributions (Sen and Srivastava 1990):

$$\mathrm{CI}_{\widehat{\beta}_j,1-2\alpha} = \left[\widehat{\beta}_j + t_{n-2}(\alpha) \cdot \mathrm{se}_{\widehat{\beta}_j}; \widehat{\beta}_j + t_{n-2}(1-\alpha) \cdot \mathrm{se}_{\widehat{\beta}_j}\right], \tag{4.20}$$

$j = 0$ (intercept) and 1 (slope). The standard errors of the estimators are (Sen and Srivastava 1990)

$$\mathrm{se}_{\widehat{\beta}_j} = [C(j,j)]^{1/2}, \tag{4.21}$$

$j = 0, 1$, where the matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \left(\mathbf{T}'\mathbf{V}^{-1}\mathbf{T}\right)^{-1}. \tag{4.22}$$

The GLS estimators are under the above assumptions also unbiased (Sen and Srivastava 1990):

$$\mathrm{bias}_{\widehat{\beta}_j} = 0, \tag{4.23}$$

$j = 0, 1$. The properties of the GLS regression parameter estimators hold of course also for WLS (a special case of GLS where $\mathbf{V}$ is diagonal with unequal diagonal elements) and OLS (a special case of GLS where $\mathbf{V}$ is diagonal with equal diagonal elements). In the case of OLS, the standard errors can be written in short explicit form (Montgomery and Peck 1992):

$$\mathrm{se}_{\widehat{\beta}_0} = MS_E^{1/2} \left\{ 1/n + \left(\sum_{i=1}^{n} t(i)/n\right)^2 \Big/ \left[\sum_{i=1}^{n} t(i)^2 - \left(\sum_{i=1}^{n} t(i)\right)^2 \Big/ n\right] \right\}^{1/2}, \tag{4.24}$$

$$\mathrm{se}_{\widehat{\beta}_1} = MS_E^{1/2} \left[\sum_{i=1}^{n} t(i)^2 - \left(\sum_{i=1}^{n} t(i)\right)^2 \Big/ n\right]^{-1/2}. \tag{4.25}$$

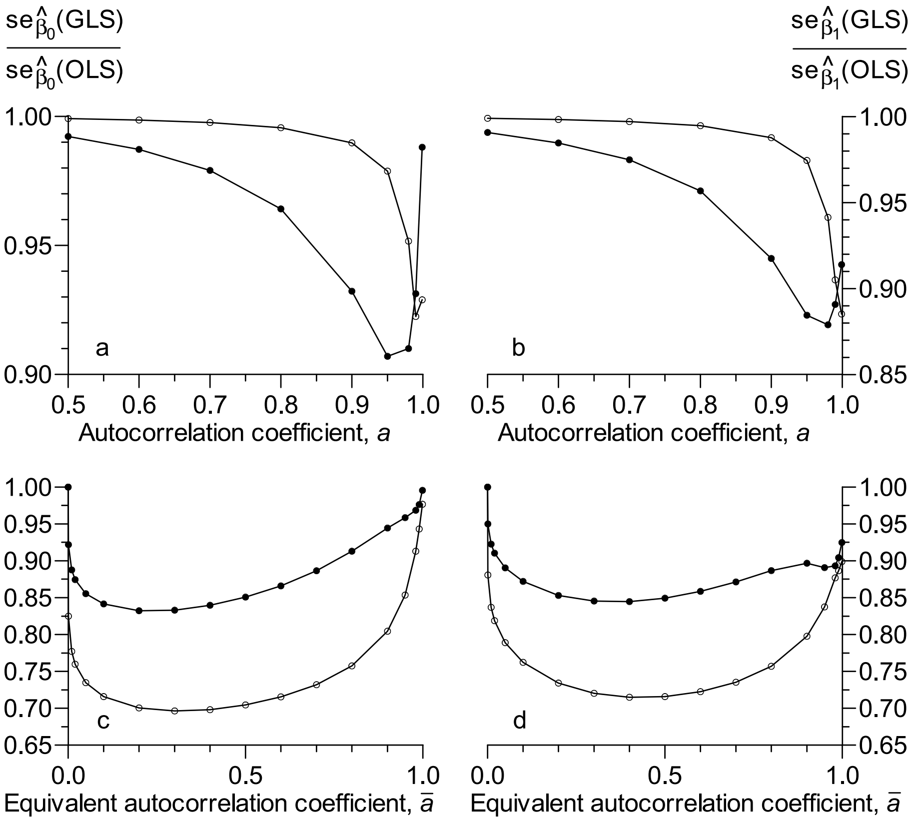


Figure 4.2. GLS versus OLS standard errors of linear regression estimators. GLS standard errors, $se_{\widehat{\beta}_j}$, $j = 0$ (intercept) and 1 (slope), are given by Eq. (4.21). OLS standard errors for the case when $\mathbf{V}$ is not diagonal with equal diagonal elements, are given (Montgomery and Peck 1992) by the square root of the diagonal elements of the matrix $(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{V}\mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}$. In each panel, the ratio of standard errors is plotted against the autocorrelation coefficient (even spacing) or the equivalent autocorrelation coefficient (uneven spacing) of the AR(1) models for $X_{\text{noise}}(i)$. **a** Even spacing, intercept; **b** even spacing, slope; **c** uneven spacing, intercept; **d** uneven spacing, slope; *filled symbols*, $n = 100$; *open symbols*, $n = 1000$. The uneven time spacing (**c, d**) was generated as $t(i) = i^2, i = 1, \ldots, n$ and scaled to $\bar{d} = 1$. Here, the coefficient of variation of the spacing, CV_d, is approximately $3^{-1/2} \approx 0.58$. The even time spacing has $d(i) = 1$.

In the case of GLS, computation of standard errors and CIs requires normally some matrix operations (Section 4.5).

One may ask what happens when the assumptions hold and OLS is used although $\mathbf{V}$ has some nonzero non-diagonal elements. The answer is that the OLS estimators are still unbiased (Montgomery and Peck 1992), but have larger standard errors than the GLS estimators. The relevance of this effect for practical applications in climatology is as follows (Fig. 4.2). The size of the error reduction by GLS depends on the autocorrelation and the spacing. For AR(1) autocorrelation, even spacing and data sizes between 100 and 1000, the reduction in standard errors (slope, intercept) is less than 13%; for uneven spacing with a typical CV_d value of somewhat above 0.5 (Fig. 1.15), the reduction is less than 31%. Interestingly, for even spacing the standard error reduction becomes sizable only for large autocorrelations (above, say, 0.9) (Fig. 4.2a, b) whereas for uneven spacing this is the case over a large autocorrelation range (Fig. 4.2c, d). Regression estimations on unevenly spaced climate time series may therefore indeed benefit from GLS no matter whether the autocorrelation is weak or strong.

4.1.4.1 Prais–Winsten procedure

The method by Prais and Winsten (1954) to apply EGLS estimation under the assumption of AR(1) persistence is of high relevance for climate time series analysis. Here, variability and size of the persistence are normally unknown and have to be estimated. In this context, the following procedure (Algorithm 4.2) can be used to obtain CIs for the regression parameters β_0 and β_1. This CI is denoted as classical because its construction assumes noise of Gaussian shape. In many practical situations, a single updating loop should already provide a satisfying regression solution: start with OLS or WLS, calculate the residuals, update $\mathbf{V}$ and perform GLS.

4.1.4.2 Cochrane–Orcutt transformation

Another idea to respect AR(1) autocorrelation of the noise process is to transform the variables. Consider

$$\beta_0^\dagger = \beta_0(1-a), \tag{4.26}$$

$$\beta_1^\dagger = \beta_1, \tag{4.27}$$

$$T^\dagger(i) = T(i) - a\,T(i-1), \qquad i = 2, \ldots, n \tag{4.28}$$

and

$$X_{\mathrm{noise}}^\dagger(i) = X_{\mathrm{noise}}(i) - a\,X_{\mathrm{noise}}(i-1), \qquad i = 2, \ldots, n. \tag{4.29}$$

Step 1	Make an initial guess of the variability, $\widehat{S}(i)$. Make an initial guess of the persistence time, $\widehat{\tau}$, for uneven spacing or the autocorrelation parameter, $\widehat{a}$, for even spacing. (For notational convenience, we omit writing a superscript.) Often the simple choices of constant variability and absent persistence are sufficient.
Step 2	Update the $\mathbf{V}$ matrix after Eq. (4.15).
Step 3	Perform GLS (Eq. 4.13) to obtain updates of regression parameter estimates. Update also $SSQG$ (Eq. 4.9).
Step 4	If regression parameters, and possibly also $SSQG$, have not changed strongly at the preceding step, take this solution and go to Step 8.
Step 5	Calculate the unweighted regression residuals $e(i)$. Update the variability estimate $\widehat{S}(i)$ by using the $e(i)$.
Step 6	Calculate the weighted regression residuals, $r(i)$, after Eq. (4.16). Update the persistence estimate $\widehat{\tau}$ or $\widehat{a}$ by using the $r(i)$. Take bias correction (Section 2.6) into account.
Step 7	Go to Step 2.
Step 8	After regression parameters, and possibly also $SSQG$, approached the solution at Step 4, calculate the residuals $r(i)$ and the standard errors (Eq. 4.21). The residuals can be used for graphical analysis (Montgomery and Peck 1992). One looks whether they fail to reveal more structure than a Gaussian AR(1) process, that means, whether the linear regression model with AR(1) noise is a suitable description of the data. The standard errors can be used for classical CI construction (Eq. 4.20).

Algorithm 4.2. Construction of classical confidence intervals, Prais–Winsten procedure.

Then

$$\begin{aligned} X^{\dagger}(i) &= X(i) - a\,X(i-1), \qquad i = 2, \ldots, n, \\ &= \beta_0^{\dagger} + \beta_1^{\dagger}\,T^{\dagger}(i) + X_{\text{noise}}^{\dagger}(i) \end{aligned} \tag{4.30}$$

yields a transformed linear regression model with independent $X_{\text{noise}}^{\dagger}(i)$ (Cochrane and Orcutt 1949), which can be solved using OLS. Because a is unknown, it has to be estimated from the residuals. This leads to a similar iterative algorithm as the Prais–Winsten procedure (Section 4.1.4.1), see also Montgomery and Peck (1992: Section 9.1.3 therein).

Step 1	Make initial guesses of the variability ($\widehat{S}(i)$) and the persistence ($\widehat{\tau}$ or $\widehat{a}$).
Step 2	Update the $\mathbf{V}$ matrix (Eq. 4.15).
Step 3	Perform GLS (Eq. 4.13) to obtain updated $\widehat{\boldsymbol{\beta}}$. Update also $SSQG$ (Eq. 4.9).
Step 4	If regression parameters, and possibly also $SSQG$, have not changed strongly at the preceding step, take this solution and go to Step 8.
Step 5	Calculate $e(i)$. Update $\widehat{S}(i)$ by using the $e(i)$.
Step 6	Calculate $r(i)$ (Eq. 4.16). Update $\widehat{\tau}$ or $\widehat{a}$ by using the $r(i)$ and bias correction.
Step 7	Go to Step 2.
Step 8	MBB: select block length (Section 3.3.1.1), guided by $\widehat{\tau}$ or $\widehat{a}$.
Step 9	Draw first resample, $\left\{t^{*b}(i), x^{*b}(i)\right\}_{i=1}^{n}$, using MBB (Algorithm 3.3) or ARB (even spacing, Algorithm 3.4; uneven spacing, Algorithm 3.5). b, counter.
Step 10	Use GLS with unchanged $\mathbf{V}$ to produce first bootstrap replication, $\widehat{\boldsymbol{\beta}}^{*b}$ with $b = 1$.
Step 11	Go to Step 9 until $b = B$ (usually $B = 2000$) replications exist.
Step 12	Calculate BCa or other CI (Section 3.4.4) from $\left\{\widehat{\boldsymbol{\beta}}^{*b}\right\}_{b=1}^{B}$.

Algorithm 4.3. Construction of bootstrap confidence intervals, Prais–Winsten procedure.

Because the transformed model has one data point less, we do not consider the Cochrane–Orcutt transformation superior to the Prais–Winsten procedure.

4.1.4.3 Approach via effective data size

A simple and fast adaption of OLS estimation to the presence of persistence can be achieved by using the effective data size for mean estimation, n'_μ. In the case of even spacing, use the general formula (Eq. 2.6) for arbitrary $X_{\text{noise}}(i)$ or Eq. (2.7) with $\widehat{a}'$ plugged in for a for AR(1) processes. In the case of uneven spacing and AR(1) dependence, use Eq. (2.7) with $\widehat{\bar{a}}' = \exp(-\bar{d}/\widehat{\tau}')$ plugged in for a.

Persistence increases the residual mean square, that is, $(n'_\mu - 2)$ has to be used instead of $(n - 2)$ in Eq. (4.8). It further increases the OLS standard errors (Eqs. 4.24 and 4.25) and widens the CI (Eq. 4.20), by a factor of $[(n-2)/(n'_\mu - 2)]^{1/2}$. Because this suggestion of a fast adaption of OLS CIs via n'_μ is ad-hoc (regression is not the same as mean estimation), its performance is tested using Monte Carlo simulations.

4.1.5 Bootstrap confidence intervals

Certain assumptions regarding the distributional shape (Gaussian) of the $X_{\text{noise}}(i)$ distribution and the persistence model (AR(1)) were made in preceding sections. This allowed to construct CIs for the regression parameter estimates using the t distribution (Eq. 4.20). In the real climate world, however, deviations from the Gaussian shape and, occasionally, the AR(1) model can be expected. In those situations, CIs may better be constructed using bootstrap resampling. We explore the MBB (Section 3.3.1) and the ARB (Section 3.3.2) resampling algorithms. We compare classical CIs with bootstrap BCa CIs (Section 3.4.4) in terms of coverage accuracy and interval length.

The application of bootstrap resampling to regression problems is straightforward. We give here (Algorithm 4.3, p. 123) the MBB and ARB algorithms for the Prais–Winsten procedure (Section 4.1.4.1). In practice, one updating loop (OLS–GLS) should already provide a satisfying regression solution.

4.1.6 Monte Carlo experiments: ordinary least-squares estimation

The performance of following CI types for OLS regression estimators was analysed by means of Monte Carlo experiments:

1. classical CI without taking persistence into account (i.e., calculating MS_E using n);
2. classical CI with persistence adjustment via n'_μ;
3. bootstrap BCa CI with ARB resampling; and
4. bootstrap BCa CI with MBB resampling.

The MBB used the block length selector after Eq. (3.28). The results of intercept estimation were similar to those of slope estimation; only the latter are therefore presented.

Ignoring persistence results in a bad coverage performance already under "ideal conditions" such as even spacing, AR(1) dependence and

Table 4.1. Monte Carlo experiment, linear OLS regression with AR(1) noise of normal shape, even spacing: CI coverage performance. $n_{\text{sim}} = 47{,}500$ random samples were generated from $X(i) = 2 + 2T(i) + X_{\text{noise}}(i)$, where $T(i) = i, i = 1, \ldots, n$ and the noise is a Gaussian AR(1) process (Eq. 2.1) with $a = 1/\mathrm{e} \approx 0.37$. Two CI types for the estimated slope were constructed, classical and bootstrap. Construction of classical CIs either ignored persistence and calculated via n (Eqs. 4.8, 4.20 and 4.25) or used n'_μ (Section 4.1.4.3). The bootstrap CIs used ARB (Algorithm 3.4) or MBB (Algorithm 3.1) resampling and the BCa method (Section 3.4.4) with $B = 1999$ and $\alpha = 0.025$.

n	$\gamma_{\widehat{\beta}_1}$ [a]				*Nominal*
	CI type				
	Classical		*Bootstrap BCa*		
	Via n	*Via n'_μ*	*ARB*	*MBB*	
10	0.851	0.900	0.809	0.718	0.950
20	0.832	0.915	0.875	0.815	0.950
50	0.819	0.932	0.915	0.867	0.950
100	0.817	0.941	0.933	0.895	0.950
200	0.819	0.945	0.941	0.913	0.950
500	0.818	0.947	0.945	0.927	0.950
1000	0.816	0.950	0.950	0.936	0.950

[a] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally 0.001.

Gaussian shape (Table 4.1). Ignoring persistence leads to underestimating the MS_E and standard errors and yields therefore too narrow CIs (Table 4.2).

Noise of AR(1) persistence and Gaussian shape is rather easy to handle for the remaining three CI types. In particular, classical CIs via n'_μ and bootstrap BCa CIs with ARB resampling performed well in terms of coverage accuracy (Table 4.1); the classical CIs did even better than the bootstrap CIs for small data sizes (less than 100).

Retaining the AR(1) persistence model but going from Gaussian to lognormal distributional shape diminished only slightly the coverage performance (Table 4.3). Interestingly, the classical CI via n'_μ performed also here excellently already for small data sizes.

Retaining the Gaussian distributional shape but adopting a persistence model more complex than AR(1) had more severe effects on coverage performance than changing the shape. In the case of AR(2) persistence (Table 4.4), the classical CI via n'_μ as well as the bootstrap BCa CI with ARB resampling failed for all data sizes tested. The reason is that the AR(1) assumption, made for n'_μ calculation as well as ARB

Table 4.2. Monte Carlo experiment, linear OLS regression with AR(1) noise of normal shape, even spacing: average CI length. Results are shown for the estimated slope. See Table 4.1 for further details.

n	$\langle$ *CI length* $\rangle$[a]			
	CI type			
	Classical		*Bootstrap BCa*	
	Via n	*Via* n'_μ	*ARB*	*MBB*
10	0.43790	0.69145	0.44414	
20	0.15132	0.24262	0.19118	
50	0.03822	0.05706	0.05316	0.04431
100	0.01354	0.02006	0.01941	0.01693
200	0.00480	0.00708	0.00697	0.00628
500	0.00121	0.00179	0.00178	0.00166
1000	0.00043	0.00063	0.00063	0.00060

[a] Average value over n_{sim} simulations.

Table 4.3. Monte Carlo experiment, linear OLS regression with AR(1) noise of lognormal shape, even spacing. $n_{\text{sim}} = 47{,}500$ random samples were generated from $X(i) = 2 + 2T(i) + X_{\text{noise}}(i)$, where $T(i) = i, i = 1, \ldots, n$. The lognormal noise was generated from a Gaussian AR(1) process (Eq. 2.1) with $a = 1/\mathrm{e} \approx 0.37$ by exponentiation and subsequent scaling (Section 3.9) to mean zero and variance unity. See Table 4.1 for further details.

n	$\gamma_{\widehat{\beta}_1}$[a]				*Nominal*
	CI type				
	Classical		*Bootstrap BCa*		
	Via n	*Via* n'_μ	*ARB*	*MBB*	
10	0.888	0.933	0.799	0.739	0.950
20	0.864	0.926	0.846	0.795	0.950
50	0.861	0.937	0.889	0.845	0.950
100	0.856	0.943	0.908	0.872	0.950
200	0.855	0.943	0.918	0.893	0.950
500	0.854	0.944	0.928	0.914	0.950
1000	0.856	0.944	0.934	0.924	0.950

[a] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally 0.001.

construction, is violated by the AR(2) model. However, both methods could in principle be adapted, the classical CI by calculating n'_μ from

Table 4.4. Monte Carlo experiment, linear OLS regression with AR(2) noise of normal shape, even spacing. $n_{\text{sim}} = 47{,}500$ random samples were generated from $X(i) = 2 + 2T(i) + X_{\text{noise}}(i)$, where $T(i) = i, i = 1, \ldots, n$. The AR(2) noise was produced after Eq. (2.14) with $a_1 = 0.5$ and $a_2 = -0.5$, and discarding the first 5000 realizations to approach asymptotic stationarity. (The a_1–a_2 setting corresponds to a quasi-cyclical noise behaviour with a period of ~ 5.2.) The noise was scaled to variance unity[a]. See Table 4.1 for further details.

n	$\gamma_{\widehat{\beta}_1}$[b]				*Nominal*
	CI type				
	Classical		*Bootstrap BCa*		
	Via n	*Via* n'_μ	*ARB*	*MBB*	
10	0.939	0.969	0.923	0.805	0.950
20	0.961	0.990	0.978	0.944	0.950
50	0.976	0.997	0.995	0.972	0.950
100	0.981	0.998	0.998	0.968	0.950
200	0.981	0.999	0.999	0.963	0.950
500	0.983	0.999	0.999	0.968	0.950
1000	0.983	0.999	0.999	0.962	0.950

[a] The asymptotically stationary AR(2) process has a mean equal to zero and a variance equal to $\sigma_\epsilon^2(1-a_2)/[(1+a_2)\,(1+a_1-a_2)\,(1-a_1-a_2)]$, where σ_ϵ^2 is the variance of the innovation term (Priestley 1981: p. 128 therein). Setting $\sigma_\epsilon^2 = 2/3$ thus yields $VAR[X_{\text{noise}}(i)] = 1$.
[b] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally 0.001.

Eq. (2.6) and the bootstrap CI by calculating the white-noise residuals (Algorithm 3.4) from an AR(2) fit. On the other hand, the bootstrap BCa CI with MBB resampling had a good coverage performance owing to some robustness against violations of the AR(1) assumption.

However, retaining the Gaussian distributional shape but adopting an ARFIMA$(0, \delta, 0)$ persistence model led to rather bad coverage performance of all four types of CIs (Table 4.5). The long-range autocorrelation of the ARFIMA model evidently cannot be captured, neither by the classical CI via n'_μ, nor by the bootstrap BCa CI with ARB resampling.

Also the bootstrap BCa CI with ordinary MBB resampling failed in the case of ARFIMA$(0, \delta, 0)$ noise (Table 4.5). This is because concatenating the independent blocks (Algorithm 3.1) introduces too strong independence and prohibits mimicking the long-term ARFIMA persistence. However, subsampling one single block (size l) and calculating the bootstrap replications by OLS regression on the simulated series of reduced size—yielded excellent coverages (Table 4.5). That means,

Table 4.5. Monte Carlo experiment, linear OLS regression with ARFIMA$(0, \delta, 0)$ noise of normal shape, even spacing. $n_{\mathrm{sim}} = 47{,}500$ random samples were generated from $X(i) = 2 + 2T(i) + X_{\mathrm{noise}}(i)$, where $T(i) = i, i = 1, \ldots, n$. The ARFIMA$(0, \delta, 0)$ noise with $\delta = 0.25$, zero mean and unity variance was generated using the algorithm of Hosking (1984). Two CI types for the estimated slope were constructed, classical and bootstrap. Construction of classical CIs either ignored persistence and calculated via n (Eqs. 4.8, 4.20 and 4.25) or used n'_{μ} (Section 4.1.4.3). The bootstrap CIs used ARB (Algorithm 3.4) or MBB (Algorithm 3.1) resampling and the BCa method (Section 3.4.4) with $B = 1999$ and $\alpha = 0.025$. Two implementations of the MBB algorithm were analysed. The ordinary MBB (Algorithm 3.3) resampled n data points by drawing random blocks of length l selected after Eq. (3.28). The subsampling MBB resampled one single block of length $l = n/2$.

n	$\gamma_{\widehat{\beta}_1}$ [a]					Nominal
	CI type					
	Classical		*Bootstrap BCa*			
	Via n	*Via n'_{μ}*	*ARB*	*MBB*		
				Ordinary	*Subsampling*	
10	0.876	0.905	0.810	0.759	0.925	0.950
20	0.816	0.876	0.825	0.786	0.942	0.950
50	0.727	0.822	0.799	0.757	0.955	0.950
100	0.650	0.770	0.757	0.726	0.955	0.950
200	0.577	0.710	0.704	0.694	0.956	0.950
500	0.481	0.618	0.615	0.644	0.957	0.950
1000	0.415	0.546	0.545	0.604	0.954	0.950

[a] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally 0.001.

reducing the data size made the MBB CIs wide enough. The critical point was to correctly guess the subsampling block length, l. The guidance offered by Lahiri (2003: Section 10.5 therein) for subsampling in the case of mean estimation, namely $l = c\,n^{1/2}$ with various c values, gave inacceptable results for slope estimation in our case of OLS regression. A trial-and-error search found the rule $l = n/2$, which worked excellently—for slope estimation and ARFIMA$(0, \delta, 0)$ noise with $\delta = 0.25$ only. This rule yielded too low coverages for intercept estimation, and it yielded too high coverages for slope estimation and ARFIMA$(0, \delta, 0)$ noise with $\delta = 0.10$.

The conclusions regarding the practice of climate time series analysis in the presence of long memory are as follows. Long-memory noise makes CIs of estimated regression parameters considerably wider than AR(1) noise. If ignored, long memory leads therefore to overstated accuracies.

A good measure against this is to use bootstrap MBB subsampling (Section 3.8). In this regard, the block length selection has a decisive impact on the coverage accuracy. A trial-and-error method, or a brute-force search as described in Section 3.3.1.1, is therefore advisable as long as no theoretical l selection rules are available.

A critical case is uneven spacing and long-memory noise. The accuracy of MBB CIs could be reduced compared to even spacing, and parametric implementations of the long-memory model, as an analogue of ARB resampling, seem not to exist.

4.1.7 Timescale errors

Up to now we have assumed that in a stochastic time series process $\{T(i), X(i)\}_{i=1}^{n}$ the times $T(i)$ were exactly known, whereas we have conceded the climate variable $X(i)$ some noise from measurement, proxy and climate uncertainties, described by the time-dependent random variable $X_{\text{noise}}(i)$. In the linear regression problem, this assumption leads to $X(i) = \beta_0 + \beta_1 T(i) + S(i) \cdot X_{\text{noise}}(i)$, with the regressor $T(i)$ fixed and known. This model is called fixed-regressor model.

For several types of climate time series, the fixed-regressor model is adequate. For example, climate model output (Fig. 1.9) or instrumental observations (Fig. 1.10) are records with exactly known $T(i)$. Also documentary data (Fig. 1.1) share this feature, potentially, as far as inhomogeneities from document loss or reporting errors can be excluded. However, for several other climate archives, namely those recording the climate via proxy variables (Table 1.2), the assumption $T(i) = T_{\text{true}}(i)$ (true time value) cannot be maintained. For example, archives such as sediment cores, speleothems or ice cores require age determinations (Section 1.6). Here we have to write the measured times as

$$T(i) = T_{\text{true}}(i) + T_{\text{noise}}(i), \tag{4.31}$$

$i = 1, \ldots, n$, where $T_{\text{noise}}(i)$ is the error owing to age uncertainties. This means further that for proxy time series we have to write the linear regression as

$$X(i) = \beta_0 + \beta_1 \left[T(i) - T_{\text{noise}}(i) \right] + S(i) \cdot X_{\text{noise}}(i), \tag{4.32}$$

$i = 1, \ldots, n$. This model is called errors-in-variables model.

What happens when we apply OLS estimation to errors-in-variables models (Eq. 4.32)? Consider first the following, simple form of the timescale error:

$$T_{\text{noise}}(i) = \mathcal{E}_{\mathrm{N}(0,\, S_T^2)}(i), \tag{4.33}$$

that means, a Gaussian random process with zero mean and variance S_T^2. Let $T_{\text{noise}}(i)$ be independent of $T_{\text{true}}(i)$ and $X_{\text{noise}}(i)$. Let $X_{\text{noise}}(i)$ be

Table 4.6. Errors and spread of time values for dated proxy time series. The time series types are listed in Table 1.2, and the records are shown in Figs. 1.2, 1.3b, 1.5a, 1.6 and 1.7.

Climate archive	*Proxy variable*	S_T	$VAR\,[T(i)]^{1/2}$	*Bias factor*[a]
Marine sediment core	$\delta^{18}O$	25 ka[b]	565 ka	0.998
Ice core	CO_2	15 ka[c]	103 ka	0.979
	Ca	13 a[d]	226 a	0.997
Tree-rings	$\Delta^{14}C$	15 a[e]	3585 a	0.99998
Speleothem	$\delta^{18}O$	82 a[f]	2367 a	0.999

[a] Also called attenuation factor; approximated by $\left(1 + S_T^2 \,/VAR\,[T(i)]\,\right)^{-1}$.
[b] Mudelsee (2005).
[c] Upper bound (Petit et al. 1999).
[d] Approximate "internal" uncertainty, i.e., within the time interval, no absolute value (Section 4.2.1.4).
[e] Average over the individual age uncertainties $S_T(j)$ (Reimer et al. 2004).
[f] Average over the individual age uncertainties $S_T(j)$ (Fleitmann et al. 2003: Table S1 therein).

given by $\mathcal{E}_{\mathrm{N}(0,\,1)}(i)$. Let further $S(i)$ be constant. It can then be shown (Draper and Smith 1981: Section 2.14 therein) that the expectation of the OLS estimator of the slope (Eq. 4.6) is

$$E\left(\widehat{\beta}_1\right) = \beta_1 \Big/ \left(1 + S_T^2 \Big/ VAR\,[T_{\mathrm{true}}(i)]\right). \tag{4.34}$$

Herein, $VAR\,[T_{\mathrm{true}}(i)]$ is the variance of the true time points, which may for practical purposes be approximated by $VAR\,[T(i)]$.

As Eq. (4.34) shows, the OLS slope estimator has a negative bias, it underestimates the true slope. On the one hand, for many practical analyses of univariate proxy climate time series, the ratio $S_T^2/VAR[T_{\mathrm{true}}(i)]$ should be negligible and the resulting bias also. This is demonstrated for some of the proxy time series analysed in this book (Table 4.6). By selecting the length of a sampled archive and the depth positions where samples are to be taken (Fig. 1.13), the experimenter has control of the variance of the depth points and, hence, some control of the variance of the time points $\{T(i)\}_{i=1}^{n}$. It is therefore sufficient to consider in this section only OLS estimation of the errors-in-variables model. The bias of $\widehat{\beta}_1$ can be taken into account using BCa CIs, which include a bias correction (Section 3.4.4).

On the other hand, bivariate proxy climate records can exhibit perhaps stronger error phenomena than univariate records. In the errors-in-variables model

$$Y(i) = \beta_0 + \beta_1\,[X(i) - S_X(i) \cdot X_{\mathrm{noise}}(i)] + S_Y(i) \cdot Y_{\mathrm{noise}}(i), \tag{4.35}$$

$i = 1, \ldots, n$, where for notational clarity $S_X(i)$ is written for $S(i)$ and $S_Y(i)$ for the variability of the $Y(i)$, and $Y_{\text{noise}}(i)$ is Y noise, the variance of the regressor (X) now cannot be controlled by designing the sampling of the archive as it is the case for univariate time series (regressor T). The OLS estimation bias may be serious to a degree that requires to analyse also other estimators for the errors-in-variables model. This is done in Chapter 8.

For climate time series from dated archives, in addition to the usual difficulties imposed by

- non-Gaussian distributional shape,

- persistence

and, to a lesser degree,

- uneven spacing (because it restricts the persistence models to types not more complex than AR(1)),

the difficulty from an

- uncertain timescale

appears. Although the resulting bias of OLS regression estimators should in most cases be negligible (Table 4.6), CI construction should take into account timescale errors to achieve better coverage accuracies. Therefore we analyse adaptions of bootstrap resampling methods: a nonparametric (via MBB), a parametric (via ARB) and also a hybrid.

4.1.7.1 Nonparametric: pairwise-moving block bootstrap

The pairwise-moving block bootstrap or pairwise-MBB algorithm (Algorithm 4.4) resamples pairs, for example, $(t^*(i), x^*(i)) = (t(j), x(j))$. This deviates from the MBB (Algorithm 3.3), where $\{t^*(i)\}_{i=1}^{n}$ is set equal to $\{t(i)\}_{i=1}^{n}$ and resampling is applied to the residuals. The idea of the pairwise-MBB is to capture the $T(i)$ uncertainties without parametrically modelling the timescale, namely by including the times into the resampling procedure.

4.1.7.2 Parametric: timescale-autoregressive bootstrap

The ARB (Algorithms 3.4 and 3.5) employed a parametric AR(1) model for resampling the noise $\{X_{\text{noise}}(i)\}_{i=1}^{n}$ and left the times $\{T(i)\}_{i=1}^{n}$ unchanged. In the presence of timescale uncertainties, the ARB can be adapted by parametric modelling; we denote this algorithm as timescale-autoregressive bootstrap or timescale-ARB (Algorithm 4.5).

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Regression residuals (Eq. 4.16)	$r(i)$
Step 3	Select block length	l
	(Section 3.3.1.1)	
	using $r(i)$	
Step 4	Apply MBB with l (Algorithm 3.1)	
	to x values	$\{x^*(i)\}_{i=1}^{n} = \{x(f(i))\}_{i=1}^{n}$
Step 5	Overtake bootstrap index	$f(i)$
	for resampled times	$\{t^*(i)\}_{i=1}^{n} = \{t(f(i))\}_{i=1}^{n}$

Algorithm 4.4. Pairwise-MBB algorithm, regression estimation. By overtaking the random bootstrap index $f(i) \in \{1, \ldots, n\}$ from x-resampling for t-resampling, $(t(j), x(j))$ pairs are resampled.

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Regression residuals (Eq. 4.16)	$r(i)$
Step 3	Resample	$\{x^*(i)\}_{i=1}^{n}$
	by applying ARB to $r(i)$ under	$\{t(i)\}_{i=1}^{n}$
Step 4	Model parametrically	$\{t^*(i)\}_{i=1}^{n}$

Algorithm 4.5. Timescale-ARB algorithm, regression estimation. The ARB (even spacing, Algorithm 3.4; uneven spacing, Algorithm 3.5) is first applied to the regression residuals using the time values $t(i)$ to produce the $x^*(i)$ resamples. Then the $t^*(i)$ are resampled from a parametric model of the accumulation process of the climate archive.

A parametric model of $\{T^*(i)\}_{i=1}^{n}$ comes from a physical description of the accumulation process of the climate archive. Consider as a simple example a linear accumulation and a number n_{date} of dated points

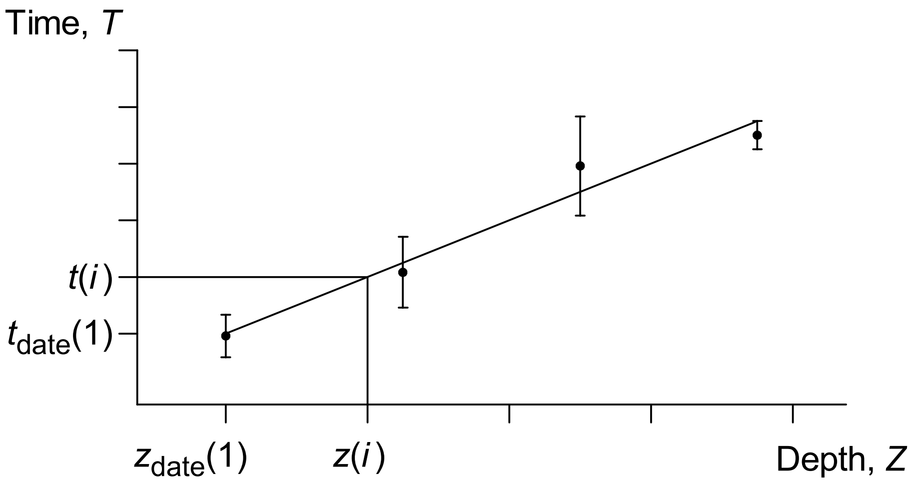


Figure 4.3. Linear timescale model. The dating points $\{z_{\text{date}}(j), t_{\text{date}}(j)\}_{j=1}^{n_{\text{date}}}$ are shown as *filled symbols*, their dating errors $S_{\text{date}}(j)$ as *vertical bars* (±). (Strictly speaking, the dating error is an unknown random variable with standard deviation $S_{\text{date}}(j)$.) The WLS regression fitted to the dating points (*solid straight line*) is used to convert the depth value, $z(i)$, of a measurement, $x(i)$, into time, $t(i)$.

(Fig. 4.3), which is a good approximation for many sedimentary and speleothem time series.

Generating the simulated time points $\{t^*(i)\}_{i=1}^{n}$ is straightforward (Algorithm 4.6). The assumptions made are:

1. linear accumulation process with

2. positive slope and

3. independent, Gaussian distributed dating errors.

The linearity assumption can be tested by analysing the residuals from the regression of the dated depth points $\{Z_{\text{date}}(j)\}_{j=1}^{n_{\text{date}}}$ as regressor on the dates $\{T_{\text{date}}(j)\}_{j=1}^{n_{\text{date}}}$. The constraint "positive slope" refers to the assumed monotonic growth of an archive; it is taken into account by retaining only those model simulations with a positive slope (Algorithm 4.6). The Gaussian assumption regarding the dating errors should be well fulfilled in most applications; if it appeared to be violated, the algorithm could be easily adapted (Algorithm 4.6, Step 4). The assumption of independent dating errors should be well fulfilled. For example, in U/Th dating the effects of dependence between dating errors owing to imper-

Step 1	Dating points,	$\{z_{\text{date}}(j), t_{\text{date}}(j)\}_{j=1}^{n_{\text{date}}}$,
	dating errors	$\{S_{\text{date}}(j)\}_{j=1}^{n_{\text{date}}}$
Step 2	Parameter estimates	$\widehat{\beta}_0, \widehat{\beta}_1$
	of WLS regression to	$\{z_{\text{date}}(j), t_{\text{date}}(j)\}_{j=1}^{n_{\text{date}}}$
	(linear model)	
Step 3	Timescale	$t(i) = \widehat{\beta}_0 + \widehat{\beta}_1 z(i),\ i = 1, \ldots, n$
Step 4	Simulated dating	
	points	$T^*_{\text{date}}(j) = T_{\text{date}}(j) + S_{\text{date}}(j) \cdot \mathcal{E}_{\text{N}(0,\,1)}(j)$,
		$j = 1, \ldots, n_{\text{date}}$
Step 5	WLS regression to	$\{z_{\text{date}}(j), t^*_{\text{date}}(j)\}_{j=1}^{n_{\text{date}}}$
Step 6	Parameter estimates,	
	simulation	$\widehat{\beta}^*_0, \widehat{\beta}^*_1$
Step 7	If $\widehat{\beta}^*_1 > 0$, then	
	calculate simulated	
	timescale	$t^*(i) = \widehat{\beta}^*_0 + \widehat{\beta}^*_1 z(i),\ i = 1, \ldots, n$

Algorithm 4.6. Timescale resampling, linear accumulation model. Note: $t^*_{\text{date}}(j)$ is the realization of $T^*_{\text{date}}(j)$.

fectly known decay constants are likely negligible compared with those from independent counting errors in the mass spectrometer.

It is also possible that the dating points reveal a break in the accumulation process, known as hiatus (Fig. 4.4). Such gaps are frequently found in sedimentary archives and speleothems. Also here, generating the points $\{t^*(i)\}_{i=1}^{n}$ is straightforward via modelling the accumulation as a two-phase regression model. The assumptions are:

1. linear accumulation in phase I and phase II,
2. positive slopes,
3. monotonic growth and
4. a correctly determined hiatus depth, $z_{\text{I/II}}$.

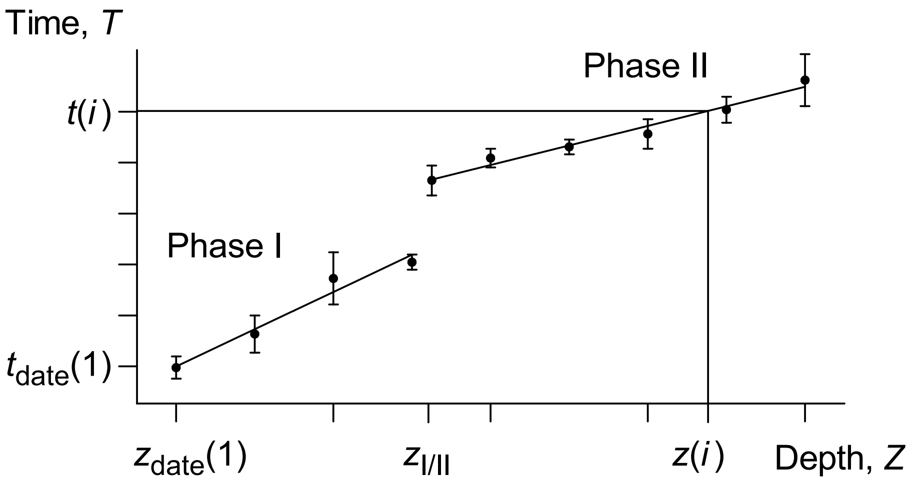


Figure 4.4. Two-phase linear timescale model. The break in the accumulation occurs at depth $z_{\mathrm{I/II}}$. See also Fig. 4.3.

The algorithm works analogously to the linear case (Algorithm 4.6). "Monotonic growth" refers to the condition that the simulated times $t^*(i)$ increase also across $z_{\mathrm{I/II}}$. If $z_{\mathrm{I/II}}$ is not clearly identifiable, it may be included as an additional parameter; the model is then a change-point regression model (Section 4.2). Evidently, multi-phase linear or smooth nonlinear accumulation models can be constructed and used for timescale resampling. The critical point is that the mathematical model describes the physical accumulation process adequately.

Reliable knowledge about dating errors $S_{\mathrm{date}}(j)$ need not always exist. This may be so for an absolutely dated archive, where, however, the conditions for a reliable age determination were not fulfilled, such as the absence of contamination with old, Δ^{14}C-poor material in radiocarbon dating. Instead of taking too small, unreliable "machine error bars" one can then estimate an $S_{\mathrm{date}}(j)$ average via the residual mean square (Eq. (4.8) for a linear model). Another example is the modelled timescale of an ice core, where the time–depth relationship is in general nonlinear and variable, and where the accuracy depends on the validity of modelling assumptions. For the ice core timescale, a simple description may be obtained via the first derivative or sedimentation rate,

$$\dot{Z}_{\mathrm{date}}(j) = \left[Z_{\mathrm{date}}(j+1) - Z_{\mathrm{date}}(j)\right] / \left[T_{\mathrm{date}}(j+1) - T_{\mathrm{date}}(j)\right]. \quad (4.36)$$

This approach is further pursued in Chapter 8.

4.1.7.3 Hybrid: timescale-moving block bootstrap

A hybrid between nonparametric resampling (the $x^*(i)$ via MBB) and parametric resampling (the $t^*(i)$ via the procedure of Section 4.1.7.2) can be easily created (Algorithm 4.7). The intention of this timescale-moving block bootstrap or timescale-MBB algorithm is to combine the advantages of the MBB (fewer parametric restrictions than ARB) with the situation in practice, where for many dated archives a parametric timescale error model can be constructed on the basis of known accumulation processes and sizes of dating errors.

Step 1	Data	$\{t(i), x(i)\}_{i=1}^n$
Step 2	Regression residuals (Eq. 4.16)	$r(i)$
Step 3	Resample	$\{x^*(i)\}_{i=1}^n$
	by applying MBB to $r(i)$ under	$\{t(i)\}_{i=1}^n$
Step 4	Model parametrically	$\{t^*(i)\}_{i=1}^n$

Algorithm 4.7. Timescale-MBB algorithm, regression estimation. The MBB (Algorithm 3.3) is first applied to the regression residuals. Then the $t^*(i)$ are resampled from a parametric model of the accumulation process of the climate archive.

4.1.7.4 Monte Carlo experiments

The Monte Carlo experiments demonstrate that even for simple models of $X_{\text{noise}}(i)$ such as Gaussian AR(1), relatively small additions of timescale noise can invalidate the coverage performance of CIs that ignore it. This is the case for the estimated slope (Table 4.7) and also the estimated intercept (Table 4.9). If we translate the numerical entry for $n = 100$ from Table 4.7 into an example of a Holocene stalagmite: $T_{\text{true}}(1) \approx 0$, $T_{\text{true}}(n) \approx 10{,}000\,\text{a}$, spacing $\bar{d} \approx 100\,\text{a}$ and $n_{\text{date}} = 2$ dating points, then Gaussian errors of $S_{\text{date}}(1) = S_{\text{date}}(2) = 50$ years, which are indeed not large values, produce gross mis-coverages of classical (via n'_μ) and bootstrap (ARB, MBB) CIs.

On the other hand, taking timescale errors into account at CI construction by parametrically modelling them (timescale-ARB, timescale-MBB) leads to excellent coverage performance already for data sizes as small as 50. Those CIs take successfully into account the increased RMSE owing to timescale errors (Table 4.8): they become wider.

Table 4.7. Monte Carlo experiment, linear OLS regression with timescale errors and AR(1) noise of normal shape: CI coverage performance, slope. $n_{\mathrm{sim}} = 475$ (in the case of timescale-ARB and timescale-MBB, $n_{\mathrm{sim}} = 47{,}500$) random samples were generated from $X(i) = 2 + 2T_{\mathrm{true}}(i) + X_{\mathrm{noise}}(i)$, where $T_{\mathrm{true}}(i) = i, i = 1, \ldots, n$ and the noise is a Gaussian AR(1) process (Eq. 2.1) with $a = 1/\mathrm{e} \approx 0.37$. Timescale errors were subsequently introduced as follows. A linear timescale model (Fig. 4.3) with $n_{\mathrm{date}} = 2$ dating points and independent, Gaussian distributed timescale errors was used to generate the $T(i)$ as $T(i) = T_{\mathrm{true}}(i) + \mathcal{E}_{\mathrm{N}(0,\,0.25)}(i)$ for $i = 1, n$, and then by linear interpolation for $i = 2, \ldots, n-1$. Two CI types for the estimated slope were constructed, classical via n'_μ and bootstrap BCa with $B = 1999$; $\alpha = 0.025$. The bootstrap CIs used following resampling methods: ARB (Algorithm 3.5), MBB (Algorithm 3.1), pairwise-MBB (Algorithm 4.4), timescale-ARB (Algorithm 4.5) and timescale-MBB (Algorithm 4.7).

n	$\gamma_{\widehat{\beta}_1}$						*Nominal*
	CI type						
	Classical	*Bootstrap BCa*					
	Via n'_μ [a]	*ARB* [a]	*MBB* [a]	*Timescale-ARB* [b]	*Timescale-MBB* [b]	*Pairwise-MBB* [a]	
10	0.81	0.66	0.56	0.925	0.907	0.70	0.950
20	0.71	0.64	0.56	0.941	0.927	0.68	0.950
50	0.59	0.57	0.47	0.950	0.943	0.54	0.950
100	0.50	0.49	0.42	0.949	0.949	0.49	0.950
200	0.39	0.38	0.34	0.949	0.948	0.36	0.950
500	0.25	0.25	0.24	0.950	0.950	0.24	0.950
1000	0.18	0.18	0.17	0.949	0.949	0.17	0.950

[a] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally 0.01.

[b] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally 0.001.

Tables 4.7 and 4.9 reveal also bad coverage performances of pairwise-MBB resampling. The reason is that this resampling algorithm is not suited for replicating the uncertainties when the timescale model has deterministic parts. Because this is inevitably the case for climate archives, which exhibit systematic accumulation processes (e.g., linear growth model, Fig. 4.3) and which are sampled not at random, the pairwise-MBB algorithm is not helpful for quantifying climatic trends. However, its potential for solving bivariate problems (where the data are given by $\{t(i), x(i), y(i)\}_{i=1}^{n}$) is further explored in Chapters 7 and 8.

For noise of lognormal shape and AR(1) persistence, both timescale-ARB and timescale-MBB resampling yielded excellent coverage performance in the presence of timescale errors (Table 4.10).

Table 4.8. Monte Carlo experiment, linear OLS regression with timescale errors and AR(1) noise of normal shape: RMSE and average CI length, slope. The entries for the case of absent timescale errors are overtaken from Table 4.2; for nonzero timescale errors the experiment described in Table 4.7 is used. See both tables also for details on regression model and noise. The number of simulations is $n_{\text{sim}} = 47{,}500$. The average CI length refers to bootstrap BCa CIs (ARB and timescale-ARB resampling).

n	$\text{RMSE}^{\text{a}}_{\widehat{\beta}_1}$		$\langle$ *CI length* $\rangle^{\text{b}}$	
	Timescale error		*Timescale error*	
	No	*Yes*	*No*[c]	*Yes*[d]
10	0.14049	0.21339	0.44414	0.81387
20	0.05332	0.09183	0.19118	0.35884
50	0.01411	0.03202	0.05316	0.12606
100	0.00506	0.01524	0.01941	0.05950
200	0.00179	0.00731	0.00697	0.02879
500	0.00046	0.00287	0.00178	0.01127
1000	0.00016	0.00142	0.00063	0.00560

[a] Empirical $\text{RMSE}_{\widehat{\beta}_1}$, given by $\left[\sum_{i=1}^{n_{\text{sim}}} \left(\widehat{\beta}_1 - \beta_1\right)^2 / n_{\text{sim}}\right]^{1/2}$.
[b] Average value over n_{sim} simulations.
[c] ARB resampling.
[d] Timescale-ARB resampling.

An interesting behaviour of the coverage performance is observed (Tables 4.11 and 4.12) for AR(2) persistence. As expected, also in the presence of timescale errors, the timescale-MBB resampling algorithm performs slightly better than the timescale-ARB algorithm, which mis-specifies the dependence (AR(1) instead of AR(2)). When the timescale is without errors, this mis-specification did make usage of the ARB nearly obsolete (and usage of MBB a duty). However, this distorting influence of the mis-specification on the coverage performance becomes smaller as the error of the timescale grows (Table 4.12). The reason is that in the experiments the timescale-ARB captures correctly the error proportion due to timescale uncertainties—this part receives more weight on coverage accuracy as the $S_{\text{date}}(j)$ values increase. This observation reiterates the importance of adequately modelling the accumulation and functional form of the age–depth curve in climate archives and accurately quantifying the size of dating errors.

Table 4.9. Monte Carlo experiment, linear OLS regression with timescale errors and AR(1) noise of normal shape: CI coverage performance, intercept. See Table 4.7 for details.

n	$\gamma_{\widehat{\beta}_0}$						*Nominal*
	CI type						
	Classical	*Bootstrap BCa*					
	Via n'_{μ} [a]	*ARB* [a]	*MBB* [a]	*Timescale-ARB* [b]	*Timescale-MBB* [b]	*Pairwise-MBB* [a]	
10	0.75	0.61	0.50	0.930	0.912	0.64	0.950
20	0.68	0.61	0.50	0.943	0.930	0.64	0.950
50	0.51	0.48	0.42	0.949	0.944	0.46	0.950
100	0.41	0.39	0.34	0.948	0.945	0.36	0.950
200	0.30	0.29	0.25	0.950	0.949	0.26	0.950
500	0.23	0.22	0.20	0.949	0.949	0.22	0.950
1000	0.15	0.15	0.15	0.951	0.951	0.15	0.950

[a] Standard error of $\gamma_{\widehat{\beta}_0}$ is nominally 0.01.
[b] Standard error of $\gamma_{\widehat{\beta}_0}$ is nominally 0.001.

Table 4.10. Monte Carlo experiment, linear OLS regression with timescale errors and AR(1) noise of lognormal shape: CI coverage performance. CI construction and generation of $T_{\text{true}}(i)$ and timescale errors were as in the experiment described in Table 4.7; likewise the generation of $X(i)$, with the difference that lognormal AR(1) noise (Table 4.3) instead of normal AR(1) noise was added; $n_{\text{sim}} = 47{,}500$.

n	$\gamma_{\widehat{\beta}_0}$ [a]		$\gamma_{\widehat{\beta}_1}$ [a]		*Nominal*
	Bootstrap BCa CI		*Bootstrap BCa CI*		
	Timescale-ARB	*Timescale-MBB*	*Timescale-ARB*	*Timescale-MBB*	
10	0.922	0.915	0.921	0.912	0.950
20	0.932	0.925	0.929	0.919	0.950
50	0.939	0.935	0.937	0.932	0.950
100	0.945	0.943	0.939	0.938	0.950
200	0.946	0.944	0.943	0.942	0.950
500	0.946	0.946	0.945	0.944	0.950
1000	0.949	0.949	0.948	0.948	0.950

[a] Standard errors of $\gamma_{\widehat{\beta}_0}$ and $\gamma_{\widehat{\beta}_1}$ are nominally 0.001.

Table 4.11. Monte Carlo experiment, linear OLS regression with timescale errors and AR(2) noise of normal shape: CI coverage performance. CI construction and generation of $T_{\mathrm{true}}(i)$ and timescale errors were as in the experiment described in Table 4.7; likewise the generation of $X(i)$, with the difference that normal AR(2) noise (parameters as in the experiment shown in Table 4.4) instead of normal AR(1) noise was added; $n_{\mathrm{sim}} = 47{,}500$.

n	$\gamma_{\widehat{\beta}_0}$ [a]		$\gamma_{\widehat{\beta}_1}$ [a]		*Nominal*
	Bootstrap BCa CI		*Bootstrap BCa CI*		
	Timescale-ARB	*Timescale-MBB*	*Timescale-ARB*	*Timescale-MBB*	
10	0.959	0.931	0.960	0.930	0.950
20	0.964	0.949	0.967	0.951	0.950
50	0.959	0.951	0.962	0.952	0.950
100	0.955	0.950	0.957	0.951	0.950
200	0.954	0.952	0.954	0.951	0.950
500	0.951	0.951	0.953	0.951	0.950
1000	0.950	0.949	0.953	0.952	0.950

[a] Standard errors of $\gamma_{\widehat{\beta}_0}$ and $\gamma_{\widehat{\beta}_1}$ are nominally 0.001.

Table 4.12. Monte Carlo experiment, linear OLS regression with AR(2) noise of normal shape: dependence on size of timescale errors. Generation of data ($n_{\mathrm{sim}} = 47{,}500$) were as in the previous experiments (Tables 4.4 and 4.7), employing a linear timescale model with $n_{\mathrm{date}} = 2$. The data size is fixed ($n = 50$). Shown are empirical coverages of bootstrap BCa CIs for the slope; $B = 1999$ and $\alpha = 0.05, 0.025$ and 0.005. Resampling algorithm is timescale-ARB.

S_{date} [a]	α					
	0.05		0.025		0.005	
	$\gamma_{\widehat{\beta}_1}$ [b]	*Nominal*	$\gamma_{\widehat{\beta}_1}$ [c]	*Nominal*	$\gamma_{\widehat{\beta}_1}$ [d]	*Nominal*
0.0	0.986	0.900	0.995	0.950	0.9994	0.9900
0.1	0.976	0.900	0.991	0.950	0.9989	0.9900
0.2	0.958	0.900	0.984	0.950	0.9981	0.9900
0.5	0.918	0.900	0.962	0.950	0.9938	0.9900
1.0	0.906	0.900	0.954	0.950	0.9907	0.9900
2.0	0.901	0.900	0.951	0.950	0.9904	0.9900

[a] Timescale error; $S_{\mathrm{date}}(1) = S_{\mathrm{date}}(2) = S_{\mathrm{date}}$.
[b] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally $\sim$0.0014.
[c] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally 0.001.
[d] Standard error of $\gamma_{\widehat{\beta}_1}$ is nominally $\sim$0.0005.

4.2 Nonlinear regression

The trend model $X_{\text{trend}}(i)$ is in climatology often more complex than a linear function. The suitability of a model is related to the time span analysed, and extending the span may require to adopt more complex models. Simple extensions are a parabola,

$$X(i) = \beta_0 + \beta_1 T(i) + \beta_2 T(i)^2 + S(i) \cdot X_{\text{noise}}(i), \tag{4.37}$$

with three parameters to be estimated, or a polynomial of general order. Because $T(i)^2$ can be viewed as a second regressor variable, the parabolic model, and the polynomial in general, can be solved using multivariate linear regression (von Storch and Zwiers 1999). Nonlinear regression models are nonlinear in the parameters. A simple example of a "real" nonlinear model is given by the exponential saturation function,

$$X(i) = \beta_0 \left\{1 - \exp\left[-\beta_1 T(i)\right]\right\} + S(i) \cdot X_{\text{noise}}(i). \tag{4.38}$$

Sometimes it is possible to transform the regressor in a way that a linear model results (von Storch and Zwiers 1999: Section 8.6.2 therein). However, this may be at the cost of the simplicity of the noise process, which is also transformed (Section 2.6). Owing to the parsimonious preference of the AR(1) noise model, it may in climatological practice be advisable to ignore transformations and carry out a nonlinear regression estimation.

The major difference to linear regression is that estimating nonlinear models normally requires elaborated numerical techniques because exact formulas, such as those for the OLS, WLS or GLS estimators (Section 4.1), do not often exist. Various fit criteria may be adopted, such as robustness or maximum likelihood.

The least-squares criterion leads to searching the point $\widehat{\boldsymbol{\beta}}$ in the parameter space where $SSQG(\boldsymbol{\beta})$ (under GLS) has a minimum. One usually makes an initial guess, $\widehat{\boldsymbol{\beta}}^{\mathbf{(0)}}$, calculates the sum of squares and its gradient and finds the next point, $\widehat{\boldsymbol{\beta}}^{\mathbf{(1)}}$, by going a step of defined size into the negative gradient direction. Gradient and step-size values are updated and the procedure repeated until stopping rules inform that the solution is sufficiently close to a minimum.

Classical CIs for estimated parameters of a nonlinear regression can be constructed using the gradient at the solution point $\widehat{\boldsymbol{\beta}}$. Making some "regularity assumptions" (Seber and Wild 1989: Chapter 12 therein), the distribution of $\widehat{\boldsymbol{\beta}}$ can be shown as asymptotically (for $n \rightarrow \infty$) normal, which enables CI construction and hypothesis tests. Because already the linear model (Section 4.1) applied to realistic climate time series (non-Gaussian shape, persistence, timescale errors) led to a tentative preference of bootstrap CIs, the additional assumptions and the difficulty to assess how far n differs from ∞ in terms of CI accuracy, cannot prevent us from considering only bootstrap CIs for the remainder of this section. In analogy to linear regression, the bootstrap resampling algorithms (ARB, MBB, timescale-ARB or timescale-MBB) are applied to the residuals from the estimated nonlinear regression.

The restriction to bootstrap CIs is further justified by our selection of two nonlinear models of climatic changes, namely the ramp and the break model. Both are not differentiable with respect to time, and hence no gradient and classical CI can be constructed for them.

4.2.1 Climate transition model: ramp

The ramp regression model (Fig. 4.5), written in continuous time as

$$\begin{aligned} X_{\text{trend}}(T) &= X_{\text{ramp}}(T) \qquad (4.39) \\ &= \begin{cases} x1 & \text{for } T \le t1, \\ x1 + (T - t1)(x2 - x1)/(t2 - t1) & \text{for } t1 < T \le t2, \\ x2 & \text{for } T > t2, \end{cases} \end{aligned}$$

has four parameters: start time $t1$, start level $x1$, end time $t2$ and end level $x2$. The attributes “start” and “end” mean that we assume without loss of generality that in Fig. 4.5 time increases from the left to the right.

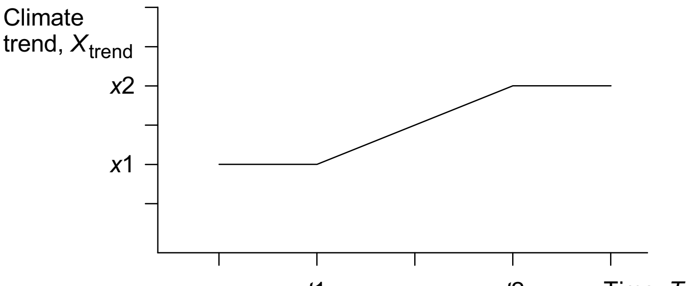


Figure 4.5. The ramp regression model. It has four parameters: $t1, x1, t2$ and $x2$.

The ramp is the simplest mathematical expression of a (climate) transition in $X_{\text{trend}}(T)$. Consider the questions: When did a transition start? When did it end? What were the levels before and after the transition? As soon as we ask those—and this is often the case in climate sciences—the ramp comes into play. As physical motivation serves a climate subsystem (described by, e.g., temperature) at equilibrium, which is disturbed by some external action (e.g., a volcanic eruption) over a period of time and then attains a new equilibrium state. The non-differentiability of the ramp with respect to time at $t1$ and $t2$ is only a mathematical inconvenience. Since the time we have accepted threshold behaviour or quantum phenomena in the physical world, we have also acknowledged that nature does indeed make jumps and requires for its description non-continuous and non-differentiable models.

4.2.1.1 Estimation

Assume the variability $S(i)$ known. The ramp regression model can then be fitted to data $\{t(i), x(i)\}_{i=1}^{n}$ by minimizing the weighted sum of squares,

$$SSQW(t1, x1, t2, x2) = \sum_{i=1}^{n} \left[x(i) - x_{\mathrm{ramp}}(i)\right]^2 / S(i)^2 \, , \tag{4.40}$$

where $x_{\mathrm{ramp}}(i)$ is the discrete-time, sample version of $X_{\mathrm{ramp}}(T)$ (Eq. 4.39).

Consider two candidate points, $\widetilde{t1} < \widetilde{t2}$, for the change-points, $t1$ and $t2$. Take them from the observed time points, that means, $\widetilde{t1} = t(\widetilde{i1})$ and $\widetilde{t2} = t(\widetilde{i2})$ with $1 \le \widetilde{i1} < \widetilde{i2} \le n$. Then the minimizers $\widehat{x1}, \widehat{x2}$ of $SSQW(\widetilde{t1}, x1, \widetilde{t2}, x2)$ follow (Mudelsee 2000) as

$$\begin{aligned} \widehat{x2} &= \left(K_3 K_4 / K_1 + K_6\right) / \left(K_2 K_4 / K_1 + K_5\right) , \\ \widehat{x1} &= \left(K_3 - \widehat{x2}\, K_2\right) \Big/ K_1, \end{aligned} \tag{4.41}$$

where

$$\begin{aligned} K_1 &= k_2 + \left(\widetilde{t1}\, k_4 - k_5\right) \Big/ \left(\widetilde{t2} - \widetilde{t1}\right), \\ K_2 &= k_3 - \left(\widetilde{t1}\, k_4 - k_5\right) \Big/ \left(\widetilde{t2} - \widetilde{t1}\right), \\ K_3 &= k_8, \\ K_4 &= k_1 + \left[\widetilde{t2}\left(\widetilde{t1} + \widetilde{t2}\right) k_4 + 2k_6 - \left(\widetilde{t1} + 3\widetilde{t2}\right) k_5\right] \Big/ \left(\widetilde{t2} - \widetilde{t1}\right)^2 , \\ K_5 &= k_3 + \left[\widetilde{t1}\left(\widetilde{t1} + \widetilde{t2}\right) k_4 + 2k_6 - \left(3\widetilde{t1} + \widetilde{t2}\right) k_5\right] \Big/ \left(\widetilde{t2} - \widetilde{t1}\right)^2 , \\ K_6 &= k_9 - k_7 - 2\left(\widetilde{t1}\, k_{10} - k_{11}\right) \Big/ \left(\widetilde{t2} - \widetilde{t1}\right) , \end{aligned} \tag{4.42}$$

and

$$k_1 = \sum_{i=1}^{\widetilde{i1}} S(i)^{-2}, \qquad k_2 = \sum_{i=1}^{\widetilde{i2}-1} S(i)^{-2}, \qquad k_3 = \sum_{i=\widetilde{i2}}^{n} S(i)^{-2},$$

$$k_4 = \sum_{i=\widetilde{i1}+1}^{\widetilde{i2}-1} S(i)^{-2}, \qquad k_5 = \sum_{i=\widetilde{i1}+1}^{\widetilde{i2}-1} t(i)\, S(i)^{-2},$$

$$k_6 = \sum_{i=\widetilde{i1}+1}^{\widetilde{i2}-1} t(i)^2\, S(i)^{-2}, \quad k_7 = \sum_{i=1}^{\widetilde{i2}-1} x(i)\, S(i)^{-2}, \tag{4.43}$$

$$k_8 = \sum_{i=1}^{n} x(i)\, S(i)^{-2}, \qquad k_9 = \sum_{i=\widetilde{i2}}^{n} x(i)\, S(i)^{-2},$$

$$k_{10} = \sum_{i=\widetilde{i1}+1}^{\widetilde{i2}-1} x(i)\, S(i)^{-2}, \quad k_{11} = \sum_{i=\widetilde{i1}+1}^{\widetilde{i2}-1} t(i)\, x(i)\, S(i)^{-2}.$$

To estimate the change-points in time, a brute-force search over all pairs of candidate points is performed because gradient techniques are inapplicable owing to the non-differentiability with respect to $t1$ and $t2$:

$$\left(\widehat{t1}, \widehat{t2}\right) = \operatorname{argmin} \left[SSQW \left(\widetilde{t1}, \widehat{x1}, \widetilde{t2}, \widehat{x2}\right)\right]. \tag{4.44}$$

Because the number of pairs of search points grows with the data size as $n(n-1)/2$, it is advisable to use computational measures to keep computing costs low (Section 4.5). A positive by-product of the brute-force search is that the solution is a global optimum. Because the candidate points $\widetilde{t1}$ and $\widetilde{t2}$ are from the set $\{t(i)\}_{i=1}^{n}$, the solution is as "coarse" as the spacing. This may be a problem when the spacing (at around $\widehat{t1}$ and $\widehat{t2}$) is larger than the standard errors, $\mathrm{se}_{\widehat{t1}}$ and $\mathrm{se}_{\widehat{t2}}$. However, in climatological applications this likely occurs only when we wish to quantify a climate transition using an archive with a hiatus located at around the place of a transition change-point.

Because in practice the variability $S(i)$ is unknown, an iterative estimation procedure via the residuals $e(i)$ is indicated (Section 4.1.1).

4.2.1.2 Example: Northern Hemisphere Glaciation

Application of the ramp model to the marine δ^{18}O record ODP 846 shows that the Northern Hemisphere Glaciation was a slow climate transition (Fig. 4.6). Whereas the $t1$ estimate of around 2.5 Ma before present is in general agreement with the climatological literature (Shackleton et al. 1984; Haug et al. 1999, 2005), the $t2$ estimate (3.7 Ma) is about 0.5 Ma earlier than what was previously thought. (Mudelsee and Raymo (2005) analysed a total of 45 δ^{18}O records using the ramp and found an average $\widehat{t2}$ of $\sim$ 3.6 Ma; the inter-record variation is likely caused by contrasting temperature trends.) As outliers (defined in the paper as more than $3S(i)$ away from the ramp fit), two prominent glaciation peaks (termed M2–MG2) appear at around 3.3 Ma (Fig. 4.6). These findings are robust against the estimation uncertainties (Fig. 4.6).

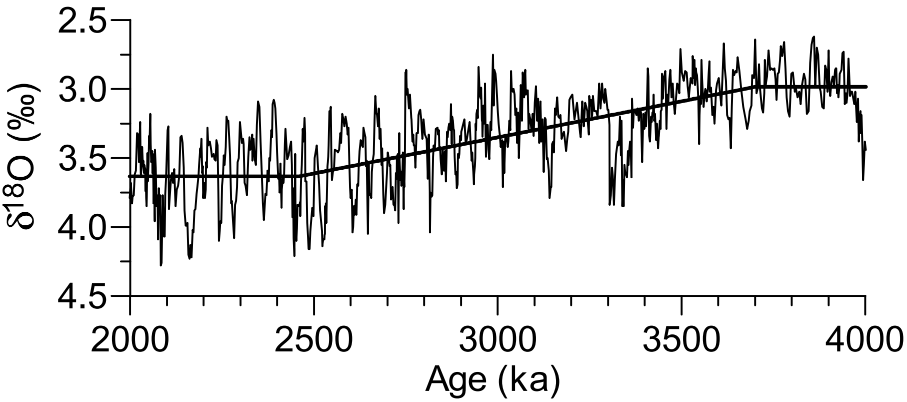


Figure 4.6. Ramp regression of the marine δ^{18}O record ODP 846 (Fig. 1.2). The long-term trend (*thick line*) documents the Northern Hemisphere Glaciation. The estimated change-points ($\pm$ standard errors) of this climate transition are: $\widehat{t1} = 2462 \pm 129$ ka, $\widehat{x1} = 3.63 \pm 0.04$‰, $\widehat{t2} = 3700 \pm 119$ ka and $\widehat{x2} = 2.99 \pm 0.04$‰. The estimates were obtained by WLS using iteratively updated variability. Also $S(i)$ was given a ramp form ($\widehat{S}(i) = 0.27$‰ for $t(i) \leq 2600$ ka, 0.18‰ for $t(i) > 3550$ ka and linearly connected between these change-points). The standard errors given here are from SB resampling (Section 3.8) using $B = 400$ and an average block length equal to $\widehat{\tau} = 7.7$ ka. A rough method to take into account the timescale uncertainty of 25 ka is Gaussian error propagation (Mudelsee and Raymo 2005), yielding total errors of $(129^2 + 25^2)^{1/2}$ ka ≈ 131 ka (for $\widehat{t1}$) and $(119^2 + 25^2)^{1/2}$ ka ≈ 122 ka (for $\widehat{t2}$). (After Mudelsee and Raymo 2005.)

4.2.1.3 Bootstrap confidence intervals

The unweighted residuals from an estimated ramp regression are given by

$$e(i) = x(i) - \widehat{x}_{\text{ramp}}(i), \qquad i = 1, \ldots, n, \tag{4.45}$$

where $\widehat{x}_{\text{ramp}}(i)$ is the discrete-time, sample version of $X_{\text{ramp}}(T)$ (Eq. 4.39), with estimates $(\widehat{t1}, \widehat{x1}, \widehat{t2}, \widehat{x2})$ plugged in for $(t1, x1, t2, x2)$. The $e(i)$ can be used to detect heteroscedasticity and assess the quality of the fit of the variability, $\widehat{S}(i)$. The weighted residuals from an estimated ramp are given by

$$r(i) = e(i) \Big/ \widehat{S}(i)\ , \qquad i = 1, \ldots, n. \tag{4.46}$$

The $r(i)$ are useful for studying model suitability of the ramp (graphically and arithmetically). They serve also for quantifying the persistence properties of the noise (e.g., autocorrelation estimation for an AR(1) process), which in turn are required for determining the bootstrap CIs (e.g., block length selection for the MBB).

The Monte Carlo experiment (Table 4.13) shows that accurate CIs for ramp parameters can be obtained when the sample size is suffiently large (above, say, 500). An interesting alternative to considering the change-points in time ($t1$, $t2$) may be analysing the parameters midpoint (given by $(t1 + t2)/2$), which performed better for smaller sample sizes in the experiment, or duration $(t2 - t1)$.

4.2.1.4 Example: onset of Dansgaard–Oeschger event 5

The onset of D–O event 5, a warming, was observed via the variables Ca content, dust content, electrical conductivity and Na content in the NGRIP ice core (Fig. 1.5). This climate transition can be excellently fitted by the ramp model (Fig. 4.7). The regression residuals, $r(i)$, exhibit somewhat right-skewed distributions with a few outliers (Fig. 1.11e–h) as well as persistence (Fig. 1.12e–h). Longer-term systematic deviations from the ramp form seem to be absent.

Answering the first question, after the synchroneity of the D–O 5 onset, Fig. 4.8 reveals that CIs of the observed change-points for all four variables do overlap. The hypothesis of synchroneity cannot be rejected. Bootstrap resampling (ARB) for CI construction did not employ timescale simulations because the four variables, measured on the same material, have identical timescales.

For answering the second question, how long the D–O 5 warming took, however, timescale uncertainties have to be taken into account (by timescale-ARB resampling) because in this context “absolute” values are sought. The covariation of sedimentation rate and δ^{18}O variations in the

Table 4.13. Monte Carlo experiment, ramp regression with timescale errors and AR(1) noise of normal shape: CI coverage performance. $n_{\text{sim}} = 475$ random samples were generated from $X(i) = X_{\text{ramp}}(i) + X_{\text{noise}}(i)$, where the prescribed ramp parameters (Eq. 4.39) are $t1 = 0.3n, x1 = 2.0, t2 = 0.7n$ and $x2 = 4.0$, the prescribed times are $T_{\text{true}}(i) = i, i = 1, \ldots, n$ and the noise is a Gaussian AR(1) process (Eq. 2.1) with $a = 1/e \approx 0.37$. Timescale errors were subsequently introduced as follows. A linear timescale model (Fig. 4.3) with $n_{\text{date}} = 2$ dating points and independent, Gaussian distributed timescale errors was used to generate the $T(i)$ as $T(i) = T_{\text{true}}(i) + \mathcal{E}_{\text{N}(0,\,25.0)}(i)$ for $i = 1$, $T(i) = T_{\text{true}}(i) + \mathcal{E}_{\text{N}(0,\,100.0)}(i)$ for $i = n$, and then by linear interpolation for $i = 2, \ldots, n-1$. The bootstrap BCa CIs were constructed with timescale-ARB resampling (Algorithm 4.5), $B = 1999$ and $\alpha = 0.025$.

n	$\gamma^{a}_{\widehat{t1}}$	$\gamma^{a}_{\widehat{x1}}$	$\gamma^{a}_{\widehat{t2}}$	$\gamma^{a}_{\widehat{x2}}$	$\gamma^{a}_{(\widehat{t1}+\widehat{t2})/2}$	$\gamma^{a}_{(\widehat{t2}-\widehat{t1})}$	*Nominal*
10	0.92	0.64	0.91	0.66	0.93	0.77	0.95
20	0.88	0.75	0.84	0.72	0.90	0.60	0.95
50	0.76	0.84	0.77	0.83	0.89	0.68	0.95
100	0.76	0.88	0.77	0.87	0.89	0.76	0.95
200	0.86	0.90	0.85	0.91	0.89	0.84	0.95
500	0.94	0.95	0.94	0.94	0.94	0.93	0.95
1000	0.96	0.96	0.93	0.93	0.95	0.95	0.95

[a] Standard error of γ is nominally 0.01.

NGRIP ice core (Fig. 4.9) reflects that timescale construction (Johnsen et al. 2001) made the reasonable assumption that elevated temperatures (indicated by higher δ^{18}O) lead to enhanced ice accumulation. The ratio is approximately 6‰ δ^{18}O change per 1.8 cm/a sedimentation rate change for the NGRIP core at around D–O 5. The δ^{18}O measurement uncertainty of 0.1‰ (North Greenland Ice Core Project members 2004) can be used to simulate timescale uncertainties as follows. The first point is fixed, $t^*(1) = t(1)$. The second point is modelled as $t^*(n) = t(1) + [t(n) - t(1)]/(1 + \mathcal{E}_{\text{N}(0,\,1)}(n) \cdot 0.1/6)$. This rough procedure is applicable for duration, but not for midpoint estimation, because only the timescale uncertainties within the D–O 5 interval were modelled. The result (Fig. 4.10) shows that the warming was completed within about 100 years.

Although the numerical results are somewhat preliminary owing to various open technical questions (logarithmic transformation of variables, linearity of δ^{18}O/sedimentation rate changes and other uncertainties at timescale construction), it is clear that ramp regression can add to the quantitative understanding of D–O events.

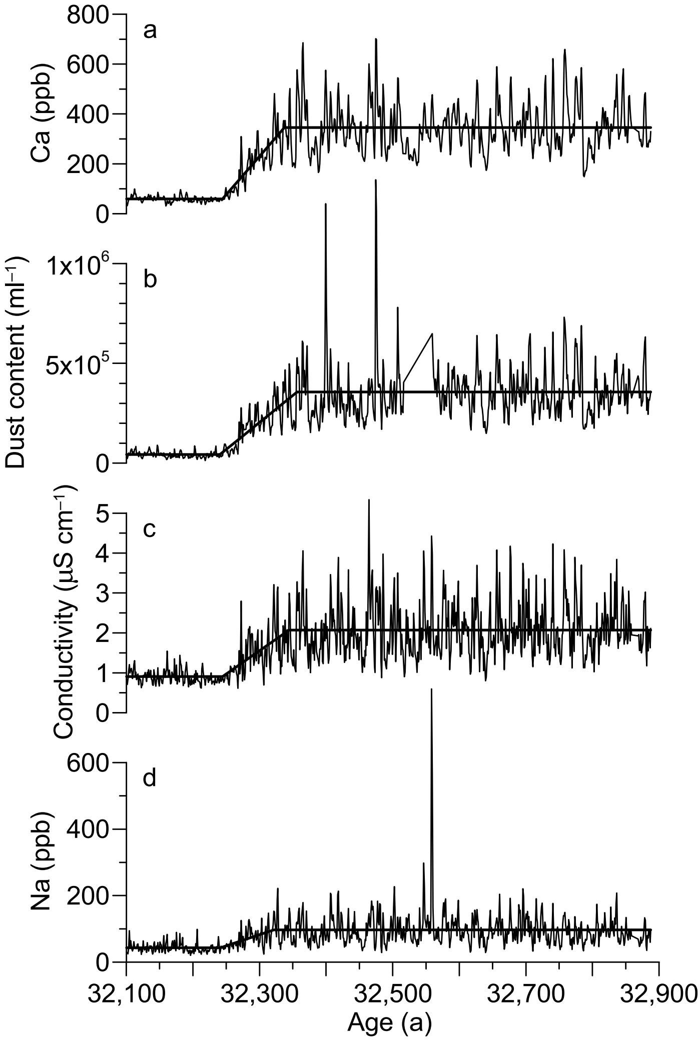


Figure 4.7. Onset of Dansgaard–Oeschger event 5, NGRIP ice core: result. Ramps (*thick lines*) were fitted to the Ca, dust content, conductivity and Na records (Fig. 1.5). The estimated change-points are **a** $\widehat{t1} = 32{,}245$ a, $\widehat{x1} = 59$ ppb, $\widehat{t2} = 32{,}338$ a, $\widehat{x2} = 346$ ppb, $n = 770$; **b** $\widehat{t1} = 32{,}241$ a, $\widehat{x1} = 0.43 \cdot 10^5$ ml^{-1}, $\widehat{t2} = 32{,}357$ a, $\widehat{x2} = 3.57 \cdot 10^5$ ml^{-1}, $n = 727$; **c** $\widehat{t1} = 32{,}245$ a, $\widehat{x1} = 0.91$ µS cm^{-1}, $\widehat{t2} = 32{,}345$ a, $\widehat{x2} = 2.07$ µS cm^{-1}, $n = 775$ and **d** $\widehat{t1} = 32{,}242$ a, $\widehat{x1} = 43$ ppb, $\widehat{t2} = 32{,}322$ a, $\widehat{x2} = 97$ ppb, $n = 774$. The estimates were obtained by WLS using iteratively updated $S(i)$. The $S(i)$ fits adopted a ramp model to account for the heteroscedasticities (higher variabilities in the earlier (colder) part); the $\widehat{S}(i)$ change-points are **a** (32,200 a, 16 ppb)–(32,350 a, 95 ppb); **b** (32,200 a, $0.2 \cdot 10^5$ ml^{-1})–(32,350 a, $1.3 \cdot 10^5$ ml^{-1}); **c** (32,200 a, 0.2 µS cm^{-1})–(32,400 a, 0.7 µS cm^{-1}) and **d** (32,200 a, 16 ppb)–(32,500 a, 55 ppb). Bootstrap CIs for the estimated time parameters are shown in Figs. 4.8 and 4.10.

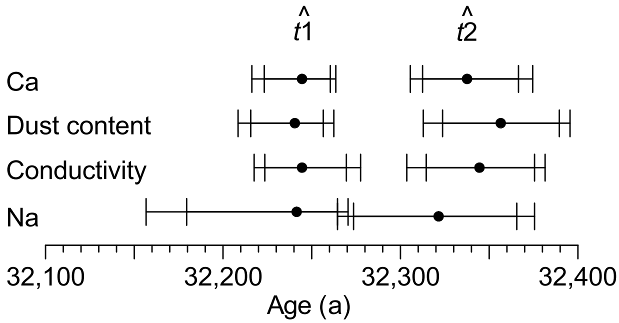


Figure 4.8. Onset of Dansgaard–Oeschger event 5, NGRIP ice core: estimated change-points with confidence intervals. Shown are 95 and 90% BCa CIs for $\widehat{t1}$ and $\widehat{t2}$, calculated with ARB resampling.

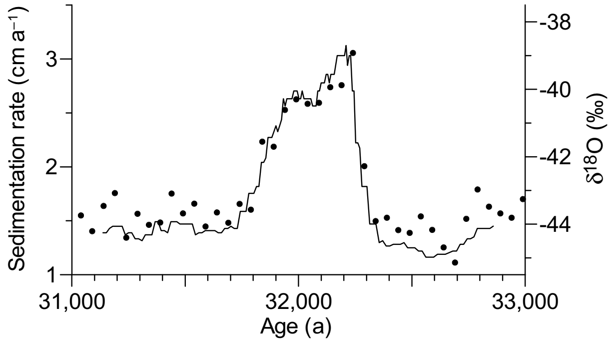


Figure 4.9. Onset of Dansgaard–Oeschger event 5, NGRIP ice core: sedimentation rate (*solid line*) and δ^{18}O (*dots*) variations. (δ^{18}O data from North Greenland Ice Core Project members 2004.)

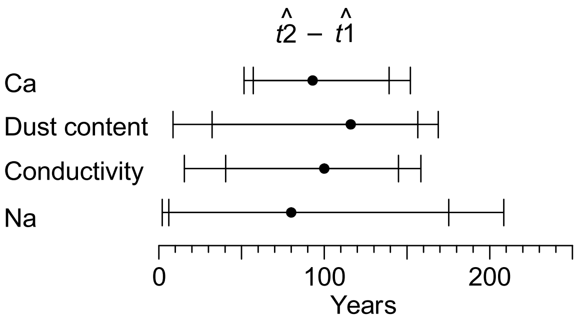


Figure 4.10. Onset of Dansgaard–Oeschger event 5, NGRIP ice core: estimated durations with confidence intervals. Shown are 95 and 90% BCa CIs for the duration of the onset, calculated with timescale-ARB resampling.

4.2.2 Trend-change model: break

The break regression model (Fig. 4.11), written in continuous time as

$$X_{\text{break}}(T) = \begin{cases} x1 + (T - t1)(x2 - x1)/(t2 - t1) & \text{for } T \le t2, \\ x2 + (T - t2)(x3 - x2)/(t3 - t2) & \text{for } T > t2, \end{cases} \quad (4.47)$$

has four free parameters: $x1$, $t2$, $x2$ and $x3$. An alternative formulation would comprise the four parameters $t2$, $x2$, $\beta_1 = (x2 - x1)/(t2 - t1)$ and $\beta_2 = (x3 - x2)/(t3 - t2)$. Also the break is a simple mathematical model. It can be useful for describing a change in linear trend at one point $(t2, x2)$, from slope β_1 to β_2.

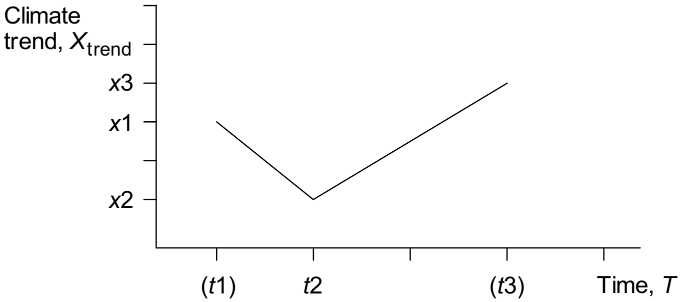


Figure 4.11. The break regression model. It has four free parameters: $x1, t2, x2$ and $x3$. (Parameter $t1$ is constrained as left, $t3$ as right bound of the time interval.)

4.2.2.1 Estimation

Assume known variability $S(i)$ and time series data $\{t(i), x(i)\}_{i=1}^{n}$. The break model can then be fitted by minimizing the weighted least-squares sum,

$$SSQW(x1, t2, x2, x3) = \sum_{i=1}^{n} \left[x(i) - x_{\text{break}}(i)\right]^2 / S(i)^2 \, , \quad (4.48)$$

where $x_{\text{break}}(i)$ is the discrete-time, sample version of $X_{\text{break}}(T)$ (Eq. 4.47).

Because we assume that the break is a suitable description over the whole record length, $t1$ and $t3$ are constrained (Fig. 4.11) and only one time point, namely $\widetilde{t2} = t(\widetilde{i2})$ with $1 \le \widetilde{i2} \le n$, needs to be considered as candidate for $t2$. Then the minimizers $\widehat{x1}, \widehat{x2}$ and $\widehat{x3}$ of

$SSQW(x1, \widetilde{t2}, x2, x3)$ follow as

$$\begin{aligned}
\widehat{x2} &= \big(K_1 K_2/K_3 - K_4 K_5/K_6 - K_7 + K_9\big) \\
&\quad \times \big(K_1^2/K_3 + K_4^2/K_6 - K_8 - K_{10}\big)^{-1}, \\
\widehat{x1} &= \widehat{x2}\, K_1\big/K_3 - K_2/K_3, \\
\widehat{x3} &= \widehat{x2}\, K_4\big/K_6 + K_5/K_6,
\end{aligned} \tag{4.49}$$

where

$$\begin{aligned}
K_1 &= \sum_{i=1}^{\widetilde{i2}} S(i)^{-2}\big[t(i) - t1\big]\big[t(i) - \widetilde{t2}\big]\big/\big[\widetilde{t2} - t1\big]^2, \\
K_2 &= \sum_{i=1}^{\widetilde{i2}} S(i)^{-2} x(i)\big[t(i) - \widetilde{t2}\big]\big/\big[\widetilde{t2} - t1\big], \\
K_3 &= \sum_{i=1}^{\widetilde{i2}} S(i)^{-2}\big[t(i) - \widetilde{t2}\big]^2\big/\big[\widetilde{t2} - t1\big]^2, \\
K_4 &= \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2}\big[t(i) - t3\big]\big[t(i) - \widetilde{t2}\big]\big/\big[t3 - \widetilde{t2}\big]^2, \\
K_5 &= \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2} x(i)\big[t(i) - \widetilde{t2}\big]\big/\big[t3 - \widetilde{t2}\big], \\
K_6 &= \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2}\big[t(i) - \widetilde{t2}\big]^2\big/\big[t3 - \widetilde{t2}\big]^2, \\
K_7 &= \sum_{i=1}^{\widetilde{i2}} S(i)^{-2} x(i)\big[t(i) - t1\big]\big/\big[\widetilde{t2} - t1\big], \\
K_8 &= \sum_{i=1}^{\widetilde{i2}} S(i)^{-2}\big[t(i) - t1\big]^2\big/\big[\widetilde{t2} - t1\big]^2, \\
K_9 &= \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2} x(i)\big[t(i) - t3\big]\big/\big[t3 - \widetilde{t2}\big], \\
K_{10} &= \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2}\big[t(i) - t3\big]^2\big/\big[t3 - \widetilde{t2}\big]^2.
\end{aligned} \tag{4.50}$$

To estimate the change-point in time, $t2$, a brute-force search over all candidate points is performed:

$$\left(\widehat{t2}\right) = \operatorname{argmin}\left[SSQW\left(\widehat{x1}, \widetilde{t2}, \widehat{x2}, \widehat{x3}\right)\right]. \tag{4.51}$$

Computing costs are clearly reduced (by a factor of $\sim n$) compared with estimating the ramp model (Section 4.2.1). The other properties the break shares: global optimum, "coarse" $\widehat{t2}$ estimate and applicability of an iterative procedure when $S(i)$ is unknown.

4.2.2.2 Example: Arctic river runoff (continued)

Application of the break trend-change regression to the modelled record of Arctic river runoff with combined anthropogenic and natural forcing reveals a change-point at $\widehat{t2} = 1973 \pm 6$ (Fig. 4.12). This date is close to the per-eye estimate (Wu et al. 2005) of 1965 (Fig. 4.1b). Before the change the trend is downwards, however, with large error bars; after the change it is strongly upwards (Fig. 4.12). The runoff modelled with natural forcing only, however, exhibits no significant slope changes of the break fit, which agrees with the result from the linear fit (Fig. 4.1a).

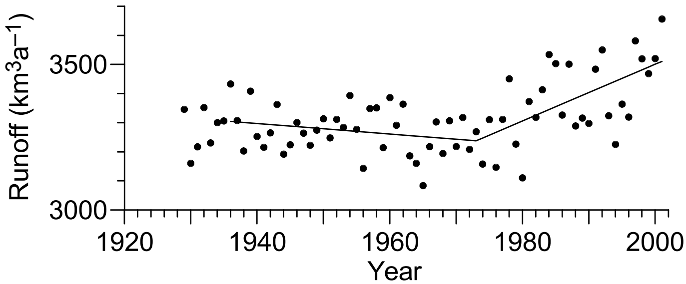


Figure 4.12. Break change-point regression fitted to modelled Arctic river runoff. Shown is the simulation with combined anthropogenic and natural forcing (Fig. 1.9b). For the interval 1936–2001 ($n = 66$), the break model was fitted using WLS (with $S(i)$ linearly increasing from 70 to 120 $\mathrm{km^3a^{-1}}$ within the fit interval). The break fit (*solid line*) has following parameter estimates with bootstrap standard errors (MBB, $B = 400$): change-point, $\widehat{t2} = 1973 \pm 6$, $\widehat{x2} = 3238 \pm 26$ $\mathrm{km^3a^{-1}}$; slopes, $\widehat{\beta}_1 = -1.8 \pm 1.6$ $\mathrm{km^3a^{-2}}$, $\widehat{\beta}_2 = 9.7 \pm 3.6$ $\mathrm{km^3a^{-2}}$.

4.2.2.3 Bootstrap confidence intervals

The unweighted residuals from an estimated break model are given by $e(i) = x(i) - \widehat{x}_{\text{break}}(i), i = 1, \ldots, n$; and the weighted residuals are given

Table 4.14. Monte Carlo experiment, break regression with timescale errors and AR(1) noise of normal shape: CI coverage performance. $n_{\text{sim}} = 475$ random samples were generated from $X(i) = X_{\text{break}}(i) + X_{\text{noise}}(i)$, where the prescribed break model parameters (Eq. 4.47) are $x1 = 2.0, t2 = 0.5n, x2 = 1.0$ and $x3 = 4.0$, the prescribed times are $T_{\text{true}}(i) = i, i = 1, \ldots, n$ and the noise is a Gaussian AR(1) process (Eq. 2.1) with $a = 1/\mathrm{e} \approx 0.37$. Timescale error simulations were performed as in the experiment on ramp regression (Table 4.13). Bootstrap BCa CIs used timescale-ARB resampling (Algorithm 4.5), $B = 1999$ and $\alpha = 0.025$.

n	$\gamma^{\mathrm{a}}_{\widehat{x1}}$	$\gamma^{\mathrm{a}}_{\widehat{t2}}$	$\gamma^{\mathrm{a}}_{\widehat{x2}}$	$\gamma^{\mathrm{a}}_{\widehat{x3}}$	$\gamma^{\mathrm{a}}_{\widehat{\beta}_1}$	$\gamma^{\mathrm{a}}_{\widehat{\beta}_2}$	*Nominal*
10	0.63	0.92	0.63	0.62	0.64	0.82	0.95
20	0.74	0.90	0.76	0.76	0.75	0.83	0.95
50	0.87	0.89	0.85	0.89	0.87	0.90	0.95
100	0.91	0.88	0.89	0.93	0.90	0.92	0.95
200	0.93	0.95	0.93	0.92	0.92	0.94	0.95
500	0.94	0.95	0.95	0.96	0.95	0.96	0.95
1000	0.95	0.95	0.95	0.95	0.95	0.94	0.95

[a] Standard error of γ is nominally 0.01.

by $r(i) = e(i)/\widehat{S}(i), i = 1, \ldots, n$; analogously to ramp regression (Section 4.2.1.3). Also here the residuals serve for studying model suitability and running the bootstrap resampling technique for CI calculation. The Monte Carlo experiment (Table 4.14) reveals that for data sizes above 100–200, the time parameter ($t2$), the level parameters ($x1, x2, x3$) and the slopes (β_1, β_2) have excellent coverage performance also for heteroscedastic timescale errors.

4.3 Nonparametric regression or smoothing

4.3.1 Kernel estimation

Instead of identifying $X_{\text{trend}}(T)$ with a specific linear or nonlinear function with parameters to be estimated, the smoothing method estimates $X_{\text{trend}}(T)$ at a time point T' by, loosely speaking, averaging the data points $X(i)$ within a neighbourhood around T'. A simple example is the running mean, where the points inside a window are averaged and the window runs along the time axis. Statistical science recommends to replace the non-smooth weighting window (points inside receive constant, positive weight and points outside zero weight) by a smooth kernel function, K. A nonparametric kernel regression estimator of the trend is given by (Priestley and Chao 1972)

$$\widehat{X}^{\text{PC}}_{\text{trend}}(T) = h^{-1} \sum_{i=1}^{n} [T(i) - T(i-1)] \; K\left[\frac{T - T(i)}{h}\right] \; X(i), \qquad (4.52)$$

where h is denoted as bandwidth (the index "PC" refers to the paper). Another kernel estimator is (Gasser and Müller 1979, 1984)

$$\widehat{X}^{\mathrm{GM}}_{\mathrm{trend}}(T) = h^{-1} \sum_{i=1}^{n} \left[\int\limits_{s(i-1)}^{s(i)} K\left(\frac{T-y}{h}\right) dy \right] X(i), \tag{4.53}$$

where $T(i-1) \leq s(i-1) \leq T(i)$ (e.g., $s(i-1) = [T(i-1) + T(i)]/2$ with $s(0)$ and $s(n)$ being the upper and lower bounds of the T interval, respectively).

The kernel is a continuous and usually positive and symmetric function, it integrates as $\int K(y)dy = 1$. Common choices are the Gaussian, $K(y) = (2\pi)^{-1/2} \exp(-y^2/2)$, and the Epanechnikov kernel, $K(y) = 0.75(1 - y^2)$ with $|y| \leq 1$. Whereas the choice of the particular kernel is more of "cosmetic" (Diggle 1985) interest, bandwidth selection is a crucial part because this determines bias and variance properties of $\widehat{X}_{\mathrm{trend}}(T)$. Several techniques exist for that purpose; one of which is cross-validation, where a cost function, composed of a bias term and a variance term, is minimized. A later paragraph here details cross-validation in the context of running median smoothing, Chapter 6 does so in the context of occurrence rate estimation of extreme events. Bandwidth selection for data with serial dependence, such as climate time series, can be considerably more difficult than in dependence-free situations (see background material). A general advice is to "play" with h and study the sensitivity of results on h. An option is to use downsampled time series for determining h on data with less serial dependence. Figure 4.13 shows a nonparametric kernel regression of the sedimentation rate in the ice core from Vostok; Fig. 4.14 analogously for the atmospheric Δ^{14}C content.

Also variability estimation can be based on nonparametric regression. For example (Gasser–Müller kernel with bandwidth h), calculate the unweighted residuals,

$$e(i) = x(i) - \widehat{x}^{\mathrm{GM},\,h}_{\mathrm{trend}}(i), \qquad i = 1, \ldots, n, \tag{4.54}$$

and either fit a parametric model of $S(i)$ to the $e(i)$ or apply again smoothing to the $e(i)$. Utilizing prior knowledge, if existent, is advisable. This may regard parametric forms of $S(i)$ or typical timescales on which $S(i)$ varies, which would then facilitate bandwidth selection for nonparametric $S(i)$ estimation. In principle, $S(i)$ can also be estimated together with $X_{\mathrm{trend}}(i)$ in the same smoothing window. (A later paragraph gives such an example in the context of $X_{\mathrm{out}}(i)$ estimation, which

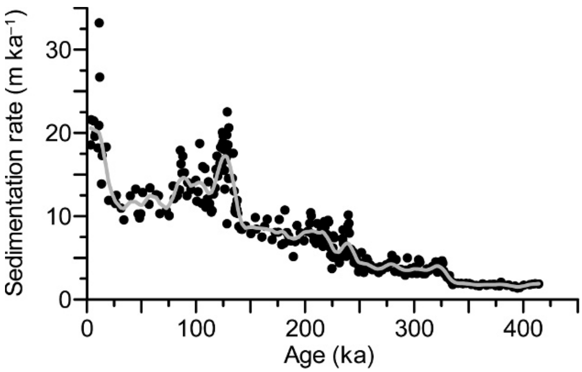


Figure 4.13. Nonparametric regression of the sedimentation rate in the Vostok record. The sedimentation rate (*dots*) $\dot{z} = [z(i+1) - z(i)]/[t(i+1) - t(i)]$, where z is depth, is calculated from the CO_2 data on the GT4 timescale (Fig. 1.3b). (Strictly speaking, $\dot{z}$ refers not to the "sedimentation" of the ice but to the derivative of the depth–age curve after ice accumulation and compaction.) The smoothed curve (*solid grey line*) is $\widehat{x}_{\text{trend}}^{\text{GM}}(t)$ calculated with a parabolic kernel, a cross-validated bandwidth of $h = 8$ ka and boundary adjustments ("boundary kernel").

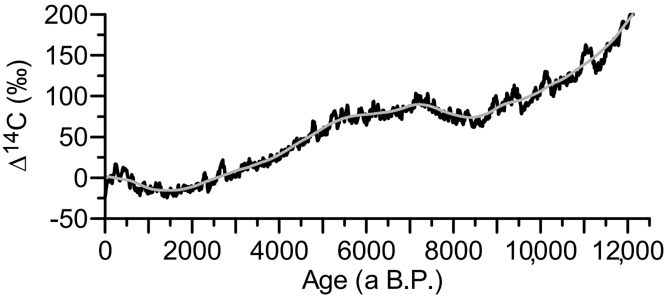


Figure 4.14. Nonparametric regression of the atmospheric radiocarbon record from tree-rings. The original data (Fig. 1.6) are shown as *black line*. The smoothed curve (*grey line*) is $\widehat{x}_{\text{trend}}^{\text{GM}}(t)$ calculated with a parabolic kernel, a bandwidth of $h = 580$ a and boundary adjustments. The bandwidth was determined by applying cross-validation to 200-year averaged segments of the original record to exlude autocorrelation effects stemming from the residence time of a CO_2 molecule in the atmosphere.

we have ignored in Eq. (4.54).) The weighted residuals follow as

$$r(i) = e(i) \Big/ \widehat{S}(i), \qquad i = 1, \ldots, n. \tag{4.55}$$

4.3.2 Bootstrap confidence intervals and bands

In principle, CI construction for $\widehat{X}_{\text{trend}}^{\text{GM},\,h}(i)$ (and for other kernels) at a point $i = i^{\dagger}$ could be tried by using the $r(i)$ from Eq. (4.55) for bootstrap resampling. Two notable improvements can be made at this step. First, because $\widehat{X}_{\text{trend}}^{\text{GM},\,h}(i)$ has larger bias in regions i where fewer data points exist, particularly near the interval bounds $i = 1$ or $i = n$, than in higher-density regions, and because the larger bias affects also $r(i)$ negatively in the lower-density regions, an adapted version of the local bootstrap may be appropriate (Davison and Hinkley 1997). In this version, the points within a neighbourhood (say, within $\pm 3h$) of the interval bounds are excluded from being resampled. Second, because the bias, which is inherent to nonparametric regression estimates, distorts also the residuals, the $e(i)$ (Eq. 4.54) and $r(i)$ (Eq. 4.55), should, for the purpose of providing samples to draw the $r^*(i)$ from, be calculated with a larger bandwidth, $h' > h$. This oversmoothing is detailed by Härdle (1990: Section 4.2 therein). It may, however, be that adopting BCa CIs reduces bias effects in nonparametric regression. Once appropriate $r(i)$ are found, resampling and CI construction (i.e., $\text{CI}_{\widehat{X}_{\text{trend}}^{\text{GM},\,h}(i^{\dagger}),1-2\alpha}$) proceeds for nonparametric regression as in the parametric cases (Sections 4.1 and 4.2).

A confidence band around the estimated nonparametric regression function helps to assess the significance of highs, lows and other features in the data. A pointwise confidence band is readily drawn by connecting the upper and lower bounds of $\text{CI}_{\widehat{X}_{\text{trend}}^{\text{GM},\,h}(i),1-2\alpha}$ for $i = 1, \ldots, n$ (Gasser–Müller kernel). Something different is a simultaneous confidence band, namely a compact set of points (T, X) that contains the line $X_{\text{trend}}(T)$ with a pre-defined probability, $1-2\alpha$. The explanation is that in the case of a pointwise band, at every position $i = 1, \ldots, n$ there is a chance of 2α to fall outside the CI. This is in analogy to the multiplicity of statistical tests in, for example, spectral analysis (Chapter 5). Construction of a simultaneous confidence band at level $1 - 2\alpha$ could be achieved by using a pointwise band constructed from CIs at level $1-2\alpha'$, with $\alpha' < \alpha$. The difficulty arises from the considerable amount of positive autocorrelation in the series of upper or lower CI bounds, which stems mainly from the smoothing procedure (plus some climate persistence). Therefore the simple setting $\alpha' = \alpha/n$ would fail. A quick and dirty setting is $\alpha' = \min\{\alpha, \alpha \cdot 3h'/[t(n) - t(1)]\}$, where h' is the bandwidth used for CI construction (see preceding paragraph). The idea is that at $3h'$ distance, the CI points are not, or at least not strongly, autocorrelated. More elaborated approaches to constructing simultaneous confidence bands are given by Härdle (1990: Section 4.3 therein).

4.3.3 Extremes or outlier detection

Nonparametric regression can also be used as a tool to detect outliers, $X_{\mathrm{out}}(i)$, which we up to now have largely assumed to be absent. Starting from the climate equation, $X(i) = X_{\mathrm{trend}}(i) + X_{\mathrm{out}}(i) + S(i) \cdot X_{\mathrm{noise}}(i)$, it is reasonable to introduce a threshold detection parameter, z, and define a positive extreme as follows. If

$$X(i) > X_{\mathrm{trend}}(i) + z \cdot S(i), \tag{4.56}$$

then

$$X_{\mathrm{out}}(i) \neq 0 \tag{4.57}$$

and, perhaps of more practical relevance, $T(i)$ is the date this positive extreme occurred. If the threshold is not exceeded, then $X_{\mathrm{out}}(i)$ is zero. Negative extremes may be defined analogously.

$X_{\mathrm{out}}(i)$ is a general description that allows to include an outlier component in the climate equation. We should employ additional, quantitative measures, such as the exceedance, $X'_{\mathrm{out}}(i) = X(i) - X_{\mathrm{trend}}(i) - z \cdot S(i)$. Or, we may define $X'_{\mathrm{out}}(i) = [X(i) - X_{\mathrm{trend}}(i)]/S(i)$ to have a dimensionless, scaled version. The extreme value analysis in Chapter 6 is based on $X'_{\mathrm{out}}(i)$ and the dates at which an extreme occurred.

To detect climate extremes on the sample level in time series data $\{t(i), x(i)\}_{i=1}^{n}$, it is essential to quantify $X_{\mathrm{trend}}(i)$, which is here denoted as "background," and $S(i)$ robustly, without interference by the "signal," $X_{\mathrm{out}}(i)$. Non-robust estimators, such as the running mean for $X_{\mathrm{trend}}(i)$ and the running standard deviation for $S(i)$ estimation, are therefore obsolete. A suitable tool for $X_{\mathrm{trend}}(i)$ estimation is the running median, calculated from $2k + 1$ points inside a pointwise shifted window (corresponding to a uniform kernel). Likewise, the running MAD is suited for a robust $S(i)$ estimation (Fig. 4.15).

Two detection parameters, z (threshold detection parameter) and k (defining the smoothing bandwidth), have to be adjusted. Hampel (1985) made extensive Monte Carlo simulations of extremes detection on background distributions "contaminated" with (distant) extreme distributions and concluded that $z = 3.5$ yields good detection rates. In a practical application it is advisable to try also more conservative (z larger) and more liberal (z smaller) settings and study the sensitivity of the results. Two cross-validation criteria for k selection in running-median smoothing seem to be useful for extremes detection owing to their robustness, namely L_1-norm (Marron 1987) and median criterion

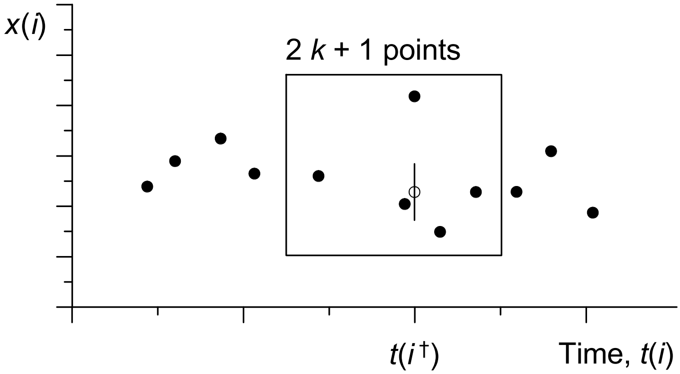


Figure 4.15. Outlier detection. The time point analysed for an outlier is $t(i^\dagger)$. The background or trend value is estimated as the median over $2k+1$ points inside a window, $\widehat{X}_{\text{trend}}(i^\dagger) = \widehat{M}\left\{X(i)\right\}_{i=i^\dagger-k}^{i^\dagger+k}$; the variability around that value is estimated as the median of absolute distances to the median (MAD) over the $2k+1$ points, $\widehat{S}(i^\dagger) = \widehat{M}\left\{\left|X(j) - \widehat{M}\left\{X(i)\right\}_{i=i^\dagger-k}^{i^\dagger+k}\right|\right\}_{j=i^\dagger-k}^{i^\dagger+k}$. In the example, the data points are shown as *filled symbols*, $k=2$ and the background $\pm$ variability estimate is shown as *open symbol* with *vertical bars*. Adopting a (likely too low) threshold detection parameter of $z=2$ would make $t(i^\dagger)$ a detected extreme because $x(i^\dagger)$ is more than $2\widehat{S}(i^\dagger)$ away from $\widehat{X}_{\text{trend}}(i^\dagger)$; a threshold detection parameter of $z=3.5$ would reject the point as an extreme. For detecting outliers in the whole series, the window is pointwise shifted. A simple solution for outlier detection near the interval bounds $(i \rightarrow 1, i \rightarrow n)$ is to extrapolate background and variability there.

(Zheng and Yang 1998):

$$C_1(k) = \left[\sum_{i=1}^{n}\left|x(i) - \widehat{m}\left\{x(j)\right\}_{j=i-k,\, j\neq i}^{i+k}\right|\right] \Big/ \, n, \tag{4.58}$$

$$C_{\mathrm{m}}(k) = \widehat{m}\left\{\left|x(i) - \widehat{m}\left\{x(j)\right\}_{j=i-k,\, j\neq i}^{i+k}\right|\right\}_{i=1}^{n}, \tag{4.59}$$

where $\widehat{m}\left\{x(j)\right\}_{j=i-k,\, j\neq i}^{i+k}$ is the delete-one background estimate. Optimal k values minimize $C_1(k)$ or $C_{\mathrm{m}}(k)$. (One leaves out the point $j=i$ to exclude the trivial solution $k=0$.) However, because those criteria assume absent serial correlation, it is important in a practical application with persistent time series to try different k values and study the sensitivity. Also local minima of the cross-validation functions may indicate some relevant structure (Marron 1988).

4.3.3.1 Example: volcanic peaks in the NGRIP sulfate record

Figure 4.16 shows detection of extremes in the annually resolved sulfate record from the NGRIP ice core (Fig. 1.4). The background variations are thought to represent fluctuations of oceanic sulfate input, against which the sulfate peaks from volcanic eruptions have to be detected. Sulfate peaks from other sources (dust and salt) have been removed prior to the analysis using information from proxy records of those disturbing variables (Fig. 1.4).

Bandwidth selection for background estimation does not resort to cross-validation because of the considerable amount of positive autocorrelation, visible already per eye (Fig. 4.16c). Instead we set $k = 750$, which means a running window of width $\sim$ 1500 a, because this is a typical timescale on which D–O climatic variations over Greenland and the North Atlantic occurred during the late Pleistocene and Holocene (Bond et al. 1997, 2001; Schulz 2002); and therefore also the oceanic sulfate input may have varied so.

Threshold setting does not follow Hampel's (1985) rule but tries two conservative values ($z = 5.0, 10.0$) in an attempt to "guarantee" that peaks do stem from heavy volcanic eruptions, at the cost of missing minor eruptions. A further point is autocorrelation within the peaks, that means, when a threshold is exceeded for a few successive years. This phenomenon is possibly (Bigler M 2003, personal communication) owing to the injection of eruption material into the stratosphere (upper part of the atmosphere), where it can reside for longer time. In such cases, only the maximum is retained (Fig. 4.16c) for further analysis of the occurrence of the volcanic peaks (Chapter 6).

4.3.3.2 Example: hurricane peaks in the Lower Mystic Lake varve thickness record

Figure 4.17 shows detection of extremes in the varve thickness record from the Lower Mystic Lake in the Boston area (Fig. 1.8). The background variations represent a combination of natural and anthropogenic (colonization, from $\sim$ 1630) factors.

The character of the laminated sedimentation changed after around 1870, owing to population growth, industrialization in the watershed and permanent alteration of the lake's natural hydraulic regime due to dam building (Besonen et al. 2008). To prevent the influence of these inhomogeneity factors on the detection of peaks in varve thickness, the twentieth century part of the lake core is not considered. The task is to detect the peaks in varve thickness for the shown interval, which are interpreted to have arisen from hurricanes that moved through the site region.

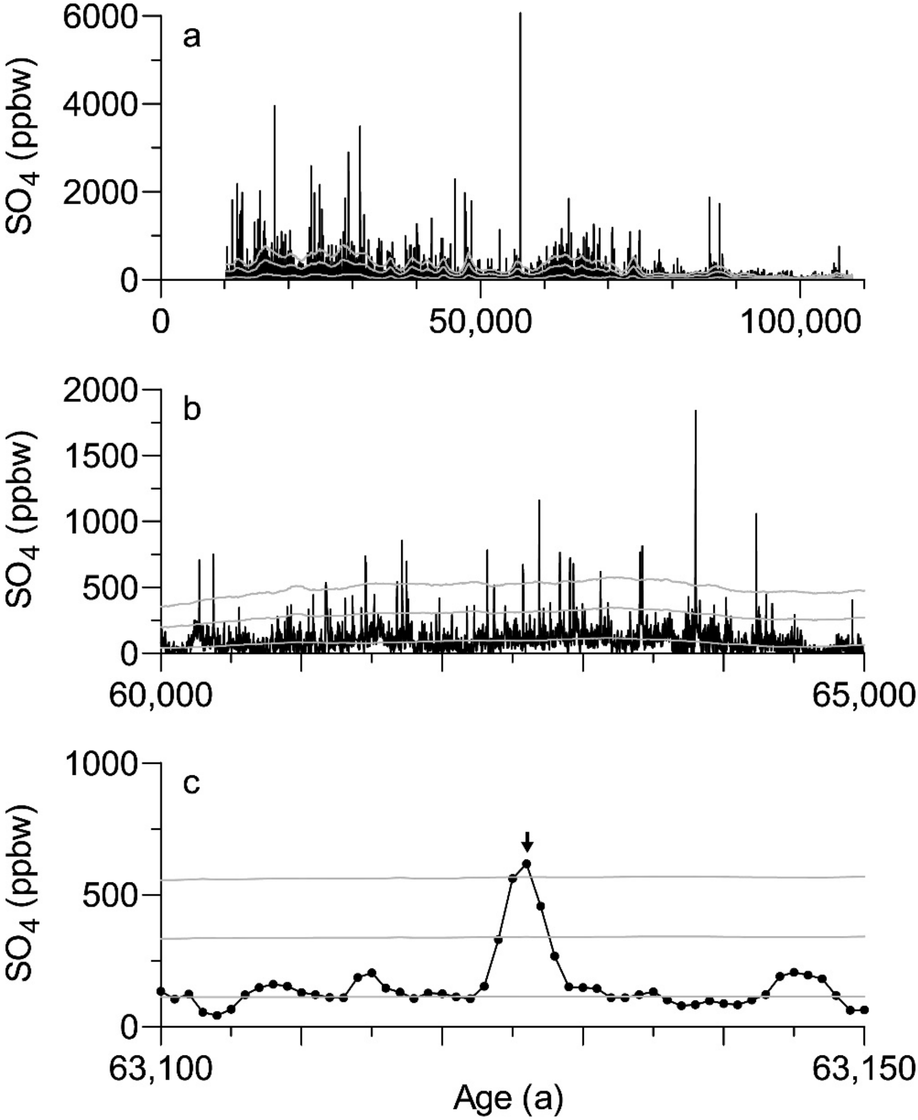


Figure 4.16. Extremes detection in the NGRIP sulfate record. **a** Full interval; **b, c** zoomed. The annual sulfate data are shown as *black lines* (additionally *dotted* in **c**). The background estimate (running median, $k = 750$) is in each panel the *lowest* of the three *grey lines*; the detection thresholds (running median plus z times running MAD, $k = 750$) are the *middle* ($z = 5.0$) and the *upper* ($z = 10.0$) *grey lines*. If a sulfate peak crosses a threshold for a few successive years (**c**), then only the maximum (*arrow*) is retained for further analysis ("declustering," Chapter 6).

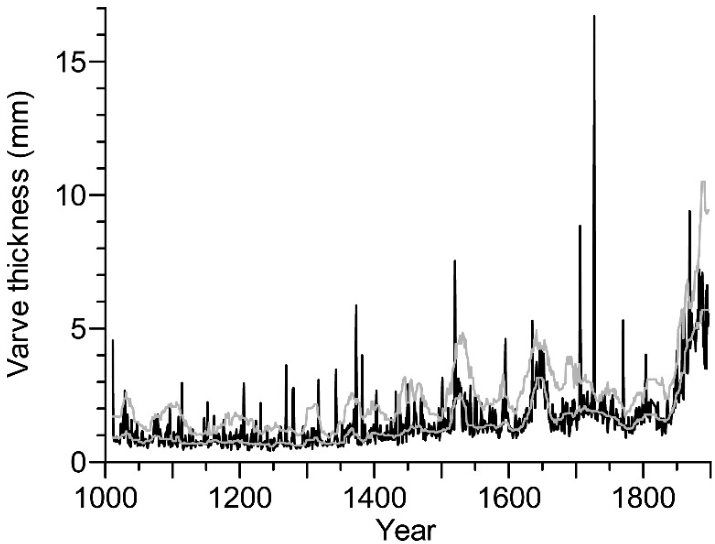


Figure 4.17. Extremes detection in the Lower Mystic Lake varve thickness record. Shown are annual varve thickness as *black line*, background estimate (running median, $k = 8$) as the *lower grey line* and the detection threshold (running median plus $z = 3.5/0.6745 \approx 5.2$ times running MAD, $k = 8$) as the *upper grey line*.

Bandwidth selection followed the median criterion; $k = 8$ means that background and variability variations on decadal timescales are "permitted." Several threshold selections were evaluated and the optimal compromise (liberal versus conservative) seen in $z = 3.5/0.6745 \approx 5.2$ (Besonen et al. 2008). This corresponds (Table 1.3, Note g) to 3.5 "robust standard deviations."

The number of detected peaks is 47. However, the second criterion imposed on a varve (graded-bed), besides thickness, led Besonen et al. (2008) to discard 11 of those events. The further analysis of the hurricane activity (Chapter 6) therefore was based on 36 events, observed between A.D. 1011 and 1897.

4.4 Background material

Textbooks on regression are numerous; accessible ones include the classic Draper and Smith (1981) as well as Sen and Srivastava (1990), Montgomery and Peck (1992), Kutner et al. (2005), Montgomery et al. (2006) and Graybill and Iyer (1994) on linear regression, Gallant (1987) and Seber and Wild (1989) on nonlinear regression, Bloomfield and Steiger (1983), Rousseeuw and Leroy (1987) and Lawrence and Arthur

(1990) on robust regression and, finally, Härdle (1990), Wand and Jones (1995), Simonoff (1996) and Wasserman (2006: Chapter 5 therein) on nonparametric regression. These books contain many ideas of relevance to the practice of climate time series regression, for example (1) the role of influential observations and leverage points, that means, points $(T(i), X(i))$ whose ex-/inclusion has a large influence on regression parameter estimates, (2) the usability of regressions for predicting time series or (3) the construction of confidence regions, that means, joint CIs for several parameters.

Model suitability is a further point covered in depth by the mentioned textbooks. We emphasize the importance of visual tools (Cook and Weisberg 1982): checking per eye how good a regression curve fits to the data or calculating the residuals ($e(i), r(i)$ and the white-noise residuals) and inspecting plots of them (e.g., $r(i)$ versus $t(i)$, or $r(i)$ versus $r(i-1)$) for how well they appear to be realizations of the assumed noise process. Such residual tests can further be performed numerically. A classical example is the test for AR(1) serial correlation in OLS regression by Durbin and Watson (1950, 1951, 1971), where the authors managed to analytically derive bounds for the null distribution of the test statistic $d_{\mathrm{DW}} = \sum_{i=2}^{n} [e(i) - e(i-1)]^2 / \sum_{i=1}^{n} e(i)^2$ (the case of straight-line subtraction is more difficult than mean subtraction). However, the Durbin–Watson test is applicable only to evenly spaced time series and therefore of minor relevance for climate time series regression. Often calculated is another parameter, the coefficient of determination, $R^2 = 1 - \sum_{i=1}^{n} e(i)^2 / \sum_{i=1}^{n} [x(i) - \bar{x}]^2$, with sample mean $\bar{x}$, which measures the proportion of variation "explained" by the regressor, $T(i)$. However, R^2 depends on $\{T(i)\}_{i=1}^{n}$ with, loosely speaking, higher spread of the time points leading to higher R^2 values (Montgomery and Peck 1992). Therefore, the coefficient of determination should be interpreted with caution in settings where the predictor points are not random variables (as in Chapter 8) but can rather be (partly) designed, for example by selecting the depth points where to take samples from a sedimentary or speleothem climate archive.

The **Gauss–Markov conditions** for linear regression are:

1. $E[X_{\mathrm{noise}}(i)] = 0, i = 1, \ldots, n$;
2. $S(i) = \mathrm{const.}, i = 1, \ldots, n$; and
3. $E[X_{\mathrm{noise}}(i) \cdot X_{\mathrm{noise}}(i-1)] = 0, i = 2, \ldots, n$.

If those conditions are fulfilled, then the OLS estimates have, among all the unbiased regression parameter estimates, the mimimum variance (Odell 1983). In climate time series analysis, the question is less whether the Gauss–Markov conditions are fulfilled than how severely they are violated.

Trimmed least-squares regression in the linear model (Ruppert and Carroll 1980) is a clearly better robust method (sensitivity to data configuration and convergence properties) than median of squares regression and today preferred (Davison AC 2009, personal communication).

Linear/nonlinear regression and bootstrap: theory. Review papers containing various details on bootstrap resampling applied to regression include the following: Efron and Tibshirani (1986), Wu (1986), Li and Maddala (1996) and Davison et al. (2003). A short expository note was written by Peters and Freedman (1984). Early, Freedman (1981) had clearly distinguished between the cases of a random regressor T, which we analyse in Chapter 8 (Regression II), and non-random T, which we consider in this chapter.

Linear regression and bootstrap: examples. Kahl et al. (1993) applied linear regression to measured records of seasonal temperature at the surface and several atmospheric heights in the Arctic Ocean region during 1950–1990; using ordinary bootstrap resampling these authors were unable to find significant upwards trends for any of the four seasons. Their ignorance of serial dependence does likely not invalidate the test result, although today presumably better data are available. Karl et al. (1995) examined linear trends in temperature and precipitation variabilities (from diurnal to interannual ranges) in records from globally distributed stations during parts of the twentieth century; using surrogate bootstrap resampling from AR(1) and ARMA models fitted to the residuals, they found, for example, that day-to-day temperature variability has decreased in the northern hemisphere. Witte et al. (1998) determined thermal gradients using proxy temperatures from beetle assemblages for the glacial–Holocene transition (Termination I) in northern Europe; the ordinary bootstrap was applied to construct percentile CIs. Kiktev et al. (2003) analysed trends in high-dimensional series of indices of daily climate extremes; the data are based on measurements and model simulations and cover the interval 1950–1995. The authors used MBB resampling for hypothesis testing in an adaption to high dimensionality (Wilks 1997). MBB resampling was applied in a linear regression analysis of Canadian low-flow runoff series covering the interval from 1954 to 2003 (Khaliq et al. 2008).

Ramp regression was elaborated by Mudelsee (2000); an early figure of the ramp as a model of a climate transition was shown by Hare (1979: Fig. 1B therein). The ramp was applied to quantify the Mid-Pleistocene Climate Transition (Mudelsee and Schulz 1997), which meant an increase in mean global ice mass of 0.29‰ (δ^{18}O-equivalent) from $\sim$ 942 to $\sim$ 902 ka; this increase initiated the late Pleistocene ice age. The ramp was also applied to quantify the onset of the Indian monsoon in the early Holocene (Fig. 4.18). Recent applications of the ramp for estimating paleoclimatic change-points include the following. Hopley et al. (2007) quantified a rapid increase in savannah grass proportions between 1.78 and 1.69 Ma ago that coincides with a pulse in African mammal turnover (proxy: δ^{13}C in a speleothem). Steffensen et al. (2008) showed that climate changes in Greenland during Termination I may have happened within a few years (δD in NGRIP ice core). Tachikawa et al. (2009) determined the onset of increases in Pacific sea-surface temperature during Termination I and analysed the spatial distribution of the change-point times (several proxy variables measured on 30 sediment cores). Wolff et al. (2009) compared Terminations I, II and V in terms of warming rates (δD in EPICA Dome C ice core). Fleitmann et al. (2009) quantified the onsets of Dansgaard–Oeschger events 1 and 3–12, between about 15 and 48 ka ago (δ^{18}O in a stalagmite from northern Turkey). The ramp could be further used for quantifying, for example, the duration of geomagnetic polarity reversals (Clement 2004) with CI.

Carbon isotopic compositions are written in the delta notation (PDB standard) as $\delta^{13}\mathrm{C} = [(^{13}\mathrm{C}/^{12}\mathrm{C})_{\mathrm{sample}}/(^{13}\mathrm{C}/^{12}\mathrm{C})_{\mathrm{PDB}} - 1] \cdot 1000‰$.

Break regression, in statistical science better known as "two-phase regression" (Hinkley 1970, 1971), was applied (Solow 1987) to southern hemisphere temperature, 1858–1985. Reinsel et al. (2002) studied GLS estimation of the break model, however, under the unrealistic assumption that the change-point in time, $t2$, is known. This method was then applied (Reinsel 2002; Reinsel et al. 2005) to detect trend changes in stratospheric ozone concentrations, 1977–2002, that means in particular, the effects of the Montreal Protocol on Substances that Deplete the Ozone Layer from 1987 and its Amendments. Hinkley (1988) mentioned and Julious (2001) studied bootstrap resampling for the two-phase regression; the latter paper devised a hypothesis test for the existence of an unknown change-point. It appears, however, that CIs for break model parameter estimators (Section 4.2.2.3; Mudelsee 2009) have not been previously studied. A recent application (Trauth et al. 2009) of break regression with bootstrap CIs examined trend changes in African aridity during the Plio-/Pleistocene (proxy: dust flux in a marine sediment core off the coast of west Africa). Tomé and Miranda (2004, 2005) presented an algorithm for fitting a continuous regression model with several break points to data and applied this method to study changes of Azores temperature, northern hemisphere temperature, the NAO index and Lisbon winter precipitation. Unfortunately, no error bars or confidence intervals for the estimated trend parameters were determined.

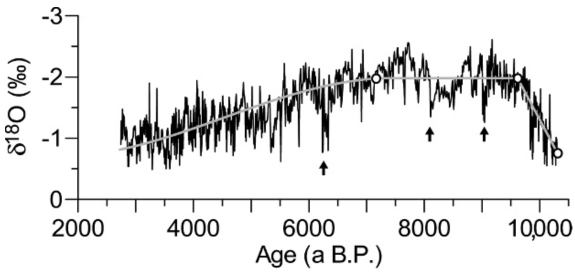


Figure 4.18. Trend estimation for the δ^{18}O record from stalagmite Q5. A combination of a portion of a ramp in the early part and a sinusoid in the late part was fitted (OLS) as trend (*grey line*) to the time series (*black line*). Choice of the sinusoid is motivated by the observation that Holocene changes in local solar insolation, induced by Earth orbital changes, influenced monsoonal rainfall amounts (Fleitmann et al. 2003). The change-point estimates (*open circles*) with bootstrap standard errors (SB resampling, $B = 2000$, time errors from statistical and dating (Fleitmann et al. 2003) uncertainties via Gaussian error propagation) are (10,300 a (fixed), -0.77 ± 0.08‰), (9617 ± 89 a, -1.98 ± 0.03‰) and (7200 ± 400 a, -1.98‰). The change at 9617 a occurred, within error bars, simultaneously with a similar change in northern temperatures, as indicated by δ^{18}O variations in the GRIP ice core (Greenland), indicating a potential influence of northern glacial boundary conditions on monsoon climate (Fleitmann et al. 2003). The sinusoid was fitted by linear OLS regression (e.g., replace $\beta_1 T(i)$ by $\beta_1 \sin(T(i))$). Climate extremes detection (Section 4.3.3) found three pronounced dry extremes (*arrows*): a longer-lasting event at $\sim$ 8.2 ka and other at $\sim$ 9.2 and $\sim$ 6.3 ka.

Other change-point estimation methods for time-dependent mean and variance exist. A collection of early Bayesian papers is Smith (1975), Cobb (1978), Menzefricke (1981), Booth and Smith (1982) and Abraham and Wei (1984). “Techniques for testing the constancy of regression relationships over time“ is the title of a paper by Brown et al. (1975), which presents “real-time” or “online” tests. Similarly did Yashchin (1995) for “real-time,” abrupt changes in $X_{\text{trend}}(T)$. The history of such tests goes back to the 1950s, when Page (1954) introduced the cumulative sum (CUSUM) chart. A CUSUM chart shows typically

$$S(r) = \sum_{i=1}^{r} \left[x(i) - \mu_{\text{target}} \right], \tag{4.60}$$

where μ_{target} is the target mean value and $S(0) \equiv 0$, plotted against $r = 0, 1, 2, \ldots$. Deviations of the “real-time” from the target mean are assessed with control limits (Barnard 1959; Goel 1982). Setting the control limits is done by taking into account the properties (shape, persistence) of the data generating process (Chen and Gupta 2000; Wu 2005). However, such tests are less useful in the context of this book, which assumes that "offline" time series are available for retrospective analysis. Tests for variance changes were presented by Hsu (1977), Tsay (1988), Inclán and Tiao (1994) and others. Popular in climatology (Karl and Riebsame 1984; Karl and Williams 1987; Yamamoto et al. 1986; Goossens and Berger 1986; Maasch 1988; Gluhovsky and Agee 1994; Lund et al. 2007) have been retrospective tests for abrupt changes in the mean, that is, points where $X_{\text{trend}}(T)$ changes from one constant level to another constant level. Those tests, performed either non-robustly on basis of the t distribution of the mean (Section 3.1.1) or robustly based on ranks (Kendall 1938; Mann 1945), may be useful when the objective is to detect inhomogeneities in the data (Section 1.6), as has been the case in many of the climatology studies cited. Because a jump in the mean is not a gradual change, however, such models are rarely useful for quantifying climatic trends. Pettitt (1979) presented a test for a "jump in the distribution function" in a sequence of random variables. Esterby and El-Shaarawi (1981) developed a maximum likelihood estimator for the change-point in a two-phase polynomial regression model with AR(1) noise component. Rodionov (2004) presented a CUSUM-like test for a jump in the mean based on the t distribution and augmented (Rodionov 2006) the test by means of prewhitening to take persistence into account. Caussinus and Mestre (2004) used a penalized log-likelihood procedure for detecting an unknown number of jumps in the mean and, notably, outliers in a time series. The latter authors performed a Monte Carlo experiment to study the test power in dependence on the size of the jumps. Also discontinuous linear models, developed in econometrics (Bai and Perron 1998), have limited applicability in climatology. Perron (2006) reviews estimation methods from the viewpoint of econometrics, where the change-points are denoted as "structural breaks." He gives results about limit distributions ($n \to \infty$) of estimators and further insight into the existing voluminous work on this topic. We also mention Transitional Generalized Linear Models, an interesting generalization of nonlinear regression, expressed in terms of conditional means and variances, which has been applied with MBB resampling to a pollen time series (Brumback et al. 2000). To summarize the paragraph, we think that parametric trend models are important for climatology. Examples include the ramp, the break, the trapezoidal model (Schulz 2002) and the piecewise linear model (Seidel and Lanzante 2004). We further think that, from the trend models discussed (Fig. 4.19), the continuous types are more realistic. One should also consider unspecified nonlinear models, which can be estimated nonparametrically (see a later paragraph).

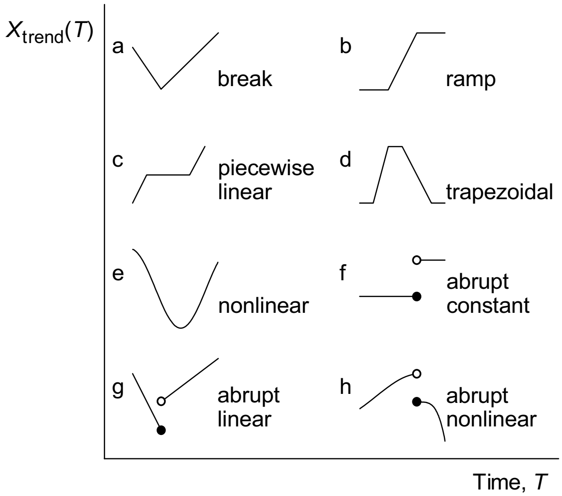


Figure 4.19. Regression models for trend estimation. *Open/filled symbols* mark discontinuities (abrupt changes).

Structural changes versus long memory: this dichotomy between trend and noise components in a stochastic process is known to several scientific disciplines. Econometrics accepts that separating models with jumps in $X_{\text{trend}}(T)$ from long-memory models is difficult (Diebold and Inoue 2001; Granger and Hyung 2004; Perron 2006). Figure 4.20 illustrates this difficulty: the trend function is such complex that one may be inclined to prefer a description as noise. Knowledge about the dichotomy exists also in hydrology (Koutsoyiannis 2006) and climatology (Rust et al. 2008). Complexity and dichotomy may lead to pitfalls in the form of fitting physically implausible models to observed climate time series. Gil-Alana (2008) analysed annual temperature (global, northern hemisphere, southern hemisphere) for the interval 1861–2002 and fitted linear regression models with ARFIMA$(0, \delta, 0)$ noise component. He also considered structural changes (two and three regimes), finding significant trends in all cases. Unfortunately, Gil-Alana (2008) did not present physical explanations of the long memory and of the change-point times (e.g., 1871 and 1974 for global temperature). We have already noted (Section 2.5.2) that aggregated series may produce spurious long memory. In another paper, Wu and Zhao (2007: p. 403 therein) were "pleased to conclude that there is no evidence for jumps in the mean trend" for monthly temperature (global), interval 1856–2000. Detected jumps (or their absence) and confirmed (or refuted) long memory should lead researchers to explain such findings. Rust et al. (2008)

give a remarkably simple explanation for detected jumps or long memory: inhomogeneities in the observed temperature records. A recently detected inhomogeneity is the jump in mean at 1945 in hemispheric temperature records. This jump is the apparent result of uncorrected instrumental biases in the sea-surface temperature measurements (Thompson et al. 2008).

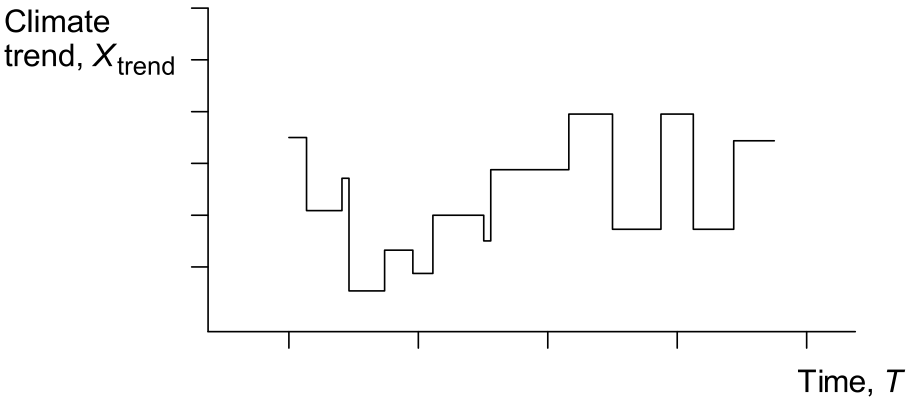


Figure 4.20. Climate trend function comprising many jumps.

The **Mann–Kendall trend test** (Mann 1945; Kendall 1938) is based on the idea of sorting. Consider a sample $\{X(i)\}_{i=1}^{n}$ in ascending order with time, $X(1) < X(2) < \cdots < X(n)$. This may be associated with a monotonically increasing trend function. A sample in non-perfect order requires a minimum, U', of interchanges (e.g., $X(1) \leftrightarrow X(2)$) to reach an ordered state. The theoretical maximum of U' is given by the number of pairs of data points, $n(n-1)/2$. The test statistic (Kendall's tau),

$$U = 1 - \frac{4U'}{n(n-1)}, \tag{4.61}$$

is theoretically between 1 (upwards trend) and -1 (downwards trend). Under an IID random process, $X(i)$, the distribution of U reaches, with increasing n, rapidly a standard normal form (Kendall 1938; Kendall and Gibbons 1990) with $E[U] = 0$ and $VAR[U] = 2(2n+5)/[9n(n-1)]$. (There are adaption formulas for the case of ties.) The advantage of the Mann–Kendall test within the context of climate time series analysis is its robustness regarding the distributional shape of $X(i)$. The major task is to find suitable adaptions to deal with serial dependence in the process. von Storch and Zwiers (1999: Section 6.6 therein) present such, and they cover also other tests of the mean. One may apply the

Cochrane–Orcutt transformation prior to the Mann–Kendall test to remove autocorrelation effects ("prewhitening"). Hamed (2009b) notes that ignoring bias in autocorrelation estimation leads to a reduced test power of this procedure. Hamed (2008) derives adaption formulas for the variance for the case $X(i)$ is a long-memory process. See Stuart (1983) for a review of Kendall's tau, which includes also references to earlier papers than that of its originator.

The **superposed epoch analysis** originates from the idea of Chree (1913, 1914) to compare the means of one variable, $X(i)$, taken from time intervals before and after the occurrence of events. The variable is assumed to be a continuous random variable, and the events are assumed to present the outlier or extreme component of a second variable, $Y_{\text{out}}(i)$. That type of analysis belongs therefore also to the bivariate setting (Part III). By testing the hypothesis of equal means before and after, it is possible to study the relation between climate variables and climate extremes. Adams et al. (2003) related the ENSO index with the occurrence of explosive volcanism in low latitudes over the past approximately 350 years. Using MBB resampling, with the selected block length equal to $2\widehat{\tau}$ (persistence time of an AR(1) process fitted to the ENSO time series), they found significant changes in the mean, that is, a multi-year El Niño-like response to volcanic forcing.

Quantile regression models (Koenker and Bassett 1978; Koenker and Hallock 2001; Yu et al. 2003) do not describe the mean conditional on time, $X_{\text{trend}}(T)$, but a quantile of the distribution of X conditional on time. Estimation can be achieved by employing a sum of asymmetrically weighted absolute residuals (instead of a least-squares sum as in WLS). Censored quantile regression (Powell 1986) is the adaption to the case when the range of values of X is restricted or not observable. This may apply to climate observations. Quantile regression can be performed to estimate robustly the time-dependent centre of location (50th percentile or median). Another application field is extreme value time series analysis (Chapter 6).

Neural networks can be viewed as complex regression models between input and output, where the numbers of regression terms and parameters are not fixed but allowed to vary. Such models fit therefore to this book's preference for putting complexity more into the trend and less into the noise component. Examples of neural networks applied to climate time series are the following: Grieger and Latif (1994) analysed ENSO dynamics and, related, Hsieh and Tang (1998) studied the prediction of Pacific sea-surface temperatures for the second half of the twentieth century. Studies on employing bootstrap resampling in neural network estimation include those by Breiman (1996) and Franke and Neumann (2000). Applications of the combined bootstrap–neural network technology to climatological problems exist few yet. Guiot and Tessier (1997) detected pollution signals in tree-ring series with bootstrap percentile confidence intervals. Giordano et al. (2005) presented, in a conference paper, the prediction of hydrological time series under usage of the sieve bootstrap, laying down their method in a later paper (Giordano et al. 2007).

Errors-in-variables models are covered in the advanced of the textbooks mentioned in a preceding paragraph, also by Jones (1979) and Fuller (1999) and here in Chapter 8.

Smoothing methods have since long been used to visualize deterministic data features; they range from intuitive drawing per hand with the "cosmic schwung" (Suess and Linick 1990) to more elaborated mathematical attempts (Härdle and Chen 1995; Simonoff 1996). An important class is formed by linear smoothers, which relate (process level) the data vector $\mathbf{X}$ (Eq. 4.11) and the estimated trend vector $\widehat{\mathbf{X}}_{\mathbf{trend}} = [X_{\text{trend}}(1), \ldots, X_{\text{trend}}(n)]'$ as $\widehat{\mathbf{X}}_{\mathbf{trend}} = \mathbf{SX}$, where the $n \times n$ matrix $\mathbf{S}$ is called a smoother matrix. Buja et al. (1989) review linear smoothers such as the running mean, running linear OLS regression, running polynomial regression and kernel smoothing. They study also a technique fashionable in climatology, namely cubic spline smoothing, which minimizes the expression

$$\sum_{i=1}^{n} \left[x(i) - x_{\text{trend}}(i)\right]^2 + \lambda \int_{-\infty}^{+\infty} \left[\ddot{x}_{\text{trend}}(t)\right]^2 dt, \tag{4.62}$$

where $\ddot{x}_{\text{trend}}(t)$ is the second derivative (curvature) of the trend function in continuous time. Herein, λ is the smoothing parameter; $\lambda = 0$ leads to interpolation and $\lambda \to +\infty$ to OLS regression.

Bandwidth selection methods for nonparametric regression can be divided into few classes, as illustrated by Gasser et al. (1991) for the Gasser–Müller kernel: the first class, cross-validation, is based on delete-one estimates (Section 4.3.3), the second, penalizing, adds a bias term to the sum of squares before minimization. Hall et al. (1995b) consider a third class, based on MBB resampling, which they use besides cross-validation in their theoretical description of how short- and long-memory serial dependence affects optimal bandwidth selection. Previously, Diggle and Hutchinson (1989) studied bandwidth selection for Gaussian AR(1) dependence. Grunwald and Hyndman (1998) considered penalized bandwidth selection under non-Gaussian errors and showed as example smoothing of daily rainfall in Melbourne, 1980–1989, with bootstrap confidence band. Francisco-Fernández et al. (2004) and Francisco-

Fernández and Vilar-Fernández (2005) presented bandwidth selectors for local polynomial regression smoothing with AR(1) and other types of serial dependence. Gijbels and Goderniaux (2004a) studied bandwidth selection with the bootstrap in the context of change-point estimation in nonparametric regression. They applied this to detect changes in the record of annual temperatures from Prague during 1775–1989 (Gijbels and Goderniaux 2004b) and the record of annual runoff from the Nile during 1871–1934 (Gijbels et al. 2004). A caveat that may be raised is that the assumed discontinuous trend model (Fig. 4.19h) is climatologically unrealistic. Previous papers on change-point estimation using nonparametric regression include Müller (1992) and Chu (1994).

Adaptive nonparametric regression is a further smoothing development, where the bandwidth, h, is allowed to be time-dependent. For example, a smaller h can be used in regions where the spacing $d(i)$ is smaller or where the variability $S(i)$ is smaller, enabling detection of systematic, local, short-term trends. Local bandwidth selection methods for the Gasser–Müller kernel regression were developed by Brockmann et al. (1993) and Herrmann (1997). A unifying approach to nonparametric regression (smoothing spline, k-nearest-neighbour, kernel) was presented by Jennen-Steinmetz and Gasser (1988).

Bootstrap confidence band construction for nonparametric regression was introduced by Härdle and Bowman (1988) and further developed to include topics such as simultaneous confidence bands (Härdle and Marron 1991) or pivotal methods (Hall 1992). An early paper on bootstrap confidence bands, together with an application to a Holocene radiocarbon record from tree-rings, was presented by Hall and Titterington (1988). The ozone time series from Arosa, Switzerland, covering the interval from the 1930s to the 1990s, was analysed by Bühlmann (1998) using an MBB confidence band for the nonparametric trend estimation and by Bühlmann (2002) using nonparametric regression and MBB tests of the hypotheses "constant mean" and "constant variability."

Detection of climate extremes has to be performed robustly because the assumed extremes should not bias estimates of trend and variability. Although Lanzante (1996) warned climatologists of this pitfall, it seems today more the rule than the exception that non-robust methods, such as running mean and running standard deviation, are employed. Mudelsee (2006) reiterated the warning using as an example a Holocene section of the sulfate record from the GISP2 ice core (Greenland). Lanzante (1996) reviewed robust techniques also for change-point estimation and presented climatological examples. The detection method from Section 4.3.3 was applied also by Fleitmann et al. (2008) to detect a climate cold anomaly at around 9.2 ka before present in paleoclimate proxy records from globally distributed sites, and by Girardin et al. (2009) to detect wildfire events in proxy records of July monthly drought code from 28 forested ecoregions of the North American and Eurasian continents, interval 1901–2002.

Timescale errors and their inclusion into the bootstrap resampling procedure depend on how the timescale for a climate archive has been constructed. (1) Timescales based on dated depth points and a regression model for the age–depth curve, constrained to strict monotonic growth (Figs. 4.3 and 4.4), can be readily used for parametrically resampling the time points by means of the dating errors (timescale-ARB and timescale-MBB algorithms). If the size of the dating error, S_T, is unknown, it may be estimated via formulas analogous to Eq. (4.8). If it is known a priori (machine error), such formulas may be used for calculating a dating error estimate and comparing it with the machine error. Agreement would then corroborate the validity of the estimated age–depth curve. Papers on this approach include Bennett (1994) on the chronology of a lake sediment core dated with ^{14}C and Spötl et al. (2006) on a stalagmite dated with U/Th. Drysdale et al. (2004) use a similar approach, also on a stalagmite, but make an additional, apparently ad-hoc assumption that stalagmite growth may vary by a factor of ten between the dating points, resulting in a wider confidence band. Bennett and Fuller (2002) study the influence of the age-model selection on the estimated date of the mid-Holocene decline of the hemlock tree in eastern North America. In the presence of hiatuses (Fig. 4.4), it should be worth applying regression models with a jump in the mean (Fig. 4.19g, h), additionally constrained to monotonic growth, to the construction of age–depth curves with confidence bands. Heegaard et al. (2005) offer an interesting extension to the case where the material in an archive at a common depth is age-inhomogeneous. This can occur in sediment cores as a result of mixing processes. (2) Timescales for laminated archives follow directly from detected lamina depths and the time period of lamina deposition. In applications, this period is almost exclusively 1 year. Examples are varved lake sediments resulting from absent bioturbation caused by low oxygen concentrations, yearly δ^{18}O cycles in ice cores, thickness and density cycles of tree-rings, growth layers in speleothems and also growth layers in "biological archives" such as corals or mollusks. The error in "absolute time," dependent on the dating error of an absolutely dated lamina, can be clearly larger than the "internal" time uncertainty. (3) Tuned timescales based on relating a record $\{t_X(i), x(i)\}_{i=1}^{n_X}$ to another, dated record $\{t_Y(i), y(i)\}_{i=1}^{n_Y}$ (see background material in Chapter 1) may depend strongly on the assumed relationship. Tuning has been often applied to erect chronologies of marine sedimentary records from the Pleistocene, with the target $y(i)$ being a time series of solar insolation or Earth orbital elements. The points $\{t_X(i)\}_{i=1}^{n_X}$ are fitted such that a cost function is minimized. Many choices can influence the design of the cost function: curvature, robustness, bounds of the inferred sedimentation rate, bounds of the maximum time shift of a time point, etc. Papers include Martinson et al. (1982), Herterich and Sarnthein (1984), Martinson et al. (1987), Brüggemann (1992), Grieger (1992), Lisiecki and Lisiecki (2002), Huybers (2002) and Huybers and Wunsch (2004). Timescale error determination may in principle be carried out using regression methods for bivariate time series (Chapter 8), but those results could be misleading when the assumptions made (cost function) are uncertain. In such cases of model uncertainty, an option would be to adopt (4) Bayesian tools for constructing chronologies (Buck and Millard 2004). A new, continuous, piecewise linear, monotone stochastic process has been suggested (Haslett and Parnell 2008) to model accumulation of a climate archive. These authors present also applications to radiocarbon-dated lake sedimentary records.

4.5 Technical issues

The **GLS standard notation** of the sum of squares to be minimized (Eq. 4.9) is with $\mathbf{V}/S^2$ instead of $\mathbf{V}$, where S is a constant "overall" standard deviation of $X_{\text{noise}}(i)$.

Matrix algebra is required for GLS estimation (Section 4.1.2). Let $\mathbf{A}$ be a $p \times q$ matrix with elements $A(i,j)$, where $i = 1, \ldots, p$ denotes the row and $j = 1, \ldots, q$ the column. For example, in Eq. (4.12), $p = n$ (data size) and $q = 2$. Let $\mathbf{B}$ be a $q \times r$ matrix. The matrix product is then a $p \times r$ matrix $\mathbf{C}$ with elements $C(i,j) = \sum_{k=1}^{q} A(i,k) \cdot B(k,j)$. A vector is a matrix with $q = 1$. The sum $\mathbf{C}$ of two $p \times q$ matrices $\mathbf{A}$ and $\mathbf{B}$ has the elements $C(i,j) = A(i,j) + B(i,j)$. The transpose $\mathbf{A}'$ of a $p \times q$ matrix $\mathbf{A}$ is given by the $q \times p$ matrix $\mathbf{C} = \mathbf{A}'$ with $C(i,j) = A(j,i)$. The matrix $\mathbf{I}$ is called unit matrix if it is a $p \times p$ matrix with $I(i,j) = 1$ for $i = j$ and 0 for $i \neq j$. The inverse $\mathbf{A}^{-1}$ of a matrix $\mathbf{A}$ has the property $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. The rule "multiplication before addition" applies also to matrices. The product of more than two matrices is calculated "from the left," $\mathbf{D} = \mathbf{ABC} = (\mathbf{AB})\mathbf{C}$ (matrix multiplication is not commutative). See Dahlquist and Björck (2008: Appendix A therein) for more details on matrix algebra.

Numerical linear algebra methods for solving OLS, WLS and GLS regression problems are explained by Gentle (1998) and Dahlquist and Björck (in press: Chapter 8 therein). LAPACK (Anderson et al.

1999) is a collection of Fortran (77 and 90) routines, available from `http://www.netlib.org/lapack` (19 December 2006). Those minimization methods often work with the gradient technique.

Search techniques have to be applied to problems where the least-squares sum is not differentiable with respect to a parameter. An example is Brent's search (Brent 1973; Press et al. 1992), which is based on interpolation and iterated interval divisions in the golden section ratio. As regards least-squares fitting of the piecewise linear regression functions (Fig. 4.19a–c), Williams (1970) extended Hudson's (1966) brute-force search for two pieces ("break") to three pieces ("ramp"); see also Schulze (1987) on "multi-phase" regression. The size of the temporal spacing at around an estimated change-point in time limits the accuracy of the change-point estimate. In practice, however, this is only a problem when $d(i)$ becomes larger than the size of the bootstrap standard error. The brute-force search with bootstrap resampling is applicable even to long climate time series (n above, say, 100,000) owing to two computational modifications (see below) and today's available computing power.

Parallel computing allows the acceleration of extensive numerical estimations on multiprocessor machines. As an example, the Monte Carlo experiment on the coverage performance of BCa CIs for ramp regression (Table 4.13) was carried out by means of a Fortran 90 program on a four-processor workstation. Constants like $k_5 = \sum_{i=\widetilde{i1}+1}^{\widetilde{i2}-1} t(i)$ (cf. Eq. (4.43) with $S(i) = 1$) were declared as two-dimensional arrays, $k_5(\widetilde{i1} = 1 : n, \widetilde{i2} = 1 : n)$, calculated as (Ellis et al. 1994; Press et al. 1996)

$$\begin{aligned}
\mathtt{b} &= (/\ \mathtt{0.0},\ \mathtt{0.0},\ (\mathtt{sum}(\mathtt{t}(2:\mathtt{j})),\ \mathtt{j} = 2, \mathtt{n}-1)\ /) \\
\mathtt{c} &= (/\ \mathtt{0.0},\ (\mathtt{sum}(\mathtt{t}(2:\mathtt{j})),\ \mathtt{j} = 2, \mathtt{n})\ /) \\
\mathtt{k5} &= \mathtt{spread}(\mathtt{b},\ \mathtt{dim} = 1,\ \mathtt{ncopies} = \mathtt{n}) - \\
&\quad\ \mathtt{spread}(\mathtt{c},\ \mathtt{dim} = 2,\ \mathtt{ncopies} = \mathtt{n})
\end{aligned} \tag{4.63}$$

and the estimation equations (4.41) and (4.42) were solved with whole-array operations.

RAMPFIT (Mudelsee 2000) is a FORTRAN 77 software for WLS fitting of a ramp (Eq. 4.39) to time series; it includes SB resampling, bootstrap standard error calculation and graphical residual analysis. Two modifications reduce computing costs of the brute-force estimation of the ramp change-points, $t1$ and $t2$. First, when searching through the $t1$–$t2$ grid, the constants (Eqs. 4.41 and 4.42) are not calculated new but their values are updated. Second, the candidate change-points $\widetilde{i1}$ and $\widetilde{i2}$ are not selected from the whole set $\{1, \ldots, n\}$ but only a subset.

Smoothing is included in RAMPFIT as a visual tool for tailoring these search ranges. By setting $\widetilde{i1} = 1$ and $\widetilde{i2} = n$, RAMPFIT can be used as a linear WLS regression tool. The software is available at the web site for this book.

Breakfit (Mudelsee 2009) is a FORTRAN 77 software for WLS fitting of a break model (Eq. 4.47) to time series; it includes MBB resampling, bootstrap standard error calculation and graphical residual analysis. The software is available at the web site for this book.

SiZer is an explorative, graphical software tool to assess the significance of zero crossings of derivatives (i.e., change-points) by means of nonparametric regression (Chaudhuri and Marron 1999). It is available via `http://www.unc.edu/~marron/DataAnalyses/SiZer_Intro.html` (8 December 2009) as Java and Matlab implementations.

The **strucchange** package for R (Zeileis et al. 2002) supports testing, monitoring and estimating structural changes in linear regression models by means of CUSUM charts and other tools. It is available at `http://cran.r-project.org/web/packages/strucchange` (9 December 2009).

Other **nonlinear regression software** with bootstrap resampling includes the following: Huet et al. (2004) have a package for S-Plus; Sherman et al. (1998) mention an S-Plus code for general regression models with MBB resampling, which can be obtained from `sherman@stat.tamu.edu` (23 January 2007).

Gasser–Müller adaptive kernel nonparametric regression can be implemented using the FORTRAN 77 subroutines and interfaces to Matlab and S-Plus, available from the following internet address: `http://www.biostat.unizh.ch/research/software/kernel.html` (3 May 2007).

Optimal median smoothing is the title of a paper by Härdle and Steiger (1995); this expression means that no faster algorithm for running median calculation has yet been found than that presented in the paper. The ideas behind the algorithm are the so-called double-heap ordering structure and updating.

CLIM-X-DETECT is a Fortran 90 program (Mudelsee 2006) for detecting extremes in time series against a time-dependent background (Section 4.3.3). The software adapts Härdle and Steiger's (1995) algorithm to calculate also the delete-one background estimate and the running MAD for setting the detection threshold. CLIM-X-DETECT is available at the web site for this book.

agedepth_1.0.zip is an archive of R functions for constructing age–depth curves using the regression approach of Heegaard et al. (2005). It is available at `http://www.eecrg.uib.no/Homepages/EinarHeegaard.htm` (25 May 2010).

WinGeol Lamination Tool is a C++ software under Windows for automatic laminae detection using digital image analysis (Meyer et al. 2006). It is available from `http://www.terramath.com` (17 January 2007).

Autocomp/Match are two C++ packages for Macintosh, Unix or Windows implementing the correlation approach of Lisiecki and Lisiecki (2002) to timescale construction. `http://www.lorraine-lisiecki.com` is the site where it can be obtained (17 January 2007).

XCM is a timescale tuning software for Matlab implementing the cross-correlation maximization algorithm of Huybers (2002). It is available from `http://web.mit.edu/~phuybers/www/Mfiles/Toolbox` (25 May 2010).

BCal is an online Bayesian tool at `http://bcal.sheffield.ac.uk` (12 February 2010) for constructing radiocarbon timescales.

Likewise, **OxCal** is an online/offline Bayesian software (Ramsey 2008), available at `http://c14.arch.ox.ac.uk/oxcal.html` (9 December 2009).

Isoplot is a geochronological toolkit for Excel (Ludwig 2003) implementing a Bayesian approach to timescale construction; it is available via `http://www.bgc.org/isoplot_etc/software.html` (17 January 2007).

Chapter 5

Spectral Analysis

Spectral analysis investigates the noise component in the climate equation (Eq. 1.2). A Fourier transformation into the frequency domain makes it possible to separate short-term from long-term variations and to distinguish between cyclical forcing mechanisms of the climate system and broad-band resonances. Spectral analysis allows to learn about the climate physics.

The task is to estimate the spectral density function, and to test for harmonic (cyclical) signals. This poses more difficulties than, for example, linear regression because now we estimate a function and not just two parameters. Spectral smoothing becomes therefore necessary, and this brings a trade-off between estimation variance and frequency resolution.

The multitaper smoothing method achieves the optimal trade-off for evenly spaced time series. The method of choice for unevenly spaced records is Lomb–Scargle, which estimates in the time domain and avoids distortions caused by interpolation.

Bootstrap resampling enhances multitaper and Lomb–Scargle methods by providing a bias correction and CIs. It supplies also a detection test for a spectral peak against realistic noise alternatives in form of an AR(1) process ("red noise"). Section 5.2.8 introduces bootstrap adaptions to take into account the effects of timescale uncertainties on detectability and frequency resolution.

5.1 Spectrum

Let us assume in this chapter that the climate process in continuous time, $X(T)$, has no trend and no outlier components and a constant

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_5,

variability, S,

$$X(T) = X_{\text{trend}}(T) + X_{\text{out}}(T) + S(T) \cdot X_{\text{noise}}(T) \\ = S \cdot X_{\text{noise}}(T). \tag{5.1}$$

Such a process could be derived from a "real" climate process, that is, with trend and so forth, by subtracting the trend and outlier components and normalizing (standard deviation). Techniques for quantifying trend and variability and detecting outliers are presented in Chapter 4.

It is then straightforward (Priestley 1981) to define a truncated process,

$$X_{T'}(T) = \begin{cases} X(T) & \text{for } -T' \le T \le T', \\ 0 & \text{elsewhere,} \end{cases} \tag{5.2}$$

and express it as a Fourier integral,

$$X_{T'}(T) = (2\pi)^{1/2} \int_{-\infty}^{\infty} G_{T'}(f) e^{2\pi i f T} df, \tag{5.3}$$

where

$$G_{T'}(f) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} X_{T'}(T) e^{-2\pi i f T} dT \\ = (2\pi)^{-1/2} \int_{-T'}^{T'} X(T) e^{-2\pi i f T} dT. \tag{5.4}$$

This introduces the frequency, f. (The symbol i in the exponent denotes $\sqrt{-1}$.) This is a useful quantity for describing phenomena that exhibit a periodic behaviour in time. The period (time units) is given by $T_{\text{period}} = 1/f$. If one associates $X(T)$ with movement and kinetic energy, then $2\pi |G_{T'}(f)^2| df$ can be seen as the energy contribution of components with frequencies within the (arbitrarily small) interval $[f; f + df]$. Regarding the truncation, because with $T' \to \infty$ also the energy goes to infinity, one defines the power, $\pi |G_{T'}(f)^2| / T'$. Because the previous formulas in this section apply to a time series rather than a stochastic process, one uses the expectation operator to define

$$h(f) = \lim_{T' \to \infty} \left\{ E \left[2\pi \left| G_{T'}(f)^2 \right| / T' \right] \right\}. \tag{5.5}$$

The function $h(f)$ is called one-sided non-normalized power spectral density function of the process $X(T)$, often denoted just as (non-normalized) spectrum. It is the average (over all realizations) of the contribution to the total power from components in $X(T)$ with frequencies within the interval $[f; f + df]$. $h(f)$ is defined for $f \geq 0$ and integrates to S^2. A closely related function is

$$g(f) = h(f) / S^2, \tag{5.6}$$

the one-sided normalized power spectral density function, which integrates to unity. A two-sided version of the spectrum, symmetric about $f = 0$, is also used (Bendat and Piersol 1986).

The functions $h(f)$ and $g(f)$ are the Fourier transforms of the autocovariance and autocorrelation functions, $R(\tau)$ and $\rho(\tau)$, respectively, provided they exist (Priestley 1981: Section 4.8 therein):

$$h(f) = \pi^{-1} \int_{-\infty}^{\infty} R(\tau) e^{-2\pi i f \tau} d\tau, \tag{5.7}$$

$$g(f) = \pi^{-1} \int_{-\infty}^{\infty} \rho(\tau) e^{-2\pi i f \tau} d\tau. \tag{5.8}$$

Herein,

$$R(\tau) = E\left[X(T) \cdot X(T + \tau)\right], \tag{5.9}$$

$$\rho(\tau) = R(\tau) / R(0) \tag{5.10}$$

and the symbol τ is used to denote a lag in continuous time. The caveat refers to the fact that not all processes $X(T)$ have a spectral representation; however, the existence of the Fourier transform of the autocovariance function $R(\tau)$ of $X(T)$ is a sufficient condition.

Turning to the discrete-time version of the climate process, $X(i)$, we assume also here absent trend, absent outliers and constant variability and find

$$X(i) = S \cdot X_{\text{noise}}(i). \tag{5.11}$$

The spectral theory is in this case similar to the continuous-time case (Priestley 1981: Section 4.8.3 therein), except that the frequency range is now restricted in both directions and the discrete Fourier transform is invoked to calculate the power spectral density functions. For example, with even time spacing, $d(i) = d > 0$,

$$g(f) = (d/\pi) \sum_{l=-\infty}^{\infty} \rho(l) e^{-2\pi i f l} dl, \qquad 0 \leq f \leq 1/(2d). \tag{5.12}$$

Herein, l denotes a lag in discrete time. The frequency $f_{\mathrm{Ny}} = (2d)^{-1}$ is denoted as Nyquist frequency; it sets the upper frequency bound.

5.1.1 Example: AR(1) process, discrete time

Consider the discrete-time AR(1) process (Section 2.1.1) with an autocorrelation parameter a on an evenly spaced timescale, $d(i) = d > 0$, with $n = \infty$ points. Then (Priestley 1981: Section 4.10 therein),

$$g(f) = 2d(1-a^2) \Big/ \left[1 - 2a\cos(2\pi f d) + a^2\right], \qquad 0 \le f \le 1/(2d). \tag{5.13}$$

Plots of the AR(1) spectrum (Fig. 5.1) show higher power at lower frequencies for $a > 0$; such a spectrum is, hence, called "red."

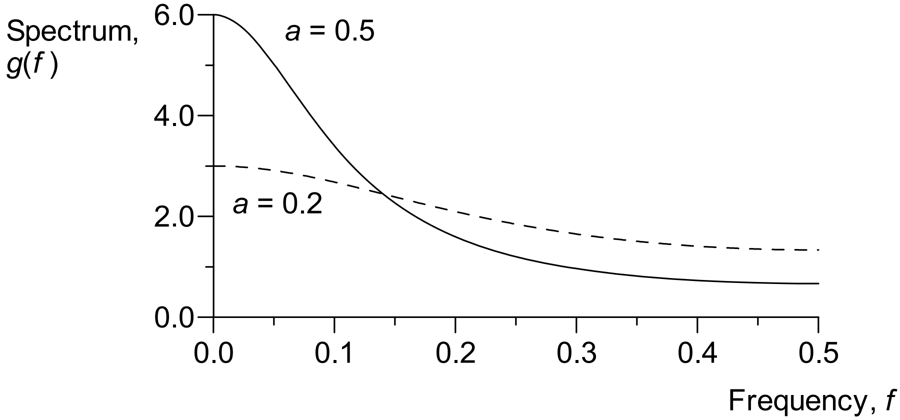


Figure 5.1. Spectrum of the AR(1) process (Eq. 5.13). Two parameter settings are shown; $d = 1$ and $f_{\mathrm{Ny}} = 0.5$.

5.1.2 Example: AR(2) process, discrete time

Consider the discrete-time AR(2) process (Section 2.2) with parameters a_1 and a_2 on an evenly spaced timescale with $d > 0$ and $n = \infty$. Then (Priestley 1981: Section 4.10 therein),

$$\begin{aligned} g(f) \;=\; & 2d(1+a_2)(1-a_2)^{-1}\left[(1-a_2)^2 - a_1^2\right]\,\big[(1+a_2)^2 \\ & + a_1^2 - 2a_1(1-a_2)\cos(2\pi f d) - 4a_2\cos(2\pi f d)^2\big]^{-1}, \end{aligned} \tag{5.14}$$

with $0 \le f \le 1/(2d)$. Plots of the AR(2) spectrum (Fig. 5.2) reveal that besides redness such spectra may exhibit quasi-cyclical behaviour (Eq. 2.15).

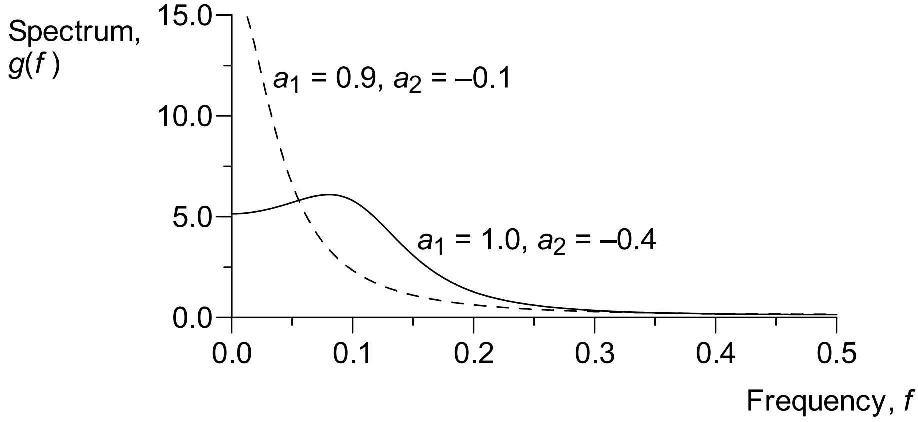


Figure 5.2. Spectrum of the AR(2) process (Eq. 5.14). Two parameter settings are shown; $d = 1$ and $f_{\mathrm{Ny}} = 0.5$.

5.1.3 Physical meaning

The importance of the power spectral density functions $h(f)$ and $g(f)$ lies in the possibility of decomposing a process into contributions from different frequency intervals. That allows to separate short-term from long-term variations and also to distinguish between cyclical forcing mechanisms of the climate system and broad-band resonances. This means that spectral analysis permits to learn about the physics of the sampled climate system. As always when having instead of a perfect knowledge only a handful of data contaminated with measurement and, perhaps, proxy errors, the task is to *estimate,* namely the spectrum. The following sections explain methods to infer $h(f)$ or $g(f)$ from $\{t(i), x(i)\}_{i=1}^{n}$.

We expect the climate spectrum either as continuous (Fig. 5.3b), reflecting a random process, or as a mixture of continuous and line components (Fig. 5.3c), the latter representing a deterministic, periodic influence. Note that estimating a spectrum is estimating a function from a finite data set. This means we can expect more difficulties and a higher susceptibility to the validness of made assumptions than for easier tasks, where only few parameters have to estimated, such as in linear regression.

A word on the notation: The literature has developed a rich variety of different notations (factors 2π, frequency versus angular velocity, etc.),

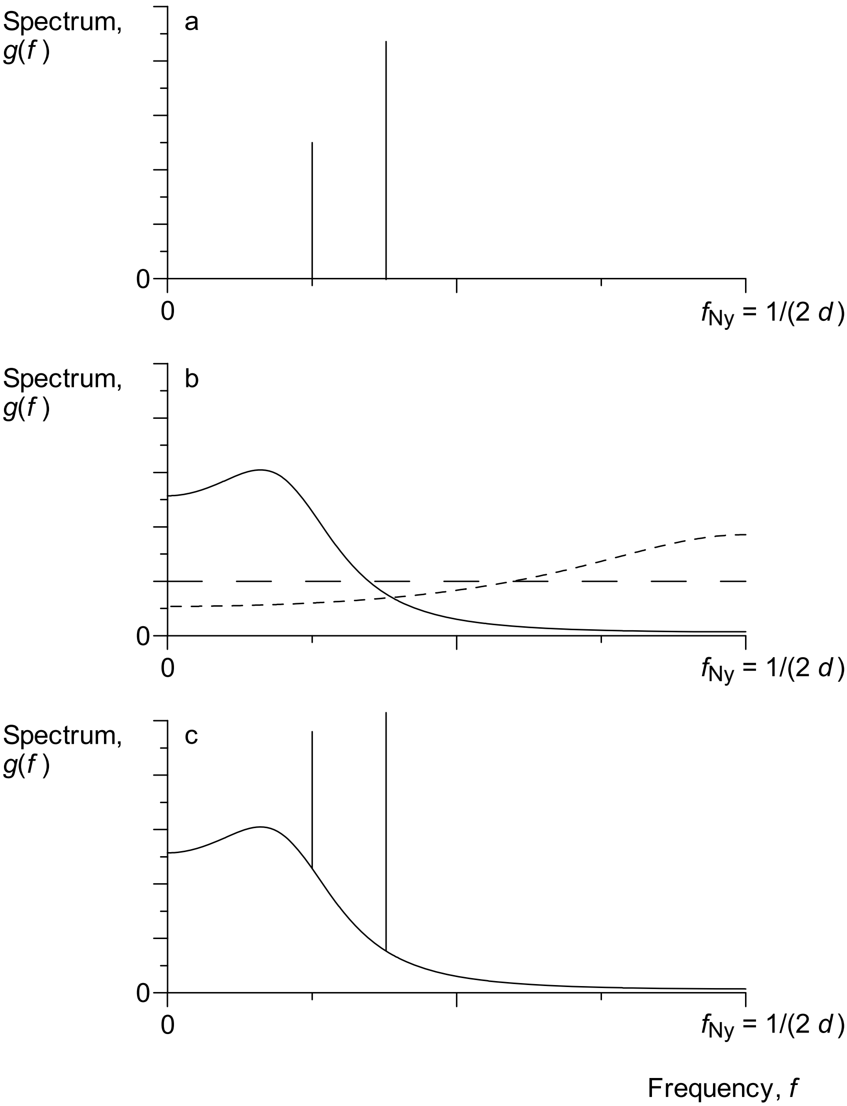


Figure 5.3. Spectrum types. **a** Line spectrum; **b** continuous spectra; **c** mixed spectrum. Three continuous spectra are shown: red noise (*solid line*), blue noise (*short-dashed line*) and white noise (*long-dashed line*). A line spectrum can be described mathematically by means of a Dirac delta function (arbitrarily narrow, arbitrarily high, finite integral).

and our is just one option. Likewise we say "spectral analysis" instead of "frequency analysis" to avoid connotations with something counted.

5.2 Spectral estimation

Estimation of the power spectral density function has in practice to be carried out using a finite data set $\{t(i), x(i)\}_{i=1}^{n}$. We expect the climate noise process $X_{\mathrm{noise}}(i)$, and respectively the trend- and outlier-free climate process $X(i)$, to exhibit persistence (time domain) or redness (frequency domain). Rarely, a blue background may be found (Section 2.6). We may possibly detect peaks superimposed on the smooth background spectrum (Fig. 5.3c), resulting from a periodic forcing process of the climate subsystem. Alternative causes of a peak could be a resonance or a noise component more complex than AR(1) (Fig. 5.2). Because the deterministic astronomical cycles (daily, annual, Milankovitch), which are harmonic processes (Section 5.2.1) with a discrete line spectrum (Fig. 5.3a), are likely not preserved without alteration in a climate archive, especially when the timescale is uncertain, it may be difficult in practice to distinguish between the alternatives “periodic forcing” and “complex noise.”

5.2.1 Periodogram

The process

$$X(i) = \sum_{j=1}^{K} \Big[A_j \cos\left(2\pi f_j T(i)\right) + B_j \sin\left(2\pi f_j T(i)\right) \Big] + \mathcal{E}_{\mathrm{N}(0,\, S^2)}(i) \quad (5.15)$$

with $i = 1, \ldots, n$ is called harmonic process. (This is done loosely, the strict definition requires $S^2 = 0$.) Its parameters are $\{A_j, B_j, f_j\}_{j=1}^{K}$, K and S^2. It has a line spectrum (frequencies f_j) sitting on a flat, constant background stemming from the persistence-free (white) noise.

If K and $\{f_j\}_{j=1}^{K}$ are known, then the other parameters can be obtained (Priestley 1981: Section 6.1.1 therein) by the least-squares technique, that is, by minimizing

$$\begin{aligned} SSQ\left(\{A_j, B_j\}_{j=1}^{K}\right) = \sum_{i=1}^{n} \Bigg\{ X(i) - \sum_{j=1}^{K} \Big[& A_j \cos\left(2\pi f_j T(i)\right) \\ & + B_j \sin\left(2\pi f_j T(i)\right) \Big] \Bigg\}^2 . \end{aligned} \quad (5.16)$$

This is in fact a regression and does not require even time spacing. However, the solution is simple if the spacing (d) is constant, n is even and $f_j = 1/(nd), 2/(nd), \ldots, 1/(2d)$. Then

$$\widehat{A}_j = (2/n) \sum_{i=1}^{n} X(i) \cos\left(2\pi f_j T(i)\right) \tag{5.17}$$

and

$$\widehat{B}_j = (2/n) \sum_{i=1}^{n} X(i) \sin\left(2\pi f_j T(i)\right). \tag{5.18}$$

For other frequencies, these expressions are approximate to $\mathcal{O}\left(1/n\right)$.

If the frequencies and other parameters of the harmonic process are unknown, which is more realistic, then we may try to find those "hidden periodicities" with a search technique called periodogram analysis. Assume from now on even n and also even spacing, $d(i) = d > 0$. (Uneven spacing is treated in Section 5.2.4.) The one-sided periodogram, $I(f_j)$, is then given by

$$I(f_j) = (nd/2) \cdot \left(\widehat{A}_j^2 + \widehat{B}_j^2\right), \tag{5.19}$$

where $\widehat{A}_j$ and $\widehat{B}_j$ are the least-squares estimators for a particular frequency, f_j.

The periodogram is calculated at trial frequencies, $f_j = 1/(nd), \ldots, 1/(2d)$. The idea is that where f_j is close to a true (but unknown) frequency of the harmonic process, the periodogram has a peak.

The expectation of the periodogram of the harmonic process (Eq. 5.15), for all $f \geq 0$, is (Bartlett 1955)

$$E\left[I(f)\right] = 2dS^2 + d(2n)^{-1} \sum_{j=1}^{K} \left(A_j^2 + B_j^2\right) \left[\frac{\sin\left(\pi n(f+f_j)\right)^2}{\sin\left(\pi(f+f_j)\right)^2} + \frac{\sin\left(\pi n(f-f_j)\right)^2}{\sin\left(\pi(f-f_j)\right)^2}\right]. \tag{5.20}$$

The covariance of the periodogram of the harmonic process, for all $f_1 \geq 0$ and $f_2 \geq 0$, is

$$COV\left[I(f_1), I(f_2)\right] = 4d^2 S^4 (n)^{-2} \left[\frac{\sin\left(\pi n(f_1+f_2)\right)^2}{\sin\left(\pi(f_1+f_2)\right)^2} + \frac{\sin\left(\pi n(f_1-f_2)\right)^2}{\sin\left(\pi(f_1-f_2)\right)^2}\right]. \tag{5.21}$$

In periodogram analysis of a harmonic process with true frequency f', the expected peak of $I(f)$ at around f', its width, its decay to a value of

$2\,d\,S^2$, and so forth, are determined by the terms within square brackets in Eq. (5.20), the sinusoids. Because the periodogram is evaluated only at discrete frequencies, $f_j = 1/(nd), 2/(nd), \ldots$, the peak at $f = f'$ may be missed. The advantage of having a larger sample size n is to search with a finer grid. However, having a larger sample size does not decrease the coefficient of variation of the periodogram, that is, the ratio of the standard deviation,

$$\begin{aligned} STD\left[I(f')\right] &= VAR\left[I(f')\right]^{1/2} \\ &= COV\left[I(f'), I(f')\right]^{1/2} \\ &= \begin{cases} \sqrt{8}\,dS^2 & \text{for } f = 0, (2d)^{-1}, \\ 2dS^2 + \mathcal{O}\left(1/n\right) & \text{elsewhere,} \end{cases} \end{aligned} \tag{5.22}$$

and the expectation (Eq. 5.20). That means, the periodogram is not a consistent estimator. The point of selecting f_1 or f_2 from $0, 1/(nd), \ldots, 1/(2d)$ is that then $COV\left[I(f_1), I(f_2)\right]$ vanishes, which allows construction of (multiple) parametric statistical tests for the existence of periodogram peaks against a white-noise background (Fisher 1929; Siegel 1980). Those tests employ the fact that for Gaussian distributed $X(i)$ (no periodic components, $K = 0$), $I(f)$ is chi-squared distributed; the degrees of freedom are 1 for $f = 0, 1/(2d)$ and 2 elsewhere (Priestley 1981: Section 6.1.3 therein). See the background material for more periodogram tests.

Also processes other than the harmonic (Eq. 5.15), for example, continuous-time processes, can be analysed using the periodogram (Priestley 1981: Section 6.2 therein). Despite the appealing property that asymptotically (for $n \to \infty$) the periodogram is also here an unbiased estimator of the spectrum, $h(f)$, it has several serious drawbacks.

1. The data size, n, for achieving acceptable levels of bias reduction for the periodogram may be extraordinarily high (Thomson (1982: p. 1058 therein) reports high bias values for n as large as $1.2 \cdot 10^6$).

2. The periodogram is not a consistent estimator of $h(f)$ (its estimation standard error does not approach zero as $n \to \infty$).

3. The periodogram has decreasing (with n) covariance between two neighbouring frequencies. This brings some erratic behaviour of the periodogram curve, which makes peak detection difficult.

In view of those points, the importance of the periodogram for climate time series analysis, for detecting peaks in the power spectral density function of $X_{\text{noise}}(T)$, is rather small. It can provide answers for discrete-time, harmonic processes, which may be found in climate driving mechanisms (daily, annual and Milankovitch cycles). But, as said, the sampled record of such a forced climate is likely influenced also by other mechanisms, making peak detection more difficult. A further limitation is that the parametric periodogram tests often assume properties (white-noise background, Gaussian distribution) that are not realistic for climatic noise (Chapters 1 and 2). This could, however, in principle be overcome by tests based on resampling, see Section 5.2.3.1.

Spacing, d, and data size, n, determine the frequency search grid of the periodogram as $f_j = 1/(nd), 2/(nd), \ldots, 1/(2d)$. (This applies to even n. For odd n, the maximum frequency is $f_j = (n-1)/[2(nd)]$.) In the case that the total record duration, $t(n) - t(1)$, is pre-determined (e.g., when a stalagmite has been sampled over its entire length), we can, by doing additional measurements, increase n to a value n', decrease d to a value d', hold $[(n-1)d] = [(n'-1)d']$ constant and therefore study higher frequencies, up to $(2d')^{-1}$. In the case we wish to have a finer search grid, we have to use a longer record because the minimum frequency resolution follows the relation $\Delta f_j = (nd)^{-1} \approx [t(n) - t(1)]^{-1}$. This frequency value is also denoted as fundamental Fourier frequency.

The advantage of the periodogram, in comparison with other spectrum estimation methods (Sections 5.2.2, 5.2.3 and 5.2.4), lies in its high frequency resolution (small Δf_j). Its use has therefore been advocated in a series of papers (Muller and MacDonald 1995, 1997a,b,c, 2000) dealing with the so-called "problem of the 100-ka cycle." These authors aimed at showing that the dominant spectral peak in late Pleistocene ice-volume proxy records (δ^{18}O) is at $T_{\text{period}} = 95$ ka instead of 100 ka. Although high frequency resolution is certainly desirable, we caution against over-interpreting periodograms of climate time series. We rather recommend to view the periodogram as one, extreme end of the smoothing technique (namely unsmoothed) in spectral analysis. The smoothing technique, described in the following (Sections 5.2.2, 5.2.3 and 5.2.4), trades resolution for standard error reduction and may lead to more reliable estimates of the power spectral density function.

5.2.2 Welch's Overlapped Segment Averaging

The periodogram is a "na-ive" (Percival and Walden 1993) estimator, $\widehat{h}(f)$, of the power spectral density function. To overcome the unfavourable variance property of the periodogram when (mis-)applied to estimate continuous spectra, Welch (1967) advanced the idea of Bartlett (1950) to divide a time series $\{t(i), x(i)\}_{i=1}^{n}$ into different segments, calculate the periodograms segment-wise and average them to obtain a reduced estimation variance. Welch (1967) allowed the segments to overlap

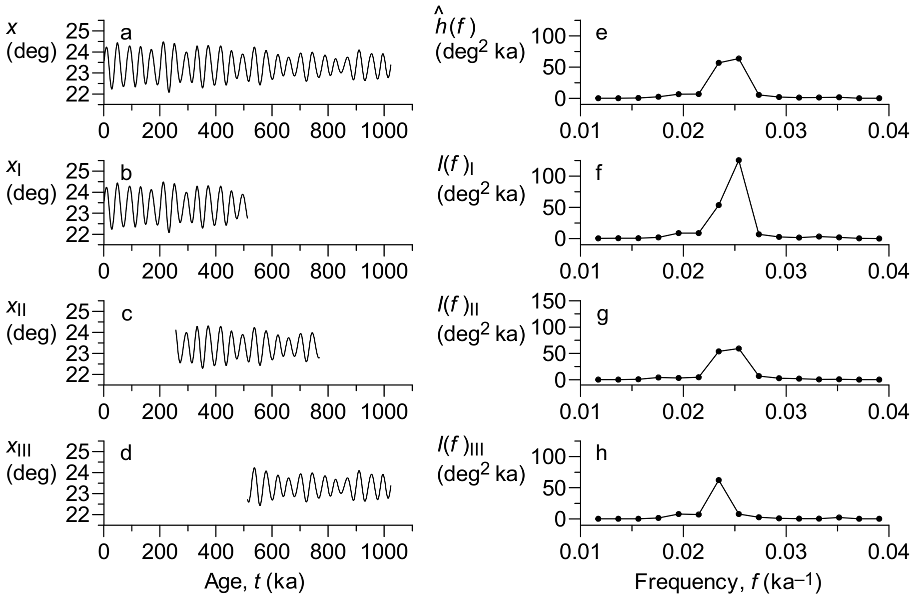


Figure 5.4. Welch's Overlapped Segment Averaging. **a** Time series of the obliquity of Earth's rotational axis over the past 1.024 Ma ($n = 1024, d(i) = d = 1$ ka). The record was produced (Berger and Loutre 1991) by solving the astronomical, kinetic equations (many-body system). The record is segmented as follows: **b** segment I, points 1 to 512; **c** segment II, points 257 to 768; and **d** segment III, points 513 to 1024. The periodograms (Eq. 5.19) are calculated for segments I (**f**), II (**g**) and III (**h**). The average of the periodograms (**e**) has a maximum at $T_{\text{period}} = 512/13$ ka ≈ 39.4 ka. Only a part of the frequency interval 0 to 1/2 ka^{-1} is shown in **e–h;** deg, degrees.

(for example, by 50%), and the method is called "Welch's Overlapped Segment Averaging" or WOSA procedure (Fig. 5.4). Overlapping has the positive effect that the number of segments, and therefore the number of averaged periodograms, is increased.

The negative effect of using WOSA (number of segments, $n_{\text{seg}} > 1$) is that the frequency points, where the periodograms are calculated, are spaced wider than for $n_{\text{seg}} = 1$. More precisely, the formula is $\Delta f_j = (n_{\text{seg}}+1)/(2nd) > 1/(nd)$ for $n_{\text{seg}} > 1$. This is the smoothing problem in the spectral domain, the trade-off between spectral estimation variance and frequency resolution. As said in the previous section, a position that advocates undersmoothing with the extreme value $n_{\text{seg}} = 1$ seems too extreme for estimating spectra of climatic processes.

Welch (1967) considered also tapering (weighting) the data points $X(i)$ within segments. Tapering is treated in the following two sections. WOSA is not the only method to obtain better bias and variance properties of spectrum estimates. Section 5.3 gives some background and references.

5.2.3 Multitaper estimation

Spectral smoothing can be accomplished with a general data weighting technique called tapering (Algorithm 5.1 and Fig. 5.5). Consider the continuous-time process $X(T)$ sampled as $\{T(i), X(i)\}_{i=1}^{n}$. The taper is a real function $w_k(T)$; k indexes the tapers; the discrete-time version is given by $\{T(i), w_k(i)\}_{i=1}^{n}$ and has the property $\sum_{i=1}^{n} w_k(i)^2 = 1$. The tapered process is then given by $\{T(i), w_k(i) \cdot X(i)\}_{i=1}^{n}$. Consider further a modified version of the periodogram (Eq. 5.19),

$$I_k(f_j) = n \cdot I(f_j). \tag{5.23}$$

A smoothed spectral estimate is then obtained by averaging a number of K modified periodograms, which are calculated from the tapered time series, $\{t(i), w_k(i) \cdot x(i)\}_{i=1}^{n}$ with $k = 0, \ldots, K-1$; see Algorithm 5.1.

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Tapers, indexed by $k = 0, \ldots, K-1$	$\{w_k(i)\}_{i=1}^{n}$
Step 3	Tapered data, $k = 0, \ldots, K-1$	$\{t(i), w_k(i) \cdot x(i)\}_{i=1}^{n}$
Step 4	Calculate modified periodograms	$I_k(f_j)$
	from tapered data, $k = 0, \ldots, K-1$	
Step 5	Average periodograms to obtain	
	smoothed spectral estimate	$\widehat{h}(f_j) = K^{-1} \sum_{k=0}^{K-1} I_k(f_j)$

Algorithm 5.1. Smoothed spectral estimation with tapering.

The periodogram is an unsmoothed spectral estimate ($K = 1, w_0(i) = n^{-1/2}$). The suggestion of Bartlett (1950) was to use $K > 1$ and non-overlapping, uniform tapers (Fig. 5.5). The recommendation of Welch (1967) was to have overlap (for example, 50%) and to allow tapers that gradually approach zero such as the Welch taper (Fig. 5.5).

It was the breakthrough of Thomson (1982) to formulate taper construction as an optimization problem, in a local least-squares sense. The solution he obtained are denoted as kth order discrete prolate spheroidal

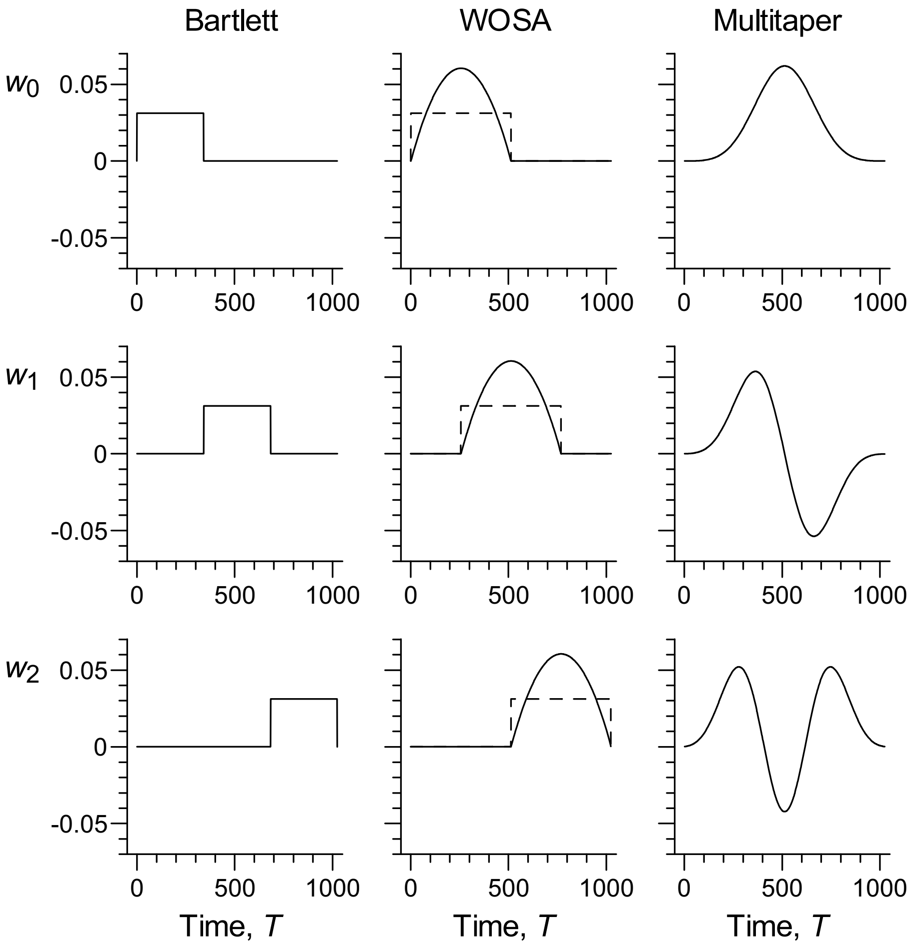


Figure 5.5. Tapers for spectral estimation. Shown are the functions $w_k(T)$, $k = 0, \ldots, K-1$, for averaging $K = 3$ (modified) periodograms, respectively eigenspectra; here $0 \leq T < 1024$ and $n = 1024$. The Bartlett type corresponds to non-overlapping and the WOSA type to (here) 50% overlapping segments. The WOSA type is shown with a uniform taper (*dashed lines*) and a normalized Welch taper (*solid lines*). The non-normalized Welch taper in continuous time is given by, for example, $w_0'(T) = 1 - [(T-256)/256]^2$ for $0 \leq T \leq 512$; the normalized Welch taper in discrete time by $w_0(i) = w_0'(i)/\sqrt{\sum_{i=1}^{n} w_0'(i)^2}$. The dpss multitaper functions have as additional parameter a resolution bandwidth of $2W = 4/(nd)$; that is, $w_k(i) = v_k^{n,W}(i)$; for convenience of presentation these discrete functions are shown as continuous plots.

sequences (dpss). The dpss had been previously described by Slepian (1978). Their calculation may be numerically difficult (Section 5.4). The dpss tapers, $v_k^{n,W}(i)$, depend on k, n and a parameter termed resolution bandwidth. The dpss have more "wiggles" than Bartlett's or Welch's suggestions (Fig. 5.5). The intuitive reason is that this leads to a smaller dependence among the modified periodograms ("eigenspectra" in the terminology of Thomson (1982, 1990a)) and, hence, to a reduced variance of their average. The resolution bandwidth, $2W$, is defined via $W = j_W/(nd)$ with j_W (not necessarily an integer) in the range from 2 to 4 (Percival and Walden 1993: Section 7.1 therein) and higher (Thomson (1990a: p. 545 therein) considers values up to 20). The resolution bandwidth limits the maximum number of eigenspectra, $K < 2\,n\,d\,W$. A larger W value has therefore the positive effect that more eigenspectra can be averaged and the spectral estimation variance reduced. On the other hand, a smaller W value lets fine details in $h(f)$ be seen better. To summarize, the combination of the multitaper parameters K and $2W$ determines estimation variance and spectral resolution.

5.2.3.1 *F* test

Thomson (1982: Section 13 therein) developed a statistical test for the existence of a line component in the spectrum against a smooth background of arbitrary shape, which is considered to be better applicable to climate time series than periodogram tests (Section 5.2.1). The idea is to compare spectral power (variance) at a frequency with the average background variance around ($\pm W$) that frequency; if the variance ratio, $F(f)$, is high, then the hypothesis of an existing line component is accepted. Under Gaussian background processes, $X(i)$, the variance ratio follows an F distribution. Thomson's recipe is as follows.

The eigenvalue problem

$$\int_{-W}^{W} \frac{\sin\left(\pi n (f - f')\right)}{\sin\left(\pi (f - f')\right)} U_k^{n,W}(f')\, df' = \lambda_k^{n,W} \cdot U_k^{n,W}(f) \tag{5.24}$$

has as solution $U_k^{n,W}(f)$ the discrete prolate spheroidal wave functions; the eigenvalues are $\lambda_k^{n,W}$ (Slepian 1978). The Fourier transform of the $U_k^{n,W}(f)$ are the dpss, $v_k^{n,W}(i)$. The eigenvalues are between 0 and 1. Let the "eigencoefficients" (Thomson 1982) of a sample be given by

$$Y_k(f) = \left(\lambda_k^{n,W}\right)^{-1} \int_{-W}^{W} U_k^{n,W}(v) \cdot Y(f+v)\, dv, \tag{5.25}$$

where $Y(f)$ is the discrete Fourier transform of the sample, $\{X(i)\}_{i=1}^{n}$.

Line components in a spectrum relate to a nonzero mean function consisting of a number of sinusoids (after detection, the sinusoidal portion can be transferred to $X_{\text{trend}}(i)$). For a single line component, the mean, $\mu(f)$, can be estimated as

$$\widehat{\mu}(f) = \sum_{k=0}^{K-1} U_k^{n,W}(0) \cdot Y_k(f) \Big/ \sum_{k=0}^{K-1} \left[U_k^{n,W}(0)\right]^2. \tag{5.26}$$

This determines the numerator of the variance ratio. The denominator is determined by the eigencoefficients and the discrete prolate spheroidal wave functions at $f = 0$. The variance ratio, finally, is

$$F(f) = \frac{(K-1)\left|\widehat{\mu}(f)\right|^2 \sum_{k=0}^{K-1} \left[U_k^{n,W}(0)\right]^2}{\sum_{k=0}^{K-1} \left|Y_k(f) - \widehat{\mu}(f) U_k^{n,W}(0)\right|^2}. \tag{5.27}$$

It is, for Gaussian $X(i)$, F-distributed with ν and $2K - \nu$ degrees of freedom (Section 5.4). For testing at a pre-defined frequency, $\nu = 2$, but if frequency is estimated as well, $\nu = 3$.

An alternative denominator for Eq. (5.27) can be used by integrating (and perhaps weighting) that expression over a frequency range of width $2W$ (Thomson 1990a: Section 5.2 therein). Obtaining the avantage of a possibly higher accuracy of the background power estimate may come at the cost of missing two line components close (within $2W$) to each other. However, such cases are likely unsolvable for noisy climate time series.

5.2.3.2 Weighted eigenspectra

Thomson (1982: Section 5 therein) advocated use not of the unweighted spectrum estimator (Algorithm 5.1) but of a scheme that puts heavier weights on eigenspectra of lower order (lower k) to reduce bias. Hence, the expression

$$\widehat{h}(f_j) = \frac{\sum_{k=0}^{K-1} b_k(f_j)^2\, \lambda_k^{n,W}\, I_k(f_j)}{\sum_{k=0}^{K-1} b_k(f_j)^2\, \lambda_k^{n,W}} \tag{5.28}$$

is the weighted multitaper spectrum estimator, where $I_k(f_j)$ is the eigenspectrum (Eq. 5.23) and

$$b_k(f_j) = h(f_j) \Big/ \left[\lambda_k^{n,W} \cdot h(f_j) + \left(1 - \lambda_k^{n,W}\right) S^2\right] \tag{5.29}$$

are the weights (Percival and Walden 1993: Section 7.4 therein). Because the true spectrum, $h(f_j)$, and the true process variance, S^2, are required to calculate the weights, an interative procedure has to be applied (Thomson 1982).

Bias reduction is a desirable aim for an estimator, but the size of the reduction seems not to have been quantified yet for spectra typical for climate processes. It may be that also usage of the simpler, unweighted estimator (Algorithm 5.1), with K not too high, leads to similar results.

5.2.3.3 Zero padding

The set of frequency values for which the covariance between spectrum estimates via the periodogram vanishes, have a spacing of $(nd)^{-1}$, the fundamental Fourier frequency (Section 5.2.1). This limit for separating two neighboured line components may be increased when spectral smoothing is applied. In multitaper estimation the limit is determined by the resolution bandwidth, $2W > (nd)^{-1}$.

Notwithstanding the limitation by $(nd)^{-1}$ is the option to "artificially" increase the resolution, by a method analogous to interpolation in the time domain. This method is zero padding (Percival and Walden 1993). That means, to the original detrended (zero mean) series, $\{t(i), x(i)\}_{i=1}^{n}$ with $d > 0$, another series, $\{t(i), x(i) = 0\}_{i=n+1}^{n^\dagger}$ with same d, is appended. Periodogram plots can now be made with resolution $(n^\dagger d)^{-1}$. F tests for line components in the multitaper spectrum (Section 5.2.3.1) can now be performed on a finer grid. We reiterate that zero padding is for such cosmetic purposes, it does not create new information.

A convenient choice is $n^\dagger$ as a power of two, because then calculations can be made fast using the Fast Fourier Transform (Section 5.4). $n^\dagger$ should not be too small; Thomson (1990a: Section 5.1 therein) recommends usage of $n^\dagger$ in the order of $4n$ to $10n$.

5.2.3.4 Jackknife

Thomson and Chave (1991) suggested a resampling approach based on the jackknife (leave one out) to evaluate the variability of the multitaper spectral estimate and construct CIs. Specifically, their approach studies not $\widehat{h}(f)$ but the natural logarithm, $\ln[\widehat{h}(f)]$. The reason is that under chi-squared distributed $\widehat{h}(f)$, taking the logarithm of the estimated spectrum leads to a symmetrical CI. (Therefore an often made advice on the graphical presentation is to plot the spectrum on a logarithmic scale.) The approach (Thomson and Chave 1991) then consists of taking the logarithms of the multitaper eigenspectra and leaving one eigenspectrum randomly out when forming the average, that is, when calculating the resampled spectrum, $\ln[\widehat{h}^{*b}(f)]$. The variability of the various $\ln[\widehat{h}^{*b}(f)]$ replications is then used by Thomson and Chave (1991) to construct

Step 1	Logarithm of eigenspectrum, $k = 0, \ldots, K-1$	$\ln\left[I_k(f_j)\right]$
Step 2	Logarithm of smoothed spectral estimate,	$\ln\left[\widehat{h}(f_j)\right] = K^{-1}\sum_{k=0}^{K-1}\ln\left[I_k(f_j)\right]$
Step 3	Jackknife version, replication, $b = 1, \ldots, K$	$\ln\left[\widehat{h}^{*b}(f_j)\right] = (K-1)^{-1}\sum_{k=0,k\neq b-1}^{K-1}\ln\left[I_k(f_j)\right]$
Step 4	Student's t CI	(Section 3.4.2)

Algorithm 5.2. Jackknife approach (Thomson and Chave 1991) to CI construction for multitaper spectrum estimate. Shown here is the case without weighted eigenspectra and without zero padding. (Cf. Algorithm 5.1.)

for each frequency point a Student's t CI. Algorithm 5.2 describes the procedure.

The advantage of evaluating uncertainties of the multitaper spectrum estimate by jackknife resampling (Thomson and Chave 1991) is twofold. First, the CIs for $\ln[\widehat{h}(f)]$ may be more robust against violations of the distributional assumptions. The jackknife CIs may therefore be more accurate than the ordinary CIs based on the chi-squared distribution. Second, what is perhaps more important in climate spectrum analysis, the variability of the $\ln[\widehat{h}^{*b}(f)]$ replications can be used to infer also the uncertainty of the estimated frequency value of a spectral peak. For the latter purpose, the advice of Thomson and Chave (1991) to use a high number of zero-padded data points (Section 5.2.3.3) is helpful because then the resolution $(n^{\dagger}d)^{-1}$ does not limit the accuracy of the estimated frequency value. The jackknife approach's drawback is that the number of eigenspectra, K, equals the number of replications, B. Since for acceptable accuracy levels of CI estimation a number B in the order of 2000 is required (Section 3.4), the limitation $K < 2\,n\,d\,W = 2j_W \leq 40$ (see beginning of Section 5.2.3) effectively means that no accurate jackknife CIs for spectrum estimates can be constructed. A situation where jackknifing leads to useful insights may be when long periods,

T_{period}, are not of interest and the whole time series can be divided into a large number of segments ("multisegmenting"). This was done by Thomson and Chave (1991) for spectral estimation of high-frequency variations of the Earth's magnetic field, observed over a total interval of 1 month. However, in climate spectrum analysis, the researcher is often in another situation, wishing to learn by means of the sampled archive also about the longer periods of climate variations.

5.2.3.5 Advanced topics: CI coverage accuracy and uneven spacing

Fodor and Stark (2000) studied the coverage performance of various bootstrap CIs in multisegment–multitaper spectral estimation for the case of a timescale with missing data. Among the interval types examined were the percentile CI, calibrated CIs and versions based on pivots; among the resampling types were the jackknife applied to the eigenspectra and the surrogate data approach applied to the time series values. The major result of the simulation experiment was the considerable inaccuracy of the various techniques for CI construction of $\widehat{h}(f)$, at two pre-defined frequencies. Only bootstrap calibration after prepivoting yielded acceptable levels of coverage accuracy (96% instead of nominal 95%). Fodor and Stark (2000) ascribed the inaccuracy to the amount of overlap among the segments, that means, the statistical dependence among the averaged eigenspectra.

Fodor and Stark (2000) presented an extension of the multitaper spectrum estimation to the special case of uneven spacing in the form of missing observations. As said before (Section 5.2.1), the periodogram estimation does not principally require even spacing. The extension of Fodor and Stark (2000) consists in an adaption of the tapers, $w_k(i)$: where $x(i)$ is missing, $w_k(i)$ is set equal to zero, and the complete $w_k(i)$ sequence is re-normalized. It may be that the problematic CI coverage performances do not stem from the application of the new concept to time series with missing values but rather from the eigenspectra dependence, as concluded by Fodor and Stark (2000). In a previous paper, Bronez (1988) introduced a new tapering scheme, called generalized prolate spheroidal sequences, to estimate the spectrum for the more general case, a process sampled at an unevenly spaced time grid. Unfortunately, studies of the CI coverage performance of this case, which often applies to climate time series, seem not to have been published in the peer-reviewed literature.

5.2.3.6 Example: radiocarbon spectrum

The radiocarbon spectrum (Fig. 5.6b) shows a number of peaks; only a few are discussed here. The low-frequency peak at 975 a period is prominent, but has a considerable frequency uncertainty. It has, on the other hand, a relatively high F value of 4.9 in the test for harmonic components. Whether it has a physical origin in variations of solar activity, the Earth's magnetic field or ocean circulation, seems not yet clear. The other low-frequency peak ($T_{\text{period}} = 556$ a) has been detected on a previous radiocarbon data set at 512 a period (Stuiver and Braziunas 1993).

The peak at 226 a period may reflect the long-term cycle named after de Vries and Suess (de Vries 1958; Münnich et al. 1958; Suess 1965), which is generally thought to present solar activity variations (Stuiver and Braziunas 1993). However, in the presented analysis (Fig. 5.6) it has an F value that is clearly lower than what Thomson (1990b) found using the same method but an older version of radiocarbon data. (To be more precise, Thomson (1990b) found two peaks, at 231 and (higher significance) 208 year period, on undetrended Δ^{14}C data with $n = 282$. Using our data without detrending (Fig. 1.6) but same parameters as in Fig. 5.6 leads to peaks at $T_{\text{period}} = 222$ a and 209 a.) Note also that solar activity variations need not form a harmonic process.

The fourth peak, at 87.6 a period, is the cycle named after the work of Gleissberg (1965), who studied nearly two millennia of auroral frequency, a proxy for solar activity variations.

It would be wrong to conclude on the sole basis of these spectral peaks that solar activity variations dominate the variations in atmospheric radiocarbon content on Holocene timescales. However, the spectrum can serve to construct a filter (Section 5.2.4.3) to extract the variations at the periods of interest and transform them into the time domain. These time series can then be compared with results from mathematical models of the Sun–climate system and enhance the quantitative physical–climatological knowledge (Solanki et al. 2004). For example, high-frequency variations such as the 10.5-year cycle in the sunspot number (Fig. 2.12), a solar activity proxy, are attenuated in tree-ring Δ^{14}C variations by exchange processes in the carbon system (Stuiver and Braziunas 1993). Furthermore, the word "dominate" should be used with caution—the area under a spectral peak is, in nearly all practical cases of climate spectrum estimation, small compared to the total area (i.e., the variance, S^2).

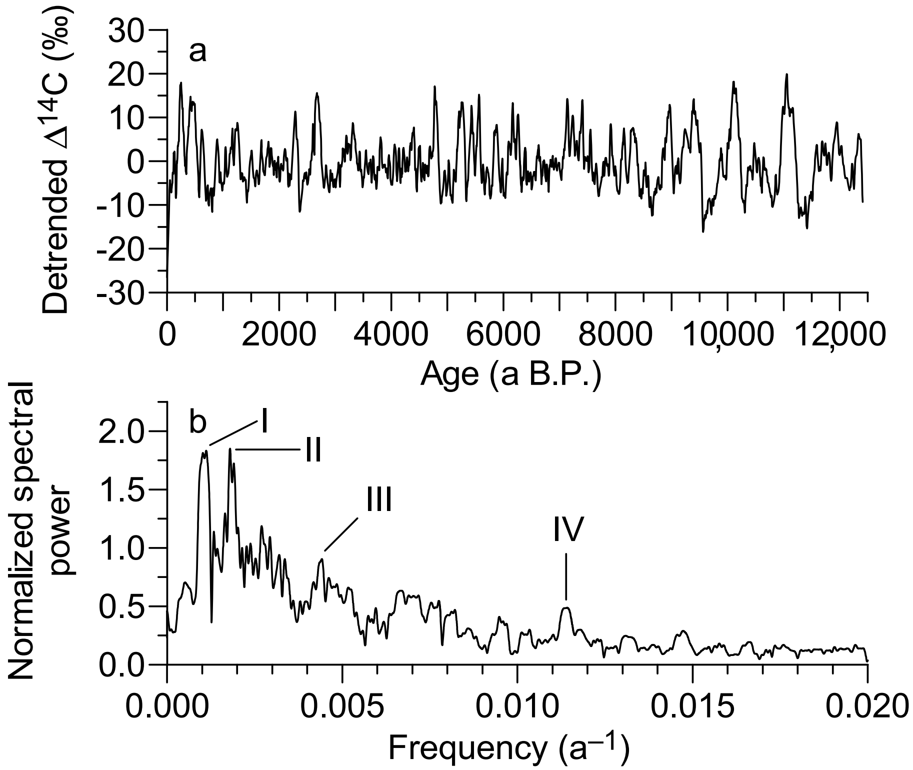


Figure 5.6. Radiocarbon spectrum, multitaper estimation. **a** Detrended radiocarbon time series, calculated by subtracting from the original data the nonparametrically estimated trend (Fig. 4.14); **b** normalized spectral power, $\widehat{g}(f)$, estimated by the multitaper technique with adaptive weighting (Section 5.2.3.2). The multitaper parameters were set as $n^\dagger = 32{,}768$, $j_W = 3.0$ and $K = 2$. Although $f_{\mathrm{Ny}} = 0.1\,\mathrm{a}^{-1}$, only the part up to $f = 0.02\,\mathrm{a}^{-1}$ is shown. Labelled are following spectral peaks: I, $T_{\mathrm{period}} = 975\,\mathrm{a}$ $\left[\left(T_{\mathrm{period}}^{-1} + W\right)^{-1} = 789\,\mathrm{a};\ \left(T_{\mathrm{period}}^{-1} - W\right)^{-1} = 1276\,\mathrm{a}\right]$, $F = 4.9$; II, $T_{\mathrm{period}} = 556\,\mathrm{a}$ $[490\,\mathrm{a}; 642\,\mathrm{a}]$, $F = 1.6$; III, $T_{\mathrm{period}} = 226\,\mathrm{a}$ $[215\,\mathrm{a}; 239\,\mathrm{a}]$, $F = 0.6$; IV, $T_{\mathrm{period}} = 87.6\,\mathrm{a}$ $[85.8\,\mathrm{a}; 89.5\,\mathrm{a}]$, $F = 2.1$; the intervals denote the uncertainty from the nonzero resolution bandwidth; also given is the F value from the test for harmonic components.

5.2.4 Lomb–Scargle estimation

The periodogram (Eq. 5.19) as spectrum estimate can also be calculated for uneven spacing, $d(i) \neq$ const., by inserting the least-squares solutions $\widehat{A}_j$ and $\widehat{B}_j$ (Eqs. 5.17 and 5.18). This was known before the work of Lomb (1976) and Scargle (1982), see the introductory parts of their papers. Scargle (1982) suggested for the case of uneven spacing, however, a new version of the periodogram,

$$I_{\mathrm{LS}}(f_j) = \bar{d} \cdot \left\{ \frac{\left[\sum_{i=1}^{n} X(i) \cos\left(2\pi f_j [T(i) - \tau_{\mathrm{Lomb}}]\right) \right]^2}{\sum_{i=1}^{n} \left[\cos\left(2\pi f_j [T(i) - \tau_{\mathrm{Lomb}}]\right) \right]^2} + \frac{\left[\sum_{i=1}^{n} X(i) \sin\left(2\pi f_j [T(i) - \tau_{\mathrm{Lomb}}]\right) \right]^2}{\sum_{i=1}^{n} \left[\sin\left(2\pi f_j [T(i) - \tau_{\mathrm{Lomb}}]\right) \right]^2} \right\}, \tag{5.30}$$

where Lomb's (1976) time shift, τ_{Lomb}, is given via

$$\tan\left(4\pi\, f_j\, \tau_{\mathrm{Lomb}}\right) = \frac{\sum_{i=1}^{n} \sin\left(4\pi f_j T(i)\right)}{\sum_{i=1}^{n} \cos\left(4\pi f_j T(i)\right)}. \tag{5.31}$$

In the case of even spacing ($d(i) = d = \bar{d}$), even n and $f_j = 1/(nd), \ldots, 1/(2d)$, it readily follows that $\tau_{\mathrm{Lomb}} = 0$ and $I_{\mathrm{LS}}(f_j) = I(f_j)$.

Scargle's objective behind introducing the Lomb–Scargle periodogram was that the distribution of $I_{\mathrm{LS}}(f_j)$ should be equal to the distribution of $I(f_j)$. Scargle (1982, 1989) showed that this is indeed so (chi-squared distribution) for a Gaussian white-noise process, $X(i) = \mathcal{E}_{\mathrm{N}(0,\, S^2)}(i)$.

5.2.4.1 Bias correction

If, what is more applicable to climate spectrum estimation, $X(i)$ is a red-noise process on an unevenly spaced timescale, perhaps with superimposed peaks, the distribution of $I_{\mathrm{LS}}(f_j)$ cannot be calculated analytically. Here, simulation methods can be used to infer the distributional properties of the Lomb–Scargle periodogram. Of particular interest is its bias.

The Monte Carlo experiment (Fig. 5.7) reveals the bias of the Lomb–Scargle periodogram for an AR(1) process and uneven spacing. The "absolute bias," of $I_{\mathrm{LS}}(f_j)$ as an estimator of non-normalized power, $h(f)$, can be substantial (Fig. 5.7a): the total area area under the curve, between zero and Nyquist frequency, nominally equal to $S^2 = 1$, is overestimated by $\sim 40\%$. The bias disappears (i.e., becomes smaller than the "simulation noise") for an AR(1) process and even spacing (Fig. 5.7b) and also, as has been shown theoretically, for a white-noise process and uneven/even spacing (Fig. 5.7c). Also in peak detection, when normalized power, $g(f)$, and its highs and lows are analysed, which is often done in climatology, the Lomb–Scargle periodogram exhibits bias. That means, even if the normalization is known, the bias of the Lomb–Scargle periodogram is significant, and it is frequency-dependent (Fig. 5.7b).

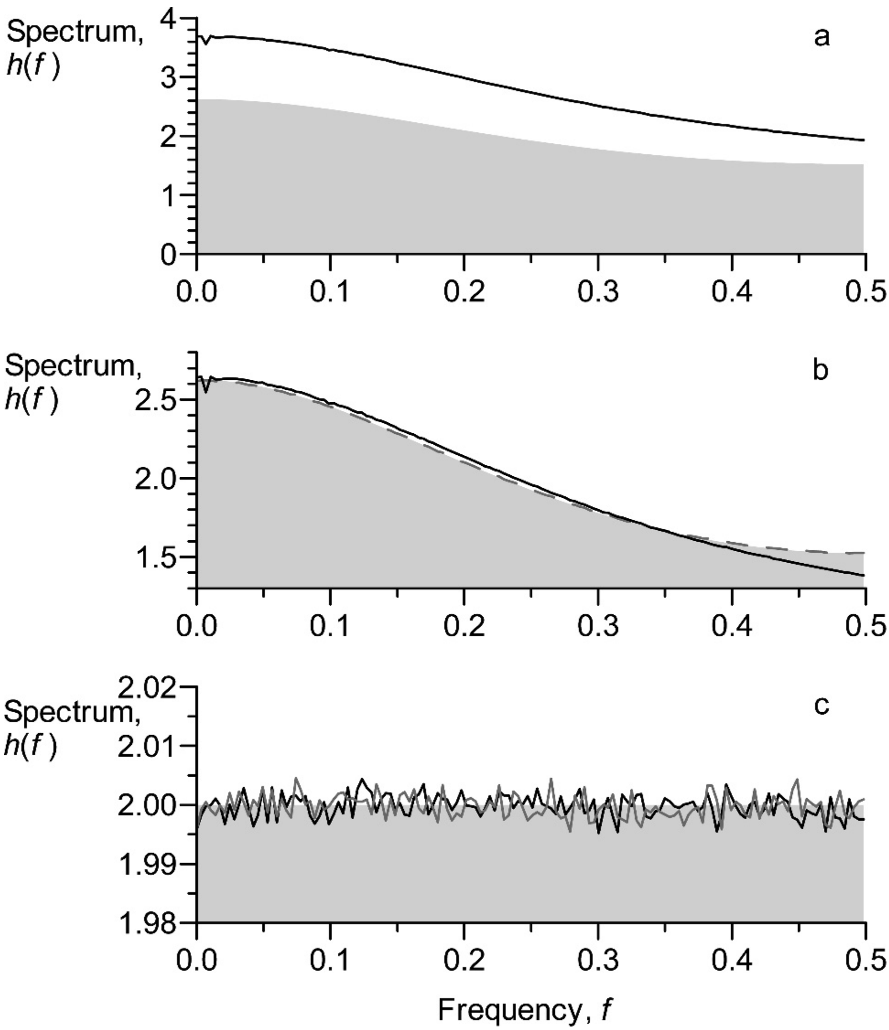


Figure 5.7. Bias of the Lomb–Scargle periodogram. Each panel shows non-normalized one-sided spectral power for selected choices of AR(1) parameters and spacings. **a** $\tau = 0.5$ and uneven spacing. The $\{t(i)\}_{i=1}^{n}$ were overtaken from the Vostok CO_2 record (Fig. 1.15b), which has $n = 283$ and $CV_d = 0.82$, and scaled to $\bar{d} = 1$. The $\{x(i)\}_{i=1}^{n}$ were generated from an AR(1) process for uneven spacing (Eq. 2.9). The *solid line* displays the average Lomb–Scargle periodogram, $I_{\mathrm{LS}}(f_j)$, on $\{t(i), x(i)\}_{i=1}^{n}$, taken from $n_{\mathrm{sim}} = 1{,}000{,}000$ simulated time series. The *shaded curve* shows the theoretical spectrum (Eq. 5.13), with $a = \exp(-\bar{d}/\tau)$. **b** $\tau = 0.5$ and uneven/even spacing. Plotted is the same as in **a,** except the following: the *solid line* shows the average $I_{\mathrm{LS}}(f_j)$, scaled such that the area under the curve equals the theoretical value (1.0); the *dashed grey line* shows the average $I_{\mathrm{LS}}(f_j)$ for data from an AR(1) process with even time spacing ($d(i) = 1$). **c** $\tau = 0$ and uneven/even spacing. The *solid black line* displays the average $I_{\mathrm{LS}}(f_j)$ (unscaled) for uneven spacing, and the *solid grey line* the average $I_{\mathrm{LS}}(f_j)$ for even spacing.

Spectrum estimation for unevenly spaced time series can be enhanced by combining the WOSA procedure with the Lomb–Scargle periodogram (Schulz and Stattegger 1997). A bias correction for such estimates was devised by Schulz and Mudelsee (2002). It uses a surrogate data approach to calculate a frequency-dependent bias correction factor from the ratio of a theoretical AR(1) spectrum and the average Lomb–Scargle spectrum of the AR(1) simulations. The bias correction (Algorithm 5.3) is thought to be well applicable to records with background spectrum coming from an AR(1) process, that is, climate records.

5.2.4.2 Covariance

The difficulty introduced to the spectrum estimation by the uneven spacing comes from the fact that the covariance for the Lomb–Scargle periodogram, $COV\left[I_{\mathrm{LS}}(f_1), I_{\mathrm{LS}}(f_2)\right]$, does not vanish in the general case (Scargle 1982). This makes not only the determination of the number of independent frequencies a problem, it may also lead to spurious peaks, having their origin in "true" power at another, not necessarily closely neighboured frequency. (In spectrum estimation on evenly spaced time series, a "true" spectral peak of the data generating process may lead not only to a peak in the estimated spectrum close to the "true" frequency, but also to a set of neighboured peaks—sidelobes (Percival and Walden 1993).)

A guide for detection of spurious peaks, oriented on the review of the Lomb–Scargle method by VanDongen et al. (1997: Section 2.4 therein), is to construct plots of $COV\left[I_{\mathrm{LS}}(f_1), I_{\mathrm{LS}}(f_2)\right]$, empirically determined using Monte Carlo simulations of AR(1) processes with same τ and sampling times as the time series under investigation. Points in the f_1–f_2 plane with large absolute covariances would then indicate where in the estimated spectrum to look for (spurious) peaks. A second option (Schulz and Stattegger 1997) when peaks at, say, f_1' and f_2' are investigated, is to use a filter (next section) to remove the signal component associated with f_1', re-calculate $I_{\mathrm{LS}}(f_j)$ and look whether the peak at f_2' still exists; and vice versa.

5.2.4.3 Harmonic filter

Scargle (1982) showed the equivalence of $I_{\mathrm{LS}}(f_j)$ estimation (frequency domain) and least-squares fitting of a harmonic model to unevenly spaced records (time domain). In analogy to that, Ferraz-Mello (1981) devised a filter algorithm for separating harmonic signal components at a predefined frequency, f', from the process $X(i)$ for uneven spacing:

$$X'(i) = X(i) - X_{f'}(i). \tag{5.32}$$

Step 1	Time series	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Spectrum estimate	$\widehat{h}(f_j)$
	from WOSA procedure with segments,	
	tapering, segment-wise linear	
	detrending and $I_{\mathrm{LS}}(f_j)$	
Step 3	Estimate persistence time segment-wise	
	with bias correction	$\widehat{\tau}'$
	and take average over segments	$\langle\widehat{\tau}'\rangle$
Step 4	Determine area	$A_{\widehat{h}}$
	under spectrum within $[0; (2\bar{d})^{-1}]$	
Step 5	Generate AR(1) data (Eq. 2.9)	$\{x^*(i)\}_{i=1}^{n}$
	with $\tau = \langle\widehat{\tau}'\rangle$	
	and times $\{t(i)\}_{i=1}^{n}$	
Step 6	Spectrum estimate	$\widehat{h}^*(f_j)$
	for AR(1) data, analogously to Step 2	
Step 7	Scale $\widehat{h}^*(f_j)$ such that area	
	within $[0; (2\bar{d})^{-1}]$ equals	$A_{\widehat{h}}$
Step 8	Repeat Steps 5–7 until n_{sim} (at least	
	1000) copies of scaled $\widehat{h}^*(f_j)$ exist	
Step 9	Take average over the n_{sim} copies,	$\langle\widehat{h}^*(f_j)\rangle$
Step 10	Theoretical AR(1) spectrum (Eq. 5.13)	
	with $a = \exp(-\bar{d}/\langle\widehat{\tau}'\rangle)$,	
	subsequently scaled to area	$A_{\widehat{h}}$
	and denoted as	$h_{\mathrm{AR(1)}}(f_j)$
Step 11	Calculate correction factor	$c(f_j) = \langle\widehat{h}^*(f_j)\rangle / h_{\mathrm{AR(1)}}(f_j)$
Step 12	Bias correction	$\widehat{h}'(f_j) = \widehat{h}(f_j)/c(f_j)$

Algorithm 5.3. Bias correction of Lomb–Scargle spectrum estimate.

In Eq. (5.32), $X'(i)$ is the process after filtering and

$$X_{f'}(i) = A\cos\left(2\pi f' T(i)\right) + B\sin\left(2\pi f' T(i)\right) + C \tag{5.33}$$

is the signal; A, B and C are parameters, which can be estimated by least squares (Ferraz-Mello 1981).

5.2.4.4 Advanced topics: degrees of freedom, bandwidth, oversampling and highest frequency

The degrees of freedom of the chi-squared distribution of a Lomb–Scargle spectrum estimate based on WOSA with 50% overlap and normally distributed $X(i)$ are

$$\nu = 2n_{\mathrm{seg}} \big/ \left(1 + 2c^2 - 2c^2/n_{\mathrm{seg}}\right), \tag{5.34}$$

where $c \leq 0.5$ is a constant representing the taper. A uniform taper has $c = 0.5$, a Welch taper $c = 0.344$; further values are listed by Harris (1978).

The discrete Fourier transform of a purely harmonic process ($S^2 = 0$) with frequency f' has (process level) a peak at f'. The decay on the flanks of the peak to a value of $10^{-6/10} \approx 0.251$ times the maximum value defines a width in frequency, the 6-dB spectral bandwidth, B_{s}. This is a useful quantity for assessing the frequency resolution, how good closely neighboured spectral peaks can theoretically be separated (Harris 1978; Nuttall 1981). The 6-dB bandwidth depends on n, n_{seg} and the type of taper.

Instead of calculating the Lomb–Scargle periodogram at frequencies $f_j = 1/(n\bar{d}), 2/(n\bar{d}), \ldots$, there is no hindrance to using a finer frequency grid. The increased number is described by the oversampling factor. This technique of artificially increasing the frequency resolution corresponds to zero padding in the time domain (Section 5.2.3.3).

Instead of letting the frequency interval end at $f_j = 1/(2\bar{d})$, it is possible to study higher frequencies because for uneven spacing there exist time intervals where $d(i) < \bar{d}$, that is, the process has been recorded at higher than average resolution. Giving guidelines for selecting the highest frequency is difficult. The choice $1/[2\min(d(i))]$ (Roberts et al. 1987) seems rather high; additionally restricting the extension of the frequency range to a value of, say, 110%, may be safer; an analysis of the $d(i)$ distribution can be helpful. Bias correction (Section 5.2.4.1) is not straightforward for $f_j > 1/(2\bar{d})$ because there the theoretical AR(1) spectrum is not defined.

5.2.5 Peak detection: red-noise hypothesis

Obtaining reliable CIs for estimated power spectral density functions is difficult in climate time series analysis because there one is usually interested also in the longer periods of variations documented in the archive and, hence, usually avoids multisegmenting. This difficulty is not restricted to the advanced multitaper estimation (Section 5.2.3.4), it applies to all spectrum estimation methods. In this situation, hypothesis tests (Section 3.6) may offer better insight by studying whether peaks (or lows) in $\widehat{h}(f)$ are significant or not. Such information is for the climate time series analyst of major relevance because it helps to filter out the variability, to construct and test conceptual climate models—the accuracy in the absolute value of $h(f)$ is less important. The typical test performed in climate spectral analysis is of H_0: "$X(i)$ is an AR(1) process, with continuous, red spectrum," the red-noise hypothesis, against H_1: "$X(i)$ has a mixed spectrum, with peak at $f_j = f_j'$ on a red-noise background."

The null distribution of $\widehat{h}(f)$ is obtained by fitting an AR(1) process to $\{t(i), x(i)\}_{i=1}^{n}$, that is, estimating a (even spacing) or τ (uneven spacing) with bias correction, followed by bootstrap resampling. The latter can be done as ARB (Algorithms 3.4 and 3.5) or via the surrogate data approach. Algorithm 5.4 shows the surrogate data simulation applied to uneven spacing (Lomb–Scargle estimation). At Step 5, $\widehat{\tau}'$ is plugged in for τ in Eq. (2.9). The peak detection test of the red-noise hypothesis can be performed also for even spacing and other spectrum estimation techniques. The usual caveat against the surrogate data approach, in favour of ARB resampling, is thought to be less severe here because deviations from the assumed Gaussian shape may have less effect in spectral estimation. This is based on the observation that spectral estimates are quadratic forms, $\sum X^2$, and the central limit theorem. Therefore, instead of (or in addition to) performing bootstrap resampling for deriving the null distribution, one may calculate upper (lower) bounds via the chi-squared distribution (Schulz and Mudelsee 2002). Of practical relevance in climatology, although conventionally ignored, should be not only peaks but also lows in $\widehat{h}(f)$, frequency intervals where less power than under an AR(1) hypothesis resides.

5.2.5.1 Multiple tests

Assume for convenience even data size, even spacing and no zero padding/oversampling. If the hypothesis test for the existence of a spectral peak is to be carried out for one single, pre-defined frequency $f_j' \in \{f_j\}_{j=1}^{n/2}$, then selection of the $100(1-\alpha)$th percentage point of the

Step 1	Time series	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Bias-corrected Lomb–Scargle	
	spectrum (Algorithm 5.3)	$\widehat{h}'(f_j)$
Step 3	Estimated, bias-corrected	
	persistence time	$\widehat{\tau}'$
Step 4	Determine area	$A_{\widehat{h}'}$
	under spectrum within $[0; (2\bar{d})^{-1}]$	
Step 5	Generate AR(1) data (Eq. 2.9)	$\{t(i), x^*(i)\}_{i=1}^{n}$
Step 6	Bias-corrected Lomb–Scargle	
	spectrum estimate	$\widehat{h}'^{*b}(f_j)$
	for generated series	
	(b, counter),	
	scaled to area	$A_{\widehat{h}'}$
Step 7	Go to Step 5 until $b = B$	
	replications exist	
Step 8	Test at each f_j whether	
	$\widehat{h}'(f_j)$ exceeds a pre-defined	
	upper percentile of $\left\{\widehat{h}'^{*b}(f_j)\right\}_{b=1}^{B}$	

Algorithm 5.4. Test of red-noise spectrum hypothesis for uneven spacing, Lomb–Scargle estimation and surrogate data resampling. The size of B depends on the size of the percentile.

empirical distribution of $\left\{\widehat{h}'^{*b}(f_j)\right\}_{b=1}^{B}$ leads to a one-sided red-noise hypothesis test with a P-value equal to α. (Alternatively, the chi-squared distribution may be used. Analogously, the 100αth percentage point is used for testing for the existence of a spectral low.) If, what is usually the case, the test is multiple, that means, it is to be carried out for many (if not all) frequencies from the set $\{f_j\}_{j=1}^{n/2}$, then a higher frequency-pointwise confidence level, $(1 - \alpha')$ with $\alpha' < \alpha$, has to be employed to yield an overall P-value of α. If a test is performed multiple times, it becomes more likely to find a significant single result.

One may define a "maximum effective number of test frequencies," M, via the overall prescribed P-value: $(1 - \alpha')^M = 1 - \alpha$. For small α and large M this leads to $\alpha' \approx \alpha/M$. The effective number of frequencies refers to a hypothetical situation where M frequencies f'_j are tested and the spectrum estimates $\widehat{h}(f'_j)$, a set of size M, are mutually independent. For the simple case of even data size, even spacing, Gaussian distributional shape and periodogram estimation (Section 5.2.1), independence is fulfilled and the maximum number is $M = n/2$. If n is odd (other setting unchanged), $M = (n - 1)/2$. Also if the Gaussian assumption is violated not too strongly, the effects on M should be negligible.

Uneven spacing with Lomb–Scargle periodogram estimation (i.e., no WOSA) can have a stronger influence on M. Since the periodogram estimates are then no longer independent, M is reduced. Horne and Baliunas (1986) and VanDongen et al. (1997) studied the effects by means of Monte Carlo simulations. If the $\{t(i)\}_{i=1}^{n}$ are more or less uniformly distributed, the approximation $M \approx n/2$ is still acceptable. This formula may also be applied to series for which the timescale is even with the exception of a few missing observations. However, if the time points are highly clustered in time, one should not use the number of points, n, but rather the number of clusters, for determining M (VanDongen et al. 1997). The effects of segmenting (WOSA) on M with Lomb–Scargle or ordinary periodogram estimation (no tapering) can be taken into account by using instead of n the number of points per segment: $M = NINT[n/(n_{\mathrm{seg}} + 1)]$, see Schulz and Mudelsee (2002). The effects of tapering (WOSA, Lomb–Scargle) could in principle be studied by means of Monte Carlo simulations. Restricting the frequency range where to study peaks is another way to reduce M, see below. Evidently, this should be done prior to the analysis (Scargle 1982).

What practical conclusions can be drawn for peak detection in climate spectra? A typical situation is an unevenly spaced timescale without strong clustering, and where the researcher is interested also in the longer periods of variations recorded by the time series. Here, Lomb–Scargle periodogram estimation with tapering, WOSA and n_{seg} not too high (less than, say, 10) is an option. To have more reliability in the low-frequency spectrum portion, one usually follows a rule of thumb (Bendat and Piersol 1986) to require at least two cycles per segment, that is, one sets the minimum test frequency f_j equal to $[(2n_{\mathrm{seg}})/(n\bar{d})]$. This also reduces M. Regarding the high-frequency spectrum portion, theoretically the uneven spacing allows inferences also for frequencies above $1/(2\bar{d})$, see Scargle (1982). On the other hand, an archive may a priori be known not to preserve a high-frequency signal, for example a marine sediment core affected by bioturbation (Fig. 1.13). Then it would make sense to ignore a part of the high frequencies, leading to a further reduction of M.

5.2.6 Example: peaks in monsoon spectrum

Figure 5.8 shows the Lomb–Scargle spectrum of the δ^{18}O record from stalagmite Q5. This allows proxy insight into the physical processes responsible for the variations of monsoonal rainfall intensities on Holocene timescales. A high oversampling provides a fine frequency grid, $\Delta f_j = (n_{\text{seg}} + 1)/(64 \cdot 2\, n\bar{d}) \approx 0.055/(n\bar{d})$.

The resulting spectrum exhibits a number of peaks above the upper bounds for the AR(1) hypothesis. Peak I ($T_{\text{period}} = 10.9$ a) is significant also when taking the test multiplicity into account. This peak from a Holocene monsoon proxy record may correspond to the sunspot cycle found in the 1716–1995 data (Fig. 5.9). Not as high confidence levels are achieved by the three periods in the centennial band (II, $T_{\text{period}} = 107$ a; III, $T_{\text{period}} = 137$ a; IV, $T_{\text{period}} = 221$ a). However, it may be unwise to ignore them at this early stage of analysis. The last peak (V, $T_{\text{period}} = 963$ a), again strong, may be related to a peak in the spectrum (Fig. 5.6) of radiocarbon variations, which contain information about changes in solar activity. Before continuing the discussion about the peaks in the monsoon spectrum in Section 5.2.9, we explore two further error sources that can exacerbate spectral peak detection: aliasing and timescale errors.

5.2.7 Aliasing

Aliasing occurs when a process $X(T)$ with a high-frequency component (f') has been sampled at insufficient temporal resolution, that is, when (even spacing) $f' > f_{\text{Ny}} = (2d)^{-1}$. Then the power associated with the high frequency is "folded" back into the analysis interval $[0; f_{\text{Ny}}]$, which produces a spurious spectral peak at frequency f_{alias}. This "alias" of f', between 0 and the Nyquist frequency, is defined via

$$f' = 2f_{\text{Ny}} \pm f_{\text{alias}}, 4f_{\text{Ny}} \pm f_{\text{alias}}, 6f_{\text{Ny}} \pm f_{\text{alias}}, \ldots, \tag{5.35}$$

see Priestley (1981: Section 7.1.1 therein) and Bendat and Piersol (1986: Section 10.3.2 therein). The sampling process thus bears the danger of a mis-interpretation of a spectrum caused by an unresolved high-frequency component.

For example, the sunspot cycle with frequency $f = 1/T_{\text{period}} = 1/10.5$ a^{-1} (Fig. 5.9) could theoretically be an alias of a true, higher frequency

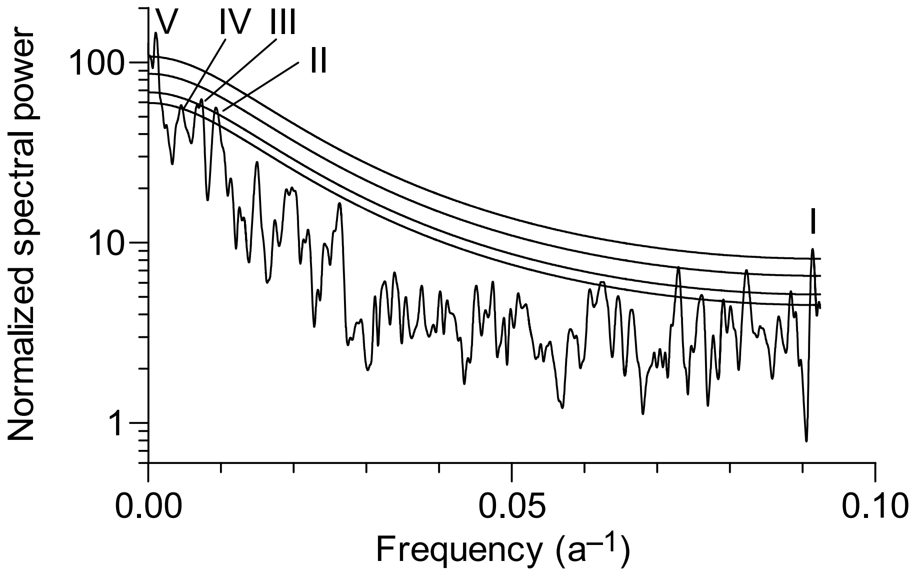


Figure 5.8. Monsoon spectrum, Lomb–Scargle estimation. Input time series is the detrended Q5 δ^{18}O record (Fig. 4.18), a proxy for Holocene Indian monsoonal rainfall. Time interval is from 8 ka B.P. to 2.7 ka B.P., that is, from after the cold and dry 8.2-ka extreme (Fig. 4.18; Rohling and Pälike (2005)) to when stalagmite growth ceased ($n = 973$, $\bar{d} = 5.4$ a). The spectrum (*wiggly line*) was calculated using the Lomb–Scargle periodogram, WOSA ($n_{\text{seg}} = 6$), tapering (Welch type) and bias correction ($n_{\text{sim}} =$ 10,000). The frequency range (number of f_j) was oversampled by a factor 64. The 6-dB bandwidth is $\sim 0.001\,\text{a}^{-1}$. *Smooth lines* show (from *bottom*) upper 90%, 95%, 99% and 99.9% chi-squared bounds for an AR(1) red-noise hypothesis ($\tau = 9.6$ a); highest bound recognizes the test multiplicity ($M = NINT[n/(n_{\text{seg}}+1)] = 139$). (Bootstrap bounds (Algorithm 5.4) are nearly identical to chi-squared bounds.) Spectral peaks labelled from I to V, possibly reflecting solar activity variations, are discussed in the text. (After Fleitmann et al. 2003.)

$f' = 2f_{\text{Ny}} + f_{\text{alias}} = 2 \cdot 0.5\,\text{a}^{-1} + 1/10.5\,\text{a}^{-1} \approx 1.095\,\text{a}^{-1}$. However, the sunspot cycle has in the past decades been observed by means of satellite measurements at much higher resolution than $d = 1$ a without any hints for such a 0.913-year cycle (Willson and Hudson 1988).

Aliasing means not only "folding" of power associated with spectral peaks, it can introduce also broad-band bias in the interval $[0; f_{\text{Ny}}]$.

The following points suggest, however, that aliasing is not a major problem in spectral analysis of climate time series.

- High time resolution. A large portion of climate time series is measured today with advanced equipment (e.g., satellites) or in laborato-

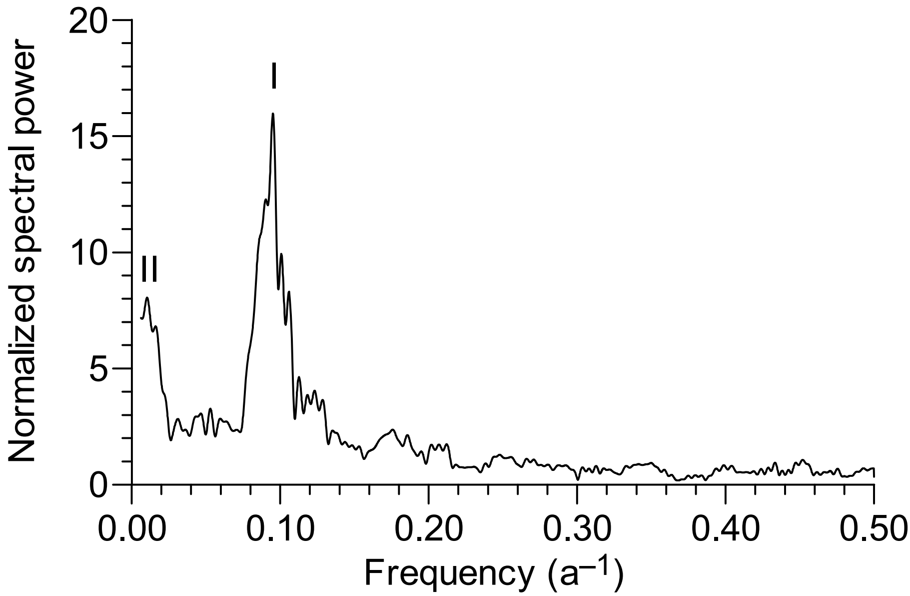


Figure 5.9. Group sunspot number spectrum. From the original data (Fig. 2.12), the time interval beginning after the Maunder Minimum was used (1716–1995, $n = 280$), a linear trend subtracted and a multitaper spectrum estimated ($n^{\dagger} = 4096$, $j_W = 3.0$, $K = 2$, $W = 0.0107\,\mathrm{a}^{-1}$). The peak (I) is at $T_{\mathrm{period}} = 10.5\,\mathrm{a}$ [9.5 a; 11.9 a]. The low-frequency peak (II) may be a remnant of a nonlinear trend.

ries with modern technology. This yields large data sizes, it allows a high sample throughput and a low sample consumption, and it therefore leads to low $d(i)$ values. This further means a reduced risk of missing high-frequency components.

- Limited degree of preservation of high-frequency variations within climate archives. A proxy time series, measured on a climate archive, is no perfect copy of a climate variable. For example: (1) Tree-ring radiocarbon variations do not well preserve the 10.5-year sunspot cycle or higher-frequency variations (Section 5.2.3.6). (2) Sediment and ice cores may be affected by diffusion-like processes in the archive (Fig. 1.13), which act as low-pass filter (Section 1.6). When high-frequency variations are not well preserved, they cannot produce large aliasing effects. To conclude, studying the physics of the climate archives to be employed is a helpful pre-sampling strategy (Wunsch and Gunn 2003).

- Uneven spacing and timescale errors. It was shown that aliasing is absent for certain types of uneven time spacing or reduced for other types (see background material). This finding applies also to the case where the timescale is assumed to be correct and have an even spacing, but dating uncertainties exist. An intuitive reason for this beneficial effect of uneven spacing is that here a Nyquist frequency is not well defined; also frequencies above $1/(2\bar{d})$ may be captured by the sampling.

The risk of spurious peaks from aliasing is likely a problem only for low-resolution, instrumental, evenly spaced time series without timescale errors. The recommendation of Madden and Jones (2001) may then be followed, namely to apply a low-pass filter (e.g., running mean) prior to the spectral analysis. Background knowledge about potential, unresolved high-frequency variations helps to design the filter.

5.2.8 Timescale errors

In the analysis of paleoclimate time series, we anticipate timescale errors. The time assigned to a sample, $T(i)$, determined by dating and possibly constructing an age–depth curve, is expected to deviate from the true time value, $T_{\mathrm{true}}(i)$. Equivalently, the spacing, $d(i)$, has an error component. This leads to a distortion of the estimated spectrum. Two effects are expected: (1) reductions of significance (detectability) of peaks compared to a situation with exact timescale and (2) increases of frequency uncertainty for a detected spectral peak.

Moore and Thomson (1991) and Thomson and Robinson (1996) studied on the process level the influence of a "jittered" spacing. The simple case of independent Gaussian jitter,

$$d(i) = d + \mathcal{E}_{\mathrm{N}(0,\,\delta_d^2)}(i), \tag{5.36}$$

is analytically tractable. (Strictly speaking, the equation refers to the spacing on the process level.) Its effect on the true continuous-time spectrum, $h(f)$, amounts (Moore and Thomson 1991; Wunsch 2000) to a multiplication by a frequency-dependent factor:

$$h_{\mathrm{distort}}(f) = \exp\left(-4\,\pi^2\,\delta_d^2\,f^2\right) \cdot h(f) + c_0, \tag{5.37}$$

where the constant c_0 serves to give the distorted spectrum, $h_{\mathrm{distort}}(f)$, the nominal area of S^2. This means, timescale errors in the form of independent jitter add white noise (c_0). As a result, spectral peaks (Section 5.2.5) have a reduced detectability.

Several assumptions went into the derivation of Eq. (5.37) by Moore and Thomson (1991), which limits its applicability to the practice of climate spectrum estimation.

- No aliasing ($h(f) = 0$ for $f > f_{\mathrm{Ny}} = 1/(2d)$). This may in practice be violated to some degree (Section 5.2.7). In addition, for unevenly spaced time series the Nyquist frequency is not well defined.
- Independent jitter. This is not realistic for many records (e.g., from ice or sediment cores). Moore and Thomson (1991) study AR(1) dependence in the jitter equation (Eq. 5.36), finding potential for larger effects on the spectrum if the dependence is strong. Still it is questionable how relevant the AR(1) jitter model is. Ice core data could exhibit heteroscedastic jitter owing to compaction. Timescales derived from layer counting may better be described by means of a random walk (Section 2.6) than by a jitter model. It is rather difficult to obtain analytical results in such cases (Section 5.3).
- Gaussian jitter distribution. This assumption is not fulfilled without imposing a constraint to guarantee a monotonic age–depth curve. (Note that Moore and Thomson (1991) studied data in the spatial domain, where no such constraint is required.)
- Process level. The mentioned paper does not study the spectrum estimators on the sample level, in particular, multitaper or Lomb–Scargle estimation.

Based on the limited relevance of available analytical results on the effects of realistic types of timescale errors on spectrum estimates (multitaper, Lomb–Scargle), we suggest two numerical simulation techniques. One quantifies the reduced detectability (Algorithm 5.5), the other the increased frequency uncertainty (Algorithm 5.6).

The following section employs both techniques. The experiments shown use the example of Lomb–Scargle estimation on the δ^{18}O record from speleothem Q5, its unevenly spaced timescale and dating errors.

5.2.9 Example: peaks in monsoon spectrum (continued)

To complete the assessment of peaks in the monsoon spectrum (Fig. 5.8), their significance and potential error sources, we consider a number of questions. Could the peaks be a result of aliasing, caused by an unresolved annual cycle? The first, statistical counterargument comes from testing whether adding an annual cycle to red noise on the Q5 time grid does lead to spectrum aliases (Fig. 5.10). Only a minor peak, at $f \approx 0.054\,\mathrm{a}^{-1}$, was found within the original frequency range 0 to $0.0924\,\mathrm{a}^{-1}$. The Q5 δ^{18}O spectrum has no corresponding signal. The peak at $f = 1\,\mathrm{a}^{-1} \gg 1/(2\bar{d})$ is rather sharp (Fig. 5.10a). However, also broader peaks (II and III) emerged. These are spurious results from the

Step 1	Time series	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Bias-corrected Lomb–Scargle	
	spectrum (Algorithm 5.3)	$\widehat{h}'(f_j)$
Step 3	Estimated, bias-corrected	
	persistence time	$\widehat{\tau}'$
Step 4	Determine area	$A_{\widehat{h}'}$
	under spectrum within $[0; (2\bar{d})^{-1}]$	
Step 5	Generate AR(1) data (Eq. 2.9)	$\{t(i), x^*(i)\}_{i=1}^{n}$
Step 6	Use timescale model	
	to resample times	$\{t^*(i)\}_{i=1}^{n}$
Step 7	Bias-corrected Lomb–Scargle	
	spectrum estimate	$\widehat{h}'^{*b}(f_j)$
	for $\{t^*(i), x^*(i)\}_{i=1}^{n}$	
	(b, counter),	
	scaled to area	$A_{\widehat{h}'}$
Step 8	Go to Step 5 until $b = B$	
	replications exist	
Step 9	Test at each f_j whether	
	$\widehat{h}'(f_j)$ exceeds a pre-defined	
	upper percentile of $\left\{\widehat{h}'^{*b}(f_j)\right\}_{b=1}^{B}$	

Algorithm 5.5. Adaption to timescale errors: test of red-noise spectrum hypothesis for uneven spacing, Lomb–Scargle estimation and surrogate data resampling. At Step 5, $\widehat{\tau}'$ is plugged in for τ. The size of B depends on the size of the percentile. The sets of frequencies f_j at Steps 2 and 7 are identical.

red noise interacting with the unevenly spaced timescale, as was found when repeating the analysis without annual cycle. Caution is therefore required when interpreting spectral peaks at frequencies much higher

Step 1	Time series	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Bias-corrected Lomb–Scargle	
	spectrum (Algorithm 5.3)	$\widehat{h}'(f_j)$
Step 3	Estimated, bias-corrected	
	persistence time	$\widehat{\tau}'$
Step 4	Determine area	$A_{\widehat{h}'}$
	under spectrum within $[0; (2\bar{d})^{-1}]$	
Step 5	Spectral peak at frequency	f'_j,
	area under peak	$\int_{f'_j-0.5B_\mathrm{s}}^{f'_j+0.5B_\mathrm{s}} \widehat{h}'(f)df = \alpha \cdot A_{\widehat{h}'}$
Step 6	Generate AR(1) data (Eq. 2.9)	$\left\{t(i), x^*_{\mathrm{AR}(1)}(i)\right\}_{i=1}^{n}$
Step 7	Generate sinusoidal data	$\{t(i), x^*_{\sin}(i)\}_{i=1}^{n}$
	with	$x^*_{\sin}(i) = (2\alpha)^{1/2} \sin\left(2\pi f'_j t(i)\right)$
Step 8	Mix series	$x^*(i) = (1-\alpha)^{1/2}\, x^*_{\mathrm{AR}(1)}(i) + x^*_{\sin}(i)$
Step 9	Use timescale model	
	to resample times	$\{t^*(i)\}_{i=1}^{n}$
Step 10	Bias-corrected Lomb–Scargle	
	spectrum for $\{t^*(i), x^*(i)\}_{i=1}^{n}$	$\widehat{h}'^{*b}(f_j)$
	(b, counter),	
	scaled to area	$A_{\widehat{h}'}$,
	peak at frequency	f'^{*b}_j
Step 11	Go to Step 6 until $b = B$	
	(usually $B = 2000$ to 10,000)	
	versions of f'^{*b}_j exist	
Step 12	Calculate $\mathrm{se}_{f'_j}$, construct CI for	f'_j
	using $\left\{f'^{*b}_j\right\}_{b=1}^{B}$	

Algorithm 5.6. Adaption to timescale errors: determination of frequency uncertainty from timescale errors for uneven spacing, Lomb–Scargle estimation and surrogate data resampling. Step 8: $VAR\,[X^*(i)] = (1-\alpha) + VAR\,[X^*_{\sin}(i)] \approx 1$.

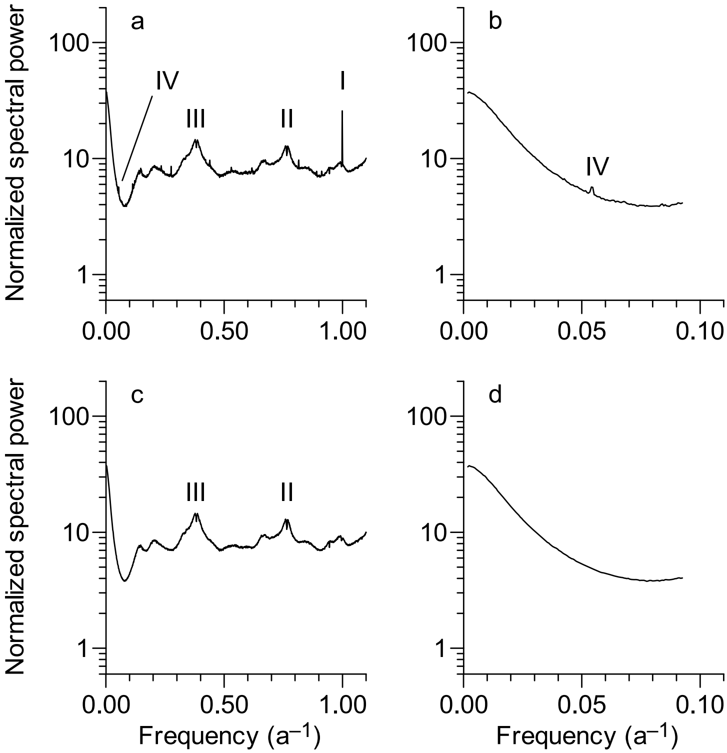


Figure 5.10. Monsoon spectrum, test for aliasing. The time grid is from the unevenly spaced Q5 δ^{18}O record (Fig. 4.18); time interval [2.7 ka B.P.; 8.0 ka B.P.], $n = 973$, $\bar{d} = 5.4$ a. **a, b** A sinusoid with frequency $1.0\,\mathrm{a}^{-1}$ and amplitude 0.2 is added to an AR(1) process (Eq. 2.9) with $\tau = 9.6$ a. A number of $n_{\mathrm{sim}} = 10{,}000$ time series were generated from this combined sinusoidal–AR(1) process and the spectrum estimated by means of the Lomb–Scargle periodogram (WOSA, $n_{\mathrm{seg}} = 6$, Welch taper, no bias correction, oversampling factor 64). Shown is the spectrum estimate, averaged over the simulations, for (**a**) an extended frequency range and (**b**) the range used for the monsoon spectrum estimation on the Q5 record (Fig. 5.8). In (**a**), the sinusoidal frequency appears as a prominent peak (I), but also other peaks (II–IV) emerged. Peak IV, an alias, is at $f \approx 0.054\,\mathrm{a}^{-1}$. **c, d** Same as in **a, b,** but without sinus component.

than the "average Nyquist frequency," $1/(2\bar{d})$. The second, physical line considers the sampling length, $L(i)$, determined by the drill diameter of 0.5 mm (Fleitmann 2001). This size in the depth domain is equivalent to an average of the sampling length in the time domain, $D(i)$, of 3.9 years—too long to capture annual variations. See Fig. 1.13 for an explanation of $L(i)$ and $D(i)$. Third, the climatological counterargument builds upon the observation that present rainfall over the Q5 site occurs not throughout the year but predominantly during the monsoon season, from July to September (Fleitmann et al. 2003). If that was also the case during the Holocene, which is a reasonable assumption, then the potential to record seasonal variations is reduced. Thus, aliasing has no influence on the monsoon spectrum.

The timescale of stalagmite Q5 is not exactly known, it has errors stemming from dating uncertainties. How does this influence the detectability and the frequency accuracy of the monsoon peaks?

We adopt a piecewise linear age–depth model for stalagmite Q5 with $n_{\mathrm{date}} = 11$ dating points and average dating uncertainty $\langle S_{\mathrm{date}}(j) \rangle \approx 70\,\mathrm{a}$ (Fleitmann et al. 2003: Table S1 therein). Resampling the times, $t^*(i)$, from this model and feeding them into the red-noise test (Algorithm 5.5) yields upper percentiles and allows a more realistic detection of spectral peaks. The percentiles obtained in this manner (Fig. 5.11) are over the whole frequency interval higher than the corresponding percentiles obtained from ignoring dating uncertainties, as expected. This effect seems in case of stalagmite Q5 not excessively large, except for higher frequencies (Fig. 5.11b). Especially the 99.9% level becomes inflated by the timescale errors, to such a degree that peak I at $T_{\mathrm{period}} = 10.9\,\mathrm{a}$ is not significant anymore in a multiple test. The only peak in the monsoon spectrum passing the hard test (timescale errors, multiplicity) is that at $T_{\mathrm{period}} = 963\,\mathrm{a}$.

Feeding the resampled times into Algorithm 5.6 allows to quantify the standard error, $\mathrm{se}_{f'_j}$, of the frequencies of the monsoon peaks owing to timescale errors. Again, the effects are larger on the high-frequency (Fig. 5.11b) than on the low-frequency side (Fig. 5.11a). There, the half of the 6-dB bandwidth is of the same order of magnitude as the frequency standard error, $\mathrm{se}_{f'_j}$. On the high-frequency side, the error interval for the period of peak I ($T_{\mathrm{period}} = 10.9\,\mathrm{a}$) is from $1/(1/10.9 + \mathrm{se}_{f'_j}) = 10.6$ years to $1/(1/10.9 - \mathrm{se}_{f'_j}) = 11.4$ years.

To summarize, the contribution of spectral analysis to answering the question after the existence of solar peaks in the spectrum of the Holocene monsoon proxy record from stalagmite Q5 is as follows. Peak I corresponds within error bars of frequency to the sunspot cycle, but taking into account timescale errors reduces its multiple test significance considerably. Peaks II ($T_{\mathrm{period}} = 107\,\mathrm{a}$), III ($T_{\mathrm{period}} = 137\,\mathrm{a}$) and IV ($T_{\mathrm{period}} = 221\,\mathrm{a}$), which are partly at periods similar to what is found for the Holocene radiocarbon record (a proxy for solar activity variations), are not statistically significant (multiple test) even when ignoring timescale errors. Only peak V at $T_{\mathrm{period}} = 963\,\mathrm{a}$, also a solar cycle candidate, is significant.

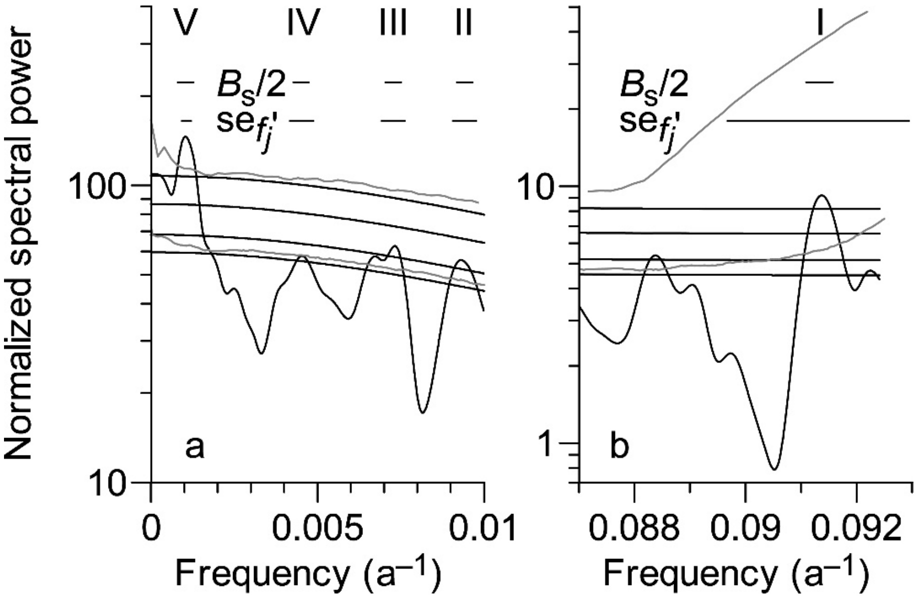


Figure 5.11. Monsoon spectrum, influence of timescale errors. Focus is on two portions (**a, b**) of the original spectrum of the proxy record Q5 (Fig. 5.8). Errors from the age determination (Fleitmann et al. 2003: Table S1 therein) were used for timescale resampling (Section 4.1.7.2) of a piecewise linear age–depth model. The 90% bootstrap bound for the red noise (the *lower* of the *grey lines* in **a** and **b**), obtained from $B = 10{,}000$ simulations, is higher than the 90% bound (the *lowest* of the four *black, smooth lines* in **a** and **b**) obtained from ignoring timescale errors. Also shown (the *upper* of the *grey lines* in **a** and **b**) is the increase in the 99.9% red-noise bound (relative to the *uppermost* of the four *black, smooth lines* in **a** and **b**), obtained from $B = 100{,}000$ simulations. The frequency uncertainties (*horizontal bars*) due to timescale errors (Algorithm 5.6, $B = 2000$), expressed as standard errors, $\mathrm{se}_{f_j'}$, are compared with the half of the 6-dB bandwidth, $B_s/2$.

It would be premature for an analysis of the Sun–monsoon relation to stop at this point. Three lines should be explored. First, the relation can be further investigated, using the same data sets, in the time domain by means of bandpass or harmonic filtering (Section 5.2.4.3). Second, the climate physics of the Sun–monsoon link can be considered. This has been done by Kodera (2004), who explained a positive correlation between solar activity and Indian monsoon strength via a weakening of the Brewer–Dobson circulation in the stratosphere. However, this was established on measurement data from 1958–1999, and the feasibility of this or other mechanisms on longer timescales is still elusive. Third, other records of Holocene monsoon variations need to be analysed. For example, Neff et al. (2001) analysed a δ^{18}O record from a stalagmite from another cave than where Q5 is from, finding monsoon peaks at $T_{\text{period}} = 10.7\,\text{a}$, $226\,\text{a}$ and $1018\,\text{a}$. Combining this evidence with the information from Q5 in a new multiple test should raise the overall statistical significance. A synopsis of evidence pro and contra the Sun–monsoon hypothesis in a multiple statistical test, with timescale errors taken into account, is a major task awaiting to be done.

5.3 Background material

Overviews of spectral analysis have been given plentiful. Accessible ones include the classic textbooks by Priestley (1981) and Percival and Walden (1993). The latter takes another route to the definition of the power spectrum than given here in Section 5.1; it also focuses on multitaper methods. John W. Tukey's work on time series and spectrum analysis, done from the 1950s to the 1990s (and summarized by Brillinger (2002)) contains useful material for the practitioner. Kay and Marple (1981) and Bendat and Piersol (1986) offer perspectives from the engineering side. A review of quadratic spectrum estimators, including multitaper methods, was given by Mullis and Scharf (1991). Reviews of spectrum analysis written by geoscientists include MacDonald (1989), Park (1992) and Ghil et al. (2002). Stratigraphy is a geoscientific subfield dealing with archived temporal changes in lithic or biotic units such as sediments. Such changes are often cyclical (Einsele et al. 1991) and quantifiable by means of spectral analysis (Weedon 2003). This area, cyclostratigraphy, deals also with series, $\{z(i), x(i)\}$, from the depth domain (Fig. 1.13). A prominent example is the identification of Milankovitch cycles in sequences such as limestone–shale from the distant geologic past (Schwarzacher 1964, 1975, 1991, 1993, 1994). There is a discussion whether we are now living in the Anthropocene (Crutzen 2002; Crutzen and Steffen 2003; Zalasiewicz et al. 2008).

Periodogram tests belong to the historically earliest tools in statistical time series analysis. A review, on which this paragraph is oriented, was given by Priestley (1981: Section 6.1.4 therein), see also Priestley (1997). The periodogram was not only invented by Schuster (1898), this man also devised a test for the significance of an $I(f_j)$ value based on the assumption of Gaussian white noise in Eq. (5.15) and the chi-squared distribution. See Brillinger (1975) for a rigorous description of CIs and other properties of the periodogram. Schuster (1898) applied his test to a "supposed 26 day period of meteorological phenomena" (the declination of the Earth's magnetic field at Prague, measured during 1870 with $d = 1$ day; the supposition had been made by Hornstein (1871)), but found little evidence in favour of a true periodicity. Later, Schuster (1906) analysed monthly sunspot time series for the interval 1749 to 1901 and detected periodicities, the major one at $T_{\text{period}} = 11.125\,\text{a}$. In that paper, Schuster also considered nonstationarity. Gilbert Walker, a physicist with contributions to meteorology, looked on $\max(I(f_j))/S^2$, with S^2 replaced by the sample variance estimator (Walker 1914), and found an asymptotic distribution. Fisher (1929) derived the exact distribution of a related test quantity for n odd, and also Hartley (1949) took Walker's test statistic, changed the denominator and derived the distribution of this re-studentized quantity. It is natural to test not only for one ($\max(I(f_j))$) but also for more harmonic components in a time series, and relevant work on this topic includes that published by Whittle (1952), Grenander and Rosenblatt (1956), Siegel (1980) and Walden (1992). A test for the number K of frequencies to include in the harmonic model (Eq. 5.15) was developed by Quinn (1989). A test for peaks in the spectrum estimated with maximum likelihood (instead of periodogram estimation) was presented by Foias et al. (1988). A serious caveat against all tests described so far in this paragraph is their assumption of a white Gaussian background noise against which to test. We assume climate processes to have rather a non-white background, that is, to exhibit a mixed spectrum (Fig. 5.3c). Statistical tests can still be constructed for mixed spectra based on analysis of $I(f_j)/h(f)$, that means, the periodogram divided by the power spectral density function of the background process. The serious practical problem here is that $h(f)$ is usually unknown and has to be replaced by an appropriate estimate, and obtaining a background estimate requires in principle the harmonic peaks to be detected. If the background spectrum has a narrow local maximum (e.g., an AR(2) spectrum), then it may be impossible to distinguish between background maximum and periodogram peak (noise and signal). Periodogram test methods to deal with such a situation require adapted background spectrum estimation (Whittle 1952; Hannan 1960, 1961; Priestley 1962a,b). An interesting alternative to periodogram tests is Thomson's F test using the background spectrum estimated with the multitaper technique (Section 5.2.3.1). The tests described so far in this paragraph were developed under the assumption of even time spacing. There exists a test using the Lomb–Scargle periodogram for unevenly spaced time series (Scargle 1982; Horne and Baliunas 1986), which is similar to Schuster's (1898)—including the restrictive assumption of white background noise. In their review of Lomb–Scargle periodogram analysis, VanDongen et al. (1997) and Van Dongen et al. (1999) mention the permutation test by Linnell Nemec and Nemec (1985). However, the permutation resampling does not preserve redness, and, hence, also this test assumes a white background. Summarizing, periodogram tests may be useful for analysing processes with line components and little/no background noise (e.g., astronomical cycles, tides), but they have little relevance for time series from climate processes with mixed, non-white spectra. We do not share the view of Muller and MacDonald (2000: p. 56 therein) that Priestley's (1981: p. 420 therein) remark of the unusefulness of the periodogram for the estimation of continuous spectra is misleading. On page 431 of his book, Priestley defends the periodogram's usefulness for estimating line spectra (Fig. 5.3a). The question raised by Muller and MacDonald (2000) is more whether their study object, Milankovitch cycles embedded in climate noise in the form of Pleistocene ice-volume changes, should indeed be analysed by means of periodogram estimation.

Superresolution refers to (almost) purely harmonic processes with a line spectrum, where a higher frequency resolution (i.e., $< \Delta f_j$) than for spectrum estimation can be achieved (Thomson 1990a). Fields for application in climatology include frequency estimation and separation of tide components (Munk and Hasselmann 1964). Also Hannan and Quinn (1989) and Quinn and Hannan (2001) studied frequency separability in dependence of n, S^2 and the amplitude of the sinusoidal components. The latter book contains further statistical tests and considers also nonstationarity in form of slowly changing frequencies.

Nonstationarity in the context of this chapter has something to do with a "time-dependent spectrum." The problem is that this is not well defined; some assumptions have to be met, and some variables to be introduced, to be able to speak of a "time-dependent frequency" or a "time-dependent power," as was reiterated by Priestley (1981, 1988). One assumption is that the time-dependences are slow and smooth. It is possible to erect a "nonstationary spectral analysis" on wavepackets or wavelets (Fig. 5.12) that have an oscillatory and a smoothly damped part (Priestley 1996). The estimation means to effectively compose a time series, $\{t(i), x(i)\}_{i=1}^{n}$, using shifted (in time) and scaled (in time) versions of the "mother wavelet," $\Psi(T)$. Most estimation algorithms seem to require (1) even spacing and (2) n to be a power of two. Evidently, interpolation methods can free the climate time series analyst from those two strong restrictions, but this seems to be at the expense of introducing heteroscedasticity and introducing or enhancing autocorrelation (Silverman 1999). The second effect could make tests of red-noise alternatives more difficult, the first require the analyst to reunite $S(T)$ with $X_{\text{noise}}(T)$. It is fair to say that a systematic and wide knowledge about interpolation effects on time-dependent spectrum estimates obtained with wavelets is not yet reached (Daubechies

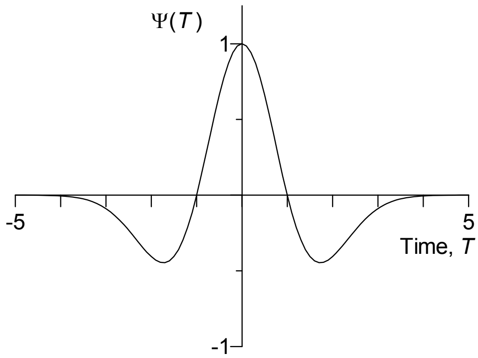


Figure 5.12. Wavelet, "Mexican hat" function, $\Psi(T) = (1 - T^2)\exp(-T^2/2)$.

et al. 1999; Vidakovic 1999; Sweldens and Schröder 2000). (For the case of nonlinear wavelet regression (time-dependent mean), Hall and Turlach (1997) studied interpolation effects theoretically and by means of a Monte Carlo simulation.) Papers on the application of wavelet estimation for unevenly spaced astronomical/climatological time series include Foster (1996a,b), Scargle (1997), Witt and Schumann (2005) and Milne and Lark (2009). A recent contribution from theory is Mondal and Percival (in press), who propose new, unbiased estimators of the wavelet power spectrum for even spacing with missing observations, analyse their large sample properties and methods for CI construction. Mondal and Percival (in press) show also an application to annual runoff minima from the Nile for the interval from A.D. 622 to 1921. To summarize, wavelet models may offer many new insights into time-dependent climate processes, but more theoretical and simulation work needs to be done, and software tools to be developed, to understand the robustness and accuracy of results with respect to uneven spacing, aliasing and timescale errors. Another technique applicable to slowly changing time-dependent spectra is to form time intervals ("windows") and estimate the spectrum separately for the windows. For example, Berger et al. (1998) studied the stability of the Milankovitch periods of variations in the Earth's orbital geometry over the interval from 1.5 Ma ago to 0.5 Ma into the future by means of a windowed multitaper estimation. Urban et al. (2000) applied the same method to look on the ENSO history within 1840–1995 as provided by the δ^{18}O proxy record from a coral ($d = 2$ months) from the central western Pacific. The ENSO spectrum exhibits power in the range of 2.2–15 years period, more broadly and not in the form of sharp peaks, and the analysis (Urban et al. 2000) shed light on the time–frequency composition of the ENSO. Schulz et al.

(1999) used windowed Lomb–Scargle estimation with WOSA for quantifying amplitude variations of the "1500-year cycle," recorded by the unevenly spaced δ^{18}O record from the GISP2 ice core. The caveats against wavelet estimation regarding robustness and accuracy of results apply also to windowed spectrum estimation techniques. One should bear in mind that the various difficulties in spectral analysis are ultimately rooted in the ambition to estimate the spectrum at frequency points f_j, which are $\mathcal{O}(n)$, on the basis of a data sample of size n. Allowing time-dependence introduces a second dimension, and estimating a quantity at $\mathcal{O}\left(n^2\right)$ time–frequency points cannot be expected to reduce the difficulties. Genton and Hall (2007) present an interesting alternative, namely estimation of parametric models for the time-dependences of frequency and amplitude, fitted in the time domain by means of kernel functions, supported by bootstrap CIs, and applicable also to uneven spacing. A notable tool for unevenly spaced series is also period analysis using robust regression in the time domain (Oh et al. 2004), which can be combined with bootstrap resampling.

The **100-ka cycle** is the dominant type of changes of global ice volume and, related, temperature and atmospheric CO_2 concentrations during the late Pleistocene (Fig. 1.3). Explaining this cycle is a challenge to the Milankovitch theorists for two reasons (Raymo and Huybers 2008). This theory of how variations in Earth orbital parameters influence climate has been successful regarding changes in the obliquity ($T_{\text{period}} \approx 40$ ka; Fig. 5.4) and precession ($T_{\text{period}} \approx 19-23$ ka) bands (Imbrie et al. 1992). First, the 100-ka cycle has a distinct sawtooth shape, which is absent in the more sinusoidal orbital time series. Second, the eccentricity component (ellipse) has a peak in that frequency range, but associated with clearly less power than the obliquity or precession components have. Ideas on how to reconcile Milankovitch theory with the 100-ka cycle include some nonlinear amplification of the eccentricity component in the climate system (Imbrie et al. 1993) and combinations of obliquity and precession components into a $\sim$ 100-ka component (Raymo 1997; Huybers and Wunsch 2005). An astronomical cause, suggested not by Milankovitch but Muller, is variations in orbital inclination ($T_{\text{period}} \approx 95$ ka), see Section 5.2.1. Non-astronomical explanations view the 1/100 ka^{-1} as kind of an eigenfrequency of the ice–bedrock–carbon cycle system (DeBlonde and Peltier 1991; Saltzman and Verbitsky 1993). It is difficult to distinguish among the various explanations on basis of statistical analyses by spectrum estimation because the 100-ka cycle came into existence as late as approximately 650 ka ago, as found by Mudelsee and Schulz (1997) using a windowed version of the harmonic filter (Section 5.4). This short time span means large bandwidth and frequency resolution, Δf_j. Spectrum estimation for the 100-ka cycle could be possibly improved by considering non-sinusoidal (i.e., sawtooth) basis functions (Thomson 1982).

The **multitaper method with jackknife resampling** for CI or standard error determination has been employed in studies of various aspects of the climate system. Among them are the following. Diaz and Pulwarty (1994) analysed centuries-long proxy records of ENSO and records of potentially related variables by means of a cross-spectral analysis (spectral analysis in a bivariate setting), and Rodó et al. (1997) tested the significance of peaks in the spectra of Iberian rainfall records from 1910 to 1994. Thomson (1997) examined variations of global temperature and the logarithm of solar irradiance during nineteenth and twentieth century and calculated jackknife CIs using $j_W = 6$. This low number and the resulting large sampling fluctuations could explain the discrepancies in CI length he noted between the jackknife approach and one based on the Gaussian assumption. Hinnov et al. (2002) investigated the interhemispheric relations among various time series of D–O variations over the past 100 ka with cross-spectral analysis, and Prokopenko et al. (2006) applied the same method to solar insolation time series and a sedimentary record from Lake Baikal covering the past ~ 1.8 Ma. The criticism regarding the low j_W values and jackknife replicates (Section 5.2.3.4) is not restricted to the paper by Thomson (1997). Also the other studies mentioned so far used similar low values. The jackknife should yield more accurate results when applied to long instrumental time series, such as oceanographic (Chave et al. 1997) or seismologic (Prieto et al. 2007), for which multisegmenting is possible.

The **Lomb–Scargle method with bootstrap resampling** for bias correction has been utilized in various climate studies. Among the analysed archives "containing" the unevenly spaced records are the following: stalagmites that provide δ^{18}O and δ^{13}C proxy evidence about changes in precipitation and temperature on Holocene and late Pleistocene timescales (Niggemann et al. 2003; Holzkämper et al. 2004; Fleitmann et al. 2007a); Antarctic ice cores that give methanesulfonic acid proxy evidence about changes in winter sea ice extent over the past 100 years (Abram et al. 2007); a loess section from Nebraska, absolutely dated with radiocarbon and dosimeter technologies, that informs via colour parameters and organic carbon content about drought variations on Holocene timescales (Miao et al. 2007); a Pacific sediment core that supplies nitrogen isotopic proxy evidence about nutrient concentrations for phytoplankton growth (i.e., carbon sequestration) over the past 70 ka (De Pol-Holz et al. 2007); and, finally, a sediment core from Bear Lake

(Utah–Idaho) that documents via pollen content the regional vegetation and climate history over the past 225 ka (Jiménez-Moreno et al. 2007).

Aliasing and uneven spacing. The effects of sampling a continuous-time climate process $X(T)$ with spectrum $h(f)$ depend on the temporal spacing of the discrete sampling points, see Priestley (1981: Section 7.1.1 therein) and Masry (1984). Even spacing bears the risk of aliasing (Section 5.2.7). Different types of uneven spacing can be distinguished.

1. An independently jittered spacing (Eq. 5.36) amounts to a "disturbed" even spacing. (This is equivalent to the timescale model given by Eqs. (4.31) and (4.33) for evenly spaced $T_{\text{true}}(i)$.) Independent jitter with $\delta_d^2 \ll d^2$ still leads to aliasing effects (Akaike 1960; Shapiro and Silverman 1960; Moore and Thomson 1991). Instead of a Gaussian, also another shape may be employed for the innovation term in the equation, the jitter. A model of a jitter uniformly distributed over the interval between $-d/2$ and $+d/2$ (excluding the endpoints) respects the condition of monotonic growth for a climate archive. Beutler (1970) shows that for frequencies below $1/(2\bar{d})$, this jitter model leads to an alias-free spectral estimation. This paper demonstrates also that spectral estimation can lead to meaningful results even when the $\{t(i)\}$ are unknown and only their rank is known, a possible situation in paleoclimate time series analysis. The independent jitter model may be applicable to climate time series when an originally even spacing is superimposed by small time uncertainties (e.g., radar measurements, which are influenced by the travel time, other instrumental observations).

2. Dependent jitter means that the innovations in the spacing equation (Eq. 5.36) are autocorrelated. Here it is more difficult than for independent jitter to obtain analytical results on second-order properties of a process such as the spectrum (Thomson and Robinson 1996). The dependent jitter model may be the norm for many climate archives such as speleothems or sediment cores (Pisias and Mix 1988). This model was also used by De Ridder et al. (2006) for modelling the shell growth of a mollusk (a climate archive).

3. Poisson sampling refers to a more irregular spacing, where the times are realizations of a homogeneous Poisson process, that is, they are uniformly distributed (Chapter 6). Then the deviation from the case of even spacing is large and spectral estimation is alias-free (Shapiro and Silverman 1960).

However, above mentioned papers on aliasing do not assume application of Lomb–Scargle spectrum estimation. Instead, they study what results when (1) the $\{x(i)\}_{i=1}^{n}$ are assumed to be realizations of a process sampled on a discrete, evenly spaced time grid and (2) a spectrum estimation method is used that assumes even spacing. Analytical results on aliasing seem hardly to exist for unevenly spaced time series analysed with the Lomb–Scargle method. Scargle (1989) states that aliasing is then diminished. This is supported by a Monte Carlo simulation study (Press et al. 1992: Fig. 13.8.1 therein), where the sampling was Poisson, $n = 100$ and $\bar{d} = 1$. The sinusoidal component with prescribed frequency 0.81, larger than $1/(2\bar{d})$, was detected with high confidence. An interesting discussion was initiated by the suggestion (Wunsch 2000) that the so-called "1500-year cycle," found in late Pleistocene and Holocene climate proxy records, is an alias of the annual cycle. Meeker et al. (2001) made it clear that, at least for the Ca record from the GISP2 ice core (Mayewski et al. 1997), interval 30 to 36 ka, the annual cycle is not preserved because of a finite sample duration ($D(i)$; Fig. 1.13) and diffusion, $D'(i) > D(i)$. Nevertheless, this time interval displays Dansgaard–Oeschger variations in Ca (Meeker et al. 2001: Fig. 1C therein), for which visual inspection infers a period of roughly 1500 years. This argument against aliasing was accepted by Wunsch (2001). A climatological objection against the existence of the "1500-year cycle," however, is that the weak stationarity assumption (time constant second-order properties) is violated: this "cycle" is restricted to this time interval (D–O events 5, 6 and 7), as was shown for the GISP2 δ^{18}O record (Schulz 2002). Amazingly, there may exist not a cycle but rather a "1500-year pacing" of the onset of D–O events, that is, the onsets (during the late Pleistocene) are not always separated by $\sim$ 1500 years but sometimes by multiples of this period (Schulz 2002; Rahmstorf 2003).

Timescale error influences. The piecewise linear age–depth model with constraint "monotonic growth," which was used for analysing the effects of timescale errors on spectrum estimates for stalagmite Q5 (Section 5.2.9), had been suggested previously (McMillan et al. 2002) for sedimentary sequences in general, where the timescale is constructed using dating points and interpolation. These authors considered also the inclusion of interpolation error models.

Sun–climate connections on timescales between those of the by everyone experienced daily and annual cycles and, on the other hand, Milankovitch cycles ($T_{\text{period}} \lessapprox 19$ ka) are not based on changes in the geometry but on solar activity variations. On shorter, decadal timescales (sunspot cycle), activity variations (Woods and Lean 2007) lead to a direct climate forcing smaller than that of greenhouse gas emissions (Hansen and Lacis 1990), although inclusion of solar activity is indispensable for climate models to reproduce the observations (Hegerl et al. 2007b). Solar irradiance changes on timescales between 11 years (sunspots) and 1500 years (Bond et al. 2001) and their role for paleoclimate where assessed by numerous authors, including Cini Castagnoli and Provenzale (1997), Hoyt and Schatten (1997) and Bard and Frank (2006). The paper by Rind (2002) considers nearly all timescales (up to the age of the Earth). Pittock (1978) took a "critical look at long-term Sun–weather relationships" and detected statistical deficits in many papers that claimed a strong Sun–climate connection. However, one may in response extend the regret for an absent comprehensive synopsis of the Sun–monsoon hypothesis (Section 5.2.9) to the Sun–climate system. Required are multiple statistical tests, which take also timescale errors into account. Regarding the spectral peak at $T_{\text{period}} = 963\,\text{a}$ (Fig. 5.8), it may correpond to reported cycles at approximately 900 year period (Schulz and Paul 2002; Rimbu et al. 2004; Wanner et al. 2008), but its nature (solar or not) deserves further analysis.

Resampling in the frequency domain. The periodogram has the property that the covariance between two points, $COV\,[I(f_1), I(f_2)]$, vanishes for $f_1, f_2 \in \{1/(nd), 2/(nd), \ldots\}$ under some conditions (normal shape, even spacing). This led to the idea to resample (ordinary bootstrap) periodogram values (frequency domain) and not residuals (time domain). One technique based on that is to resample $I(f_j)$ locally, that means close to a frequency of interest f', in order to determine a confidence interval for the spectrum estimate, $\widehat{h}(f')$, see Paparoditis (2002). This technique can be applied also to spectrum estimation with tapers (Politis et al. 1992; Politis and Romano 1992b). We have written the harmonic process (Eq. 5.15) with sinus and cosinus components; an alternative notation uses terms $A_j \cos(2\pi f_j T(i) + \Phi_j)$, where Φ_j is the phase. Estimation of the phase over the frequency range, the phase spectrum, becomes important for climatology in bivariate settings. There one is interested in leads and lags at a certain period between two climate variables. The second resampling technique mentioned in this paragraph resamples the phase spectrum estimates, while leaving the amplitude spectrum estimates $(\widehat{A}_j)$ intact. The resampled data are transformed back into the time domain and serve as surrogate bootstrap data (Nordgaard 1992; Theiler et al. 1992; Kantz and Schreiber 1997; Hidalgo 2003). This surrogate data technique has been extended into the "wavelet domain" (Angelini et al. 2005). These techniques apply to evenly spaced time series, and adapting them to the case of uneven spacing should be worth the effort. One problem with spectrum estimation then, however, is that the Lomb–Scargle periodogram does not exhibit vanishing covariances (Section 5.2.4.2).

Other spectrum estimation methods than multitaper (even spacing) or Lomb–Scargle (uneven spacing) are not recommended. The Blackman–Tukey approach (Jenkins and Watts 1968) goes via Eqs. (5.7), (5.8), (5.9) and (5.10) on the sample level and takes the autocovariance or autocorrelation function, truncates it at a certain lag to smooth and transforms into the frequency domain. Because of the estimation bias and variance involved therein (Chapter 2), we subscribe to Thomson's (1990a: p. 543 therein) remark that "if your spectrum estimate explicitly requires sample autocorrelations you are almost certainly doing something wrong." Multitaper and Lomb–Scargle are nonparametric methods. Parametric methods include fitting AR(p) models in the time domain and taking the fitted model in the frequency domain. Various procedures of fitting led to various names associated with such methods: Yule–Walker, Levinson–Durbin, maximum entropy or Burg's algorithm (Percival and Walden 1993: Chapter 9 therein). We do not dispute their usefulness for fitting time-series models to data, but we remain cautious regarding parametric estimation of climate spectra. In the case of even spacing, the bias and variance properties of the estimates these parametric methods produce, should be less good than those of the optimal (least-squares sense) multitaper method. In the case of uneven spacing, these methods are not available without interpolation.

Interpolation of an unevenly spaced time series to equidistance for spectrum estimation is not recommended. Besides the ambiguity which interpolation type (linear, cubic spline, Akima spline, etc.) to take, interpolation leads generally to smoothing and distortion of the data (Belcher et al. 1994). It may introduce spurious peaks, especially at higher frequencies, as was demonstrated by Horowitz (1974) or Schulz and Stattegger (1997). Interpolation may also corrupt the test of the red-noise alternative (Section 5.2.5) because of the amount of serial dependence artificially introduced.

Bispectrum. A zero-mean process $X(i)$ can be written in the Volterra expansion, to second order, as

$$\begin{aligned} X(i) = & \sum_j g_j \cdot \mathcal{E}_{\mathrm{F}(0,\,\sigma^2)}(i-j) \\ & + \sum_{j,k} g_{jk} \cdot \mathcal{E}_{\mathrm{F}(0,\,\sigma^2)}(i-j) \cdot \mathcal{E}_{\mathrm{F}(0,\,\sigma^2)}(i-k), \end{aligned} \tag{5.38}$$

where the g_j and g_{jk} are parameters and $\mathcal{E}_{\mathrm{F}(0,\,\sigma^2)}(i)$ is a random variable with mean zero, variance σ^2 and distribution function F (Stine 1997). The linear term (the first on the right-hand side) of the expansion gives a complete description if F is Gaussian. The parameters g_j of this term are related to the spectral density function $h(f)$. The nonlinear term is required for describing non-Gaussian or nonlinear processes. It is related (Stine 1997) to a function $h(f_1, f_2)$, the bispectrum. Estimation of the bispectrum (Subba Rao and Gabr 1984) can add to a spectral characterization of an observed process. Muller and MacDonald (1997b,c) applied bispectral analysis to evenly spaced climatological and astronomical records in order to support their hypothesis that the 100-ka cycle corresponds to variations of orbital inclination. A limitation of bispectral analysis is that currently available implementations seem to be restricted to even time spacing. Notwithstanding this, researchers studying late Pleistocene global climate changes applied bispectral analysis to time series originally unevenly spaced (Hagelberg et al. 1991; King 1996; Rutherford and D'Hondt 2000). Little knowledge seems to exist on testing noise alternatives and quantifying the robustness of bispectral estimates against timescale errors.

The **trade-off between variance and bandwidth** of spectrum estimators has been referred to by many authors as "Heisenberg's uncertainty principle," although the latter is a concept from quantum physics. It would be more apt to speak of "Grenander's uncertainty principle," after, for example, the paper by Grenander (1958).

5.4 Technical issues

The **calculation of the dpss** multitapers (Section 5.2.3) can be done in various ways. Percival and Walden (1993: Chapter 8 therein) note numerical integration and bypassing the problem by using substitutes in form of trigonometric polynomials, but they favour two other calculation types, namely via a tridiagonal formulation or directly from the defining eigenvalue problem (Eq. 5.24). For solving the latter, Bell et al. (1993) developed an iterative algorithm, written in FORTRAN 77 and available from `http://lib.stat.cmu.edu/jcgs/bell-p-w` (29 January 2008). Own experiments with a Fortran 90 translation on 32-bit and 64-bit machines, where the numerical precision of real numbers can be adjusted conveniently, attest the robustness of the algorithm.

The **Fast Fourier Transform** or FFT is a numerical algorithm (Cooley and Tukey 1965) that reduces the number of operations from $\mathcal{O}\left(n^2\right)$ to $\mathcal{O}\left(n\log(n)\right)$. Data size n must be a power of two. The FFT was the technical basis of the scientific revolution that came with spectral analysis.

The ***F* distribution** with ν_Y and ν_Z degrees of freedom has following PDF:

$$f(x) = \frac{\left(\nu_Y / \nu_Z\right)^{\nu_Y/2}}{B\left(\nu_Y/2\, , \nu_Z/2\right)} \cdot \frac{x^{(\nu_Y/2)-1}}{\left(1 + x \cdot \nu_Y / \nu_Z\right)^{(\nu_Y+\nu_Z)/2}}, \tag{5.39}$$

where $x > 0$ and B is the beta function (Section 2.7). It arises as the distribution of the ratio of two chi-squared variables (Section 3.9). Let Y and Z be independent and chi-squared distributed with ν_Y and ν_Z degrees of freedom, respectively. Then $X = [(Y/\nu_Y) \cdot (Z/\nu_Z)^{-1}]$ is F-distributed (Eq. 5.39). See Johnson et al. (1995: Chapter 27 therein) for more details on the F distribution.

Lees and Park (1995) published a C subroutine for multitaper estimation. It can be obtained via `http://www.iamg.org` (29 January 2008). This software was used in an influential paper on signal detection against a red-noise background of climate spectra (Mann and Lees 1996).

Multitaper.zip is a Matlab implementation of multitaper estimation in the presence of missing data (Fodor and Stark 2000). It can be downloaded from `http://www.stat.berkeley.edu/~stark/Code/` (29 January 2008).

mwlib is a Fortran 90 library of subroutines for multitaper estimation (Prieto et al. 2009). It is available at the internet address `http://wwwprof.uniandes.edu.co/~gprieto/software/mwlib.html` (11 December 2009).

SSA-MTM Toolkit is a compiled software that includes multitaper estimation in connection with SSA (`http://www.atmos.ucla.edu/tcd/ssa/`, 29 January 2008). Version exist for DEC, Linux, Macintosh, SGI and Sun systems.

CYSTRATI is a FORTRAN 77 package, developed and listed by Pardo-Igúzquiza et al. (1994), for cyclostratigraphic data analysis, including multitaper and maximum entropy spectrum estimation.

REDFIT is a Fortran 90 program (code, Windows binaries) for Lomb–Scargle spectrum estimation with bootstrap bias correction and test of the AR(1) red-noise alternative (Schulz and Mudelsee 2002). It is based on SPECTRUM (Schulz and Stattegger 1997), which has a graphical interface but no bias correction or red-noise test. An option is interactively working with SPECTRUM to find out suitable smoothing parameters and then performing with REDFIT the final calculations. RED2CON is a recent Matlab implementation of REDFIT with graphical interface. The core of the programs lies in the routines for $I_{\mathrm{LS}}(f_j)$ calculation (Scargle 1989). REDFIT, RED2CON and SPECTRUM are available at the site `http://www.geo.uni-bremen.de/geomod/staff/mschulz/` (29 March 2010), REDFIT also at the web site for this book.

ENVELOPE is a DOS/Windows software implementing a windowed version of the harmonic filter (Section 5.2.4.3) for analysing slowly changing sinusoidal components (frequency f'). The time-dependent amplitude is $(A^2 + B^2)^{1/2}$, see Eq. (5.33). It is estimated using a least-squares criterion (Ferraz-Mello 1981; Schulz 1996). The software can be obtained from http://www.geo.uni-bremen.de/geomod/staff/mschulz/ (29 January 2008).

MATLAB Recipes for Earth Sciences is the title of a book (Trauth 2007) with software that includes Lomb–Scargle estimation.

AutoSignal is a commercial package containing spectral analysis tools, including multitaper and Lomb–Scargle estimation. It can be obtained from Systat (http://www.systat.com, 29 January 2008).

CLEAN is a deconvolution algorithm for switching between frequency and time domains while collecting iteratively the strongest spectral peaks and their time-domain representation, respectively (Roberts et al. 1987). It can be applied to unevenly spaced time series for spectrum estimation. A surrogate data resampling approach to derive significance levels (Heslop and Dekkers 2002) is available as Matlab package MC-CLEAN at http://www.geo.uu.nl/~forth/Software/mc_clean.zip (29 January 2008).

REDFITmc2 (Mudelsee et al. 2009) is an adaption of REDFIT, which implements Algorithms 5.5 and 5.6. See the web site for this book.

Chapter 6

Extreme Value Time Series

Extreme value time series refer to the outlier component in the climate equation (Eq. 1.2). Quantifying the tail probability of the PDF of a climate variable—the risk of climate extremes—is of high socioeconomical relevance. In the context of climate change, it is important to move from stationary to nonstationary (time-dependent) models: with climate changes also risk changes may be associated.

Traditionally, extreme value data are evaluated in two forms: first, block extremes such as annual maxima, and second, exceedances of a high threshold. A stationary model of great flexibility for the first and the second form is the Generalized Extreme Value distribution and the generalized Pareto distribution, respectively. Classical estimation techniques based on maximum likelihood exist for both distributions.

Nonstationary models can be constructed parametrically, by writing the extreme value models with time-dependent parameters. Maximum likelihood estimation may impose numerical difficulties here. The inhomogeneous Poisson process constitutes an interesting nonparametric model of the time-dependence of the occurrence of an extreme. Here, bootstrap confidence bands can be constructed and hypothesis tests performed to assess the significance of trends in climate risk. A recent development is a hybrid, which estimates the time-dependence nonparametrically and, conditional on the occurrence of an extreme, models the extreme value parametrically.

6.1 Data types

We distinguish among several types of extreme value data. One guide for doing so is the accuracy of $X_{\mathrm{out}}(i)$, the outlier or extreme component in the climate equation (Eq. 1.2). Even data with a very low accuracy

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_6,

can be analysed, for example, cases where only the time an extreme occurred is known. A related guide comes from considering how the extreme data were obtained. An example is outlier detection by imposing a threshold (Section 4.3.3).

6.1.1 Event times

In the low-accuracy case it is just known about an event that it did occur, that means, $X_{\text{out}}(i) \neq 0$. The time points of the events recorded by a time series are

$$\{T_{\text{out}}(j)\}_{j=1}^{m} = \{T(i) \big| X_{\text{out}}(i) \neq 0\}_{i=1}^{n}. \tag{6.1}$$

On the sample level, the set of time points inferred from analysing $\{t(i), x(i)\}_{i=1}^{n}$ is written as $\{t_{\text{out}}(j)\}_{j=1}^{m}$. The number of extreme events is m; it is $m \leq n$.

A second constraint imposed on $X_{\text{out}}(i)$, besides being unequal to zero, is independence. The observed extreme should have occurred because a climate process generated it and not because there had previously been another, interfering event.

6.1.1.1 Example: Elbe winter floods

The winter floods of the river Elbe (Fig. 1.1) were recorded with a slightly higher accuracy ($x'_{\text{out}}(j) = 1$, 2 or 3). For the documentary period (up to 1850), independence of events was achieved by studying the historical sources (Mudelsee et al. 2003). Consider the ice flood in 1784, for which Weikinn (2000) gives 32 source texts that report about the breaking ice cover in the last week of February, the rising water levels, the considerable damages this and the moving ice floes caused and, finally, the decreasing water levels in the first week of March 1784. Mudelsee et al. (2003) considered this as one single event ($t_{\text{out}}(j) = 1784.167$) and not two (February, March).

The question after the flood risk, whether winter floods occur at a constant rate or there exist instead changes, is analysed by means of occurrence rate estimation (Section 6.3.2).

6.1.2 Peaks over threshold

If $X(i)$ is known with higher accuracy, a threshold criterion may be applied to detect extremes.

$$\{T_{\text{out}}(j), X'_{\text{out}}(j)\}_{j=1}^{m} = \{T(i), X(i) \big| X(i) > u\}_{i=1}^{n} \tag{6.2}$$

is a rule for detecting maxima with a constant threshold, u. The extension to detecting minima is straightforward.

The peaks-over-threshold (POT) data can be analysed in two ways. Occurrence rate estimation (Section 6.3.2) uses the sample $\{t_{\mathrm{out}}(j)\}_{j=1}^{m}$ to infer trends in the occurrence of extremes. Fitting a generalized Pareto distribution (Section 6.2.2) to $\{x'_{\mathrm{out}}(j)\}_{j=1}^{m}$ is helpful for studying the risk of an event of pre-defined size, $\mathrm{prob}(X(i) > u + v)$ with $v > 0$.

In climatology it is also useful to consider a time-dependent threshold to take into account effects of trends in mean, $X_{\mathrm{trend}}(T)$, and variability, $S(T)$. To fulfill the assumption of mutual independence of the POT data, imposing further criteria than passing the threshold may be necessary.

6.1.2.1 Example: volcanic peaks in the NGRIP sulfate record (continued)

Outlier/extremes detection in the NGRIP sulfate record (Fig. 4.16) employed a time-dependent threshold, $X_{\mathrm{trend}}(i) + z \cdot S(i)$, and robust estimates of trend ("background") and variability, to take into account variable oceanic input. A second criterion was the absence of contemporaneous Ca and Na peaks to extract the extremes caused by volcanic eruptions (Fig. 1.4). To satisfy the independence assumption, further threshold exceedances closely neighboured in time were discarded (third criterion). In general, the size of such a neighbourhood can be estimated using persistence models (Chapter 2). Instead of taking $\{X'_{\mathrm{out}}(j)\}_{j=1}^{m}$ from $\{X(i)\}_{i=1}^{n}$, one may also collect scaled extremes $\{X'_{\mathrm{out}}(j)\}_{j=1}^{m}$ from $\{[X(i) - X_{\mathrm{trend}}(i)]/S(i)\}_{i=1}^{n}$. Scaling is one form of taking nonstationarity into account (Section 6.3).

6.1.3 Block extremes

It may sometimes be that climate or weather data are in the form of extremes over a certain time period. An example of such a block extreme is the annual maximum,

$$X'_{\mathrm{out}}(j) = \max\Big(\{X(i)\}_{T(i) \text{ within } j\text{th year of time series}}\Big), \tag{6.3}$$

$$T_{\mathrm{out}}(j) = j\text{th year of time series}. \tag{6.4}$$

The block extremes $X'_{\mathrm{out}}(j)$ are the input for fitting a Generalized Extreme Value distribution (Section 6.2.1). The estimation result sheds light on the risk at which an extreme of a pre-defined size and at a pre-defined block length occurs.

Risk estimation (Section 6.2.1) assumes that an extreme is taken from a block with a large number k (at least, say, 100) of independent observations. This can be done explicitly, by segmenting or "blocking" an original series $\{X(i)\}_{i=1}^{n}$. Alternatively, the blocking may have already been done implicitly. An example is documentary data in form of maximum annual water stage in a river, where original daily observations have not been preserved or have simply not been made. Another possibility, theoretically also conceivable, are proxy measurements with a machine that records not the mean value (e.g., of a concentration) but the extreme value. In any case, the independence assumption should be approximately fulfilled if the block length (time units) is large compared with $\max(\tau, D'(i))$ (Fig. 1.13). For practical applications, τ and $D(i)$ have to be estimated.

6.1.4 Remarks on data selection

The rules for selecting $\{X'_{\mathrm{out}}(j)\}_{j=1}^{m}$ from $\{X(i)\}_{i=1}^{n}$ are not uniquely determined. This allows the analyst to explore various climate system properties regarding extremes.

One area is threshold selection in the POT approach. Besides allowing time-dependence, the size can be adjusted. A high (low) threshold size for maxima detection leads evidently to fewer (more) cases and, hence, to more conservative (liberal) results but likely also to wider (narrower) CIs. Furthermore, a too low threshold may lead to violations of the conditions of convergence to an extreme value distribution. Data in form of event times have implicitly also undergone a threshold selection. The documentary data about Elbe floods, for example, were critically screened (Mudelsee et al. 2003) whether there is enough evidence that merits inclusion into the flood record or there had instead been just an elevated water level noticed by a hypercritical observer.

For block extremes, the adjustable parameter is the block length. In the case of original data $X(i)$ with even spacing, this corresponds to a fixed number, k, of $X(i)$ values per block. In the case of uneven spacing, besides leaving the block length constant, one may also fix k. The connection to nonparametric regression and the smoothing problem (Section 4.3) is evident.

Henceforth we omit for convenience the prime and write $\{X_{\mathrm{out}}(j)\}_{j=1}^{m}$ on the process and $\{x_{\mathrm{out}}(j)\}_{j=1}^{m}$ on the sample level.

6.2 Stationary models

In stationary models, the distribution parameters and related quantities, such as risk, do not change over time.

6.2.1 Generalized Extreme Value distribution

The Generalized Extreme Value (GEV) distribution is suitable for analysing block extremes. Our treatment follows closely that of Coles (2001b: Chapter 3 therein).

6.2.1.1 Model

The GEV distribution function is given by

$$F_{\mathrm{GEV}}(x_{\mathrm{out}}) = \begin{cases} \exp\Big\{-\big[1+\xi\,(x_{\mathrm{out}}-\mu)\,/\sigma\big]^{-1/\xi}\Big\} & (\xi \neq 0), \\ \exp\Big\{-\exp\big[-(x_{\mathrm{out}}-\mu)\,/\sigma\big]\Big\} & (\xi = 0), \end{cases} \tag{6.5}$$

where $1+\xi\,(x_{\mathrm{out}}-\mu)\,/\sigma > 0$, $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. The parameters μ and σ identify location and scale, respectively, while the shape parameter, ξ, determines the tail behaviour of $F_{\mathrm{GEV}}(x_{\mathrm{out}})$.

The importance of the GEV distribution lies in the fact that it is the limiting distribution of the block maximum (for k large). Under mild conditions, nearly irrespective of what the common, but generally unknown distributional shape of the individual variables $X(i)$ is, the distribution of $X_{\mathrm{out}}(j)$ approaches the GEV (Fig. 6.1). This is in essence the extreme value analogue of the central limit theorem (Coles 2001b).

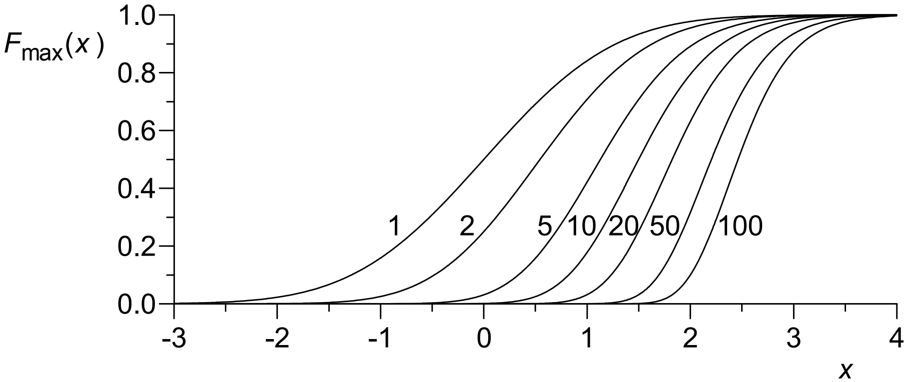


Figure 6.1. Distribution of the maximum of k independent standard normal variates. The plotted distribution functions, $F_{\max}(x)$, are labelled with k. For $k = 1$, the symmetric form of the standard normal distribution, $F_{\mathrm{N}}(x)$ (Eq. 3.49), appears. In general, $F_{\max}(x) = [F_{\mathrm{N}}(x)]^k$. Letting k increase has three effects: the location (average) is shifted to the right, the scale (standard deviation) is decreased and the right-skewness (shape parameter) is increased. With increasing k, $F_{\max}(x)$ approaches $F_{\mathrm{GEV}}(x)$. This is a theoretical example, with prescribed $F_{\mathrm{N}}(x)$ and exactly determined $F_{\max}(x)$. In a practical setting, with distribution and parameters of the independent variables unknown, $F_{\max}(x)$ can still be approximated by $F_{\mathrm{GEV}}(x)$.

6.2.1.2 Maximum likelihood estimation

Assume that the approximation is perfect and the block maxima $\{x_{\mathrm{out}}(j)\}_{j=1}^{m}$ do come from a GEV distribution (Eq. 6.5). Assume further that $\xi \neq 0$. Adopting the maximum likelihood principle (Section 2.6,

p. 58) requires then to maximize the (logarithm of the) likelihood function (Coles 2001b),

$$\ln\left[L(\mu,\sigma,\xi)\right] = -m\ln(\sigma) - (1+1/\xi)\sum_{j=1}^{m}\ln\bigl[y(j)\bigr] - \sum_{j=1}^{m} y(j), \qquad (6.6)$$

where

$$y(j) = 1 + \xi\left[\frac{x_{\text{out}}(j)-\mu}{\sigma}\right]. \qquad (6.7)$$

The additional condition is that $y(j) > 0\ \forall j$.

Section 6.6 explains the case $\xi = 0$, for which another log-likelihood function is used. That section covers also the regularity conditions, the properties of the maximum likelihood estimators in dependence on the shape parameter, ξ. To summarize, if $\xi > -0.5$, which is the usual case in applications according to Coles (2001b), and assumed also here, the maximum likelihood estimators can be applied without technical problems.

Under the assumptions made regarding ξ and $y(i)$, the distribution of the maximum likelihood estimators $(\widehat{\mu},\widehat{\sigma},\widehat{\xi})$ approaches with $m\to\infty$ multivariate normality. The covariance matrix is given by the inverse of the Fisher expected information matrix, evaluated at the maximum likelihood estimate (Coles 2001b). The elements of the latter matrix are (Prescott and Walden 1980):

$$\begin{aligned}
E\left[-\frac{\partial^2\ln(L)}{\partial\mu^2}\right] &= \frac{m}{\sigma^2}p,\\
E\left[-\frac{\partial^2\ln(L)}{\partial\sigma^2}\right] &= \frac{m}{\sigma^2\xi^2}\left[1-2\,\Gamma(2+\xi)+p\right],\\
E\left[-\frac{\partial^2\ln(L)}{\partial\xi^2}\right] &= \frac{m}{\xi^2}\left[\pi^2/6+(1-\gamma+1/\xi)^2-2q/\xi+p/\xi^2\right],\\
E\left[-\frac{\partial^2\ln(L)}{\partial\mu\,\partial\sigma}\right] &= \frac{m}{\sigma^2\xi}\left[\Gamma(2+\xi)-p\right], \qquad (6.8)\\
E\left[-\frac{\partial^2\ln(L)}{\partial\mu\,\partial\xi}\right] &= -\frac{m}{\sigma\xi}\left(q-p/\xi\right),\\
E\left[-\frac{\partial^2\ln(L)}{\partial\sigma\,\partial\xi}\right] &= -\frac{m}{\sigma\xi^2}\bigl\{1-\gamma+\left[1-\Gamma(2+\xi)\right]/\xi-q+p/\xi\bigr\},
\end{aligned}$$

where

$$\begin{aligned}
p &= (1+\xi)^2\ \Gamma(1+2\xi),\\
q &= \Gamma(2+\xi)\left[\Psi(1+\xi)+(1+\xi)/\xi\right],
\end{aligned} \qquad (6.9)$$

the constant $\gamma \approx 0.5772157$ is Euler's and $\Psi(\cdot)$ is the digamma function (Section 6.6). Classical CIs for the maximum likelihood estimates follow immediately from the covariance matrix and the percentage points of the normal distribution (Section 3.9).

6.2.2 Generalized Pareto distribution

The generalized Pareto (GP) distribution is suitable for analysing POT extremes. The relation of the GP to the GEV distribution is as follows. If the data generating process $X(i)$ and the extremes selection satisfy the assumptions, such that the block extremes have (approximately) a GEV distribution, then the POT extremes have (approximately) a GP distribution (Leadbetter et al. 1983). This is illustrated by Fig. 6.2.

6.2.2.1 Model

The GP distribution function is given by

$$F_{\mathrm{GP}}(x_{\mathrm{out}}) = \begin{cases} 1 - \Big\{ 1 + \xi \left(x_{\mathrm{out}} - u\right) / \big[\sigma + \xi\left(u - \mu\right)\big] \Big\}^{-1/\xi} & (\xi \neq 0), \\ 1 - \exp\Big[- \left(x_{\mathrm{out}} - u\right) / \sigma \Big] & (\xi = 0), \end{cases} \tag{6.10}$$

where $\sigma > 0$, $x_{\mathrm{out}} > u$, $\{1 + \xi\left(x_{\mathrm{out}} - u\right)/[\sigma + \xi\left(u - \mu\right)]\} > 0$, $-\infty < \xi < \infty$ and u is large (compared with the centre of location of the distribution of $X(i)$). Notably, the GP parameter shape, ξ, is the same as for the GEV distribution (Hosking and Wallis 1987). Again, the tail behaviour of the GP distribution is determined by ξ. If $\xi < 0$, then $F_{\mathrm{GP}}(x_{\mathrm{out}})$ has an upper bound of $x_{\mathrm{out}} = -\sigma/\xi + \mu$, if $\xi \geq 0$, then the GP distribution has no upper bound.

6.2.2.2 Maximum likelihood estimation

Analogously to Section 6.2.1.2, assume that the approximation is perfect and the POT data $\{x_{\mathrm{out}}(j)\}_{j=1}^{m}$ do come from a GP distribution (Eq. 6.10) and that $\xi \neq 0$. The log-likelihood function to be maximized is then (Coles 2001b)

$$\ln\left[L(\widetilde{\sigma}, \xi)\right] = -m \ln\left(\widetilde{\sigma}\right) - \left(1 + 1/\xi\right) \sum_{j=1}^{m} \ln\big[y(j)\big], \tag{6.11}$$

where

$$y(j) = 1 + \xi \left[\frac{x_{\mathrm{out}}(j) - u}{\widetilde{\sigma}} \right] \tag{6.12}$$

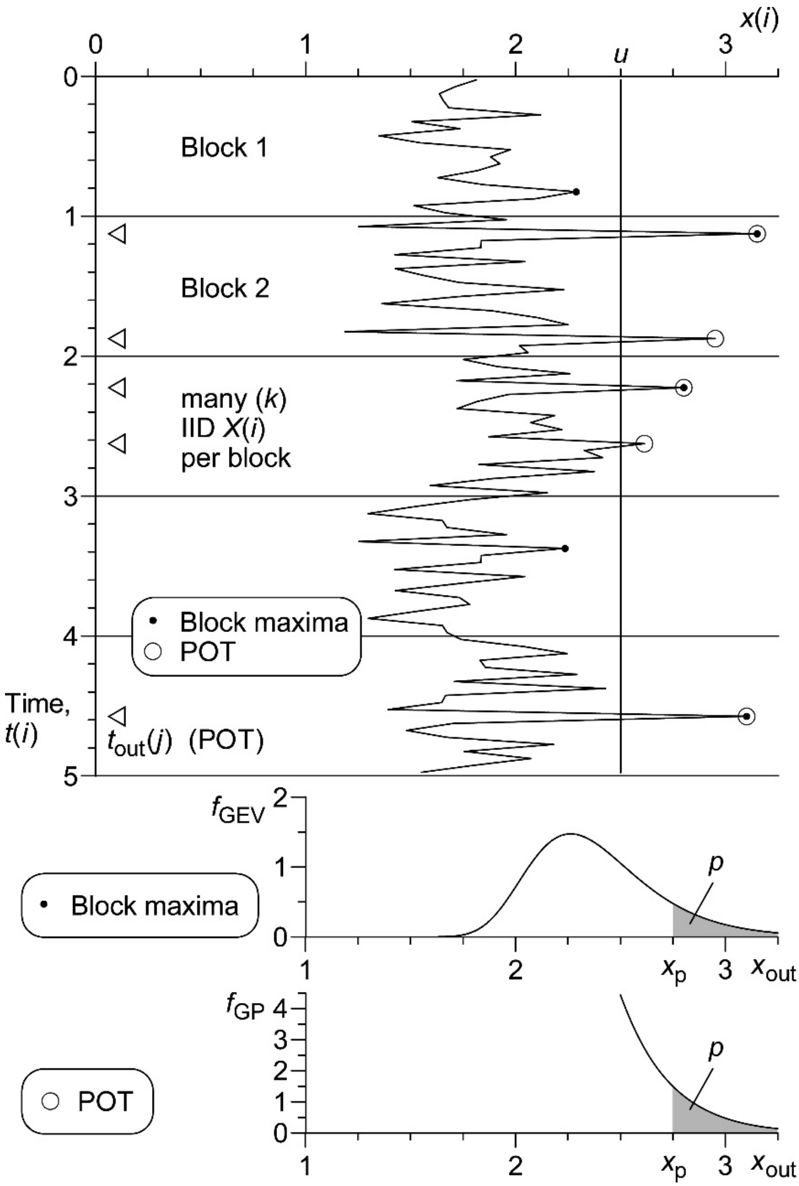


Figure 6.2. Block maxima, POT data, GEV and GP distributions. The basic assumption is that many (above, say, 100) independent and identically distributed (IID) random processes $X(i)$ contribute to each block. The threshold is denoted as u. The extremes, $X_{\text{out}}(j)$, have a GEV distribution (block maxima) or a GP distribution (POT). Shown are density functions, $f_{\text{GEV}}(x_{\text{out}})$ and $f_{\text{GP}}(x_{\text{out}})$; the related distribution functions ($F = \int f$) are given by Eqs. (6.5) and (6.10). The tail probability, or risk, for $X_{\text{out}} > x_p = 2.75$ (*shaded areas*) is $p = 11\%$ (GEV) and 31% (GP), respectively. x_p is the return level, $1/p$ the return period (in time units).

and

$$\widetilde{\sigma} = \sigma + \xi\,(u - \mu)\,. \tag{6.13}$$

Required is also that $y(j) > 0\ \forall j$. The convention of writing $\widetilde{\sigma}$ leads to only two parameters to be estimated (Coles 2001b).

Standard likelihood theory leads, analogously to Section 6.2.1.2, to classical CIs for the estimated parameters. The covariance matrix for the maximum likelihood estimators of $(\widetilde{\sigma}, \xi)$ on the process level is, for the usual case $\xi > -0.5$, given by (Davison and Smith 1990)

$$\frac{1+\xi}{m}\begin{bmatrix} 2\widetilde{\sigma}^2 & \widetilde{\sigma} \\ \widetilde{\sigma} & 1+\xi \end{bmatrix}. \tag{6.14}$$

In practice, on the sample level, the estimator values are plugged in. For example, $\widehat{\mathrm{se}}_{\widehat{\xi}} = (1+\widehat{\xi})/m^{1/2}$, from which CIs can be calculated using percentage points of the normal distribution.

6.2.2.3 Model suitability

Several conditions regarding the data have to be fulfilled to derive (Leadbetter et al. 1983) the GP distribution.

1. The process $X(i)$ generating the time series is serially independent.
2. The process $X(i)$ is stationary, the distributional shape of $X(i)$ does not change with time, $T(i)$.

These first two are the IID conditions.

3. The extremes or outliers are POT data, with the threshold, u, being large.

In addition to those three, the regularity condition,

4. the parameter ξ is greater than -0.5,

leads to obtainable maximum likelihood estimators (Section 6.2.2.2) with asymptotic ($m \to \infty$) properties, such as the covariance matrix, Eq. (6.14).

In practice the question is not whether preceding conditions are fulfilled but rather how strongly they are violated.

Precautionary measures employed during the data selection procedure (Section 6.1) can reduce the degree of violation and enhance the applicability of the GP model. Regarding serial independence, this can be achieved by taking into account the persistence properties of $X(i)$.

Regarding stationarity of location (and scale) of $X(i)$, by using a time-dependent threshold (Section 6.1.2) it is possible to correct for some nonstationarity. The alternative would be usage of a nonstationary GP model (similarly as in Section 6.3.1). Violations of the stationarity assumption may also be detected by adopting a point-process approach and analysing $\{t_{\text{out}}(j)\}_{j=1}^{m}$, as is done in Section 6.3.2. Regarding large u, it is necessary to recognize that, on the other hand, a lower u means a higher m and therefore smaller statistical uncertainties of estimated parameters—a typical dilemma between systematic and statistical errors that can be tackled by analysing a range of thresholds, u, and studying the sensitivity of the results. Regarding $\xi > -0.5$, this condition is said to be less a problem in practice, but nevertheless it should be tested.

In addition to the precautionary measures, it is helpful to assess the suitability of the GP model by analysis of diagnostic plots. Such analyses are presented by textbooks such as Coles (2001b: Section 4.4 therein) and here in the example of Elbe summer floods (Section 6.2.4). Statistical tests for model suitability augment graphical tools. For example, Van Montfort and Witter (1985) present a test for $\xi = 0$ in the GP model.

The question of model suitability applies also to fitted GEV distributions. Similar or same methods as for the GP distribution are applied. Instead of sensitivity studies of the threshold, the dependence on block length selection (k) is analysed.

6.2.2.4 Return period

Consider some large value, x_p, for a positive extreme (maximum) or outlier. This defines the tail probability, p, as $p = \int_{x_p}^{\infty} f_{\text{GP}}(x_{\text{out}}) dx_{\text{out}} = 1 - F_{\text{GP}}(x_p)$. The function $f_{\text{GP}}(x_{\text{out}})$ is the density function of the GP distribution. The return period has the numerical value $1/p$, its units are the same as of the original time values. The return period is approximately the expected time span required for observing one extreme event, X_{out}, in excess of x_p. The parameter x_p is called return level. (Strictly speaking, x_p is defined as the level that is exceeded once with probability p in one time unit.)

The generalization to negative extremes (minima) is straightforward. Obviously, the concept is applicable also to other distributions such as the GEV. Figure 6.2 illustrates cases with return level $x_p = 2.75$ and tail probabilities $p = 11\%$ (GEV) or 31% (GP). If the time units were years, then the return periods associated with x_p were approximately 9 years (GEV) or 3 years (GP), and an observation $x_{\text{out}} > x_p$ would be called a 9-year (GEV) or a 3-year (GP) event.

We define the term "risk" as tail probability, p, following the Encyclopedia of statistical sciences (Gardenier and Gardenier 1988). Because a variety of application fields of risk analysis exist, such as actuarial sciences, econometrics and climatology, many risk definitions are in usage; Thywissen (2006) lists 22, although not completely mutually exclusive, definitions currently employed. The definition via the probability has the advantage that this is a fundamental, real number, from which the other parameters of interest, for example, the expected economic loss, can be derived.

6.2.2.5 Probability weighted moment estimation

The method of probability weighted moments (PWMs) (Greenwood et al. 1979) offers an alternative to maximum likelihood estimation of the GP distribution parameters. In general, the PWM of a continuous random variable X is the quantity $M_{q,r,s} = E[X^q\{F(x)\}^r\{1 - F(x)\}^s]$. For the GP distribution, it is convenient (Hosking and Wallis 1987) to use the parameters

$$\alpha_s = M_{1,0,s} = E[X\{1 - F_{\text{GP}}(x_{\text{out}})\}^s] = \widetilde{\sigma}\,/[(s+1)\,(s+1-\xi)]\,, \tag{6.15}$$

which exist for $\xi < 1$. The GP parameters expressed in terms of PWMs are

$$\widetilde{\sigma} = 2\,\alpha_0\,\alpha_1\,/(\alpha_0 - 2\,\alpha_1) \tag{6.16}$$

and

$$\xi = 2 - \alpha_0\,/(\alpha_0 - 2\,\alpha_1)\,. \tag{6.17}$$

The PWM method plugs in estimates for α_0 and α_1 into Eqs. (6.16) and (6.17) to estimate the GP parameters. For example (Landwehr et al. 1979),

$$\widehat{\alpha}_s = m^{-1}\sum_{j=1}^{m}\frac{(m-j)(m-j-1)\cdots(m-j-s+1)}{(m-1)(m-2)\cdots(m-s)}\,x_{\text{out, sort}}(j), \tag{6.18}$$

where $\{x_{\text{out, sort}}(j)\}_{j=1}^{m}$ is the sample sorted in ascending order. (Hosking and Wallis (1987) give a second α_s estimator.)

Asymptotically, for $m \to \infty$, and under the condition $\xi < 0.5$, the PWM estimators for the GP parameters $(\widetilde{\sigma}, \xi)$ have a normal distribution and, on the process level, following covariance matrix (Hosking et al. 1985; Hosking and Wallis 1987):

$$\frac{1}{m(1-2\xi)(3-2\xi)} \times \begin{bmatrix} \widetilde{\sigma}^2\left(7-18\xi+11\xi^2-2\xi^3\right) & \widetilde{\sigma}\left(2-\xi\right)\left(2-6\xi+7\xi^2-2\xi^3\right) \\ \widetilde{\sigma}\left(2-\xi\right)\left(2-6\xi+7\xi^2-2\xi^3\right) & \left(1-\xi\right)\left(2+\xi\right)^2\left(1-\xi+2\xi^2\right) \end{bmatrix}. \tag{6.19}$$

In practice, the estimator values are plugged in. CIs for the PWM estimates of the GP parameters, and also for related quantities such as return levels, follow directly from the covariance matrix (Hosking and Wallis 1987).

The PWM method can also be applied to estimating parameters and related quantities of the GEV distribution (Hosking et al. 1985; Lu and Stedinger 1992).

A method closely related to PWM is the estimation with so-called L-moments (Hosking 1990; Hosking and Wallis 1997).

6.2.3 Bootstrap confidence intervals

The classical CIs for GEV or GP parameter estimates, calculated from the covariance matrices, are not exact because the sample size of the extremes, m, is not infinite. In addition to that, violations of the underlying model assumptions (Section 6.2.2.3) may increase the inexactness. In general, such situations favour the bootstrap approach to deliver more accurate results. However, in the case of CI construction for GEV or GP parameters and related quantities such as quantiles and return periods, bootstrap resampling may not always be preferable.

The problem with the nonparametric bootstrap resampling (Section 3.3.1) is that the distribution of the bootstrap replications does not uniformly converge with m to the true distribution when the parameter of interest is a quantile, see Bickel and Freedman (1981), Davison and Smith (1990: p. 440 therein) and Angus (1993). The alternative resampling, parametric surrogate data simulation (Section 3.3.3), has been found in Monte Carlo experiments (Caers et al. 1999a; Kyselý 2008) to give CIs with acceptable accuracies—better than from nonparametric bootstrap resampling. The caveat against the method of parametric simulation, however, is that it prescribes a certain distribution model (GEV, GP) to draw data from and assumes its suitability. In practice, where $m < \infty$ and the limiting model distribution has been only approximately approached, there comes additional uncertainty, which should widen the CIs obtained from parametric simulation. It is difficult to quantify how much wider accurate CIs would be.

6.2.4 Example: Elbe summer floods, 1852–2002

The example of the Elbe summer floods (Fig. 6.3) explores the suitability of the GEV model for hydrological time series. Prior to the analysis, the stationarity of the data generating process was positively tested by methods explained in Section 6.3.

The sensitivity plots for estimated shape parameter and 100-year return level (Fig. 6.3b, c) show some variations with the block length; a value of 2 months for taking block maxima was assessed as appropriate. For all block lengths tested, the regularity condition $\xi > -0.5$ is very likely fulfilled.

In light of the variations with the block length, the possible violation of the independence assumption of $X(i)$ by the long-memory property of runoff time series (Section 2.5.3) and the limited size of $m = 444$, some model error has to be considered. This would add to the errors deduced from the estimated covariance matrix for the maximum likelihood estimation (Fig. 6.3). However, the density plot (Fig. 6.3d) and the probability plot (Fig. 6.3e) reveal agreement between data and fitted model and do not allow to reject the GEV assumption.

The Elbe flood in August 2002 was a devastating event causing 36 deaths and over 15 billion EUR economic damages (Mueller 2003; Sercl and Stehlik 2003). What is the risk, p, of an event of this or larger size? Although the measured water level at gauge station Dresden was relatively accurately determined, the associated maximum runoff value of $Q = 4700 \text{ m}^3\text{s}^{-1}$ for the August 2002 flood (Engel et al. 2002) is less certain; the true value may have been larger. The reason preventing a direct Q determination was that water velocity measurements over the entire river cross section were not possible (Engel H 2002, personal communication). This left the stage–runoff calibration inaccurate at such high values.

Bearing the caveat regarding data accuracy and the model errors in mind, the inspection of Fig. 6.3f leaves the impression that the Elbe flooding in August 2002 was clearly a larger event than a 100-year flood—in Dresden. Statements about return periods of 200 years and more are likely rather inaccurate. Analyses of the event in August 2002 performed for several stations along the Elbe, shed more light on the flood risk (Engel et al. 2002; Mudelsee et al. 2004).

We remark that it is mandatory to distinguish between winter and summer floods because they have different meteorological–hydrological causes (see background material). The winter floods of the Elbe, for example, do not share the stationarity property of the summer floods for the same interval (past $\sim 150\,\text{a}$).

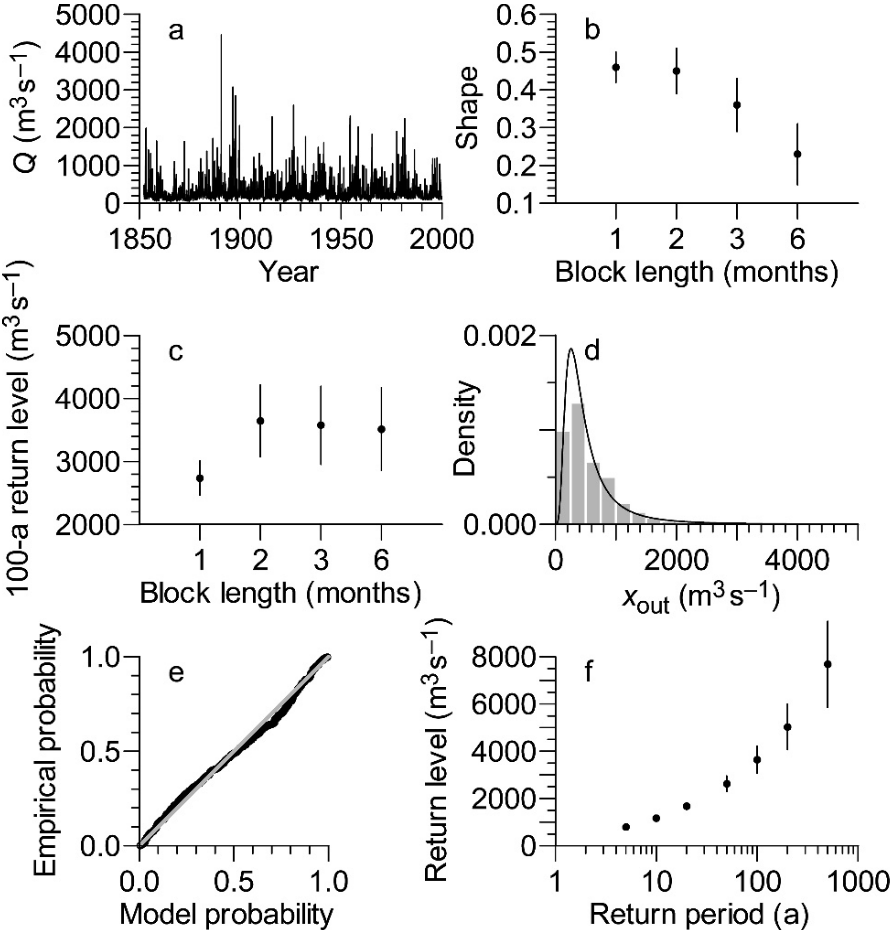


Figure 6.3. Elbe summer floods 1852–1999, GEV estimation applied to block maxima. **a** Daily runoff, $x(i)$, at station Dresden for the hydrological summer, from May to October; $n = 27{,}232$. Data from Global Runoff Data Centre, Koblenz, Germany. **b** Maximum likelihood estimates of the shape parameter, ξ, of a GEV distribution in dependence on block length; standard errors (*vertical lines*) from the estimated covariance matrix. (The block length of 2 months corresponds to $m = 444$.) **c** 100-year return level, x_p for $p = 0.01$, in hydrology also denoted as HQ_{100}, in dependence on block length; standard errors from error propagation (Section 6.5) using the estimated covariance matrix. **d** Estimated GEV density function (*solid line*) and histogram estimate (Section 1.6) of empirical density for block length 2 months. **e** Empirical probability, $j/(m+1)$ for $j = 1, \ldots, m$, against model probability, $\exp\{-[1 + \widehat{\xi}(x_{\text{out, sort}}(j) - \widehat{\mu}) / \widehat{\sigma}]^{-1/\widehat{\xi}}\}$, shown as (*closely spaced*) *dots*; 1:1 line (*grey*). **f** Return level with standard errors for fitted GEV model in dependence on return period.

6.2.5 Persistence

Climate processes, $X(i)$, often show persistence (Chapter 2). This violates the independence assumption made for deriving the GEV and GP distribution models for extremes. For short-memory persistence such as AR(1) processes, typical for climate (Section 2.1), however, this violation does not invalidate GEV or GP estimation when the sampling of the extremes (Section 6.1) is done appropriately. Even for certain types of long-memory persistence, GEV or GP estimation may still be applied. Our exposition follows closely that of Coles (2001b: Chapter 5 therein).

6.2.5.1 Condition $D(u_n)$

A stationary process $\{X(i)\}_{i=1}^{n}$ satisfies the condition $D(u_n)$ if for all $1 \leq i_1 < \cdots < i_p < j_1 < \cdots < j_q \leq n$ with $j_1 - i_p > l > 0$,

$$\begin{aligned} \Big| \text{prob}\,\{X(i_1) \leq u_n, \ldots, X(i_p) \leq u_n,\ X(j_1) \leq u_n, \ldots, X(j_q) \leq u_n\} \\ -\ \text{prob}\,\{X(i_1) \leq u_n, \ldots, X(i_p) \leq u_n\} \\ \times\ \text{prob}\,\{X(j_1) \leq u_n, \ldots, X(j_q) \leq u_n\} \Big| \leq \alpha(n, l), \end{aligned} \tag{6.20}$$

where the sequence $\alpha(n, l_n) \to 0$ and $l_n/n \to 0$ as $n \to \infty$. An independent process $X(i)$ has zero difference, α, in probabilities. The condition $D(u_n)$ generalizes this concept. The integer l plays a similar role as the persistence time.

6.2.5.2 Extremal index

Consider the stationary process $\{X(i)\}_{i=1}^{n}$ with persistence and the related process $\{X^*(i)\}_{i=1}^{n}$ without persistence (but identical data distributions). As explained in Section 6.2.1.1, under suitable conditions the distribution of the block maxima of $\{X^*(i)\}_{i=1}^{n}$ approaches a GEV distribution. Denote the distribution function as $F_{\text{GEV},1}(x_{\text{out}})$. It may be shown (Leadbetter et al. 1983) that under the same conditions also the distribution of the block maxima of $\{X(i)\}_{i=1}^{n}$ approaches a GEV distribution, with other parameters μ and σ (but with identical ξ). Denote this distribution function as $F_{\text{GEV},2}(x_{\text{out}})$. It is (Leadbetter et al. 1983)

$$F_{\text{GEV},2}(x_{\text{out}}) = \left[F_{\text{GEV},1}(x_{\text{out}})\right]^{\theta}, \tag{6.21}$$

where $0 < \theta \leq 1$. The parameter θ linking the dependence and independence cases is called extremal index.

Equation (6.21) has considerable practical consequences because it allows to apply the GEV and GP estimation methods also to data from short-memory processes. A caveat here is that the number of independent observations is reduced. Coles (2001b: Section 5.3.1 therein) gives the number of $n\theta$ as effective data size with respect to the quality of the GEV approximation.

For fitting a GP distribution to threshold extremes from short-memory processes, Coles (2001b: Section 5.3.3 therein) notes the technique of declustering. This takes into account that under persistence the extremes tend to occur in clusters (consecutive times). Within a cluster, only the maximum excess over a threshold is retained for GP estimation. Declustering is equivalent to POT data selection (Section 6.1.2) with an imposed secondary selection criterion. To prohibit the information loss associated with declustering, it may be worth instead to consider to retain all POT values and account for the persistence by either modelling it or adjusting the covariance matrix (see background material).

6.2.5.3 Long memory

Even if $X(i)$ is a long-memory process (Section 2.4.1), the GEV or GP model may be applicable. Smith (1989: pp. 392–393 therein) remarks that if $X(i)$ has a Gaussian distributional shape and $\rho(n)\log(n) \to 0$ for $n \to \infty$, then the long-range dependence "does not matter," referring to a paper by Berman (1964). The autocorrelation function $\rho(h)$ for an ARFIMA process does indeed fulfill the condition, and a suitable transformation of $X(i)$ may yield approximately a Gaussian shape. See the example of river runoff (Section 2.5.3).

The major problem from long memory could be that the number of independent observations is reduced—to a stronger degree than for short memory. This makes the GEV or GP approximation less accurate than in the no-memory or short-memory cases.

6.2.6 Remark: tail estimation

In practice, when analysing a sample of extremes $\{t_{\text{out}}(j), x_{\text{out}}(j)\}_{j=1}^{m}$ obtained from a climate time series $\{t(i), x(i)\}_{i=1}^{n}$, the questions often regard the tails of the distribution of X_{out}. It is here at the upper values, where the "climate risk" is located, where tail probabilities p, return periods $1/p$ (in time units) and return levels x_p are often associated with events of high socioeconomical relevance (Fig. 6.2). For example, authorities dealing with flood protection may be interested in HQ_{1000}, the 1000-year return level of runoff at a certain river station.

The requirement of accurate methods of tail estimation was also noticed in the Earth Sciences literature, for example, by Dargahi-Noubary

(1989). He argued in favour of the POT–GP and against the block extremes–GEV approach because of information wastage caused by the latter. Notably, he also remarked that methods suited for estimating distribution parameters need not be optimal for estimating tail probabilities. An example in the Earth Sciences literature, where accurate tail estimation was an objective, was given by Caers et al. (1999b), who applied parametric bootstrap resampling to earthquake, diamond (an extreme with a "positive" connotation) and impact crater size data.

Smith (1987) proposed to estimate the tail probability of a GP distribution (Fig. 6.2) as follows:

$$\widehat{p} = m\, n^{-1} \left[1 + \widehat{\xi}\,(x_{\mathrm{out}} - u) \middle/ \widehat{\widetilde{\sigma}}\right]. \tag{6.22}$$

Herein, $\widehat{\xi}$ is the estimate of the shape parameter and $\widehat{\widetilde{\sigma}}$ the estimate of the transformed scale parameter (Eq. 6.13). This tail probability estimator is defined for $x_{\mathrm{out}} > u$ (if $\widehat{\xi} > 0$) or $0 < (x_{\mathrm{out}} - u) < -\widehat{\widetilde{\sigma}}/\widehat{\xi}$ (if $\widehat{\xi} < 0$). Smith (1987) showed by means of theoretical studies of its asymptotic properties that this estimator has "often" a better performance than a previous tail estimator suggested by Hill (1975).

In Eq. (6.22), the expression $m\, n^{-1}$ is an estimate of the time-constant rate of occurrence of an extreme event within a time unit. The term within square brackets is the tail probability conditional on that an extreme occurred. Eq. (6.22) can therefore be seen (Kallache M 2009, personal communication) as a manifestation of the hybrid Poisson–extreme value distribution approach (Section 6.3.3) in the stationary setting, as a counterpart to Eq. (6.42).

Smith's (1987) estimator (Eq. 6.22) applies to "within-sample" thresholds. If $u > \max(\{x_{\mathrm{out}}(j)\}_{j=1}^{m})$ for positive extremes, then $m = 0$ and $\widehat{p} = 0$, which is not a helpful estimation. When confronted with the task to estimate such "out-of-sample" probabilities or quantiles, other methods (Hall and Weissman 1997; El-Aroui and Diebolt 2002; Ferreira et al. 2003) can be tried. These methods are based on estimating a quantile "within" ($< \max(\{x_{\mathrm{out}}(j)\}_{j=1}^{m})$ for positive extremes) and transforming it to "outside" ($> \max(\{x_{\mathrm{out}}(j)\}_{j=1}^{m})$). A related task is to make long-range predictions of extremes, based on extrapolation; for that purpose Hall et al. (2002) found good coverage performance of calibrated bootstrap CIs (Section 3.8). For the analysis of climate risk, however, long-range predictions based on "out-of-sample" estimations bear the danger of considerable errors caused by nonstationarities. The assumption made so far, namely that the distribution of $X(i)$ does not change with time, is rather strong. It may, for example, be questionable to estimate an HQ_{1000} based on 150 years of data.

6.2.7 Remark: optimal estimation

It is helpful for the analyst to recall at this point what he or she is doing. Given is a data sample $\{x(i)\}_{i=1}^{n}$ and a question about the system the data document, the climate system. Often, instead of a single question there is a whole complex of questions. Sometimes the questions existed before the data, sometimes the data were earlier, and often questions and data have evolved together, in loops. In quantitative climatology, the questions can be translated into a parameter, θ, or a set of parameters, which need to be estimated using the data.

The estimator, $\widehat{\theta}$, appears as a third actor besides data and parameter. For more difficult questions, the construction of an estimator is not straightforward but a work that requires creativity. The fourth is the confidence interval, $\mathrm{CI}_{\widehat{\theta},1-2\alpha}$. Also CI construction is not straightforward, there may exist bootstrap versions, and there may exist classical versions.

The aim of the estimation is, of course, to come close to the truth with the parameter estimate produced by data and estimator. This can be judged by various measures: bias, standard error, RMSE, CI length, CI coverage accuracy, robustness, and so forth. It becomes apparent that an optimal estimation requires to cycle through more, nested loops. In dependence on data and parameter, the methods of estimation and CI construction are selected that have the desired properties. This selection may feed back and lead to other parameters to be tried, refined questions to be asked.

As an example with regard to the content of this chapter: asking for the optimal estimators of the parameters of a GEV distribution of runoff maxima is not the same as asking for the optimal estimator of a quantile (like HQ_{100}). Knowledge about the properties of a combination of data, estimator and CI requires usually ($n < \infty$ and not too simple question) evidence from Monte Carlo simulations because of theoretical intractability. It may be expected that existing Monte Carlo evidence from previous studies cannot always be applied or generalized to the combination at hand. The nested route towards an optimal estimation can, therefore, require the climate analyst to carry out new Monte Carlo experiments.

6.3 Nonstationary models

In the extreme value analysis of climate time series, however, it is more realistic to assume time-dependent models: with climate changes also risk changes may come. Already before the contribution of IPCC–WG I to the Fourth Assessment Report (Solomon et al. 2007) appeared, the nonstationary analysis of climate extremes had been an active research field. After the report, the developments in this field have received growing attention, also by lay people and the media.

One may argue that by estimating time-dependent trend and time-dependent variability (Section 6.1.2), taking extremes from the scaled data, $\left\{[X(i) - \widehat{X}_{\text{trend}}(i)]/\widehat{S}(i)\right\}_{i=1}^{n}$, and fitting a stationary model, the nonstationarity is taken into account, but such an analysis could miss trends in the tail behaviour and therefore be insufficient.

One route towards a more complete analysis is to retain the extreme value distribution models and introduce time-dependence into their parameters. We present the time-dependent GEV distribution, where the mean, scale and shape are allowed to exhibit trends described by parameters. The other route is to think of the time points when an extreme occurred, $\{t_{\text{out}}(j)\}_{j=1}^{m}$, as a realization of a nonstationary model of the occurrence of an event (an inhomogeneous Poisson process). We show estimation of the time-dependent occurrence rate by means of a nonparametric technique (kernel estimation).

6.3.1 Time-dependent Generalized Extreme Value distribution

The nonstationary GEV model is the same as the stationary (Eq. 6.5), except that allowed now are time-dependences in location, scale and shape. A simple form of dependence is:

$$\mu(T_{\text{out}}) = \beta_0 + \beta_1 T_{\text{out}}, \tag{6.23}$$

$$\sigma(T_{\text{out}}) = \exp\left(\gamma_0 + \gamma_1 T_{\text{out}}\right), \tag{6.24}$$

$$\xi(T_{\text{out}}) = \delta_0 + \delta_1 T_{\text{out}}. \tag{6.25}$$

(The exponential function ensures a positive scale parameter.) The log-likelihood function (Coles 2001b) depends now on six parameters; on the sample level,

$$\ln\left[L(\mu,\sigma,\xi)\right] = \ln\left[L(\beta_0,\beta_1,\gamma_0,\gamma_1,\delta_0,\delta_1)\right] = -\sum_{j=1}^{m}\Big\{\ln\big[\sigma(t_{\text{out}}(j))\big] + \big[1 + 1/\xi(t_{\text{out}}(j))\big]\,\ln\big[y(j)\big] + \big[y(j)\big]^{-1/\xi(t_{\text{out}}(j))}\Big\}, \tag{6.26}$$

where

$$y(j) = 1 + \xi(t_{\text{out}}(j))\left\{\frac{x_{\text{out}}(j) - \mu(t_{\text{out}}(j))}{\sigma(t_{\text{out}}(j))}\right\}. \tag{6.27}$$

It is assumed that $\xi(t_{\text{out}}(j)) \neq 0$; for the j for which this is not the case, the equivalent to the "Gumbel likelihood" (Section 6.6) has to be used. The additional condition is that $y(j) > 0 \; \forall j$.

In principle, the log-likelihood function is maximized using numerical techniques. The point in the parameter space (6-dimensional in the case here) defines the maximum likelihood estimate. Approximate standard errors and confidence intervals follow from the information matrix, analogously to Section 6.2.1.2. The parameter estimates define the time-dependent GEV distribution, from which in turn time-dependent tail probabilities (risks), return periods and return levels can be calculated with error bars.

It appears, however, that in practice the first step, maximization can cause numerical problems (Smith 1989). Specifically, when $\xi(T_{\text{out}})$ is close to zero, the selection of the appropriate likelihood function may be difficult. Further numerical challenges (start values, stopping rule, local versus global maxima) may be encountered when many parameters are used to model the time-dependence, or when the time-functions contain jumps. This should not be interpreted as a criticism of the maximum likelihood approach but rather as an indication to employ suitable techniques to difficult numerical problems. The climatological applications of the time-dependent GEV distribution (Section 6.5) seem to have gained their successes from limiting the number of parameters, using simple time-dependences and constraining the values of the shape parameter, ξ. There is no hindrance to applying maximum likelihood estimation with trends in the parameters to the GP model (background material).

A second critical point is the inevitable extrapolation from within the observation time interval to outside (in practice: the future). The parameter values or the functional form of the nonstationary GEV model may change outside, and this may then bias the estimations severely.

6.3.2 Inhomogeneous Poisson process

6.3.2.1 Model

Consider the time points $\{T_{\text{out}}(j)\}_{j=1}^{m}$ when an extreme has occurred. The points may have resulted from POT data or block extremes, or those event times may constitute the only information left about the occurrence of an extreme or outlier. Consider the number of extreme events to be described by a discrete random variable, M. Let the number of extremes at continuous time T be described by the random process $M(T)$. Its realization consists of step functions with "unit jumps" at $T_{\text{out}}(j)$. The process $M(T)$ is called a point process (see Karr (1986) for a complete definition).

Consider the incremental process, $dM(T) = M(T + \delta T) - M(T)$, which represents the number of events in the time interval $[T; T + \delta T]$. Let δT be arbitrarily small, so that not two or more events occur within the interval and $dM(T)$ takes only two values, namely,

$$dM(T) = \begin{cases} 1 & \text{with probability } \lambda \cdot \delta T, \\ 0 & \text{with probability } 1 - \lambda \cdot \delta T, \end{cases} \tag{6.28}$$

where $\lambda \geq 0$ is a constant. Then

$$E\left[dM(T)\right] = \lambda \cdot \delta T \tag{6.29}$$

and

$$VAR\left[dM(T)\right] = \lambda \cdot \delta T. \tag{6.30}$$

Assume further that the events occur independently of each other,

$$COV\left[dM(T_1), dM(T_2)\right] = 0 \text{ for } T_1 \neq T_2. \tag{6.31}$$

The point process $M(T)$ is then specified as a homogeneous Poisson process with occurrence rate parameter λ.

The parameter of interest for the analysis of climate extremes is λ. Its units are one over time units. It gives the probability per time interval that an extreme occurs. For studying nonstationarity and trends in climate risk, we now introduce time-dependence and denote the function $\lambda(T)$ as occurrence rate. The process is then denoted as inhomogeneous Poisson process (Cox and Lewis 1966).

6.3.2.2 Nonparametric occurrence rate estimation

The kernel approach (Diggle 1985) estimates the occurrence rate as

$$\widehat{\lambda}(T) = h^{-1} \sum_{j=1}^{m} K\big(\left[T - T_{\text{out}}(j)\right]/h\big), \tag{6.32}$$

where h is the bandwidth and K is the kernel function.

Consider for heuristic reasons the following primitive occurrence rate estimator. Divide the observation interval $[T(1); T(n)]$ in two halves of equal length, $H = [T(n) - T(1)]/2$. Let the number of events in the first and second halve be m_1 and m_2, respectively. Estimate $\lambda(T)$ in the first halve as m_1/H and in the second halve as m_2/H. This estimator corresponds to a uniform kernel ($K(y) = 1$ for $|y| \leq 1/2$ and $K(y) = 0$ otherwise) with bandwidth $h = H$ and merely two estimation time points.

Estimations of practical relevance employ therefore quasi-continuously distributed (many) estimation time points, T, a smooth kernel function, K, and a suitably selected bandwidth, h—in exact analogy to kernel density estimation and kernel smoothing (Sections 1.6 and 4.3.1; see also Diggle (1985) and Diggle and Marron (1988)). Bandwidth selection is treated in Section 6.3.2.4 and usage of a Gaussian kernel function, $K(y) = (2\pi)^{-1/2}\exp(-y^2/2)$, is motivated in the technical issues.

6.3.2.3 Boundary bias reduction

Usage of Eq. (6.32) may lead to bias in the form of underestimation of $\lambda(T)$ near the boundaries, $T = T(1)$ and $T = T(n)$, because of "missing data" outside of the observation interval. One option to reduce this bias is to let h decrease towards the boundaries, to use a "boundary kernel" (Gasser and Müller 1979). The other, adopted here, is to generate pseudodata (Cowling and Hall 1996) outside of $[T(1); T(n)]$ and estimate $\lambda(T)$ using a constant bandwidth and the original data augmented by the pseudodata:

$$\widehat{\lambda}(T) = h^{-1}\sum_{j=1}^{m^\dagger} K\big(\left[T - T^\dagger_{\text{out}}(j)\right]/h\big). \tag{6.33}$$

This is the equation on which the occurrence rate estimates in this chapter are based.

The original event data are $\{T_{\text{out}}(j)\}_{j=1}^{m}$. Let the (left) pseudodata for $T < T(1)$ be denoted as $\{T'_{\text{out}}(j)\}_{j=1}^{m'}$ and the (right) pseudodata for $T > T(n)$ as $\{T''_{\text{out}}(j)\}_{j=1}^{m''}$. Then the augmented set of event data,

$$\left\{T^\dagger_{\text{out}}(j)\right\}_{j=1}^{m^\dagger = m + m' + m''} = \{T_{\text{out}}(j)\}_{j=1}^{m} \cup \left\{T'_{\text{out}}(j)\right\}_{j=1}^{m'} \cup \left\{T''_{\text{out}}(j)\right\}_{j=1}^{m''}, \tag{6.34}$$

is the set union of original data, left and right pseudodata.

How can the pseudodata be generated? Cowling and Hall (1996) show the equivalence of pseudodata generation and extrapolation of the empirical distribution function of $\{T_{\text{out}}(j)\}_{j=1}^{m}$ and give rules how to generate the pseudodata. Consider the left boundary, the start of the observation interval, $T(1)$. The simplest rule is "reflection,"

$$T'_{\text{out}}(j') = T(1) - \left[T_{\text{out}}(j) - T(1)\right]. \tag{6.35}$$

Setting $j = 1$ gives the rightmost of the left pseudodata points. How many pseudodata points should be generated? Since the objective in this chapter is to estimate $\lambda(T)$ within $[T(1); T(n)]$, pseudodata coverage of a time interval extending to, say, $3\,h$ below $T(1)$ is sufficient. The "reflection" rule is analogously applied to produce right pseudodata, for $T > T(n)$. If the objective of the analysis is forecasting, an interval extending to beyond $T(n) + 3\,h$ should be covered, that is, more pseudodata be generated. The "reflection" rule corresponds to an extrapolation of the empirical distribution function with a constant rate.

Cowling and Hall (1996) give other rules, which may be applicable when the rate, $\lambda(T)$, is expected to change at the boundaries. Of particular relevance for climatological applications is when $T(n)$ is the present and a future upwards trend in climate risk may exist. We note the "two-point" rule,

$$T'_{\text{out}}(j') = T(1) - 9\,[T_{\text{out}}(j/3) - T(1)] + 2\,[T_{\text{out}}(j) - T(1)]\,, \tag{6.36}$$

where the fractional data $T_{\text{out}}(j/3)$ are determined by linear interpolation and the setting $T_{\text{out}}(0) \equiv T(1)$; analogously for the right pseudodata.

It is evident that pseudodata generation is a crucial step on the way to an improved occurrence rate estimate. As with any extrapolation method, care is required in the interpretation of the results. On the other hand, it is inevitable to make assumptions when analysing a problem. This applies not only to the statistical "extrapolability," but also to the actualism that is assumed when using physical climate models for future projections.

6.3.2.4 Bandwidth selection

Bandwidth (h) selection determines bias and variance properties of the occurrence rate estimator (Eq. 6.33) and is therefore the second crucial step. Brooks and Marron (1991) developed the cross-validation bandwidth selector for kernel occurrence rate estimation. This is the minimizer of

$$C(h) = \int\limits_{T(1)}^{T(n)} \left[\widehat{\lambda}(T)\right]^2 dT - 2 \sum_{j=1}^{m} \widehat{\lambda}_j\big(T_{\text{out}}(j)\big), \tag{6.37}$$

where

$$\widehat{\lambda}_j(T) = \sum_{k=1,\ k\neq j}^{m^\dagger} h^{-1}\, K\big(\big[T - T^\dagger_{\text{out}}(k)\big]\,/h\big) \tag{6.38}$$

is the delete-one estimate.

The cross-validated bandwidth can be seen as a compromise between small h (large variance and small bias of $\widehat{\lambda}$) and large h (small variance and large bias).

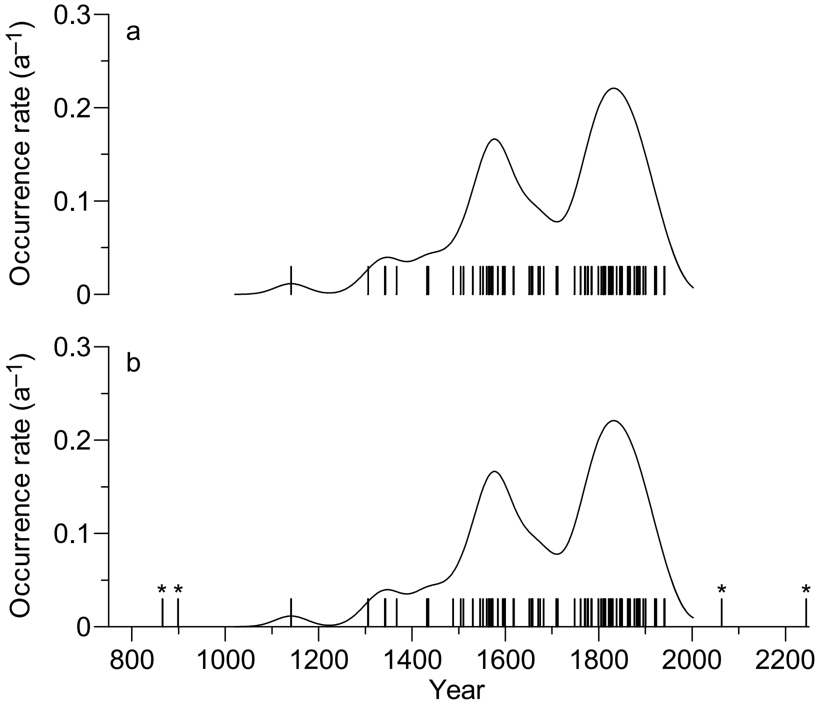


Figure 6.4. Elbe winter floods, pseudodata generation. The heavy events (magnitudes 2–3) are taken from the complete record (Fig. 1.1) and plotted (**a, b**) as *bars* ($m = 73$). The "twopoint" rule is used to generate four pseudodata points (**b,** *asterisks*) outside the observation interval. Occurrence rates are estimated with $h = 35\,\mathrm{a}$, and without (**a** $m^\dagger = 73$) or with (**b** $m^\dagger = 77$) pseudodata.

6.3.2.5 Example: Elbe winter floods (continued)

The number of heavy (magnitudes 2–3) floods of the Elbe in winter is $m = 73$. The first event was in 1141. However, the historical information back to 1021 was analysed (Mudelsee et al. 2003), and the observation interval is $[1021; 2002]$. Pseudodata generation (Fig. 6.4) uses the "twopoint" rule to take (climatic and other) trends in flood risk at the boundaries into account.

The cross-validation function (Fig. 6.5) has a minimum at $h = 41\,\mathrm{a}$. For suppressing potential extrapolation effects (Section 6.3.2.3) and further reducing the bias (Section 6.3.2.4) it may be advisable to undersmooth slightly. For this reason and for achieving consistency with results from other flood records (Elbe, summer; Oder, winter and summer), Mudelsee et al. (2003) set the analysis bandwith to $h = 35\,\mathrm{a}$. The estimated flood occurrence rate (Fig. 6.4) reveals—in the case of heavy

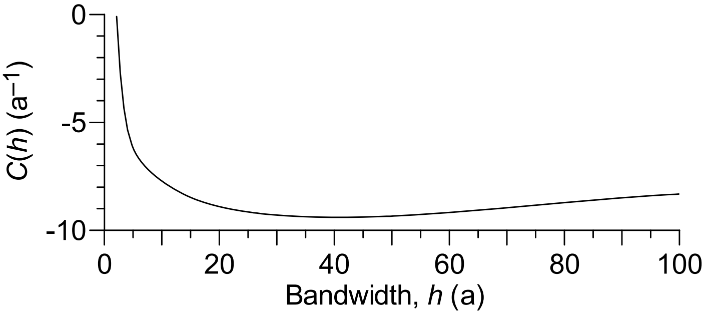


Figure 6.5. Elbe winter floods, cross-validation function, heavy events (magnitudes 2–3).

winter floods of the Elbe—little boundary bias. The reason is that the occurrence rate at the boundaries is rather low.

Bandwidth selection has large effects on flood occurrence rate estimation. Too strong undersmoothing with $h = 5$ a (Fig. 6.6a) allows too many variations. Within the bootstrap confidence band (Section 6.3.2.6), most of these wiggles are not significant (not shown). Too strong oversmoothing with $h = 100$ a (Fig. 6.6b) reduces the estimation variance but enhances the bias: too many significant variations in flood occurrence rate are smoothed away. The right amount of smoothing appears to be indicated by cross-validation; an only slight undersmoothing with $h = 35$ a (Fig. 6.6c) lets the significant variations appear. The example of the heavy Elbe winter floods is pursued further in Section 6.3.2.7.

6.3.2.6 Bootstrap confidence band

A measure of the uncertainty of $\widehat{\lambda}(T)$ (Eq. 6.33) is essential for interpreting results. For example, it might be asked if the low in $\widehat{\lambda}(T)$ at $T \approx 1700$ for the heavy winter floods of the Elbe (Fig. 6.6c) is real or the mere product of sampling variability. Cowling et al. (1996) devised bootstrap algorithms for constructing a confidence band around $\widehat{\lambda}(T)$; one is shown as Algorithm 6.1.

Step 2 of the algorithm, discretization of T, uses a large number, N_T, in the order of several hundred, to render a smooth estimate. For Step 4, alternative bootstrap methods, where also the size of the simulated set is a random variable, were tested by Cowling et al. (1996). Studentization (Step 8) draws advantage from the fact that the auxiliary variable $T_{\text{stud}}(T, b)$ is approximately pivotal (independent of T). Alternative CI

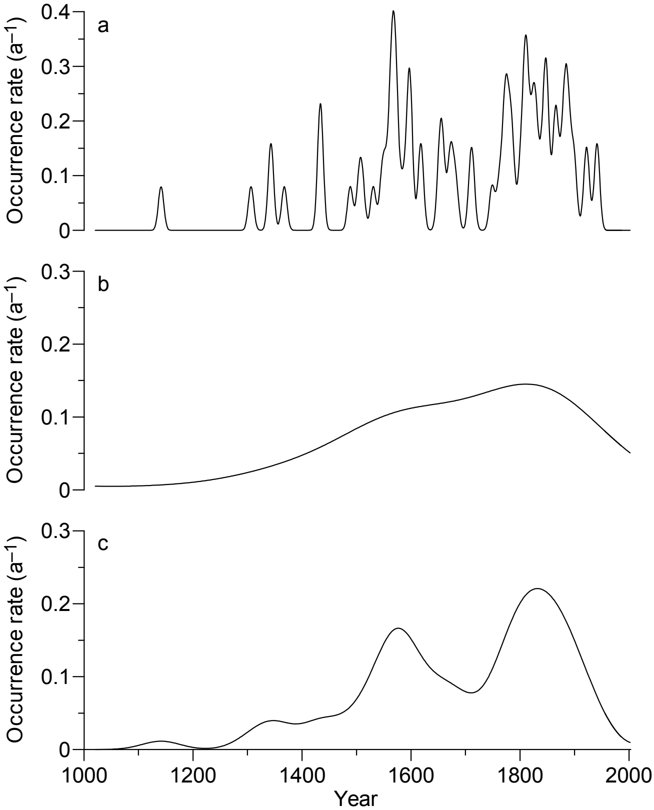


Figure 6.6. Elbe winter floods, bandwidth selection, heavy events (magnitudes 2–3). The occurrence rate is estimated with pseudodata ("twopoint" rule) and bandwidth $h = 5\,\mathrm{a}$ (**a**), $100\,\mathrm{a}$ (**b**) and $35\,\mathrm{a}$ (**c**).

construction methods (percentile) at this step were tested by Cowling et al. (1996). The resulting confidence band (Step 12) is a pointwise.

The coverage performance of the confidence band (Algorithm 6.1) was tested by means of Monte Carlo simulations (Cowling 1995; Cowling et al. 1996; Hall P 2008, personal communication). The prescribed $\lambda(T)$ functions had the form of a sinusoid added to a linear trend. This nonmonotonic curve resembles what may be found in climate (Fig. 6.6c). This makes the experiments relevant in the context of this book. The number of extreme data, j, was in the order of a few hundreds. The Monte Carlo results revealed good coverage performance of the method (Algorithm 6.1), and also of the alternatives in resampling or CI construction.

Step 1	Event times, augmented by pseudodata (Eq. 6.34)	$\left\{T^{\dagger}_{\text{out}}(j)\right\}_{j=1}^{m^{\dagger}}$
Step 2	Discretization of time T (N_T points)	$T \in [T(1); T(n)]$
Step 3	Kernel occurrence rate estimate (Eq. 6.33)	$\widehat{\lambda}(T)$
Step 4	From data set (Step 1), draw with replacement a simulated set of size $m^{\dagger}$	$\left\{T^{\dagger *}_{\text{out}}(j)\right\}_{j=1}^{m^{\dagger}}$
Step 5	Kernel occurrence rate estimate, simulated data, using same h as in Step 3	$\widehat{\lambda}^{*b}(T)$ (b, counter)
Step 6	Go to Step 4 until $b = B$ (usually $B = 2000$) replications exist	
Step 7	Average	$A(T) = B^{-1} \sum_{b=1}^{B} \widehat{\lambda}^{*b}(T)$
Step 8	Studentize	$T_{\text{stud}}(T, b) = \left[\widehat{\lambda}^{*b}(T) - A(T)\right] \left[\widehat{\lambda}^{*b}(T)\right]^{-1/2}$
Step 9	Determine t_α as	$\#\left\{\left\lvert T_{\text{stud}}(T, b)\right\rvert \leq t_\alpha\right\} = (1 - 2\alpha)\, N_T B$
Step 10	Lower CI bound at T	$\max\left\{0, A(T) - t_\alpha \left[\widehat{\lambda}(T)\right]^{1/2}\right\}$
Step 11	Upper CI bound at T	$A(T) + t_\alpha \left[\widehat{\lambda}(T)\right]^{1/2}$
Step 12	Confidence band is given by joint CIs over T	

Algorithm 6.1. Construction of a bootstrap confidence band for kernel occurrence rate estimation (Cowling et al. 1996). (Step 9 requires interpolation because the number of cases, #, is discrete.) The CI type is called percentile-t.

6.3.2.7 Example: Elbe winter floods (continued)

Figure 6.7 shows the occurrence rate of heavy Elbe winter floods with 90% confidence band. A very long increase starting from the beginning of the millennium culminated in a high during the second half of the sixteenth century, when $\lambda(T) \approx 0.17\,\mathrm{a}^{-1}$, corresponding to a return period of about 6 years. The changes to a low at around 1700 ($\lambda(T) \approx 0.08\,\mathrm{a}^{-1}$) and a subsequent high in the first half of the nineteenth century ($\lambda(T) \approx 0.22\,\mathrm{a}^{-1}$) are significant, as attested by the confidence band. The upper CI bound for that high is approximately $0.31\,\mathrm{a}^{-1}$. Elbe winter flood risk then decreased, and this trend has continued until the present.

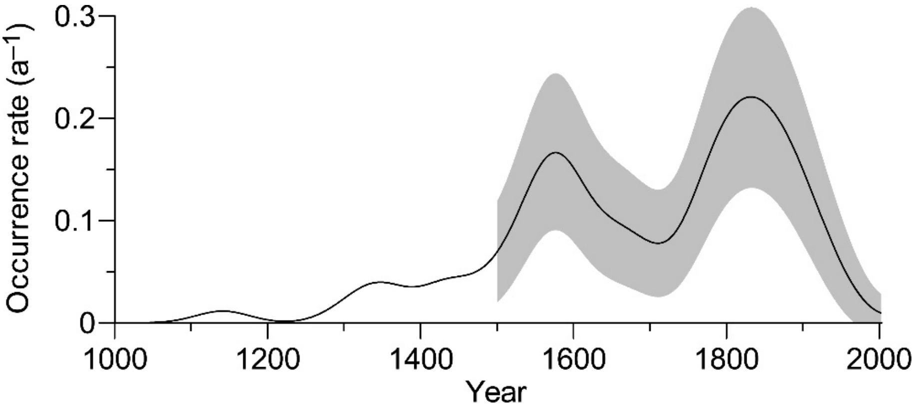


Figure 6.7. Elbe winter floods, occurrence rate estimation, heavy events (magnitudes 2–3). The confidence band is *shaded*. Estimation parameters as in Fig. 6.6c: pseudodata generation rule "twopoint," $h = 35\,\mathrm{a}$; $N_T = 1322$, $B = 2000$; confidence level: $1 - 2\alpha = 90\%$.

In a short interpretation of the mathematical finding, the long-term increase is a result of data inhomogeneity in the form of document loss. It is likely that documents from before the invention of printing in Europe (fifteenth century) were not many, and information about past floods may have been lost before finding entrance into secondary compilations. Therefore the confidence band is drawn only for the interval after A.D. 1500. The end of the sixteenth century was reportedly wet also in other parts of central and southwest Europe (Brázdil et al. 1999). The relatively low flood risk in the decades around $T = 1700$ may be a manifestation of the dry (and cold) European climate (Luterbacher et al. 2001) of the Late Maunder Minimum (Fig. 2.12). The downwards trend from $T \approx 1830$ to the present reflects a reduced risk of ice floods (like that in

1784), which in turn is a product of surface warming in the Elbe region (Mudelsee et al. 2003, 2004).

6.3.2.8 Example: volcanic peaks in the NGRIP sulfate record (continued)

Figure 6.8 shows a number of highs and lows in occurrence of extreme sulfate peaks in the NGRIP ice core record from $\sim$ 10 to $\sim$ 110 ka. Applying a more liberal detection threshold ($z = 5.0$) leads to more events, smaller relative errors ($\propto m^{-1/2}$) and higher significances of the changes in $\widehat{\lambda}(T)$, but also with a more conservative threshold ($z = 10.0$) the changes appear as significant. Estimates close to the boundaries of the observation interval depend on the pseudodata generation rule (not shown) and should be interpreted cautiously.

Construction of the "excess" sulfate record (Fig. 1.4) and extremes detection (Fig. 4.16) had the purpose of extracting from the ice core record the information about the times major volcanic eruptions occurred. For bandwidth selection, we ignore cross-validation and set $h = 5$ ka to be able to inspect changes in volcanic activity on Milankovitch timescales ($\gtrapprox$ 19 ka). Ice-age climate varied on such orbital timescales (Chapter 5), and studying causal relationships between volcanic activity and ice-age climate is facilitated by having common dynamical scales. See background material (Section 6.5).

6.3.2.9 Example: hurricane peaks in the Lower Mystic Lake varve thickness record (continued)

Figure 6.9 shows the occurrence rate of hurricanes in the Boston area (Lower Mystic Lake). Bandwidth selection imposes a slight undersmoothing ($h = 50\,\mathrm{a}$); a further undersmoothing would produce too many nonsignificant wiggles. There has been a significantly higher hurricane activity during the thirteenth century; the upper bound of the 90% CI is close to one event per decade. Hurricane activity after, and likely also before, that period was lower. The Cox–Lewis test (Section 6.3.2.10) about an overall trend is inconclusive ($u = -1.15$, $p = 0.12$) due to the nonmonotonic risk curve and the limited sample size.

The climatic interpretation may notice a relation between the high in hurricane activity and the Medieval Warm Period. The elevated hurricane risk may thus be a result of the Carnot machine in the tropical Atlantic region (Emanuel 1987, 1999), fuelled by higher sea-surface temperatures during that time (Keigwin 1996). However, Besonen et al. (2008: Section 4 therein) recognized "that the LML [Lower Mystic Lake] record is a single point source record representative for the greater Boston area,

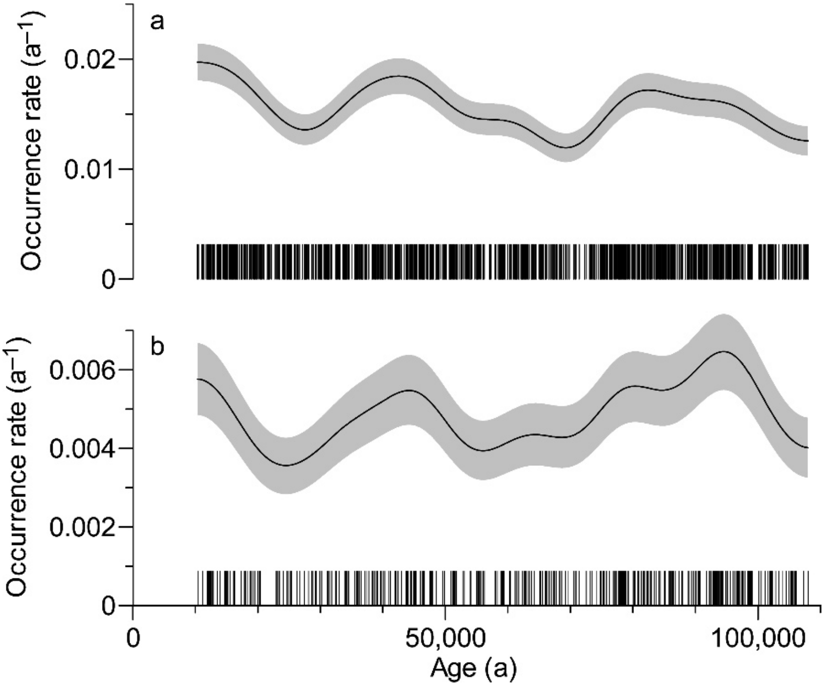


Figure 6.8. NGRIP sulfate record, volcanic activity estimation. Sulfate extremes stemming from volcanic eruptions were detected (Fig. 4.16) by applying thresholds of $z = 5.0$ (**a**) and $z = 10.0$ (**b**) and declustering. Event times (**a** $m = 1525$; **b** $m = 475$) are shown as *bars*, occurrence rate as *solid line*, confidence band *shaded*. Estimation parameters: pseudodata generation rule "reflection," $h = 5000\,\mathrm{a}$; $N_T = 574$, $B = 2000$; confidence level: $1 - 2\alpha = 90\%$.

and hurricanes that passed a few hundred km to the east or west may not have produced the very heavy rainfall amounts and vegetation disturbance in the lake watershed necessary to produce a strong signal within the LML sediments."

6.3.2.10 Parametric Poisson models and hypothesis tests

It is possible to formulate a parametric regression model (Chapter 4) for the occurrence rate. Since $\lambda(T)$ cannot be negative, it is convenient to employ the exponential function. A particularly simple model is

$$\lambda(T) = \exp\left(\beta_0 + \beta_1 T\right). \tag{6.39}$$

Another is the logistic model,

$$\lambda(T) = \frac{\exp\left(\beta_0 + \beta_1 T\right)}{1 + \exp\left(\beta_0 + \beta_1 T\right)}. \tag{6.40}$$

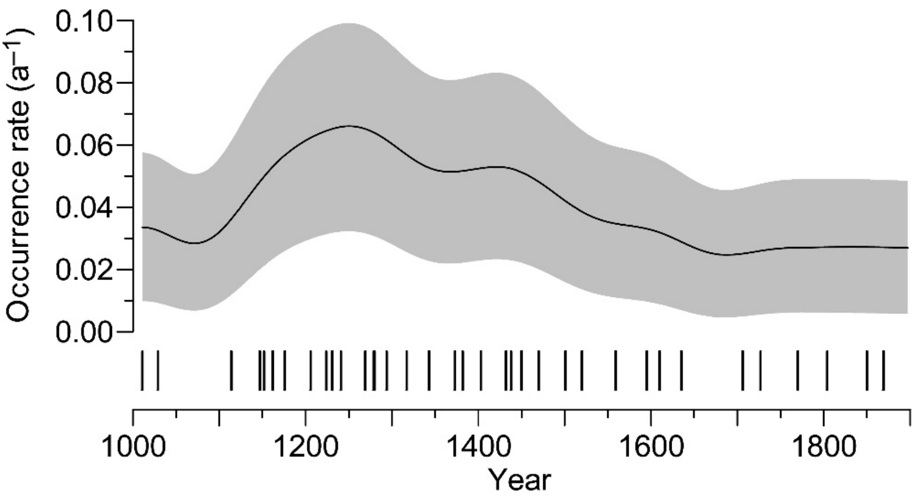


Figure 6.9. Lower Mystic Lake varve thickness record, hurricane activity estimation. Hurricane events were detected (Fig. 4.17) by applying a threshold of $z = 5.2$ and imposing a second condition (graded bed). Event times ($m = 36$) are shown as *bars*, occurrence rate as *solid line*, confidence band *shaded*. Estimation parameters: pseudodata generation rule "reflection," $h = 50$ a; $N_T = 616$, $B = 2000$; confidence level: $1 - 2\alpha = 90\%$.

These two are monotonic functions, and they can be used to model simple increases (decreases) of the occurrence rate. Section 6.5 lists more parametric occurrence rate models. These models do not offer the flexibility of the nonparametric kernel approach (Section 6.3.2.2). The parametric models are better suited to a situation where the task is not quantification of $\lambda(T)$ but rather testing whether $\lambda(T)$ shows an increase (decrease) or not. Cox and Lewis (1966) use the simple model (Eq. 6.39) to test the hypothesis H_1: "$\beta_1 > 0$" (increasing occurrence rate) against H_0: "$\beta_1 = 0$" (constant occurrence rate). Their test statistic is

$$U = \frac{\sum_{j=1}^{m} T_{\text{out}}(j)\,/m - [T(n) + T(1)]\,/2}{[T(n) - T(1)]\,(12m)^{-1/2}}, \tag{6.41}$$

which becomes, with increasing m, rapidly standard normally distributed in shape (Cramér 1946: p. 245 therein). On the sample level, plug in $\{t_{\text{out}}(j)\}_{j=1}^{m}$, $t(1)$ (observation interval, start) and $t(n)$ (observation interval, end) to obtain u.

6.3.2.11 Monte Carlo experiment: Cox–Lewis test versus Mann–Kendall test

The Cox–Lewis statistic (Eq. 6.41) can be used to test for monotonic trends in the occurrence of extremes, the Mann–Kendall statistic (Eq. 4.61) was developed to test for changes in $X_{\text{trend}}(T)$. This theoretical unsuitability of the Mann–Kendall test (Zhang et al. 2004) has, however, not hindered climatologists and hydrologists to apply it for studying extremes.

We analyse the performance of both tests in a Monte Carlo experiment with climatologically realistic properties of the data generating process: a persistent noise component with non-normal distributional shape and an outlier or extreme component that exhibits an upwards trend in occurrence rate. That means, we study the performance of the overall procedure that is employed in practice: detecting extremes and testing for trends in their occurrence.

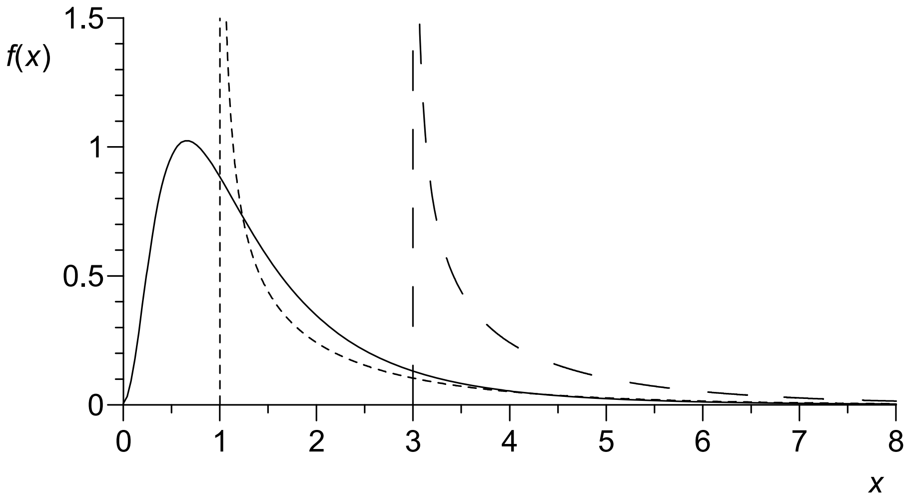


Figure 6.10. Density functions used in Monte Carlo experiment (Tables 6.1, 6.2, 6.3 and 6.4). The PDF of the noise component (*solid line*) is a lognormal, the PDF of the extreme component (which replaces the noise component in the case an extreme occurs) is a chi-squared distribution with $\nu = 1$ degrees of freedom and shifted in x-direction by a value of 1.0 (*short-dashed line*) and 3.0 (*long-dashed line*), respectively.

Figure 6.10 shows that in one simulation setting (the outlier component shifted by 1.0) the PDFs of outlier and noise components overlap to a good degree, while in the other (shifted by 3.0) the PDFs overlap only to a strongly reduced degree.

Table 6.1. Monte Carlo experiment, hypothesis tests for trends in occurrence of extremes. $n_{\mathrm{sim}} = 90{,}000$ random samples were generated from $X(i) = X_{\mathrm{out}}(i) + X_{\mathrm{noise}}(i)$, where $T(i) = i, i = 1, \ldots, n$ and the noise is an AR(1) process with $a = 1/\mathrm{e} \approx 0.37$, lognormal shape, mean 1.0 and standard deviation 0.5 (Table 3.5). The number of extremes, m_{true}, was prescribed. The extreme event times, $T_{\mathrm{out}}(j)$, were generated by taking a random variable uniformly distributed over $[0; 1]$ to the power of κ and mapping it linearly on $[T(1); T(n)]$; the parameter κ served to prescribe the trend in occurrence rate. $X_{\mathrm{out}}(j)$ was drawn from a shifted (+1 in x-direction) chi-squared distribution with $\nu = 1$; this extreme value replaced the value $X_{\mathrm{noise}}(i)$ for which the time, $T(i)$, was closest to $T_{\mathrm{out}}(j)$. Extremes detection employed a constant threshold of median + 3.5 MAD (Fig. 4.15) for the POT approach and a block length of $k = 12$ (Fig. 6.2) for the block extremes approach. The Cox–Lewis test was applied to the detected POT data, the Mann-Kendall test to the POT data and also the block extremes. The significance level of the one-sided tests was $\alpha = 0.10$.

n	m_{true} [a]	κ [b]	*Empirical power* [c]		
			Test		
			Cox–Lewis (POT)	*Mann–Kendall (block extremes)*	*Mann–Kendall (POT)*
120	10	0.75	0.161	0.089	0.041
240	20	0.75	0.181	0.099	0.069
600	50	0.75	0.236	0.150	0.077
1200	100	0.75	0.313	0.217	0.096
2400	200	0.75	0.434	0.338	0.125
6000	500	0.75	0.680	0.619	0.194
12,000	1000	0.75	0.883	0.868	0.300
120	10	0.9	0.129	0.065	0.036
240	20	0.9	0.132	0.066	0.058
600	50	0.9	0.147	0.080	0.059
1200	100	0.9	0.166	0.093	0.065
2400	200	0.9	0.195	0.115	0.073
6000	500	0.9	0.268	0.176	0.089
12,000	1000	0.9	0.357	0.261	0.109

[a] True (prescribed) number of extremes.
[b] Prescribed occurrence rate trend parameter, $\lambda(T) \propto T^{1/\kappa - 1}$.
[c] Number of simulations where H_0: "no trend" is rejected against H_1: "upwards trend," divided by n_{sim}. Standard error is (Efron and Tibshirani 1993) nominally $[\alpha(1-\alpha)/n_{\mathrm{sim}}]^{1/2} = 0.001$.

The results (Tables 6.1, 6.2, 6.3 and 6.4) can be summarized as follows.

1. Higher numbers of extremes allow better detectability of trends in $\lambda(T)$.

Table 6.2. Monte Carlo experiment, hypothesis tests for trends in occurrence of extremes (continued). The number of simulations was in each case $n_{\mathrm{sim}} = 47{,}500$. The significance level of the one-sided tests was $\alpha = 0.05$. The shift parameter of the outlier component was 1.0. See Table 6.1 for further details.

n	m_{true} [a]	κ [b]	*Empirical power* [c] *Test*		
			Cox–Lewis (POT)	*Mann–Kendall (block extremes)*	*Mann–Kendall (POT)*
120	10	0.75	0.085	0.043	0.020
240	20	0.75	0.101	0.058	0.029
600	50	0.75	0.139	0.092	0.040
1200	100	0.75	0.195	0.142	0.051
2400	200	0.75	0.297	0.238	0.070
6000	500	0.75	0.542	0.494	0.120
12,000	1000	0.75	0.797	0.785	0.200
120	10	0.9	0.065	0.030	0.018
240	20	0.9	0.069	0.035	0.026
600	50	0.9	0.079	0.044	0.029
1200	100	0.9	0.089	0.052	0.033
2400	200	0.9	0.111	0.069	0.038
6000	500	0.9	0.166	0.108	0.049
12,000	1000	0.9	0.234	0.173	0.061

[a] True (prescribed) number of extremes.
[b] Prescribed occurrence rate trend parameter, $\lambda(T) \propto T^{1/\kappa - 1}$.
[c] Number of simulations where H_0: "no trend" is rejected against H_1: "upwards trend," divided by n_{sim}. Standard error is nominally $[\alpha(1-\alpha)/n_{\mathrm{sim}}]^{1/2} = 0.001$.

2. Giving the extremes larger values (shift parameter) enhances their detectability and the power of the tests for trends in $\lambda(T)$.

3. Performing a test at a lower significance level (α) reduces the power (as for hypothesis tests in general).

4. A stronger trend in $\lambda(T)$ (parameter κ) can be easier detected (higher power).

5. The best performance, for all settings studied, was achieved by the Cox–Lewis test. For example, when the data size is $n = 1200$, the shift parameter is 3.0, the prescribed number of extremes is $m_{\mathrm{true}} = 100$, which is equivalent to an average $\lambda(T)$ of 1/12, and $\kappa = 0.75$, which means an increase of $\lambda(T) \propto T^{0.333}$, then can this upwards trend be detected by the Cox–Lewis test at the 10% level in approximately 84.2% of all cases (Table 6.3).

Table 6.3. Monte Carlo experiment, hypothesis tests for trends in occurrence of extremes (continued). The number of simulations was in each case $n_{\mathrm{sim}} = 90{,}000$. The significance level of the one-sided tests was $\alpha = 0.10$. The shift parameter of the outlier component was 3.0. See Table 6.1 for further details.

n	m_{true} [a]	κ [b]	*Empirical power* [c] *Test* *Cox–Lewis (POT)*	*Mann–Kendall (block extremes)*	*Mann–Kendall (POT)*
120	10	0.75	0.267	0.145	0.064
240	20	0.75	0.377	0.200	0.069
600	50	0.75	0.622	0.379	0.080
1200	100	0.75	0.842	0.603	0.098
2400	200	0.75	0.977	0.857	0.125
6000	500	0.75	1.000	0.996	0.188
12,000	1000	0.75	1.000	1.000	0.280
120	10	0.9	0.143	0.080	0.056
240	20	0.9	0.169	0.088	0.057
600	50	0.9	0.229	0.122	0.062
1200	100	0.9	0.308	0.167	0.065
2400	200	0.9	0.442	0.246	0.070
6000	500	0.9	0.709	0.450	0.089
12,000	1000	0.9	0.909	0.696	0.106

[a] True (prescribed) number of extremes.
[b] Prescribed occurrence rate trend parameter, $\lambda(T) \propto T^{1/\kappa - 1}$.
[c] Number of simulations where H_0: "no trend" is rejected against H_1: "upwards trend," divided by n_{sim}. Standard error is nominally $[\alpha(1-\alpha)/n_{\mathrm{sim}}]^{1/2} = 0.001$.

6. The Mann–Kendall test may be applied to the block extreme data, $\{T_{\mathrm{out}}(j), X_{\mathrm{out}}(j)\}_{j=1}^{m}$, where the central time of a block is taken as $T_{\mathrm{out}}(j)$. This leads to power levels that may be acceptable in practice. However, in all simulation settings the Cox–Lewis test performed significantly better than the Mann–Kendall test. (Note that the tuning of the block length, k, resulted in $m = m_{\mathrm{true}}$. This may have elevated the test power compared to a situation where k has to be adjusted.)

7. The Mann-Kendall test applied to the POT data leads to an inacceptable test power.

We therefore recommend to use the Cox–Lewis test rather than any form of the Mann–Kendall test for studying trends in the occurrence of extreme events.

Table 6.4. Monte Carlo experiment, hypothesis tests for trends in occurrence of extremes (continued). The number of simulations was in each case $n_{\mathrm{sim}} = 47{,}500$. The significance level of the one-sided tests was $\alpha = 0.05$. The shift parameter of the outlier component was 3.0. See Table 6.1 for further details.

n	m_{true} [a]	κ [b]	*Empirical power* [c]		
			Test		
			Cox–Lewis (POT)	*Mann–Kendall (block extremes)*	*Mann–Kendall (POT)*
120	10	0.75	0.154	0.072	0.028
240	20	0.75	0.240	0.116	0.035
600	50	0.75	0.465	0.262	0.043
1200	100	0.75	0.728	0.474	0.055
2400	200	0.75	0.944	0.771	0.072
6000	500	0.75	1.000	0.991	0.117
12,000	1000	0.75	1.000	1.000	0.185
120	10	0.9	0.073	0.036	0.025
240	20	0.9	0.090	0.046	0.028
600	50	0.9	0.128	0.068	0.031
1200	100	0.9	0.188	0.099	0.034
2400	200	0.9	0.298	0.157	0.037
6000	500	0.9	0.567	0.329	0.048
12,000	1000	0.9	0.831	0.577	0.060

[a] True (prescribed) number of extremes.
[b] Prescribed occurrence rate trend parameter, $\lambda(T) \propto T^{1/\kappa - 1}$.
[c] Number of simulations where H_0: "no trend" is rejected against H_1: "upwards trend," divided by n_{sim}. Standard error is nominally $[\alpha(1-\alpha)/n_{\mathrm{sim}}]^{1/2} = 0.001$.

6.3.3 Hybrid: Poisson–extreme value distribution

Let us consider the estimation problem for extreme value time series in a more general manner. In principle, the distribution function of $X_{\mathrm{out}}(i)$ may change with time, $T(i)$. The sample is used to estimate properties of the time-dependent PDF. Fitting a stationary distribution (GEV or GP; Section 6.2) corresponds to using an "estimation area" (Fig. 6.11) with dx_{out} arbitrarily small and δt (sample level) equal to the whole observation interval, $[t(n) - t(1)]$. Fitting an inhomogeneous Poisson process (Section 6.3.2) corresponds to using an estimation area with δt small (in the order of the bandwidth, h) and dx_{out} arbitrarily large (interval from u to ∞). Fitting a GEV or GP model with time-dependent parameters (Section 6.3.1) means using an estimation area with dx_{out} arbitrarily small and δt in principle also small. By writing "in principle" we acknowledge that here the comparison is flawed and the

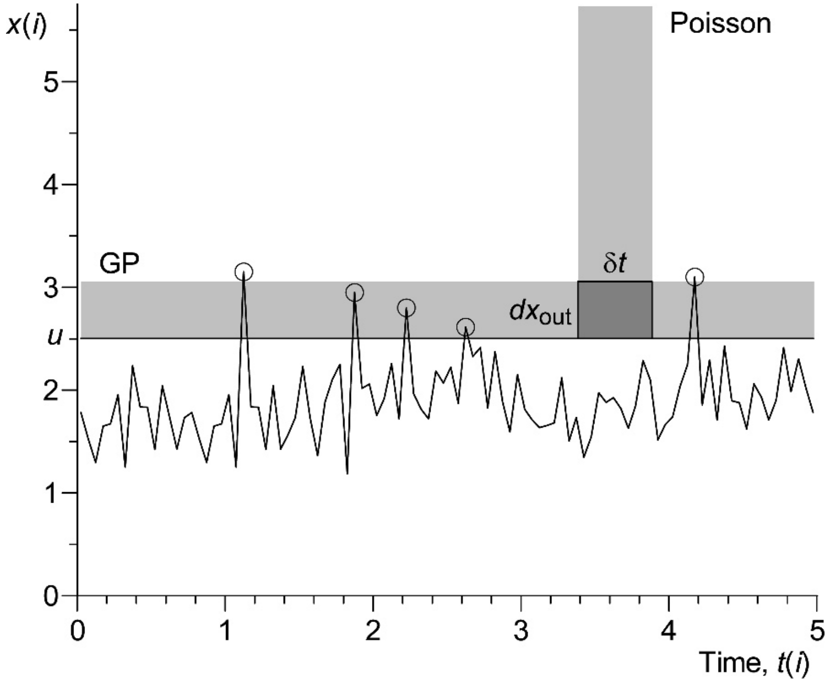


Figure 6.11. Estimation area for extreme value time series. The area (*dark shading*) is given by $dx_{\mathrm{out}} \cdot \delta t$. In the GP case, where POT values (*circles*) are analysed, $\delta t = [t(n) - t(1)]$; in the inhomogeneous Poisson case, where event times are analysed, $dx_{\mathrm{out}} \to \infty$. u, threshold; $x_{\mathrm{out}}(i) = x(i) - u$.

time-dependence is actually modelled parametrically and not estimated nonparametrically.

The hybrid model (Smith 1989, 2004) is a mixture between a nonparametric description of the time-dependence via the inhomogeneous Poisson process and a parametric extreme value distribution model such as the GP. The probability that an event occurs is multiplied with the probability that the extreme has a size within a certain interval. The Poisson–GP hybrid model corresponds to a combined rate measure,

$$\Lambda(T, x_{\mathrm{out}}) = \lambda(T) \cdot f_{\mathrm{GP}}(x_{\mathrm{out}}). \tag{6.42}$$

Analogously, the Poisson–GEV hybrid uses the GEV density function, $f_{\mathrm{GEV}}(x_{\mathrm{out}})$.

The Monte Carlo experiment (Section 6.3.2.11) employed a hybrid model to generate the data. Conditional on the existence of an event at $T_{\mathrm{out}}(j)$, which is described by the occurrence rate, we drew $x_{\mathrm{out}}(j)$ from a chi-squared distribution with $\nu = 1$.

Fitting the hybrid model corresponds to using an estimation area with both δt and dx_{out} small (Fig. 6.11). Davison and Ramesh (2000) developed an estimation method for the hybrid model based on kernel smoothing and maximum likelihood fitting (Section 6.5). We remark that such "two-dimensional" (T, X) estimations may require a large sample size to achieve acceptably small error bars.

6.4 Sampling and time spacing

The sampling of a climate archive (Fig. 1.13) can influence the detectability of extreme events. Table 6.5 lists the notation for this section.

Consider the case that the spacing, $d(i)$, is large compared with the sample duration, $D(i)$, or its diffusion-extended form, $D'(i)$, and also large compared with the persistence time, τ. It may then be that the time series fails to record information about an extreme event, $X_{\text{out}}(i)$. This would render the series useless for the purpose of risk analysis. Another case is a hiatus, where $d(i^*) \gg D(i^*)$ and $d(i^*) \gg \tau$ only at a certain point, i^*. For fitting an extreme value distribution, the log-likelihood function may be adapted (Coles 2004) to take the absent portion of information into account. For estimating the occurrence rate (Section 6.3.2), it is indicated to "exclude" the hiatus prior to the analysis, that means, to shift artificially the portion of the time series before or after the hiatus. The calculations (kernel estimation, cross-validation) are carried out on those time-transformed data and the hiatus "included" by inserting the time-gap in the results.

A further case is uneven spacing when hiatuses are absent. Block extremes detection for fitting a GEV distribution may then be enhanced by fixing the number of observations, k, per block rather than the length of a block.

This section focuses on still another, "ice core" case, where the sample duration is large and the age–depth relation is strongly nonlinear, leading to large changes in $D(i)$. (In ice cores, which are influenced by compaction, $D(i)$ can exhibit strong trends.) This poses a detection problem for extremes because with $D(i)$ changes also the recording quality downcore (inhomogeneity). Note that the NGRIP sulfate record (Section 6.3.2.8) does not suffer strongly from inhomogeneities of this kind owing to a very high time resolution that allowed to have $D(i) \approx$ const. We follow Mudelsee (1999) and study the physics of the recording system to derive a data transformation that corrects for the inhomogeneity.

Suppose that the archive is an ice core with a segmented sampling and the measured variable is, for example, sulfate. The objective is to detect the extremes stemming from an event of short duration (e.g., volcanic

Table 6.5. Notation for Section 6.4.

$i = 1, \ldots, n$	Index, segment (top: $i = 1$)
n	Number of segments
k	Index, extreme events within segment i
$D(i)$	Duration, segment i
$D_{\mathrm{out}}(i)_k$	Duration, extreme event k within segment i
$N_{\mathrm{trend}}(i)$	Number of particles of interest (sulfate) from background, segment i
$\bar{N}_{\mathrm{trend}}(i)$	Number of particles not of interest (non-sulfate) from background, segment i
$N_{\mathrm{out}}(i)$	Number of particles of interest (sulfate) from extreme events, segment i
$N_{\mathrm{out}}(i)_k$	Number of particles of interest (sulfate) from kth extreme event, segment i
$\bar{N}_{\mathrm{out}}(i)$	Number of particles not of interest (non-sulfate) from extreme events, segment i
$F_{\mathrm{out}}(i)_k$	Flux of particles of interest (sulfate) from kth extreme event, segment i
$\bar{F}_{\mathrm{trend}}(i)$	Flux of particles not of interest (non-sulfate) from background, segment i
$A(i)$	Exposure area to flux of particles of interest (sulfate), segment i
$\bar{A}(i)$	Exposure area to flux of particles not of interest (non-sulfate), segment i
$X(i)$	Concentration of particles of interest (sulfate), segment i
$X_{\mathrm{trend}}(i)$	Concentration of particles of interest (sulfate) from background, segment i
$X'(i)$	Transformed concentration of particles of interest (sulfate), segment i

eruption) against the background trend. The sulfate concentration is

$$X(i) = \frac{N_{\mathrm{trend}}(i) + N_{\mathrm{out}}(i)}{N_{\mathrm{trend}}(i) + \bar{N}_{\mathrm{trend}}(i) + N_{\mathrm{out}}(i) + \bar{N}_{\mathrm{out}}(i)}. \tag{6.43}$$

Assumption 1. The number of non-sulfate particles from the background dominates,

$$\bar{N}_{\mathrm{trend}}(i) \gg N_{\mathrm{trend}}(i) + N_{\mathrm{out}}(i) + \bar{N}_{\mathrm{out}}(i). \tag{6.44}$$

This is certainly fulfilled because the bulk of the material is water. Then we have

$$X(i) - X_{\mathrm{trend}}(i) \approx \frac{N_{\mathrm{out}}(i)}{\bar{N}_{\mathrm{trend}}(i)}, \tag{6.45}$$

where $X_{\mathrm{trend}}(i)$ is the time-dependent background sulfate concentration,

$$X_{\mathrm{trend}}(i) = \frac{N_{\mathrm{trend}}(i)}{N_{\mathrm{trend}}(i) + \bar{N}_{\mathrm{trend}}(i) + N_{\mathrm{out}}(i) + \bar{N}_{\mathrm{out}}(i)}. \tag{6.46}$$

Consider first the case of a single event recorded in segment i. For a short duration, $D_{\mathrm{out}}(i)_{k=1} \ll D(i)$, there is input of material (sulfate and non-sulfate) from the event (eruption) added to the background. The flux of sulfate particles from the event is

$$F_{\mathrm{out}}(i)_{k=1} = \frac{N_{\mathrm{out}}(i)_{k=1}}{A(i) \cdot D_{\mathrm{out}}(i)_{k=1}}, \tag{6.47}$$

where $N_{\mathrm{out}}(i)_{k=1}$ is the number of such particles; here $N_{\mathrm{out}}(i)_{k=1} = N_{\mathrm{out}}(i)$. The area $A(i)$ represents the susceptibility of segment i exposed to the flux of incoming particles, mainly the perpendicular component. The flux of non-sulfate particles from the background is

$$\bar{F}_{\mathrm{trend}}(i) = \frac{\bar{N}_{\mathrm{trend}}(i)}{\bar{A}(i) \cdot D(i)}. \tag{6.48}$$

Then it follows from Eq. (6.45),

$$X(i) - X_{\mathrm{trend}}(i) \approx \frac{F_{\mathrm{out}}(i)_{k=1} \cdot A(i) \cdot D_{\mathrm{out}}(i)_{k=1}}{\bar{F}_{\mathrm{trend}}(i) \cdot \bar{A}(i) \cdot D(i)}. \tag{6.49}$$

Assumption 2. The susceptibility to incoming sulfate particles is proportional to the susceptibility to incoming non-sulfate particles,

$$A(i) \propto \bar{A}(i). \tag{6.50}$$

The degree to which this is fulfilled depends on the site and the changes of the type of deposition (Wagenbach et al. 1996; Fischer 1997). For wet deposition, where the particles are scavenged by precipitation, the assumption should be well fulfilled because water and particles are transported to the exposed segment by the same carrier.

This leads from Eq. (6.49) to

$$[X(i) - X_{\mathrm{trend}}(i)]\, D(i) \propto \frac{F_{\mathrm{out}}(i)_{k=1} \cdot D_{\mathrm{out}}(i)_{k=1}}{\bar{F}_{\mathrm{trend}}(i)}. \tag{6.51}$$

This formula for one single event is only approximate. The generalization to several extreme events per segment i is straightforward: approximately,

$$[X(i) - X_{\mathrm{trend}}(i)]\, D(i) \propto \frac{\sum_k F_{\mathrm{out}}(i)_k \cdot D_{\mathrm{out}}(i)_k}{\bar{F}_{\mathrm{trend}}(i)}. \tag{6.52}$$

Assumption 3. The background trend is constant over the whole time interval.

$$\bar{F}_{\mathrm{trend}}(i) = \text{ const.} \tag{6.53}$$

This is certainly violated. However, there is confidence that the temporal variability of the numerator of the term on the right-hand side of Eq. (6.52) is clearly larger (event versus non-event) than the variability of the denominator. For example, Wagenbach (1989) reported $\bar{F}_{\text{trend}}(i)$ to be in the interval from 0.2 to 0.4 (arbitrary units) for an Alpine ice core site.

Assumption 3 and Eq. (6.52) lead to the transformed values,

$$X'(i) = [X(i) - X_{\text{trend}}(i)]\, D(i), \tag{6.54}$$

which should approximate the climatic signal of the events,

$$X'(i) \propto \sum_k F_{\text{out}}(i)_k \cdot D_{\text{out}}(i)_k. \tag{6.55}$$

The transformation corrects for the dilution of the extreme values by the background values; the degree of the dilution depends on $D(i)$.

For the practice of extreme value analysis on segmented ice core data (unevenly spaced timescale), a multi-stage approach is indicated. (1) Estimate the trend component (e.g., by using the running median). (2) Transform the data (Eq. 6.54). (3) Estimate trend and variability on the transformed data to select the threshold for detecting the POT data.

6.5 Background material

The **literature** on extreme value distributions is extensive. The GEV as the limiting distribution of a block extreme was introduced by Jenkinson (1955). The GP and its relation to the GEV distribution was established by Pickands (1975). The following is a selection of statistical books and papers that contain practical examples and are accessible to climatologists. Hydrology had a leading role in posing questions to statistical science and motivating theoretical research. Books were written by Gumbel (1958), Embrechts et al. (1997), Reiss and Thomas (1997), Coles (2001b) and Beirlant et al. (1996, 2004). See also Clarke (1994: Chapter 3 therein) for a hydrological perspective. A classic is the paper (with discussion) by Smith (1989) on extreme value analysis and trend detection in an ozone time series. Overviews with relevance for climatology (serial dependence, nonstationarity) were given by Buishand (1989), Coles (2004), Smith (2004) and Khaliq et al. (2006). The review on statistical approaches in flood research by Cox et al. (2002) contains a comparison of the efficiencies of estimators for block extremes and POT data. More theoretical are the book by Leadbetter et al. (1983) and their review (Leadbetter and Rootzén 1988) on the asymptotic distributional properties of extreme values from serially dependent

Table 6.6. GEV distribution, parameter notations.

Reference	*Parameter*		
	Location	*Scale*	*Shape*
Hosking (1985)	ξ	α	$-k$
Smith (1985)	μ	σ	$-k$
Lu and Stedinger (1992)	ξ	α	$-\kappa$
Johnson et al. (1995: Eq. (22.4) therein)	ξ	θ	$1/\alpha$
Johnson et al. (1995: Eq. (22.183) therein)	ξ	θ	$-\gamma$
Reiss and Thomas (1997)	μ	σ	γ
Kotz and Nadarajah (2000); Coles (2001b); this book	μ	σ	ξ

stochastic processes. Those two and also the books by Galambos (1978) and Resnick (1987) contain details about how the distribution of the extreme value approaches ("gets attracted to") one of the "families" of the GEV distribution (i.e., special parameter cases) and the role of the "normalizing constants" (Pickands 1975) used to scale the extreme value—concepts this book does not present. We finally mention Johnson et al. (1995: Chapter 22 therein) and Kotz and Nadarajah (2000) on the GEV and Johnson et al. (1994: Chapter 20 therein) on the GP distribution.

The **scaling method** of taking nonstationarity into account for the POT approach (Section 6.1.2.1) has been supported in a recent Monte Carlo study (Eastoe and Tawn 2009).

The **extreme value analogue of the central limit theorem** was established by Fréchet (1927), Fisher and Tippett (1928) and Gnedenko (1943).

The **naming and notation** of the distribution types, special cases of extreme value distributions and their parameters have developed a rich variety over the past decades. Table 6.6 gives a selection of parameter notations used for the GEV distribution. Our abbreviation of the generalized Pareto (GP) distribution is not following convention; this abbreviates generalized Pareto distribution as GPD. Writing the GP distribution (Eq. 6.10) with $(x_{\text{out}} - u)$ instead of realizations of an excess variable is non-standard. The GEV distribution is sometimes also referred to as von Mises–Jenkinson distribution. The special case of a GEV distribution with shape parameter $\xi = 0$ (Eq. 6.5) is also called Gumbel distribution; the special cases $\xi > 0$ and $\xi < 0$ are also known as Fréchet and Weibull distribution, respectively. A GP distribution with shape parameter $\xi = 0$ (Eq. 6.10) is an exponential distribution.

Error propagation or the delta method is performed for the return level for a GEV distribution as follows. The return level is for $\xi \neq 0$ given by $x_p = \mu - (\sigma/\xi)\{1 - [-\ln(1-p)]^{-\xi}\}$. The standard error of its estimate is $\mathrm{se}_{\widehat{x}_p} = [VAR(\widehat{x}_p)]^{1/2}$. The variance is given by (Coles 2001b: Section 3.3.3 therein)

$$
\begin{aligned}
VAR(x_p) \approx & \left(\frac{\partial x_p}{\partial \mu}\right)^2 VAR\,[\mu] + \left(\frac{\partial x_p}{\partial \sigma}\right)^2 VAR\,[\sigma] + \left(\frac{\partial x_p}{\partial \xi}\right)^2 VAR\,[\xi] \\
& + 2\left(\frac{\partial x_p}{\partial \mu}\right)\left(\frac{\partial x_p}{\partial \sigma}\right) COV\,[\mu,\sigma] + 2\left(\frac{\partial x_p}{\partial \mu}\right)\left(\frac{\partial x_p}{\partial \xi}\right) COV\,[\mu,\xi] \\
& + 2\left(\frac{\partial x_p}{\partial \sigma}\right)\left(\frac{\partial x_p}{\partial \xi}\right) COV\,[\sigma,\xi]\,,
\end{aligned}
\tag{6.56}
$$

where the "hats" have to be inserted. On the sample level, the parameter estimates and the elements of the estimated covariance matrix are plugged in.

Declustering records prior to fitting a GP distribution discards excess data and loses information, as noted by Coles (2001b). A more efficient GP estimation may come from retaining all excess data (also those within a cluster) and modelling the serial dependence. Fawcett and Walshaw (2006) present Monte Carlo evidence supporting this approach and an example where the AR(1) persistence model is applied to hourly wind-speed data from central and northern England (time interval 1974–1991, 10 sites; 1975–1984, 2 sites). An alternative for efficient GP estimation (Fawcett and Walshaw 2007) may be inflating the covariance matrix (Eq. 6.14). Referring to a preprint by Smith RL, this paper advises to replace the covariance matrix by $\mathbf{H}^{-1}\mathbf{V}\mathbf{H}^{-1}$, where $\mathbf{H}$ is the Fisher observed information matrix and $\mathbf{V}$ is the covariance matrix of the likelihood gradient vector. Ferro and Segers (2003) devised an automatic declustering scheme that relies on the extremal index (Section 6.2.5.2), which is estimated before declustering. Ledford and Tawn (2003) developed a diagnostic tool (autocorrelation measure for extreme values), which helps to assess whether declustering of a series has been successful.

The **efficiency** of a statistical estimator refers to its standard error under a particular parent distribution. Higher efficiency means smaller se.

Fisher information is a measure of the amount of information provided by a sample about an unknown parameter (Kullback 1983). In case of maximum likelihood estimation of the parameters of the GEV distribution (Eq. 6.8), the information is related to the expectation of the negative of the matrix that gives the curvature of the log-likelihood function. Efron and Hinkley (1978) gave the advice, and this is also the modern tendency (Davison AC 2009, personal communication), to use instead of the Fisher expected information the observed information matrix, that means, not to use the expectation (Eq. 6.8), but rather the numerically determined derivatives of the log-likelihood function.

Optimal estimation is a general theme, it opens a wide research field (Section 9.5). Regarding the GEV and GP models, in addition to estimation methods presented in Section 6.2, many articles devoted to improving the estimation appeared, including the following. Castillo and Hadi (1997) reviewed GP estimation methods and suggested a new one (elemental percentile method), which is based on a two-stage procedure. Castillo and Hadi (1997: p. 1611 therein) wrote: "In the first stage, preliminary initial estimates of the parameters are calculated [based on $\{x_{\text{out, sort}}(j)\}_{j=1}^{m}$]. These initial estimates are combined in the second stage to give overall estimates of the parameters. These final estimates are then substituted in the quantile function to obtain estimates of all desired quantiles." They provided theoretical asymptotic as well as Monte Carlo evidence in support of their estimator. Martins and Stedinger (2000) augmented maximum likelihood estimation of GEV parameters with a Bayesian method to restrict values of the "critical" shape parameter, ξ, to "statistically/physically reasonable" ranges. Subsequently, Martins and Stedinger (2001) extended this "generalized maximum likelihood estimation" to the case of GP parameters and quantiles. A full Bayesian estimation method with computing-intensive determination of the distribution of the GEV parameter estimates was presented by Reis and Stedinger (2005). Bayesian methods for GP parameter estimation were developed also to include either prior expert knowledge (Parent and Bernier 2003b) or additional historical information (Parent and Bernier 2003a). Documentary data, although less accurate than runoff measurements, may cover longer time intervals (Section 6.1) and can therefore lead to an improved tail estimation. Hewa et al. (2007) applied an adaption of PWM estimation of the GEV model, where weighting is imposed on the extreme part of the distribution, to study low river flows in Australia.

Time-dependent extreme value distributions have been applied in a number of climatological and environmental studies, with the GEV model seemingly preferred over the GP. Smith (1989) fitted the GEV model with linearly time-dependent location (Eq. 6.23) and constant scale and shape parameters to hourly ground-level ozone concentration time series from a station in Texas, April 1973 to December 1986 ($n =$ 119,905). Despite this simple form of time-dependence, he reported that the maximization of the log-likelihood function was nontrivial and that numerical techniques that approximate the second derivatives (instead of explicitly calculating them) performed better. In his later review, Smith (2004) applied the same GEV model to wind-speed extremes, where he allowed seasonality in the time-dependence of the location by including terms $\sim \sin(2\pi T/T_0)$, where $T_0 = 1\,\mathrm{a}$. In that paper, he compared this model with a more elaborate model (exponential increases with time in location and scale) for rainfall extremes from a number of stations in the United States of America, period 1951–1997. Coles (2001a) used the GEV model with time-dependent location and scale parameters and a constant shape for analysing annual maxima of wind speed recorded at stations in the United States of America between 1912 and 1987. Regarding time-dependence in shape, Coles (2001a: Section 2.2 therein) remarks that "such a model is likely to be difficult to identify." Katz et al. (2002) considered the GEV model with linear trend in location, exponential increase in scale and constant shape for studying extreme precipitation and runoff events in a changing climate. Seasonality can be taken into account at the stage of data selection by setting the block size to 1 year or by dividing the year into seasons and building separate models. Coles and Pericchi (2003) and Coles (2004) formulated the division of the year into two seasons as an inference problem of the 2 days of change. These papers present also an adaption of the likelihood function for a GEV model to a situation where partly only annual maxima were recorded and partly daily values exist. Their example, rainfall in Venezuela, with $d(i) = 1$ year for 1951–1961 and $d(i) = 1$ day for 1961–1999, is not unusual within the context of direct meteorological observations. Naveau et al. (2005) applied the GEV model with exponential trend in the location parameter to time series of lichen size from a moraine formation in the Andes mountains with the objective to study glacier retreats over the past approximately 700 years. Kharin and Zwiers (2005) studied the global, gridded near-surface air temperature and precipitation for the interval 1990–2100 using the CGCM2 climate model driven by various greenhouse gas emission scenarios. These authors applied the GEV model with linear trends in location and exponential trends in scale. Interestingly, they allowed for a linear trend in the shape parameter and found no "serious computational obstacle" to solving the maximum likelihood estimation, although $\widehat{\xi}(T_{\mathrm{out}}) \approx 0$ was found as a result for most of the grid-point time series. Kharin and Zwiers (2005) also preferred error bars from nonparametric bootstrap resampling over the more traditional estimates from the covariance matrix. Rust et al. (2009) fitted the GEV model with seasonal trends in location and scale (and constant shape) to daily rainfall at 689 stations across the United Kingdom. From their analysis of the interval from 1 January 1900–31 December 2006, they concluded that during the winter season (Rust et al. 2009: p. 106 therein) "the entire west coast shows a band of return levels larger than the inland and the east coast." Pujol et al. (2007) tested for trends in the GEV distribution fitted with maximum likelihood to time series of monthly and annual rainfall maxima from 92 stations in the French Mediterranean region. The competing models were the stationary (three parameters), and the model with linear trends in location and scale and constant shape (five parameters). The test statistic employed was the deviance, which is defined as $D = 2\ln(L_1 - L_0)$, where L_1 and L_0 is the maximized log-likelihood of the linear and the stationary model, respectively. Under stationarity and for large m, D is approximately chi-squared distributed with the degrees of freedom equal to the difference in number of parameters (Coles 2001b), that is, two in this case. Zhang et al. (2004) analysed the test power by means of Monte Carlo simulations and showed the superiority over the Mann–Kendall test for detecting trends in GEV parameters. In a series of papers, Strupczewski et al. (2001a), Strupczewski and Kaczmarek (2001) and Strupczewski et al. (2001b) developed the methodology of time-dependent moments and analysed runoff extremes from Polish rivers, interval 1921–1990. Trends in location and scale of various degree of complexity were fitted by maximum likelihood or an adaption of weighted least squares, and model selection was based on the AIC, similar to the deviance test. Instead of letting, say, the location parameter depend directly on time, one may let it depend on another, informative variable (covariate): $\mu(T_{\mathrm{out}}) = \beta_0 + \beta_1 Y(T_{\mathrm{out}})$ is a linear model. Smith and Shively (1995) analysed trends in ground-level ozone concentration, $X(T_{\mathrm{out}})$, by means of GP distributions dependent on time and other covariates, $Y(T_{\mathrm{out}})$, such as maximum temperature or average wind speed. The GP distribution with time-dependent scale parameter was applied in other work dealing with surface-air temperature extremes in the North Atlantic region during 1948–2004 (Nogaj et al. 2006) or river floods in the Czech Republic during 1825–2003 (Yiou et al. 2006). Time-dependent GP and GEV models were fitted to runoff records from Germany during 1941–2000 (Kallache 2007). Assuming a constant shape parameter this author found no major numerical problems in likelihood maximization using the simplex method, even for polynomial time-dependences in location and scale of orders up to four (Kallache M 2008, personal communication).

Covariates, $Y(i)$, bear information about the extremal part of the climate variable of interest, $X(i)$. This chapter focuses on the time, $T(i)$, as covariate. However, other covariates as well may help, also jointly, to predict $X(i)$ extremes. This leads to methods of regression between two processes (Chapter 8). In particular, a climate model may perform better at predicting the $Y(i)$ than the extremal part of $X(i)$. Better climate risk forecasts should then come from model-predicted $Y(i)$. For example, Cooley et al. (2007) use as covariates (1) mean precipitation and (2) topography to model extreme precipitation return levels for Colorado (time series from 56 stations, interval 1948–2001).

Semi-parametric estimation of the time-dependent GEV distribution based on kernel weighting and local likelihood estimation was introduced by Davison and Ramesh (2000) and Hall and Tajvidi (2000). The unweighted local log-likelihood function, see Eq. (6.6), is written as $\ln[L(\mu,\sigma,\xi;y(j))]$, where μ,σ and ξ are the GEV parameters and $y(j)$ is a scaled extreme (Eq. 6.7). The weighted log-likelihood function is formed by putting a kernel weight, K, to the local log-likelihood:

$$\ln\left[L(\mu,\sigma,\xi;T)\right] = \sum_{j=1}^{m} K\big(\left[T - T_{\text{out}}(j)\right]/h\big) \cdot \ln\left[L(\mu,\sigma,\xi;y(j))\right], \quad (6.57)$$

where h is the bandwidth. Hall and Tajvidi (2000) present several bandwidth selectors. Maximization of the weighted log-likelihood function produces the local (in T) maximum likelihood estimates. Davison and Ramesh (2000) further adapted bootstrap resampling by studentizing to determine the estimation uncertainty. They presented Monte Carlo experiments for sample size $m = 100$, which demonstrated acceptable coverage performance. The semi-parametric method was then applied to the central England temperature time series (Section 2.6), which showed that (Davison and Ramesh 2000: p. 202 therein) "the change in upper extremes is mostly due not to changes in the location or in the shape of their distribution but in their variability." In a later paper (Ramesh and Davison 2002), the authors applied semi-parametric local likelihood estimation to study time-dependent extremes in sea-level data from Venice, 1887–1981. Butler et al. (2007) employed local likelihood estimation to quantify trends in extremes of modelled North Sea surges for the period 1955–2000. Another semi-parametric estimation method (Pauli and Coles 2001; Chavez-Demoulin and Davison 2005) uses spline functions (Eq. 4.62) to model the time-dependences of the GEV parameters. This was applied to annual temperature maxima between 1900 and 1980 at two stations in England (Pauli and Coles 2001) and daily winter temperature minima between 1971 and 1997 at 21 stations in Switzerland (Chavez-Demoulin and Davison 2005).

Poisson and point processes are treated in the books by Cox and Lewis (1966), Cox and Isham (1980) and Karr (1986).

Occurrence rate is the name employed in this book for the parameter λ or the function $\lambda(T)$ of the Poisson process, prohibiting misunderstandings from the alternatively used "intensity."

Parametric occurrence rate models are often used in combination with statistical tests. Loader (1992) developed tests, based on maximum likelihood estimation, to choose among three models. The first is a gradual change-point model

$$\lambda(T) = \begin{cases} \exp(\beta_0 + \beta_1 T) & \text{for } T(1) \leq T \leq T_{\text{change}}, \\ \exp(\beta_0 + \beta_1 T + \beta_2) & \text{for } T_{\text{change}} < T \leq T(n), \end{cases} \tag{6.58}$$

where T_{change} is the change-point in time. It includes the second, abrupt change-point model, which has $\beta_1 = 0$, and it includes also the simple model (Eq. 6.39), which has $\beta_2 = 0$. Loader (1992) derived analytical approximations of the test powers. Worsley (1986) had previously devised a test for the abrupt change-point model with null hypothesis "constant occurrence rate." Frei and Schär (2001) constructed a test for increasing (decreasing) occurrence rate in the logistic model (Eq. 6.40) and carried out Monte Carlo simulations to evaluate the test power. A caveat is that their experiments do not simulate serial dependence. This may lead to an overestimated power when applied to a climate time series that stems from a persistent process.

Model suitability of the inhomogeneous Poisson process can theoretically be tested using methods (Solow 1991; Smith and Shively 1994, 1995) based on the spacing of the event times, $S_{\text{out}}(j) = T_{\text{out}}(j) - T_{\text{out}}(j-1)$. One procedure is to construct a probability plot (as in Fig. 6.3e) to test the shape of the distribution function, the other is to calculate the correlation (Chapter 7) between successive $S_{\text{out}}(j)$ to assess the statistical independence. Further tests are reviewed by Lang et al. (1999).

Quantile regression (Section 4.4) may in principle be used for estimating time-dependent quantiles. Few studies exist yet in climatology. Sankarasubramanian and Lall (2003) presented a Monte Carlo experiment that compares this method with the semi-parametric local likelihood estimation (Davison and Ramesh 2000). Both methods exhibited similar bias and RMSE values of quantile estimates. Sankarasubramanian and Lall (2003) further applied both methods to estimate time-dependent risk of floods in the river Clark Fork, based on daily runoff data from the interval 1930–2000. Elsner et al. (2008) found an increasing magnitude of Atlantic tropical cyclones for the period from 1981 to 2006. This result may be interpreted with caution as the study did deliberately not take persistence into account. Allamano et al. (2009) found that "global warming increases flood risk in mountainous areas" on basis of quantile regression analyses of annual maxima of 27 Swiss runoff series over the past approximately 100 years. Unfortunately, their paper did not provide the details required to reproduce their finding (station names, data sizes and missing values). For example, spurious upwards (downwards) trends might arise if missing values cluster in the earlier (later) period. A second caveat against accepting the found significance of the increased flood risk comes from the authors' deliberate ignorance of the Hurst phenomenon of long-term persistence (Section 2.5.3).

Timescale-uncertainty effects on extreme value analyses seem not to have been studied yet. For stationary models (Section 6.2), we anticipate sizable effects on block extremes–GEV estimates only when the uncertainties distort strongly the blocking procedure. For nonstationary models (Section 6.3), one may augment confidence band construction by inserting a timescale simulation step (after Step 4 in Algorithm 6.1).

The **Elbe flood in August 2002** has received extensive scientific coverage. Ulbrich et al. (2003b) analyse the meteorological situation that led to this extreme event. Engel et al. (2002) and Ulbrich et al. (2003a) explain the hydrographical development. Grünewald et al. (2003) and Becker and Grünewald (2003) assess the damages caused by the catastrophe and consider consequences such as improving the risk protection.

The **Elbe flood occurrence rate since 1021** was estimated by Mudelsee et al. (2003). This paper and Mudelsee et al. (2004) consider besides climatological influences the following other potential factors: deforestation, solar activity variations, river engineering, reservoir construction and land-use changes. Analyses of flood risk, not only of the Elbe, benefit from considering seasonal effects. In many parts of central Europe, the floods in hydrological summer are caused by heavy rainfall, in the winter additionally by thawing snow (Fischer 1907; Grünewald et al. 1998). Breaking river ice may function as barrier, enhancing winter floods severely (Grünewald et al. 1998). Elbe summer flood risk during the instrumental period (from 1852) does not show trends in occurrence of heavy floods (Mudelsee et al. 2003). This season can therefore be analysed using a stationary model (Fig. 6.3). Elbe winter flood risk decreased significantly during the instrumental period (Fig. 6.7).

Volcanism and climate are coupled: a volcanic eruption releases material into the atmosphere, which changes the radiative forcing and leads generally to cooling. This and other mechanisms have been observed for the past millennium via proxy variables (Robock 2000). Volcanic influences on climate act also on longer timescales: the Holocene (Zielinski et al. 1994), the late Pleistocene (Zielinski et al. 1996) and the Pliocene (Prueher and Rea 2001). The results obtained with kernel occurrence rate estimation on sulfate data from the NGRIP ice core (Fig. 6.8), interval 10–110 ka, may be compared with the findings (Zielinski et al. 1996) from histogram estimation on sulfate data from the GISP2 ice core. These authors report elevated levels of activity during [6 ka; 17 ka] and [22 ka; 35 ka]. These time intervals, and possibly also that of another high during [55 ka; 70 ka] (Zielinski et al. 1996: Fig. 5 therein), agree qualitatively well with the results from NGRIP. Quantitative agreement (at maximum a few tens of eruptions per ka) is approached when adopting the more liberal detection threshold (Fig. 6.8a). The occurrence rate of volcanic eruptions, restricted to the tropical region and shorter timescales (period 1400–1998), was estimated by application of a parametric logistic model to sulfate records from ice cores (Ammann and Naveau 2003). These authors found indications for the existence of a cycle of 76 year period in occurrence rate and adapted the logistic model (Eq. 6.40) by adding a sinusoidal time-dependence.

A **hurricane activity** peak during medieval times was also found on proxy data in the form of overwash sediment records from sites along the North American East Coast (Mann et al. 2009), confirming the previous finding by Besonen et al. (2008). A hurricane is a tropical cyclone in the North Atlantic–West Indies region with near-surface wind speed equal to or larger than 64 knots or about 119 $\mathrm{m\,s^{-1}}$ (Elsner and Kara 1999). There is a considerable, partly heated debate in the scientific literature, before and after the Katrina hurricane in August 2005, on the trend in hurricane risk during the twentieth century. Papers on data and analysis include Landsea (1993), Bengtsson et al. (1996), Landsea et al. (1996, 1997), Michener et al. (1997), Wilson (1997), Pielke and Landsea (1998), Elsner et al. (1999), Landsea et al. (1999), Easterling et al. (2000), Meehl et al. (2000), Goldenberg et al. (2001), Cutter and Emrich (2005), Emanuel (2005), Pielke et al. (2005), Elsner (2006), Mann and Emanuel (2006), Chang and Guo (2007), Holland (2007), Landsea (2007), Mann et al. (2007a,b), Nyberg et al. (2007), Elsner et al. (2008), Landsea et al. (2008), Vecchi and Knutson (2008), Knutson et al. (2010) and Landsea et al. (2010). While the issue of the trend seems not resolved, it appears clear that (1) economic losses are not a good proxy variable of hurricane occurrence or magnitude and (2) there is room for enhancing the analyses by means of advanced statistical methods.

Heatwaves are events of extreme temperature lasting several days to weeks. An example is the summer heat 2003 in Europe (Beniston 2004). To capture the intensity and duration aspects of a heatwave, various index variables (Kyselý 2002; Meehl and Tebaldi 2004; Khaliq et al. 2005; Alexander et al. 2006; Della-Marta et al. 2007) can be constructed from measured daily temperature series. A direct approach is the exceedance product (Kürbis et al. 2009), an index variable formed by multiplying the exceedance of a previous record temperature by the number of days an exceedance occurs within a summer season. Kürbis et al. (2009) devise a hypothesis test based on MBB resampling to evaluate trends in the exceedance product and apply it to long instrumental records from Potsdam (1893–2005) and Prague–Klementinum (1775–2004). An open research field is the analysis of the distributional properties of functionals like the heatwave index variables within the context of multivariate extremes (Beirlant et al. 2004: Chapters 8 and 9 therein). In an application to daily minimum temperature from a station in Ohio, interval 1893–1987, Smith et al. (1997) studied various functionals such as the length of a cluster of cold extremes.

Applications of a fitted inhomogeneous Poisson process with bootstrap confidence band to extreme events in the climate system include the following. Solow (1991) studied explosive volcanism in the northern hemisphere, 1851–1985, and linked the upwards trend in occurrence rate to the increase in northern hemisphere temperature. Mudelsee et al. (2006) estimated flood risk of the German river Werra over the past 500 years and found trends that partly deviate from trends of neighboured rivers Elbe and Oder (Mudelsee et al. 2003). This demonstrates the spatial variability of river flood risk. Fleitmann et al. (2007b) explored, via Ba/Ca proxy evidence from a coral, events of extreme soil erosion in Kenya, 1700–2000, and detected upwards trends that set in around 1900, after the colonization. Girardin et al. (2006b) inferred dendroclimatically a record of wildfires in Canada that goes back to 1769. Augmenting this data set with other series from the region and climate model output, Girardin and Mudelsee (2008) studied past and possible future (up to 2100) trends in wildfire risk and conclude that past high levels ($\widehat{\lambda}(T) \approx 0.2\,\mathrm{a}^{-1}$) may again be reached. Abram et al. (2008) explored the Indian Ocean Dipole (IOD, east–west sea-surface temperature gradient), 1846–2008, using coral proxy evidence and find an increase in occurrence of extreme IOD events during the past decades.

6.6 Technical issues

Maximum likelihood estimation of the GEV distribution has the following regularity conditions (Smith 1985):

- for $\xi > -0.5$, the estimators have the asymptotic properties of multivariate normality with the covariance matrix as described in Section 6.2.1.2;

- for $-1 < \xi \leq -0.5$, the estimators may exist but do not have the asymptotic properties;
- for $\xi \leq -1$, consistent maximum likelihood estimators do not exist.

The log-likelihood function to be employed for the GEV model with $\xi = 0$ ("Gumbel likelihood") is (Coles 2001b),

$$\ln\left[L(\mu, \sigma)\right] = -m \ln(\sigma) - \sum_{j=1}^{m} y(j) - \sum_{j=1}^{m} \exp\left[-y(j)\right], \tag{6.59}$$

where

$$y(j) = \left[\frac{x_{\text{out}}(j) - \mu}{\sigma}\right]. \tag{6.60}$$

Kharin and Zwiers (2005: Appendix therein) describe details (starting values, local minima) of the numerical maximization of the log-likelihood function of the GEV model. Van Montfort and Witter (1985: Appendix B therein) do similar for the GP model.

The **digamma function** $\Psi(x)$ is the logarithmic derivative of the gamma function, $\Psi(x) = d \ln\left[\Gamma(x)\right]/dx$. See Abramowitz and Stegun (1965: Section 6.3 therein) for more details on the digamma function.

The **simplex method** is a numerical search technique applicable to optimization problems (Press et al. 1992: Section 10.4 therein) such as high-dimensional maximum likelihood estimation. Consider a space of dimension (number of estimation parameters) k. A simplex is a non-degenerate geometric figure spanned by $k+1$ points (starting values) in the space. The task is to move and shrink the simplex in the space in a way that it includes with sufficient precision the maximum likelihood solution. The method does not perform gradient calculation for deciding how to move/shrink, it selects among possible steps more in a brute-force manner. It may be slower than gradient search techniques but, on the other hand, also more robust.

Gaussian kernel functions for occurrence rate estimation offer the advantage that Eq. (6.33) can be computed fast in the Fourier domain (Silverman 1982; Jones and Lotwick 1984). Fourier transform algorithms (FFT) are abundant (Monro 1975, 1976; Press et al. 1996).

Cross-validation function evaluation for kernel occurrence rate estimation (Eqs. 6.37 and 6.38) is computationally expensive. The second term on the right-hand side of Eq. (6.37) constitutes a sum of exponentials over a rectangle ($j = 1, \ldots, m; k = 1, \ldots, m^{\dagger}$). Because of the symmetry only approximately half of the summands have to be determined. The summands near the upper left or lower right corner of the rectangle are small ($\propto \exp\{-[(T_{\text{out}}(j) - T^{\dagger}_{\text{out}}(k))/h]^2/2\}$), the summands near the 1:1 line are around unity. The following approximation could in principle reduce further computing costs. Calculate the summands only in the intermediate range, set the summands near ("near" defined by machine precision) the 1:1 line equal to unity, and omit the summands near the two corners. However, for typical sample sizes, m, in climatology (less than a few thousand) and typical machine precisions (PC and workstation systems with 32- or 64-bit processors), the reduction is negligible (Mudelsee 2001, unpublished manuscript).

Software tools for fitting stationary extreme value distributions to data are abundant, while programs for estimating nonstationary extreme value models are rare.

MLEGEV is a Fortran subroutine (Hosking 1985; Macleod 1989) for maximum likelihood estimation of the parameters of the stationary GEV model. It serves as a basis for many software tools developed later. A download site is `http://lib.stat.cmu.edu/apstat/215` (14 July 2008).

Statistical Modelling in Hydrology is the title of a book (Clarke 1994) that contains Genstat and Matlab programs implementing various estimation methods for stationary extreme value distributions.

Xtremes (Reiss and Thomas 1997) is a compiled Windows software package for analysing stationary extreme value models by means of several estimation methods, bootstrap resampling and model suitability tests.

Flood Frequency Analysis is the title of a book (Rao and Hamed 2000) that includes Matlab programs for maximum likelihood and PWM estimation of stationary GEV and GP distributions.

WAFO is a Matlab package (WAFO group 2000) that includes maximum likelihood and PWM estimation of stationary GEV and GP distributions. The software can be downloaded from the following site: `http://www.maths.lth.se/matstat/wafo` (7 July 2008).

The **ismev** package for the R computing environment supports the computations carried out in the book by Coles (2001b). It is available at `http://cran.r-project.org/web/packages/ismev` (7 July 2008).

The **evd** package for the R computing environment augments ismev. It is available at `http://cran.r-project.org/web/packages/evd` (7 July 2008).

EVIM is a Matlab package (Gençay et al. 2001) for stationary extreme value analysis: declustering, fitting GEV and GP models and assessing suitability. It is available at the following internet address: `http://www.bilkent.edu.tr/~faruk/evim.htm` (7 July 2008).

Dataplot is a software (Unix, Linux, Windows) for fitting stationary extreme value distributions with bootstrap CIs and performing model suitability analysis. It can be obtained from the following internet address: `http://www.itl.nist.gov/div898/winds/dataplot.htm` (4 July 2008).

Extremes is a software tool (R language), based on ismev and evd routines, for analysing interactively stationary extreme value models. It is available at `http://www.isse.ucar.edu/extremevalues/evtk.html` (4 July 2008).

GEVFIT is a module (Stata computing environment) for maximum likelihood estimation of a GEV model. It resides on the following internet address: `http://ideas.repec.org/c/boc/bocode/s456892.html` (4 July 2008).

The **declustering** method for GP estimation (Fawcett and Walshaw 2006) was implemented as an R code. It is available at the internet address `http://www.mas.ncl.ac.uk/~nlf8` (25 May 2010).

VGAM is a mixed package (C, Fortran 77 and 90, S-Plus/R) for fitting a wide class of regression models, so-called vector generalized additive models (Yee and Wild 1996), to time series. This includes not only estimation of stationary extreme value distributions but also quantile regression (nonstationarity). The software can be downloaded from `http://www.stat.auckland.ac.nz/~yee/VGAM` (7 July 2008).

Statistics of Extremes is the title of a book (Beirlant et al. 2004) that is accompanied by a set of routines written in S-Plus and FORTRAN 77. Besides fitting stationary models and estimating distribution parameters and quantiles, the routines for Chapter 7 of the book allow for covariates and may be used for fitting nonstationary models. The software resides at `http://lstat.kuleuven.be/Wiley` (7 July 2008).

Caliza is a Fortran 90 software for fitting a nonstationary inhomogeneous Poisson process with bootstrap confidence band to POT data. It includes CLIM-X-DETECT for threshold selection and extremes detection (Chapter 4). Caliza also performs the Cox–Lewis test for trends in the occurrence of extreme events. A demo version is available at the web site for this book.

Part III

Bivariate Time Series

Chapter 7

Correlation

The correlation measures how strong a coupling is between the noise components of two processes, $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$. Using a bivariate time series sample, $\{t(i), x(i), y(i)\}_{i=1}^{n}$, this measure allows to study the relationship between two climate variables, each described by its own climate equation (Eq. 1.2).

Pearson's correlation coefficient (Section 7.1) estimates the degree of the *linear* relationship. It is one of the most widely used statistical quantities in all branches of the natural sciences. Spearman's correlation coefficient (Section 7.2) estimates the degree of the *monotonic* relationship. Although clearly less often used, it offers robustness against violations of the Gaussian assumption, as also the Monte Carlo experiments (Section 7.3) show.

Explorative climate data analyses should strongly benefit from correlation estimates that are supported by a CI and not only a P-value of a test of the null hypothesis of no correlation. It is then possible to take several pairs of variables and rank the associations. One finding may be, for example, that global temperature changes are stronger associated to variations of CO_2 than to those of solar activity (background material). The challenge of providing accurate CIs is met by pairwise bootstrap resampling (MBB or ARB), which takes into account the serial dependence structures of both climate processes.

A second, rarely mentioned challenge appears when the processes differ in their sampling times (Section 7.5). This book introduces two novel estimators, denoted as binned and synchrony correlation, respectively. These are able (and outperform interpolation) to recover correlation information under the conditions of (1) persistence in the system, which is realistic for climate, and (2) not too large spacings of the time series.

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_7,

7.1 Pearson's correlation coefficient

Let us assume in this chapter, for simplicity of exposition, that the climate process, $X(i)$, has a constant trend function at level μ_X, a constant variability, S_X, and no outlier component. In discrete time,

$$\begin{aligned} X(i) &= X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i) \\ &= \mu_X + S_X \cdot X_{\text{noise}}(i). \end{aligned} \tag{7.1}$$

Assume analogously for the second climate process, $Y(i)$, which is on the same time points, $T(i)$, as the first climate process,

$$Y(i) = \mu_Y + S_Y \cdot Y_{\text{noise}}(i). \tag{7.2}$$

The correlation coefficient is then defined as

$$\rho_{XY} = \frac{E\left[\{X(i) - \mu_X\} \cdot \{Y(i) - \mu_Y\}\right]}{S_X \cdot S_Y}. \tag{7.3}$$

The correlation measures the degree of the linear relationship between the variables X and Y; ρ_{XY} is between -1 ("anti-correlation") and 1.

For convenience of presentation we introduce here the correlation operator,

$$CORR\left[X(i), Y(i)\right] = \frac{COV\left[X(i), Y(i)\right]}{\left\{VAR\left[X(i)\right] \cdot VAR\left[Y(i)\right]\right\}^{1/2}}. \tag{7.4}$$

The definition of the correlation coefficient is thus based on the assumption of time-constancy of $CORR\left[X(i), Y(i)\right] = \rho_{XY}$.

Let $\{X(i), Y(i)\}_{i=1}^{n}$ be a bivariate sample (process level). Pearson's (1896) estimator of ρ_{XY} is

$$r_{XY} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X(i) - \bar{X}}{S_{n,X}}\right) \cdot \left(\frac{Y(i) - \bar{Y}}{S_{n,Y}}\right), \tag{7.5}$$

where

$$\bar{X} = \sum_{i=1}^{n} X(i) \,/\, n \tag{7.6}$$

and

$$\bar{Y} = \sum_{i=1}^{n} Y(i) \,/\, n \tag{7.7}$$

are the sample means and

$$S_{n,X} = \left\{ \sum_{i=1}^{n} \left[X(i) - \bar{X} \right]^2 / \, n \right\}^{1/2} \tag{7.8}$$

and

$$S_{n,Y} = \left\{ \sum_{i=1}^{n} \left[Y(i) - \bar{Y} \right]^2 / \, n \right\}^{1/2} \tag{7.9}$$

are the sample standard deviations calculated with the denominator n (instead of $n - 1$). On the sample level, given a bivariate sample $\{x(i), y(i)\}_{i=1}^{n}$, plug in those values for $X(i)$ and $Y(i)$ in Eqs. (7.5), (7.6), (7.7), (7.8) and (7.9). The estimator r_{XY} is called Pearson's correlation coefficient. Also r_{XY} is between -1 and 1.

7.1.1 Remark: alternative correlation measures

It is of course possible to employ other estimators. For example, S_{n-1} (Eq. 3.19) may replace S_n for estimating S_X or S_Y, leading to an (unfortunate) correlation estimator that can have values < -1 or > 1. Another option may be to subtract the sample medians (Galton 1888) and not the sample means (Eqs. 7.6 and 7.7). More complex examples arise when time-dependent trend functions are subtracted or time-dependent variability functions used for normalization. Such cases may be relevant for climate time series analysis. All those examples lead to other correlation measures than ρ_{XY} and other correlation estimators than r_{XY}. Their properties and CI performance can in principle be studied in the same manner with Monte Carlo methods. Here we focus on r_{XY}, stationary trends and variabilities. Another measure (Spearman's) is analysed in Section 7.2.

7.1.2 Classical confidence intervals, non-persistent processes

Let $X(i)$ and $Y(i)$ both be a stochastic process without persistence or "memory." Let further $X(i)$ and $Y(i)$ both have a Gaussian distributional shape; their joint distribution is then denoted as bivariate normal or binormal distribution (Section 7.1.3.1). The PDF of Pearson's correlation coefficient is then (Fisher 1915):

$$f(r_{XY}) = \frac{\left(1-\rho_{XY}^2\right)^{(n-1)/2} \left(1-r_{XY}^2\right)^{(n-4)/2}}{\sqrt{\pi}\,\Gamma[(n-1)/2]\,\Gamma[(n-2)/2]} \times \sum_{j=0}^{\infty} \frac{\{\Gamma[(n-1+j)/2]\}^2}{j!} \left(2\,\rho_{XY}\,r_{XY}\right)^j . \quad (7.10)$$

Numerous discussions on, and much work in the implementation of, this celebrated formula exist in statistical science. Hotelling (1953) gave approximations for the moments of r_{XY}. In particular,

$$\mathrm{bias}_{r_{XY}} = \left(1-\rho_{XY}^2\right)\left[-\frac{\rho_{XY}}{2n} + \frac{\rho_{XY} - 9\rho_{XY}^3}{8n^2} + \frac{\rho_{XY} + 42\rho_{XY}^3 - 75\rho_{XY}^5}{16n^3} + \mathcal{O}\left(n^{-4}\right)\right] \quad (7.11)$$

and

$$\mathrm{se}_{r_{XY}} = \left(1-\rho_{XY}^2\right)\left[\frac{1}{n^{1/2}} + \frac{11\rho_{XY}^2}{4n^{3/2}} - \frac{192\rho_{XY}^2 - 479\rho_{XY}^4}{32n^{5/2}} + \mathcal{O}\left(n^{-7/2}\right)\right] . \quad (7.12)$$

Regarding the focus of this chapter, CI construction, it is common practice to employ Fisher's (1921) transformation. The quantity

$$z = \tanh^{-1}\left(r_{XY}\right) \quad (7.13)$$

approaches with increasing n a normal distributional shape considerably faster than r_{XY}, particularly when $\rho_{XY} \neq 0$. Fisher's z has for large n the following properties (Rodriguez 1982):

$$E[z] \approx \tanh^{-1}\left(\rho_{XY}\right) \quad (7.14)$$

and

$$\mathrm{se}_z \approx (n-3)^{-1/2} . \quad (7.15)$$

This leads to the approximate classical CI for r_{XY},

$$\mathrm{CI}_{r_{XY},1-2\alpha} = \Big[\tanh\left[z + z(\alpha)\cdot \mathrm{se}_z\right] ; \tanh\left[z - z(\alpha)\cdot \mathrm{se}_z\right]\Big], \quad (7.16)$$

where $z(\alpha)$ is the percentage point of the normal distribution (Section 3.9).

If we keep the assumption of absence of persistence for processes $X(i)$ and $Y(i)$, but drop the Gaussian assumption, less is known, and no exact formula for the distribution of r_{XY} has been found. One recipe is then to work with higher-moment properties of the distributions and approximate solutions (Section 7.6). The alternative recipe is to use still the formulas for the Gaussian case (Eqs. 7.13, 7.14, 7.15 and 7.16) and assume robustness of this method. Johnson et al. (1995: Chapter 32 therein) give an account of the bewildering diversity of opinions in the research literature on the suitability of this approach.

7.1.3 Bivariate time series models

A bivariate model describes not only the distributional and persistence properties of two processes, $X(i)$ and $Y(i)$, but also the correlation between them. The bivariate white-noise model characterizes persistence-free processes and serves to build bivariate autoregressive and higher-order processes.

7.1.3.1 Bivariate white noise

The bivariate Gaussian white noise model is given by

$$\begin{aligned} X(i) &= \mathcal{E}^{X}_{\mathrm{N}(0,\,1)}(i), \qquad i = 1, \ldots, n, \\ Y(i) &= \mathcal{E}^{Y}_{\mathrm{N}(0,\,1)}(i), \qquad i = 1, \ldots, n. \end{aligned} \tag{7.17}$$

The Gaussian random processes $\mathcal{E}^{X}_{\mathrm{N}(0,\,1)}(\cdot)$ and $\mathcal{E}^{Y}_{\mathrm{N}(0,\,1)}(\cdot)$ are indexed. The correlation coefficient between them is denoted as $\rho_{\mathcal{E}}$.

The moments of this special case of the bivariate Gaussian white noise model are by definition

$$E\left[X(i)\right] = E\left[Y(i)\right] = 0, \tag{7.18}$$

$$VAR\left[X(i)\right] = VAR\left[Y(i)\right] = 1 \tag{7.19}$$

and

$$CORR\left[X(i), Y(i)\right] = \rho_{XY} = \rho_{\mathcal{E}}. \tag{7.20}$$

In the general case, $X(i)$ has mean μ_X and variance S_X^2, and $Y(i)$ has mean μ_Y and variance S_Y^2. The binormal PDF of $X(i)$ and $Y(i)$ (Section 7.6) is uniquely determined by the means, variances and correlation.

The bivariate lognormal white noise model is given by

$$\begin{aligned} X(i) &= \exp\left[\mathcal{E}^{X}_{\mathrm{N}(0,\,1)}(i)\right], \qquad i = 1, \ldots, n, \\ Y(i) &= \exp\left[\mathcal{E}^{Y}_{\mathrm{N}(0,\,1)}(i)\right], \qquad i = 1, \ldots, n. \end{aligned} \tag{7.21}$$

The moments are (Section 3.9)

$$E[X(i)] = E[Y(i)] = \exp(1/2), \tag{7.22}$$

$$VAR[X(i)] = VAR[Y(i)] = \mathrm{e}(\mathrm{e} - 1) \tag{7.23}$$

and (Section 7.6)

$$CORR\,[X(i), Y(i)] = \rho_{XY} = [\exp(\rho_{\mathcal{E}}) - 1]/(\mathrm{e} - 1). \tag{7.24}$$

7.1.3.2 Bivariate first-order autoregressive process

Extending the univariate Gaussian AR(1) process (Section 2.1) to two dimensions yields a simple bivariate persistence model. The version for even time spacing is

$$\begin{aligned} X(1) &= \mathcal{E}^{X}_{\mathrm{N}(0,\,1)}(1), \\ Y(1) &= \mathcal{E}^{Y}_{\mathrm{N}(0,\,1)}(1), \\ X(i) &= a_X \cdot X(i-1) + \mathcal{E}^{X}_{\mathrm{N}(0,\,1-a_X^2)}(i), \qquad i = 2, \ldots, n, \\ Y(i) &= a_Y \cdot Y(i-1) + \mathcal{E}^{Y}_{\mathrm{N}(0,\,1-a_Y^2)}(i), \qquad i = 2, \ldots, n, \end{aligned} \tag{7.25}$$

where the white-noise innovation terms are correlated as

$$\begin{aligned} CORR\left[\mathcal{E}^{X}_{\mathrm{N}(0,\,1)}(1), \mathcal{E}^{Y}_{\mathrm{N}(0,\,1)}(1)\right] &= \rho_{\mathcal{E}}, \\ CORR\left[\mathcal{E}^{X}_{\mathrm{N}(0,\,1)}(i), \mathcal{E}^{Y}_{\mathrm{N}(0,\,1)}(i)\right] &= \frac{1 - a_X \cdot a_Y}{\left[\left(1 - a_X^2\right)\left(1 - a_Y^2\right)\right]^{1/2}}\, \rho_{\mathcal{E}}, \\ & \quad\; i = 2, \ldots, n, \\ CORR\left[\mathcal{E}^{X}_{\mathrm{N}(0,\,1)}(i), \mathcal{E}^{Y}_{\mathrm{N}(0,\,1)}(j)\right] &= 0, \qquad i, j = 1, \ldots, n, \qquad i \neq j. \end{aligned} \tag{7.26}$$

This model requires the autocorrelation parameters a_X and a_Y to have the same sign.

The bivariate AR(1) process for even spacing (Eq. 7.25) is strictly stationary. Its properties are

$$E\left[X(i)\right] = E\left[Y(i)\right] = 0, \tag{7.27}$$

$$VAR\left[X(i)\right] = VAR\left[Y(i)\right] = 1 \tag{7.28}$$

and

$$CORR\left[X(i), Y(i)\right] = \rho_{XY} = \rho_{\mathcal{E}}. \tag{7.29}$$

The bivariate AR(1) process for uneven time spacing is obtained in the usual manner: replace a_X by $\exp\{-[T(i) - T(i-1)]/\tau_X\}$ and a_Y by $\exp\{-[T(i) - T(i-1)]/\tau_Y\}$. This leads to heteroscedastic innovation terms, as already noticed in the univariate case. The model is given in the background material (Eq. 7.53).

To summarize, the simple bivariate Gaussian AR(1) model, written for even (Eq. 7.25) or uneven spacing, has three parameters. Two describe the persistence properties of the processes $X(i)$ and $Y(i)$, one describes the correlation between both. The more general case in form of means unequal to zero and variances unequal to unity, is less relevant in the context of this chapter because the correlation estimation (Eq. 7.5) eliminates such effects.

Interesting is, however, the general formulation of the bivariate AR(1) model, where the variable at a time, say, $X(i)$, depends not only on its own immediate past, $X(i-1)$, but also on the past of the second variable, $Y(i-1)$. This general model has more than three parameters, and it can give rise to "identifiability" problems (Priestley 1981: Section 9.4 therein). These difficulties how to uniquely determine the number of parameters and their values do certainly not decrease when considering uneven instead of even spacing, see the univariate embedding problem (Section 2.1.2.1). We therefore ignore the general formulation and avoid the identifiability and embedding problems when describing parametric bootstrap resampling (Sections 7.1.5.2 and 7.2.3.2) and designing Monte Carlo experiments (Section 7.3). A further conclusion is that the possible existence, and theoretical applicability to climatology, of the general bivariate AR(1) model with more than three parameters supports the selection of the nonparametric bootstrap resampling algorithms (Sections 7.1.5.1 and 7.2.3.1).

7.1.4 Classical confidence intervals, persistent processes

If we continue from before the excursion on bivariate stochastic processes and drop not only the Gaussian assumption about the distributional shape of $X(i)$ and $Y(i)$, but also leave away the assumption that

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Pearson's r_{XY} (Eq. 7.5)	
Step 3	Fisher's z-transformation	$z = \tanh^{-1}(r_{XY})$
Step 4	Estimated, bias-corrected	
	persistence time, process $X(i)$,	$\widehat{\tau}'_X$
	using mean-detrended	
	time series, $\{x(i) - \bar{x}\}_{i=1}^{n}$	
Step 5	Analogously, process $Y(i)$	$\widehat{\tau}'_Y$
Step 6	Estimated, bias-corrected	
	equivalent autocorrelation	
	coefficient, process $X(i)$	$\widehat{\bar{a}}'_X = \exp\left(-\bar{d}/\widehat{\tau}'_X\right)$
Step 7	Analogously, process $Y(i)$	$\widehat{\bar{a}}'_Y = \exp\left(-\bar{d}/\widehat{\tau}'_Y\right)$
Step 8	Effective data size,	n'_ρ
	obtained by plugging in	
	$\widehat{\bar{a}}'_X$ for a_X and $\widehat{\bar{a}}'_Y$ for a_Y	
	in Eq. (2.38)	
Step 9	Approximate, classical	
	normal CI for r_{XY},	$\mathrm{CI}_{r_{XY},1-2\alpha} =$
	obtained from	$\Big[\tanh\left[z + z(\alpha)\cdot\left(n'_\rho - 3\right)^{-1/2}\right];$
	re-transforming z	$\tanh\left[z - z(\alpha)\cdot\left(n'_\rho - 3\right)^{-1/2}\right]\Big]$

Algorithm 7.1. Construction of classical confidence intervals for Pearson's correlation coefficient, bivariate AR(1) model. Steps 3 and 9: z is Fisher's transformed correlation, $z(\alpha)$ is the percentage point of the normal distribution.

these processes are persistence-free white noise, we approach the reality for the majority of processes occurring in the climate system. A classical CI for r_{XY}, which is approximate, is obtained readily by invoking the effective data size (Chapter 2). The complete procedure is given (Algorithm 7.1) for a bivariate AR(1) process on an unevenly spaced time grid.

7.1.5 Bootstrap confidence intervals

For processes with persistence, possibly of more complex form than AR(1), and distributional shapes more complex than Gaussian, that means, for the majority of climate processes, the classical CI for r_{XY} is not exact but approximate. The accuracy of it is expected to depend on how strongly the properties of the process deviate from the assumed properties. This brings naturally the bootstrap into play. Two algorithms are analysed that resample data pairs, $(x(i), y(i))$, namely the pairwise-MBB and the pairwise-ARB algorithm. Both resampling types serve to construct bootstrap CIs.

7.1.5.1 Pairwise-moving block bootstrap

The pairwise-MBB (Algorithm 7.2), already introduced for regression (Section 4.1.7.1), extends the ability of the MBB to preserve persistence of a single process, $X(i)$, over the length of a block, to the bivariate setting. Because also $Y(i)$ can exhibit the memory phenomenon, expressed by the persistence time, τ_Y, block length selection may be more difficult in the bivariate than in the univariate setting. Mudelsee (2003) suggested the block length selector

$$l_{\mathrm{opt}} = 4\ \max\left(\tau_X, \tau_Y\right). \tag{7.30}$$

In practice, the (bias-corrected) persistence-time estimates $\widehat{\tau}'_X$ and $\widehat{\tau}'_Y$ are plugged in. Although Monte Carlo experiments (Mudelsee 2003) revealed acceptable coverage performance of BCa CIs for r_{XY}, it should be worth testing other block length selectors. This is also in line with the "optimal estimation" strategy (Section 6.2.7).

The second selector is, thus, based on combining the bias-corrected equivalent autororrelation coefficients, $\widehat{\bar{a}}'_X$ and $\widehat{\bar{a}}'_Y$, in a new expression,

$$\widehat{\bar{a}}'_{XY} = \left[\widehat{\bar{a}}'_X \cdot \widehat{\bar{a}}'_Y\right]^{1/2}, \tag{7.31}$$

and employing the univariate selector (Eq. 3.28). This yields

$$l_{\mathrm{opt}} = NINT\left\{\left[6^{1/2} \cdot \widehat{\bar{a}}'_{XY} \Big/ \left(1 - \widehat{\bar{a}}'^{\,2}_{XY}\right)\right]^{2/3} \cdot n^{1/3}\right\}. \tag{7.32}$$

Another technical measure is the z-transformation. First, a bootstrap CI is constructed for z and then the CI bounds are re-transformed to

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Pearson's r_{XY} (Eq. 7.5)	
Step 3	Fisher's z-transformation	$z = \tanh^{-1}(r_{XY})$
Step 4	Estimated, bias-corrected	
	persistence time, process $X(i)$,	$\widehat{\tau}'_X$
	using mean-detrended	
	time series, $\{x(i) - \bar{x}\}_{i=1}^{n}$	
Step 5	Analogously, process $Y(i)$	$\widehat{\tau}'_Y$
Step 6	Select block length	l
Step 7	Apply MBB with l	$\left\{x^{*b}(i)\right\}_{i=1}^{n} = \left\{x(f(i))\right\}_{i=1}^{n}$
	(Algorithm 3.1) to x values	(b, counter)
Step 8	Overtake bootstrap index	$f(i)$
	for resampled y values	$\left\{y^{*b}(i)\right\}_{i=1}^{n} = \left\{y(f(i))\right\}_{i=1}^{n}$
Step 9	Resample	$\left\{x^{*b}(i), y^{*b}(i)\right\}_{i=1}^{n}$
Step 10	Bootstrap replications,	
	Pearson's r_{XY} and Fisher's z	$r_{XY}^{*b}, \quad z^{*b} = \tanh^{-1}\left(r_{XY}^{*b}\right)$
Step 11	Go to Step 7 until $b = B$	
	(usually $B = 2000$)	
	replications exist	$\left\{z^{*b}\right\}_{b=1}^{B}$
Step 12	Calculate CI (Section 3.4)	
	for Fisher's z	$\mathrm{CI}_{z,1-2\alpha} = \left[z_{\mathrm{l}}; z_{\mathrm{u}}\right]$
Step 13	Re-transform lower and upper	
	endpoints to obtain	
	pairwise-MBB CI for r_{XY}	$\mathrm{CI}_{r_{XY},1-2\alpha} = \left[\tanh(z_{\mathrm{l}}); \tanh(z_{\mathrm{u}})\right]$

Algorithm 7.2. Construction of bootstrap confidence intervals for Pearson's correlation coefficient, pairwise-MBB resampling. Step 8: By overtaking the random bootstrap index $f(i) \in \{1, \ldots, n\}$ from x-resampling for y-resampling, $(x(j), y(j))$ pairs are resampled.

obtain a CI for r_{XY}. The idea (Hall et al. 1989) is to enhance CI construction by supplying replications (z^*) that are in shape closer to a normal distribution than the alternative (r^*_{XY}).

7.1.5.2 Pairwise-autoregressive bootstrap

The pairwise-ARB (Algorithm 7.3) resamples pairs $(x(i), y(i))$ by overtaking the random index from x-resampling for y-resampling. Also this algorithm employs the z-transformation.

The coverage performance of CIs from the pairwise-ARB and the pairwise-MBB algorithms are explored by means of Monte Carlo experiments (Section 7.3).

7.2 Spearman's rank correlation coefficient

Consider instead of process $X(i)$ its rank, $R(i) = \text{rank}[X(i)]$. For example, if from all $X(i)$, it is $X(7)$ that has the lowest value, then $R(7) = 1$. Let analogously $S(i) = \text{rank}[Y(i)]$. The rank correlation coefficient is then given by

$$\rho_{\text{S}} = \frac{E\left[\{R(i) - \mu_R\} \cdot \{S(i) - \mu_S\}\right]}{\{VAR\left[R(i)\right] \cdot VAR\left[S(i)\right]\}^{1/2}}, \tag{7.33}$$

where $\mu_R = E[R(i)]$ and $\mu_S = E[S(i)]$. That means, the rank correlation coefficient between X and Y is equal to the correlation coefficient between the rank of X and the rank of Y. The rank correlation measures the degree of the monotone relationship between X and Y; also ρ_{S} is between -1 and 1.

Strictly speaking, Eq. (7.33) applies only to discrete random variables $X(i)$ and $Y(i)$ because continuous variables cannot be ranked (Gibbons and Chakraborti 2003). However, for continuous variables it is possible to define ρ_{S} as the grade correlation coefficient (background material). This distinction between rank and grade correlation coefficient is of theoretical importance (Kruskal 1958) but of limited practical relevance in the context of this book.

Spearman's (1904, 1906) estimator of ρ_{S} uses a bivariate sample as follows:

$$r_{\text{S}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{R(i) - \bar{R}}{S_{n,R}} \right) \cdot \left(\frac{S(i) - \bar{S}}{S_{n,S}} \right), \tag{7.34}$$

where $\bar{R}$ and $\bar{S}$ are the sample means, and $S_{n,R}$ and $S_{n,S}$ are the sample standard deviations calculated with the denominator n.

Assume the absence of ties in $R(i)$ and $S(i)$. This means a negligible loss of generality for continuous climate variables. The case of a binary variable, which is relevant for analysing the climate extremes component,

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Pearson's r_{XY} (Eq. 7.5)	
Step 3	Fisher's z-transformation	$z = \tanh^{-1}(r_{XY})$
Step 4	Estimated, bias-corrected persistence time, process $X(i)$, using mean-detrended time series, $\{x(i) - \bar{x}\}_{i=1}^{n}$	$\widehat{\tau}'_X$
Step 5	Analogously, process $Y(i)$	$\widehat{\tau}'_Y$
Step 6	Climate equation residuals, process $X(i)$	$\{q_X(i)\}_{i=1}^{n} = \{[x(i) - \bar{x}] / s_{n,X}\}_{i=1}^{n}$
Step 7	Analogously, process $Y(i)$	$\{q_Y(i)\}_{i=1}^{n} = \{[y(i) - \bar{y}] / s_{n,Y}\}_{i=1}^{n}$
Step 8	Abbreviation, process $X(i)$	$\widehat{a}'_X(i) = \exp\{-[t(i) - t(i-1)]/\widehat{\tau}'_X\}$, $i = 2, \ldots, n$
Step 9	Analogously, process $Y(i)$	$\widehat{a}'_Y(i) = \exp\{-[t(i) - t(i-1)]/\widehat{\tau}'_Y\}$, $i = 2, \ldots, n$
Step 10	White-noise residuals, process $X(i)$	$\epsilon_X(i) = [q_X(i) - \widehat{a}'_X(i) \cdot q_X(i-1)] \times \left\{1 - [\widehat{a}'_X(i)]^2\right\}^{-1/2}$, $i = 2, \ldots, n$
Step 11	Analogously, process $Y(i)$	$\epsilon_Y(i) = [q_Y(i) - \widehat{a}'_Y(i) \cdot q_Y(i-1)] \times \left\{1 - [\widehat{a}'_Y(i)]^2\right\}^{-1/2}$, $i = 2, \ldots, n$
Step 12	Centering, process $X(i)$	$\tilde{\epsilon}_X(i) = \epsilon_X(i) - \sum_{i=2}^{n} \epsilon_X(i)/(n-1)$
Step 13	Analogously, process $Y(i)$	$\tilde{\epsilon}_Y(i) = \epsilon_Y(i) - \sum_{i=2}^{n} \epsilon_Y(i)/(n-1)$

Algorithm 7.3. Construction of bootstrap confidence intervals for Pearson's correlation coefficient, pairwise-ARB resampling. Steps 6 and 7 employ the sample standard deviations (sample level), $s_{n,X} = \left\{\sum_{i=1}^{n} [x(i) - \bar{x}]^2 / n\right\}^{1/2}$ and $s_{n,Y} = \left\{\sum_{i=1}^{n} [y(i) - \bar{y}]^2 / n\right\}^{1/2}$, calculated with the denominator n.

Step 14	Draw $\tilde{\epsilon}_X^{*b}(j), j = 2, \ldots, n$,	
	with replacement from	$\{\tilde{\epsilon}_X(i)\}_{i=2}^{n}$
	(b, counter), process $X(i)$	
Step 15	Bootstrap index, $f(j)$,	$f(j) \in \{1, \ldots, n\}, j = 2, \ldots, n$
	process $X(i)$, defined via	$\tilde{\epsilon}_X^{*b}(j) = \tilde{\epsilon}_X(f(j))$
Step 16	Resampled climate equation	$q_X^{*b}(1)$ drawn from $\{q_X(i)\}_{i=1}^{n}$,
	residuals, process $X(i)$	$q_X^{*b}(i) = \widehat{a}'_X(i) \cdot q_X^{*b}(i-1)$
		$+ \left\{1 - [\widehat{a}'_X(i)]^2\right\}^{1/2} \cdot \tilde{\epsilon}_X^{*b}(i)$,
		$i = 2, \ldots, n$
Step 17	Bootstrap index, $f(j = 1)$,	$f(1) \in \{1, \ldots, n\}$
	process $X(i)$, defined via	$q_X^{*b}(1) = q_X(f(1))$
Step 18	Overtake random	
	bootstrap index,	
	process $Y(i)$	$\left\{\tilde{\epsilon}_Y^{*b}(i)\right\}_{i=2}^{n} = \{\tilde{\epsilon}_Y(f(i))\}_{i=2}^{n}$
Step 19	Resampled climate equation	$q_Y^{*b}(1) = q_Y(f(1))$,
	residuals, process $Y(i)$	$q_Y^{*b}(i) = \widehat{a}'_Y(i) \cdot q_Y^{*b}(i-1)$
		$+ \left\{1 - [\widehat{a}'_Y(i)]^2\right\}^{1/2} \cdot \tilde{\epsilon}_Y^{*b}(i)$,
		$i = 2, \ldots, n$
Step 20	Resampled data,	
	process $X(i)$	$x^{*b}(i) = \bar{x} + s_{n,X} \cdot q_X^{*b}(i),\ i = 1, \ldots, n$
Step 21	Analogously, process $Y(i)$	$y^{*b}(i) = \bar{y} + s_{n,Y} \cdot q_Y^{*b}(i),\ i = 1, \ldots, n$
Step 22	Resample	$\left\{x^{*b}(i), y^{*b}(i)\right\}_{i=1}^{n}$
Step 23	Bootstrap replications,	
	Pearson's r_{XY} and	
	Fisher's z	$r_{XY}^{*b}, \quad z^{*b} = \tanh^{-1}\left(r_{XY}^{*b}\right)$

Algorithm 7.3. Construction of bootstrap confidence intervals for Pearson's correlation coefficient, pairwise-ARB resampling (continued). Step 18: By overtaking the random bootstrap index $f(i) \in \{1, \ldots, n\}$ from x-resampling for y-resampling, $(x(j), y(j))$ pairs are resampled.

Step 24	Go to Step 14 until $b = B$	
	(usually $B = 2000$)	
	replications exist	$\left\{z^{*b}\right\}_{b=1}^{B}$
Step 25	Calculate CI (Section 3.4)	
	for Fisher's z	$\mathrm{CI}_{z,1-2\alpha} = \left[z_{\mathrm{l}}; z_{\mathrm{u}}\right]$
Step 26	Re-transform lower and upper	
	endpoints to obtain	
	pairwise-ARB CI for r_{XY}	$\mathrm{CI}_{r_{XY},1-2\alpha} = \left[\tanh\left(z_{\mathrm{l}}\right); \tanh\left(z_{\mathrm{u}}\right)\right]$

Algorithm 7.3. Construction of bootstrap confidence intervals for Pearson's correlation coefficient, pairwise-ARB resampling (continued).

is treated in Section 7.6. Since the set of values $R(i)$ and $S(i)$ can take, is known,

$$\bar{R} = \sum_{i=1}^{n} R(i)/n = \bar{S} = \sum_{i=1}^{n} S(i)/n = (n+1)/2 \tag{7.35}$$

and

$$\begin{aligned} S_{n,R} &= \left\{\sum_{i=1}^{n}\left[R(i)-\bar{R}\right]^2/n\right\}^{1/2} = S_{n,S} = \left\{\sum_{i=1}^{n}\left[S(i)-\bar{S}\right]^2/n\right\}^{1/2} \\ &= \left[\left(n^2-1\right)/12\right]^{1/2}. \end{aligned} \tag{7.36}$$

This facilitates the computation of r_{S} (Gibbons and Chakraborti 2003: Section 11.3 therein):

$$r_{\mathrm{S}} = 1 - \frac{6\sum_{i=1}^{n}\left[R(i)-S(i)\right]^2}{n\left(n^2-1\right)}. \tag{7.37}$$

The estimator r_{S} is called Spearman's rank correlation coefficient. Also r_{S} is between -1 and 1.

7.2.1 Classical confidence intervals, non-persistent processes

Let $X(i)$ and $Y(i)$ both be a stochastic process without persistence. Consider the "null case" that the true rank correlation, ρ_{S}, equals zero.

Then each combination of rank $R(i) \in \{1, \ldots, n\}$ and rank $S(i) \in \{1, \ldots, n\}$ has the same probability. The distribution of r_S (Eq. 7.37) is independent of the distributions of $X(i)$ and $Y(i)$; hence r_S is called a distribution-free statistic. The null distribution of r_S can in theory be exactly deduced by means of combinatorics. In practice (finite computing power), this is done only if n is not larger than, say, 15, and approximations are used for larger data sizes (van de Wiel and Di Bucchianico 2001; Gibbons and Chakraborti 2003). The null distribution of r_S is essential for performing statistical tests of H_0: "$\rho_S = 0$."

Consider the "non-null case" that ρ_S is not specified to be zero. This is of relevance within the context of this chapter, CI construction. In principle, the PDF of Spearman's correlation coefficient can be derived exactly via calculating the unequal probabilities of the $(n!)^2$ pairs of $R(i)$ and $S(i)$, given a bivariate distribution of the ranks (Henze 1979). This method is for typical data sizes n in climatology not feasible. The practice employs therefore approximations.

Fieller et al. (1957) suggested use of the z-transformation: the quantity

$$z_S = \tanh^{-1}(r_S) \tag{7.38}$$

approaches with increasing n normal distributional shape and has

$$\mathrm{se}_{z_S} \approx 1.03\,(n-3)^{-1/2}\,. \tag{7.39}$$

The approximation for the expectation of z_S is less accurate and makes the assumption of binormally distributed $(X(i), Y(i))$ with correlation coefficient ρ_{XY}. Then (Fieller et al. 1957)

$$E[z_S] \approx \tanh^{-1}(\bar{r}_S) \;+\; \bar{r}_S\, VAR[r_S] \Big/ \left(1 - \bar{r}_S^2\right)^2 , \tag{7.40}$$

where (Moran 1948)

$$\bar{r}_S = \frac{6}{(n+1)\,\pi}\left[\sin^{-1}(\rho_{XY}) \;+\; (n-2)\,\sin^{-1}(\rho_{XY}/2)\right] \tag{7.41}$$

and $VAR[r_S]$ is approximated by David and Mallows' (1961) formula, given in the technical issues (Eq. 7.65). On the sample level, plug in r_{XY} for ρ_{XY}. A classical CI for z_S follows from using $E[z_S]$, se_{z_S} and a percentage point of the normal distribution; and a classical CI for r_S follows from re-transforming the bounds of $\mathrm{CI}_{z_S,1-2\alpha}$. Note that the binormal assumption is not strong (Fieller et al. 1957) because the ranks, $R(i)$ and $S(i)$, are robust and apply to a wider class of bivariate distributions of $X(i)$ and $Y(i)$.

7.2.2 Classical confidence intervals, persistent processes

A realistic CI for r_S should take persistence into account. The construction algorithm for a bivariate AR(1) process on an unevenly spaced time grid (Algorithm 7.4) is analogous to the case of Pearson's r_{XY}. It employs the effective data size. Also $\mathrm{CI}_{r_S,1-2\alpha}$ is approximate.

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Spearman's r_S (Eq. 7.37)	
Step 3	Fisher's z-transformation	$z_S = \tanh^{-1}(r_S)$
Step 4	Perform Steps 4–8 of the construction algorithm of classical normal CI for r_{XY} (Algorithm 7.1)	$\widehat{\tau}'_X, \widehat{\tau}'_Y, \widehat{\bar{a}}'_X, \widehat{\bar{a}}'_Y, n'_\rho$
Step 5	Pearson's r_{XY} (Eq. 7.5)	
Step 6	Plug in r_{XY} for ρ_{XY} and n'_ρ for n in Eq. (7.41)	$\bar{r}_S$
Step 7	Plug in r_{XY} for ρ_{XY} and n'_ρ for n in Eq. (7.65)	$VAR[r_S]$
Step 8	Plug in $\bar{r}_S$ and $VAR[r_S]$ in Eq. (7.40)	$E[z_S]$
Step 9	Approximate, classical normal CI for r_S, obtained from re-transforming z_S	$\mathrm{CI}_{r_S,1-2\alpha} = \left[\tanh\left[E[z_S] + z(\alpha)\cdot 1.03\left(n'_\rho - 3\right)^{-1/2}\right];\ \tanh\left[E[z_S] - z(\alpha)\cdot 1.03\left(n'_\rho - 3\right)^{-1/2}\right]\right]$

Algorithm 7.4. Construction of classical confidence intervals for Spearman's rank correlation coefficient, bivariate AR(1) models. Steps 3 and 9: z_S is Fisher's transformed rank correlation, $z(\alpha)$ is the percentage point of the normal distribution.

7.2.3 Bootstrap confidence intervals

The motivation for considering bootstrap CIs for r_S is similar to the case of r_{XY}: possibly more complex persistence forms than AR(1) and distributional shapes than bivariate Gaussian. The robustness of ranking methods with respect to violations of the distributional assumption (Kendall and Gibbons 1990; Gibbons and Chakraborti 2003) may weaken the motivation. On the other hand, the inaccuracy of the formula for the expectation of z_S (Eq. 7.40) strengthens it. The two bootstrap algorithms, pairwise-MBB and pairwise-ARB, are for Spearman's r_S completely analogous to the case of Pearson's r_{XY}.

7.2.3.1 Pairwise-moving block bootstrap

The pairwise-MBB algorithm for constructing a bootstrap $\mathrm{CI}_{r_S,1-2\alpha}$ is displayed in shortened form (Algorithm 7.5). Also for Spearman's rank correlation we apply Fisher's z-transformation of r_S to bring the distribution of the replications (z_S^*) in shape closer to a normal distribution. Also the two block length selectors (Eqs. 7.30 and 7.32) are overtaken.

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Spearman's r_S (Eq. 7.37)	
Step 3	Fisher's z-transformation	$z_S = \tanh^{-1}(r_S)$
Step 4	Estimated, bias-corrected	
	persistence times,	$\widehat{\tau}'_X, \widehat{\tau}'_Y$
	block length selection	l
Step 5	Resample, pairwise-MBB with l	$\left\{x^{*b}(i), y^{*b}(i)\right\}_{i=1}^{n}$ (b, counter)
Step 6	Bootstrap replications	$r_S^{*b}, \quad z_S^{*b} = \tanh^{-1}\left(r_S^{*b}\right)$
Step 7	Go to Step 5 until $b = B$	
	(usually $B = 2000$)	
	replications exist	$\left\{z_S^{*b}\right\}_{b=1}^{B}$
Step 8	Calculate CI for Fisher's z_S	$\mathrm{CI}_{z_S,1-2\alpha} = \left[z_{S,l}; z_{S,u}\right]$
Step 9	Re-transformation	$\mathrm{CI}_{r_S,1-2\alpha} = \left[\tanh(z_{S,l}); \tanh(z_{S,u})\right]$

Algorithm 7.5. Construction of bootstrap confidence intervals for Spearman's rank correlation coefficient, pairwise-MBB resampling (cf. Algorithm 7.2).

7.2.3.2 Pairwise-autoregressive bootstrap

The pairwise-ARB algorithm for constructing a bootstrap $\mathrm{CI}_{r_\mathrm{S},1-2\alpha}$ is displayed in shortened form (Algorithm 7.6). Also here we apply Fisher's z-transformation of r_S.

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Spearman's r_S (Eq. 7.37)	
Step 3	Fisher's z-transformation	$z_\mathrm{S} = \tanh^{-1}(r_\mathrm{S})$
Step 4	Estimated, bias-corrected	
	persistence times,	$\widehat{\tau}'_X, \widehat{\tau}'_Y$
	climate equation	
	residuals,	$\{q_X(i)\}_{i=1}^{n}, \{q_Y(i)\}_{i=1}^{n}$
	abbreviations,	$\widehat{a}'_X(i), \widehat{a}'_Y(i)$
	white-noise residuals,	$\epsilon_X(i), \epsilon_Y(i)$
	centering	$\tilde{\epsilon}_X(i), \tilde{\epsilon}_Y(i)$
Step 5	Resample, pairwise-ARB	$\left\{x^{*b}(i), y^{*b}(i)\right\}_{i=1}^{n}$
	(cf. Algorithm 7.3:	
	Steps 14–22 therein); b, counter	
Step 6	Bootstrap replications	$r_\mathrm{S}^{*b}, \quad z_\mathrm{S}^{*b} = \tanh^{-1}\left(r_\mathrm{S}^{*b}\right)$
Step 7	Go to Step 5 until $b = B$	
	(usually $B = 2000$)	
	replications exist	$\left\{z_\mathrm{S}^{*b}\right\}_{b=1}^{B}$
Step 8	Calculate CI for Fisher's z_S	$\mathrm{CI}_{z_\mathrm{S},1-2\alpha} = \left[z_{\mathrm{S,l}}; z_{\mathrm{S,u}}\right]$
Step 9	Re-transformation	$\mathrm{CI}_{r_\mathrm{S},1-2\alpha} = \left[\tanh(z_{\mathrm{S,l}}); \tanh(z_{\mathrm{S,u}})\right]$

Algorithm 7.6. Construction of bootstrap confidence intervals for Spearman's rank correlation coefficient, pairwise-ARB resampling (cf. Algorithm 7.3).

7.3 Monte Carlo experiments

The performance of CIs for r_{XY} and r_S was analysed by means of Monte Carlo simulations. The experiments focused on identifying CI

Table 7.1. Monte Carlo experiment, Spearman's correlation coefficient with Fisher's z-transformation for bivariate lognormal AR(1) processes. $n_{\mathrm{sim}} = 47{,}500$ random samples were generated from the binormal AR(1) process, $\{X(i), Y(i)\}_{i=1}^{n}$, after Eqs. (7.53) and (7.54) with $\tau_X = 1, \tau_Y = 2$ and $\rho_{\mathcal{E}}$ given by Table 7.8. The lognormal shape was generated by taking $\exp[X(i)]$ and $\exp[Y(i)]$. The start was set to $t(1) = 1$; the time spacing, $d(i)$, was drawn from a gamma distribution (Eq. 2.48) with order parameter 16, that means, a distribution with a coefficient of variation equal to $(16)^{-1/2} = 0.25$, and subsequently scaled to $\bar{d} = 1$. Two CI types for ρ_{S} were constructed, classical and bootstrap. The classical CI (Algorithm 7.4) used the effective data size, n'_{ρ}. The bootstrap CI (Algorithm 7.5) used pairwise-MBB resampling, block length selection after Eqs. (7.31) and (7.32), Student's t ($\nu = 2n - 5$) and BCa interval types and $B = 2000$. Confidence level is 95%.

n	$\gamma_{r_{\mathrm{S}}}$[a]						*Nominal*
	True rank correlation, ρ_{S}						
	0.3			0.8			
	CI type			*CI type*			
	Bootstrap		*Classical*	*Bootstrap*		*Classical*	
	Student's t	*BCa*		*Student's t*	*BCa*		
10	0.824	0.671	0.693	0.946	0.842	0.833	0.950
20	0.867	0.800	0.764	0.977	0.900	0.842	0.950
50	0.912	0.893	0.858	0.944	0.929	0.729	0.950
100	0.957	0.922	0.823	0.930	0.920	0.645	0.950
200	0.964	0.928	0.718	0.935	0.930	0.471	0.950
500	0.964	0.927	0.608	0.946	0.945	0.228	0.950
1000	0.943	0.942	0.293	0.943	0.941	0.098	0.950

[a] Standard error of $\gamma_{r_{\mathrm{S}}}$ is nominally 0.001.

types, resampling schemes and correlation measures that perform well, in terms of coverage accuracy, in situations typical for climate time series, namely in the presence of

- non-Gaussian distributional shapes

and

- nonzero, possibly different persistence times of processes $X(i)$ and $Y(i)$.

The two major findings are the following.

Table 7.2. Monte Carlo experiment, Spearman's correlation coefficient with Fisher's z-transformation for bivariate lognormal AR(1) processes: influence of block length selection. The number of Monte Carlo simulations, the properties of $\{T(i), X(i), Y(i)\}_{i=1}^{n}$ and the construction of bootstrap CIs are identical to those in the first experiment (Table 7.1), with the exception that here block length is selected after Eq. (7.30) instead of Eqs. (7.31) and (7.32).

n	$\gamma_{r_S}^{a}$				*Nominal*
	True rank correlation, ρ_S				
	0.3		0.8		
	Bootstrap CI type		*Bootstrap CI type*		
	Student's t	*BCa*	*Student's t*	*BCa*	
10	0.599	0.380	0.756	0.559	0.950
20	0.723	0.592	0.837	0.712	0.950
50	0.894	0.849	0.929	0.904	0.950
100	0.954	0.915	0.924	0.909	0.950
200	0.964	0.927	0.934	0.928	0.950
500	0.963	0.926	0.947	0.946	0.950
1000	0.940	0.939	0.943	0.942	0.950

[a] Standard error of γ_{r_S} is nominally 0.001.

1. Spearman's rank correlation coefficient performed clearly better than Pearson's correlation coefficient. The latter's use is advisable only in situations where the Gaussian assumption is likely to be fulfilled or where computing power allows to calibrate the CIs.

2. Classical CIs failed completely in the presence of non-Gaussian distributions.

Spearman's r_S in combination with pairwise-MBB resampling produced acceptable coverage accuracies (deviations from the nominal level of less than, say, five percentage points) for the bivariate lognormal AR(1) process with unequal, nonzero persistence times (Table 7.1, p. 303). It seems that data size requirements for achieving such coverages are slightly less demanding for Student's t CIs ($n \gtrapprox 100$) than for BCa CIs ($n \gtrapprox 200$).

The choice of the block length selector had minor influence. This is seen by comparing results from using the selector after Eqs. (7.31) and (7.32) in Table 7.1, with those from using Eq. (7.30) in Table 7.2.

Table 7.3. Monte Carlo experiment, Spearman's correlation coefficient without Fisher's z-transformation for bivariate lognormal AR(1) processes. The number of Monte Carlo simulations and the properties of $\{T(i), X(i), Y(i)\}_{i=1}^{n}$ are identical to those in the first experiment (Table 7.1). Also the construction of bootstrap CIs used pairwise-MBB resampling, block length selection after Eqs. (7.31) and (7.32), Student's t and BCa interval types, $B = 2000$ and a confidence level of 95%. The difference is that here no Fisher's z-transformation and no re-transformation (Algorithm 7.5, Steps 3 and 9, respectively) are performed and the bootstrap replications consist not of z_S^{*b}, but of r_S^{*b}.

n	$\gamma_{r_S}^{a}$				*Nominal*
	True rank correlation, ρ_S				
	0.3		0.8		
	Bootstrap CI type		*Bootstrap CI type*		
	Student's t	*BCa*	*Student's t*	*BCa*	
10	0.831	0.821	0.905	0.840	0.950
20	0.877	0.891	0.938	0.900	0.950
50	0.893	0.904	0.942	0.929	0.950
100	0.916	0.921	0.935	0.919	0.950
200	0.926	0.928	0.940	0.930	0.950
500	0.926	0.927	0.942	0.945	0.950
1000	0.941	0.942	0.946	0.941	0.950

[a] Standard error of γ_{r_S} is nominally 0.001.

Fisher's z-transformation of r_S had been advocated by Fieller et al. (1957). However, it seems not to have a major effect on bootstrap CI coverage accuracies. This is seen by comparing Tables 7.1 and 7.3. An early simulation study (Kraemer 1974) with data sizes $n = 10$ and 20 already concluded that using the z-transformation in combination with the normal approximation of the distribution of z_S is less accurate than using instead Student's t distribution.

Pearson's r_{XY} did not produce acceptable coverage accuracies for bivariate lognormal processes—neither with uncalibrated classical CIs nor with uncalibrated CIs from pairwise-MBB resampling (Table 7.4). The reason is likely that the distribution of r_{XY} is, despite the z-transformation, skewed itself when the distributions of the input data, $X(i)$ and $Y(i)$, are skewed. The second experiment with r_{XY}, in which $\{X(i), Y(i)\}$ were binormally distributed (Table 7.5), exhibited reason-

Table 7.4. Monte Carlo experiment, Pearson's correlation coefficient with Fisher's z-transformation for bivariate lognormal AR(1) processes. The number of Monte Carlo simulations and the properties of $\{T(i), X(i), Y(i)\}_{i=1}^{n}$ are identical to those in the first experiment (Table 7.1), with the exception that $\rho_{\mathcal{E}}$ is here given via Eq. (7.24). The construction of CIs followed Algorithms 7.1 and 7.2.

n	$\gamma_{r_{XY}}$ [a]						Nominal
	True correlation, ρ_{XY}						
	0.3			0.8			
	CI type			*CI type*			
	Bootstrap		*Classical*	*Bootstrap*		*Classical*	
	Student's t	*BCa*		*Student's t*	*BCa*		
10	0.820	0.701	0.748	0.864	0.778	0.875	0.950
20	0.876	0.808	0.805	0.904	0.859	0.807	0.950
50	0.932	0.875	0.848	0.898	0.864	0.737	0.950
100	0.939	0.866	0.836	0.895	0.856	0.684	0.950
200	0.941	0.879	0.781	0.897	0.853	0.633	0.950
500	0.907	0.876	0.767	0.899	0.846	0.554	0.950
1000	0.911	0.885	0.730	0.913	0.866	0.551	0.950

[a] Standard error of $\gamma_{r_{XY}}$ is nominally 0.001.

ably good coverage accuracies for $n \gtrapprox 100$. In the second experiment, bootstrap Student's t CIs for r_{XY} performed slightly better than bootstrap BCa (too low coverage) or classical CIs (too high coverage).

The dependence of the coverage results on the true values, ρ_{XY} or ρ_{S}, which were prescribed as 0.3 and 0.8, seems weak.

The better performance of r_{S} in comparison with r_{XY} is rooted in its robustness, which in turn stems from the use of the ranks of the values instead of the values themselves (Fieller et al. 1957). The Monte Carlo simulations reveal that the robustness influences positively also the property of coverage accuracy.

Pairwise-ARB resampling in combination with Student's t CIs performed less good than pairwise-MBB resampling (results not shown), but the coverage error (some percentage points) may be acceptable in climate sciences. Pairwise-ARB resampling in combination with BCa CIs, on the other hand, produced clearly too large coverage errors.

Table 7.5. Monte Carlo experiment, Pearson's correlation coefficient with Fisher's z-transformation for binormal AR(1) processes. The number of Monte Carlo simulations, the properties of $\{T(i)\}_{i=1}^{n}$ and the construction of CIs (Algorithm 7.2) are identical to those in the previous r_{XY} experiment (Table 7.4). The process $\{X(i), Y(i)\}_{i=1}^{n}$ is binormal AR(1) after Eqs. (7.53) and (7.54) with $\tau_X = 1, \tau_Y = 2$ and $\rho_{\mathcal{E}}$ given by Eq. (7.57).

n	$\gamma_{r_{XY}}$[a]						*Nominal*
	True correlation, ρ_{XY}						
	0.3			0.8			
	CI type			*CI type*			
	Bootstrap		*Classical*	*Bootstrap*		*Classical*	
	Student's t	*BCa*		*Student's t*	*BCa*		
10	0.701	0.566	0.752	0.840	0.735	0.973	0.950
20	0.721	0.629	0.812	0.904	0.840	0.974	0.950
50	0.890	0.861	0.931	0.886	0.859	0.974	0.950
100	0.934	0.893	0.959	0.910	0.900	0.972	0.950
200	0.961	0.922	0.953	0.930	0.925	0.968	0.950
500	0.949	0.932	0.953	0.935	0.931	0.968	0.950
1000	0.937	0.934	0.953	0.942	0.941	0.968	0.950

[a] Standard error of $\gamma_{r_{XY}}$ is nominally 0.001.

Calibrating the CI increased the coverage performance dramatically (Table 7.6), especially for the problematic case of Pearson's r_{XY} and bivariate lognormal AR(1) processes. Those accurate results demonstrate that CI lengths (Table 7.7) of correlation estimates are rather large if

- sample sizes are small,
- persistence exists and
- true correlation coefficients are small in size.

Table 7.6. Monte Carlo experiment, Pearson's and Spearman's correlation coefficients with Fisher's z-transformation for bivariate lognormal AR(1) processes: calibrated CI coverage performance. The number of Monte Carlo simulations and the properties of $\{T(i), X(i), Y(i)\}_{i=1}^{n}$ are identical to those in the first experiment (Table 7.1), with $\rho_{\mathcal{E}}$ given by Eq. (7.24) and Table 7.8, respectively. Calibrated Student's t CIs were constructed after Eq. (3.47) using two loops of pairwise-MBB resampling with block length selection after Eqs. (7.31) and (7.32). The first loop (bootstrap of samples) used $B = 2000$ resamplings, the second loop (bootstrap of resamples) used 1000 resamplings. In the second loop, the block length was not re-estimated but overtaken from the first loop. The spacing of the λ values for the calibration (Eq. 3.45) is 0.001.

n	$\gamma_{r_{XY}}^{\mathrm{a}}$		$\gamma_{r_{\mathrm{S}}}^{\mathrm{a}}$		*Nominal*
	True correlation, ρ_{XY}		*True rank correlation*, ρ_{S}		
	0.3	0.8	0.3	0.8	
10	0.917	0.836	0.912	0.959	0.950
20	0.959	0.937	0.960	0.939	0.950
50	0.964	0.947	0.944	0.929	0.950
100	0.969	0.947	0.951	0.945	0.950
200	0.966	0.946	0.970	0.948	0.950

[a] Standard errors of $\gamma_{r_{XY}}$ and $\gamma_{r_{\mathrm{S}}}$ are nominally 0.001.

Table 7.7. Monte Carlo experiment, Pearson's and Spearman's correlation coefficients with Fisher's z-transformation for bivariate lognormal AR(1) processes: average calibrated CI length. The number of Monte Carlo simulations is $n_{\mathrm{sim}} = 47{,}500$. See Table 7.6 for further details.

n	$\langle\ CI_{r_{XY},95\%}\ \mathit{length}\ \rangle^{\mathrm{a}}$		$\langle\ CI_{r_{\mathrm{S}},95\%}\ \mathit{length}\ \rangle^{\mathrm{a}}$	
	True correlation, ρ_{XY}		*True rank correlation*, ρ_{S}	
	0.3	0.8	0.3	0.8
10	1.691	1.028	1.743	1.744
20	1.713	0.506	1.789	0.834
50	1.593	0.291	1.709	0.263
100	1.203	0.229	1.367	0.167
200	0.929	0.184	0.890	0.135

[a] Average value over n_{sim} simulations.

7.4 Example: Elbe runoff variations

Dresden is a station on the river Elbe, from which long measured runoff records are available (Fig. 6.3). We study how strongly the random component in Dresden runoff variations correlate with variations in records from other stations on the river, namely Děčín (70 km upstream) and Barby (240 km downstream). The raw data are shown in Fig. 7.1.

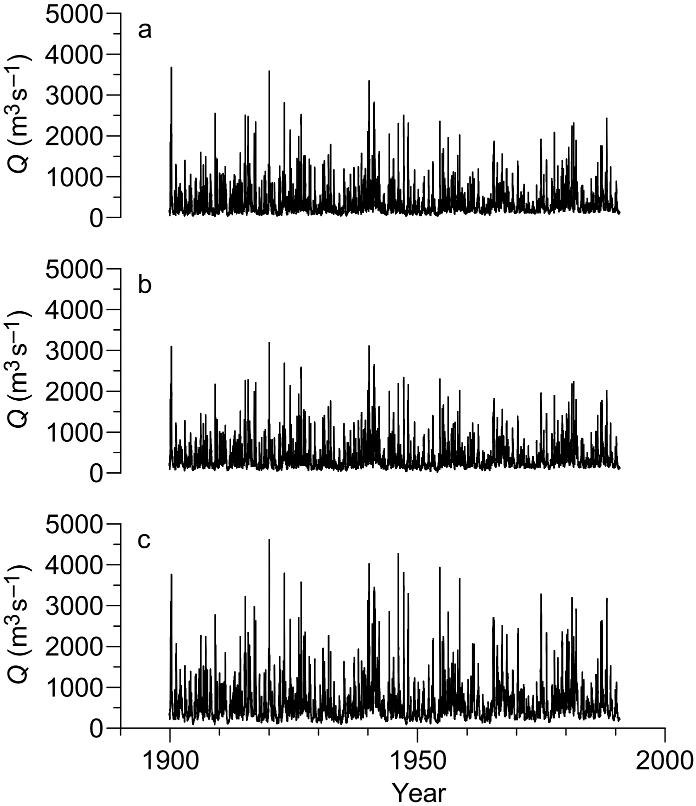


Figure 7.1. Elbe runoff 1899–1990, time series. The raw data are from stations (**a**) Děčín (Czech Republic), (**b**) Dresden and (**c**) Barby (both Germany); they cover the interval from 1 November 1899 to 31 October 1990 in daily resolution, without missing values. (The Czech name of the river is Labe. Data from Global Runoff Data Centre, Koblenz, Germany.)

To extract the random component, we remove the annual cycle from the raw time series. To correct for effects of nonzero travel times of the water (Engel et al. 2002), we bin the daily series into monthly mean records. The resulting data size (n = 1092) is large enough to let us expect a high accuracy of the CIs for the correlation measures.

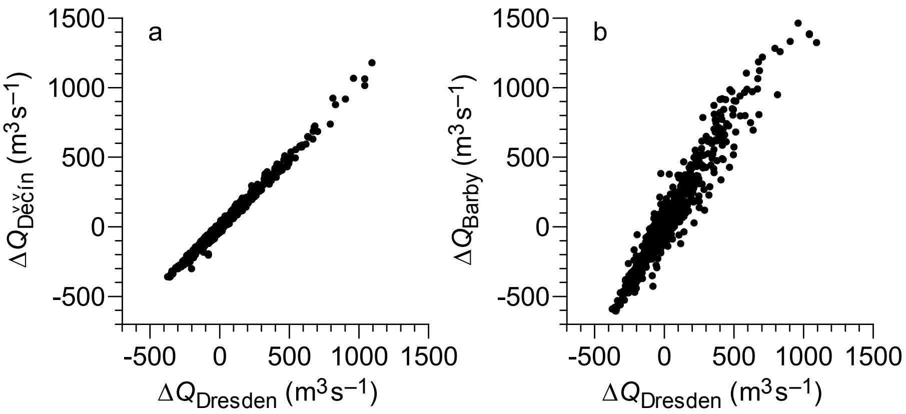


Figure 7.2. Elbe runoff 1899–1990, correlations. **a** Děčín versus Dresden; **b** Barby versus Dresden. Prior to correlation estimation, (1) the annual cycles were removed from the raw data (Fig. 7.1) by subtracting the day-wise long-term averages and (2) monthly resolved records (denoted as ΔQ) constructed by binning. Each record has a sample size of $n = 1092$. (See text for r_{XY} and r_S values.)

The resulting correlation values with 95% CI are as follows. Děčín versus Dresden (Fig. 7.2a), $r_{XY} = 0.995$ [0.993; 0.997] and $r_S = 0.991$ [0.986; 0.995]; Barby versus Dresden (Fig. 7.2b), $r_{XY} = 0.964$ [0.954; 0.972] and $r_S = 0.955$ [0.942; 0.965]. These are calibrated Student's t CIs, obtained using Fisher's transformation, pairwise-MBB resampling (first loop, $B = 2000$ resamplings; second loop, 1000 resamplings) and a λ-spacing of 0.001. The selected block lengths (first loop, overtaken for second loop) after Eqs. (7.31) and (7.32) are (Fig. 7.2a) $l = 13$ and (Fig. 7.2b) $l = 14$.

The significantly higher correlation values of Děčín–Dresden compared with Barby–Dresden bivariate runoff variations can in terms of hydrology be interpreted to reflect the growing catchment area of the river Elbe. While the increase between Děčín and Dresden is moderate (from 51,104 to 53,096 km^2), it is larger when going further downstream, to Barby (94,060 km^2). That means, between Dresden and Barby clearly more "random innovations" in the form of confluencing tributaries "disturb" the water flow than between Děčín and Dresden. This examples demonstrates also the superiority of CIs over mere correlation testing.

7.5 Unequal timescales

Consider the situation that the set of time points for the X values is not identical to the set of time points for the Y values. This is ubiquitous in paleoclimatology, where we study the relation between variations of one variable, measured on one dated climate archive, and a second variable, from a second archive that is independent of the first archive. The challenge imposed by these unequal timescales roots in the fact that Pearson's or Spearman's recipes for estimating the degree of the relation between the fluctuations of both variables cannot be readily applied.

The method of adaption to the case of unequal timescales that is conventionally used in climatology, is to interpolate both time series to a common time grid and then apply the usual estimation procedure of r_{XY} or r_{S}. (We denote these interpolation correlation estimators as $\grave{r}_{XY}$ and $\grave{r}_{\mathrm{S}}$.) One danger with that method is that the freedom of how to select the time grid translates into an arbitrariness regarding the size of the interpolated sample and, in turn, a risk of a biased error determination. The other danger stems from the serial dependences caused by the interpolations, which have to be taken correctly into account.

The adaption method developed in statistical science (Hocking and Smith 1968) focuses on *missing* observations, which is a special case of unequal timescales. There, a number of common time points exist, which allows inference of the covariance because information on the "mixing" exists. The other points, for which data exist from either X or Y, but not from both, are used for inference of the means and standard deviations. This research, summarized by Kotz et al. (2000: pp. 298–305 therein), is, however, of limited relevance for climatological applications.

1. If the unequal timescales do not at all have common time points, which may occur with paleoclimate samples, the correlation estimation is prohibited because no "mixing information" can be used.

2. Instationarities of the first or second moment may bias the estimation. In particular, heteroscedasticity may lead to underestimated standard deviations and, hence, to absolute correlation values greater than unity (Kotz et al. 2000).

3. The assumptions made in the statistical literature, namely multivariate normal distributional shape and serial independence, are typically violated in climatological applications. The properties of the suggested estimators seem not to be known for such a more realistic setting.

This book suggests therefore a seemingly novel estimation approach for climate data samples, which is denoted as binned correlation. It rests on the concept that the nonzero persistence times (Chapter 2), seemingly a genuine property of climate time series, allow to recover the "mixing information" even when the two timescales differ. The condition is that the time spacing is not much larger than the persistence times. Then enough common data points fall within a time bin and knowledge can be acquired on the covariance. We give a second estimation procedure, denoted as synchrony correlation, based on selecting only those X–Y pairs that consist of values close to each other in time.

7.5.1 Binned correlation

To consider the binned correlation (Fig. 7.3), let the two time series have sample size n_X and n_Y, and let the data be given on the process level as

$$\{T_X(i), X(i)\}_{i=1}^{n_X} \text{ and } \{T_Y(j), Y(j)\}_{j=1}^{n_Y}, \tag{7.42}$$

respectively. Let the persistence time be denoted as τ_X and τ_Y, respectively. The first step is to divide the time interval from the leftmost of the time points (denoted as $\bar{T}_{\mathrm{min}}$) to the rightmost of the time points (denoted as $\bar{T}_{\mathrm{max}}$) into bins of a constant length, $\bar{\tau}$. Three selection rules for $\bar{\tau}$ are given in the following paragraph. If $(\bar{T}_{\mathrm{max}} - \bar{T}_{\mathrm{min}})/\bar{\tau}$ has a remainder, let the rightmost bin be smaller than $\bar{\tau}$. The second step is to evaluate whether a time bin contains both more than zero X points and more than zero Y points. For example, in Fig. 7.3 the first bin, $[\bar{T}_{\mathrm{min}}; \bar{T}_{\mathrm{min}} + \bar{\tau}]$, contains three X points and one Y point. If the evaluation is positive, then form the average (denoted as $\bar{X}(k)$) of the X points within the bin and the average (denoted as $\bar{Y}(k)$) of the Y points within the bin, and assign as time value (denoted as $\bar{T}(k)$) the central time for the bin. In Fig. 7.3, $\bar{T}(1) = \bar{T}_{\mathrm{min}} + (1/2) \cdot \bar{\tau}$ and $\bar{T}(2) = \bar{T}_{\mathrm{min}} + (5/2) \cdot \bar{\tau}$ (there are no Y points contained in the second time bin). The resulting time series is

$$\left\{\bar{T}(k), \bar{X}(k), \bar{Y}(k)\right\}_{k=1}^{\bar{n}}. \tag{7.43}$$

This binned sample has size $\bar{n}$. The binned Pearson's correlation coefficient, denoted as $\bar{r}_{XY}$, is calculated as r_{XY} on the binned time series (Eq. 7.43).

The bin width $\bar{\tau}$ is selected such that it is permissible to compare X and Y values within the same bin. This means that the selection takes into account the persistence times of both processes. Simple rules are

$$\bar{\tau} = \tau_X + \tau_Y \tag{7.44}$$

and

$$\bar{\tau} = \max(\tau_X, \tau_Y). \tag{7.45}$$

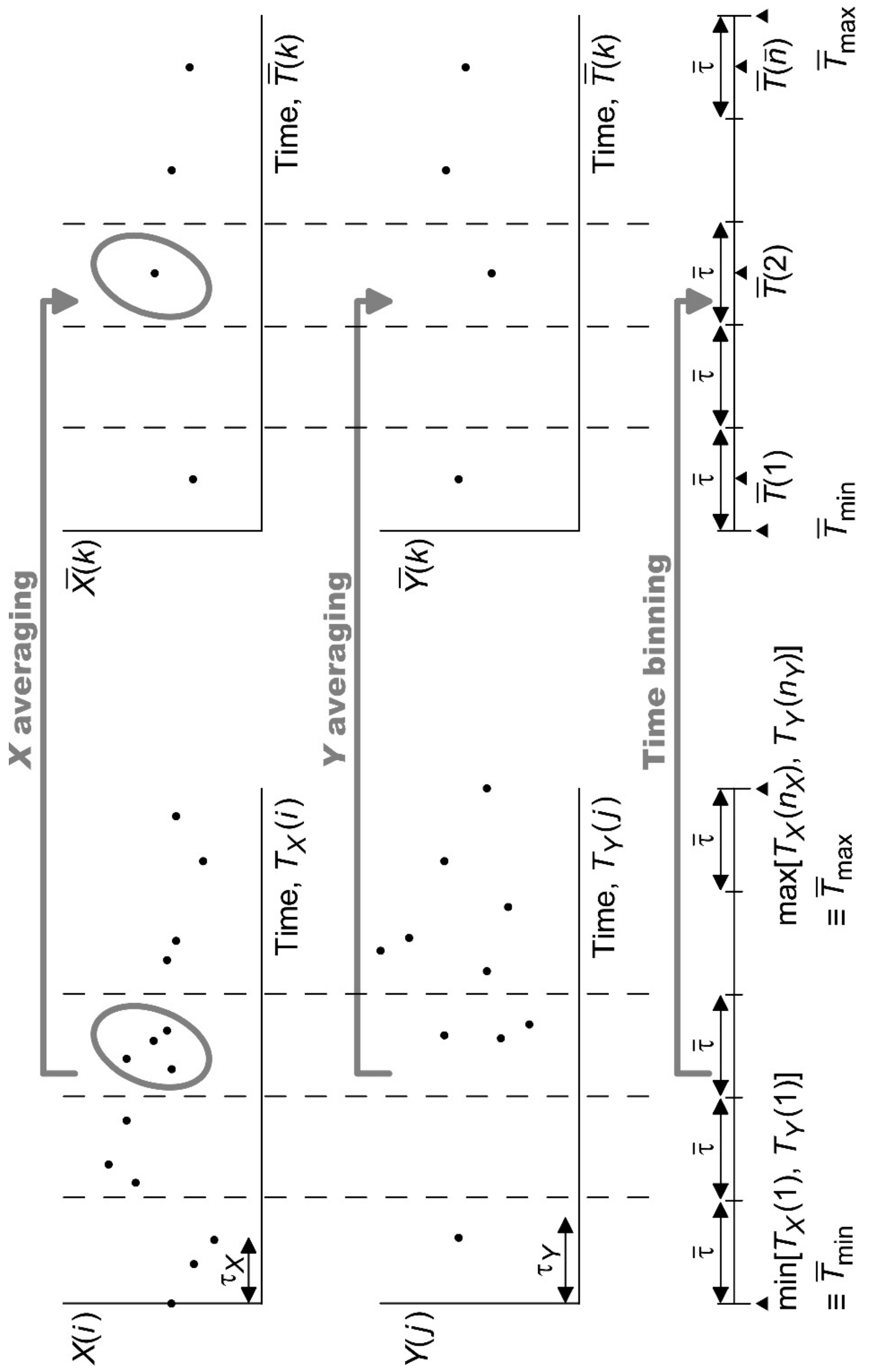


Figure 7.3. Binning for correlation estimation in the presence of unequal timescales.

Monte Carlo experiments, similar to those presented in Section 7.5.3, show the superiority (in terms of $\text{RMSE}_{\widehat{\rho}_{XY}}$) of a third rule, based on the average spacings,

$$\begin{aligned}\bar{d}_X &= \left[T_X(n_X) - T_X(1)\right]/(n_X - 1)\ ,\\ \bar{d}_Y &= \left[T_Y(n_Y) - T_Y(1)\right]/(n_Y - 1)\ ,\\ \bar{d}_{XY} &= \left[\bar{T}_{\max} - \bar{T}_{\min}\right]/(n_X + n_Y - 1)\ ,\end{aligned} \tag{7.46}$$

the bias-corrected equivalent autocorrelation coefficients,

$$\begin{aligned}\widehat{\bar{a}}'_X &= \exp\left(-\bar{d}_X \,/\widehat{\tau}'_X\right),\\ \widehat{\bar{a}}'_Y &= \exp\left(-\bar{d}_Y \,/\widehat{\tau}'_Y\right),\\ \widehat{\bar{a}}'_{XY} &= \left(\widehat{\bar{a}}'_X \cdot \widehat{\bar{a}}'_Y\right)^{1/2},\end{aligned} \tag{7.47}$$

and

$$\bar{\tau} = -\bar{d}_{XY} \Big/ \ln\left(\widehat{\bar{a}}'_{XY}\right). \tag{7.48}$$

Selection of $\bar{\tau}$ determines the sample size, $\bar{n}$, of the binned series and the statistical properties of $\bar{r}_{XY}$. In the case of unequal timescales, the existence of (climate-induced) persistence may have a beneficial effect. If no persistence would be in the (climate) system, and no common time points of X and Y exist, then $\bar{n}$ would be equal to zero and no information on the correlation between X and Y could be recovered.

7.5.2 Synchrony correlation

The synchrony correlation estimation (Algorithm 7.7) starts with selecting the pair $(X(i_{\min}), Y(j_{\min}))$, for which the absolute time difference, $|T_X(i_{\min}) - T_Y(j_{\min})|$, is minimal. The algorithm takes only a percentage, $\beta \cdot 100\%$, of the maximum possible number of X–Y pairs; this maximum number equals $\min(n_X, n_Y)$, and the pairs have the smallest absolute time differences. The correlation estimation is then made by using those n_k "synchrony pairs" and calculating Pearson's or Spearman's correlation coefficient. We denote the synchrony Pearson's correlation coefficient as $\tilde{r}_{XY\beta 100\%}$.

The synchrony correlation estimation avoids making a step away from the original data as interpolation does. The synchrony procedure avoids also the averaging step as the binned correlation procedure (Section 7.5.1) does. The statistical properties of $\tilde{r}_{XY\beta 100\%}$ as an estimator of ρ_{XY} are therefore potentially better (e.g., smaller $\text{RMSE}_{\widehat{\rho}_{XY}}$) than those of $\bar{r}_{XY}$ or $\grave{r}_{XY}$; analogous expectations can be raised for Spearman's correlation coefficient.

Step 1	Processes	$\{T_X(i), X(i)\}_{i=1}^{n_X}$,
		$\{T_Y(j), Y(j)\}_{j=1}^{n_Y}$
Step 2	Initialize counter of "synchrony pairs"	$c_{XY} = 1$
Step 3	Prescribe number of synchrony pairs	$n_k = NINT\left[\beta \cdot \min\left(n_X, n_Y\right)\right]$
Step 4	Absolute time differences	$\Delta T(i,j) = \lvert T_X(i) - T_Y(j)\rvert$
Step 5	Determine combination,	$(i_{\min}, j_{\min})$,
	for which	$\Delta T(i_{\min}, j_{\min})$
	is minimal	
Step 6	Add pair	$(X(i_{\min}), Y(j_{\min}))$
	as number $k = c_{XY}$	
	to the set of synchrony pairs,	
	remove points	$(T_X(i_{\min}), X(i_{\min}))$
	and	$(T_Y(j_{\min}), Y(j_{\min}))$
	from processes, renumber,	
	decrease by one data sizes	n_X, n_Y
Step 7	Increase counter,	$c_{XY} = c_{XY} + 1$
Step 8	If $c_{XY} = n_k$	go to Step 9
	else	go to Step 4
Step 9	Calculate r_{XY} or r_S on	
	the set of synchrony pairs,	$\{X(k), Y(k)\}_{k=1}^{n_k}$

Algorithm 7.7. Synchrony correlation estimation (process level).

The choice of β is crucial because it determines the bias and variance properties of the synchrony correlation estimator. A smaller β means a more restrictive selection of synchrony pairs, leading to a smaller n_k value and, hence, to a larger estimation variance. On the other hand, a smaller β means that the synchrony pairs have smaller absolute time differences and the loss of information, caused by the unequal times (and scaled by τ_X and τ_Y), is smaller; this means further that $\tilde{r}_{XY\beta 100\%}$ has then a smaller estimation bias. The choice of β is a smoothing problem. This book cannot offer a theoretical solution but note that the optimal β value should depend on n_X, n_Y, τ_X, τ_Y and the spacings, $d_X(i) = T_X(i) - T_X(i-1)$ and $d_Y(j) = T_Y(j) - T_Y(j-1)$. For example, larger n_X or n_Y values should allow a more restrictive selection of synchrony pairs, that is, a smaller optimal β. We explore the smoothing problem for a range of dependence factors (n_X, n_Y, τ_X, τ_Y, $d_X(i)$ and $d_Y(j)$) using Monte Carlo simulations.

7.5.3 Monte Carlo experiments

The binned correlation and synchrony correlation are seemingly novel estimation procedures for the situation of unequal timescales. It is therefore appropriate to learn about their basic statistical properties, such as bias or standard error. This is achieved by means of Monte Carlo simulations. In case of the synchrony correlation coefficient, the simulation results help also to assess the influence of the choice of the percentage of "synchrony pairs." Both novel estimators are compared with the conventional interpolation estimator (equidistance, n data pairs).

The simulation experiment studies ρ_{XY} and employs the bivariate Gaussian AR(1) process. The unequal timescales for $X(i)$ and $Y(j)$ are generated (Fig. 7.4) by producing a large number ($10n$) of data pairs on an evenly spaced grid, discarding the majority of points and retaining only small numbers of X and Y points ($n_X = n_Y = n$). The time points for $X(i)$ and $Y(j)$ can be either "well mixed" (Fig. 7.4b) or "wildly mixed" (Fig. 7.4c); the control case of equal time points is included. The results (Figs. 7.5, 7.6 and 7.7) and conclusions are as follows.

1. Because there is memory (persistence) in the climate system, it is in general possible to recover information about the correlation of two processes, $X(T)$ and $Y(T)$, that have been sampled at unequal time points. However, the uncertainties associated with the estimation, in particular, the absolute value of the bias, may be substantially larger than in the case of equal timescales. The bias of the estimated correlation is negative (underestimation) because of the noise introduced by the random innovations between two time points $T_X(i')$ and $T_Y(j')$ ("loss of mixing information").

2. The $\text{RMSE}_{\widehat{\rho}_{XY}}$ descreases for all three estimation procedures (interpolation, binned and synchrony) with the data size (Fig. 7.5). The rate of the decrease is for unequal timescales (Fig. 7.5d–i) similar to the rate of the decrease for equal timescales (Fig. 7.5a–c).

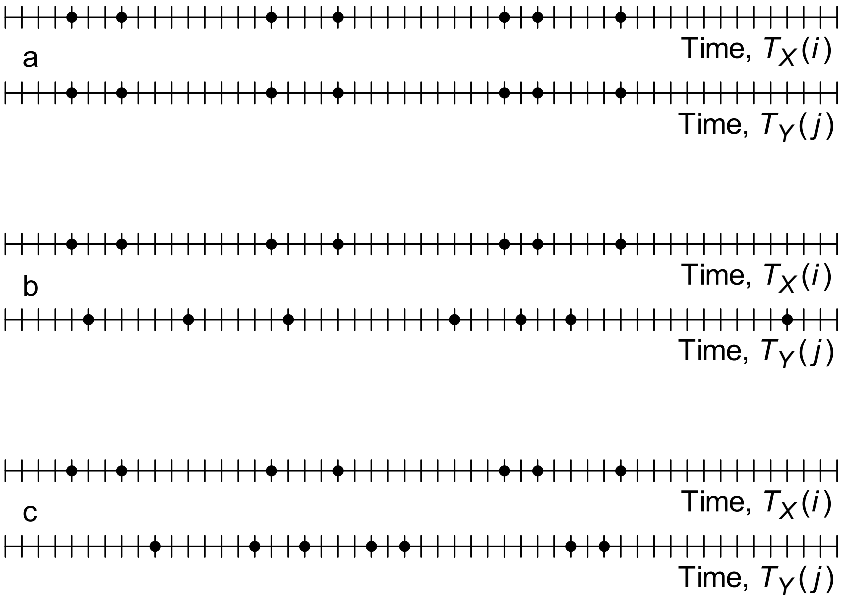


Figure 7.4. Monte Carlo study of correlation estimation, generation of unequal timescales. Time series from the bivariate Gaussian AR(1) process (Eq. 7.25) were generated on a time grid with even spacing (of 1.0), a large data size ($10n$) and prescribed values for $a_X = \exp(-1/\tau_X)$, $a_Y = \exp(-1/\tau_Y)$ and ρ_{XY}. (The evenly spaced grid is displayed as *vertical bars* on the time axes.) First, drawing $n_X = n$ random integers without replacement from the set $\{1, 2, \ldots, 10n\}$ generated the $\{T_X(i)\}_{i=1}^{n_X}$ (process level). The three types of timescales were subsequently generated as follows. **a** Equal timescales (control case) resulted from setting $n_Y = n_X$ and $\{T_Y(j)\}_{j=1}^{n_Y} = \{T_X(i)\}_{i=1}^{n_X}$. **b** Well mixed timescales resulted from drawing $n_Y = n$ random integers without replacement from the set $\{1, 2, \ldots, 10n\}$ and imposing the constraints (1) $T_Y(j) \neq T_X(i)\ \forall\ i, j$ and (2) $T_X(1) < T_Y(1) < T_X(2) < T_Y(2) < T_X(3) < \cdots < T_X(n_X) < T_Y(n_Y)$. **c** Wildly mixed timescales resulted in the same manner as the well mixed timescales but without imposing constraint (2). (The three timescale types are displayed as *filled symbols* on the time axes.) The time series values $\{X(i)\}_{i=1}^{n_X}$ and $\{Y(j)\}_{j=1}^{n_Y}$ (process level) were, finally, taken from the large, evenly spaced bivariate series (size $10n$) according to the random integers from the generated timescales. (For example, if $T_X(1)$ is the seventh time value of the evenly spaced grid (size $10n$), then $X(1)$ is the seventh X value of the evenly spaced series.)

3. Longer persistence times lead in the control case of equal times (Fig. 7.5a–c) to a smaller effective data size, n'_ρ, and a larger standard error. The size of the bias is not strongly influenced (Fig. 7.6a–c). This is described by Eq. (7.11), with n'_ρ plugged in for n.

4. Longer persistence times have in the case of unequal times a twofold effect. First, the standard error is increased (smaller effective data

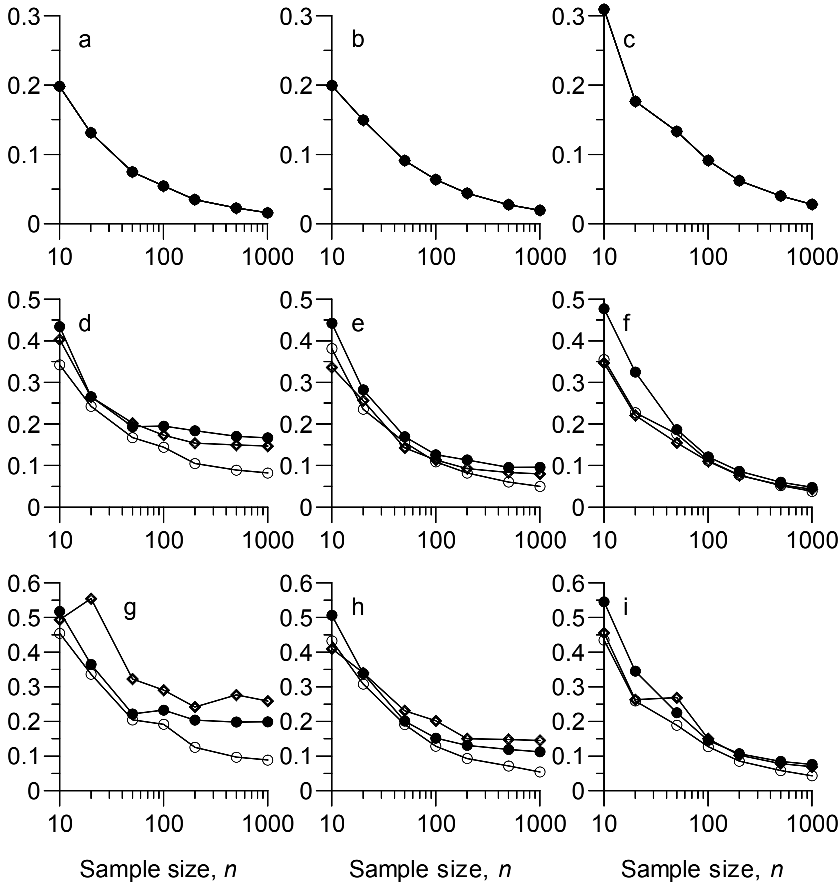


Figure 7.5. Monte Carlo study of correlation estimation in the presence of unequal timescales, dependence on sample size. **a–c** Equal timescales; **d–f** well mixed unequal timescales; **g–i** wildly mixed unequal timescales; **a, d, g** persistence time $\tau_X = \tau_Y = 10$; **b, e, h** $\tau_X = \tau_Y = 20$; **c, f, i** $\tau_X = \tau_Y = 50$; **a–i** $\rho_{XY} = 0.8$; see Fig. 7.4 as regards generation of timescales and of time series after the bivariate Gaussian AR(1) process. Each panel shows the empirical $\mathrm{RMSE}_{\widehat{\rho}_{XY}}$, determined via averaging $(\widehat{\rho}_{XY} - \rho_{XY})^2$ over $n_{\mathrm{sim}} = 10{,}000$ simulations, in dependence on the sample size. The analysed tools are the interpolation correlation estimator $\grave{r}_{XY}$ (shown as *open diamonds*, connected with *lines*), the binned Pearson's correlation coefficient $\bar{r}_{XY}$ (*filled circles*) and the synchrony Pearson's correlation coefficient $\tilde{r}_{XY\beta 100\%}$ (*open circles*) with optimized β (i.e., minimal $\mathrm{RMSE}_{\tilde{r}_{XY\beta 100\%}}$). The optimization of β was done by a brute-force search from the set $\{10\%, 20\%, \ldots, 100\%\}$. Optimal β values for $n = 10, 20, 50, 100, 200, 500, 1000$ were (**d**) $\beta = 1.0, 0.8, 0.4, 0.4, 0.4, 0.4, 0.4$; (**e**) $\beta = 0.9, 0.9, 0.7, 0.6, 0.5, 0.4, 0.4$; (**f**) $\beta = 0.9, 0.8, 0.9, 0.4, 0.7, 0.4, 0.4$; (**g**) $\beta = 0.9, 0.6, 0.4, 0.4, 0.2, 0.2, 0.2$; (**h**) $\beta = 1.0, 0.5, 0.8, 0.3, 0.4, 0.2, 0.2$; and (**i**) $\beta = 0.7, 0.9, 0.8, 0.5, 0.4, 0.2, 0.2$. For the control case of equal timescales (**a–c**), all three estimators (with optimized $\beta = 1.0 \; \forall \; n$) yielded nearly identical results. The relative error is in the order of $n_{\mathrm{sim}}^{-1/2}$.

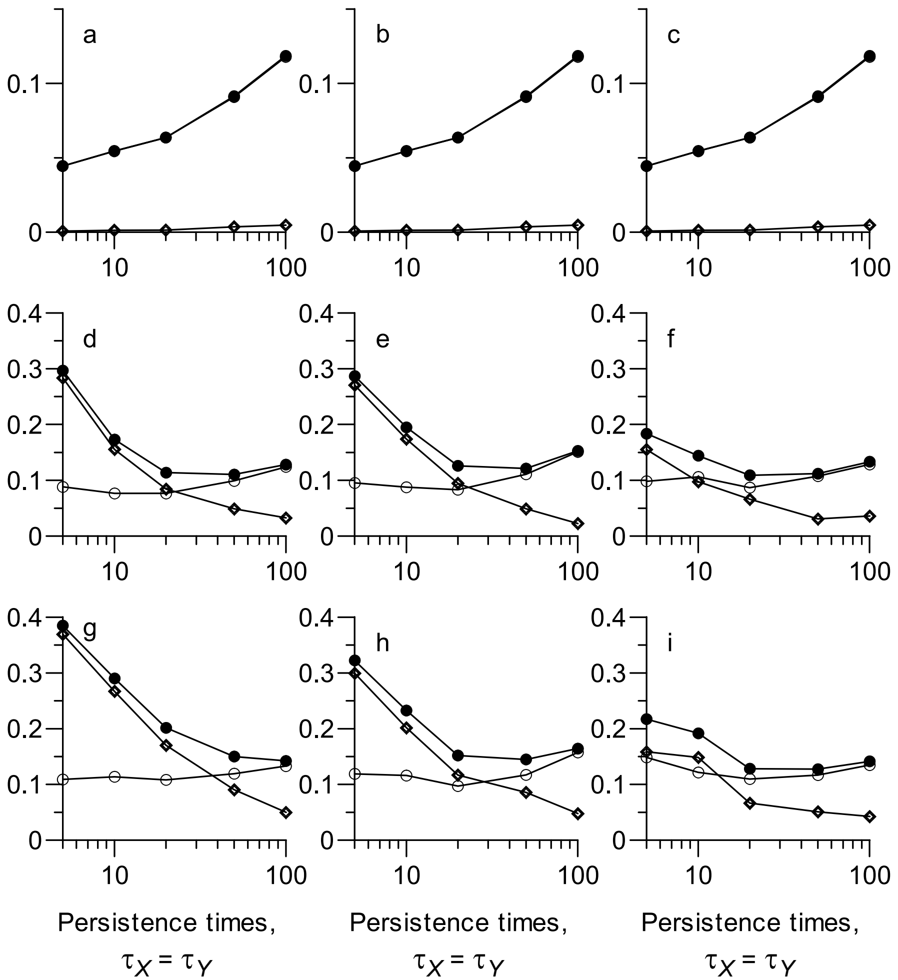


Figure 7.6. Monte Carlo study of correlation estimation in the presence of unequal timescales, dependence on persistence times. **a–c** Equal timescales; **d–f** well mixed unequal timescales; **g–i** wildly mixed unequal timescales; see Fig. 7.4 as regards timescale generation; **a, d, g** interpolation correlation estimator $\grave{r}_{XY}$; **b, e, h** binned Pearson's correlation coefficient $\bar{r}_{XY}$; **c, f, i** optimized synchrony Pearson's correlation coefficient $\tilde{r}_{XY\beta 100\%}$; **a–i** $n = 100$ and $\rho_{XY} = 0.8$; see Fig. 7.4 as regards time series generation after the bivariate Gaussian AR(1) process. Each panel shows the empirical $\mathrm{RMSE}_{\widehat{\rho}_{XY}}$ (as *filled circles*), the negative empirical $\mathrm{bias}_{\widehat{\rho}_{XY}}$ (as *open diamonds*), determined via averaging $(\rho_{XY} - \widehat{\rho}_{XY})$ over $n_{\mathrm{sim}} = 10{,}000$ simulations, and the empirical $\mathrm{se}_{\widehat{\rho}_{XY}} = (\mathrm{RMSE}^2_{\widehat{\rho}_{XY}} - \mathrm{bias}^2_{\widehat{\rho}_{XY}})^{1/2}$ (as *open circles*), in dependence on the persistence times. Optimal β values for $\tau_X = \tau_Y = 5, 10, 20, 50, 100$ were (**f**) $\beta = 0.5, 0.4, 0.6, 0.4, 1.0$; and (**i**) $\beta = 0.2, 0.4, 0.3, 0.5, 0.8$. For the control case of equal timescales (**a–c**), all three estimators (with optimized $\beta = 1.0\ \forall\ \tau_X, \tau_Y$) yielded nearly identical results.

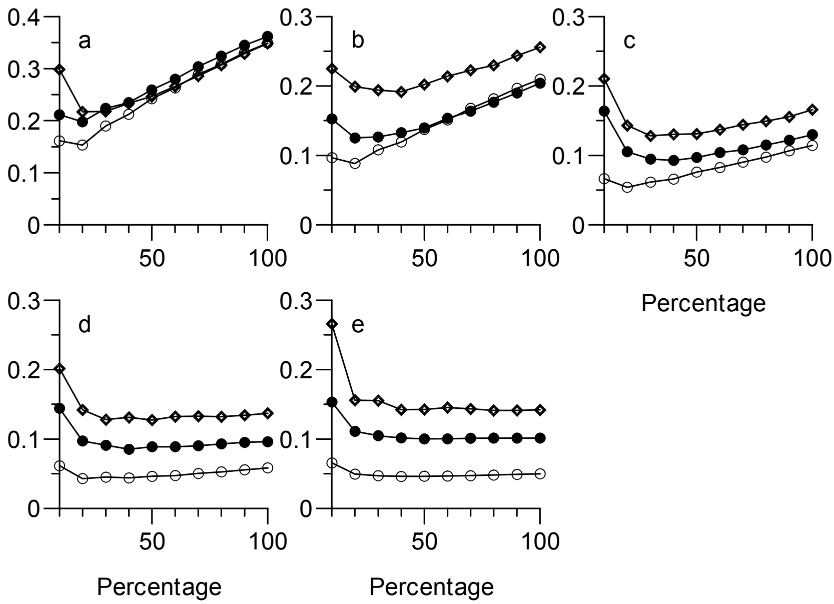


Figure 7.7. Monte Carlo study of synchrony Pearson's correlation coefficient for unequal timescales, dependence on percentage. **a** $\tau_X = \tau_Y = 5$; **b** $\tau_X = \tau_Y = 10$; **c** $\tau_X = \tau_Y = 20$; **d** $\tau_X = \tau_Y = 50$; **e** $\tau_X = \tau_Y = 100$; **a–e** wildly mixed unequal timescales and $\rho_{XY} = 0.8$. See Fig. 7.4 as regards timescale and time series generation. Each panel shows for $n = 100$ (*open diamonds*), 200 (*filled circles*) and 1000 (*open circles*) the empirical $\mathrm{RMSE}_{\tilde{r}_{XY\beta 100\%}}$ in dependence on the percentage (i.e., β).

size). Second, the size of the bias is reduced because a larger amount of the "mixing information" is preserved. The $\mathrm{RMSE}_{\widehat{\rho}_{XY}}$, composed of standard error and bias, has a minimum for intermediate values of τ_X and τ_Y. For the studied Monte Carlo designs (Fig. 7.6d–i) the optimum persistence times are in the range between 20 and 50 time units.

5. A better mixing between $\{T_X(i)\}_{i=1}^{n_X}$ and $\{T_Y(j)\}_{j=1}^{n_Y}$, that is, the well mixed case compared with the wildly mixed case, increases the "mixing information." This leads to a smaller size of the negative bias (Fig. 7.6d–f compared with Fig. 7.6g–i) and to smaller $\mathrm{RMSE}_{\widehat{\rho}_{XY}}$ values (Fig. 7.5d–f compared with Fig. 7.5g–i). For the studied Monte Carlo designs, however, this effect is not very large.

6. The order of the best (in terms of $\mathrm{RMSE}_{\widehat{\rho}_{XY}}$) correlation estimators is: first synchrony method, second binning method and third interpolation method. This is pronounced in the case of wildly mixed

timescales, which is climatologically relevant, and persistence times in the order of 10–20 time units (Fig. 7.5g, h). For larger persistence times, the methods perform more similarly, with a small edge for the synchrony estimator (Fig. 7.5f, i). The synchrony correlation coefficient outperforms the other estimators because it achieves a smaller size of the bias by discarding data pairs with a large loss of mixing information (Fig. 7.6g–i).

7. The choice of the optimal β value for the synchrony method is important. With larger data sizes, the selection of "synchrony pairs" should be more restrictive, as previously expected. The results for the studied Monte Carlo designs (Fig. 7.7) show: while for $n_X = n_Y = 100$ a value of $\beta = 0.5$ to 0.3 or 0.2 is suitable, for $n_X = n_Y = 1000$ one should use a value of closer to 0.2. The curves of $\mathrm{RMSE}_{\tilde{r}_{XY\beta 100\%}}$ versus β are steeper (and hence the importance of the choice of β stronger) for smaller persistence times (Fig. 7.7).

Monte Carlo experiments with $\rho_{XY} = 0.3$ (results not shown) and unequal timescales indicate that a reduced correlation leads to a smaller size of the negative bias and, notably, to an increased standard error.

Regarding bootstrap CI construction for $\bar{r}_{XY}$ or $\tilde{r}_{XY\beta 100\%}$, this may be achievable by adapting the pairwise-MBB (Algorithm 7.2) such that a resampling block need not contain a specified number of points but covers a certain time span. It is important to perform bootstrap resampling on the original data points, $X(i)$ and $Y(j)$, and not on the processed (e.g., binning) points, $X(k)$ and $Y(k)$, to capture fully all aspects of the estimation (e.g., binning). Using the BCa interval type may be advantageous because this CI takes estimation bias into account.

7.5.3.1 Optimal estimation

The example of a suitable selection of a correlation estimator in the case of unequal times (e.g., choice of β) illustrates the concept of optimal estimation (Section 6.2.7). Given two time series from processes $\{T_X(i), X(i)\}_{i=1}^{n_X}$ and $\{T_Y(j), Y(j)\}_{j=1}^{n_Y}$, answering accurately the question about the correlation between both recorded random variables requires first to find out the suitable estimation technique. This can be tackled by analysing the persistence and distributional properties and performing Monte Carlo simulations to explore the hyperspace spanned by the model properties ($\tau_X, \tau_Y, n_X, n_Y, d_X(i), d_Y(j)$, etc.) and, as additional dimensions, the estimation properties of interest (e.g., $\mathrm{RMSE}_{\widehat{\rho}_{XY}}$). It may, especially for more complex estimation problems (e.g., climatology), only then be possible to make a guided inference with a realistic chance of coming close to the truth about the underlying processes.

7.5.4 Example: Vostok ice core records

The Vostok ice core records (Fig. 1.3) of $X(i)$: CO_2 and $Y(j)$: deuterium (proxy for temperature variations) over the past 420 ka display a high correlation (Fig. 7.8). This sheds some light on the coupling of those two variables, which govern (among others) the Pleistocene climate.

The data sizes are $n_X = 283$ and $n_Y = 3311$; the average time spacings $\bar{d}_X = 1.46$ ka and $\bar{d}_Y = 0.128$ ka; and the estimated, bias-corrected persistence times $\widehat{\tau}'_X = 38.1$ ka and $\widehat{\tau}'_Y = 25.6$ ka. For performing the binning procedure after Eq. (7.48), these values lead to $\bar{\tau} = 5.4$ ka. Most, but not all of the time bins contain both CO_2 and deuterium values, and the resulting binned bivariate sample has a size of $\bar{n} = 77$. The correlation coefficient, calculated on the binned sample, is $\bar{r}_{XY} = 0.876$.

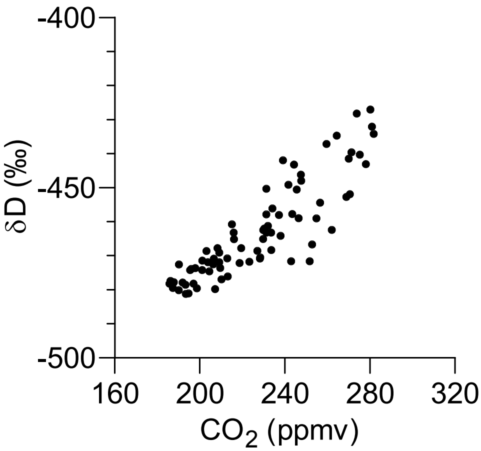


Figure 7.8. Vostok deuterium and CO_2 over the past 420 ka, correlation. Shown is the binned bivariate sample. The binned Pearson's correlation coefficient is $\bar{r}_{XY} = 0.876$.

For assessing the accuracy of the correlation estimation, Monte Carlo simulations were performed such as in Section 7.5.3, but with the design $\rho_{XY} = 0.9$, $\tau_X = \widehat{\tau}'_X$, $\tau_Y = \widehat{\tau}'_Y$ and identical data sizes and timescales ($\{T_X(i)\}_{i=1}^{n_X}$ and $\{T_Y(j)\}_{j=1}^{n_Y}$). This experiment ($n_{\mathrm{sim}} = 10{,}000$) resulted in empirical values of $\mathrm{RMSE}_{\bar{r}_{XY}} = 0.039$, $\mathrm{bias}_{\bar{r}_{XY}} = -0.025$ and $\mathrm{se}_{\bar{r}_{XY}} = 0.030$. We may safely conclude that the true correlation coefficient between temperature and CO_2 variations at Vostok is somewhere between 0.85 and 0.9.

Two aspects raised by the ice core example may be studied further. First, timescale uncertainties should in principle be amenable to analysis by means of parametric modelling (Chapter 4). The modelled times, $\{T_X^*(i)\}_{i=1}^{n_X}$ and $\{T_Y^*(j)\}_{j=1}^{n_Y}$, may find entrance into Monte Carlo simulations for calculating the empirical $\text{RMSE}_{\bar{r}_{XY}}$, and so forth. Regarding the Vostok records, both are from the same ice core and only the uncertainty in the age difference between ice and gas needs to be considered. This uncertainty is, however, clearly smaller (Chapter 8) than the bin width of 5.4 ka. The effect on the accuracy of $\bar{r}_{XY}$ is therefore rather small. Second, for climatological purposes it makes sense to allow for time lags between variations of $X(i)$ and $Y(j)$. This point is pursued in Chapter 8. Those two methodical expansions are neither restricted to the binned coefficient nor to Pearson's version of correlation estimation.

7.6 Background material

The **binormal distribution** has following PDF:

$$\begin{aligned} f(x,y) &= (2\pi S_X S_Y)^{-1}\left(1-\rho_{XY}^2\right)^{-1/2} \\ &\quad \times \exp\left[-\frac{1}{2\left(1-\rho_{XY}^2\right)}\left(\frac{(x-\mu_X)^2}{S_X^2}\right.\right. \\ &\quad \left.\left. -\frac{2\rho_{XY}(x-\mu_X)(y-\mu_Y)}{S_X S_Y}+\frac{(y-\mu_Y)^2}{S_Y^2}\right)\right]. \end{aligned} \tag{7.49}$$

μ_X and μ_Y is the mean, S_X^2 and S_Y^2 the variance of the univariate processes $X(i)$ and $Y(i)$, respectively; $\rho_{XY} = \rho_{\mathcal{E}}$ is the correlation coefficient. The PDF is "slanted" for $\rho_{XY} \neq 0$ (Fig. 7.9). See Priestley (1981: Section 2.12.9 therein), Patel and Read (1996: Chapter 9 therein) and Kotz et al. (2000: Chapter 46 therein) for more details on the binormal distribution.

The **bivariate lognormal distribution** is in the more general case, with shape parameters σ_X and σ_Y and scale parameters b_X and b_Y, given by

$$\begin{aligned} X(i) &= \exp\left[\sigma_X \cdot \mathcal{E}^X_{\mathrm{N}(0,\,1)}(i) + \ln(b_X)\right], \qquad i = 1, \ldots, n, \\ Y(i) &= \exp\left[\sigma_Y \cdot \mathcal{E}^Y_{\mathrm{N}(0,\,1)}(i) + \ln(b_Y)\right], \qquad i = 1, \ldots, n; \end{aligned} \tag{7.50}$$

see also Section 3.9. It has the correlation (Mostafa and Mahmoud 1964)

$$\rho_{XY} = \left[\exp\left(\sigma_X \cdot \sigma_Y \cdot \rho_{\mathcal{E}}\right) - 1\right] \Big/ \sqrt{\left[\exp\left(\sigma_X^2\right) - 1\right]\ \left[\exp\left(\sigma_Y^2\right) - 1\right]} \tag{7.51}$$

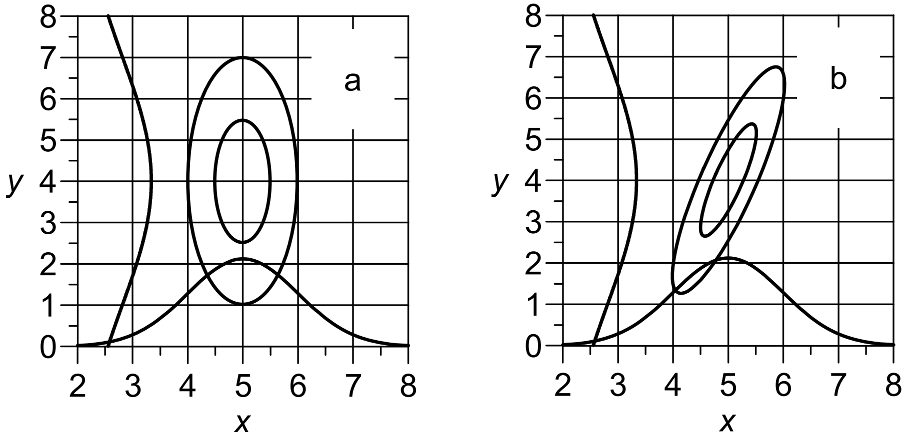


Figure 7.9. Binormal probability density function: contour lines and marginal distributions. Parameters: $\mu_X = 5, S_X = 1, \mu_Y = 4, S_Y = 3$ and (**a**) $\rho_{XY} = 0$ or (**b**) $\rho_{XY} = 0.6$.

and the PDF

$$\begin{aligned} f(x,y) \;\; &= \;\; (2\pi\sigma_X\sigma_Y\, xy)^{-1}\left(1-\rho_{\mathcal{E}}^2\right)^{-1/2} \\ & \quad \times \exp\left[-\frac{1}{2\left(1-\rho_{\mathcal{E}}^2\right)}\left(\frac{[\ln(x/b_X)]^2}{\sigma_X^2}\right.\right. \\ & \quad \left.\left. -\frac{2\rho_{\mathcal{E}}\ln(x/b_X)\ln(y/b_Y)}{\sigma_X\sigma_Y}+\frac{[\ln(y/b_Y)]^2}{\sigma_Y^2}\right)\right]. \end{aligned} \tag{7.52}$$

The **bivariate AR(1) process for uneven time spacing** is given by

$$\begin{aligned} X(1) &= \mathcal{E}^X_{\mathrm{N}(0,\,1)}(1), \\ Y(1) &= \mathcal{E}^Y_{\mathrm{N}(0,\,1)}(1), \\ X(i) &= \exp\left\{-\left[T(i)-T(i-1)\right]/\tau_X\right\}\cdot X(i-1) \\ & \quad + \mathcal{E}^X_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau_X\})}(i), \qquad i = 2,\ldots,n, \\ Y(i) &= \exp\left\{-\left[T(i)-T(i-1)\right]/\tau_Y\right\}\cdot Y(i-1) \\ & \quad + \mathcal{E}^Y_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau_Y\})}(i), \qquad i = 2,\ldots,n, \end{aligned} \tag{7.53}$$

where the white-noise innovation terms are correlated as

$$
\begin{aligned}
& CORR\Big[\mathcal{E}^X_{\mathrm{N}(0,\,1)}(1), \mathcal{E}^Y_{\mathrm{N}(0,\,1)}(1)\Big] = \rho_{\mathcal{E}}, \\
& CORR\Big[\mathcal{E}^X_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau_X\})}(i), \\
& \qquad \mathcal{E}^Y_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau_Y\})}(i)\Big] \\
& \qquad\qquad = \Big(1 - \exp\left\{- \left[T(i) - T(i-1)\right] \cdot (1/\tau_X + 1/\tau_Y)\right\}\Big) \\
& \qquad\qquad\quad \times \Big(1 - \exp\left\{-2\left[T(i) - T(i-1)\right]/\tau_X\right\}\Big)^{-1/2} \\
& \qquad\qquad\quad \times \Big(1 - \exp\left\{-2\left[T(i) - T(i-1)\right]/\tau_Y\right\}\Big)^{-1/2} \rho_{\mathcal{E}}, \\
& \qquad\qquad\quad i = 2, \ldots, n,
\end{aligned}
\tag{7.54}
$$

$$
\begin{aligned}
& CORR\Big[\mathcal{E}^X_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau_X\})}(i), \\
& \qquad \mathcal{E}^Y_{\mathrm{N}(0,\,1-\exp\{-2[T(j)-T(j-1)]/\tau_Y\})}(j)\Big] \\
& \qquad\qquad = 0, \qquad i, j = 2, \ldots, n, \qquad i \neq j,
\end{aligned}
$$

$$
\begin{aligned}
& CORR\Big[\mathcal{E}^X_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau_X\})}(i), \mathcal{E}^Y_{\mathrm{N}(0,\,1)}(1)\Big] \\
& \qquad\qquad = 0, \qquad i = 2, \ldots, n,
\end{aligned}
$$

$$
\begin{aligned}
& CORR\Big[\mathcal{E}^X_{\mathrm{N}(0,\,1)}(1), \mathcal{E}^Y_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau_Y\})}(i),\Big] \\
& \qquad\qquad = 0, \qquad i = 2, \ldots, n.
\end{aligned}
$$

This process is strictly stationary. Its properties are

$$E\left[X(i)\right] = E\left[Y(i)\right] = 0, \tag{7.55}$$

$$VAR\left[X(i)\right] = VAR\left[Y(i)\right] = 1 \tag{7.56}$$

and

$$CORR\left[X(i), Y(i)\right] = \rho_{XY} = \rho_{\mathcal{E}}. \tag{7.57}$$

Bias and standard error of Pearson's correlation coefficient for distributions of $X(i)$ or $Y(i)$ that deviate from the Gaussian shape can theoretically be approximated using the parameters (cumulants of higher order) describing the deviations. Lengthy approximation formulas were given by Gayen (1951) and Nakagawa and Niki (1992). The relevance of the formulas for practical climatological purposes seems to be limited because of the considerable uncertainties in the estimation of those cumulants from data sets limited in size.

An **alternative to Fisher's transformation** is (Hotelling 1953)

$$z_{\mathrm{H}} = z - \frac{3z + r_{XY}}{4\,(n-1)}. \tag{7.58}$$

For small n, Hotelling's z_{H} is in distribution closer to a Gaussian shape than Fisher's z (Rodriguez 1982).

Spearman's rank correlation coefficient is reviewed by Pirie (1988). Fisher's z-transformation and usage of the normal distribution is not the only method for constructing classical, approximate CIs for r_{S}. Kraemer (1974) suggested an alternative transformation and usage of Student's t distribution. Otten (1973) gave for the null case $\rho_{\mathrm{S}} = 0$ the exact PDF of r_{S} for $n = 13$ to 16. Franklin (1988) examined the convergence of the exact null distribution of r_{S} to normality for $n = 9$ to 18. As regards the originator of r_{S}, Pearson (1924: p. 393 therein) thinks that there is "sufficient evidence that Galton dealt with the correlation of ranks before he even reached the correlation of variates, and the claim that it is a contribution of the psychologists [i.e., Spearman] some thirty or forty years later to the conception of correlation does not seem to me valid."

The **grade correlation coefficient** between two continuous variables X and Y is (Gibbons and Chakraborti 2003: Section 11 therein)

$$\begin{aligned} \rho_{\mathrm{S}} &= 12\; E\left[F_X(X)\cdot F_Y(Y)\right] - 3 \\ &= 12 \int F_X(x)\cdot F_Y(y)\; f(x,y)\; dx\, dy - 3. \end{aligned} \tag{7.59}$$

Herein, $F_X(x)$ and $F_Y(y)$ are the (marginal) distribution functions and $f(x,y)$ is the bivariate PDF. The case of a binormal PDF (Eq. 7.49) with correlation coefficient $\rho_{XY} = \rho_{\mathcal{E}}$ can be analytically solved (Pearson 1907):

$$\rho_{\mathrm{S}} = \frac{6}{\pi}\, \sin^{-1}\left(\rho_{\mathcal{E}}/2\right). \tag{7.60}$$

The case of a bivariate lognormal PDF (Eq. 7.52), where ρ_{XY} is related to $\rho_{\mathcal{E}}$ via Eq. (7.51), was solved (Table 7.8) by means of simulations. Consider the normal distribution function (Eq. 3.49), denoted as $F_{\mathrm{N}}(x)$. Consider further the lognormal distribution function, $F_{\mathrm{LN}}(x) = \int_{-\infty}^{x} f_{\mathrm{LN}}(x')dx'$. Herein, $f_{\mathrm{LN}}(x)$ is the lognormal PDF (Eq. 3.61). Then, $F_{\mathrm{LN}}(x) = F_{\mathrm{N}}(\ln(x))$ for $x > 0$.

Table 7.8. Grade correlation coefficient, bivariate lognormal distribution. ρ_{S} was determined from its definition (Eq. 7.59) by drawing random bivariate numbers from the density (Eq. 7.52) and calculating the average and its standard error over n_{sim} = 1,000,000,000 simulations. Lognormal parameters: $b_X = b_Y = 0.0$ and $\sigma_X = \sigma_Y = 1.0$.

ρ_{S}[a]	*Accuracy*[b] *of* ρ_{S}	ρ_{ε}
0.3	$< 10^{-4}$	0.3129
0.8	$< 10^{-4}$	0.8135

[a] Average over n_{sim} simulations.
[b] Standard error over n_{sim} simulations.

The **point biserial correlation coefficient** can be used as an estimator of the degree of the linear relationship between a continuous variable, $X(i)$, and a dichotomous (binary) variable, $Y(i)$. A field for climatological applications is the analysis of outliers or climate extremes (Chapter 6), where, for example, $Y(i) = 0$ means the absence and $Y(i) = 1$ the occurrence of an extreme at time $T(i)$. Let (on the sample level) p denote the proportion of $y(i)$ values equal to 0; $q = 1-p$; $\bar{x}_0$ and $\bar{x}_1$ be the mean $x(i)$ value with $y(i) = 0$ and 1, respectively; and $s_{n,X}$ be the sample analogue of the standard deviation estimator (Eq. 7.8). The point biserial correlation coefficient is then defined (Kraemer 1982) as

$$r_{\mathrm{pb}} = (pq)^{1/2} \left(\bar{x}_1 - \bar{x}_0\right) / s_{n,X} \; . \tag{7.61}$$

It readily follows that $r_{\mathrm{pb}} = r_{XY}$. It may be shown (Tate 1954) that if (1) $\rho_{XY} = 0$ and (2) the standard deviation of $X(i)$ is independent of whether $Y(i)$ equals 1 or 0, the statistic

$$t_{\mathrm{pb}} = (n-2)^{1/2} \, r_{\mathrm{pb}} \left(1 - r_{\mathrm{pb}}^2\right)^{-1/2} \tag{7.62}$$

is distributed as Student's t with $n-2$ degrees of freedom (Section 3.9). This statistic was used by Mudelsee et al. (2004) to study whether a relation exists between atmospheric variables (sea-level pressure, geopotential height) and the occurrence of Elbe floods for the interval from 1658 to 1999. Because of the persistence of the processes that generated the atmospheric time series, Eq. (7.62) was adapted by replacing n with the effective data size (determined as $0.85\,n$ to $0.90\,n$). Other climatological examples of usage of r_{pb} are the following. Ruiz and Vargas (1998) study the relation between an atmospheric variable (vorticity) and the occurrence of large rainfall at South American stations, interval 1983–1987, and Giaiotti and Stel (2001) relate thunderstorm occurrence to geopotential height in northeast Italy, interval 1998–1999. A caveat that applies to the interpretation of the results from both studies is that persistence was ignored in the analyses. Bootstrap CIs for r_{pb} were studied by Sievers (1996), who found good coverage performance of calibrated CIs already for small data sizes ($n = 10$).

Kendall's tau, employed in Section 4.4 (p. 168) for trend testing, can also be used (and this was historically earlier) as a correlation measure. For trend testing, we count the number of interchanges to bring $\{X(i)\}_{i=1}^{n}$ into the same (monotone) order as $\{i\}_{i=1}^{n}$; for correlation estimation, we have to bring $\{X(i)\}_{i=1}^{n}$ into the same order as $\{Y(i)\}_{i=1}^{n}$. Hamed (2009a) presented adaptions of the statistical test of H_0: "zero correlation" to take into account serial dependences (short- and long-term).

The **Monte Carlo performance of bootstrap CIs** for correlation coefficients was studied by the following. Hall et al. (1989) found that one loop of calibration of the percentile CI for r_{XY} brings a dramatic increase in coverage accuracy for bivariate lognormal white noise and small data sizes ($n = 8, 10, 12$). Sievers (1996) confirmed this finding for $n = 19$ and eight types of the distributional shape of the white noise. Above studies used ordinary bootstrap resampling because of the absence of serial dependence. Mudelsee (2003) analysed bootstrap BCa CIs for r_{XY} on bivariate Gaussian and lognormal AR(1) processes with n between 10 and 1000. He used pairwise-MBB resampling and concluded that acceptable levels of coverage accuracy can be achieved but that the serial dependence reduces the effective data size to a considerable degree. Two caveats against this study are that block length selection was done in an ad-hoc manner (Eq. 7.30) and that the studied process was not identical to the strictly stationary model of Eq. (7.53). The observation that nonzero persistence has detrimental effects (larger bias and RMSE) of correlation estimators was also quantified by Park and Lee (2001), who analysed r_{S} on bivariate Gaussian AR(1) processes with $n = 137$. These authors tried several resampling methods in combination with a brute-force block length selection (in terms of RMSE of the standard deviation of r_{S}). One of their conclusions is that pairwise-ARB resampling performed better than a pairwise version of the nonparametric SB resampling (Section 3.8). Papers from the psychology literature report about coverage performances of bootstrap CIs for quantities that are related to r_{XY} and are of relevance to that branch of investigation, namely (1) correlation coefficients that account for range restriction or censoring of one variable (Chan and Chan 2004) and (2) the difference of correlation coefficients in overlapping data sequences (Zou 2007).

The **Monte Carlo performance of bootstrap hypothesis tests** about correlation coefficients was studied by the following. Martin (2007) considered H_0: "$\rho_{XY} = \rho'_{XY}$" with nonzero ρ'_{XY}. This constitutes an important test case, not only for the climate sciences, because it does not consider the "straw man" H_0: "$\rho_{XY} = 0$," but instead a more realistic H_0. Such a test may supply a quality of information similar to that of a CI (Section 3.6), with additional information regarding the test power. To resample under the null of nonzero ρ_{XY}, a "rotated" version of the array of the original bivariate sample, $\{y(i)\}_{i=1}^{n}$ versus $\{x(i)\}_{i=1}^{n}$, is used (Beasley et al. 2007: Fig. 1 therein). The two cited papers study the empirical significances and powers for bivariate white noise with a range of data sizes (from 10 up to 100), various distributional shapes and several ρ_{XY} values $(-0.5, 0, 0.4, 0.8)$. Belaire-Franch and Contreras-Bayarri (2002) performed an analogous experiment of the empirical test significances and powers, employing AR(1) and MA(1) models of serial dependence and using SB resampling. Summarizing the results of the above mentioned Monte Carlo experiments, we conclude that testing realistic null hypotheses about ρ_{XY} can be accurately performed using bootstrap resampling. Regarding the test of H_0: "$\rho_{XY} = 0$," Ebisuzaki (1997) studied the frequency-domain bootstrap (Section 5.3) and the classical approach (via n'_ρ) using even time spacing and bivariate AR(1) and AR(2) models with $n = 8$, 16, 32 and 64. He found the bootstrap variant to produce acceptably small deviations between nominal and empirical rejection rates, not only for the AR(1) but also the AR(2) model. Larger deviations occurred for small n and the AR(2) parameter approaching with $a_2 \rightarrow -1$ the boundary of the stationarity regime (Fig. 2.4). Ebisuzaki (1997) ascribed this deficit to the poor properties of the periodogram as spectrum estimator (Chapter 5). Pyper and Peterman (1998) did not study the bootstrap but rather the classical approach (via n'_ρ) to take into account serial dependence. They explored various persistence models (AR(1), AR(2) and ARIMA), sample sizes (between 15 and 50), autocorrelation estimators and the effects of smoothing prior to the correlation estimation. Prior smoothing constitutes a special case of an alternative correlation measure (Section 7.1.1). One of their conclusions is that prior smoothing may reduce considerably the effective data size and lead to a reduced power of the statistical test, see also the paragraph here on the Sun–climate relationship.

Binned and synchrony correlation coefficients seem to be novel estimation tools applicable to the case of unequal timescales. Davison and Hinkley (1997: Example 3.12 therein) consider an example from a closely related case, where some values (of, say, $Y(i)$) are missing. They consider the imputation of the missing values "to obtain a suitable bivariate $\widehat{F}$ [estimate of the distribution function], next estimate θ [i.e., ρ_{XY}] with the usual sample correlation $t(\widehat{F})$ [i.e., r_{XY}], and then resample appropriately" (Davison and Hinkley 1997: p. 90 therein). An imputation method is to make a regression of $X(i)$ on $Y(i)$ (Chapter 8) using the bivariate subsample without the missing values. The assumption here is that the values are missing at random (Rubin 1976), that means, for example, that no range restriction or censoring occurred. Algorithms for imputing missing data for general estimation purposes were presented by Dempster et al. (1977) and Efron (1994). With regard to the temporal spacing, we have a focus on the general case of irregularity and not on the case of missing observations from an evenly spaced grid. This is why we almost exclusively do not consider imputation. In the bivariate setting, however, imputation may be an interesting estimation alternative. We note that when X and Y have no common time points, imputation is not straightforward to implement, whereas binned and synchrony correlation are so and may (if persistence exists and the time points of X and Y are well "mixed") help to recover information about the underlying correlation.

Timescale uncertainty was also identified by Haam and Huybers (2010) as a problem affecting the estimation of the relationship between two processes. These authors selected the covariance measure, assumed even spacing and allowed only one of the two processes to be influenced by timescale errors. Furthermore, the timescale errors were assumed to take discrete values only. For this simplified setting, they obtained analytical and numerical results on the distribution of the maximum of the covariance. This was in turn used as a measure of the significance of the empirical covariance. Finally, Haam and Huybers (2010) used this test to study the relation between variations of δ^{18}O in a stalagmite (with timescale errors) and atmospheric radiocarbon content during the Holocene. They were unable to reject the null hypothesis of zero covariance. This result should be assessed with some caution because prior to the analysis the series had been interpolated to achieve even spacing.

The **smoothed bootstrap** consists of adding an amount of (normally distributed) noise to resampled values, $x^*(i)$ and $y^*(i)$. The idea (Efron 1982) is to circumvent the discrete distribution of the bootstrap samples, which may lead for quantities such as the sample median to a bad performance (Davison and Hinkley 1997: Section 3.4 therein). A Monte Carlo study (Silverman and Young 1987) of the RMSE of the sample standard deviation of r_{XY} and z demonstrated the superiority of smoothing, especially for z and small data sizes ($n \leq 50$). Young (1988) gave a rule for adjusting the amount of smoothing. It may be that the coverage of bootstrap CIs for correlation estimators could benefit from smoothing. However, more theoretical knowledge on the application of the smoothed bootstrap to time series from serially dependent bivariate processes would be helpful.

Climatological applications of bootstrap CIs for Pearson's r_{XY} include the following. Kumar et al. (1999) used MBB resampling and percentile CIs to study the "weakening relationship between the Indian monsoon and ENSO" during the interval from 1856 to 1997. They took a running window of length 21 years and determined the correlation using the points of all-Indian summer monsoon rainfall (June to September average) and equatorial Pacific sea-surface temperature anomalies (June to August average) within the window. The obtained correlation confidence band is pointwise. Such "running correlations" are often used in explorative climatology, despite the absence of a theoretical framework for nonconstant $CORR\,[X(i), Y(i)]$. Girardin et al. (2006a) used pairwise-MBB resampling and BCa CIs (PearsonT software, Section 7.7) to find a highly significant correlation between (a transformation of) the Pacific sea-surface temperature and west-east atmospheric flow over Canada during the past approximately 150 years. Boessenkool et al. (2007) used the same method to relate the (proxy-derived) speed of the water flow near the ocean bottom at the Iceland-Scotland ridge with the NAO index, interval 1885–2004. They found that a positive index (stronger sea-level pressure gradient) had reduced the water flow; the value of $r_{XY} = -0.42$ with 95% CI $[-0.60; -0.20]$ serves to quantify the amount of covariation. This finding has implications for our knowledge about the meridional overturning of the Atlantic in response to climate change—a currently debated point in scientific discussions. Prior to correlation estimation, the NAO time series had been pre-processed by filtering and interpolating the record (unequal times) to the time points of the flow record ($\bar{d} = 2.2$ a); the alternative method would have been the binning procedure (Section 7.5.1). Röthlisberger et al. (2008) also used PearsonT to study the coupling between variations of Antarctic temperature and sea ice extent in the ice-age climate over the past 800 ka. The proxy information, about temperature from δD and about sea ice from the flux of seasalt Na (Fig. 1.5), comes from EPICA Dome C, that is, the ice core with the currently longest time span of climate information. The finding was that during mild climate stages the correlation is strong, while during cold glacial conditions it is weaker (but still significant). Mudelsee (2003), introducing PearsonT, re-assessed the Sun–monsoon relation on Holocene timescales documented by δ^{18}O variations measured in a stalagmite from Oman (Neff et al. 2001). He chose the interpolation (unequal times) instead of the more appropriate binned or synchrony methods. He showed that the tuning of the $\{t(i)\}$ of the monsoon proxy changed a nonsignificant correlation into a significant value, emphasizing, however, that the size of the time shifts of the tuning was smaller than the dating errors.

The **Sun–climate relationship** on decadal timescales, in the interval from roughly the middle of the nineteenth century to the present, has been the subject of intense discussions over the past years. In addition to the anthropogenic warming signal (Section 3.8), there may exist a warming signal caused by solar activity variations (Fig. 2.12). Two original papers were in the focus. Friis-Christensen and Lassen (1991) claimed the existence of a close association between the changes of the period of the sunspot cycle and the variations of northern hemisphere land surface-air temperature for the interval from 1866 to 1985. The estimation of changes of period may be done using methods from nonstationary spectral analysis (Section 5.3), although Friis-Christensen and Lassen (1991) preferred a simpler method via smoothing (Gleissberg 1944) and taking the time differences between maxima and minima. The missing link in the Sun–climate relationship was later (Svensmark and Friis-Christensen 1997) suggested to consist of variations of the galactic cosmic ray flux influencing global cloud coverage. A series of comments, criticisms and replies to those findings were published. The impression of the author of this book is that the latest accusations by Laut (2003) and Damon and Laut (2004) against the two original papers, which include “unacceptable handling of observational data” (Damon and Laut 2004: p. 374 therein), have not been refuted in the peer-reviewed literature. This impression has been supported by Laut P (2009, personal communication), while Friis-Christensen E (2009, personal communication) has added that an earlier exchange of arguments (Laut and Gundermann 2000; Lassen and Friis-Christensen 2000) already includes his reply. One point is that the northern hemisphere temperature record has been smoothed with a filter, apparently using pseudodata at the lower and upper interval bounds (Jones et al. 1986: Fig. 5 therein), and also the sunspot cycle record has been smoothed with a filter, using a technique equivalent to a boundary kernel. Both methods of boundary-bias reduction are described in the context of the inhomogeneous Poisson process (Section 6.3.2.3). They are standard techniques in time series analysis, and insofar as the quotation regards the use of those for graphical purposes, we think, contrary to Damon and Laut (2004), that the usage is acceptable. Regarding the influence of galactic cosmic rays on climate, IPCC–WG I (Forster et al. 2007: p. 193 therein) reports: "However, there appears to be a small but statistically significant positive correlation between cloud over the UK and galactic cosmic ray flux during 1951 to 2000 (Harrison and Stephenson 2006). Contrarily, cloud cover anomalies from 1900 to 1987 over the USA do have a signal at 11 years that is anti-phased with the galactic cosmic ray flux (Udelhofen and Cess 2001). Because the mechanisms are uncertain, the apparent relationship between solar variability and cloud cover has been interpreted to result not only from changing cosmic ray fluxes modulated by solar activity in the heliosphere (Usoskin et al. 2004) and solar-induced changes in ozone (Udelhofen and Cess 2001), but also from sea surface temperatures altered directly by changing total solar irradiance (Kristjánsson et al. 2002) and by internal variability due to the El Niño–Southern Oscillation (Kernthaler et al. 1999). In reality, different direct and indirect physical processes [...] may operate simultaneously." A statistical analysis of the association between solar cycle length and temperature on basis of the original data (Fig. 7.10) may shed some light on the issue. Let $X(i)$ denote cycle length and $Y(i)$ temperature. Using the digitized data and omitting the earliest solar data point, for which no corresponding temperature point exists in the original paper, yields $n = 23$; the spacing of the resulting bivariate series is, as the cycle length, not constant. Pearson's correlation coefficient is $r_{XY} = -0.956$. Persistence time estimation with bias correction yields $\widehat{\tau}'_X = 45\,\mathrm{a}$ with 90% percentile CI $[9\,\mathrm{a}; 78\,\mathrm{a}]$ and $\widehat{\tau}'_Y = 106$ a $[8\,\mathrm{a}; 128\,\mathrm{a}]$. The lag-1 scatterplots (Fig. 7.10c, d) show the residuals (Eq. 2.12) to reflect clearly less autocorrelation than the original data, attesting to the suitability of the AR(1) persistence model. The large persistence times come obviously from the high amount of smoothing performed on both records. The effective data size is $n'_\rho = 2.13$. This tiny value prohibits any interpretation of a determined association; the large absolute value of r_{XY} may well be spurious. Insofar as the quotation from Damon and Laut (2004) regards the criticism of oversmoothing prior to correlation estimation, we think they are completely right; see also Pyper and Peterman (1998). A recent review (Lockwood and Fröhlich 2007) found that since 1987, trends in solar climate forcings and the global mean surface-air temperature go in opposite directions. Evidently, it should make sense to study the unsmoothed records and other (proxy) documents of solar and climate variability. One may extend the view further back in time and employ also climate models as analysis tools (Meehl et al. 2003). Bootstrap CIs should be helpful for assessing better the results quantitatively—not only those of claimed associations but also those of opposing trends. A second recommendation is to consider the application of multiple tests (Section 5.2.5.1), since the selection of the time intervals and the type of pre-processing gives the researcher additional freedom. These measures would allow to test harder the Sun–climate relationship on decadal timescales, despite the fact that the existing knowledge seems not to allow to expect a physically significant effect.

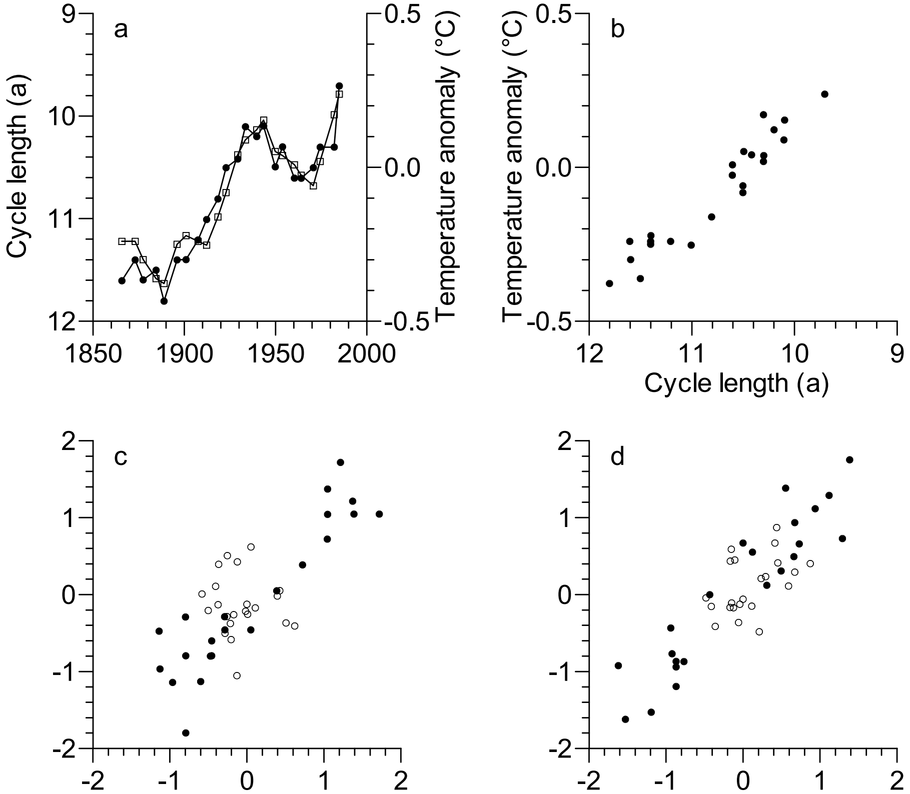


Figure 7.10. Solar cycle length and northern hemisphere land surface-air temperature anomalies, 1866–1985. **a** Time series (smoothed) of cycle length (*open symbols*) and temperature anomaly (*filled symbols*); **b** scatterplot between cycle length and temperature anomalies; **c** lag-1 scatterplots, standardized cycle length (*filled symbols*) and standardized cycle-length residuals (*open symbols*); **d** lag-1 scatterplots, standardized temperature anomalies (*filled symbols*) and standardized temperature-anomaly residuals (*open symbols*). (The time series are digitized values from Friis-Christensen and Lassen (1991: Fig. 2 therein).)

Climatological applications of hypothesis tests with autocorrelation adjustment include the following. Rothman (2002) examined the correlation between $X(i)$: strontium isotopic ratio and $Y(j)$: isotopic fractionation between total organic carbon and sedimentary carbonates over the past 500 Ma. Both variables ($n_X = 48, n_Y = 46$) were measured on independent samples of marine sedimentary rocks and have therefore independent timescales. The objective of the correlation analysis was to derive a proxy for variations of atmospheric CO_2 concentration over such long geological periods. The author interpolated the $X(i)$ values to the $T_Y(j)$ times and calculated Spearman's rank correlation coefficient as $r_S = -0.4$. He used the frequency-domain bootstrap (Section 5.3) to take autocorrelation effects into account (Rothman 2001) and determined a one-sided P-value of 0.17. The accuracy of the P-value may be influenced by the following factors: (1) prior smoothing of $X(i)$ had been applied, (2) possibly a second interpolation (to equidistance) for calculating the periodogram (Chapter 5) was necessary, (3) the data size is limited and (4) the unequal timescales and the interpolation may have introduced a negative bias into r_S (Section 7.5). The net effect is not clear: while factor (1) would tend to let P increase, factor (4) would let P decrease and, hence, raise the level of confidence. Pyper and Peterman (1998) used their approach via n'_ρ to test H_0: "$\rho_{XY} = 0$" for bivariate samples of the survival rate of different stocks of salmon from a bay in Alaska (time interval from 1957 to 1989, annual resolution). Edwards and Richardson (2004) study with the same approach the relation between the interannual variation in the timing of the seasonal cycle for various functional groups (e.g., diatoms or dinoflagellates) and sea-surface temperature, time interval 1958–2002. They find significant correlations, which underline the impact of climate change on marine pelagic phenology.

Causality is not the same as correlation. However, that philosophical concept (Chapter 1) of the association between an action (variable X) and a reaction (variable Y) should require the time arrow and be related to the theme of this chapter. Overviews of this relation have been published in statistics (Barnard 1982; Glymour 1998) and physics literature (Paluš and Vejmelka 2007). One quantitative formulation of the concept of causality comes from information theory and uses the idea of predictability: "We say that $Y(i)$ is causing $X(i)$ if we are better able to predict $X(i)$ using all available information than if the information apart from $Y(i)$ had been used" (Granger 1969: p. 428 therein). Inference about this "Granger causality" requires the analysis of time series, $\{t(i), x(i), y(i)\}_{i=1}^{n}$, and may employ statistical models, linear or nonlinear, possibly with a time-lag parameter (Granger and Lin 1994; Stern and Kaufmann 2000; Triacca 2007), see also Chapter 8. Climatological examples are the following. A bivariate linear regression model was fitted to temperature time series from the northern ($X(i)$) and the southern ($Y(i)$) hemisphere, covering the interval from 1865 to 1964 (Kaufmann and Stern 1997; Stern and Kaufmann 1999). From the estimated time lag between the two variables, the authors concluded the existence of a south-to-north causal order "generated by anthropogenic activities that increase the concentration of greenhouse gases globally, but which increase the concentration and effects of sulphate aerosols mainly in the Northern Hemisphere" (Kaufmann and Stern 1997: p. 42 therein). This conclusion was criticized as inconclusive by Triacca (2001), who preferred the direct demonstration of Granger causality of CO_2 changes on temperature changes; that, however, had been done by Tol and de Vos (1998) using a linear regression model with a prescribed lag. This simple model type has also been utilized for demonstrating an ocean feedback (daily wintertime sea-surface temperature) on the NAO, performed (Mosedale et al. 2006) using a 50-year long simulation from the climate model HadCM3 (Fig. 1.9). More advanced, nonlinear descriptions result from employing the mutual information (Fraser and Swinney 1986; Granger and Lin 1994),

$$I_{XY} = \int \int f(x,y) \ \log \left[\frac{f(x,y)}{f(x)\, f(y)} \right] \, dx \, dy. \tag{7.63}$$

Assuming that the logarithm is taken to the base of two, I_{XY} quantifies how many bits of information about X can be predicted on the basis of a sample of Y. The concept of mutual information has been extended to higher dimensions and related to properties of chaotic systems (Prichard and Theiler 1995). One such extension, which is called generalized redundancy, was employed (Diks and Mudelsee 2000) to study causal relations between variables of the Plio- to Pleistocene climate. The P-value of the test of the null hypothesis "zero information content" (no Granger causality) was determined using SB resampling (Diks and DeGoede 2001). One finding (Diks and Mudelsee 2000), from interpolated series, was that changes of δ^{18}O (a proxy for ice volume) do Granger cause changes of δ^{13}C (a proxy for the strength of formation of North Atlantic Deep Water), and that this coupling did increase towards the late Pleistocene. Other information-theoretic measures can be applied when three variables, X, Y and Z, are available; an analysis of data covering the past 400 years found that solar activity variations seem to "account for a smaller-scale behavior of global temperatures than greenhouse gases" (Verdes 2005: p. 026222-7 therein). A recent review of causality detection using information-theoretic methods (Hlaváčková-Schindler et al. 2007) gives more examples from climatology. Barnard (1982: p. 387 therein) notes also that "causation does not necessarily imply correlation as the latter is usually measured." He gives the simple nonlinear model,

$$Y(i) = \sin\left(\pi X(i)\right), \tag{7.64}$$

with $X(i)$ uniformly distributed over $[-1; +1]$, where also $Y(i)$ varies between -1 and $+1$; this model has $\rho_{XY} = 0$. The design of suitable dependence measures for nonlinear processes, alternatives to Pearson's or Spearman's linear measures, has something of an art. Granger et al. (2004: pp. 651–652 therein) propose that a measure should have the following properties:

1. It is well defined for both continuous and discrete variables.

2. It is *normalized* to zero if X and Y are independent, and lies between 0 and $+1$.

3. The modulus of the measure is equal to unity (or a maximum) if there is a *measurable* exact (nonlinear) relationship, $Y = m(X)$ say, between the random variables.

4. It is equal to or has a simple relationship with the (linear) correlation coefficient in the case of a bivariate normal distribution.

5. It is *metric,* i.e., it is a true measure of 'distance' and not just of divergence.

6. The measure is *invariant* under continuous and strictly increasing transformations $\Psi(\cdot)$. This is useful since X and Y are independent if and only if $\Psi(X)$ and $\Psi(Y)$ are independent. Invariance is important since otherwise clever or inadvertent transformations would produce different levels of dependence."

Granger et al. (2004) studied several dependence measures for many nonlinear models by means of Monte Carlo simulations.

7.7 Technical issues

The **variance of Spearman's rank correlation coefficient** is for binormal processes (David and Mallows 1961: Eq. (Z) therein)

$$
\begin{aligned}
VAR\left[r_{\mathrm{S}}\right] \approx & \frac{1}{n-1}+\frac{36}{\pi^2 n\,(n-1)\,(n+1)^2} \\
& \times \Big[n^3\Big(-0.42863279\rho_{XY}^2+0.08354697\rho_{XY}^4+0.04257246\rho_{XY}^6 \\
& +0.01687474\rho_{XY}^8+0.00664071\rho_{XY}^{10}+0.00270655\rho_{XY}^{12}\Big) \\
& +n^2\Big(0.1551301\rho_{XY}^2-0.057362293\rho_{XY}^4-0.18443407\rho_{XY}^6 \\
& -0.02271732\rho_{XY}^8+0.00757524\rho_{XY}^{10}+0.01329883\rho_{XY}^{12}\Big) \\
& +n\Big(0.36837259\rho_{XY}^2+0.44738882\rho_{XY}^4-0.08427574\rho_{XY}^6 \\
& -0.27929901\rho_{XY}^8-0.19943375\rho_{XY}^{10}-0.1386106\rho_{XY}^{12}\Big) \\
& +0.07179677\rho_{XY}^2+0.06467162\rho_{XY}^4+0.21015257\rho_{XY}^6 \\
& +0.28589798\rho_{XY}^8+0.31704425\rho_{XY}^{10}+0.07923733\rho_{XY}^{12}\Big].
\end{aligned}
\tag{7.65}
$$

PearsonT (Mudelsee 2003) is a Fortran 90 program for calculating r_{XY} with BCa CI from pairwise-MBB resampling. The software is available at the web site for this book.

Chapter 8

Regression II

Regression serves in this chapter to relate two climate variables, $X(i)$ and $Y(i)$. This is a standard tool for formulating a quantitative "climate theory" based on equations. Owing to the complexity of the climate system, such a theory can never be derived alone from the pure laws of physics—it requires to establish empirical relations between observed climate processes.

Since not only $Y(i)$ but also $X(i)$ are observed with error, the relation has to be formulated as an errors-in-variables model, and the estimation has to be carried out using adaptions of the OLS technique. This chapter focuses on the linear model and studies three estimation techniques (denoted as OLSBC, WLSXY and Wald–Bartlett procedure). It presents a novel bivariate resampling approach (pairwise-MBBres), which enhances the coverage performance of bootstrap CIs for the estimated regression parameters.

Monte Carlo simulations allow to assess the role of various aspects of the estimation. First, prior knowledge about the size of the measurement errors is indispensable to yield a consistent estimation. If this knowledge is not exact, which is typical for a situation in the climatological practice, it contributes to the estimation error of the slope (RMSE and CI length). This contribution persists even when the data size goes to infinity; the RMSE does then not approach zero. Second, autocorrelation has to be taken into account to prevent estimation errors unrealistically small and CIs too narrow.

This chapter studies two extensions of high relevance for climatological applications: linear prediction and lagged regression.

Regression as a method to estimate the trend in the climate equation (Eq. 1.2) is presented in Chapter 4.

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_8,

8.1 Linear regression

To make a regression of the predictor variable, X, on the response variable, Y, we re-apply the errors-in-variables model (Section 4.1.7),

$$Y(i) = \beta_0 + \beta_1 \left[X(i) - S_X(i) \cdot X_{\text{noise}}(i)\right] + S_Y(i) \cdot Y_{\text{noise}}(i), \qquad (8.1)$$

$i = 1, \ldots, n$. The variability of process $X(i)$ and $Y(i)$ is denoted as $S_X(i)$ and $S_Y(i)$, respectively; the noise component, $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$, is of assumed AR(1) type with persistence time τ_X and τ_Y, respectively. One task is to estimate the regression parameters, β_0 and β_1, given a bivariate sample, $\{t(i), x(i), y(i)\}_{i=1}^{n}$. Another, related task is to make a prediction of an unknown Y for a given value of X.

The errors-in-variables model (Eq. 8.1) differs from the simple model (Eq. 4.3) in its nonzero noise component of the predictor. Several estimators for the errors-in-variables model have been developed to deal with this more complex situation.

8.1.1 Ordinary least-squares estimation

The simple OLS estimation minimizes the unweighted sum of squares,

$$SSQ(\beta_0, \beta_1) = \sum_{i=1}^{n} \left[y(i) - \beta_0 - \beta_1 x(i)\right]^2 . \qquad (8.2)$$

This yields the estimators

$$\widehat{\beta}_0 = \left[\sum_{i=1}^{n} y(i) - \widehat{\beta}_1 \sum_{i=1}^{n} x(i)\right] \Big/ n \qquad (8.3)$$

and

$$\widehat{\beta}_1 = \left\{ \left[\sum_{i=1}^{n} x(i)\right] \left[\sum_{i=1}^{n} y(i)\right] \Big/ n - \sum_{i=1}^{n} x(i)\, y(i) \right\} \times \left\{ \left[\sum_{i=1}^{n} x(i)\right]^2 \Big/ n - \sum_{i=1}^{n} x(i)^2 \right\}^{-1} . \qquad (8.4)$$

Using OLS means ignoring heteroscedasticity, persistence and errors in the predictor variable, X. However, heteroscedasticity and persistence can successfully be taken into account by employing WLS and GLS estimation, respectively. The success of ignoring errors in X depends on how large these are relative to the spread of the "true" X values (Eq.

4.34), which are given by $X_{\text{true}}(i) = X(i) - S_X(i) \cdot X_{\text{noise}}(i)$. If $S_X(i) = S_X$ is constant and $S_X^2 \ll VAR[X_{\text{true}}(i)]$, the estimation bias should be negligible. If $S_X(i)$ is not constant, one may expect a similar condition to the average of $S_X(i)$. The decisive quantity is $VAR[X_{\text{true}}(i)]$, which may be difficult to control for an experimenter prior to sampling the process.

If $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ are independent, the estimator $\widehat{\beta}_1$ is biased downwards (Section 4.1.7) as $E\left(\widehat{\beta}_1\right) = \kappa \cdot \beta_1$, where $\kappa \leq 1$ is the attenuation factor or reliability ratio,

$$\kappa = \left(1 + S_X^2 \,/ VAR\left[X_{\text{true}}(i)\right]\right)^{-1}. \tag{8.5}$$

The intuitive reason of the bias downwards is that "smearing" the "true" predictor variable, $X_{\text{true}}(i)$, leads to a situation where the "cheapest fit solution" in terms of SSQ is a line that is horizontally tilted (Fig. 8.1).

8.1.1.1 Bias correction

Eq. (8.5) points to a bias-corrected slope estimation. Let $S_X(i) = S_X$ be constant and known, and let the variance of the "true" predictor values be given by $VAR[X_{\text{true}}(i)] = VAR[X(i)] - S_X^2$. This leads to

$$\widehat{\beta}_1 = \widehat{\beta}_{1,\text{OLS}} \,/ \left\{1 - S_X^2 \,/ VAR\left[X(i)\right]\right\}, \tag{8.6}$$

where $\widehat{\beta}_{1,\text{OLS}}$ is the simple OLS slope estimator (Eq. 8.4). We denote this estimation method (Eq. 8.6) as ordinary least squares with bias correction (OLSBC). The OLSBC intercept estimator equals the OLS intercept estimator (Eq. 8.3). In practice (sample level), plug in $x(i)$ for $X(i)$.

8.1.1.2 Prior knowledge about standard deviations

Assume homoscedastic noise components, $S_Y(i) = S_Y$ and $S_X(i) = S_X$, and denote their squared ratio as

$$\lambda = S_Y^2 \,/ S_X^2. \tag{8.7}$$

Knowledge prior to the estimation about S_X, S_Y or λ can increase the estimation accuracy.

If S_X is known, then OLSBC can be readily performed (Eq. 8.6). Such prior knowledge may be acquired, for example, by repeating measurements. Or there may exist theoretical information about the measuring device and, hence, S_X.

If S_X is only known within bounds, OLSBC estimation can still be applied. CI construction has then to take into account the limited prior

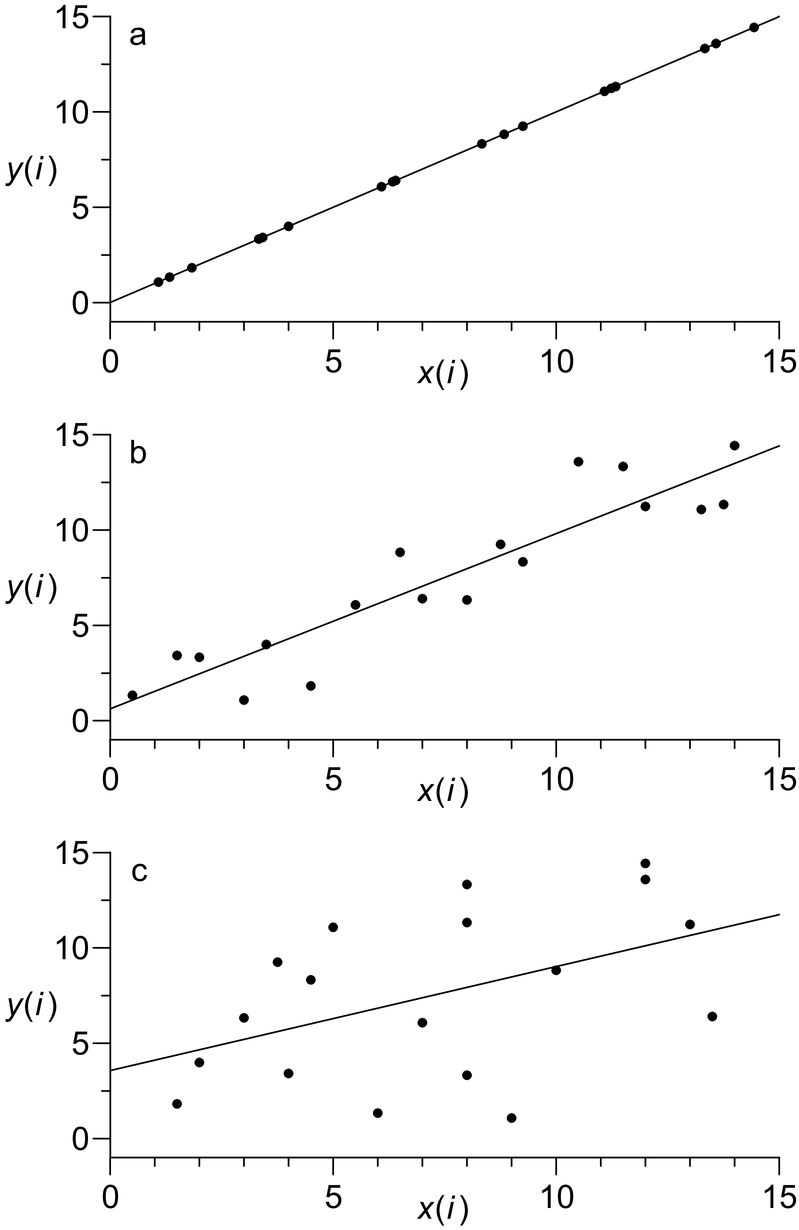


Figure 8.1. Linear errors-in-variables regression model, OLS estimation. The $\{y(i)\}_{i=1}^{n}$ are identical in panels **a–c;** the data size is $n = 18$; and the $\{x(i)\}_{i=1}^{n}$ are realizations of a predictor variable, $X(i)$, with constant zero (**a**), small (**b**) and large (**c**) noise components, $S_X(i) \cdot X_{\text{noise}}(i)$. The true slope is $\beta_1 = 1.0$ (**a**). The OLS fits (*solid lines*) exhibit slope estimates that are unbiased (**a** $\widehat{\beta}_1 = 1.0$) or biased (**b** $\widehat{\beta}_1 = 0.92$; **c** $\widehat{\beta}_1 = 0.55$).

knowledge. The result is a wider CI compared to the situation of perfect prior knowledge (Section 8.3).

If only the ratio, λ, is known, then one may be tempted to employ the method of moments estimator from the background material (Eq. 8.26) and plug in $\widehat{S}_X$ for S_X in Eq. (8.6). Similarly, if only S_Y is known, then one may be tempted to employ Eq. (8.26), replace therein $\delta = \lambda^{1/2}$ by $S_Y/\widehat{S}_X$ and solve the equation for $\widehat{S}_X$. However, own Monte Carlo experiments (results not shown) revealed completely inacceptable coverage accuracies of bootstrap confidence intervals for the slope (but acceptable accuracies for the intercept). The reason is the inaccurate $\widehat{S}_X$ estimation (Fuller 1987: Section 2.5 therein). Our recommendation for the case of known λ (or S_Y) is the weighted least-squares estimation (Section 8.1.2).

If no knowledge at all exists about S_X, S_Y or λ, then we face difficulties. One may simply try OLS but risk a biased slope estimation. One may resort to the Wald–Bartlett procedure (Section 8.1.3), but also this does not produce accurate results when so little is known. We discourage from adopting an OLS regression of Y an X and estimating S_X via the residual mean square (Eq. 4.8), an idea found occasionally in the literature. Own Monte Carlo experiments (similar to those in Section 8.3, results not shown) revealed inacceptable coverage performance of bootstrap CIs.

8.1.2 Weighted least-squares for both variables estimation

Studying the combination of both noise components in Eq. (8.1) in the form of $S_Y(i) \cdot Y_{\text{noise}}(i) - \beta_1\, S_X(i) \cdot X_{\text{noise}}(i)$ makes clear the estimation approach via attaching weights to the observations of both variables (Deming 1943; Lindley 1947). The variant by York (1966) and others, who suggested minimization of the weighted least-squares sum,

$$SSQWXY(\beta_0, \beta_1) = \sum_{i=1}^{n} \frac{[y(i) - \beta_0 - \beta_1 x(i)]^2}{S_Y(i)^2 + \beta_1^2\, S_X(i)^2}, \tag{8.8}$$

was included in the Numerical Recipes (Press et al. 1992: Section 15.3 therein). However, no general analytical solution exists and some numerical difficulties have to be circumnavigated (Section 8.8). We abbreviate this estimation procedure as WLSXY (Fig. 8.2).

8.1.2.1 Prior knowledge about standard deviation ratio

Assume $S_Y(i)$ and $S_X(i)$ to be unknown, but their (squared) ratio,

$$\lambda = S_Y(i)^2 / S_X(i)^2, \tag{8.9}$$

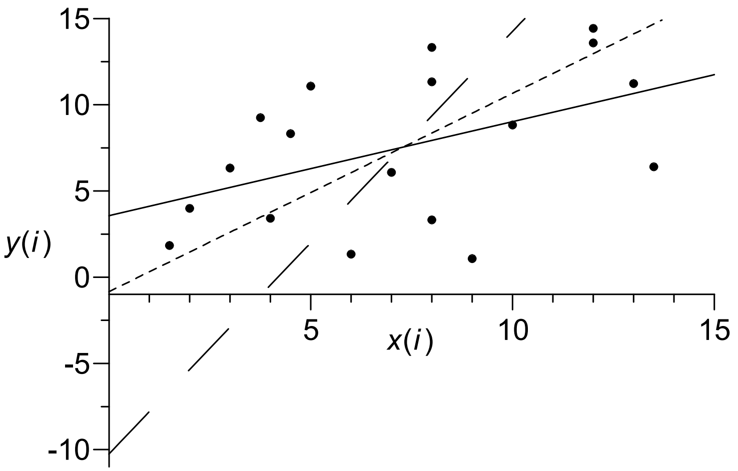


Figure 8.2. Linear errors-in-variables regression model, WLSXY and OLS estimations. The $\{x(i), y(i)\}_{i=1}^{n}$ are overtaken from Fig. 8.1c. The OLS fit of X on Y (*solid line*) has a slope of $\widehat{\beta}_1 = 0.55$, the OLS fit of Y on X (*long-dashed line*) has $1/\widehat{\beta}_1' = 2.45$ and the WLSXY fit of X on Y (*short-dashed line*) has $\widehat{\beta}_1 = 1.15$. (The model for the regression of Y on X is $X(i) = \beta_0' + \beta_1' Y(i) + S_X \cdot X_{\text{noise}}(i)$.)

to be constant and known. Such type of knowledge may be available in climatological applications. Then,

$$SSQWXY(\beta_0, \beta_1) = \sum_{i=1}^{n} \frac{\left[y(i) - \beta_0 - \beta_1 x(i)\right]^2}{\left(1 + \beta_1^2 / \lambda\right) S_Y(i)^2}, \tag{8.10}$$

which is minimized (Section 8.8). The sub-case of constant $S_Y(i)$ (or $S_X(i)$) is considerably easier to treat than the heteroscedastic sub-case.

Under the assumption of Gaussian distributional shapes of $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$, the WLSXY estimators equal the maximum likelihood estimators (Madansky 1959; Fuller 1987).

8.1.2.2 Geometric interpretation

WLSXY minimizes the sum of squares of distances between the fit line and the data points. How to measure the distance depends on the ratio, $\lambda = S_Y(i)^2 / S_X(i)^2$. The geometric interpretation is straightforward (Fig. 8.3) and generalizable to higher dimensions (background material).

If $S_X(i) = 0$, that means, the $X(i)$ values are exact, then $\lambda = \infty$ and we use WLS regression of X on Y (Section 4.1.1); if further $S_Y(i)$ is constant, this amounts to OLS regression. On the other hand, if $S_Y(i) = 0$, then $\lambda = 0$ and we use WLS regression of Y on X. (See Fig.

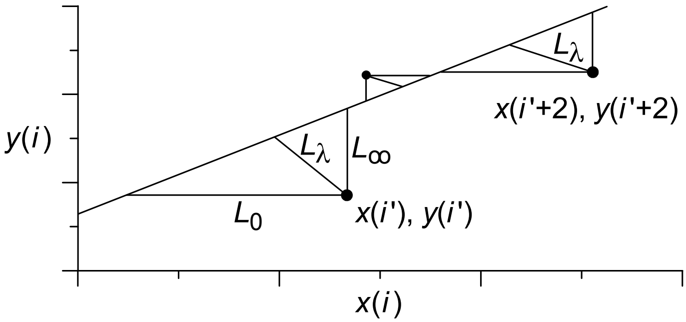


Figure 8.3. Geometric interpretation of WLSXY. The lines L_0, L_λ and L_∞ measure the distance from a data point to the fit line for $\lambda = 0$, $0 < \lambda < \infty$ and $\lambda = \infty$, respectively.

8.2 for the regression of Y on X.) If the standard deviations are nonzero and $0 < \lambda < \infty$, we measure the distance along the line L_λ (Fig. 8.3). The slope of this line is equal to $-\lambda/\widehat{\beta}_1$ (York 1967).

If heteroscedasticity is in one or both of the noise components, then the ratio λ may vary with time (i) and, hence, the line L_λ may vary in its slope. The difficulty of non-identifiability is introduced by unknown λ because then it is not unequivocally determined how to measure the distance and minimize the sum of squares.

8.1.3 Wald–Bartlett procedure

A straightforward estimation idea (Draper and Smith 1981: Section 2.14 therein) is to build two groups of the bivariate sample according to the size of the x values, then to take for each group the centres defined by the x and y averages and, finally, to connect the centres using a straight line—defining the estimate of the slope. The intercept estimate is found via the centre of the complete bivariate sample and the slope estimate. This goes back to Wald (1940), who grouped the sample into two halves of same size (if n is even) and Bartlett (1949), who showed that taking three groups improves the accuracy of the regression estimators. (Intuitively, the means of the two groups are further apart for taking thirds than for taking halves, outweighing the deficit of reduced data sizes.) We call this estimation Wald–Bartlett procedure (Fig. 8.4).

The Wald–Bartlett procedure can in principle be applied to *any* grouping of the set of data points, not only according to the size of the x values. A point to note is that the grouping has to be independent of $X_{\mathrm{noise}}(i)$ for achieving consistency of the estimators (Wald 1940). This condition is violated when the $\{X_{\mathrm{true}}(i)\}_{i=1}^{n}$ are unknown and the size ordering

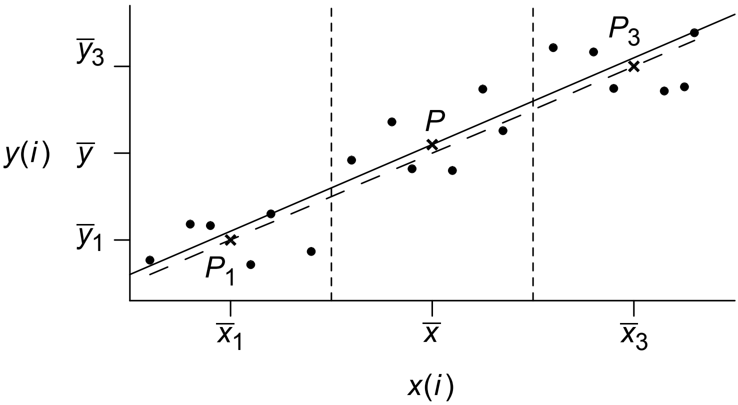


Figure 8.4. Wald–Bartlett procedure. The bivariate sample $\{x(i), y(i)\}_{i=1}^{n}$ is divided into three groups of same size according to the size of the x values; if n is not divisible by 3, then take the closest grouping. Let j index the size-sorted sample. Let the averages of $\{x(j)\}_{j=1}^{n/3}$ and $\{y(j)\}_{j=1}^{n/3}$, denoted as $\bar{x}_1$ and $\bar{y}_1$, define the first group's centre (P_1, *cross*), and let the averages of $\{x(j)\}_{j=2n/3+1}^{n}$ and $\{y(j)\}_{j=2n/3+1}^{n}$, denoted as $\bar{x}_3$ and $\bar{y}_3$, define the third group's centre (P_3). The line $P_1 P_3$ (*long-dashed*) defines the Wald–Bartlett regression estimate of the slope, $\widehat{\beta}_1 = (\bar{y}_3 - \bar{y}_1)/(\bar{x}_3 - \bar{x}_1)$. The centre of the complete sample (P) is defined via the averages of $\{x(j)\}_{j=1}^{n}$ and $\{y(j)\}_{j=1}^{n}$, denoted as $\bar{x}$ and $\bar{y}$. The Wald–Bartlett intercept estimate, $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$, completes the linear fit (*solid line*).

is made on the noise-influenced observations. Monte Carlo simulations, similar to those in Section 8.3, reveal that the inconsistency leads to an inacceptably poor coverage performance of bootstrap CIs (for $\widehat{\beta}_0$ and $\widehat{\beta}_1$) (not shown). This limits severely the applicability of the Wald–Bartlett procedure to real-world climatological problems, where the $X_{\text{true}}(i)$ are usually unknown.

Wald (1940: p. 298 therein) notes that if prior knowledge exists on the standard deviation ratio, then a consistent estimation could be constructed. This situation is similar to WLSXY estimation (Section 8.1.2.1).

The calculation of classical CIs (Wald 1940; Bartlett 1949) via the Student's t distribution assumed prior knowledge to be available, allowing a consistent estimation, and the errors, $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$, to be serially independent and of Gaussian shape.

8.2 Bootstrap confidence intervals

Classical CIs are based on the PDF of an estimator (Chapter 3). The PDF can be analytically determined unless the situation (estimation problem, noise properties) becomes too complex. The construction of classical CIs for the linear errors-in-variables model (Wald 1940; Bartlett 1949; York 1966; Fuller 1987) made a number of assumptions from the following:

1. Gaussian distributional shapes of the noise components, $X_{\mathrm{noise}}(i)$ and $Y_{\mathrm{noise}}(i)$;

2. absence of autocorrelation in the noise components;

3. absence of correlation between $X(i)$ and $X_{\mathrm{noise}}(i)$ as well as between $Y(i)$ and $Y_{\mathrm{noise}}(i)$;

4. absence of correlation between $X_{\mathrm{noise}}(i)$ and $Y_{\mathrm{noise}}(i)$.

Some authors treat the correlation effects (points 3 and 4) and non-Gaussian errors (point 1), see the background material (Section 8.7). However, allowance for autocorrelations (point 2) seems to have been made by none.

Here we are interested in linearly relating two climate processes, $X(i)$ and $Y(i)$, and our sample, $\{t(i), x(i), y(i)\}_{i=1}^{n}$, contains the time. The previous chapters document that non-Gaussian distributions and persistence phenomena are typical of climate processes. We cannot therefore expect the classical method to yield accurate results for climate data. This is, as in previous chapters, the reason to consider the bootstrap method. An additional point is incomplete knowledge about the noise components. Often we have no or only limited information about $S_Y(i)$, $S_X(i)$ or their (squared) ratio, λ. Such incomplete knowledge, which may widen the CI, is quantifiable using bootstrap resampling (Booth and Hall 1993).

One resampling algorithm is the pairwise-MBB, which has been found useful in the context of correlation estimation (Algorithm 7.2).

The other algorithm, introduced here for the purpose of enhancing the coverage performance in the context of fitting errors-in-variables models, is called pairwise-moving block bootstrap resampling of residuals or pairwise-MBBres. It is based on the observation that the (linear) errors-in-variables regression (Eq. 8.1) is a model with a deterministic (linear) component. Since pairwise resampling seems to be handicapped in the presence of deterministic components (Chapter 4), the idea of the pairwise-MBBres algorithm is to take the fit and the regression residuals, apply pairwise-MBB to the residuals and add the resampled residuals to the fit. The new approach is that the residuals, $e_X(i)$ and $e_Y(i)$, are in two dimensions (Fig. 8.5). The pairwise-MBBres is given as Algorithm 8.1.

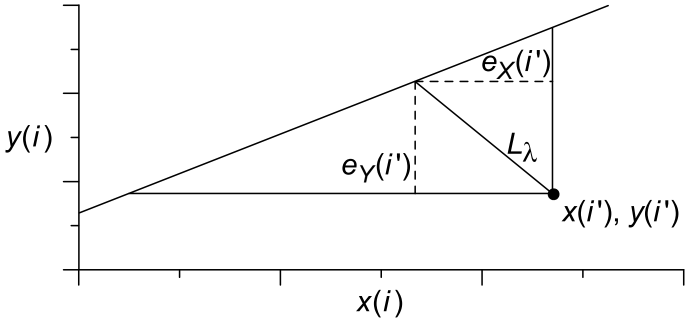


Figure 8.5. Pairwise-MBBres algorithm, definition of residuals. The line L_λ measures the distance from a data point to the fit line (Fig. 8.3). The residuals (*dashed lines*) are given by $e_X(i) = [\widehat{\beta}_0 + \widehat{\beta}_1 \cdot x(i) - y(i)]/[\lambda/\widehat{\beta}_1 + \widehat{\beta}_1]$ and $e_Y(i) = -\lambda \cdot e_X(i)/\widehat{\beta}_1$.

8.2.1 Simulating incomplete prior knowledge

Assume for the convenience of exposition homoscedastic noise components, $S_X(i) = S_X$ and $S_Y(i) = S_Y$. For achieving an identifiable problem, OLSBC estimation requires information, not contained in the sample, about S_X (Section 8.1.1.2); analogously, WLSXY estimation requires information about both S_X and S_Y, or about their ratio, $\delta = \lambda^{1/2} = S_Y/S_X$ (Section 8.1.2.1).

In practical applications such prior knowledge is not always exact. S_X or δ are then described by random variables. Bootstrap resampling can be augmented by a simulation step, where random numbers are drawn from the distribution of S_X or δ. This increases the uncertainty of the OLSBC or WLSXY estimates, leading to wider bootstrap CIs compared to a situation with exact prior knowledge (Booth and Hall 1993).

In the Monte Carlo experiments studying incomplete prior knowledge, we use the model

$$\sqrt{\lambda^*} = \delta^* = \delta \cdot \mathcal{E}_{\mathrm{U}[1-\Delta;\,1+\Delta]}(i), \tag{8.11}$$

$$S_X^* = S_X/\sqrt{\delta^*}, \tag{8.12}$$

where $\mathcal{E}_{\mathrm{U}[1-\Delta;\,1+\Delta]}(i)$ is an IID random process with a uniform distribution over the interval $[1 - \Delta;\ 1 + \Delta]$. For example, $\Delta = 0.5$ specifies a situation where we only know δ to lie between 0.5 and 1.5 times its true value. Other models are possible.

The construction of bootstrap CIs (Algorithm 8.1) is adapted at Step 8, the calculation of the replications. Instead of applying to the resampled data the same estimation procedure that is used for the original data, an adapted estimation is performed (Steps 8a and 8b).

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Parameter estimates	$\widehat{\beta}_0, \widehat{\beta}_1$
	from OLSBC, WLSXY or	
	Wald–Bartlett procedure	
Step 3	Residuals (Fig. 8.5)	$e_X(i), e_Y(i)$
Step 4	Fit values	$x_{\text{fit}}(i) = x(i) - e_X(i)$,
		$y_{\text{fit}}(i) = y(i) - e_Y(i)$
Step 5	Bias-corrected AR(1)	
	parameters,	$\widehat{a}'_X = \widehat{a}'_Y$
	estimated on residuals,	
	block length selection	l
	after Eqs. (7.31) and (7.32)	
Step 6	Resampled residuals,	
	pairwise-MBB with l	$\left\{e_X^{*b}(i), e_Y^{*b}(i)\right\}_{i=1}^{n}$ (b, counter)
Step 7	Resample	$x^{*b}(i) = x_{\text{fit}}(i) + e_X^{*b}(i)$,
		$y^{*b}(i) = y_{\text{fit}}(i) + e_Y^{*b}(i), i = 1, \ldots, n$
Step 8	Bootstrap replications	$\widehat{\beta}_0^{*b}, \widehat{\beta}_1^{*b}$
Step 9	Bootstrap prediction	$\widehat{y}^{*b}(n+1) = \widehat{\beta}_0^{*b} + \widehat{\beta}_1^{*b} x(n+1)$
Step 10	Go to Step 6 until $b = B$	
	(usually $B = 2000$)	
	replications exist	$\left\{\widehat{\beta}_0^{*b}\right\}_{b=1}^{B}, \left\{\widehat{\beta}_1^{*b}\right\}_{b=1}^{B}, \left\{\widehat{y}^{*b}(n+1)\right\}_{b=1}^{B}$
Step 11	Calculate CIs	
	(Section 3.4)	

Algorithm 8.1. Construction of bootstrap confidence intervals for parameters of the linear errors-in-variables regression model, pairwise-MBBres resampling, even spacing. In case of uneven spacing, Step 5 uses $\widehat{\tau}'_X = \widehat{\tau}'_Y$. Step 8 can be adapted as follows for taking incomplete prior knowledge into account. Step 8a: simulate λ^*, S_X^*; Step 8b: use OLSBC with S_X^* instead of S_X (Eq. 8.6), use WLSXY with λ^* instead of λ (Eq. 8.10). Steps 9 and 10: $\widehat{y}^{*b}(n+1)$ refers to prediction (Section 8.5).

8.3 Monte Carlo experiments

The first group of experiments (Section 8.3.1) adopts an "easy" setting, where distributional shapes are Gaussian and prior knowledge is exact. The results confirm the success of block resampling methods in preserving serial dependence. The second group of experiments (Section 8.3.2) shows that for realistic settings, with non-Gaussian distributions or incomplete knowledge, the results are less exact. It appears that then WLSXY estimation combined with pairwise-MBBres resampling yields the relatively best results. The third group (Section 8.3.3) quantifies coverage accuracy and RMSE in dependence on the accuracy of the prior knowledge (standard deviation ratio). It demonstrates that even with $n \to \infty$, the RMSE (and the CI length) for $\widehat{\beta}_1$ (slope) does not go to zero, but rather approaches finite values. On the other hand (intercept), with $n \to \infty$ does $\text{RMSE}_{\widehat{\beta}_0} \to 0$. The last group (Section 8.3.4), finally, explores what happens when we mis-specify the degree of how accurately we know the standard deviation ratio.

8.3.1 Easy setting

The easy setting (Gaussian shapes of $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$, complete prior knowledge) is further simplified when no autocorrelation resides in the noise components. Table 8.1 exhibits excellent coverage performance of bootstrap CIs with pairwise-MBBres resampling already for sample sizes as small as 20. The excellent performance regards both parameters (intercept and slope) and all estimation procedures (OLSBC, WLSXY and Wald–Bartlett). Similar results for OLSBC and WLSXY were obtained using pairwise-MBB resampling (results not shown).

If there exists autocorrelation, then pairwise-MBBres resampling successfully preserves it, where it does not matter whether the AR(1) parameters are known or, which is more realistic, have to be estimated. However, there is one exception: the Wald–Bartlett estimation of the intercept fails completely, independent of the sample size. Also OLSBC and WLSXY estimations of the intercept are of reduced CI coverage accuracy, but become acceptable for n above, say, 200. It is not clear why intercept estimation is more problematic than slope estimation.

We remark that the inacceptable performance of the Wald–Bartlett procedure occurred even in the presence of knowledge of the size of the true predictor values, which in turn enabled a perfect grouping (independent of the predictor noise). The three points of the (1) inacceptable performance of CIs for the intercept, (2) not better (compared with OLSBC and WLSXY) performance of CIs for the slope and (3) rather strong requirement of knowledge of the size of true predictor values—

Table 8.1. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape and complete prior knowledge: CI coverage performance. $n_{\text{sim}} = 47{,}500$ random samples were generated from $X_{\text{true}}(i) = \mathcal{E}_{\text{N}(0,\,1)}(i)$ and $Y(i) = \beta_0 + \beta_1 X_{\text{true}}(i) + S_Y \cdot Y_{\text{noise}}(i)$, $i = 1, \ldots, n$, with $\beta_0 = 1.0$ and $\beta_1 = 2.0$. Predictor noise was subsequently added, $X(i) = X_{\text{true}}(i) + S_X \cdot X_{\text{noise}}(i)$. The $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ are mutually independent Gaussian AR(1) processes for even spacing (Eq. 2.1) with parameters a_X and a_Y, respectively. Construction of bootstrap CIs used pairwise-MBBres resampling (Algorithm 8.1), block length selection after Eqs. (7.31) and (7.32), the Student's t interval type ($\nu = n - 2$), $B = 2000$ and confidence level 95%. Prior knowledge of $S_X = 0.25, S_Y = 0.5$ and the size of the $\{x_{\text{true}}(i)\}_{i=1}^{n}$ was exact and utilized in the estimations; AR(1) parameters are in two cases known, in one case unknown and estimated with bias correction.

n	$\gamma_{\widehat{\beta}_0}$[a]			$\gamma_{\widehat{\beta}_1}$[a]			*Nominal*
	Estimation method			*Estimation method*			
	OLSBC	*WLSXY*	*WB*[b]	*OLSBC*	*WLSXY*	*WB*[b]	
	$a_X = a_Y = 0.0$ (*known*)						
10	0.933	0.928	0.933	0.955	0.928	0.938	0.950
20	0.939	0.939	0.942	0.949	0.938	0.942	0.950
50	0.944	0.947	0.948	0.949	0.946	0.947	0.950
100	0.944	0.946	0.947	0.949	0.947	0.946	0.950
200	0.945	0.949	0.950	0.949	0.947	0.945	0.950
500	0.946	0.949	0.949	0.950	0.946	0.946	0.950
1000	0.945	0.948	0.949	0.949	0.948	0.945	0.950
	$a_X = a_Y = 0.3$ (*known*)						
10	0.839	0.831	0.785	0.949	0.919	0.912	0.950
20	0.863	0.862	0.816	0.947	0.936	0.926	0.950
50	0.895	0.896	0.827	0.948	0.945	0.930	0.950
100	0.909	0.911	0.827	0.947	0.945	0.933	0.950
200	0.921	0.924	0.836	0.947	0.945	0.939	0.950
500	0.931	0.933	0.840	0.948	0.944	0.941	0.950
1000	0.935	0.938	0.844	0.950	0.949	0.945	0.950
	$a_X = a_Y = 0.3$ (*unknown, estimated*)						
10	0.807	0.797	0.841	0.922	0.892	0.935	0.950
20	0.871	0.869	0.853	0.943	0.932	0.947	0.950
50	0.896	0.897	0.852	0.946	0.943	0.947	0.950
100	0.911	0.913	0.854	0.948	0.947	0.949	0.950
200	0.921	0.923	0.853	0.949	0.948	0.949	0.950
500	0.930	0.932	0.850	0.949	0.945	0.948	0.950
1000	0.934	0.937	0.849	0.949	0.948	0.949	0.950

[a] Standard errors of $\gamma_{\widehat{\beta}_0}$ and $\gamma_{\widehat{\beta}_1}$ are nominally 0.001.

[b] Wald–Bartlett procedure.

Table 8.2. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape and complete prior knowledge: CI coverage performance (continued). Design is identical to the previous experiment (Table 8.1), with the exception that autocorrelation is ignored at CI construction.

n	$\gamma_{\widehat{\beta}_0}$ [a]			$\gamma_{\widehat{\beta}_1}$ [a]			*Nominal*
	Estimation method			*Estimation method*			
	OLSBC	*WLSXY*	*WB* [b]	*OLSBC*	*WLSXY*	*WB* [b]	
	$a_X = a_Y = 0.3$ (*ignored*)						
10	0.846	0.838	0.846	0.955	0.928	0.939	0.950
20	0.842	0.840	0.845	0.948	0.937	0.942	0.950
50	0.845	0.847	0.849	0.948	0.944	0.946	0.950
100	0.842	0.846	0.846	0.947	0.946	0.946	0.950
200	0.845	0.848	0.849	0.947	0.946	0.945	0.950
500	0.846	0.849	0.849	0.948	0.945	0.947	0.950
1000	0.846	0.848	0.849	0.946	0.948	0.946	0.950

[a] Standard errors of $\gamma_{\widehat{\beta}_0}$ and $\gamma_{\widehat{\beta}_1}$ are nominally 0.001.

[b] Wald–Bartlett procedure.

provide enough support to exclude the Wald–Bartlett procedure from consideration in the further experiments (which have more realistic settings).

One experiment under the easy setting studied what happens when autocorrelation is ignored (Table 8.2). This was achieved by prescribing positive AR(1) parameters, a_X and a_Y, of both noise components and resetting their estimate values to zero, $\widehat{a}'_X \equiv 0$ and $\widehat{a}'_Y \equiv 0$. The detrimental effect, again on $\widehat{\beta}_0$ but not $\widehat{\beta}_1$, was an underestimated bootstrap standard error, which led to too narrow CIs and too low coverages. (Similar results where found when replacing pairwise-MBBres by pairwise-MBB resampling.)

The major findings of the first experiment on CI coverage accuracy (Table 8.1) are reflected in the results on empirical RMSE (Table 8.3). Autocorrelation increases the estimation error of the intercept, but not of the slope. A larger data size means a smaller estimation error of the intercept and the slope. For $n \to \infty$, both $\mathrm{RMSE}_{\widehat{\beta}_0}$ and $\mathrm{RMSE}_{\widehat{\beta}_1}$ appear to go to zero.

OLSBC and WLSXY estimation methods perform similarly; only for small n and slope estimation does WLSXY seem to give smaller error bars than OLSBC.

Table 8.3. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape and complete prior knowledge: RMSE. Design is identical to the experiment shown in Table 8.1.

n	$\mathrm{RMSE}^{a}_{\widehat{\beta}_0}$		$\mathrm{RMSE}^{b}_{\widehat{\beta}_1}$	
	Estimation method		*Estimation method*	
	OLSBC	*WLSXY*	*OLSBC*	*WLSXY*
	$a_X = a_Y = 0.0$			
10	0.237	0.245	0.645	0.289
20	0.161	0.164	0.190	0.178
50	0.099	0.101	0.110	0.105
100	0.070	0.071	0.076	0.073
200	0.049	0.050	0.053	0.052
500	0.031	0.032	0.033	0.033
1000	0.022	0.022	0.023	0.023
	$a_X = a_Y = 0.3$			
10	0.300	0.310	5.759	0.275
20	0.211	0.216	0.188	0.174
50	0.134	0.137	0.110	0.105
100	0.094	0.096	0.076	0.073
200	0.067	0.068	0.053	0.052
500	0.042	0.043	0.034	0.033
1000	0.030	0.031	0.024	0.023

[a] Empirical $\mathrm{RMSE}_{\widehat{\beta}_0}$, given by $\left[\sum_{i=1}^{n_{\mathrm{sim}}}\left(\widehat{\beta}_0 - \beta_0\right)^2 / n_{\mathrm{sim}}\right]^{1/2}$.

[b] Empirical $\mathrm{RMSE}_{\widehat{\beta}_1}$, given by $\left[\sum_{i=1}^{n_{\mathrm{sim}}}\left(\widehat{\beta}_1 - \beta_1\right)^2 / n_{\mathrm{sim}}\right]^{1/2}$.

8.3.2 Realistic setting: incomplete prior knowledge

The setting becomes more complex, or realistic, when the prior knowledge about the standard deviations of the measurement noise is not complete. We study (Table 8.4) a situation where the true ratio is $\delta = S_Y/S_X = 2.0$ but one knows only that δ is between 1.0 and 3.0. The adapted bootstrap CI construction (WLSXY with $\lambda^* = (\delta^*)^2$), for both $\widehat{\beta}_0$ and $\widehat{\beta}_1$, yields acceptable accuracies for normal shape and $n \gtrsim 200$—under the condition that pairwise-MBBres resampling is employed. (For climatological purposes, a 95% CI may be "acceptable" if the true coverage is between, say, 92 and 98%.) The pairwise-MBBres resampling method (Fig. 8.5) is clearly superior to the pairwise-MBB method.

Bootstrap CI construction for OLSBC estimates failed to achieve the accuracies for WLSXY—for both resampling methods. Rather large

Table 8.4. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal/lognormal shape and incomplete prior knowledge: CI coverage performance. Design is identical to the first experiment (Table 8.1), with the following exceptions: (1) autocorrelation parameters are unknown (and estimated with bias correction) and (2) $Y_{\text{noise}}(i)$ has normal or lognormal shape. Estimation and CI construction is identical to the first experiment (Table 8.1), with the following exceptions: (1) the Wald–Bartlett procedure is omitted; (2) prior knowledge of $S_X = 0.25, S_Y = 0.5$ ($\delta = 2.0$) is incomplete after Eqs. (8.11) and (8.12) with $\Delta = 0.5$; and (3) CI construction is adapted accordingly (Section 8.2.1).

n	$\gamma^{\text{a}}_{\widehat{\beta}_0}$		$\gamma^{\text{a}}_{\widehat{\beta}_1}$		*Nominal*
	Estimation method		*Estimation method*		
	OLSBC	*WLSXY*	*OLSBC*	*WLSXY*	
	$a_X = a_Y = 0.3$, $Y_{\text{noise}}(i)$ *normal shape*, *pairwise-MBBres*				
10	0.809	0.802	0.941	0.895	0.950
20	0.871	0.871	0.956	0.931	0.950
50	0.898	0.899	0.955	0.940	0.950
100	0.909	0.913	0.952	0.942	0.950
200	0.921	0.924	0.948	0.943	0.950
500	0.930	0.934	0.939	0.947	0.950
1000	0.937	0.939	0.927	0.952	0.950
2000	0.936	0.939	0.919	0.958	0.950
5000	0.941	0.944	0.909	0.960	0.950
	$a_X = a_Y = 0.3$, $Y_{\text{noise}}(i)$ *normal shape*, *pairwise-MBB*				
10	0.856	0.875	0.980	0.947	0.950
20	0.864	0.867	0.972	0.944	0.950
50	0.862	0.862	0.958	0.943	0.950
100	0.866	0.866	0.955	0.943	0.950
200	0.865	0.865	0.954	0.948	0.950
500	0.871	0.871	0.943	0.947	0.950
1000	0.873	0.871	0.932	0.952	0.950
2000	0.880	0.880	0.922	0.957	0.950
5000	0.891	0.889	0.915	0.962	0.950
	$a_X = 0.3, a_Y = 0.8$, $Y_{\text{noise}}(i)$ *lognormal shape*, *pairwise-MBBres*				
10	0.689	0.673	0.942	0.871	0.950
20	0.738	0.738	0.956	0.902	0.950
50	0.788	0.793	0.961	0.904	0.950
100	0.817	0.824	0.960	0.897	0.950
200	0.849	0.856	0.959	0.897	0.950
500	0.880	0.887	0.949	0.902	0.950
1000	0.894	0.901	0.936	0.917	0.950
2000	0.912	0.919	0.925	0.929	0.950
5000	0.924	0.931	0.917	0.947	0.950

[a] Standard errors of $\gamma_{\widehat{\beta}_0}$ and $\gamma_{\widehat{\beta}_1}$ are nominally 0.001.

sample sizes ($n = 2000$ and 5000) reveal the "worrisome" behaviour of $\gamma_{\widehat{\beta}_1}$ for the OLSBC estimates: they do not saturate and approach the nominal value of 0.95 but seem rather to drift away for large n.

It becomes clear that for realistic settings (autocorrelation, incomplete prior knowledge), WLSXY estimation combined with pairwise-MBBres resampling is the only one-loop option to achieve acceptable levels of CI accuracy. A second loop of resampling (calibration or bootstrap-t) may in principle improve the accuracy, also for errors-in-variables regression (Booth and Hall 1993).

The combination of WLSXY and pairwise-MBBres performed well (Table 8.4) also for a rather difficult setting (stronger, unequal autocorrelations, lognormal shape). It is interesting to note that slope estimation yielded more accurate results than intercept estimation. The data size requirements, however, become rather strong (Table 8.4). Obtaining accurate results for data sizes in the range of 500 and below may require calibration methods.

8.3.3 Dependence on accuracy of prior knowledge

In practical situations, our prior knowledge about the measurement standard errors or their ratio, $\delta = S_Y/S_X$, may depend to a considerable degree on how good we know the measurement devices (calibration standards, replication analyses, etc.) or the archives "containing" the data (sampling error). The accuracy of that knowledge, parameterized here in form of Δ (Eqs. 8.11 and 8.12), should influence the estimation RMSE and possibly also the CI accuracy. This is explored by means of a set of simulation experiments (Tables 8.5 and 8.6), where Δ is varied.

The selection of the other setting parameters follows the previous Monte Carlo experiments in this section: intermediate sizes of autocorrelation, Gaussian shape and a true standard deviation ratio of $\delta = 2.0$. The data size may take relatively large values ($n = 2000$ and 5000) because also the limiting behaviour is of interest. We employ the WLSXY estimation and Student's t CI constructed by means of pairwise-MBBres resampling.

The resulting coverages (Table 8.5) approach with increasing data size the nominal value—as they should. In general, the levels are acceptable from n above, say, 200 (slope estimation) or 500 (intercept estimation). In the case of slope estimation, a highly inaccurate prior knowledge ($\Delta = 0.9$) may require more data points for achieving a coverage level similar to values found for smaller inaccuracies ($\Delta \leq 0.7$).

The resulting RMSE values (Table 8.6) for intercept estimation approach zero with increasing data size. The rate of this convergence seems not to depend on the accuracy of the prior knowledge (Δ). The RMSE

Table 8.5. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape: influence of accuracy of prior knowledge on CI coverage performance. Design is identical to the previous experiment (Table 8.4), with the following fixed setting: (1) autocorrelation parameters are $a_X = a_Y = 0.3$, (2) both noise components have normal shape. Estimation and CI construction is identical to the previous experiment (Table 8.4), with the following exceptions: (1) only WLSXY estimation with pairwise-MBBres resampling is considered; (2) prior knowledge of $S_X = 0.25, S_Y = 0.5$ ($\delta = 2.0$) is incomplete after Eq. (8.11) with various Δ values.

n	γ[a]					*Nominal*
	Accuracy of prior knowledge					
	$\Delta = 0.1$	$\Delta = 0.3$	$\Delta = 0.5$	$\Delta = 0.7$	$\Delta = 0.9$	
	Intercept estimation					
10	0.774	0.772	0.802	0.773	0.782	0.950
20	0.868	0.867	0.871	0.869	0.873	0.950
50	0.895	0.898	0.899	0.897	0.902	0.950
100	0.913	0.912	0.913	0.914	0.915	0.950
200	0.924	0.924	0.924	0.925	0.928	0.950
500	0.931	0.933	0.934	0.934	0.934	0.950
1000	0.936	0.936	0.939	0.939	0.941	0.950
2000	0.940	0.939	0.939	0.942	0.944	0.950
5000	0.945	0.945	0.944	0.944	0.946	0.950
	Slope estimation					
10	0.873	0.877	0.895	0.877	0.875	0.950
20	0.933	0.931	0.931	0.925	0.916	0.950
50	0.944	0.942	0.940	0.936	0.916	0.950
100	0.948	0.945	0.942	0.936	0.918	0.950
200	0.947	0.948	0.943	0.938	0.920	0.950
500	0.947	0.948	0.947	0.942	0.924	0.950
1000	0.948	0.947	0.952	0.945	0.925	0.950
2000	0.946	0.950	0.958	0.946	0.924	0.950
5000	0.950	0.963	0.960	0.945	0.923	0.950

[a] Standard errors of γ for intercept and slope estimations are nominally 0.001.

values for slope estimation show an interesting behaviour: they do not vanish with increasing data size but rather approach a finite value. The reason is that the inaccurate prior knowledge about the measurement standard errors (nonzero Δ) persists to influence the slope estimation—an error source independent of the data size. Similar behaviours were found also for OLSBC estimation of intercept and slope (results not shown). The saturation value of $\mathrm{RMSE}_{\widehat{\beta}_1}$ depends on the accuracy of the prior knowledge (Δ), seemingly in a close-to-linear relation.

Table 8.6. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape: influence of accuracy of prior knowledge on RMSE. The experiment is the same as described in Table 8.5.

n	RMSE				
	Accuracy of prior knowledge				
	$\Delta = 0.1$	$\Delta = 0.3$	$\Delta = 0.5$	$\Delta = 0.7$	$\Delta = 0.9$
	Intercept estimation[a]				
10	0.310	0.313	0.315	0.329	0.348
20	0.217	0.219	0.220	0.227	0.241
50	0.137	0.137	0.139	0.143	0.151
100	0.096	0.097	0.098	0.100	0.106
200	0.068	0.069	0.069	0.071	0.075
500	0.043	0.043	0.044	0.045	0.048
1000	0.031	0.031	0.031	0.032	0.034
2000	0.022	0.022	0.022	0.023	0.024
5000	0.014	0.014	0.014	0.014	0.015
	Slope estimation[b]				
10	0.279	0.300	0.428	0.303	0.339
20	0.173	0.177	0.181	0.191	0.223
50	0.105	0.107	0.114	0.126	0.166
100	0.074	0.077	0.085	0.099	0.146
200	0.052	0.056	0.066	0.084	0.135
500	0.033	0.040	0.052	0.073	0.127
1000	0.024	0.032	0.047	0.069	0.125
2000	0.018	0.028	0.043	0.067	0.123
5000	0.012	0.025	0.042	0.066	0.124

[a] Empirical $\mathrm{RMSE}_{\widehat{\beta}_0}$, given by $\left[\sum_{i=1}^{n_{\mathrm{sim}}} \left(\widehat{\beta}_0 - \beta_0\right)^2 / n_{\mathrm{sim}}\right]^{1/2}$.

[b] Empirical $\mathrm{RMSE}_{\widehat{\beta}_1}$, given by $\left[\sum_{i=1}^{n_{\mathrm{sim}}} \left(\widehat{\beta}_1 - \beta_1\right)^2 / n_{\mathrm{sim}}\right]^{1/2}$.

To summarize, measurement error in the predictor requires to modify the OLS method to yield a bias-free slope estimation: OLSBC or WLSXY. These modified estimation methods require prior knowledge about the size of the measurement error. If this knowledge is not exact, which is a typical situation in the climatological practice, then it contributes to the estimation error of the slope (RMSE and CI length). This contribution persists even when the data size goes to infinity.

8.3.4 Mis-specified prior knowledge

What happens if we make a wrong specification of the accuracy of our prior knowledge? We study (Table 8.7) a situation where (1) the

Table 8.7. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape: influence of mis-specified prior knowledge on CI coverage performance. Design and estimation (WLSXY) are identical to the previous experiment (Table 8.5). CI construction (via pairwise-MBBres resampling) is identical to that in the previous experiment, with the following exceptions: (1) prior knowledge of $S_X = 0.25, S_Y = 0.5$ ($\delta = 2.0$) is incomplete after Eq. (8.11) with $\Delta = 0.5$; (2) the adaptive Steps 8a and 8b of Algorithm 8.1 are allowed to mis-specify Δ.

n	$\gamma^{a}_{\widehat{\beta}_0}$			$\gamma^{a}_{\widehat{\beta}_1}$			*Nominal*
	True $\Delta = 0.5$ *Specified* Δ			*True* $\Delta = 0.5$ *Specified* Δ			
	0.3	0.5	0.7	0.3	0.5	0.7	
10	0.801	0.802	0.803	0.893	0.895	0.898	0.950
20	0.870	0.871	0.871	0.928	0.931	0.935	0.950
50	0.899	0.899	0.900	0.932	0.940	0.950	0.950
100	0.912	0.913	0.913	0.924	0.942	0.960	0.950
200	0.923	0.924	0.925	0.908	0.943	0.970	0.950
500	0.933	0.934	0.934	0.870	0.947	0.986	0.950
1000	0.939	0.939	0.940	0.827	0.952	0.994	0.950
2000	0.939	0.939	0.940	0.783	0.958	0.998	0.950
5000	0.944	0.944	0.944	0.744	0.960	1.000	0.950

[a] Standard errors of $\gamma_{\widehat{\beta}_0}$ and $\gamma_{\widehat{\beta}_1}$ are nominally 0.001.

true standard deviation ratio is $\delta = S_Y/S_X = 2.0$, (2) the estimation on the sample is done with an incomplete knowledge of δ, modelled as a uniform distribution over the interval between 1.0 and 3.0 ($\Delta = 0.5$), and (3) the bootstrap CI construction is allowed to mis-specify the incomplete knowledge by letting δ^* be uniformly distributed over the intervals between 1.4 and 2.6 (specified $\Delta = 0.3$) or between 0.6 and 3.4 (specified $\Delta = 0.7$). Specifying $\Delta = 0.3$ (instead of the correct $\Delta = 0.5$) constitutes a case of overestimation of the accuracy of the prior knowledge, $\Delta = 0.7$ means an underestimation and $\Delta = 0.5$ is an unbiased estimation.

The first result is that such a mis-specification has no effect on the accuracy of CIs for the intercept. Table 8.7 displays results (for $\Delta = 0.3$, 0.5 and 0.7) that are, within the bounds of the "simulation noise," indistinguishable.

The second result is that mis-specified prior knowledge has a clear effect on the accuracy of CIs for the slope. Table 8.7 shows results for $\Delta = 0.3$ and 0.7 to deviate from those for the correct value of $\Delta = 0.5$. If we underestimate the accuracy of the prior knowledge about the size of the measurement standard deviations ($\Delta = 0.3$ instead of 0.5), then the

CIs become too narrow and the coverage is reduced; if we overestimate the accuracy ($\Delta = 0.7$ instead of 0.5), then the CIs become too wide and the coverage is inflated.

8.4 Example: climate sensitivity

The effective climate sensitivity, denoted here as Λ_S^{-1}, is a parameter that relates changes in annual-mean surface temperature to changes in the radiative forcing (greenhouse gases, etc.) of the climate system. Its units are °C (or K) per Wm^{-2}. Climate sensitivity may vary with forcing history and climatic state, reflecting the influence of varying feedback mechanisms (amplifying or attenuating) in the climate system (Mitchell et al. 1987). The lack of an accurate knowledge of Λ_S^{-1} in the recent past (since, say, 1850) is one of the major obstacles for making accurate projections of future temperatures by means of AOGCMs (Forster et al. 2007).

The traditional estimation method for Λ_S^{-1} seems to be via perturbed climate models experiments, where the temperature response of the system is studied for a range of variations of model parameters and forcing scenarios (Forster et al. 2007). Due to the limited performance of climate models, it may be helpful to consider estimations that are based entirely on direct observations. We therefore relate variable $Y(i)$, the observed temperature changes from 1850 to 2001, to variable $X(i)$, the radiative forcing variations. The time series with standard errors are shown in Fig. 8.6. Since the predictor (forcing) has been determined with error, our model is the linear errors-in-variables regression (Eq. 8.1). The estimation objective is the slope, $\beta_1 = \Lambda_\mathrm{S}^{-1}$.

The result (Fig. 8.7) from WLSXY estimation of Λ_S^{-1} is $0.85\ \mathrm{K\,W^{-1}m^2}$. The 95% CI, a Student's t interval obtained from pairwise-MBBres resampling with $B = 2000$ (Algorithm 8.1), is $[0.47\ \mathrm{K\,W^{-1}m^2};\ 1.24\ \mathrm{K\,W^{-1}m^2}]$. The 90% CI, a level often used in the IPCC–WG I Report's chapter on radiative forcing (Forster et al. 2007), is $[0.53\ \mathrm{K\,W^{-1}m^2};\ 1.17\ \mathrm{K\,W^{-1}m^2}]$. The climate literature often uses the "equilibrium climate sensitivity," which is defined as the temperature change that would be approached in a (hypothetical) equilibrium following a doubling of the atmospheric "equivalent carbon dioxide concentration" (representing all greenhouse gases). This other sensitivity value is around $(4\ \mathrm{W^{-1}m^2})\Lambda_\mathrm{S}^{-1}$, at least in the climate world of the E-R AOGCM of the National Aeronautics and Space Administration Goddard Institute for Space Studies, New York (Foster et al. 2008). Thus, the WLSXY result suggests that a CO_2 doubling will lead to a temperature increase of 3.4 K.

What are the effects of autocorrelation? The block bootstrap resampling took into account the relatively strong memory of temperature and

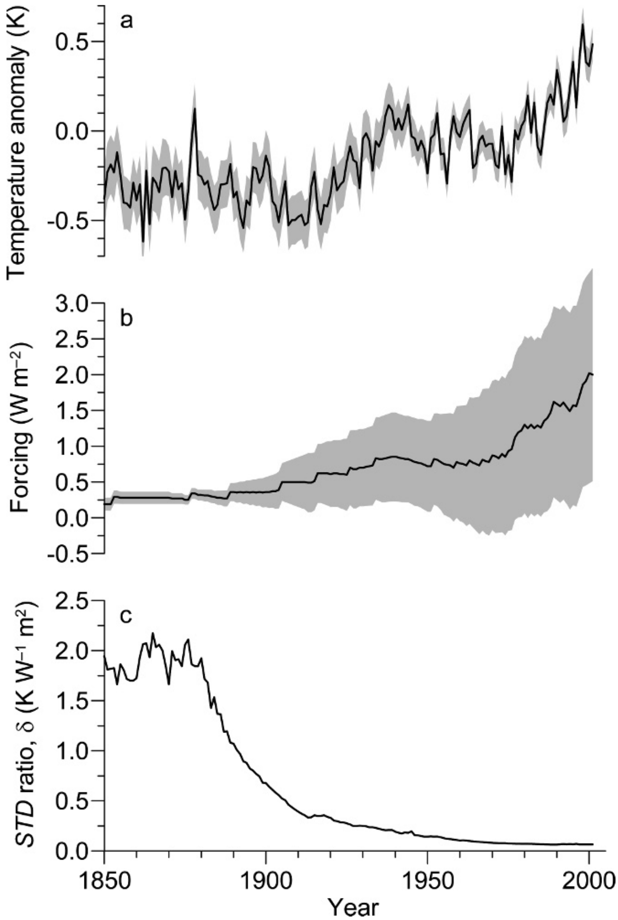


Figure 8.6. Northern hemisphere temperature anomalies and climate forcing, 1850–2001: data. **a** The temperature time series, $y(i)$, is shown (*solid line*) as deviation from the 1961–1990 average ($n = 152$). The annual-mean composite was derived using instrumental data from several thousand stations on land and sea (HadCRUT3 data set). The temperature standard error, $s_Y(i)$, is shown (*shaded band*) as $\pm 2 s_Y(i)$ interval around $y(i)$; it reflects following sources of uncertainty (Brohan et al. 2006): measurements, reporting, inhomogeneity correction, sampling, station coverage and bias correction of sea-surface temperatures. **b** The radiative forcing time series, $x(i)$, is shown (*solid line*) with $\pm 2 s_X(i)$ uncertainty band (*shaded*); it comprises following components thought to influence temperature changes (Hegerl et al. 2006; Forster et al. 2007): changes of atmospheric concentrations of greenhouse gases, solar activity variations (Fig. 2.12) and changes of sulfate and other aerosol constituents in the troposphere (lower part of the atmosphere). **c** Standard deviation ratio, $\delta = s_Y(i)/s_X(i)$. (Data from (**a**) Brohan et al. (2006) and (**b**) Hegerl et al. (2006).)

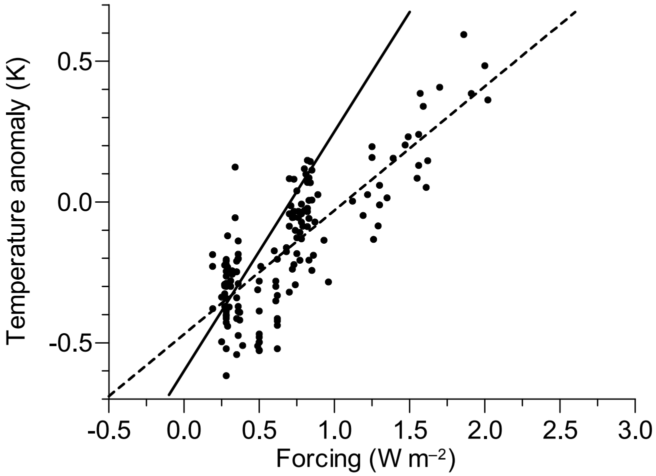


Figure 8.7. Northern hemisphere temperature anomalies and climate forcing, 1850–2001: fit. WLSXY estimation yields a straight regression line (*solid*) with a slope (i.e., effective climate sensitivity) of $\widehat{\beta}_1 = 0.85\ \mathrm{K\,W^{-1}m^2}$. Also shown is OLS regression line (*dashed*).

forcing noise components ($\widehat{a}'_X = \widehat{a}'_Y = 0.82$) by selecting a block length of $l = 18$. Ignoring autocorrelation (setting $l = 1$) would make the CI too narrow; for example, the 90% CI would become $[0.56\ \mathrm{K\,W^{-1}m^2}; 1.14\ \mathrm{K\,W^{-1}m^2}]$.

The estimate and, more, the CI for $\Lambda_{\mathrm{S}}^{-1}$ should be assessed, however, with caution.

- CI construction (Algorithm 8.1) used pairwise-MBBres resampling with an assumed constant standard deviation ratio of $\delta = 0.66$ (time-average). This was done because of the absence of Monte Carlo tests of adaptions of pairwise-MBBres resampling with respect to heteroscedastic errors. Instead we imposed an uncertainty of δ measured by the "incomplete prior knowledge" parameter Δ (Eq. 8.11). The employed value of $\Delta = 0.5$ may have been too small and produced a too narrow CI. Particularly, unrecognized temperature variations not caused by measurement error or forcing changes, that is, "internal temperature variability," may let δ increase and reduce the sensitivity estimate (Laepple T 2010, personal communication). Note that WLSXY estimation itself recognized heteroscedasticity.

- The HadCRUT3 temperature data (Brohan et al. 2006) are down-biased between about 1940 and the mid-1960s because of an unrecognized change in 1945 in the sea-surface measurement techniques

(Thompson et al. 2008). Since this interval is short relative to the total observation interval, the influence of the inhomogeneity on the Λ_S^{-1} estimate should be small.

- The tropospheric aerosol component of the forcing is known only with a "low" to "medium–low" scientific understanding (Forster et al. 2007). The aerosol contribution to $X(i)$ and $S_X(i)$ may be large in error. Consequently, the error in $S_X(i)$ and δ may be large, and the parameter Δ may be larger than 0.5 (or even another model of the incomplete prior knowledge required). We stress that the large error of the predictor necessitates fitting an errors-in-variables regression model. Ignoring this error (i.e., using OLS estimation) would strongly underestimate the climate sensitivity (Fig. 8.7).
- Volcanic eruptions, providing large negative forcing components (cooling) have been ignored in the estimation (because of the many unknowns), although the observed temperature time series (Fig. 8.6a) includes this effect. Since the number of large eruptions during the 151-year interval (Hegerl et al. 2006) is assessed as relatively small (about 8 eruptions with < -2.0 Wm^{-2} in the northern hemisphere), this omission should have a minor influence on the Λ_S^{-1} estimate.
- Ocean heat uptake has similarly been ignored, although observed temperatures may show this influence. Assuming that it cannot be neglected would (1) increase the Λ_S^{-1} estimate and (2) widen its CI.
- The analysis focused on the temperature of the northern hemisphere, while the concept of climate sensitivity applies to the globe. The superiority of temperature data quality for the northern part (more stations) suggested this restriction. Obviously, other geographic parts, including the globe, may be analysed in an analogous manner.

8.5 Prediction

A prediction is a statement about an uncertain event. In climate sciences the events lie often in the future (forecast) but frequently also in the past (hindcast), see the introductory examples (p. 3). In the context of the present chapter, we wish to predict an unobserved value, $y(n+1)$, given a sample, $\{t(i), x(i), y(i)\}_{i=1}^n$, and a new observation, $x(n+1)$, of the predictor variable made at time $t(n+1)$.

A typical situation is when a relation between a climate variable, $Y(i)$, and a proxy variable, $X(i)$, is to be established. Suppose we observed $\{t(i), x(i), y(i)\}_{i=1}^n$ over a time interval $[t(1); t(n)]$ but have available a longer proxy time series, $\{t(i), x(i)\}_{i=1}^m$ with $m > n$ (often $m \gg n$). If $t(i)$ denotes age and $t(i) > t(n)$ for $i > n$, then we wish "to hindcast"

$y(i)$ for $i > n$. An example is δ^{18}O as precipitation proxy; $y(i)$ is precipitation, $x(i)$ is δ^{18}O from a speleothem, $[t(1); t(n)] = [0\,\text{a};\ 50\,\text{a}]$ is the interval for which we have instrumental measurements of $y(i)$ (the past 50 years) and $[t(1); t(m)] = [0\,\text{a};\ 500\,\text{a}]$ is the interval covered by the speleothem samples (the past 500 years). If $t(i)$ denotes time, then we wish to forecast. An example is climate model projections; $y(i)$ is precipitation, $x(i)$ is modelled precipitation (AOGCM), $[t(1); t(n)] = [1950; 2010]$ is the interval for which we have instrumental measurements of $y(i)$ and $[t(1); t(m)] = [1950;\ 2100]$ is the interval analysed by means of the climate model (a typical value for the upper bound used by IPCC–WG I (Houghton et al. 2001; Solomon et al. 2007) in its reports).

Prediction can be performed by fitting a regression model and utilizing the estimated regression parameters. In the linear case (Fuller 1987: Section 1.6.3 therein):

$$\widehat{y}(n+1) = \widehat{\beta}_0 + \widehat{\beta}_1\, x(n+1), \tag{8.13}$$

where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ have been estimated using the sample $\{t(i), x(i), y(i)\}_{i=1}^{n}$ and the new observation is $x(n+1)$.

Which method is suitable for estimating $\widehat{\beta}_0$ and $\widehat{\beta}_1$?

Fuller (1987: pp. 75–76 therein) explains that usage of OLS, ignoring measurement errors of the predictor, is justified when $x(n+1)$ is drawn from the same distribution that generated $\{x(i)\}_{i=1}^{n}$. This means effectively that two conditions have to be met:

1. $S_X(n+1) \cdot X_{\text{noise}}(n+1)$ has the same properties (range, shape, etc.) as $S_X(i) \cdot X_{\text{noise}}(i), i = 1, \ldots, n$;

2. $X_{\text{true}}(n+1)$ has the same properties as $X_{\text{true}}(i), i = 1, \ldots, n$.

Fuller advises further to take measurement error into account when not both conditions are satisfied. This can be done, for example, by using WLSXY (or OLSBC) estimation. Treating the regression estimates as if they were known parameters, Fuller (1987: p. 76 therein) gives the following expression for the prediction standard error:

$$\widehat{\text{se}}_{\widehat{Y}(n+1)} = \left[S_Y(n+1)^2 + \widehat{\beta}_1^2\, S_X(n+1)^2 \right]^{1/2}. \tag{8.14}$$

We argue that in climatology the above conditions are almost exclusively not satisfied, and we advise to use WLSXY (or OLSBC) as a more conservative approach. In the majority of applications, $t(n+1)$, the time value related to the new measurement, is outside of $[t(1); t(n)]$, and $x(n+1)$ does not necessarily originate from a random drawing from the process $X_{\text{true}}(i), i = 1, \ldots, n$. The new measurement may rather constitute a step in a new direction of the course of climate, and it is safer to allow for that possibility by using WLSXY (or OLSBC).

However, the "machine error bar" (Eq. 8.14) may be too small because it does not include the estimation errors of the regression parameters. Therefore, it is advisable to use the bootstrap prediction error:

$$\widehat{\text{se}}_{\widehat{Y}(n+1)} = \left\{ \sum_{b=1}^{B} \left[\widehat{Y}^{*b}(n+1) - \left\langle \widehat{Y}^{*b}(n+1) \right\rangle \right]^2 \Big/ (B-1) \right\}^{1/2}, \quad (8.15)$$

where $\left\langle \widehat{Y}^{*b}(n+1) \right\rangle = \sum_{b=1}^{B} \widehat{Y}^{*b}(n+1)/B$ and the determination of $\widehat{Y}^{*b}(n+1)$ is explained (sample level) within Algorithm 8.1.

Another source of prediction error, difficult to quantify, stems from the extrapolation. This regards (1) the standard deviations, $S_X(n+1)$ and $S_Y(n+1)$, under heteroscedasticity but also (2) the possibility that with $x(n+1)$ outside of the observation interval, from $\min(x(i))$ to $\max(x(i))$, or with $t(n+1)$ outside of $[t(1); t(n)]$, new laws set in and, if unrecognized, may bias the prediction. A physical theory behind the regression model may guard against such errors (background material).

8.5.1 Example: calibration of a proxy variable

Calibrating a proxy variable, $X(i)$, means quantifying the relation with a climate variable, $Y(i)$, by means of regression. Since $X(i)$ is usually observed with measurement error, the errors-in-variables equation (8.1) has to be considered. The fitted regression curve serves for predicting an uncertain value, $y(n+1)$, given a new proxy measurement, $x(n+1)$. Calibration is ubiquitous in quantitative paleoclimatology. Examples: oxygen isotopic composition in a marine sediment core is a proxy for temperature (Fig. 1.2), hydrogen isotopes in an ice core indicate temperature (Fig. 1.3a). Here we look at δ^{18}O in a coral as a proxy for air temperature.

We make two further remarks. First, the calibration methodology applies also to predicting future climate values by means of climate models. Second, the core of the interest lies usually in relative variations, changes of a variable—the slope (which itself is susceptible to estimation bias).

Draschba et al. (2000) calibrated δ^{18}O, measured in a coral taken from a site off the coast of Bermuda, against observations of air-temperature on that island (Fig. 8.8). The calibration curve, established for the time interval from 1856 to 1920, was then used to make a hindcast of temperature for the interval from 1350 to 1630 (by using measurements from another coral located close to the site of the "calibration coral").

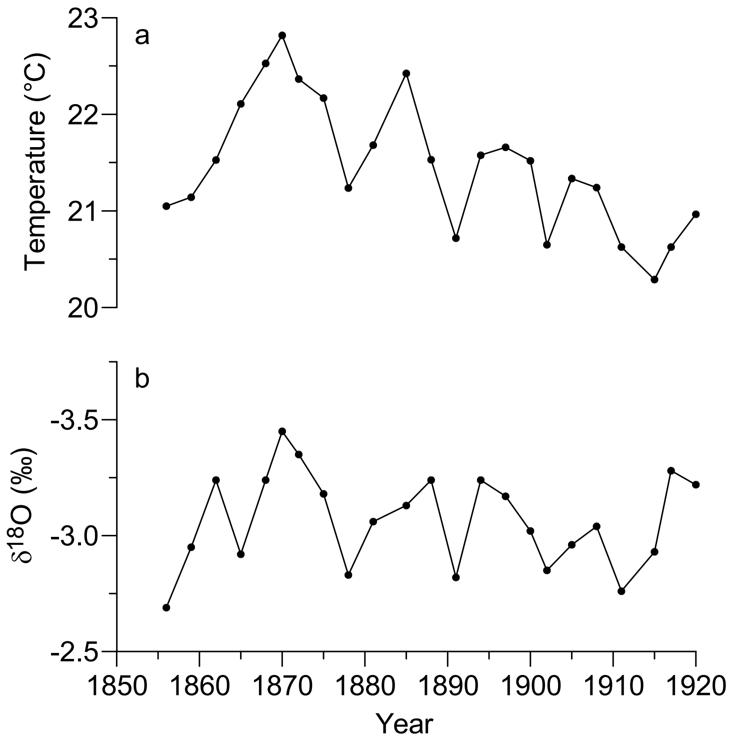


Figure 8.8. Bermuda air temperature and coral δ^{18}O, 1856–1920: data. **a** The annual-mean temperature time series, $y(i)$, is shown only at those time points for which δ^{18}O values (**b**) are available ($n = 23$). The temperature standard error, $s_Y(i)$, is assumed to be constant and equal to 0.03°C (Table 1.3). **b** The δ^{18}O time series, $x(i)$, is unevenly spaced ($d(i) = 2$ or 3 a). The δ^{18}O values have a constant measurement error of $s_X = 0.07$‰ (Draschba et al. 2000). (The temperature data are digitized values from Draschba et al. (2000: Fig. 2c therein), the δ^{18}O data were downloaded from `http://doi.pangaea.de/10.1594/PANGAEA.88200` (17 September 2009).)

The accuracy of the δ^{18}O timescale, crucial for a successful calibration, is excellent owing to the presence of seasonal density banding (visible on X-ray photographs). Measurement procedures and errors (Fig. 8.8) are described in detail by Draschba et al. (2000). Sample material requirements led to an unevenly spaced δ^{18}O time series, with $D'(i) = D(i) = d(i) = 2$ or 3 a (see Fig. 1.13 for definitions). Draschba et al. (2000) transformed the temperature record (monthly observations) by binning to either an annual resolution or a 3-year resolution. Their

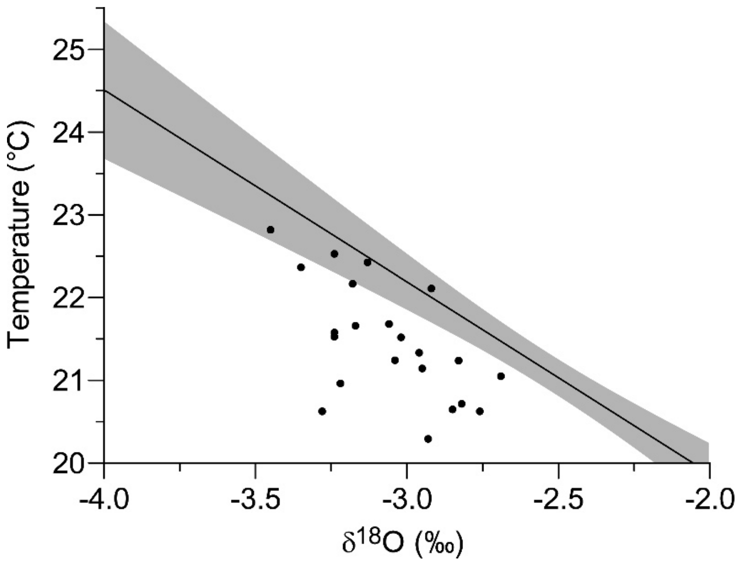


Figure 8.9. Bermuda air temperature and coral δ^{18}O, 1856–1920: prediction. OLSBC estimation yields a straight prediction line (*solid*) with an intercept of $\widehat{\beta}_0 = 15.2$°C and a slope of $\widehat{\beta}_1 = -2.3$°C‰$^{-1}$. Also shown is 95% Student's t confidence band (*shaded*), obtained from bootstrap resampling (pairwise-MBBres with $\widehat{\tau}'_X = \widehat{\tau}'_Y = 6.9\,\mathrm{a}$, $l = 6, B = 2000$).

calibration result did not strongly depend on that choice. Here we use the annual values from those years for which also δ^{18}O values exist (Fig. 8.8).

The calibration curve (Fig. 8.9) has a slope that is in size larger by a factor of approximately 1.3 than that estimated by Draschba et al. (2000). This considerable deviation is likely the result of an ignored bias correction in the original paper. The bias-corrected OLSBC fit curve deviates considerably from a na-ive per-eye fit through the points (Fig. 8.9). (Interestingly, the authors considered already their slope estimate as rather large in absolute size.) The pointwise bootstrap confidence band allows to quantify the prediction uncertainty, also outside of the original range of observations (Fig. 8.9).

The bootstrap prediction error (Eq. 8.15), averaged over the interval of x values shown in Fig. 8.9, is equal to 0.25°C, while the "machine error bar" (Eq. 8.14) is 0.16°C.

Two further remarks ought to be made. First, the confidence band assumes a time-independent calibration relation and homoscedastic errors. This assumption may be violated. Second, the confidence band may be inaccurate owing to the limited data size, as the Monte Carlo experiments (Section 8.3) show.

8.6 Lagged regression

Let us reconsider the linear errors-in-variables model (Eq. 8.1) in continuous time, T. Assume for convenience homoscedasticity. Introduce a time lag parameter, H, to shift the predictor variable, such that

$$Y(T) = \beta_0 + \beta_1 \left[X(T+H) - S_X \cdot X_{\text{noise}}(T+H)\right] + S_Y \cdot Y_{\text{noise}}(T). \quad (8.16)$$

A lag $H > 0$ ($H < 0$) means that variations of "true" Y lead over (lag behind) variations of "true" X. This is a lagged errors-in-variables regression model.

Measured time series are discrete in time and finite in size. Assume for convenience even time spacing ($d(i) = d = \text{const.}$) and introduce a dimensionless time lag, $h = H/d$, such that

$$Y(i) = \beta_0 + \beta_1 \left[X(i+h) - S_X \cdot X_{\text{noise}}(i+h)\right] + S_Y \cdot Y_{\text{noise}}(i), \quad (8.17)$$

$i = 1, \ldots, n-h$. Given a bivariate sample, $\{x(i), y(i)\}_{i=1}^{n}$, the task is to estimate β_0, β_1 and h.

WLSXY estimation should in principle be possible by minimizing a normalized sum,

$$SSQWXY(\beta_0, \beta_1, h) = (n-h)^{-1} \sum_{i=1}^{n-h} \frac{\left[y(i) - \beta_0 - \beta_1 x(i+h)\right]^2}{S_Y^2 + \beta_1^2\, S_X^2}. \quad (8.18)$$

This may be achieved technically by numerical minimization (Section 8.8) of $SSQWXY(\beta_0, \beta_1, \widetilde{h})$ for a fixed (candidate) lag, $\widetilde{h}$, and a brute-force search over a range of $\widetilde{h}$ values. Intuitively, if $1 \ll h \ll n$, then the error due to the discretization of the time should be smaller than when h is close to either bound.

A more general, realistic situation arises when the two time series were observed at mutually unequal times. This has been explored in the context of correlation estimation (Section 7.5), where the time gaps could be bridged owing to the presence of persistence. The situation becomes even more realistic (difficult), when timescale errors are introduced. We analyse such an example (Section 8.6.1), where we resort to interpolation. The taken approach is somewhat ad-hoc. The theoretical knowledge about estimators and their properties, let alone CI construction, is rather limited for such situations, and the given literature (background material) does not cover this issue exhaustively.

8.6.1 Example: CO_2 and temperature variations in the Pleistocene

One of the major contributions of ice cores as climate archives is information about CO_2 variations far back in time (late Pleistocene). The Vostok core's record, first drilled and measured over the past 160 ka (Barnola et al. 1987), was later extended to the full span of 420 ka (Petit et al. 1999). The longest CO_2 record currently available (past 800 ka) comes from the EPICA Dome C ice core (Siegenthaler et al. 2005; Lüthi et al. 2008). The major finding from those ice core studies was that not only temperature and ice volume underwent large changes during the ice age (100-ka cycle), but also the atmospheric CO_2 concentration. We explore here the full Vostok span of changes of CO_2 and temperature (inferred via δD), shown in Fig. 1.3, to estimate the phase relations between these changes. Such relations constitute a basis for erecting a causal climatological theory of the late Pleistocene ice age—which does not yet exist in sufficient detail. We follow the paper by Mudelsee (2001b), who used lagged regression as a tool for phase relationship estimation.

Mudelsee (2001b) deviated in some technical points from the errors-in-variables methodology developed in the previous sections. These and some additional points are discussed first, the results shown thereafter.

First, the time values of the predictor variable, $\{t_X(i)\}_{i=1}^{3311}$, are not identical to those of the response variable, $\{t_Y(j)\}_{j=1}^{283}$. Allowing for a candidate lag, $\widetilde{H}$, requires a time shift. For those reasons, the lag estimation used linear interpolation of the x values (Fig. 8.10), $t_X = t_Y(j) - \widetilde{H}$. The fact that the lag is imposed for computational reasons on $t_Y(j)$ rather than $t_X(i)$ (Eq. 8.16), is not relevant for the estimation.

Second, the lagged regression employed a parabolic model. This performed slightly better than the linear one, as evaluated by means of the reduced sum of squares (fourth point). (A logarithmic model would yield similar values as the parabolic (Fig. 8.12).)

Third, the predictor's error has an upper limit of $s_X = 1‰$ (Petit et al. 1999), which is clearly smaller than the standard deviation (spread) of the $x(i)$ values of 17‰. This means only a small estimation bias when ignoring measurement error (Mudelsee 2001b).

Fourth, the estimation (Mudelsee 2001b) used GLS (Section 4.1.2) with the $\mathbf{V}$ matrix elements given by $\exp[-|t_Y(j_1) - t_Y(j_2)|/\tau_Y]$. Because the persistence time, τ_Y, was unknown, a second brute-force loop for τ_Y was nested and the overall minimum taken as solution. The resulting reduced least-squares sum is

$$SSQG_\nu(\boldsymbol{\beta}, H, \tau_Y) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \, \mathbf{V}^{-1} \, (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \, / \nu, \tag{8.19}$$

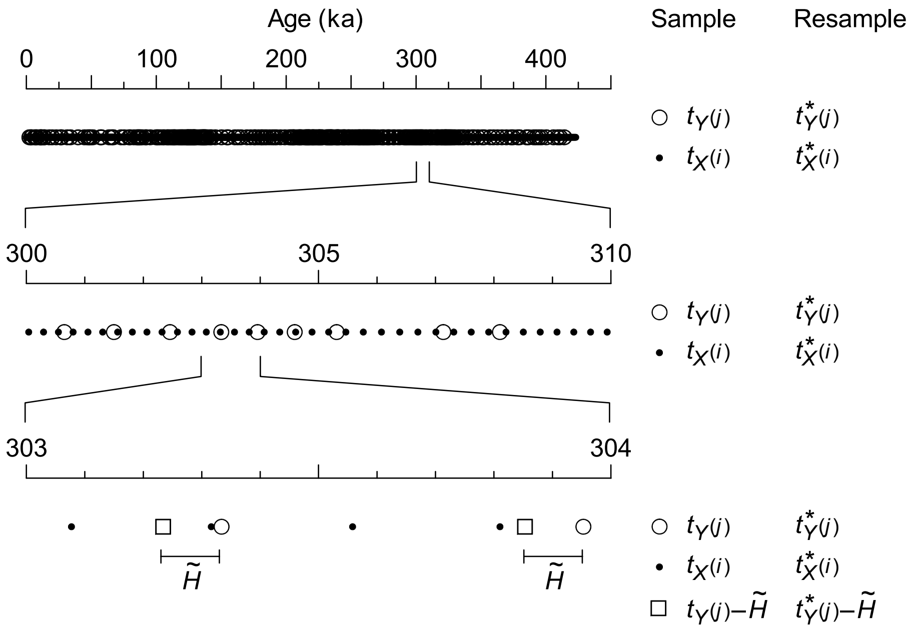


Figure 8.10. Vostok deuterium and CO_2, timescales for lag estimation. The time interval [303 ka; 304 ka] illustrates the relation between the predictor (X) variable, deuterium, and the response (Y) variable, lagged CO_2. The candidate lag in time is $\widetilde{H}$. The predictor values are obtained by linear interpolation of the x and x^* values, $t_X = t_Y(j) - \widetilde{H}$ and $t_X^* = t_Y^*(j) - \widetilde{H}$, respectively. (Original data shown in Fig. 1.3.)

where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \text{ (parameter vector)}, \tag{8.20}$$

$$\mathbf{y} = \begin{bmatrix} y(1) \\ \vdots \\ y(n-h) \end{bmatrix} \text{ (response vector)}, \tag{8.21}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x'(1) & x'(1)^2 \\ \vdots & \vdots & \vdots \\ 1 & x'(n-h) & x'(n-h)^2 \end{bmatrix} \text{ (predictor matrix)}, \tag{8.22}$$

$\nu = n - h - 3$ (degrees of freedom) and x' is interpolated x (Fig. 8.10). The linear model has no β_2 parameter and $\nu = n - h - 2$. The step size of the brute-force search for $\widehat{H}$ was 5 a.

The resulting lag estimate is $\widehat{H} = -1.3$ ka, that is, a lag of CO_2 variations behind temperature variations. The resulting persistence time is $\widehat{\tau}_Y = 0.92$ ka. The reduced least-squares sum in dependence on $\widetilde{H}$ is shown in Fig. 8.11. The resulting parabolic fit is shown in Fig. 8.12.

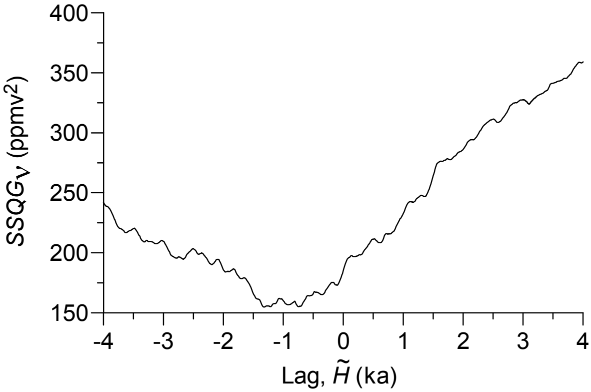


Figure 8.11. Vostok deuterium and CO_2, reduced sum of squares. The minimum (i.e., lag estimate) is at $\widetilde{H} = -1.3$ ka. (After Mudelsee 2001b.)

Both predictor and response (x, y) exhibit measurement and proxy errors, and both timescales (t_X, t_Y) show dating uncertainties. These four error sources propagate into the estimation standard error of the lag, $\widehat{\mathrm{se}}_{\widehat{H}}$. Mudelsee (2001b) determined $\widehat{\mathrm{se}}_{\widehat{H}}$ by means of a parametric surrogate data approach (Algorithm 8.2).

The first error source (x) was simulated (Mudelsee 2001b) as

$$x^*(i) = x(i) + x_{\mathrm{noise}}(i), \tag{8.23}$$

where $x_{\mathrm{noise}}(i)$ is a realization of a Gaussian AR(1) process with standard deviation $s_X = 1.0‰$ (Petit et al. 1999) and persistence time $\tau_X = 2.1$ ka (Chapter 2).

The second error source (y) was simulated analogously as

$$y^*(i) = y(i) + y_{\mathrm{noise}}(i), \tag{8.24}$$

where the noise process had a standard deviation $s_Y = 2.5$ ppmv (Petit et al. 1999) and persistence time $\tau_Y = \widehat{\tau}_Y = 0.92$ ka.

The third error source (t_X) was simulated (Mudelsee 2001b) using the depth points of the ice core samples (Petit et al. 1999) and a non-parametric fit of the "sedimentation rate" (Fig. 4.13). The simulated

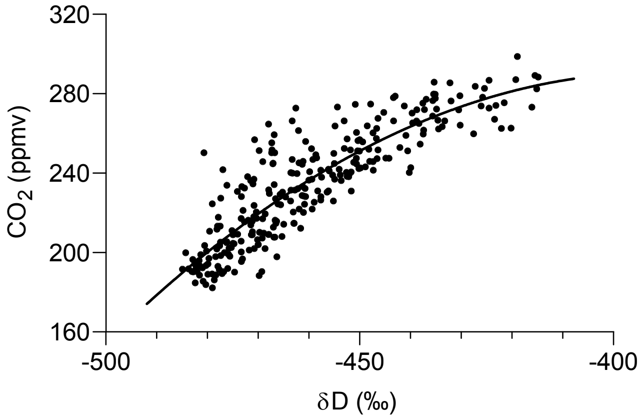


Figure 8.12. Vostok deuterium and CO_2, parabolic fit. Data points are lagged CO_2 ($\widehat{H} = -1.3$ ka) against interpolated δD ($n - h = 280$). The fit line is given by $y = -1482 - 9.05x - 0.012x^2$. (After Mudelsee 2001b.)

Step 1	Time series	$\{t_X(i), x(i)\}_{i=1}^{n_X}$,
		$\{t_Y(j), y(j)\}_{j=1}^{n_Y}$
Step 2	Lag estimate	$\widehat{H}$
	via minimization of $SSQG_\nu(\boldsymbol{\beta}, H, \tau_Y)$	
Step 3	Simulated time series;	$\left\{t_X^{*b}(i), x^{*b}(i)\right\}_{i=1}^{n_X}$,
	b, counter	$\left\{t_Y^{*b}(j), y^{*b}(j)\right\}_{j=1}^{n_Y}$
Step 4	Replication	$\widehat{H}^{*b}$
Step 5	Go to Step 3 until $b = B$ (usually $B = 2000$)	
	replications exist	$\left\{\widehat{H}^{*b}\right\}_{b=1}^{B}$
Step 6	Calculate standard error and CIs	

Algorithm 8.2. Determination of bootstrap standard error and construction of CIs for lag estimate in lagged regression, surrogate data approach (Sections 3.3.3 and 3.4). The algorithm is applicable also to other estimation techniques than $SSQG_\nu$ minimization (Step 2).

sedimentation rate was obtained parametrically (Mudelsee 2001b) by imposing a relative, Gaussian error of 1.2%. The simulated sedimen-

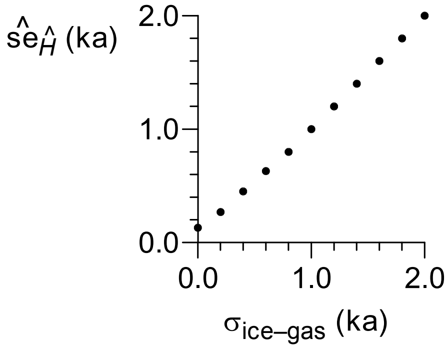


Figure 8.13. Vostok deuterium and CO_2, sensitivity study of lag estimation error.

tation rate, combined with the depth points, resulted in a simulated timescale (Section 4.1.7). In a final step, the simulated timescale was randomly compressed or expanded to fit into the GT4 timescale error range (Petit et al. 1999), which is ≤ 5 ka for the last 110 ka, ≤ 10 ka for "most of the record" (interpreted as 110–300 ka by Mudelsee (2001b)) and ≤ 15 ka in the early part.

The fourth error source (t_Y) was simulated on basis of the simulated ice-ages (t_X^*). The additional error contribution comes from the uncertainty in the ice–gas age difference,

$$t_Y^* = t_X^* + \mathcal{E}_{\mathrm{N}(0,\, \sigma^2_{\text{ice–gas}})}(\cdot). \tag{8.25}$$

Petit et al. (1999: p. 434 therein) reported $\sigma_{\text{ice–gas}}$ to be 1 ka or more.

The surrogate data approach yielded (Mudelsee 2001b) $\widehat{\text{se}}_{\widehat{H}} = 1.0$ ka. To restate, the lag estimation result is that temperature variations occurred 1.3 ± 1.0 ka before CO_2 variations.

The crucial point for achieving such a small estimation error is that x and y were measured on the same core (Vostok). This means a rather close coupling of t_X^* and t_Y^* (Eq. 8.25). Only the uncertainty in the ice–gas age difference weakens the coupling. Had CO_2 been measured on a core from a different site, no coupling would exist and t_X^* and t_Y^* had to be simulated independently of each other, leading to a clearly larger lag estimation error than in the present case.

The lag estimation result underlines the importance of the uncertainty, $\sigma_{\text{ice–gas}}$, in the ice–gas age difference, which contributes nearly 100% to the lag estimation error of 1.0 ka. A sensitivity study (Fig. 8.13) quantifies this contribution over a range of prescribed $\sigma_{\text{ice–gas}}$ values. For example, in the case of $\sigma_{\text{ice–gas}} = 0.2$ ka the lag estimation error would be $\widehat{\text{se}}_{\widehat{H}} = 0.27$ ka. In the case of a perfectly known difference

($\sigma_{\text{ice–gas}}$ equal to zero), the remaining error sources would propagate into $\widehat{\text{se}}_{\widehat{H}} = 0.13$ ka.

As regards causal explanations of the late Pleistocene glacial cycles, Mudelsee (2001b) noted that Vostok's air temperature (δD) represents, at best, the southern hemisphere and that there exists a time lag of the variations relative to the northern hemisphere (Blunier et al. 1998). However, the complexity of the ice-age climate may be better understood, that is, the set of feasible causal scenarios (Broecker and Henderson 1998: Table 1 therein) further constrained, with the help of quantified phase relations.

8.7 Background material

OLSBC estimation of the slope has also been denoted as attenuation-corrected OLS (ACOLS) estimation (Ammann et al. 2009).

The **method of moments** estimator of the standard deviation of the predictor in the case of homoscedasticity, S_X, is (Fuller 1987: Eq. (1.3.10) therein):

$$\widehat{S}_X = (2\delta)^{-1} \left\{ m_{YY} + \delta m_{XX} - \left[(m_{YY} - \delta m_{XX})^2 + 4\delta m_{XY}^2 \right]^{1/2} \right\}, \tag{8.26}$$

where

$$\delta = \lambda^{1/2} = S_Y / S_X, \tag{8.27}$$

the moments are

$$m_{YY} = \sum_{i=1}^{n} \left[y(i) - \bar{y} \right]^2 / (n-1) , \tag{8.28}$$

$$m_{XX} = \sum_{i=1}^{n} \left[x(i) - \bar{x} \right]^2 / (n-1) , \tag{8.29}$$

$$m_{XY} = \sum_{i=1}^{n} \left[x(i) - \bar{x} \right] \left[y(i) - \bar{y} \right] / (n-1) \tag{8.30}$$

and the sample means are

$$\bar{y} = \sum_{i=1}^{n} y(i) / n \tag{8.31}$$

and

$$\bar{x} = \sum_{i=1}^{n} x(i)\,/n\,. \tag{8.32}$$

$\widehat{S}_X$ is plugged in for S_X (Eq. 8.6).

WLSXY estimation of a linear relationship between two variables that are both subject to error has been studied, and the geometric interpretation been made, already before and at the beginning of the twentieth century (Adcock 1877, 1878; Pearson 1901); see also Wald (1940) and Fuller (1987, 1999). The method to fit a hyperplane to data with errors in all their coordinates (possibly more than two) is also denoted as total least squares (Nievergelt 1998).

Non-Gaussian, heteroscedastic noise components in the linear errors-in-variables regression model can be taken into account in the estimation using GLS, that is, using the covariance matrix, analogously to Section 4.1.2. In practical applications to climatological problems, where the covariance matrix is unknown and has to be estimated, an iterative procedure may be used. Fuller (1987: Section 3.1 therein) describes GLS estimation for serially independent noise components and gives a result (standard errors of parameters) that is valid for large data sizes. He advises to consider developing a model for the error structure if the data size is small. However, it is not clear whether such a classical approach to parameter error determination can be applied also to serially dependent noise components.

Correlated noise components in the linear errors-in-variables regression model can be taken into account. York (1969) adapts a least-squares criterion to recognize correlation between $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ and gives an example from radiometric dating. Freedman (1984) and Freedman and Peters (1984) present two-stage regression with bootstrap resampling as a method to treat a correlation between $X_{\text{noise}}(i)$ and $Y(i)$. Fuller (1987: Section 3.4 therein) presents a transformation for dealing with correlation between $X_{\text{noise}}(i)$ and $X(i)$.

Multiplicative measurement error may occur in form of $X(i) = X_{\text{true}}(i) \cdot X'_{\text{noise}}(i)$, where the primed noise component is dimensionless. Carroll et al. (2006: Section 4.5 therein) mention transformation methods that may be applied in this case.

Nonlinear errors-in-variables models can be estimated on basis of several assumptions about the model and the noise properties, by using numerical techniques for solving the maximum likelihood or least-squares optimizations (Fuller 1987: Section 3.3 therein). A recent book (Carroll et al. 2006) gives more details.

The **pairwise-MBBres** algorithm from Section 8.2 is a response to resolving the "quite nonstandard" (Hall and Ma 2007: p. 2621 therein) situation, where neither the true predictor variable, $X_{\text{true}}(i)$, nor the errors, $X_{\text{noise}}(i)$, "can be directly accessed." Previously, Efron and Tibshirani (1993: Section 9.5 therein) and Davison and Hinkley (1997: Section 6.2.4 therein) considered that pairwise bootstrap resampling is applicable to errors-in-variables regression problems. Linder and Babu (1994) presented another alternative to the simple pairwise resampling. These authors scaled the residuals in both dimensions (X, Y) and resampled independently from both sets. They analysed maximum likelihood estimation with known standard deviation ratio and tested the accuracy of bootstrap CIs (percentile and Student's t) by means of Monte Carlo experiments, finding acceptable levels of accuracy. This was confirmed in an additional simulation study of slope estimation (Musekiwa 2005) with small data sizes ($n = 20, 30$). It should be interesting to investigate further the approach of Linder and Babu (1994), adapted to the climatologically more realistic situation where the standard deviation ratio is not exactly known and the errors exhibit serial dependence.

The **approaching of finite RMSE values** or, equivalently, the absence of shrinking CIs with $n \to \infty$ was verified for slope estimation and falsified for intercept estimation (Section 8.3.3). Previously, Booth and Hall (1993) found a non-shrinking bootstrap confidence band in a Monte Carlo experiment on prediction (Section 8.5), $\widehat{y}(n+1) = \widehat{\beta}_0 + \widehat{\beta}_1\, x(n+1)$. Thus, it appears that this observation (Booth and Hall 1993) has its origin in the non-shrinking of $\text{RMSE}_{\widehat{\beta}_1}$.

The **simulation–extrapolation** algorithm (Carroll et al. 2006: Chapter 5 therein) is a bias correction method based on Monte Carlo simulations. The idea is to add artificial measurement error to the data and study the dependence of an estimate (say, $\widehat{\beta}_1$) in dependence of the size of the artificial error. Extrapolation to zero size should, so the idea, provide an unbiased estimate.

Prediction and forecasting by means of regression and other models is reviewed by Ledolter (1986). The success of prediction depends, of course, on the validness of the regression model and the absence of disturbing "latent" variables (Box 1966). A physical theory behind the model is a guard against wrong conclusions based on such disturbances. For example, radiation physics and meteorology support the concept of climate sensitivity (estimated by means of a regression of changes in temperature on changes in radiative forcing, see Section 8.4) and refute claims that time acts as a latent variable. On the other hand, a regression of annual temperature on the annual output of scientific papers on global warming over the past, say, 150 years, would find a strong relation (highly significant slope)—however, a spurious relation because the latent variable time is acting and no theory exists that supports the model.

Lagged regression as presented in Section 8.6 (that means, with one single lag parameter, h), is a special case of rational distributed lag models (Koyck 1954; Dhrymes 1981; Doran 1983; Pankratz 1991), where

$$Y(i) = \beta_0 + \beta_{1,0} X(i) + \beta_{1,1} X(i+1) + \beta_{1,2} X(i+2) + \cdots + S_Y \cdot Y_{\text{noise}}(i). \tag{8.33}$$

The sequence $\{\beta_{1,0}, \beta_{1,1}, \beta_{1,2}, \ldots\}$ is called impulse response function. Note that the equation ignores errors in the predictor. Fitting such models may be performed using maximum likelihood (Dhrymes 1981) or frequency-domain techniques (Hannan and Robinson 1973; Hannan and Thomson 1974; Hamon and Hannan 1974; Foutz 1980). This technique is frequently applied in econometrics. One of the rare applications to climatology is the work by Tol and de Vos (1993), who estimated a lagged regression of annual-mean temperature, 1951–1979, on atmospheric CO_2 concentration. Insofar climate is a dispersive system, where the response of one variable on the input of another is frequency-dependent, it should be worth developing further such models and fitting techniques that take into account typical properties of paleoclimatic series (measurement errors, unequal times and uncertain timescales).

The **effective climate sensitivity** is usually denoted as $\lambda_{\mathrm{S}}^{-1}$. Various estimation approaches have been carried out, Table 8.8 gives a short overview. The approach via the heat capacity (Schwartz 2007) opened an interesting exchange of arguments in the Journal of Geophysical Research. Let C denote the effective heat capacity (change in heat per change in temperature) per unit area that is coupled to the relevant timescale of a perturbation (i.e., years to decades). The perturbation regards the radiative balance of the Earth (change in forcing). Schwartz (2007) estimated C (with standard error) to be 17 ± 7 W a m^{-2}K^{-1}. The C value reflects mainly the upper part of the ocean, which can take up heat on the discussed timescale of (anthropogenically enhanced) radiative perturbations. The simple equation,

$$\tau = C \cdot \Lambda_{\mathrm{S}}^{-1}, \tag{8.34}$$

describes the time span (relaxation, τ) over which a radiative perturbation (C) has an effect on the temperature ($\Lambda_{\mathrm{S}}^{-1}$). Schwartz (2007) estimated τ by fitting an AR(1) model (Chapter 2) to observational data. The criticism on this approach (Foster et al. 2008; Knutti et al.

2008; Scafetta 2008) was centred on the AR(1) model as over-simplistic and estimation bias. In his reply, Schwartz (2008) kept the AR(1) model but conceded τ to be larger (8.5 ± 2.5 a) than in his original contribution (5 ± 1 a). The revised estimate for τ leads to the entry in Table 8.8.

Table 8.8. Estimates of the effective climate sensitivity.

Λ_S^{-1} *Estimate*[a] (K W^{-1}m^2)	*Approach*	*Reference*
0.29 [0.05; 0.53][b,c]	Direct observations, 2000–2006	Chylek et al. (2007)
0.48 [0.24; 0.72][b,d]	Direct observations, 2000–2006	Chylek et al. (2007)
0.51 [−0.01; 1.03][b]	Physical principles (heat capacity)	Schwartz (2008)
0.65 [0.09; 1.21][b]	Thermodynamical model	Scafetta and West (2007)
0.70 [0.38; 1.55][e]	Climate model and observations, 1000–2001	Hegerl et al. (2006)
0.85 [0.53; 1.17][c]	Direct observations, 1850–2001	This book
1.53 [0.40; ∞][e,f]	Direct observations, 1861–1900 and 1957–1994	Gregory et al. (2002)

[a] With 90% CI.
[b] CI calculated as ±2 standard error interval.
[c] Ignoring ocean heat uptake.
[d] Assuming strong ocean heat uptake.
[e] Calculated from originally estimated equilibrium climate sensitivity.
[f] Median given instead of estimate.

The **leads and lags of carbon dioxide** variations relative to those of temperature in the Pleistocene have been studied by several researchers on time series from ice cores from Antarctica. Previously to Mudelsee (2001b), who estimated $\widehat{H} = -1.3 \pm 1.0$ ka (a lag of CO_2), the original authors of the 0–420 ka Vostok paper (Petit et al. 1999) had found, seemingly by per-eye inspection, that CO_2 decreases lag behind temperature decreases by several ka. Cuffey and Vimeux (2001: p. 523 therein) believed that the lag, "especially during the onset of the last glaciation, about 120 ka ago," is largely an "artefact caused by variations of climate in the water vapour source regions." They presented model simulations that correct for this effect and lead to $\widehat{H} \approx 0$ ka using the Vostok data, 0–150 ka and 150–350 ka. Subsequently, Monnin et al. (2001) analysed high-resolution records ($\bar{d} \approx 0.18$ ka for CO_2) from the EPICA Dome C site over the interval 9–22 ka by means of an explorative technique ("correlation maximization") similar to the brute-force search (Section 8.6.1). They obtained an estimate of $\widehat{H} = -0.41$ ka, which was assessed as not significant owing to the uncertainty of the ice–gas age difference. Caillon et al. (2003) revisited the Vostok ice core, inspected the time interval around Termination III (230–255 ka) and took argon isotopes instead of deuterium as proxy for temperature changes. The motivation for performing the new measurements was the idea that the poorer proxy quality of argon isotopes would be more than compensated by the smaller timescale uncertainties. Since argon is, as CO_2, contained in the enclosed air bubbles in the ice, no uncertainty of the ice–gas age difference influences lag estimation (Eq. 8.25). The result of correlation maximization (Caillon et al. 2003) was a lag of CO_2, $\widehat{H} = -0.8 \pm 0.2$ ka, where the error bar value is based on an assessment of the uncertainty of the ice accumulation (but not on an additional consideration of the proxy errors). The "EPICA challenge" (Wolff et al. 2005), issued to paleoclimatologists, was to predict CO_2 concentration for the interval 420–740 ka on basis of the then available EPICA Dome C deuterium (temperature) and dust records (EPICA community members 2004). The simple model, lagged regression of Vostok CO_2 on EPICA deuterium (temperature), calibrated on the 0–420 ka records, did not produce the worst of the eight predictions, as was found when the EPICA Dome C CO_2 series became known. The complete interval back to 800 ka from the EPICA ice core archive of past changes in temperature (Jouzel et al. 2007) and CO_2 (Siegenthaler et al. 2005; Lüthi et al. 2008) enables to analyse also temporal changes of the lag, H, concepts that the ice-age climate relationships changed for a while after a glacial termination (Raynaud et al. 1993). To summarize, the overall lag of CO_2 variations behind temperature variations during the late Pleistocene appears significant. This is a quantitative basis for developing and testing concepts about causes and effects of long-term climate changes (Broecker and Henderson 1998; Saltzman 2002), about how the external astronomical forcing (Milankovitch cycles) propagates into the geographic regions, and how the climate variables respond. Further refining the ice-age theory is currently an active research field (Kawamura et al. 2007; Huybers and Denton 2008; Wolff et al. 2009). It is important to note that the characteristic timescales on which the analysed Pleistocene climate changes occurred, are relatively long: the average spacing ($\bar{d}$), the estimated lag ($\widehat{H}$) and its estimation error ($\widehat{\mathrm{se}}_{\widehat{H}}$) are between several hundred and a few thousand years. The late Pleistocene lag estimates are therefore hardly relevant as regards concepts about the ongoing climate change, which is anthropogenically enhanced since, say, 150 years. This recent change is considerably faster than the late Pleistocene change, it leads to CO_2 levels not experienced during at least the past 800 ka and it affects other physical processes. The consideration from physics and meteorology that the recent change has a positive time lag (CO_2 rise before temperature rise) is not contradicted by the finding that the late Pleistocene change had a negative time lag. The scientifically interesting question is whether humans are able to put a (temporary) end to the succession of glacials and interglacials (Berger and Loutre 2002).

Errors-in-variables regression models for climatology have not often been formulated in an explicit manner in the research literature. Allen and Stott (2003) showed theoretically the importance of an unbiased slope estimation for linear models that relate temperature changes predicted by an AOGCM with observed temperature changes. Hegerl et al. (2007a) studied in that manner proxy calibration to reconstruct 30–90°N mean annual land-surface temperature for the past 1500 years. Kwon et al. (2002) fitted a nonlinear model to dating samples,

$$Y(i) = \frac{\exp\left(\lambda_{^{40}\mathrm{K}} \cdot \tau_{\mathrm{FCs}}\right) - 1}{\exp\left\{\lambda_{^{40}\mathrm{K}}\left[X(i) - S_X \cdot X_{\mathrm{noise}}(i)\right]\right\} - 1} + S_Y \cdot Y_{\mathrm{noise}}(i), \quad (8.35)$$

$i = 1, \ldots, n$. They used five paired samples of $Y(i) = {}^{40}\mathrm{Ar}/{}^{39}\mathrm{Ar}$ ratio and $X(i)$ = reference age, observed with small, homoscedastic (S_X, S_Y), mutually independent measurement errors of assumed Gaussian shape. The estimation parameters were $\lambda_{^{40}\mathrm{K}}$ (decay constant of $^{40}\mathrm{K}$) and τ_{FCs} (age of Fish Canyon sanidine age standard). Kwon et al. (2002) derived maximum likelihood estimators and analysed bootstrap standard errors based on the surrogate data approach. Their Monte Carlo study showed that the estimates do not strongly depend on the Gaussian assumption. The result, $\widehat{\lambda}_{^{40}\mathrm{K}} = 5.4755 \pm 0.0170 \cdot 10^{-10}\ \mathrm{a}^{-1}$, leads to a half-life estimate (Section 1.6) of $\widehat{T}_{1/2} = \ln(2)/\widehat{\lambda}_{^{40}\mathrm{K}} = 1.266 \pm 0.004$ Ga. Bloomfield et al. (1996) made a multivariate nonlinear regression of daily tropospheric ozone concentration in the Chicago metropolitan area, 1981–1991, on a variety of predictors, including temperature, wind speed and relative humidity. The interesting point in the context of this chapter is that also lagged predictors (H prescribed as 1 or 2 days) were included. Bloomfield et al. (1996) used GLS estimation (Gallant 1987: Sections 2.1 and 2.2 therein) and obtained parameter standard errors by means of jackknife resampling (Section 3.8), which takes serial dependence into account.

8.8 Technical issues

WLSXY minimization of $SSQWXY(\beta_0, \beta_1)$ is numerically difficult because the slope, β_1, appears in the denominator of the least-squares sum (Eq. 8.8). The routine Fitexy (Press et al. 1992) parameterizes the slope as $\beta_1' = \tan^{-1}(\beta_1)$, scales the y values and uses Brent's search (Section 4.5) with a starting value for the slope from an initial OLS estimation. (A more recent Numerical Recipes edition is Press et al.

(2007).) Papers on the way from the work of Deming (1943) and York (1966) to the routine Fitexy include Reed (1989, 1992) and Squire (1990). This book follows those authors in usage of WLSXY for estimation, but it does not so for parameter error determination; for that purpose it uses instead bootstrap resampling. Extensions of WLSXY to nonlinear regression problems (nonlinear in the parameters) were considered by Jefferys (1980, 1981) and Lybanon (1984). A review of least-squares fitting when both variables are subject to error (Macdonald and Thompson 1992) studied besides WLSXY also other weighting techniques. It appears that a generalized least-squares estimation method for the case of (1) serial correlations in both X and Y errors and (2) correlation between X and Y errors, supported by a proof of optimality (in terms of, say, RMSE) under the Gaussian distributional assumption, has not yet been developed.

LEIV1 is another Fortran implementation of WLSXY estimation (York 1966), available at `http://lib.stat.edu/multi/leiv1` (26 October 2009).

LEIV3 is a Fortran software for maximum likelihood fitting of linear errors-in-variables models with heteroscedastic noise components (Ripley and Thompson 1987), available at `http://lib.stat.edu/multi/leiv3` (26 October 2009).

Bootstrap software for errors-in-variables regression is not abundant. Carroll et al. (2006) and Hardin et al. (2003) mention Stata software, available from the site `http://www.stat.tamu.edu/~carroll` (26 October 2009). Software for block bootstrap resampling seems to be unavailable—limiting the ability to study errors-in-variables regression with autocorrelated noise components.

Part IV

Outlook

Chapter 9

Future Directions

What changes may bring the future to climate time series analysis? First we outline (Sections 9.1, 9.2 and 9.3) more short-term objectives of "normal science" (Kuhn 1970), extensions of previous material (Chapters 1, 2, 3, 4, 5, 6, 7 and 8). Then we take a chance (Sections 9.4 and 9.5) and look on paradigm changes in climate data analysis that may be effected by virtue of strongly increased computing power (and storage capacity). Whether this technological achievement comes in the form of grid computing (Allen 1999; Allen et al. 2000; Stainforth et al. 2007) or quantum computing (Nielsen and Chuang 2000; DiCarlo et al. 2009; Lanyon et al. 2009)—the assumption here is the availability of machines that are faster by a factor of ten to the power of, say, twelve, by a mid-term period of, say, less than a few decades.

9.1 Timescale modelling

Climate time series consist not only of measured values of a climate variable, but also of observed time values. Often the latter are not evenly spaced and also influenced by dating uncertainties. Conventional time series analysis largely ignored uneven and uncertain timescales, climate time series analysis has to take them into account.

The process that generated the times, $\{t_X(i)\}$ for univariate and also $\{t_Y(j)\}$ for bivariate series, depends on the climate archive. We have studied linear and piecewise linear processes for speleothem or sedimentary archives (Section 4.1.7) and nonparametric models for ice cores (Section 8.6.1). Such types of models are the basis for including uncertain timescales in the error determination by means of bootstrap resampling ($\{t_X^*(i)\}$ and also $\{t_Y^*(j)\}$). In bivariate and higher dimensional estimation problems, also the joint distributions of the timescale processes are

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_9,

important. See the example of the Vostok ice core (Section 8.6.1) with the coupled timescales for the ice and the gas.

Climate archive modelling should be enhanced in the future to provide accurate descriptions of uncertain timescales. Archive models should evidently include the physics of the accumulation of the archive. One may even think of physiological models describing the performance of humans in layer counting of regular sequences such as varves (Table 1.3). A second ingredient of climate archive modelling are statistical constraints, for example, a strictly monotonically increasing age–depth curve in a speleothem archive or an absolutely dated fixpoint in a marine sediment core. An exemplary paper (Parrenin et al. 2007) of climate archive modelling studies the accumulation and flow in an ice sheet, into which a core is drilled. The Bayesian approach may be suitable for combining the inputs from physics and statistical constraints (Buck and Millard 2004).

9.2 Novel estimation problems

Chapters 2, 3, 4, 5 and 6 presented stochastic processes and estimation algorithms for inferring the fundamental properties of univariate climate processes in the climate equation (Eq. 1.2): trend, variability, persistence, spectrum and extremes. Chapters 7 and 8 studied bivariate processes: correlation and the regression relation between two univariate processes. We believe to have covered with these chapters the vast majority of application fields for the climate sciences.

However, in science there is always room for asking more questions, that means in a quantitative approach, for attempting to estimate different climate parameters in the uni- or bivariate setting.

An obvious example of such a novel estimation problem is SSA, mentioned in the background material of Chapter 1. This decomposition method has been formulated so far only for evenly spaced, discrete time series. Interpolation to equidistance is obsolete because it biases the objectives of the decomposition (estimates of trend, variability, etc.). SSA formulations applicable to unevenly spaced records should therefore be developed.

Other novel estimation approaches are expected to come from the array of nonlinear dynamical systems theory (Section 1.6). This field has a focus more on application data from controlled measurements or computer experiments and less on unevenly spaced, short paleoclimatic time series. A breakthrough, also with respect to SSA, may come from techniques of reconstructing the phase space at irregular points.

9.3 Higher dimensions

Climate is a complex, high-dimensional system, comprising many variables. Therefore it makes sense to study not only univariate processes (Part II), X, or bivariate processes (Part III), X and Y, but also trivariate processes, X and Y and Z, and so forth. A simple estimation problem for such high-dimensional processes is the multivariate regression, mentioned occasionally in previous chapters (Sections 4.2 and 8.7),

$$Y(i) = \theta_0 + \theta_1 X(i) + \theta_2 Z(i) + \cdots + S_Y(i) \cdot Y_{\text{noise}}(i). \qquad (9.1)$$

The higher number of dimensions may also result from describing the climate evolution in the spatial domain (e.g., X is temperature in the northern, Y in the southern hemisphere). There is a variety of high-dimensional, spatial estimation problems: multivariate regression, PCA and many more (von Storch and Zwiers 1999: Part V therein).

As regards the bootstrap method, there is no principle obstacle to perform resampling in higher dimensions. An important point is that resampling the marginal distributions, of X and Y and Z separately, is not sufficient; the joint distribution of (X, Y, Z), including dependences among variables, has to be resampled to preserve the original covariance structure. This requires adaptions of the block bootstrap (MBB) approach. A further point, which may considerably exacerbate the estimation as well as the bootstrap implementation, is unequal observation times. The sets

$$\{t_X(i)\}_{i=1}^{n_X}, \ \{t_Y(j)\}_{j=1}^{n_Y}, \ \{t_Z(k)\}_{k=1}^{n_Z} \qquad (9.2)$$

need not be identical. Depending on the estimation problem and the properties of the joint climate data generating process (e.g., persistence times), the algorithm for determining $\theta_0, \theta_1, \theta_2$, and so forth, has to be adapted. This is a step into new territory. An example from the bivariate setting is the "synchrony correlation coefficient" (Section 7.5.2). A final point of complication from the move into higher dimensions is dependence among the timescale variables. Since this type of complication can occur already in two-dimensional problems (Section 8.6.1), we expect it in higher dimensions as well. This challenge must be met by means of timescale modelling (Section 9.1).

9.4 Climate models

Computer models render the climate system in the form of mathematical equations. The currently most sophisticated types, AOGCMs (Fig. 1.9), require the most powerful computers. Nevertheless, the rendered spatial and temporal scales are bounded by finite resolutions and finite domain sizes. Also the number of simulated climate processes is limited.

The problem of a finite spatial resolution is currently tackled by means of using an AOGCM (grid size several tens to a few hundred kilometres) for the global domain and nesting into it a regional model or RM (grid size reduced by a factor ~ 20) for a sub-domain of interest (say, Europe). The AOGCM "forces" the RM (Meehl et al. 2007; Christensen et al. 2007), that means, prescribes the conditions at the boundaries of the sub-domain. Sub-grid processes, not resolved even by the RM (e.g., cloud processes) and therefore not explicitly renderable by the AOGCM–RM combination, can be implicitly included by employing inferred parametric relations (e.g., between cloud formation and temperature). The AOGCM–RM combination includes many variables, $(X, Y, Z, \ldots)' \equiv \mathbf{X}$, from the climate at grid points, and many parameters, $(\theta_0, \theta_1, \theta_2, \ldots)' \equiv \boldsymbol{\theta}$, from the parameterizations (Stensrud 2007) and other model equations. For convenience of presentation, we consider the climate variable vector, $\mathbf{X}$, and the climate model parameter vector, $\boldsymbol{\theta}$.

Our premise of a future "quantum boost" by a factor $\sim 10^{12}$ can make regionalization dispensable and let more realistic AOGCMs (grid size several tens to a few hundred metres) become calculable with computing times reduced from, say, a year to less than a month. Regarding the sophistication of a climate model, the increased computing power can also be utilized for including processes from the fields of biology and economy (greenhouse gas emissions (Moss et al. 2010) and "climate engineering" measures). Indeed, a finer spatial grid does require more processes to be explicitly included. Regarding the temporal scale, the boost should allow to simulate much larger spans (transient paleoclimate runs) by the means of AOGCMs and their successors.

There exists, however, another field where to invest computing power, namely the uncertainty determination of climate model results. We sketch this area in light of the methodology presented in this book, statistical estimation and bootstrap resampling.

Physics describes climate dynamics by means of nonlinear coupled differential equations,

$$\dot{\mathbf{X}} = f\left(\mathbf{X}, \mathbf{R}, \boldsymbol{\theta}\right), \tag{9.3}$$

where the dot denotes time derivative, f is a function, and $\mathbf{R}$ represents uncoupled, external forcing variables (e.g., solar activity). Time discretization yields

$$\mathbf{X}(i+1) = \mathbf{X}(i) + \Delta T \cdot \dot{\mathbf{X}}, \tag{9.4}$$

where ΔT is a time step, in an AOGCM typically in the order of minutes to hours. From an initial climate state, $\mathbf{X}(1)$, the climate evolution is derived. This sample from the climate model "archive" is

$$\{\mathbf{x}(i)\}_{i=1}^{n}\,. \tag{9.5}$$

The climate evolution can also be observed, yielding a multivariate time series sample,

$$\{\mathbf{x}_{\mathrm{o}}(i)\}_{i=1}^{n}\,. \tag{9.6}$$

The observations are, of course, strongly limited in the number of climate variables, geographic locations and time resolutions. There have been few observations made of, say, temperature in 1000 m height above sea-level at 130°W, 30°S for the time interval from 1850 to 2010 and a spacing of $d(i) = \Delta T = 30$ minutes.

9.4.1 Fitting climate models to observations

Let us view climate modelling as an estimation problem. The task is to estimate the model parameters, $\boldsymbol{\theta}$, given observations, $\{\mathbf{x}_{\mathrm{o}}(i)\}_{i=1}^{n}$. This set shall include the "missing observations." The task requires to run the model and produce $\{\mathbf{x}(i)\}_{i=1}^{n}$. The less distant the model output is to the observations, the better the fit.

Let us introduce a cost function to measure the distance,

$$SSQGXYZ_{\nu}(\boldsymbol{\theta}) = g\big(\,\{\mathbf{x}_{\mathrm{o}}(i)\}_{i=1}^{n}\,, \{\mathbf{x}(i)\}_{i=1}^{n}\,\big). \tag{9.7}$$

g may be a form of a generalized least-squares cost function that takes into account predictor uncertainty and the degrees of freedom; Section 9.4.3 considers the design of g in more detail. The parameter estimate minimizes the cost function,

$$\widehat{\boldsymbol{\theta}} = \mathrm{argmin}\Big\{g\big(\,\{\mathbf{x}_{\mathrm{o}}(i)\}_{i=1}^{n}\,, \{\mathbf{x}(i)\}_{i=1}^{n}\,\big)\Big\}. \tag{9.8}$$

The parameter vector is included in the right-hand side of the equation because the model output, $\{\mathbf{x}(i)\}_{i=1}^{n}$, depends on it.

The outlined procedure is with current computing power not feasible for a full estimation of AOGCM parameters. It has been performed for a simple climate model containing only three variables (Hargreaves and Annan 2002) and an Earth system model of intermediate complexity (Paul and Schäfer-Neth 2005). The concept of fitting climate models to data is also denoted as data assimilation or state estimation (Wunsch 2006).

Subsequent to the estimation, we should like to know the parameter uncertainties for the fitted climate model. This knowledge may be achieved by means of bootstrap methods, producing the replications,

$$\widehat{\boldsymbol{\theta}}^{*} = \mathrm{argmin}\Big\{g\big(\,\{\mathbf{x}_{\mathrm{o}}^{*}(i)\}_{i=1}^{n}\,, \{\mathbf{x}^{*}(i)\}_{i=1}^{n}\,\big)\Big\}. \tag{9.9}$$

The observation resample, $\mathbf{x}_{\mathrm{o}}^{*}(i)$, can be obtained via the surrogate data bootstrap (Section 3.3.3), taking into account the errors of the observation devices, the distributional shapes (which may be Gaussian or not), the covariances (which may be rather small) and the "internal climate variability" (which may have to be estimated by means of separate model experiments). The model output resample, $\mathbf{x}^{*}(i)$, incorporates a new (trial) set of parameters, $\boldsymbol{\theta}^{*}$. However, it should also be based on a random initial state, $\mathbf{x}^{*}(1)$, because the initial conditions are not exactly known. $\mathbf{x}^{*}(1)$ may be taken randomly from a set of time series values, $\{\mathbf{x}_{\mathrm{unforced}}^{*}(j)\}_{j=1}^{m}$, of a climate model run without forcing components (stationarity). This "ensemble technique" is already currently being applied to quantify the uncertainty component owing to imperfectly known initial conditions (Randall et al. 2007; van der Linden and Mitchell 2009). Also the forcing variable, $\mathbf{R}(i)$, may have to be described stochastically for being included in the surrogate data approach.

The replications, $\{\widehat{\boldsymbol{\theta}}^{*b}\}_{b=1}^{B}$, serve in the usual manner (Section 3.4) for constructing CIs. Of particular interest should be the joint PDF of the climate model parameter estimators, which may be described by means of confidence regions in the parameter hyperspace (Smith et al. 2009; Tebaldi and Sansó 2009). Realistic climate model error and CI determination do not require a handful of runs (current ensemble technique) but rather B runs, with B in the usual order of 2000 or even higher (because of the dimensionality).

9.4.2 Forecasting with climate models

Models are employed to forecast future climate, $\mathbf{x}(n+1)$, at time $t(n+1)$. (Indeed, forecasts are made for many time steps to cover the typical range from the present to the year 2100.) This is achieved in our vision by a run of the model employing the estimated, optimal parameters, $\widehat{\boldsymbol{\theta}}$. That run has to use also a guess of the future forcing, $\mathbf{R}(n+1)$.

Of crucial importance, scientifically and socioeconomically, is to determine the size of the forecasting error. The bootstrap methodology, utilized for that purpose in the bivariate setting (Section 8.5), should be helpful also in the high-dimensional setting.

The recommendation is to produce forecast resamples, $\mathbf{x}^{*}(n+1)$, from which to calculate standard errors, CIs, confidence bands (over a time span), and so forth.

How are the $\mathbf{x}^{*}(n+1)$ produced to reflect the full range of the various sources of uncertainty?

- The parameterization uncertainty can be taken into account by resampling from the set of replications, $\{\widehat{\boldsymbol{\theta}}^{*b}\}_{b=1}^{B}$. This preserves the covariance structure of the parameter estimates.
- The initial-condition uncertainty can be taken into account by means of the ensemble technique.
- The forcing uncertainty may be difficult to include in a quantitative manner. This step does likely necessitate the usage of separate forcing models.

9.4.3 Design of the cost function

Designing the cost function (Eq. 9.7) is important for achieving small standard errors and narrow CIs for the climate forecasts and the model parameter estimates. It is rather difficult to demonstrate theoretically the optimality of a certain cost function. One should perform Monte Carlo simulations to find "empirically" a suitable function. The following points may guide the design endeavour.

- A least-squares technique is mandatory. It seems impossible to write down a likelihood function (for maximization) owing to the size of the body of the climate model equations. One may wish to make the sum of squares more robust with respect to "outliers." On the other hand, one may give the "outliers" instead *more* weight in situations where the focus is on modelling the climate extremes.
- GLS, employing the covariance matrices (variability, persistence) of the many climate variables, is a possible technique to reduce the estimation standard errors. The normalization (variability) produces dimensionless $SSQG$ terms for each variable, which can be processed further (e.g., summed up).
- A problem is multicollinearity (correlated predictors), stemming from spatial dependence among the climate variables (neighboured grid points). This may indicate to reduce the number of variables in the cost function by means of spatial binning. PCA techniques should help evaluating geographically meaningful bins (regions).
- Errors in the observations $(S_X, S_Y, S_Z, \ldots)$ should lead researchers to consider techniques like WLSXY estimation (Section 8.1.2) to reduce estimation bias.
- Further weighting could be performed "in the time domain" to enforce, for example, the most recent years to be more accurately simulated.

- The degrees of freedom, ν, of the observation–model combination can be taken into account (a simple division by ν).
- One may put bounds to the $\boldsymbol{\theta}$ hyperspace to exclude estimation results that are inconsistent with physics (hard bounds) or prior knowledge (soft bounds). Bayesian formulas may help here.

The envisaged availability of "quantum computing power" does not release us from the task of constructing efficient methods to search through the hyperspace, to locate the minimum of the cost function: gradient techniques, Brent's search, hybrid procedures or Bayesian approaches (Monte Carlo Markov Chain, see Hargreaves and Annan (2002) and Leith and Chandler (2010)).

9.4.4 Climate model bias

Climate model bias regards, generally speaking, a function of the climate variable vector,

$$\eta = h\left(\mathbf{X}\right). \tag{9.10}$$

The function, h, can be used to make η an index variable or extract a geographic region. For example, we may wish to study time-dependent, annual-mean, regional-mean, land-surface precipitation in central Europe,

$$\eta(j) = n_k^{-1}\, n_i^{-1} \sum_{k\, \in\, \text{region}} \;\; \sum_{T(i)\, \in\, \text{year } j} X_k(i), \tag{9.11}$$

where $X_k(i)$ is precipitation at grid point k and time $T(i)$, n_i is the number of time values within year j and n_k is the number of model grid points within central Europe.

Let us now view the *modelled* sequence as an estimate obtained by means of a climate model, $\widehat{\eta}(j)$. Next we consider the true sequence. Since the truth is hidden, we take instead an observed sequence, $\eta_{\mathrm{o}}(j)$. This leads, in analogy to Eq. (3.2), to the climate model bias,

$$\mathrm{bias}_{\widehat{\eta}}(j) = E\left[\widehat{\eta}(j)\right] - \eta_{\mathrm{o}}(j). \tag{9.12}$$

In the example of precipitation in central Europe, there are indications from a range of AOGCM–RM combinations that $\mathrm{bias}_{\widehat{\eta}}(j) > 0$ for the time interval from 1950 to the very recent past (Jacob D 2009, personal communication), that is, the climate models systematically overestimate precipitation. Similar overestimations were found for the region of Scandinavia (Goodess et al. 2009).

In the context of climate forecasting (Section 9.4.2), better predictions may therefore include a climate model bias correction. For example, if the model bias is simply a constant, $\text{bias}_{\widehat{\eta}}$, then

$$\eta'(j_{\text{future}}) = \eta(j_{\text{future}}) - \text{bias}_{\widehat{\eta}}, \tag{9.13}$$

where j_{future} indicates future (unobserved) time and the prime denotes bias correction. Evidently, the time-dependence of the bias and also its form (additive, multiplicative) should be analysed in such situations. Further developments may employ more complex stochastic models of the climate model bias (Jun et al. 2008).

9.5 Optimal estimation

Increased computer power would also allow to perform optimal estimation. We have sketched this concept in previous parts of this book (Sections 6.2.7 and 7.5.3.1). Not only climatology, other science branches as well may benefit from optimal estimation.

Central to the investigation in natural sciences, such as climatology, is to infer the truth from the data. This calls for the statistical language. In quantitative climatology, the investigative questions can be translated into a parameter, θ, which needs to be estimated using the data. The investigation cycles through loops: question, estimation, refined question based on the estimation result, new estimation, and so forth.

An estimator, $\widehat{\theta}$, is a recipe how to guess θ using the data. Since the sample size is less than infinity and the sampled climate system contains unknown influences (noise), we cannot expect that $\widehat{\theta}$ equals θ. However, we can calculate the size of that error, the uncertainty. This leads to the measures $\text{se}_{\widehat{\theta}}$, $\text{bias}_{\widehat{\theta}}$, $\text{RMSE}_{\widehat{\theta}}$ and the confidence interval, $\text{CI}_{\widehat{\theta},1-2\alpha}$, which is thought to include θ with probability $1 - 2\alpha$. Without having the information contained in such measures, it is difficult to assess how close $\widehat{\theta}$ is to θ: estimates without error bars are useless.

For simple estimation problems (e.g., mean estimation) and simple noise properties (e.g., Gaussian distributional shape), the error measures can be analytically derived via the PDF of an estimator. However, climate is more complex—as regards the noise as well as the estimation problem. This book advocates therefore the bootstrap resampling approach, which allows to analyse complex problems for realistic (i.e., complex) properties such as non-Gaussian shape or serial dependence.

For the most part of this book, we have assumed the uncertainty to have its origin in the complex climate system and the measured variables (proxy, measurement and dating errors). We have occasionally considered (Sections 4.1.7.4, 4.4 and 8.3.4) another error source, a misspecified model. Statistical science refers to this error source as model uncertainty; see Chatfield (1995), Draper (1995), Candolo et al. (2003) and Chatfield (2004: Section 13.5 therein). By fitting a range of candidate models it is possible to infer the range of feasible estimation outcomes. For example, one may compare the estimated 100-year return level, HQ_{100}, from a Weibull fit with the estimated HQ_{100} from a GEV fit to observed runoff data, and look whether the difference of the results is comparable to the statistical standard errors. Note that model uncertainty may regard also the assumed noise model (e.g., short versus long memory). A method to reduce model uncertainty is to employ graphical and computational tests of model suitability. As a method to quantify model uncertainty, we may study not only the range of the estimation outcomes but impose a weighting according to the probability that a particular model is correct. The "model probability" may be based in a Bayesian approach on a prior consultation of experts (Smith et al. 2009). In the example of HQ_{100}, there is hope that the hydrologists would put more weight on the GEV model than on the Weibull. It is principally possible to add model uncertainty as a new dimension to the hyperspace of climate estimation (Fig. 9.1).

Climate is a paradigm of a complex system that requires for its analysis the bootstrap. In addition, climate opens the new problem dimensions of unequally spaced series and timescale errors. This book has presented various bootstrap algorithms to adapt closely to the estimation problem imposed by the data: ARB, MBB, SB, surrogate data, timescale-ARB, timescale-MBB, pairwise-ARB, pairwise-MBB and pairwise-MBBres. It also described algorithms to support bootstrap resampling and CI construction: block length selection, calibration, the CI types normal, Student's t, percentile and BCa.

The critical question is: What is the best method for inferring the truth from the data? What is the optimal estimation method, and how are the most accurate CIs constructed?

Future, strongly increased computing power allows to approach that question by means of Monte Carlo experiments. We outline this optimal estimation approach (Fig. 9.1). We reiterate that optimal estimation is not limited to the field of climate sciences.

The hyperspace of climate estimation has many, but not infinite dimensions. It consists of the three subspaces Monte Carlo design, method and measure.

The Monte Carlo design (Fig. 9.1) describes the data generating process. The design is used to generate artificial data, to which the method is applied. The design should, in some sense, cover the estimation problem (data and estimation) to be carried out. One group of dimensions is occupied by the type of estimation model and the parameters. For example, one may be interested in a linear regression model with the two parameters intercept and slope (Chapters 4 and 8). To restate, the

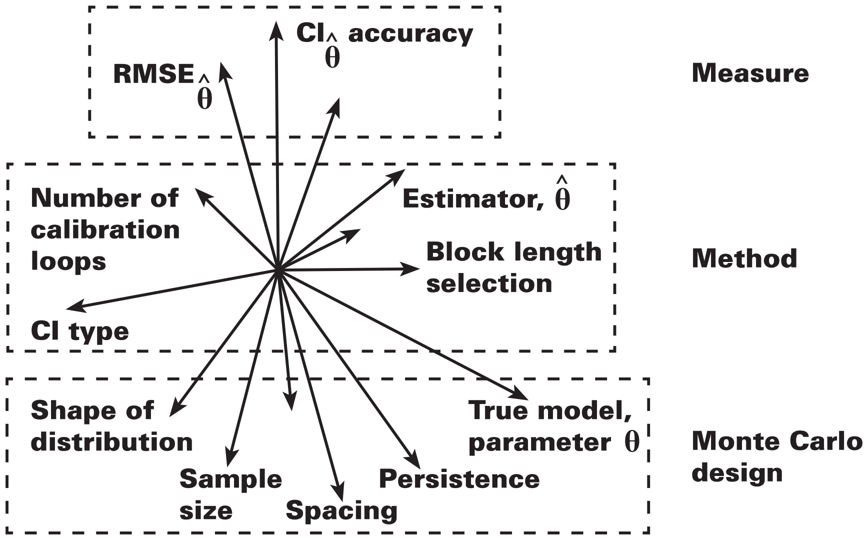


Figure 9.1. Hyperspace of climate parameter estimation. The Monte Carlo experiment prescribes the stochastic model, parameters and other properties (shape, sample size, spacing, persistence, etc.) in a way that the problem at hand (data and estimation) is covered. The method regards estimation and CI construction. The optimal estimation is determined by using a measure.

Monte Carlo parameters (e.g., prescribed intercept and slope) should be close to the estimated parameters (estimated intercept and slope). The other group of dimensions in the Monte Carlo subspace describe the sample size (prescribed n, which should be close to the size of the sample at hand), the spacing (again, similar to the spacing of the sample) and the noise properties (also similar). An option is to invest three dimensions to model the persistence of the noise as an ARFIMA(p, δ, q) process (which contains the simpler types such as AR(1)) and one or two to model the shape (skewness, kurtosis). Heteroscedasticity may also be modelled. The ARFIMA process contains the preferred parsimonious, embedding-problem free AR(1) process ($p = 1, \delta = 0, q = 0$). Some dimensions have integer values (e.g., the ARFIMA parameter p), some have real values (e.g., the slope parameter). Timescale errors may also be modelled (additional dimensions).

The method subspace (Fig. 9.1) describes the estimation and CI construction. The ticks along the estimator dimension are named least squares, maximum likelihood, and so forth. CI construction requires more dimensions: one for distinguishing between classical and bootstrap CIs, and several for detailing the bootstrap methodology (block length selection for MBB, calibration, subsampling, etc.) and calculating the interval bounds from the replications. Consider, for example, the brute-force block length selector (Berkowitz and Kilian 2000): one dimension with integer values between 1 and $n-1$.

The measure subspace (Fig. 9.1) describes how to detect the optimal estimation method for the Monte Carlo experiment: CI accuracy and width, RMSE, bias, robustness, and so forth. It should make sense to consider also joint measures (e.g., CI accuracy and robustness).

The hyperspace of climate parameter estimation is large. Present computing power limits our ability to explore it and find the optimal method for solving a (climate) estimation problem. This book has examined many important estimation problems (regression, spectrum, extremes and correlation) but visited only parts of the hyperspace by means of Monte Carlo experiments. For example, in linear regression (Chapter 4), we have studied

- $\theta = \beta_0$ (intercept) and β_1 (slope);
- prescribed $\beta_0 = 2, \beta_1 = 2$;
- $n \in \{10, 20, 50, 100, 200, 500, 1000\}$;
- spacing: even and uneven (timescale errors);
- shape: Gaussian and lognormal;
- persistence: AR(1), AR(2) and ARFIMA$(0, 0.25, 0)$;
- estimator: least squares only;
- resampling: ARB, MBB, subsampling, timescale-ARB, timescale-MBB and pairwise-MBB;
- CI type: classical and bootstrap BCa;
- confidence level: 90, 95 and 99%;
- calibration loop: none;

and

- measure: RMSE, CI accuracy and CI length.

We have found "acceptable" results (mainly judged via CI accuracy) from the bootstrap method applied to Monte Carlo samples generated from designed processes that are considered as close to the climate processes. These positive results have given us confidence that the results (estimate with CI) from analysing the observed, real climate time series are valid. However, we have to concede that there may exist more accurate methods, resulting in particular from (computing-intensive) CI calibration. This may be of relevance especially for small sample sizes.

The envisaged large increase in computing power may bring the following idea of optimal climate estimation into existence. Given a time series, $\{t(i), x(i)\}_{i=1}^{n}$, some prior information (e.g., measurement standard errors, age–depth curve) and a set of questions (parameters to be estimated), the first task is simple: perform an initial estimation on basis of existing knowledge and experience with such types of estimation problems. The second task requires the computing power: explore the hyperspace (Fig. 9.1) to find the suitable method, that is, the mode of estimation and CI construction that optimizes a selected measure for prescribed values close to the initial estimates. Also here, intelligent exploration methods (gradient, Brent, etc.) are useful. The third task is to apply the optimal estimation method to the climate time series.

References

Abarbanel HDI, Brown R, Sidorowich JJ, Tsimring LS (1993) The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics* 65(4): 1331–1392.

Abraham B, Wei WWS (1984) Inferences about the parameters of a time series model with changing variance. *Metrika* 31(3–4): 183–194.

Abram NJ, Gagan MK, Cole JE, Hantoro WS, Mudelsee M (2008) Recent intensification of tropical climate variability in the Indian Ocean. *Nature Geoscience* 1(12): 849–853.

Abram NJ, Mulvaney R, Wolff EW, Mudelsee M (2007) Ice core records as sea ice proxies: An evaluation from the Weddell Sea region of Antarctica. *Journal of Geophysical Research* 112(D15): D15101. [doi:10.1029/2006JD008139]

Abramowitz M, Stegun IA (Eds) (1965) *Handbook of Mathematical Functions.* Dover, New York, 1046 pp.

Adams JB, Mann ME, Ammann CM (2003) Proxy evidence for an El Niño-like response to volcanic forcing. *Nature* 426(6964): 274–278.

Adcock RJ (1877) Note on the method of least squares. *Analyst* 4(6): 183–184.

Adcock RJ (1878) A problem in least squares. *Analyst* 5(2): 53–54.

Agrinier P, Gallet Y, Lewin E (1999) On the age calibration of the geomagnetic polarity timescale. *Geophysical Journal International* 137(1): 81–90.

Ahrens JH, Dieter U (1974) Computer methods for sampling from gamma, beta, Poisson and binomial distributions. *Computing* 12(3): 223–246.

Aitchison J, Brown JAC (1957) *The Lognormal Distribution.* Cambridge University Press, Cambridge, 176 pp.

Akaike H (1960) Effect of timing-error on the power spectrum of sampled-data. *Annals of the Institute of Statistical Mathematics* 11: 145–165.

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csáki F (Eds) *Second International Symposium on Information Theory.* Akadémiai Kiadó, Budapest, pp 267–281.

Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Rupa Kumar K, Revadekar J, Griffiths G, Vincent L, Stephenson DB, Burn J, Aguilar E, Brunet M, Taylor M, New M, Zhai P, Rusticucci M, Vazquez-Aguirre JL (2006) Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research* 111(D5): D05109. [doi:10.1029/2005JD006290]

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7,

Allamano P, Claps P, Laio F (2009) Global warming increases flood risk in mountainous areas. *Geophysical Research Letters* 36(24): L24404. [doi:10.1029/2009GL041395]

Allen M (1999) Do-it-yourself climate prediction. *Nature* 401(6754): 642.

Allen MR, Smith LA (1994) Investigating the origins and significance of low-frequency modes of climate variability. *Geophysical Research Letters* 21(10): 883–886.

Allen MR, Stott PA (2003) Estimating signal amplitudes in optimal fingerprinting, Part I: Theory. *Climate Dynamics* 21(5–6): 477–491.

Allen MR, Stott PA, Mitchell JFB, Schnur R, Delworth TL (2000) Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* 407(6804): 617–620.

Ammann CM, Genton MG, Li B (2009) Technical note: Correcting for signal attenuation from noise: Sharpening the focus on past climate. *Climate of the Past Discussions* 5(3): 1645–1657.

Ammann CM, Naveau P (2003) Statistical analysis of tropical explosive volcanism occurrences over the last 6 centuries. *Geophysical Research Letters* 30(5): 1210. [doi:10.1029/2002GL016388]

Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Du Croz J, Greenbaum A, Hammarling S, McKenney A, Sorensen D (1999) *LAPACK Users' Guide*. Third edition. SIAM, Philadelphia, PA.

Anderson TW (1971) *The Statistical Analysis of Time Series*. Wiley, New York, 704 pp.

Andrews DWK, Buchinsky M (2000) A three-step method for choosing the number of bootstrap repetitions. *Econometrica* 68(1): 23–51.

Andrews DWK, Buchinsky M (2002) On the number of bootstrap repetitions for BC_a confidence intervals. *Econometric Theory* 18(4): 962–984.

Andrews DWK, Lieberman O (2002) *Higher-order Improvements of the Parametric Bootstrap for Long-memory Gaussian Processes*. Cowles Foundation for Research in Economics, Yale University, New Haven, CT, 40 pp. [Discussion Paper No. 1378]

Angelini C, Cava D, Katul G, Vidakovic B (2005) Resampling hierarchical processes in the wavelet domain: A case study using atmospheric turbulence. *Physica D* 207(1–2): 24–40.

Angus JE (1993) Asymptotic theory for bootstrapping the extremes. *Communications in Statistics—Theory and Methods* 22(1): 15–30.

Antle CE (1985) Lognormal distribution. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 5. Wiley, New York, pp 134–136.

Appleby PG, Oldfield F (1992) Application of lead-210 to sedimentation studies. In: Ivanovich M, Harmon RS (Eds) *Uranium-series Disequilibrium: Applications to Earth, Marine, and Environmental Sciences*, second edition. Clarendon Press, Oxford, pp 731–778.

Arnold L (2001) Hasselmann's program revisited: The analysis of stochasticity in deterministic climate models. In: Imkeller P, von Storch J-S (Eds) *Stochastic Climate Models*. Birkhäuser, Basel, pp 141–158.

Atkinson AC, Cox DR (1988) Transformations. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 9. Wiley, New York, pp 312–318.

Bai J, Perron P (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66(1): 47–78.

Baker A, Smart PL, Edwards RL, Richards DA (1993) Annual growth banding in a cave stalagmite. *Nature* 364(6437): 518–520.

Bard E, Frank M (2006) Climate change and solar variability: What's new under the sun? *Earth and Planetary Science Letters* 248(1–2): 1–14.

Barnard GA (1959) Control charts and stochastic processes (with discussion). *Journal of the Royal Statistical Society, Series B* 21(2): 239–271.

Barnard GA (1982) Causation. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 1. Wiley, New York, pp 387–389.

Barnett T, Zwiers F, Hegerl G, Allen M, Crowley T, Gillett N, Hasselmann K, Jones P, Santer B, Schnur R, Stott P, Taylor K, Tett S (2005) Detecting and attributing external influences on the climate system: A review of recent advances. *Journal of Climate* 18(9): 1291–1314.

Barnola JM, Raynaud D, Korotkevich YS, Lorius C (1987) Vostok ice core provides 160,000-year record of atmospheric CO_2. *Nature* 329(6138): 408–414.

Bartlett MS (1946) On the theoretical specification and sampling properties of autocorrelated time-series. *Journal of the Royal Statistical Society, Supplement* 8(1): 27–41. [Corrigendum: 1948 Vol. 10(1)]

Bartlett MS (1949) Fitting a straight line when both variables are subject to error. *Biometrics* 5(3): 207–212.

Bartlett MS (1950) Periodogram analysis and continuous spectra. *Biometrika* 37(1–2): 1–16.

Bartlett MS (1955) *An introduction to stochastic processes with special reference to methods and applications.* Cambridge University Press, Cambridge, 312 pp.

Basseville M, Nikiforov IV (1993) *Detection of Abrupt Changes: Theory and Application.* Prentice-Hall, Englewood Cliffs, NJ, 447 pp.

Bayley GV, Hammersley JM (1946) The "effective" number of independent observations in an autocorrelated time series. *Journal of the Royal Statistical Society, Supplement* 8(2): 184–197.

Beasley WH, DeShea L, Toothaker LE, Mendoza JL, Bard DE, Rodgers JL (2007) Bootstrapping to test for nonzero population correlation coefficients using univariate sampling. *Psychological Methods* 12(4): 414–433.

Becker A, Grünewald U (2003) Flood risk in central Europe. *Science* 300(5622): 1099.

Beer J, Baumgartner S, Dittrich-Hannen B, Hauenstein J, Kubik P, Lukasczyk C, Mende W, Stellmacher R, Suter M (1994) Solar variability traced by cosmogenic isotopes. In: Pap JM, Fröhlich C, Hudson HS, Solanki SK (Eds) *The Sun as a Variable Star: Solar and Stellar Irradiance Variations.* Cambridge University Press, Cambridge, pp 291–300.

Beer J, Tobias S, Weiss N (1998) An active sun throughout the Maunder Minimum. *Solar Physics* 181(1): 237–249.

Beersma JJ, Buishand TA (1999) A simple test for equality of variances in monthly climate data. *Journal of Climate* 12(6): 1770–1779.

Beirlant J, Goegebeur Y, Teugels J, Segers J (2004) *Statistics of Extremes: Theory and Applications.* Wiley, Chichester, 490 pp.

Beirlant J, Teugels JL, Vynckier P (1996) *Practical Analysis of Extreme Values.* Leuven University Press, Leuven, 137 pp.

Belaire-Franch J, Contreras-Bayarri D (2002) Improving cross-correlation tests through re-sampling techniques. *Journal of Applied Statistics* 29(5): 711–720.

Belcher J, Hampton JS, Tunnicliffe Wilson G (1994) Parameterization of continuous time autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Society, Series B* 56(1): 141–155.

Bell B, Percival DB, Walden AT (1993) Calculating Thomson's spectral multitapers by inverse iteration. *Journal of Computational and Graphical Statistics* 2(1): 119–130.

Bendat JS, Piersol AG (1986) *Random Data: Analysis and Measurement Procedures.* Second edition. Wiley, New York, 566 pp.

Bengtsson L, Botzet M, Esch M (1996) Will greenhouse gas-induced warming over the next 50 years lead to higher frequency and greater intensity of hurricanes? *Tellus, Series A* 48(1): 57–73.

Beniston M (2004) The 2003 heat wave in Europe: A shape of things to come? An analysis based on Swiss climatological data and model simulations. *Geophysical Research Letters* 31(2): L02202. [doi:10.1029/2003GL018857]

Bennett KD (1994) Confidence intervals for age estimates and deposition times in late-Quaternary sediment sequences. *The Holocene* 4(4): 337–348.

Bennett KD, Fuller JL (2002) Determining the age of the mid-Holocene *Tsuga canadensis* (hemlock) decline, eastern North America. *The Holocene* 12(4): 421–429.

Beran J (1994) *Statistics for Long-Memory Processes.* Chapman and Hall, Boca Raton, FL, 315 pp.

Beran J (1997) Long-range dependence. In: Kotz S, Read CB, Banks DL (Eds) *Encyclopedia of statistical sciences*, volume U1. Wiley, New York, pp 385–390.

Beran J (1998) Fractional ARIMA models. In: Kotz S, Read CB, Banks DL (Eds) *Encyclopedia of statistical sciences*, volume U2. Wiley, New York, pp 269–271.

Beran R (1987) Prepivoting to reduce level error of confidence sets. *Biometrika* 74(3): 457–468.

Beran R (1988) Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83(403): 687–697.

Berger A, Loutre MF (1991) Insolation values for the climate of the last 10 million years. *Quaternary Science Reviews* 10(4): 297–317.

Berger A, Loutre MF (2002) An exceptionally long interglacial ahead? *Science* 297(5585): 1287–1288.

Berger A, Loutre MF, Mélice JL (1998) Instability of the astronomical periods from 1.5 Myr BP to 0.5 Myr AP. *Paleoclimates* 2(4): 239–280.

Berggren WA, Hilgen FJ, Langereis CG, Kent DV, Obradovich JD, Raffi I, Raymo ME, Shackleton NJ (1995a) Late Neogene chronology: New perspectives in high-resolution stratigraphy. *Geological Society of America Bulletin* 107(11): 1272–1287.

Berggren WA, Kent DV, Swisher III CC, Aubry M-P (1995b) A revised Cenozoic geochronology and chronostratigraphy. In: Berggren WA, Kent DV, Aubry M-P, Hardenbol J (Eds) *Geochronology, Time Scales and Global Stratigraphic Correlation.* Society for Sedimentary Geology, Tulsa, OK, pp 129–212. [SEPM Special Publication No. 54]

Berkowitz J, Kilian L (2000) Recent developments in bootstrapping time series. *Econometric Reviews* 19(1): 1–48.

Berman SM (1964) Limit theorems for the maximum term in stationary sequences. *Annals of Mathematical Statistics* 35(2): 502–516.

Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM, West M (Eds) (2003) *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting.* Clarendon Press, Oxford, 750 pp.

Bernardo JM, Smith AFM (1994) *Bayesian theory.* Wiley, Chichester, 586 pp.

Besonen MR (2006) *A 1,000 year high-resolution hurricane history for the Boston area based on the varved sedimentary record from the Lower Mystic Lake (Medford/Arlington, MA).* Ph.D. Dissertation. University of Massachusetts at Amherst, Amherst, MA, 297 pp.

Besonen MR, Bradley RS, Mudelsee M, Abbott MB, Francus P (2008) A 1,000-year, annually-resolved record of hurricane activity from Boston, Massachusetts. *Geophysical Research Letters* 35(14): L14705. [doi:10.1029/2008GL033950]

Beutler FJ (1970) Alias-free randomly timed sampling of stochastic processes. *IEEE Transactions on Information Theory* 16(2): 147–152.

Bickel P (1988) Robust estimation. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 8. Wiley, New York, pp 157–163.

Bickel PJ, Freedman DA (1981) Some asymptotic theory for the bootstrap. *The Annals of Statistics* 9(6): 1196–1217.

Bigler M, Wagenbach D, Fischer H, Kipfstuhl J, Miller H, Sommer S, Stauffer B (2002) Sulphate record from a northeast Greenland ice core over the last 1200 years based on continuous flow analysis. *Annals of Glaciology* 35(1): 250–256.

Blaauw M, Christen JA (2005) Radiocarbon peat chronologies and environmental change. *Applied Statistics* 54(4): 805–816.

Bloomfield P, Royle JA, Steinberg LJ, Yang Q (1996) Accounting for meteorological effects in measuring urban ozone levels and trends. *Atmospheric Environment* 30(17): 3067–3077.

Bloomfield P, Steiger WL (1983) *Least Absolute Deviations: Theory, Applications, and Algorithms.* Birkhäuser, Boston, 349 pp.

Blunier T, Chappellaz J, Schwander J, Dällenbach A, Stauffer B, Stocker TF, Raynaud D, Jouzel J, Clausen HB, Hammer CU, Johnsen SJ (1998) Asynchrony of Antarctic and Greenland climate change during the last glacial period. *Nature* 394(6695): 739–743.

Boessenkool KP, Hall IR, Elderfield H, Yashayaev I (2007) North Atlantic climate and deep-ocean flow speed changes during the last 230 years. *Geophysical Research Letters* 34(13): L13614. [doi:10.1029/2007GL030285]

Bolch BW (1968) More on unbiased estimation of the standard deviation. *The American Statistician* 22(3): 27.

Bond G, Kromer B, Beer J, Muscheler R, Evans MN, Showers W, Hoffmann S, Lotti-Bond R, Hajdas I, Bonani G (2001) Persistent solar influence on North Atlantic climate during the Holocene. *Science* 294(5549): 2130–2136.

Bond G, Showers W, Cheseby M, Lotti R, Almasi P, deMenocal P, Priore P, Cullen H, Hajdas I, Bonani G (1997) A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates. *Science* 278(5341): 1257–1266.

Booth JG, Hall P (1993) Bootstrap confidence regions for functional relationships in errors-in-variables models. *The Annals of Statistics* 21(4): 1780–1791.

Booth JG, Hall P (1994) Monte Carlo approximation and the iterated bootstrap. *Biometrika* 81(2): 331–340.

Booth NB, Smith AFM (1982) A Bayesian approach to retrospective identification of change-points. *Journal of Econometrics* 19(1): 7–22.

Bose A (1988) Edgeworth correction by bootstrap in autoregressions. *The Annals of Statistics* 16(4): 1709–1722.

Box GEP (1953) Non-normality and tests on variances. *Biometrika* 40(3–4): 318–335.

Box GEP (1966) Use and abuse of regression. *Technometrics* 8(4): 625–629.

Box GEP, Jenkins GM, Reinsel GC (1994) *Time Series Analysis: Forecasting and Control.* Third edition. Prentice-Hall, Englewood Cliffs, NJ, 598 pp.

Box GEP, Muller ME (1958) A note on the generation of random normal deviates. *Annals of Mathematical Statistics* 29(2): 610–611.

Bradley RS (1999) *Paleoclimatology: Reconstructing Climates of the Quaternary.* Second edition. Academic Press, San Diego, 610 pp.

Brázdil R, Glaser R, Pfister C, Dobrovolný P, Antoine J-M, Barriendos M, Camuffo D, Deutsch M, Enzi S, Guidoboni E, Kotyza O, Rodrigo FS (1999) Flood events of selected European rivers in the sixteenth century. *Climatic Change* 43(1): 239–285.

Brázdil R, Pfister C, Wanner H, von Storch H, Luterbacher J (2005) Historical climatology in Europe—the state of the art. *Climatic Change* 70(3): 363–430.

Breiman L (1996) Bagging predictors. *Machine Learning* 24(2): 123–140.

Brent RP (1973) *Algorithms for minimization without derivatives.* Prentice-Hall, Englewood Cliffs, NJ, 195 pp.

Brillinger DR (1975) *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, New York, 500 pp.

Brillinger DR (2002) John W. Tukey's work on time series and spectrum analysis. *The Annals of Statistics* 30(6): 1595–1618.

Brockmann M, Gasser T, Herrmann E (1993) Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association* 88(424): 1302–1309.

Brockwell PJ, Davis RA (1991) *Time Series: Theory and Methods.* Second edition. Springer, New York, 577 pp.

Brockwell PJ, Davis RA (1996) *Introduction to Time Series and Forecasting.* Springer, New York, 420 pp.

Broecker WS, Henderson GM (1998) The sequence of events surrounding Termination II and their implications for the cause of glacial–interglacial CO_2 changes. *Paleoceanography* 13(4): 352–364.

Broecker WS, Peng T-H (1982) *Tracers in the Sea.* Eldigio Press, New York, 690 pp.

Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research* 111(D12): D12106. [doi:10.1029/2005JD006548]

Bronez TP (1988) Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36(12): 1862–1873.

Brooks MM, Marron JS (1991) Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications* 38(1): 157–165.

Broomhead DS, King GP (1986) Extracting qualitative dynamics from experimental data. *Physica D* 20(2–3): 217–236.

Brown RL, Durbin J, Evans JM (1975) Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society, Series B* 37(2): 149–192.

Brückner E (1890) Klimaschwankungen seit 1700 nebst Bemerkungen über die Klimaschwankungen der Diluvialzeit. *Geographische Abhandlungen* 4(2): 153–484.

Brüggemann W (1992) A minimal cost function method for optimizing the age–depth relation of deep-sea sediment cores. *Paleoceanography* 7(4): 467–487.

Brumback BA, Ryan LM, Schwartz JD, Neas LM, Stark PC, Burge HA (2000) Transitional regression models, with application to environmental time series. *Journal of the American Statistical Association* 95(449): 16–27.

Buck CE, Millard AR (Eds) (2004) *Tools for Constructing Chronologies: Crossing Disciplinary Boundaries.* Springer, London, 257 pp.

Bühlmann P (1994) Blockwise bootstrapped empirical process for stationary sequences. *The Annals of Statistics* 22(2): 995–1012.

Bühlmann P (1997) Sieve bootstrap for time series. *Bernoulli* 3(2): 123–148.

Bühlmann P (1998) Sieve bootstrap for smoothing in nonstationary time series. *The Annals of Statistics* 26(1): 48–83.

Bühlmann P (2002) Bootstraps for time series. *Statistical Science* 17(1): 52–72.

Bühlmann P, Künsch HR (1999) Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis* 31(3): 295–310.

Buishand TA (1989) Statistics of extremes in climatology. *Statistica Neerlandica* 43(1): 1–30.

Buja A, Hastie T, Tibshirani R (1989) Linear smoothers and additive models. *The Annals of Statistics* 17(2): 453–510.

Bunde A, Eichner JF, Havlin S, Koscielny-Bunde E, Schellnhuber HJ, Vyushin D (2004) Comment on "Scaling of atmosphere and ocean temperature correlations in observations and climate models." *Physical Review Letters* 92(3): 039801. [doi:10.1103/PhysRevLett.92.039801]

Bunde A, Eichner JF, Kantelhardt JW, Havlin S (2005) Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Physical Review Letters* 94(4): 048701. [doi:10.1103/PhysRevLett.94.048701]

Burns SJ, Fleitmann D, Mudelsee M, Neff U, Matter A, Mangini A (2002) A 780-year annually resolved record of Indian Ocean monsoon precipitation from a speleothem from south Oman. *Journal of Geophysical Research* 107(D20): 4434. [doi:10.1029/2001JD001281]

Butler A, Heffernan JE, Tawn JA, Flather RA (2007) Trend estimation in extremes of synthetic North Sea surges. *Applied Statistics* 56(4): 395–414.

Caers J, Beirlant J, Maes MA (1999a) Statistics for modeling heavy tailed distributions in geology: Part I. Methodology. *Mathematical Geology* 31(4): 391–410.

Caers J, Beirlant J, Maes MA (1999b) Statistics for modeling heavy tailed distributions in geology: Part II. Application. *Mathematical Geology* 31(4): 411–434.

Caillon N, Severinghaus JP, Jouzel J, Barnola J-M, Kang J, Lipenkov VY (2003) Timing of atmospheric CO_2 and Antarctic temperature changes across Termination III. *Science* 299(5613): 1728–1731.

Cande SC, Kent DV (1992) A new geomagnetic polarity time scale for the late Cretaceous and Cenozoic. *Journal of Geophysical Research* 97(B10): 13917–13951.

Cande SC, Kent DV (1995) Revised calibration of the geomagnetic polarity timescale for the late Cretaceous and Cenozoic. *Journal of Geophysical Research* 100(B4): 6093–6095.

Candolo C, Davison AC, Demétrio CGB (2003) A note on model uncertainty in linear regression. *The Statistician* 52(2): 165–177.

Carlstein E (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics* 14(3): 1171–1179.

Carlstein E, Do K-A, Hall P, Hesterberg T, Künsch HR (1998) Matched-block bootstrap for dependent data. *Bernoulli* 4(3): 305–328.

Carpenter J, Bithell J (2000) Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine* 19(9): 1141–1164.

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) *Measurement Error in Nonlinear Models: A Modern Perspective.* Second edition. Chapman and Hall, Boca Raton, FL, 455 pp.

Casella G (Ed) (2003) *Silver Anniversary of the Bootstrap*, volume 18(2) of *Statistical Science*. [Special issue]

Castillo E, Hadi AS (1997) Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association* 92(440): 1609–1620.

Caussinus H, Mestre O (2004) Detection and correction of artificial shifts in climate series. *Applied Statistics* 53(3): 405–425.

Chan K-S, Tong H (2001) *Chaos: A Statistical Perspective*. Springer, New York, 300 pp.

Chan KS, Tong H (1987) A note on embedding a discrete parameter ARMA model in a continuous parameter ARMA model. *Journal of Time Series Analysis* 8(3): 277–281.

Chan W, Chan DW-L (2004) Bootstrap standard error and confidence intervals for the correlation corrected for range restriction: A simulation study. *Psychological Methods* 9(3): 369–385.

Chang EKM, Guo Y (2007) Is the number of North Atlantic tropical cyclones significantly underestimated prior to the availability of satellite observations? *Geophysical Research Letters* 34(14): L14801. [doi:10.1029/2007GL030169]

Chatfield C (1995) Model uncertainty, data mining and statistical inference (with discussion). *Journal of the Royal Statistical Society, Series A* 158(3): 419–466.

Chatfield C (2004) *The Analysis of Time Series: An Introduction*. Sixth edition. Chapman and Hall, Boca Raton, FL, 333 pp.

Chaudhuri P, Marron JS (1999) SiZer for exploration of structures in curves. *Journal of the American Statistical Association* 94(447): 807–823.

Chave AD, Luther DS, Filloux JH (1997) Observations of the boundary current system at 25.5°N in the subtropical North Atlantic Ocean. *Journal of Physical Oceanography* 27(9): 1827–1848.

Chavez-Demoulin V, Davison AC (2005) Generalized additive modelling of sample extremes. *Applied Statistics* 54(1): 207–222.

Chen J, Gupta AK (2000) *Parametric Statistical Change Point Analysis*. Birkhäuser, Boston, 184 pp.

Choi E, Hall P (2000) Bootstrap confidence regions computed from autoregressions of arbitrary order. *Journal of the Royal Statistical Society, Series B* 62(3): 461–477.

Chree C (1913) Some phenomena of sunspots and of terrestrial magnetism at Kew observatory. *Philosophical Transactions of the Royal Society of London, Series A* 212: 75–116.

Chree C (1914) Some phenomena of sunspots and of terrestrial magnetism—Part II. *Philosophical Transactions of the Royal Society of London, Series A* 213: 245–277.

Christensen JH, Hewitson B, Busuioc A, Chen A, Gao X, Held I, Jones R, Kolli RK, Kwon W-T, Laprise R, Magaña Rueda V, Mearns L, Menéndez CG, Räisänen J, Rinke A, Sarr A, Whetton P (2007) Regional climate projections. In: Solomon S, Qin D, Manning M, Marquis M, Averyt K, Tignor MMB, Miller Jr HL, Chen Z (Eds) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, pp 847–940.

Chu CK (1994) Estimation of change-points in a nonparametric regression function through kernel density estimation. *Communications in Statistics—Theory and Methods* 23(11): 3037–3062.

Chu JT (1955) On the distribution of the sample median. *Annals of Mathematical Statistics* 26(1): 112–116.

Chylek P, Lohmann U, Dubey M, Mishchenko M, Kahn R, Ohmura A (2007) Limits on climate sensitivity derived from recent satellite and surface observations. *Journal of Geophysical Research* 112(D24): D24S04. [doi:10.1029/2007JD008740]

Cini Castagnoli G, Provenzale A (Eds) (1997) *Past and Present Variability of the Solar–Terrestrial System: Measurement, Data Analysis and Theoretical Models.* Società Italiana di Fisica, Bologna, 491 pp.

Clarke RT (1994) *Statistical Modelling in Hydrology.* Wiley, Chichester, 412 pp.

Clement BM (2004) Dependence of the duration of geomagnetic polarity reversals on site latitude. *Nature* 428(6983): 637–640.

Cobb GW (1978) The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika* 65(2): 243–251.

Cochrane D, Orcutt GH (1949) Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* 44(245): 32–61.

Coles S (2001a) Improving the analysis of extreme wind speeds with information-sharing models. *Institut Pierre Simon Laplace des Sciences de l'Environnement Global, Notes des Activitês Instrumentales* 11: 23–34.

Coles S (2001b) *An Introduction to Statistical Modeling of Extreme Values.* Springer, London, 208 pp.

Coles S (2004) The use and misuse of extreme value models in practice. In: Finkenstädt B, Rootzén H (Eds) *Extreme Values in Finance, Telecommunications, and the Environment.* Chapman and Hall, Boca Raton, FL, pp 79–100.

Coles S, Pericchi L (2003) Anticipating catastrophes through extreme value modelling. *Applied Statistics* 52(4): 405–416.

Comte F, Renault E (1996) Long memory continuous time models. *Journal of Econometrics* 73(1): 101–149.

Cook RD, Weisberg S (1982) *Residuals and Influence in Regression.* Chapman and Hall, New York, 230 pp.

Cooley D, Nychka D, Naveau P (2007) Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* 102(479): 824–840.

Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation* 19(90): 297–301.

Cowling A, Hall P (1996) On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society, Series B* 58(3): 551–563.

Cowling A, Hall P, Phillips MJ (1996) Bootstrap confidence regions for the intensity of a Poisson point process. *Journal of the American Statistical Association* 91(436): 1516–1524.

Cowling AM (1995) *Some problems in kernel curve estimation.* Ph.D. Dissertation. Australian National University, Canberra, 130 pp.

Cox A (1969) Geomagnetic reversals. *Science* 163(3864): 237–245.

Cox DR, Isham V (1980) *Point Processes.* Chapman and Hall, London, 188 pp.

Cox DR, Isham VS, Northrop PJ (2002) Floods: Some probabilistic and statistical approaches. *Philosophical Transactions of the Royal Society of London, Series A* 360(1796): 1389–1408.

Cox DR, Lewis PAW (1966) *The Statistical Analysis of Series of Events.* Methuen, London, 285 pp.

Cramér H (1946) *Mathematical Methods of Statistics.* Princeton University Press, Princeton, 575 pp.

Cronin TM (2010) *Paleoclimates: Understanding Climate Change Past and Present.* Columbia University Press, New York, 441 pp.

Crow EL, Shimizu K (Eds) (1988) *Lognormal Distributions: Theory and Applications.* Marcel Dekker, New York, 387 pp.

Crowley TJ, North GR (1991) *Paleoclimatology.* Oxford University Press, New York, 339 pp.

Crutzen PJ (2002) Geology of mankind. *Nature* 415(6867): 23.

Crutzen PJ, Steffen W (2003) How long have we been in the Anthropocene era? *Climatic Change* 61(3): 251–257.

Cuffey KM, Vimeux F (2001) Covariation of carbon dioxide and temperature from the Vostok ice core after deuterium-excess correction. *Nature* 412(6846): 523–527.

Cureton EE (1968a) Priority correction to "Unbiased estimation of the standard deviation." *The American Statistician* 22(3): 27.

Cureton EE (1968b) Unbiased estimation of the standard deviation. *The American Statistician* 22(1): 22.

Cutter SL, Emrich C (2005) Are natural hazards and disaster losses in the U.S. increasing? *Eos, Transactions of the American Geophysical Union* 86(41): 381, 389.

Dahlquist G, Björck Å (2008) *Numerical Methods in Scientific Computing*, volume 1. SIAM, Philadelphia, PA, 717 pp.

Dahlquist G, Björck Å (in press) *Numerical Methods in Scientific Computing*, volume 2. SIAM, Philadelphia, PA.

Dalfes HN, Schneider SH, Thompson SL (1984) Effects of bioturbation on climatic spectra inferred from deep sea cores. In: Berger A, Imbrie J, Hays J, Kukla G, Saltzman B (Eds) *Milankovitch and Climate*, volume 1. D. Reidel, Dordrecht, pp 481–492.

Dalrymple GB, Lanphere MA (1969) *Potassium–Argon Dating.* Freeman, San Francisco, 258 pp.

Damon PE, Laut P (2004) Pattern of strange errors plagues solar activity and terrestrial climate data. *Eos, Transactions of the American Geophysical Union* 85(39): 370, 374.

Dansgaard W, Oeschger H (1989) Past environmental long-term records from the Arctic. In: Oeschger H, Langway Jr CC (Eds) *The Environmental Record in Glaciers and Ice Sheets.* Wiley, Chichester, pp 287–317.

Daoxian Y, Cheng Z (Eds) (2002) *Karst Processes and the Carbon Cycle.* Geological Publishing House, Beijing, 220 pp.

Dargahi-Noubary GR (1989) On tail estimation: An improved method. *Mathematical Geology* 21(8): 829–842.

Daubechies I, Guskov I, Schröder P, Sweldens W (1999) Wavelets on irregular point sets. *Philosophical Transactions of the Royal Society of London, Series A* 357(1760): 2397–2413.

David FN, Mallows CL (1961) The variance of Spearman's rho in normal samples. *Biometrika* 48(1–2): 19–28.

Davis JC (1986) *Statistics and Data Analysis in Geology.* Second edition. Wiley, New York, 646 pp.

Davison AC (2003) *Statistical models.* Cambridge University Press, Cambridge, 726 pp.

Davison AC, Hinkley DV (1997) *Bootstrap methods and their application.* Cambridge University Press, Cambridge, 582 pp.

Davison AC, Hinkley DV, Schechtman E (1986) Efficient bootstrap simulation. *Biometrika* 73(3): 555–566.

Davison AC, Hinkley DV, Young GA (2003) Recent developments in bootstrap methodology. *Statistical Science* 18(2): 141–157.

Davison AC, Ramesh NI (2000) Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society, Series B* 62(1): 191–208.

Davison AC, Smith RL (1990) Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B* 52(3): 393–442.

DeBlonde G, Peltier WR (1991) A one-dimensional model of continental ice volume fluctuations through the Pleistocene: Implications for the origin of the mid-Pleistocene climate transition. *Journal of Climate* 4(3): 318–344.

Deep Sea Drilling Project (Ed) (1969–1986) *Initial Reports of the Deep Sea Drilling Project*, volume 1–96. U.S. Govt. Printing Office, Washington, DC.

Della-Marta PM, Haylock MR, Luterbacher J, Wanner H (2007) Doubled length of western European summer heat waves since 1880. *Journal of Geophysical Research* 112(D15): D15103. [doi:10.1029/2007JD008510]

Deming WE (1943) *Statistical Adjustment of Data.* Wiley, New York, 261 pp.

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the *EM* algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* 39(1): 1–38.

De Pol-Holz R, Ulloa O, Lamy F, Dezileau L, Sabatier P, Hebbeln D (2007) Late Quaternary variability of sedimentary nitrogen isotopes in the eastern South Pacific Ocean. *Paleoceanography* 22(2): PA2207. [doi:10.1029/2006PA001308]

De Ridder F, de Brauwere A, Pintelon R, Schoukens J, Dehairs F (2006) Identification of the accretion rate for annually resolved archives. *Biogeosciences Discussions* 3(2): 321–344.

de Vries H (1958) Variation in concentration of radiocarbon with time and location on Earth. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series B* 61(2): 94–102.

Dhrymes PJ (1981) *Distributed Lags: Problems of Estimation and Formulation.* Second edition. North-Holland, Amsterdam, 470 pp.

Diaz HF, Pulwarty RS (1994) An analysis of the time scales of variability in centuries-long ENSO-sensitive records in the last 1000 years. *Climatic Change* 26(2–3): 317–342.

DiCarlo L, Chow JM, Gambetta JM, Bishop LS, Johnson BR, Schuster DI, Majer J, Blais A, Frunzio L, Girvin SM, Schoelkopf RJ (2009) Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* 460(7252): 240–244.

DiCiccio T, Efron B (1992) More accurate confidence intervals in exponential families. *Biometrika* 79(2): 231–245.

DiCiccio TJ, Efron B (1996) Bootstrap confidence intervals (with discussion). *Statistical Science* 11(3): 189–228.

Diebold FX, Inoue A (2001) Long memory and regime switching. *Journal of Econometrics* 105(1): 131–159.

Diggle P (1985) A kernel method for smoothing point process data. *Applied Statistics* 34(2): 138–147.

Diggle P, Marron JS (1988) Equivalence of smoothing parameter selectors in density and intensity estimation. *Journal of the American Statistical Association* 83(403): 793–800.

Diggle PJ (1990) *Time Series: A Biostatistical Introduction.* Clarendon Press, Oxford, 257 pp.

Diggle PJ, Hutchinson MF (1989) On spline smoothing with autocorrelated errors. *Australian Journal of Statistics* 31(1): 166–182.

Diks C (1999) *Nonlinear Time Series Analysis: Methods and Applications.* World Scientific, Singapore, 209 pp.

Diks C, DeGoede J (2001) A general nonparametric bootstrap test for Granger causality. In: Broer HW, Krauskopf B, Vegter G (Eds) *Global Analysis of Dynamical Systems.* Institute of Physics Publishing, Bristol, pp 391–403.

Diks C, Mudelsee M (2000) Redundancies in the Earth's climatological time series. *Physics Letters A* 275(5–6): 407–414.

Divine DV, Polzehl J, Godtliebsen F (2008) A propagation-separation approach to estimate the autocorrelation in a time-series. *Nonlinear Processes in Geophysics* 15(4): 591–599.

Donner RV, Barbosa SM (Eds) (2008) *Nonlinear Time Series Analysis in the Geosciences: Applications in Climatology, Geodynamics and Solar–Terrestrial Physics.* Springer, Berlin, 390 pp.

Doornik JA, Ooms M (2001) *A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima package 1.01 for Ox.* Nuffield College, University of Oxford, Oxford, 32 pp.

Doornik JA, Ooms M (2003) Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics and Data Analysis* 42(3): 333–348.

Doran HE (1983) Lag models, distributed. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 4. Wiley, New York, pp 440–448.

Douglass AE (1919) *Climatic Cycles and Tree-Growth: A Study of the Annual Rings of Trees in Relation to Climate and Solar Activity*, volume 1. Carnegie Institution of Washington, Washington, DC, 127 pp.

Douglass AE (1928) *Climatic Cycles and Tree-Growth: A Study of the Annual Rings of Trees in Relation to Climate and Solar Activity*, volume 2. Carnegie Institution of Washington, Washington, DC, 166 pp.

Douglass AE (1936) *Climatic Cycles and Tree Growth: A Study of Cycles*, volume 3. Carnegie Institution of Washington, Washington, DC, 171 pp.

Doukhan P, Oppenheim G, Taqqu MS (Eds) (2003) *Theory and Applications of Long-Range Dependence.* Birkhäuser, Boston, 719 pp.

Draper D (1995) Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B* 57(1): 45–97.

Draper NR, Smith H (1981) *Applied Regression Analysis.* Second edition. Wiley, New York, 709 pp.

Draschba S, Pätzold J, Wefer G (2000) North Atlantic climate variability since AD 1350 recorded in δ^{18}O and skeletal density of Bermuda corals. *International Journal of Earth Sciences* 88(4): 733–741.

Drysdale RN, Zanchetta G, Hellstrom JC, Fallick AE, Zhao J, Isola I, Bruschi G (2004) Palaeoclimatic implications of the growth history and stable isotope (δ^{18}O and δ^{13}C) geochemistry of a middle to late Pleistocene stalagmite from central-western Italy. *Earth and Planetary Science Letters* 227(3–4): 215–229.

Durbin J, Watson GS (1950) Testing for serial correlation in least squares regression I. *Biometrika* 37(3–4): 409–428.

Durbin J, Watson GS (1951) Testing for serial correlation in least squares regression II. *Biometrika* 38(1–2): 159–178.

Durbin J, Watson GS (1971) Testing for serial correlation in least squares regression III. *Biometrika* 58(1): 1–19.

Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO (2000) Climate extremes: Observations, modeling, and impacts. *Science* 289(5487): 2068–2074.

Eastoe EF, Tawn JA (2009) Modelling non-stationary extremes with application to surface level ozone. *Applied Statistics* 58(1): 25–45.

Ebisuzaki W (1997) A method to estimate the statistical significance of a correlation when the data are serially correlated. *Journal of Climate* 10(9): 2147–2153.

Eckmann J-P, Ruelle D (1992) Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. *Physica D* 56(2–3): 185–187.

Edgington ES (1986) Randomization tests. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 7. Wiley, New York, pp 530–538.

Edwards M, Richardson AJ (2004) Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* 430(7002): 881–884.

Efron B (1979) Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7(1): 1–26.

Efron B (1982) *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia, PA, 92 pp.

Efron B (1987) Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82(397): 171–185.

Efron B (1994) Missing data, imputation, and the bootstrap (with discussion). *Journal of the American Statistical Association* 89(426): 463–479.

Efron B, Hinkley DV (1978) Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion). *Biometrika* 65(3): 457–487.

Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statistical Science* 1(1): 54–77.

Efron B, Tibshirani RJ (1993) *An Introduction to the Bootstrap.* Chapman and Hall, London, 436 pp.

Einsele G, Ricken W, Seilacher A (Eds) (1991) *Cycles and Events in Stratigraphy.* Springer, Berlin, 955 pp.

Einstein A (1949) Autobiographisches—Autobiographical notes. In: Schilpp PA (Ed) *Albert Einstein: Philosopher–Scientist.* Library of Living Philosophers, Evanston, IL, pp 1–95.

El-Aroui M-A, Diebolt J (2002) On the use of the peaks over thresholds method for estimating out-of-sample quantiles. *Computational Statistics and Data Analysis* 39(4): 453–475.

Ellis TMR, Philips IR, Lahey TM (1994) *Fortran 90 Programming.* Addison-Wesley, Harlow, 825 pp.

Elsner JB (2006) Evidence in support of the climate change–Atlantic hurricane hypothesis. *Geophysical Research Letters* 33(16): L16705. [doi:10.1029/2006GL026869]

Elsner JB, Kara AB (1999) *Hurricanes of the North Atlantic: Climate and Society.* Oxford University Press, New York, 488 pp.

Elsner JB, Kara AB, Owens MA (1999) Fluctuations in North Atlantic hurricane frequency. *Journal of Climate* 12(2): 427–437.

Elsner JB, Kossin JP, Jagger TH (2008) The increasing intensity of the strongest tropical cyclones. *Nature* 455(7208): 92–95.

Emanuel K (2005) Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* 436(7051): 686–688.

Emanuel KA (1987) The dependence of hurricane intensity on climate. *Nature* 326(6112): 483–485.

Emanuel KA (1999) Thermodynamic control of hurricane intensity. *Nature* 401(6754): 665–669.

Embrechts P, Klüppelberg C, Mikosch T (1997) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, 648 pp.

Emiliani C (1955) Pleistocene temperatures. *Journal of Geology* 63(6): 538–578.

Engel H, Krahé P, Nicodemus U, Heininger P, Pelzer J, Disse M, Wilke K (2002) *Das Augusthochwasser 2002 im Elbegebiet*. Bundesanstalt für Gewässerkunde, Koblenz, 48 pp.

EPICA community members (2004) Eight glacial cycles from an Antarctic ice core. *Nature* 429(6992): 623–628.

Esterby SR, El-Shaarawi AH (1981) Inference about the point of change in a regression model. *Applied Statistics* 30(3): 277–285.

Fairchild IJ, Frisia S, Borsato A, Tooth AF (2007) Speleothems. In: Nash DJ, McLaren SJ (Eds) *Geochemical Sediments and Landscapes*. Blackwell, Malden, MA, pp 200–245.

Fan J, Yao Q (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York, 551 pp.

Fawcett L, Walshaw D (2006) A hierarchical model for extreme wind speeds. *Applied Statistics* 55(5): 631–646.

Fawcett L, Walshaw D (2007) Improved estimation for temporally clustered extremes. *Environmetrics* 18(1–2): 173–188.

Ferraz-Mello S (1981) Estimation of periods from unequally spaced observations. *The Astronomical Journal* 86(4): 619–624.

Ferreira A, de Haan L, Peng L (2003) On optimising the estimation of high quantiles of a probability distribution. *Statistics* 37(5): 401–434.

Ferro CAT, Segers J (2003) Inference for clusters of extreme values. *Journal of the Royal Statistical Society, Series B* 65(2): 545–556.

Fieller EC, Hartley HO, Pearson ES (1957) Tests for rank correlation coefficients I. *Biometrika* 44(3–4): 470–481.

Findley DF (1986) On bootstrap estimates of forecast mean square errors for autoregressive processes. In: Allen DM (Ed) *Computer Science and Statistics*. North-Holland, Amsterdam, pp 11–17.

Fine TL (1983) Foundations of probability. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 3. Wiley, New York, pp 175–184.

Fischer H (1997) *Räumliche Variabilität in Eiskernzeitreihen Nordostgrönlands*. Ph.D. Dissertation. University of Heidelberg, Heidelberg, 188 pp.

Fischer K (1907) Die Sommerhochwasser der Oder von 1813 bis 1903. *Jahrbuch für die Gewässerkunde Norddeutschlands, Besondere Mitteilungen* 1(6): 1–101.

Fisher DA, Reeh N, Clausen HB (1985) Stratigraphic noise in time series derived from ice cores. *Annals of Glaciology* 7(1): 76–83.

Fisher RA (1915) Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10(4): 507–521.

Fisher RA (1921) On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron* 1(4): 3–32.

Fisher RA (1922) On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A* 222: 309–368.

Fisher RA (1929) Tests of significance in harmonic analysis. *Proceedings of the Royal Society of London, Series A* 125(796): 54–59.

Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24(2): 180–190.

Fishman GS (1996) *Monte Carlo: Concepts, Algorithms, and Applications*. Springer, New York, 698 pp.

Fleitmann D (2001) *Annual to millennial Indian Ocean monsoon variability recorded in Holocene and Pleistocene stalagmites from Oman*. Ph.D. Dissertation. University of Bern, Bern, 236 pp.

Fleitmann D, Burns SJ, Mangini A, Mudelsee M, Kramers J, Villa I, Neff U, Al-Subbary AA, Buettner A, Hippler D, Matter A (2007a) Holocene ITCZ and Indian monsoon dynamics recorded in stalagmites from Oman and Yemen (Socotra). *Quaternary Science Reviews* 26(1–2): 170–188.

Fleitmann D, Burns SJ, Mudelsee M, Neff U, Kramers J, Mangini A, Matter A (2003) Holocene forcing of the Indian monsoon recorded in a stalagmite from southern Oman. *Science* 300(5626): 1737–1739.

Fleitmann D, Burns SJ, Neff U, Mudelsee M, Mangini A, Matter A (2004) Paleoclimatic interpretation of high-resolution oxygen isotope profiles derived from annually laminated speleothems from southern Oman. *Quaternary Science Reviews* 23(7–8): 935–945.

Fleitmann D, Cheng H, Badertscher S, Edwards RL, Mudelsee M, Göktürk OM, Fankhauser A, Pickering R, Raible CC, Matter A, Kramers J, Tüysüz O (2009) Timing and climatic impact of Greenland interstadials recorded in stalagmites from northern Turkey. *Geophysical Research Letters* 36(19): L19707. [doi:10.1029/2009GL040050]

Fleitmann D, Dunbar RB, McCulloch M, Mudelsee M, Vuille M, McClanahan TR, Cole JE, Eggins S (2007b) East African soil erosion recorded in a 300 year old coral colony from Kenya. *Geophysical Research Letters* 34(4): L04401. [doi:10.1029/2006GL028525]

Fleitmann D, Mudelsee M, Burns SJ, Bradley RS, Kramers J, Matter A (2008) Evidence for a widespread climatic anomaly at around 9.2 ka before present. *Paleoceanography* 23(1): PA1102. [doi:10.1029/2007PA001519]

Fligge M, Solanki SK, Beer J (1999) Determination of solar cycle length variations using the continuous wavelet transform. *Astronomy and Astrophysics* 346(1): 313–321.

Fodor IK, Stark PB (2000) Multitaper spectrum estimation for time series with gaps. *IEEE Transactions on Signal Processing* 48(12): 3472–3483.

Foias C, Frazho AE, Sherman PJ (1988) A geometric approach to the maximum likelihood spectral estimator for sinusoids in noise. *IEEE Transactions on Information Theory* 34(5): 1066–1070.

Folland CK, Sexton DMH, Karoly DJ, Johnson CE, Rowell DP, Parker DE (1998) Influences of anthropogenic and oceanic forcing on recent climate change. *Geophysical Research Letters* 25(3): 353–356.

Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW, Haywood J, Lean J, Lowe DC, Myhre G, Nganga J, Prinn R, Raga G, Schulz M, Van Dorland R (2007) Changes in atmospheric constituents and in radiative forcing. In: Solomon S, Qin D, Manning M, Marquis M, Averyt K, Tignor MMB, Miller Jr HL, Chen Z (Eds) *Climate Change 2007: The Physical Science Basis. Contribution of*

Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 129–234.

Foster G (1996a) Time series analysis by projection. I. Statistical properties of Fourier analysis. *The Astronomical Journal* 111(1): 541–554.

Foster G (1996b) Wavelets for period analysis of unevenly sampled time series. *The Astronomical Journal* 112(4): 1709–1729.

Foster G, Annan JD, Schmidt GA, Mann ME (2008) Comment on "Heat capacity, time constant, and sensitivity of Earth's climate system" by S. E. Schwartz. *Journal of Geophysical Research* 113(D15): D15102. [doi:10.1029/2007JD009373]

Foutz RV (1980) Estimation of a common group delay between two multiple time series. *Journal of the American Statistical Association* 75(372): 779–788.

Fraedrich K, Blender R (2003) Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Physical Review Letters* 90(10): 108501. [doi:10.1103/PhysRevLett.90.108501]

Fraedrich K, Blender R (2004) Fraedrich and Blender reply. *Physical Review Letters* 92(3): 039802. [doi:10.1103/PhysRevLett.92.039802]

Francisco-Fernández M, Opsomer J, Vilar-Fernández JM (2004) Plug-in bandwidth selector for local polynomial regression estimator with correlated errors. *Nonparametric Statistics* 16(1–2): 127–151.

Francisco-Fernández M, Vilar-Fernández JM (2005) Bandwidth selection for the local polynomial estimator under dependence: A simulation study. *Computational Statistics* 20(4): 539–558.

Frangos CC, Schucany WR (1990) Jackknife estimation of the bootstrap acceleration constant. *Computational Statistics and Data Analysis* 9(3): 271–281.

Franke J, Neumann MH (2000) Bootstrapping neural networks. *Neural Computation* 12(8): 1929–1949.

Franklin LA (1988) A note on approximations and convergence in distribution for Spearman's rank correlation coefficient. *Communications in Statistics—Theory and Methods* 17(1): 55–59.

Fraser AM, Swinney HL (1986) Independent coordinates for strange attractors from mutual information. *Physical Review A* 33(2): 1134–1140.

Fréchet M (1927) Sur la loi probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique* 6: 93–116.

Freedman D (1984) On bootstrapping two-stage least-squares estimates in stationary linear models. *The Annals of Statistics* 12(3): 827–842.

Freedman DA (1981) Bootstrapping regression models. *The Annals of Statistics* 9(6): 1218–1228.

Freedman DA, Peters SC (1984) Bootstrapping an econometric model: Some empirical results. *Journal of Business & Economic Statistics* 2(2): 150–158.

Frei C, Schär C (2001) Detection probability of trends in rare events: Theory and application to heavy precipitation in the Alpine region. *Journal of Climate* 14(7): 1568–1584.

Freund RJ, Minton PD (1979) *Regression Methods: A Tool for Data Analysis*. Marcel Dekker, New York, 261 pp.

Friis-Christensen E, Lassen K (1991) Length of the solar cycle: An indicator of solar activity closely associated with climate. *Science* 254(5032): 698–700.

Fuller WA (1987) *Measurement Error Models*. Wiley, New York, 440 pp.

Fuller WA (1996) *Introduction to Statistical Time Series*. Second edition. Wiley, New York, 698 pp.

Fuller WA (1999) Errors in variables. In: Kotz S, Read CB, Banks DL (Eds) *Encyclopedia of statistical sciences*, volume U3. Wiley, New York, pp 213–216.

Galambos J (1978) *The Asymptotic Theory of Extreme Order Statistics.* Wiley, New York, 352 pp.

Gallant AR (1987) *Nonlinear Statistical Models.* Wiley, New York, 610 pp.

Galton F (1888) Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London* 45(245): 135–145.

Gardenier JS, Gardenier TK (1988) Statistics of risk management. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 8. Wiley, New York, pp 141–148.

Gasser T, Kneip A, Köhler W (1991) A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association* 86(415): 643–652.

Gasser T, Müller H-G (1979) Kernel estimation of regression functions. In: Gasser T, Rosenblatt M (Eds) *Smoothing Techniques for Curve Estimation.* Springer, Berlin, pp 23–68.

Gasser T, Müller H-G (1984) Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics* 11(3): 171–185.

Gayen AK (1951) The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika* 38(1–2): 219–247.

Gençay R, Selçuk F, Ulugülyağci A (2001) EVIM: A software package for extreme value analysis in MATLAB. *Studies in Nonlinear Dynamics & Econometrics* 5(3): 213–239.

Gentle JE (1998) *Numerical Linear Algebra for Applications in Statistics.* Springer, New York, 221 pp.

Genton MG, Hall P (2007) Statistical inference for evolving periodic functions. *Journal of the Royal Statistical Society, Series B* 69(4): 643–657.

Geyh MA, Schleicher H (1990) *Absolute Age Determination: Physical and Chemical Dating Methods and Their Application.* Springer, Berlin, 503 pp.

Ghil M, Allen MR, Dettinger MD, Ide K, Kondrashov D, Mann ME, Robertson AW, Saunders A, Tian Y, Varadi F, Yiou P (2002) Advanced spectral methods for climatic time series. *Reviews of Geophysics* 40(1): 1003. [doi:10.1029/2000RG000092]

Giaiotti D, Stel F (2001) A comparison between subjective and objective thunderstorm forecasts. *Atmospheric Research* 56(1–4): 111–126.

Gibbons JD, Chakraborti S (2003) *Nonparametric Statistical Inference.* Fourth edition. Marcel Dekker, New York, 645 pp.

Giese H-J, Albeverio S, Stabile G (1999) Stochastic and deterministic methods in the analysis of the δ^{18}O record in the core V28-239. *Chemical Geology* 161(1–3): 271–289.

Gijbels I, Goderniaux A-C (2004a) Bandwidth selection for changepoint estimation in nonparametric regression. *Technometrics* 46(1): 76–86.

Gijbels I, Goderniaux A-C (2004b) Bootstrap test for change-points in nonparametric regression. *Nonparametric Statistics* 16(3–4): 591–611.

Gijbels I, Hall P, Kneip A (2004) Interval and band estimation for curves with jumps. *Journal of Applied Probability* 41A: 65–79.

Gil-Alana LA (2008) Time trend estimation with breaks in temperature time series. *Climatic Change* 89(3–4): 325–337.

Gillieson D (1996) *Caves: Processes, Development and Management.* Blackwell, Oxford, 324 pp.

Gilman DL, Fuglister FJ, Mitchell Jr JM (1963) On the power spectrum of "red noise." *Journal of the Atmospheric Sciences* 20(2): 182–184.

Giordano F, La Rocca M, Perna C (2005) Neural network sieve bootstrap prediction intervals for hydrological time series. *Geophysical Research Abstracts* 7: 02801.

Giordano F, La Rocca M, Perna C (2007) Forecasting nonlinear time series with neural network sieve bootstrap. *Computational Statistics and Data Analysis* 51(8): 3871–3884.

Girardin M-P, Tardif JC, Flannigan MD, Bergeron Y (2006a) Synoptic-scale atmospheric circulation and boreal Canada summer drought variability of the past three centuries. *Journal of Climate* 19(10): 1922–1947.

Girardin MP, Ali AA, Carcaillet C, Mudelsee M, Drobyshev I, Hély C, Bergeron Y (2009) Heterogeneous response of circumboreal wildfire risk to climate change since the early 1900s. *Global Change Biology* 15(11): 2751–2769.

Girardin MP, Bergeron Y, Tardif JC, Gauthier S, Flannigan MD, Mudelsee M (2006b) A 229-year dendroclimatic-inferred record of forest fire activity for the Boreal Shield of Canada. *International Journal of Wildland Fire* 15(3): 375–388.

Girardin MP, Mudelsee M (2008) Past and future changes in Canadian boreal wildfire activity. *Ecological Applications* 18(2): 391–406.

Glaser R (2001) *Klimageschichte Mitteleuropas.* Wissenschaftliche Buchgesellschaft, Darmstadt, 227 pp.

Gleissberg W (1944) A table of secular variations of the solar cycle. *Terrestrial Magnetism and Atmospheric Electricity* 49(4): 243–244.

Gleissberg W (1965) The eighty-year solar cycle in auroral frequency numbers. *Journal of the British Astronomical Association* 75(4): 227–231.

Gluhovsky A, Agee E (1994) A definitive approach to turbulence statistical studies in planetary boundary layers. *Journal of the Atmospheric Sciences* 51(12): 1682–1690.

Glymour C (1998) Causation (update). In: Kotz S, Read CB, Banks DL (Eds) *Encyclopedia of statistical sciences*, volume U2. Wiley, New York, pp 97–109.

Gnedenko B (1943) Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* 44(3): 423–453. [English translation in: Kotz S, Johnson NL (Eds) (1992) *Breakthroughs in Statistics*, volume 1. Springer, New York, pp 195–225]

Goel AL (1982) Cumulative sum control charts. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 2. Wiley, New York, pp 233–241.

Goldenberg SB, Landsea CW, Mestas-Nuñez AM, Gray WM (2001) The recent increase in Atlantic hurricane activity: Causes and implications. *Science* 293(5529): 474–479.

Goldstein RB (1973) Algorithm 451: Chi-square quantiles. *Communications of the ACM* 16(8): 483–485.

Good PI (2005) *Resampling Methods: A Practical Guide to Data Analysis.* Third edition. Birkhäuser, Boston, 218 pp.

Goodess CM, Jacob D, Déqué M, Guttiérrez JM, Huth R, Kendon E, Leckebusch GC, Lorenz P, Pavan V (2009) Downscaling methods, data and tools for input to impacts assessments. In: van der Linden P, Mitchell JFB (Eds) *ENSEMBLES: Climate change and its impacts at seasonal, decadal and centennial timescales.* Met Office Hadley Centre, Exeter, pp 59–78.

Goodman LA (1953) A simple method for improving some estimators. *Annals of Mathematical Statistics* 24(1): 114–117.

Goossens C, Berger A (1986) Annual and seasonal climatic variations over the northern hemisphere and Europe during the last century. *Annales Geophysicae, Series B* 4(4): 385–399.

Gordon C, Cooper C, Senior CA, Banks H, Gregory JM, Johns TC, Mitchell JFB, Wood RA (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics* 16(2–3): 147–168.

Goreau TJ (1980) Frequency sensitivity of the deep-sea climatic record. *Nature* 287(5783): 620–622.

Gosse JC, Phillips FM (2001) Terrestrial in situ cosmogenic nuclides: Theory and application. *Quaternary Science Reviews* 20(14): 1475–1560.

Götze F, Künsch HR (1996) Second-order correctness of the blockwise bootstrap for stationary observations. *The Annals of Statistics* 24(5): 1914–1933.

Govindan RB, Vyushin D, Bunde A, Brenner S, Havlin S, Schellnhuber H-J (2002) Global climate models violate scaling of the observed atmospheric variability. *Physical Review Letters* 89(2): 028501. [doi:10.1103/PhysRevLett.89.028501]

Gradshteyn IS, Ryzhik IM (2000) *Tables of Integrals, Series, and Products.* Sixth edition. Academic Press, San Diego, 1163 pp.

Gradstein FM, Ogg JG, Smith AG (Eds) (2004) *A Geologic Time Scale 2004.* Cambridge University Press, Cambridge, 589 pp.

Granger C, Lin J-L (1994) Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis* 15(4): 371–384.

Granger CW, Maasoumi E, Racine J (2004) A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* 25(5): 649–669.

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37(3): 424–438.

Granger CWJ (1980) Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14(2): 227–238.

Granger CWJ, Hyung N (2004) Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* 11(3): 399–421.

Granger CWJ, Joyeux R (1980) An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1(1): 15–29.

Grassberger P (1986) Do climatic attractors exist? *Nature* 323(6089): 609–612.

Graybill FA, Iyer HK (1994) *Regression Analysis: Concepts and Applications.* Duxbury Press, Belmont, CA, 701 pp.

Greenwood JA, Landwehr JM, Matalas NC, Wallis JR (1979) Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research* 15(5): 1049–1054.

Gregory JM, Stouffer RJ, Raper SCB, Stott PA, Rayner NA (2002) An observationally based estimate of the climate sensitivity. *Journal of Climate* 15(22): 3117–3121.

Grenander U (1958) Bandwidth and variance in estimation of the spectrum. *Journal of the Royal Statistical Society, Series B* 20(1): 152–157.

Grenander U, Rosenblatt M (1956) *Statistical Analysis of Stationary Time Series.* Almqvist & Wiksell, Stockholm, 300 pp.

Grieger B (1992) *Orbital tuning of marine sedimentary cores: An automatic procedure based on a general linear model.* Max Planck Institute for Meteorology, Hamburg, 30 pp. [Report No. 79]

Grieger B, Latif M (1994) Reconstruction of the El Niño attractor with neural networks. *Climate Dynamics* 10(6–7): 267–276.

Grün R (1989) *Die ESR-Altersbestimmungsmethode.* Springer, Berlin, 132 pp.

Grünewald U, Chmielewski R, Kaltofen M, Rolland W, Schümberg S, Ahlheim M, Sauer T, Wagner R, Schluchter W, Birkner H, Petzold R, Radczuk L, Eliasiewicz R, Bjarsch B, Paus L, Zahn G (1998) *Ursachen, Verlauf und Folgen des Sommer-Hochwassers 1997 an der Oder sowie Aussagen zu bestehenden Risikopotentialen. Eine interdisziplinäre Studie — Langfassung.* Deutsches IDNDR-Komitee für Katastrophenvorbeugung e.V., Bonn, 187 pp.

Grünewald U, Kaltofen M, Schümberg S, Merz B, Kreibich H, Petrow T, Thieken A, Streitz W, Dombrowsky WR (2003) *Hochwasservorsorge in Deutschland: Lernen aus der Katastrophe 2002 im Elbegebiet.* Deutsches Komitee für Katastrophenvorsorge, Bonn, 144 pp. [Schriftenreihe des DKKV 29]

Grunwald GK, Hyndman RJ (1998) Smoothing non-Gaussian time series with autoregressive structure. *Computational Statistics and Data Analysis* 28(2): 171–191.

Guiot J, Tessier L (1997) Detection of pollution signals in tree-ring series using AR processes and neural networks. In: Subba Rao T, Priestley MB, Lessi O (Eds) *Applications of Time Series Analysis in Astronomy and Meteorology.* Chapman and Hall, London, pp 413–426.

Gumbel EJ (1958) *Statistics of Extremes.* Columbia University Press, New York, 375 pp.

Haam E, Huybers P (2010) A test for the presence of covariance between time-uncertain series of data with application to the Dongge cave speleothem and atmospheric radiocarbon records. *Paleoceanography* 25(2): PA2209. [doi:10.1029/2008PA001713]

Hagelberg T, Pisias N, Elgar S (1991) Linear and nonlinear couplings between orbital forcing and the marine δ^{18}O record during the late Neogene. *Paleoceanography* 6(6): 729–746.

Haldane JBS (1942) Moments of the distributions of powers and products of normal variates. *Biometrika* 32(3–4): 226–242.

Hall P (1985) Resampling a coverage pattern. *Stochastic Processes and their Applications* 20(2): 231–246.

Hall P (1986) On the bootstrap and confidence intervals. *The Annals of Statistics* 14(4): 1431–1452.

Hall P (1988) Theoretical comparison of bootstrap confidence intervals (with discussion). *The Annals of Statistics* 16(3): 927–985.

Hall P (1992) On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics* 20(2): 695–711.

Hall P, Horowitz JL, Jing B-Y (1995a) On blocking rules for the bootstrap with dependent data. *Biometrika* 82(3): 561–574.

Hall P, Jing B-Y, Lahiri SN (1998) On the sampling window method for long-range dependent data. *Statistica Sinica* 8(4): 1189–1204.

Hall P, Lahiri SN, Polzehl J (1995b) On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. *The Annals of Statistics* 23(6): 1921–1936.

Hall P, Ma Y (2007) Testing the suitability of polynomial models in errors-in-variables problems. *The Annals of Statistics* 35(6): 2620–2638.

Hall P, Martin MA (1988) On bootstrap resampling and iteration. *Biometrika* 75(4): 661–671.

Hall P, Martin MA (1996) Comment on "Bootstrap confidence intervals" by DiCiccio TJ and Efron B. *Statistical Science* 11(3): 212–214.

Hall P, Martin MA, Schucany WR (1989) Better nonparametric bootstrap confidence intervals for the correlation coefficient. *Journal of Statistical Computation and Simulation* 33(3): 161–172.

Hall P, Peng L, Tajvidi N (2002) Effect of extrapolation on coverage accuracy of prediction intervals computed from Pareto-type data. *The Annals of Statistics* 30(3): 875–895.

Hall P, Tajvidi N (2000) Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science* 15(2): 153–167.

Hall P, Titterington DM (1988) On confidence bands in nonparametric density estimation and regression. *Journal of Multivariate Analysis* 27(1): 228–254.

Hall P, Turlach BA (1997) Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *The Annals of Statistics* 25(5): 1912–1925.

Hall P, Weissman I (1997) On the estimation of extreme tail probabilities. *The Annals of Statistics* 25(3): 1311–1326.

Hall P, Wilson SR (1991) Two guidelines for bootstrap hypothesis testing. *Biometrics* 47(2): 757–762.

Hamed KH (2008) Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. *Journal of Hydrology* 349(3–4): 350–363.

Hamed KH (2009a) Effect of persistence on the significance of Kendall's tau as a measure of correlation between natural time series. *European Physical Journal Special Topics* 174(1): 65–79.

Hamed KH (2009b) Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. *Journal of Hydrology* 368(1–4): 143–155.

Hammer C, Mayewski PA, Peel D, Stuiver M (Eds) (1997) *Greenland Summit Ice Cores GISP2/GRIP*, volume 102(C12) of *Journal of Geophysical Research*. [Special issue]

Hamon BV, Hannan EJ (1974) Spectral estimation of time delay for dispersive and non-dispersive systems. *Applied Statistics* 23(2): 134–142.

Hampel FR (1985) The breakdown points of the mean combined with some rejection rules. *Technometrics* 27(2): 95–107.

Hann J (1901) *Lehrbuch der Meteorologie*. Tauchnitz, Leipzig, 805 pp.

Hannan EJ (1960) *Time Series Analysis*. Methuen, London, 152 pp.

Hannan EJ (1961) Testing for a jump in the spectral function. *Journal of the Royal Statistical Society, Series B* 23(2): 394–404.

Hannan EJ, Quinn BG (1989) The resolution of closely adjacent spectral lines. *Journal of Time Series Analysis* 10(1): 13–31.

Hannan EJ, Robinson PM (1973) Lagged regression with unknown lags. *Journal of the Royal Statistical Society, Series B* 35(2): 252–267.

Hannan EJ, Thomson PJ (1974) Estimating echo times. *Technometrics* 16(1): 77–84.

Hansen AR, Sutera A (1986) On the probability density distribution of planetary-scale atmospheric wave amplitude. *Journal of the Atmospheric Sciences* 43(24): 3250–3265.

Hansen JE, Lacis AA (1990) Sun and dust versus greenhouse gases: An assessment of their relative roles in global climate change. *Nature* 346(6286): 713–719.

Hardin JW, Schmiediche H, Carroll RJ (2003) The regression–calibration method for fitting generalized linear models with additive measurement error. *The Stata Journal* 3(4): 361–372.

Härdle W (1990) *Applied nonparametric regression*. Cambridge University Press, Cambridge, 333 pp.

Härdle W, Bowman AW (1988) Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association* 83(401): 102–110.

Härdle W, Chen R (1995) Nonparametric time series analysis, a selective review with examples. *Bulletin of the International Statistical Institute* 56(1): 375–394.

Härdle W, Marron JS (1991) Bootstrap simultaneous error bars for nonparametric regression. *The Annals of Statistics* 19(2): 778–796.

Härdle W, Steiger W (1995) Optimal median smoothing. *Applied Statistics* 44(2): 258–264.

Hare FK (1979) Climatic variation and variability: Empirical evidence from meteorological and other sources. In: Secretariat of the World Meteorological Organization (Ed) *Proceedings of the World Climate Conference*. World Meteorological Organization, Geneva, pp 51–87. [WMO Publication No. 537]

Hargreaves JC, Annan JD (2002) Assimilation of paleo-data in a simple Earth system model. *Climate Dynamics* 19(5–6): 371–381.

Harris FJ (1978) On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE* 66(1): 51–83.

Harrison RG, Stephenson DB (2006) Empirical evidence for a nonlinear effect of galactic cosmic rays on clouds. *Proceedings of the Royal Society of London, Series A* 462(2068): 1221–1233.

Hartley HO (1949) Tests of significance in harmonic analysis. *Biometrika* 36(1–2): 194–201.

Haslett J, Parnell A (2008) A simple monotone process with application to radiocarbon-dated depth chronologies. *Applied Statistics* 57(4): 399–418.

Hasselmann K (1976) Stochastic climate models: Part I. Theory. *Tellus* 28(6): 473–485.

Hasselmann K (1993) Optimal fingerprints for the detection of time-dependent climate change. *Journal of Climate* 6(10): 1957–1971.

Hasselmann K (1997) Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics* 13(9): 601–611.

Hasselmann K (1999) Linear and nonlinear signatures. *Nature* 398(6730): 755–756.

Haug GH, Ganopolski A, Sigman DM, Rosell-Mele A, Swann GEA, Tiedemann R, Jaccard SL, Bollmann J, Maslin MA, Leng MJ, Eglinton G (2005) North Pacific seasonality and the glaciation of North America 2.7 million years ago. *Nature* 433(7028): 821–825.

Haug GH, Sigman DM, Tiedemann R, Pedersen TF, Sarnthein M (1999) Onset of permanent stratification in the subarctic Pacific Ocean. *Nature* 401(6755): 779–782.

Hays JD, Imbrie J, Shackleton NJ (1976) Variations in the Earth's orbit: Pacemaker of the ice ages. *Science* 194(4270): 1121–1132.

Heegaard E, Birks HJB, Telford RJ (2005) Relationships between calibrated ages and depth in stratigraphical sequences: An estimation procedure by mixed-effect regression. *The Holocene* 15(4): 612–618.

Hegerl GC, Crowley TJ, Allen M, Hyde WT, Pollack HN, Smerdon J, Zorita E (2007a) Detection of human influence on a new, validated 1500-year temperature reconstruction. *Journal of Climate* 20(4): 650–666.

Hegerl GC, Crowley TJ, Hyde WT, Frame DJ (2006) Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* 440(7087): 1029–1032.

Hegerl GC, Hasselmann K, Cubasch U, Mitchell JFB, Roeckner E, Voss R, Waszkewitz J (1997) Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Climate Dynamics* 13(9): 613–634.

Hegerl GC, North GR (1997) Comparison of statistically optimal approaches to detecting anthropogenic climate change. *Journal of Climate* 10(5): 1125–1133.

Hegerl GC, von Storch H, Hasselmann K, Santer BD, Cubasch U, Jones PD (1996) Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *Journal of Climate* 9(10): 2281–2306.

Hegerl GC, Zwiers FW, Braconnot P, Gillett NP, Luo Y, Marengo Orsini JA, Nicholls N, Penner JE, Stott PA (2007b) Understanding and attributing climate change. In: Solomon S, Qin D, Manning M, Marquis M, Averyt K, Tignor MMB, Miller Jr HL, Chen Z (Eds) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, pp 663–745.

Heisenberg W (1969) *Der Teil und das Ganze.* Piper, Munich, 334 pp.

Henderson GM (2002) New oceanic proxies for paleoclimate. *Earth and Planetary Science Letters* 203(1): 1–13.

Henze FH-H (1979) The exact noncentral distributions of Spearman's r and other related correlation coefficients. *Journal of the American Statistical Association* 74(366): 459–464. [Corrigendum: 1980 Vol. 75(371): 765]

Herrmann E (1997) Local bandwidth choice in kernel regression estimation. *Journal of Computational and Graphical Statistics* 6(1): 35–54.

Herterich K, Sarnthein M (1984) Brunhes time scale: Tuning by rates of calcium-carbonate dissolution and cross spectral analyses with solar insolation. In: Berger A, Imbrie J, Hays J, Kukla G, Saltzman B (Eds) *Milankovitch and Climate*, volume 1. D. Reidel, Dordrecht, pp 447–466.

Heslop D, Dekkers MJ (2002) Spectral analysis of unevenly spaced climatic time series using CLEAN: Signal recovery and derivation of significance levels using a Monte Carlo simulation. *Physics of the Earth and Planetary Interiors* 130(1–2): 103–116.

Hewa GA, Wang QJ, McMahon TA, Nathan RJ, Peel MC (2007) Generalized extreme value distribution fitted by LH moments for low-flow frequency analysis. *Water Resources Research* 43(6): W06301. [doi:10.1029/2006WR004913]

Hidalgo J (2003) An alternative bootstrap to moving blocks for time series regression models. *Journal of Econometrics* 117(2): 369–399.

Hill BM (1975) A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* 3(5): 1163–1174.

Hinkley DV (1970) Inference about the change-point in a sequence of random variables. *Biometrika* 57(1): 1–17.

Hinkley DV (1971) Inference about the change-point from cumulative sum tests. *Biometrika* 58(3): 509–523.

Hinkley DV (1988) Bootstrap methods. *Journal of the Royal Statistical Society, Series B* 50(3): 321–337.

Hinnov LA, Schulz M, Yiou P (2002) Interhemispheric space–time attributes of the Dansgaard–Oeschger oscillations between 100 and 0 ka. *Quaternary Science Reviews* 21(10): 1213–1228.

Hlaváčková-Schindler K, Paluš M, Vejmelka M, Bhattacharya J (2007) Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports* 441(1): 1–46.

Hocking RR, Smith WB (1968) Estimation of parameters in the multivariate normal distribution with missing observations. *Journal of the American Statistical Association* 63(321): 159–173.

Holland GJ (2007) Misuse of landfall as a proxy for Atlantic tropical cyclone activity. *Eos, Transactions of the American Geophysical Union* 88(36): 349–350.

Holton JR, Curry JA, Pyle JA (Eds) (2003) *Encyclopedia of Atmospheric Sciences*, volume 1–6. Academic Press, Amsterdam, 2780 pp.

Holtzman WH (1950) The unbiased estimate of the population variance and standard deviation. *American Journal of Psychology* 63(4): 615–617.

Holzkämper S, Mangini A, Spötl C, Mudelsee M (2004) Timing and progression of the last interglacial derived from a high Alpine stalagmite. *Geophysical Research Letters* 31(7): L07201. [doi:10.1029/2003GL019112]

Hopley PJ, Weedon GP, Marshall JD, Herries AIR, Latham AG, Kuykendall KL (2007) High- and low-latitude orbital forcing of early hominin habitats in South Africa. *Earth and Planetary Science Letters* 256(3–4): 419–432.

Horne JH, Baliunas SL (1986) A prescription for period analysis of unevenly sampled time series. *The Astrophysical Journal* 302(2): 757–763.

Hornstein C (1871) Über die Abhängigkeit des Erdmagnetismus von der Rotation der Sonne. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Classe, Zweite Abtheilung* 64(6): 62–74.

Horowitz LL (1974) The effects of spline interpolation on power spectral density. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 22(1): 22–27.

Hosking JRM (1981) Fractional differencing. *Biometrika* 68(1): 165–176.

Hosking JRM (1984) Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research* 20(12): 1898–1908.

Hosking JRM (1985) Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Applied Statistics* 34(3): 301–310.

Hosking JRM (1990) L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B* 52(1): 105–124.

Hosking JRM, Wallis JR (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29(3): 339–349.

Hosking JRM, Wallis JR (1997) *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge, 224 pp.

Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the generalized extreme value distribution by the method of probability-weighted moments. *Technometrics* 27(3): 251–261.

Hotelling H (1953) New light on the correlation coefficient and its transforms (with discussion). *Journal of the Royal Statistical Society, Series B* 15(2): 193–232.

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (Eds) (2001) *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 881 pp.

Houseman EA (2005) A robust regression model for a first-order autoregressive time series with unequal spacing: Application to water monitoring. *Applied Statistics* 54(4): 769–780.

Hoyt DV, Schatten KH (1997) *The Role of the Sun in Climate Change*. Oxford University Press, New York, 279 pp.

Hoyt DV, Schatten KH (1998) Group sunspot numbers: A new solar activity reconstruction. *Solar Physics* 179(1): 189–219. [Corrigendum: 1998 Vol. 181(2): 491–512]

Hsieh WW, Tang B (1998) Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society* 79(9): 1855–1870.

Hsu DA (1977) Tests for variance shift at an unknown time point. *Applied Statistics* 26(3): 279–284.

Huber PJ (1964) Robust estimation of location parameter. *Annals of Mathematical Statistics* 35(1): 73–101.

Huber PJ (1981) *Robust Statistics*. Wiley, New York, 308 pp.

Hudson DJ (1966) Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association* 61(316): 1097–1129.

Huet S, Bouvier A, Poursat M-A, Jolivet E (2004) *Statistical Tools for Nonlinear Regression: A Practical Guide With S-PLUS and R Examples*. Second edition. Springer, New York, 232 pp.

Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science* 269(5524): 676–679.

Hurst HE (1951) Long-term storage capacity of reservoirs (with discussion). *Transactions of the American Society of Civil Engineers* 116: 770–808.

Hurvich CM, Tsai C-L (1989) Regression and time series model selection in small samples. *Biometrika* 76(2): 297–307.

Huybers P (2002) *Depth and Orbital Tuning: A New Chronology of Glaciation and Nonlinear Orbital Climate Change*. M.Sc. Thesis. Massachusetts Institute of Technology, Cambridge, MA, 119 pp.

Huybers P, Denton G (2008) Antarctic temperature at orbital timescales controlled by local summer duration. *Nature Geoscience* 1(11): 787–792.

Huybers P, Wunsch C (2004) A depth-derived Pleistocene age model: Uncertainty estimates, sedimentation variability, and nonlinear climate change. *Paleoceanography* 19(1): PA1028. [doi:10.1029/2002PA000857]

Huybers P, Wunsch C (2005) Obliquity pacing of the late Pleistocene glacial terminations. *Nature* 434(7032): 491–494.

Hwang S (2000) The effects of systematic sampling and temporal aggregation on discrete time long memory processes and their finite sample properties. *Econometric Theory* 16(3): 347–372.

Imbrie J, Berger A, Boyle EA, Clemens SC, Duffy A, Howard WR, Kukla G, Kutzbach J, Martinson DG, McIntyre A, Mix AC, Molfino B, Morley JJ, Peterson LC, Pisias NG, Prell WL, Raymo ME, Shackleton NJ, Toggweiler JR (1993) On the structure and origin of major glaciation cycles 2. The 100,000-year cycle. *Paleoceanography* 8(6): 699–735.

Imbrie J, Boyle EA, Clemens SC, Duffy A, Howard WR, Kukla G, Kutzbach J, Martinson DG, McIntyre A, Mix AC, Molfino B, Morley JJ, Peterson LC, Pisias NG, Prell WL, Raymo ME, Shackleton NJ, Toggweiler JR (1992) On the structure and origin of major glaciation cycles 1. Linear responses to Milankovitch forcing. *Paleoceanography* 7(6): 701–738.

Imbrie J, Hays JD, Martinson DG, McIntyre A, Mix AC, Morley JJ, Pisias NG, Prell WL, Shackleton NJ (1984) The orbital theory of Pleistocene climate: Support from a revised chronology of the marine δ^{18}O record. In: Berger A, Imbrie J, Hays J, Kukla G, Saltzman B (Eds) *Milankovitch and Climate*, volume 1. D. Reidel, Dordrecht, pp 269–305.

Inclán C, Tiao GC (1994) Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* 89(427): 913–923.

Ivanovich M, Harmon RS (Eds) (1992) *Uranium-series Disequilibrium: Applications to Earth, Marine, and Environmental Sciences*. Second edition. Clarendon Press, Oxford, 910 pp.

Jansson M (1985) A comparison of the detransformed logarithmic regressions and power function regressions. *Geografiska Annaler* 67A(1–2): 61–70.

Jarrett RF (1968) A minor exercise in history. *The American Statistician* 22(3): 25–26.

Jefferys WH (1980) On the method of least squares. *The Astronomical Journal* 85(2): 177–181. [Corrigendum: 1988 Vol. 95(4): 1299]

Jefferys WH (1981) On the method of least squares. II. *The Astronomical Journal* 86(1): 149–155. [Corrigendum: 1988 Vol. 95(4): 1300]

Jenkins GM, Watts DG (1968) *Spectral Analysis and its Applications*. Holden-Day, San Francisco, 525 pp.

Jenkinson AF (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81(348): 158–171.

Jennen-Steinmetz C, Gasser T (1988) A unifying approach to nonparametric regression estimation. *Journal of the American Statistical Association* 83(404): 1084–1089.

Jiménez-Moreno G, Anderson RS, Fawcett PJ (2007) Orbital- and millennial-scale vegetation and climate changes of the past 225 ka from Bear Lake, Utah–Idaho (USA). *Quaternary Science Reviews* 26(13–14): 1713–1724.

Johns TC, Carnell RE, Crossley JF, Gregory JM, Mitchell JFB, Senior CA, Tett SFB, Wood RA (1997) The second Hadley Centre coupled ocean–atmosphere GCM: Model description, spinup and validation. *Climate Dynamics* 13(2): 103–134.

Johnsen SJ, Dahl-Jensen D, Gundestrup N, Steffensen JP, Clausen HB, Miller H, Masson-Delmotte V, Sveinbjörnsdottir AE, White J (2001) Qxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP. *Journal of Quaternary Science* 16(4): 299–307.

Johnson NL, Kotz S, Balakrishnan N (1994) *Continuous Univariate Distributions*, volume 1. Second edition. Wiley, New York, 756 pp.

Johnson NL, Kotz S, Balakrishnan N (1995) *Continuous Univariate Distributions*, volume 2. Second edition. Wiley, New York, 719 pp.

Johnson NL, Kotz S, Kemp AW (1993) *Univariate Discrete Distributions*. Second edition. Wiley, New York, 565 pp.

Johnson RG (1982) Brunhes–Matuyama magnetic reversal dated at 790,000 yr B.P. by marine–astronomical correlations. *Quaternary Research* 17(2): 135–147.

Jones MC, Lotwick HW (1984) A remark on algorithm AS 176. Kernel density estimation using the Fast Fourier Transform. *Applied Statistics* 33(1): 120–122.

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

Jones PD, Raper SCB, Bradley RS, Diaz HF, Kelly PM, Wigley TML (1986) Northern hemisphere surface air temperature variations: 1851–1984. *Journal of Climate and Applied Meteorology* 25(2): 161–179.

Jones RH (1981) Fitting a continuous time autoregression to discrete data. In: Findley DF (Ed) *Applied Time Series Analysis II*. Academic Press, New York, pp 651–682.

Jones RH (1985) Time series analysis with unequally spaced data. In: Hannan EJ, Krishnaiah PR, Rao MM (Eds) *Handbook of Statistics*, volume 5. Elsevier, Amsterdam, pp 157–177.

Jones RH (1986) Time series regression with unequally spaced data. *Journal of Applied Probability* 23A: 89–98. [Special volume]

Jones RH, Tryon PV (1987) Continuous time series models for unequally spaced data applied to modeling atomic clocks. *SIAM Journal on Scientific and Statistical Computing* 8(1): 71–81.

Jones TA (1979) Fitting straight lines when both variables are subject to error. I. Maximum likelihood and least-squares estimation. *Mathematical Geology* 11(1): 1–25.

Jouzel J, Masson-Delmotte V, Cattani O, Dreyfus G, Falourd S, Hoffmann G, Minster B, Nouet J, Barnola JM, Chappellaz J, Fischer H, Gallet JC, Johnsen S, Leuenberger M, Loulergue L, Luethi D, Oerter H, Parrenin F, Raisbeck G, Raynaud D, Schilt A, Schwander J, Selmo E, Souchez R, Spahni R, Stauffer B, Steffensen JP, Stenni B, Stocker TF, Tison JL, Werner M, Wolff EW (2007) Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* 317(5839): 793–796.

Julious SA (2001) Inference and estimation in a changepoint regression problem. *The Statistician* 50(1): 51–61.

Jun M, Knutti R, Nychka DW (2008) Spatial analysis to quantify numerical model bias and dependence: How many climate models are there? *Journal of the American Statistical Association* 103(483): 934–947.

Kahl JD, Charlevoix DJ, Zaitseva NA, Schnell RC, Serreze MC (1993) Absence of evidence for greenhouse warming over the Arctic Ocean in the past 40 years. *Nature* 361(6410): 335–337.

Kallache M (2007) *Trends and Extreme Values of River Discharge Time Series*. Ph.D. Dissertation. University of Bayreuth, Bayreuth, 125 pp.

Kallache M, Rust HW, Kropp J (2005) Trend assessment: Applications for hydrology and climate research. *Nonlinear Processes in Geophysics* 12(2): 201–210.

Kandel ER (2006) *In Search of Memory: The Emergence of a New Science of Mind*. W. W. Norton, New York, 510 pp.

Kant I (1781) *Critik der Reinen Vernunft*. Hartknoch, Riga, 856 pp.

Kantz H, Schreiber T (1997) *Nonlinear time series analysis*. Cambridge University Press, Cambridge, 304 pp.

Karl TR, Knight RW, Plummer N (1995) Trends in high-frequency climate variability in the twentieth century. *Nature* 377(6546): 217–220.

Karl TR, Riebsame WE (1984) The identification of 10- to 20-year temperature and precipitation fluctuations in the contiguous United States. *Journal of Climate and Applied Meteorology* 23(6): 950–966.

Karl TR, Williams Jr CN (1987) An approach to adjusting climatological time series for discontinuous inhomogeneities. *Journal of Climate and Applied Meteorology* 26(12): 1744–1763.

Kärner O (2002) On nonstationarity and antipersistency in global temperature series. *Journal of Geophysical Research* 107(D20): 4415. [doi:10.1029/2001JD002024]

Karr AF (1986) *Point Processes and Their Statistical Inference*. Marcel Dekker, New York, 490 pp.

Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. *Advances in Water Resources* 25(8–12): 1287–1304.

Kaufmann RK, Stern DI (1997) Evidence for human influence on climate from hemispheric temperature relations. *Nature* 388(6637): 39–44.

Kawamura K, Parrenin F, Lisiecki L, Uemura R, Vimeux F, Severinghaus JP, Hutterli MA, Nakazawa T, Aoki S, Jouzel J, Raymo ME, Matsumoto K, Nakata H, Motoyama H, Fujita S, Goto-Azuma K, Fujii Y, Watanabe O (2007) Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* 448(7156): 912–916.

Kay SM, Marple Jr SL (1981) Spectrum analysis—a modern perspective. *Proceedings of the IEEE* 69(11): 1380–1419.

Keigwin LD (1996) The Little Ice Age and Medieval Warm Period in the Sargasso Sea. *Science* 274(5292): 1504–1508.

Kendall M, Gibbons JD (1990) *Rank Correlation Methods*. Fifth edition. Edward Arnold, London, 260 pp.

Kendall MG (1938) A new measure of rank correlation. *Biometrika* 30(1–2): 81–93.

Kendall MG (1954) Note on bias in the estimation of autocorrelation. *Biometrika* 41(3–4): 403–404.

Kennett JP (1982) *Marine Geology*. Prentice-Hall, Englewood Cliffs, NJ, 813 pp.

Kernthaler SC, Toumi R, Haigh JD (1999) Some doubts concerning a link between cosmic ray fluxes and global cloudiness. *Geophysical Research Letters* 26(7): 863–865.

Khaliq MN, Ouarda TBMJ, Gachon P, Sushama L (2008) Temporal evolution of low-flow regimes in Canadian rivers. *Water Resources Research* 44(8): W08436. [doi:10.1029/2007WR006132]

Khaliq MN, Ouarda TBMJ, Ondo J-C, Gachon P, Bobée B (2006) Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology* 329(3–4): 534–552.

Khaliq MN, St-Hilaire A, Ouarda TBMJ, Bobée B (2005) Frequency analysis and temporal pattern of occurrences of southern Quebec heatwaves. *International Journal of Climatology* 25(4): 485–504.

Kharin VV, Zwiers FW (2005) Estimating extremes in transient climate change simulations. *Journal of Climate* 18(8): 1156–1173.

Kiktev D, Sexton DMH, Alexander L, Folland CK (2003) Comparison of modeled and observed trends in indices of daily climate extremes. *Journal of Climate* 16(22): 3560–3571.

King T (1996) Quantifying nonlinearity and geometry in time series of climate. *Quaternary Science Reviews* 15(4): 247–266.

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

Klemeš V (1978) Physically based stochastic hydrologic analysis. *Advances in Hydroscience* 11: 285–356.

Knuth DE (2001) *The Art of Computer Programming*, volume 2. Third edition. Addison-Wesley, Boston, 762 pp.

Knutson TR, McBride JL, Chan J, Emanuel K, Holland G, Landsea C, Held I, Kossin JP, Srivastava AK, Sugi M (2010) Tropical cyclones and climate change. *Nature Geoscience* 3(3): 157–163.

Knutti R, Krähenmann S, Frame DJ, Allen MR (2008) Comment on "Heat capacity, time constant, and sensitivity of Earth's climate system" by S. E. Schwartz. *Journal of Geophysical Research* 113(D15): D15103. [doi:10.1029/2007JD009473]

Kodera K (2004) Solar influence on the Indian Ocean monsoon through dynamical processes. *Geophysical Research Letters* 31(24): L24209. [doi:10.1029/2004GL020928]

Koen C, Lombard F (1993) The analysis of indexed astronomical time series — I. Basic methods. *Monthly Notices of the Royal Astronomical Society* 263(2): 287–308.

Koenker R, Bassett Jr G (1978) Regression quantiles. *Econometrica* 46(1): 33–50.

Koenker R, Hallock KF (2001) Quantile regression. *Journal of Economic Perspectives* 15(4): 143–156.

Kolmogoroff A (1933) Grundbegriffe der Wahrscheinlichkeitsrechnung. *Ergebnisse der Mathematik und ihrer Grenzgebiete* 2(3): 195–262.

Köppen W (1923) *Die Klimate der Erde: Grundriss der Klimakunde.* de Gruyter, Berlin, 369 pp.

Koscielny-Bunde E, Bunde A, Havlin S, Goldreich Y (1996) Analysis of daily temperature fluctuations. *Physica A* 231(4): 393–396.

Koscielny-Bunde E, Bunde A, Havlin S, Roman HE, Goldreich Y, Schellnhuber H-J (1998a) Indication of a universal persistence law governing atmospheric variability. *Physical Review Letters* 81(3): 729–732.

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

Koscielny-Bunde E, Roman HE, Bunde A, Havlin S, Schellnhuber H-J (1998b) Long-range power-law correlations in local daily temperature fluctuations. *Philosophical Magazine B* 77(5): 1331–1340.

Kotz S, Balakrishnan N, Johnson NL (2000) *Continuous Multivariate Distributions*, volume 1. Second edition. Wiley, New York, 722 pp.

Kotz S, Johnson NL, Read CB (Eds) (1982a) *Encyclopedia of statistical sciences*, volume 1. Wiley, New York, 480 pp.

Kotz S, Johnson NL, Read CB (Eds) (1982b) *Encyclopedia of statistical sciences*, volume 2. Wiley, New York, 613 pp.

Kotz S, Johnson NL, Read CB (Eds) (1983a) *Encyclopedia of statistical sciences*, volume 3. Wiley, New York, 722 pp.

Kotz S, Johnson NL, Read CB (Eds) (1983b) *Encyclopedia of statistical sciences*, volume 4. Wiley, New York, 657 pp.

Kotz S, Johnson NL, Read CB (Eds) (1985a) *Encyclopedia of statistical sciences*, volume 5. Wiley, New York, 741 pp.

Kotz S, Johnson NL, Read CB (Eds) (1985b) *Encyclopedia of statistical sciences*, volume 6. Wiley, New York, 758 pp.

Kotz S, Johnson NL, Read CB (Eds) (1986) *Encyclopedia of statistical sciences*, volume 7. Wiley, New York, 714 pp.

Kotz S, Johnson NL, Read CB (Eds) (1988a) *Encyclopedia of statistical sciences*, volume 8. Wiley, New York, 870 pp.

Kotz S, Johnson NL, Read CB (Eds) (1988b) *Encyclopedia of statistical sciences*, volume 9. Wiley, New York, 762 pp.

Kotz S, Johnson NL, Read CB (Eds) (1989) *Encyclopedia of statistical sciences*, volume S. Wiley, New York, 283 pp.

Kotz S, Nadarajah S (2000) *Extreme value distributions: Theory and applications.* Imperial College Press, London, 187 pp.

Kotz S, Read CB, Banks DL (Eds) (1997) *Encyclopedia of statistical sciences*, volume U1. Wiley, New York, 568 pp.

Kotz S, Read CB, Banks DL (Eds) (1998) *Encyclopedia of statistical sciences*, volume U2. Wiley, New York, 745 pp.

Kotz S, Read CB, Banks DL (Eds) (1999) *Encyclopedia of statistical sciences*, volume U3. Wiley, New York, 898 pp.

Koutsoyiannis D (2002) The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrological Sciences Journal* 47(4): 573–595.

Koutsoyiannis D (2005a) Hydrological persistence and the Hurst phenomenon. In: Lehr JH, Keeley J (Eds) *Water Encyclopedia: Surface and Agricultural Water.* Wiley, New York, pp 210–220.

Koutsoyiannis D (2005b) Uncertainty, entropy, scaling and hydrological stochastics. 2. Time dependence of hydrological processes and time scaling. *Hydrological Sciences Journal* 50(3): 405–426.

Koutsoyiannis D (2006) Nonstationarity versus scaling in hydrology. *Journal of Hydrology* 324(1–4): 239–254.

Koyck LM (1954) *Distributed Lags and Investment Analysis.* North-Holland, Amsterdam, 111 pp.

Kraemer HC (1974) The non-null distribution of the Spearman rank correlation coefficient. *Journal of the American Statistical Association* 69(345): 114–117.

Kraemer HC (1982) Biserial correlation. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 1. Wiley, New York, pp 276–280.

Kreiss J-P (1992) Bootstrap procedures for AR(∞)-processes. In: Jöckel K-H, Rothe G, Sendler W (Eds) *Bootstrapping and Related Techniques.* Springer, Berlin, pp 107–113.

Kreiss J-P, Franke J (1992) Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis* 13(4): 297–317.

Kristjánsson JE, Staple A, Kristiansen J, Kaas E (2002) A new look at possible connections between solar activity, clouds and climate. *Geophysical Research Letters* 29(23): 2107. [doi:10.1029/2002GL015646]

Kruskal WH (1958) Ordinal measures of association. *Journal of the American Statistical Association* 53(284): 814–861.

Kuhn TS (1970) *The Structure of Scientific Revolutions.* Second edition. University of Chicago Press, Chicago, 210 pp.

Kullback S (1983) Fisher information. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 3. Wiley, New York, pp 115–118.

Kumar KK, Rajagopalan B, Cane MA (1999) On the weakening relationship between the Indian monsoon and ENSO. *Science* 284(5423): 2156–2159.

Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* 17(3): 1217–1241.

Kürbis K, Mudelsee M, Tetzlaff G, Brázdil R (2009) Trends in extremes of temperature, dew point, and precipitation from long instrumental series from central Europe. *Theoretical and Applied Climatology* 98(1–2): 187–195.

Kürschner WM, van der Burgh J, Visscher H, Dilcher DL (1996) Oak leaves as biosensors of late Neogene and early Pleistocene paleoatmospheric CO_2 concentrations. *Marine Micropaleontology* 27(1–4): 299–312.

Kutner MH, Nachtsheim CJ, Neter J, Li W (2005) *Applied Linear Statistical Models.* Fifth edition. McGraw-Hill/Irwin, Boston, 1396 pp.

Kwon J, Min K, Bickel PJ, Renne PR (2002) Statistical methods for jointly estimating the decay constant of ^{40}K and the age of a dating standard. *Mathematical Geology* 34(4): 457–474.

Kyselý J (2002) Temporal fluctuations in heat waves at Prague–Klementinum, the Czech Republic, from 1901–97, and their relationships to atmospheric circulation. *International Journal of Climatology* 22(1): 33–50.

Kyselý J (2008) A cautionary note on the use of nonparametric bootstrap for estimating uncertainties in extreme-value models. *Journal of Applied Meteorology and Climatology* 47(12): 3236–3251.

Lahiri SN (1993) On the moving block bootstrap under long range dependence. *Statistics & Probability Letters* 18(5): 405–413.

Lahiri SN (1999) Theoretical comparisons of block bootstrap methods. *The Annals of Statistics* 27(1): 386–404.

Lahiri SN (2003) *Resampling Methods for Dependent Data*. Springer, New York, 374 pp.

Lakatos I, Musgrave A (Eds) (1970) *Criticism and the Growth of Knowledge*. Cambridge University Press, Cambridge, 282 pp.

Lanczos C (1964) A precision approximation of the gamma function. *SIAM Journal on Numerical Analysis* 1: 86–96.

Landsea CW (1993) A climatology of intense (or major) Atlantic hurricanes. *Monthly Weather Review* 121(6): 1703–1713.

Landsea CW (2007) Counting Atlantic tropical cyclones back to 1900. *Eos, Transactions of the American Geophysical Union* 88(18): 197, 202.

Landsea CW, Glenn DA, Bredemeyer W, Chenoweth M, Ellis R, Gamache J, Hufstetler L, Mock C, Perez R, Prieto R, Sánchez-Sesma J, Thomas D, Woolcock L (2008) A reanalysis of the 1911–20 Atlantic hurricane database. *Journal of Climate* 21(10): 2138–2168.

Landsea CW, Nicholls N, Gray WM, Avila LA (1996) Downward trends in the frequency of intense Atlantic hurricanes during the past five decades. *Geophysical Research Letters* 23(13): 1697–1700.

Landsea CW, Nicholls N, Gray WM, Avila LA (1997) Reply. *Geophysical Research Letters* 24(17): 2205.

Landsea CW, Pielke Jr RA, Mestas-Nuñez AM, Knaff JA (1999) Atlantic basin hurricanes: Indices of climatic changes. *Climatic Change* 42(1): 89–129.

Landsea CW, Vecchi GA, Bengtsson L, Knutson TR (2010) Impact of duration thresholds on Atlantic tropical cyclone counts. *Journal of Climate* 23(10): 2508–2519. [doi:10.1175/2009JCLI3034.1]

Landwehr JM, Matalas NC, Wallis JR (1979) Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research* 15(5): 1055–1064.

Lang M, Ouarda TBMJ, Bobée B (1999) Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* 225(3–4): 103–117.

Lanyon BP, Barbieri M, Almeida MP, Jennewein T, Ralph TC, Resch KJ, Pryde GJ, O'Brien JL, Gilchrist A, White AG (2009) Simplifying quantum logic using higher-dimensional Hilbert spaces. *Nature Physics* 5(2): 134–140.

Lanzante JR (1996) Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology* 16(11): 1197–1226.

Lassen K, Friis-Christensen E (2000) Reply. *Journal of Geophysical Research* 105(A12): 27493–27495.

Laurmann JA, Gates WL (1977) Statistical considerations in the evaluation of climatic experiments with Atmospheric General Circulation Models. *Journal of the Atmospheric Sciences* 34(8): 1187–1199.

Laut P (2003) Solar activity and terrestrial climate: An analysis of some purported correlations. *Journal of Atmospheric and Solar-Terrestrial Physics* 65(7): 801–812.

Laut P, Gundermann J (2000) Solar cycle lengths and climate: A reference revisited. *Journal of Geophysical Research* 105(A12): 27489–27492.

Lawrence KD, Arthur JL (Eds) (1990) *Robust Regression: Analysis and Applications.* Marcel Dekker, New York, 287 pp.

Leadbetter MR, Lindgren G, Rootzén H (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York, 336 pp.

Leadbetter MR, Rootzén H (1988) Extremal theory for stochastic processes. *The Annals of Probability* 16(2): 431–478.

Ledford AW, Tawn JA (2003) Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society, Series B* 65(2): 521–543.

Ledolter J (1986) Prediction and forecasting. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 7. Wiley, New York, pp 148–158.

Lees JM, Park J (1995) Multiple-taper spectral analysis: A stand-alone C-subroutine. *Computers and Geosciences* 21(2): 199–236.

Lehmann EL, Casella G (1998) *Theory of Point Estimation.* Second edition. Springer, New York, 589 pp.

Lehmann EL, Romano JP (2005) *Testing Statistical Hypotheses.* Third edition. Springer, New York, 784 pp.

Leith CE (1973) The standard error of time-average estimates of climatic means. *Journal of Applied Meteorology* 12(6): 1066–1069.

Leith NA, Chandler RE (2010) A framework for interpreting climate model outputs. *Applied Statistics* 59(2): 279–296.

LePage R, Billard L (Eds) (1992) *Exploring the Limits of Bootstrap.* Wiley, New York, 426 pp.

Li H, Maddala GS (1996) Bootstrapping time series models (with discussion). *Econometric Reviews* 15(2): 115–195.

Linden M (1999) Time series properties of aggregated AR(1) processes with uniformly distributed coefficients. *Economics Letters* 64(1): 31–36.

Linder E, Babu GJ (1994) Bootstrapping the linear functional relationship with known error variance ratio. *Scandinavian Journal of Statistics* 21(1): 21–39.

Lindley DV (1947) Regression lines and the linear functional relationship. *Journal of the Royal Statistical Society, Supplement* 9(2): 218–244.

Lindley DV (1965) *Introduction to Probability and Statistics.* Cambridge University Press, Cambridge, 259 pp.

Linnell Nemec AF, Nemec JM (1985) A test of significance for periods derived using phase-dispersion-minimization techniques. *The Astronomical Journal* 90(11): 2317–2320.

Lisiecki LE, Lisiecki PA (2002) Application of dynamic programming to the correlation of paleoclimate records. *Paleoceanography* 17(4): 1049. [doi:10.1029/2001PA000733]

Lisiecki LE, Raymo ME (2005) A Pliocene–Pleistocene stack of 57 globally distributed benthic δ^{18}O records. *Paleoceanography* 20(1): PA1003. [doi:10.1029/2004PA001071]

Liu RY, Singh K (1992) Moving blocks jackknife and bootstrap capture weak dependence. In: LePage R, Billard L (Eds) *Exploring the Limits of Bootstrap.* Wiley, New York, pp 225–248.

Loader CR (1992) A log-linear model for a Poisson process change point. *The Annals of Statistics* 20(3): 1391–1411.

Lockwood M, Fröhlich C (2007) Recent oppositely directed trends in solar climate forcings and the global mean surface air temperature. *Proceedings of the Royal Society of London, Series A* 463(2086): 2447–2460.

Loh W-Y (1987) Calibrating confidence coefficients. *Journal of the American Statistical Association* 82(397): 155–162.

Loh W-Y (1991) Bootstrap calibration for confidence interval construction and selection. *Statistica Sinica* 1(2): 477–491.

Lomb NR (1976) Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science* 39(2): 447–462.

Lomnicki ZA (1967) On the distribution of products of random variables. *Journal of the Royal Statistical Society, Series B* 29(3): 513–524.

Lorenz EN (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 20(2): 130–141.

Lorenz EN (1991) Dimension of weather and climate attractors. *Nature* 353(6341): 241–244.

Lovelock JE, Kump LR (1994) Failure of climate regulation in a geophysiological model. *Nature* 369(6483): 732–734.

Lu L-H, Stedinger JR (1992) Variance of two- and three-parameter GEV/PWM quantile estimators: Formulae, confidence intervals, and a comparison. *Journal of Hydrology* 138(1–2): 247–267.

Ludwig KR (2003) *User's Manual for Isoplot 3.00: A Geochronological Toolkit for Microsoft Excel.* Berkeley Geochronology Center, Berkeley, CA, 70 pp. [Special Publication No. 4]

Lund R, Wang XL, Lu Q, Reeves J, Gallagher C, Feng Y (2007) Changepoint detection in periodic and autocorrelated time series. *Journal of Climate* 20(20): 5178–5190.

Luterbacher J, Rickli R, Xoplaki E, Tinguely C, Beck C, Pfister C, Wanner H (2001) The late Maunder Minimum (1675–1715)—A key period for studying decadal scale climatic change in Europe. *Climatic Change* 49(4): 441–462.

Lüthi D, Le Floch M, Bereiter B, Blunier T, Barnola J-M, Siegenthaler U, Raynaud D, Jouzel J, Fischer H, Kawamura K, Stocker TF (2008) High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* 453(7193): 379–382.

Lybanon M (1984) A better least-squares method when both variables have uncertainties. *American Journal of Physics* 52(1): 22–26.

Maasch KA (1988) Statistical detection of the mid-Pleistocene transition. *Climate Dynamics* 2(3): 133–143.

MacDonald GJ (1989) Spectral analysis of time series generated by nonlinear processes. *Reviews of Geophysics* 27(4): 449–469.

Macdonald JR, Thompson WJ (1992) Least-squares fitting when both variables contain errors: Pitfalls and possibilities. *American Journal of Physics* 60(1): 66–73.

Macleod AJ (1989) A remark on algorithm AS 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Applied Statistics* 38(1): 198–199.

Madansky A (1959) The fitting of straight lines when both variables are subject to error. *Journal of the American Statistical Association* 54(285): 173–205.

Madden RA, Jones RH (2001) A quantitative estimate of the effect of aliasing in climatological time series. *Journal of Climate* 14(19): 3987–3993.

Maidment DR (Ed) (1993) *Handbook of Hydrology.* McGraw-Hill, New York, 1400 pp.

Mandelbrot BB (1983) Fractional Brownian motions and fractional Gaussian noises. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 3. Wiley, New York, pp 186–189.

Mandelbrot BB, Wallis JR (1969) Some long-run properties of geophysical records. *Water Resources Research* 5(2): 321–340.

Mankinen EA, Dalrymple GB (1979) Revised geomagnetic polarity time scale for the interval 0–5 m.y. B.P. *Journal of Geophysical Research* 84(B2): 615–626.

Manley G (1974) Central England temperatures: Monthly means 1659 to 1973. *Quarterly Journal of the Royal Meteorological Society* 100(425): 389–405.

Mann HB (1945) Nonparametric tests against trend. *Econometrica* 13(3): 245–259.

Mann ME, Emanuel KA (2006) Atlantic hurricane trends linked to climate change. *Eos, Transactions of the American Geophysical Union* 87(24): 233, 238, 241.

Mann ME, Emanuel KA, Holland GJ, Webster PJ (2007a) Atlantic tropical cyclones revisited. *Eos, Transactions of the American Geophysical Union* 88(36): 349–350.

Mann ME, Lees JM (1996) Robust estimation of background noise and signal detection in climatic time series. *Climatic Change* 33(3): 409–445.

Mann ME, Sabbatelli TA, Neu U (2007b) Evidence for a modest undercount bias in early historical Atlantic tropical cyclone counts. *Geophysical Research Letters* 34(22): L22707. [doi:10.1029/2007GL031781; corrigendum: 2007 Vol. 34(24): L24704 (doi:10.1029/2007GL032798)]

Mann ME, Woodruff JD, Donnelly JP, Zhang Z (2009) Atlantic hurricanes and climate over the past 1,500 years. *Nature* 460(7257): 880–883.

Maraun D, Rust HW, Timmer J (2004) Tempting long-memory—on the interpretation of DFA results. *Nonlinear Processes in Geophysics* 11(4): 495–503.

Markowitz E (1968a) Minimum mean-square-error estimation of the standard deviation of the normal distribution. *The American Statistician* 22(3): 26.

Markowitz E (1968b) Priority acknowledgement to "Minimum mean-square-error estimation of the standard deviation of the normal distribution." *The American Statistician* 22(4): 42.

Marquardt DW, Acuff SK (1982) Direct quadratic spectrum estimation from unequally spaced data. In: Anderson OD, Perryman MR (Eds) *Applied Time Series Analysis*. North-Holland, Amsterdam, pp 199–227.

Marriott FHC, Pope JA (1954) Bias in the estimation of autocorrelations. *Biometrika* 41(3–4): 390–402.

Marron JS (1987) What does optimal bandwidth selection mean for nonparametric regression estimation? In: Dodge Y (Ed) *Statistical Data Analysis Based on the L_1-Norm and Related Methods*. North-Holland, Amsterdam, pp 379–392.

Marron JS (1988) Automatic smoothing parameter selection: A survey. *Empirical Economics* 13(3–4): 187–208.

Martin MA (1990) On bootstrap iteration for coverage correction in confidence intervals. *Journal of the American Statistical Association* 85(412): 1105–1118.

Martin MA (2007) Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties. *Computational Statistics and Data Analysis* 51(12): 6321–6342.

Martin RJ (1998) *Irregularly sampled signals: Theories and techniques for analysis.* Ph.D. Dissertation. University College London, London, 158 pp.

Martins ES, Stedinger JR (2000) Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research* 36(3): 737–744.

Martins ES, Stedinger JR (2001) Generalized maximum likelihood Pareto–Poisson estimators for partial duration series. *Water Resources Research* 37(10): 2551–2557.

Martinson DG, Menke W, Stoffa P (1982) An inverse approach to signal correlation. *Journal of Geophysical Research* 87(B6): 4807–4818.

Martinson DG, Pisias NG, Hays JD, Imbrie J, Moore Jr TC, Shackleton NJ (1987) Age dating and the orbital theory of the ice ages: Development of a high-resolution 0 to 300,000-year chronostratigraphy. *Quaternary Research* 27(1): 1–29.

Masry E (1984) Spectral and probability density estimation from irregularly observed data. In: Parzen E (Ed) *Time Series Analysis of Irregularly Observed Data.* Springer, New York, pp 224–250.

Matalas NC, Langbein WB (1962) Information content of the mean. *Journal of Geophysical Research* 67(9): 3441–3448.

Matteucci G (1990) Analysis of the probability distribution of the late Pleistocene climatic record: Implications for model validation. *Climate Dynamics* 5(1): 35–52.

Matyasovszky I (2001) A nonlinear approach to modeling climatological time series. *Theoretical and Applied Climatology* 69(3–4): 139–147.

Mayewski PA, Meeker LD, Twickler MS, Whitlow S, Yang Q, Lyons WB, Prentice M (1997) Major features and forcing of high-latitude northern hemisphere atmospheric circulation using a 110,000-year-long glaciochemical series. *Journal of Geophysical Research* 102(C12): 26345–26366.

McAvaney BJ, Covey C, Joussaume S, Kattsov V, Kitoh A, Ogana W, Pitman AJ, Weaver AJ, Wood RA, Zhao Z-C (2001) Model evaluation. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (Eds) *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, pp 471–523.

McGuffie K, Henderson-Sellers A (1997) *A Climate Modelling Primer.* Second edition. Wiley, Chichester, 253 pp.

McMillan DG, Constable CG, Parker RL (2002) Limitations on stratigraphic analyses due to incomplete age control and their relevance to sedimentary paleomagnetism. *Earth and Planetary Science Letters* 201(3–4): 509–523.

Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper SCB, Watterson IG, Weaver AJ, Zhao Z-C (2007) Global climate projections. In: Solomon S, Qin D, Manning M, Marquis M, Averyt K, Tignor MMB, Miller Jr HL, Chen Z (Eds) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, pp 747–845.

Meehl GA, Tebaldi C (2004) More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* 305(5686): 994–997.

Meehl GA, Washington WM, Wigley TML, Arblaster JM, Dai A (2003) Solar and greenhouse gas forcing and climate response in the twentieth century. *Journal of Climate* 16(3): 426–444.

Meehl GA, Zwiers F, Evans J, Knutson T, Mearns L, Whetton P (2000) Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. *Bulletin of the American Meteorological Society* 81(3): 427–436.

Meeker LD, Mayewski PA, Grootes PM, Alley RB, Bond GC (2001) Comment on "On sharp spectral lines in the climate record and the millennial peak" by Carl Wunsch. *Paleoceanography* 16(5): 544–547.

Menzefricke U (1981) A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. *Applied Statistics* 30(2): 141–146.

Mesa OJ, Poveda G (1993) The Hurst effect: The scale of fluctuation approach. *Water Resources Research* 29(12): 3995–4002.

Meyer MC, Faber R, Spötl C (2006) The WinGeol Lamination Tool: New software for rapid, semi-automated analysis of laminated climate archives. *The Holocene* 16(5): 753–761.

Miao X, Mason JA, Johnson WC, Wang H (2007) High-resolution proxy record of Holocene climate from a loess section in southwestern Nebraska, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology* 245(3–4): 368–381.

Michener WK, Blood ER, Bildstein KL, Brinson MM, Gardner LR (1997) Climate change, hurricanes and tropical storms, and rising sea level in coastal wetlands. *Ecological Applications* 7(3): 770–801.

Miller DM (1984) Reducing transformation bias in curve fitting. *The American Statistician* 38(2): 124–126.

Mills TC (2007) Time series modelling of two millennia of northern hemisphere temperatures: Long memory or shifting trends? *Journal of the Royal Statistical Society, Series A* 170(1): 83–94.

Milly PCD, Wetherald RT (2002) Macroscale water fluxes 3. Effects of land processes on variability of monthly river discharge. *Water Resources Research* 38(11): 1235. [doi:10.1029/2001WR000761]

Milne AE, Lark RM (2009) Wavelet transforms applied to irregularly sampled soil data. *Mathematical Geosciences* 41(6): 661–678.

Mitchell JFB, Wilson CA, Cunnington WM (1987) On CO_2 climate sensitivity and model dependence of results. *Quarterly Journal of the Royal Meteorological Society* 113(475): 293–322.

Mondal D, Percival DB (in press) Wavelet variance analysis for gappy time series. *Annals of the Institute of Statistical Mathematics*. [doi:10.1007/s10463-008-0195-z]

Monnin E, Indermühle A, Dällenbach A, Flückiger J, Stauffer B, Stocker TF, Raynaud D, Barnola J-M (2001) Atmospheric CO_2 concentrations over the last glacial termination. *Science* 291(5501): 112–114.

Monro DM (1975) Complex discrete Fast Fourier Transform. *Applied Statistics* 24(1): 153–160.

Monro DM (1976) Real discrete Fast Fourier Transform. *Applied Statistics* 25(2): 166–172.

Montanari A (2003) Long-range dependence in hydrology. In: Doukhan P, Oppenheim G, Taqqu MS (Eds) *Theory and Applications of Long-Range Dependence*. Birkhäuser, Boston, pp 461–472.

Montanari A, Rosso R, Taqqu MS (1997) Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. *Water Resources Research* 33(5): 1035–1044.

Montgomery DC, Peck EA (1992) *Introduction to Linear Regression Analysis*. Second edition. Wiley, New York, 527 pp.

Montgomery DC, Peck EA, Vining GG (2006) *Introduction to Linear Regression Analysis*. Fourth edition. Wiley, Hoboken, NJ, 612 pp.

Moore MI, Thomson PJ (1991) Impact of jittered sampling on conventional spectral estimates. *Journal of Geophysical Research* 96(C10): 18519–18526.

Moore PD, Webb JA, Collinson ME (1991) *Pollen analysis*. Second edition. Blackwell, Oxford, 216 pp.

Moran PAP (1948) Rank correlation and product-moment correlation. *Biometrika* 35(1–2): 203–206.

Mosedale TJ, Stephenson DB, Collins M, Mills TC (2006) Granger causality of coupled climate processes: Ocean feedback on the North Atlantic Oscillation. *Journal of Climate* 19(7): 1182–1194.

Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA, Mitchell JFB, Nakicenovic N, Riahi K, Smith SJ, Stouffer RJ, Thomson AM, Weyant JP, Wilbanks TJ (2010) The next generation of scenarios for climate change research and assessment. *Nature* 463(7282): 747–756.

Mostafa MD, Mahmoud MW (1964) On the problem of estimation for the bivariate lognormal distribution. *Biometrika* 51(3–4): 522–527.

Mudelsee M (1999) *On an interesting statistical problem imposed by an ice core*. Institute of Mathematics and Statistics, University of Kent, Canterbury, 12 pp. [Technical Report UKC/IMS/99/21]

Mudelsee M (2000) Ramp function regression: A tool for quantifying climate transitions. *Computers and Geosciences* 26(3): 293–307.

Mudelsee M (2001a) Note on the bias in the estimation of the serial correlation coefficient of AR(1) processes. *Statistical Papers* 42(4): 517–527.

Mudelsee M (2001b) The phase relations among atmospheric CO_2 content, temperature and global ice volume over the past 420 ka. *Quaternary Science Reviews* 20(4): 583–589.

Mudelsee M (2002) TAUEST: A computer program for estimating persistence in unevenly spaced weather/climate time series. *Computers and Geosciences* 28(1): 69–72.

Mudelsee M (2003) Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. *Mathematical Geology* 35(6): 651–665.

Mudelsee M (2005) A new, absolutely dated geomagnetic polarity timescale for the Late Pliocene to Early Pleistocene. In: Berger A, Ercegovac M, Mesinger F (Eds) *Milutin Milankovitch Anniversary Symposium: Paleoclimate and the Earth Climate System*. Serbian Academy of Sciences and Arts, Belgrade, pp 145–149.

Mudelsee M (2006) CLIM-X-DETECT: A Fortran 90 program for robust detection of extremes against a time-dependent background in climate records. *Computers and Geosciences* 32(1): 141–144.

Mudelsee M (2007) Long memory of rivers from spatial aggregation. *Water Resources Research* 43(1): W01202. [doi:10.1029/2006WR005721]

Mudelsee M (2009) Break function regression: A tool for quantifying trend changes in climate time series. *European Physical Journal Special Topics* 174(1): 49–63.

Mudelsee M, Alkio M (2007) Quantifying effects in two-sample environmental experiments using bootstrap confidence intervals. *Environmental Modelling and Software* 22(1): 84–96.

Mudelsee M, Börngen M, Tetzlaff G, Grünewald U (2003) No upward trends in the occurrence of extreme floods in central Europe. *Nature* 425(6954): 166–169. [Corrigendum: Insert in Eq. (1) on the right-hand side a factor h^{-1} before the sum sign.]

Mudelsee M, Börngen M, Tetzlaff G, Grünewald U (2004) Extreme floods in central Europe over the past 500 years: Role of cyclone pathway "Zugstrasse Vb." *Journal of Geophysical Research* 109(D23): D23101. [doi:10.1029/2004JD005034; corrigendum: Eq. (5): replace $n^\dagger$ by n, Eq. (6): replace $K(t - T^\dagger(j))$ by $h^{-1}K([t - T^\dagger(j)]h^{-1})$.]

Mudelsee M, Deutsch M, Börngen M, Tetzlaff G (2006) Trends in flood risk of the River Werra (Germany) over the past 500 years. *Hydrological Sciences Journal* 51(5): 818–833.

Mudelsee M, Raymo ME (2005) Slow dynamics of the Northern Hemisphere Glaciation. *Paleoceanography* 20(4): PA4022. [doi:10.1029/2005PA001153]

Mudelsee M, Scholz D, Röthlisberger R, Fleitmann D, Mangini A, Wolff EW (2009) Climate spectrum estimation in the presence of timescale errors. *Nonlinear Processes in Geophysics* 16(1): 43–56.

Mudelsee M, Schulz M (1997) The Mid-Pleistocene Climate Transition: Onset of 100 ka cycle lags ice volume build-up by 280 ka. *Earth and Planetary Science Letters* 151(1–2): 117–123.

Mudelsee M, Stattegger K (1994) Plio-/Pleistocene climate modeling based on oxygen isotope time series from deep-sea sediment cores: The Grassberger–Procaccia algorithm and chaotic climate systems. *Mathematical Geology* 26(7): 799–815.

Mudelsee M, Stattegger K (1997) Exploring the structure of the mid-Pleistocene revolution with advanced methods of time-series analysis. *Geologische Rundschau* 86(2): 499–511.

Mueller M (2003) Damages of the Elbe flood 2002 in Germany—A review. *Geophysical Research Abstracts* 5: 12992.

Müller H-G (1992) Change-points in nonparametric regression analysis. *The Annals of Statistics* 20(2): 737–761.

Muller RA, MacDonald GJ (1995) Glacial cycles and orbital inclination. *Nature* 377(6545): 107–108.

Muller RA, MacDonald GJ (1997a) Glacial cycles and astronomical forcing. *Science* 277(5323): 215–218.

Muller RA, MacDonald GJ (1997b) Simultaneous presence of orbital inclination and eccentricity in proxy climate records from Ocean Drilling Program Site 806. *Geology* 25(1): 3–6.

Muller RA, MacDonald GJ (1997c) Spectrum of the 100 kyr glacial cycle: Orbital inclination, not eccentricity. *Proceedings of the National Academy of Sciences of the United States of America* 94(16): 8329–8334.

Muller RA, MacDonald GJ (2000) *Ice Ages and Astronomical Causes: Data, spectral analysis and mechanisms*. Springer, London, 318 pp.

Mullis CT, Scharf LL (1991) Quadratic estimators of the power spectrum. In: Haykin S (Ed) *Advances in Spectrum Analysis and Array Processing*, volume 1. Prentice-Hall, Englewood Cliffs, NJ, pp 1–57.

Munk W, Hasselmann K (1964) Super-resolution of tides. In: Yoshida K (Ed) *Studies on Oceanography: A Collection of Papers dedicated to Koji Hidaka*. University of Washington Press, Seattle, WA, pp 339–344.

Münnich KO, Östlund HG, de Vries H (1958) Carbon-14 activity during the past 5,000 years. *Nature* 182(4647): 1432–1433.

Musekiwa A (2005) *Estimating the slope in the simple linear errors-in-variables model*. M.Sc. Thesis. University of Johannesburg, Johannesburg, South Africa, 85 pp.

Nakagawa S, Niki N (1992) Distribution of the sample correlation coefficient for non-normal populations. *Journal of the Japanese Society of Computational Statistics* 5(1): 1–19.

Naveau P, Nogaj M, Ammann C, Yiou P, Cooley D, Jomelli V (2005) Statistical methods for the analysis of climate extremes. *Comptes Rendus Geoscience* 337(10–11): 1013–1022.

Neff U, Burns SJ, Mangini A, Mudelsee M, Fleitmann D, Matter A (2001) Strong coherence between solar variability and the monsoon in Oman between 9 and 6 kyr ago. *Nature* 411(6835): 290–293.

Negendank JFW, Zolitschka B (Eds) (1993) *Paleolimnology of European Maar Lakes.* Springer, Berlin, 513 pp.

Neuendorf KKE, Mehl Jr JP, Jackson JA (2005) *Glossary of Geology.* Fifth edition. American Geological Institute, Alexandria, VA, 779 pp.

Neumann MH, Kreiss J-P (1998) Regression-type inference in nonparametric autoregression. *The Annals of Statistics* 26(4): 1570–1613.

Newton HJ, North GR, Crowley TJ (1991) Forecasting global ice volume. *Journal of Time Series Analysis* 12(3): 255–265.

Nicolis C, Nicolis G (1984) Is there a climatic attractor? *Nature* 311(5986): 529–532.

Nielsen MA, Chuang IL (2000) *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, 676 pp.

Nierenberg WA (Ed) (1992) *Encyclopedia of Earth System Science*, volume 1–4. Academic Press, San Diego, 2825 pp.

Nievergelt Y (1998) Total least squares. In: Kotz S, Read CB, Banks DL (Eds) *Encyclopedia of statistical sciences*, volume U2. Wiley, New York, pp 666–670.

Niggemann S, Mangini A, Mudelsee M, Richter DK, Wurth G (2003) Sub-Milankovitch climatic cycles in Holocene stalagmites from Sauerland, Germany. *Earth and Planetary Science Letters* 216(4): 539–547.

Nogaj M, Yiou P, Parey S, Malek F, Naveau P (2006) Amplitude and frequency of temperature extremes over the North Atlantic region. *Geophysical Research Letters* 33(10): L10801. [doi:10.1029/2005GL024251]

Nordgaard A (1992) Resampling stochastic processes using a bootstrap approach. In: Jöckel K-H, Rothe G, Sendler W (Eds) *Bootstrapping and Related Techniques.* Springer, Berlin, pp 181–185.

North Greenland Ice Core Project members (2004) High-resolution record of northern hemisphere climate extending into the last interglacial period. *Nature* 431(7005): 147–151.

Nuttall AH (1981) Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29(1): 84–91.

Nyberg J, Malmgren BA, Winter A, Jury MR, Kilbourne KH, Quinn TM (2007) Low Atlantic hurricane activity in the 1970s and 1980s compared to the past 270 years. *Nature* 447(7145): 698–701.

Ocean Drilling Program (Ed) (1986–2004) *Proceedings of the Ocean Drilling Program, Initial Reports*, volume 101–210. Ocean Drilling Program, College Station, TX.

Ocean Drilling Program (Ed) (1988–2007) *Proceedings of the Ocean Drilling Program, Scientific Results*, volume 101–210. Ocean Drilling Program, College Station, TX.

Odeh RE, Evans JO (1974) The percentage points of the normal distribution. *Applied Statistics* 23(1): 96–97.

Odell PL (1983) Gauss–Markov theorem. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 3. Wiley, New York, pp 314–316.

Oeschger H, Langway Jr CC (Eds) (1989) *The Environmental Record in Glaciers and Ice Sheets.* Wiley, Chichester, 401 pp.

Oh H-S, Nychka D, Brown T, Charbonneau P (2004) Period analysis of variable stars by robust smoothing. *Applied Statistics* 53(1): 15–30.

Otten A (1973) The null distribution of Spearman's S when $n = 13(1)16$. *Statistica Neerlandica* 27(1): 19–20.

Packard NH, Crutchfield JP, Farmer JD, Shaw RS (1980) Geometry from a time series. *Physical Review Letters* 45(9): 712–716.

Page ES (1954) Continuous inspection schemes. *Biometrika* 41(1–2): 100–115.

Palm FC, Smeekes S, Urbain J-P (2008) Bootstrap unit-root tests: Comparison and extensions. *Journal of Time Series Analysis* 29(2): 371–401.

Paluš M, Vejmelka M (2007) Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections. *Physical Review E* 75(5): 056211. [doi:10.1103/PhysRevE.75.056211]

Pankratz A (1991) *Forecasting with Dynamic Regression Models.* Wiley, New York, 386 pp.

Paparoditis E (2002) Frequency domain bootstrap for time series. In: Dehling H, Mikosch T, Sørensen M (Eds) *Empirical Process Techniques for Dependent Data.* Birkhäuser, Boston, pp 365–381.

Paparoditis E, Politis DN (2001) Tapered block bootstrap. *Biometrika* 88(4): 1105–1119.

Paparoditis E, Politis DN (2002) Local block bootstrap. *Comptes Rendus Mathematique* 335(11): 959–962.

Pardo-Igúzquiza E, Chica-Olmo M, Rodríguez-Tovar FJ (1994) CYSTRATI: A computer program for spectral analysis of stratigraphic successions. *Computers and Geosciences* 20(4): 511–584.

Parent E, Bernier J (2003a) Bayesian POT modeling for historical data. *Journal of Hydrology* 274(1–4): 95–108.

Parent E, Bernier J (2003b) Encoding prior experts judgments to improve risk analysis of extreme hydrological events via POT modeling. *Journal of Hydrology* 283(1–4): 1–18.

Park E, Lee YJ (2001) Estimates of standard deviation of Spearman's rank correlation coefficients with dependent observations. *Communications in Statistics—Simulation and Computation* 30(1): 129–142.

Park J (1992) Envelope estimation for quasi-periodic geophysical signals in noise: A multitaper approach. In: Walden AT, Guttorp P (Eds) *Statistics in the Environmental & Earth Sciences.* Edward Arnold, London, pp 189–219.

Park SK, Miller KW (1988) Random number generators: Good ones are hard to find. *Communications of the ACM* 31(10): 1192–1201.

Parrenin F, Barnola J-M, Beer J, Blunier T, Castellano E, Chappellaz J, Dreyfus G, Fischer H, Fujita S, Jouzel J, Kawamura K, Lemieux-Dudon B, Loulergue L, Masson-Delmotte V, Narcisi B, Petit J-R, Raisbeck G, Raynaud D, Ruth U, Schwander J, Severi M, Spahni R, Steffensen JP, Svensson A, Udisti R, Waelbroeck C, Wolff E (2007) The EDC3 chronology for the EPICA Dome C ice core. *Climate of the Past* 3(3): 485–497.

Parthasarathy B, Munot AA, Kothawale DR (1994) All-India monthly and seasonal rainfall series: 1871–1993. *Theoretical and Applied Climatology* 49(4): 217–224.

Parzen E (Ed) (1984) *Time Series Analysis of Irregularly Observed Data.* Springer, New York, 363 pp.

Patel JK, Read CB (1996) *Handbook of the Normal Distribution.* Second edition. Marcel Dekker, New York, 431 pp.

Paul A, Schäfer-Neth C (2005) How to combine sparse proxy data and coupled climate models. *Quaternary Science Reviews* 24(7–9): 1095–1107.

Pauli F, Coles S (2001) Penalized likelihood inference in extreme value analyses. *Journal of Applied Statistics* 28(5): 547–560.

Pearson K (1896) Mathematical contributions to the theory of evolution—III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London, Series A* 187: 253–318.

Pearson K (1901) On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2(11): 559–572.

Pearson K (1907) Mathematical contributions to the theory of evolution—XVI. On further methods for determining correlation. *Drapers' Company Research Memoirs, Biometric Series* 4: 1–39.

Pearson K (1924) *The Life, Letters and Labours of Francis Galton*, volume 2. Cambridge University Press, Cambridge, 425 pp.

Pelletier JD, Turcotte DL (1997) Long-range persistence in climatological and hydrological time series: Analysis, modeling and application to drought hazard assessment. *Journal of Hydrology* 203(1–4): 198–208.

Peng C-K, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. *Physical Review E* 49(2): 1685–1689.

Peng C-K, Havlin S, Stanley HE, Goldberger AL (1995) Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 5(1): 82–87.

Penner JE, Andreae M, Annegarn H, Barrie L, Feichter J, Hegg D, Jayaraman A, Leaitch R, Murphy D, Nganga J, Pitari G (2001) Aerosols, their direct and indirect effects. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (Eds) *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, pp 289–348.

Percival DB, Walden AT (1993) *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques.* Cambridge University Press, Cambridge, 583 pp.

Percival DB, Walden AT (2000) *Wavelet Methods for Time Series Analysis.* Cambridge University Press, Cambridge, 594 pp.

Perron P (2006) Dealing with structural breaks. In: Mills TC, Patterson K (Eds) *Palgrave Handbook of Econometrics*, volume 1. Palgrave Macmillan, Houndmills, Basingstoke, pp 278–352.

Pestiaux P, Berger A (1984) Impacts of deep-sea processes on paleoclimatic spectra. In: Berger A, Imbrie J, Hays J, Kukla G, Saltzman B (Eds) *Milankovitch and Climate*, volume 1. D. Reidel, Dordrecht, pp 493–510.

Peters SC, Freedman DA (1984) Some notes on the bootstrap in regression problems. *Journal of Business & Economic Statistics* 2(4): 406–409.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Førland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D (1998a) Homogeneity adjustments of *in situ* atmospheric climate data: A review. *International Journal of Climatology* 18(13): 1493–1517.

Peterson TC, Vose R, Schmoyer R, Razuvaëv V (1998b) Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *International Journal of Climatology* 18(11): 1169–1179.

Petit JR, Jouzel J, Raynaud D, Barkov NI, Barnola J-M, Basile I, Bender M, Chappellaz J, Davis M, Delaygue G, Delmotte M, Kotlyakov VM, Legrand M, Lipenkov VY, Lorius C, Pépin L, Ritz C, Saltzman E, Stievenard M (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* 399(6735): 429–436.

Pettitt AN (1979) A non-parametric approach to the change-point problem. *Applied Statistics* 28(2): 126–135.

Pfister C (1999) *Wetternachhersage*. Paul Haupt, Bern, 304 pp.

Pickands III J (1975) Statistical inference using extreme order statistics. *The Annals of Statistics* 3(1): 119–131.

Pielke Jr RA, Landsea C, Mayfield M, Laver J, Pasch R (2005) Hurricanes and global warming. *Bulletin of the American Meteorological Society* 86(11): 1571–1575.

Pielke Jr RA, Landsea CW (1998) Normalized hurricane damages in the United States: 1925–95. *Weather and Forecasting* 13(3): 621–631.

Pirie W (1988) Spearman rank correlation coefficient. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 8. Wiley, New York, pp 584–587.

Pisias NG, Mix AC (1988) Aliasing of the geologic record and the search for long-period Milankovitch cycles. *Paleoceanography* 3(5): 613–619.

Pittock AB (1978) A critical look at long-term Sun–weather relationships. *Reviews of Geophysics and Space Physics* 16(3): 400–420.

Polansky AM (1999) Upper bounds on the true coverage of bootstrap percentile type confidence intervals. *The American Statistician* 53(4): 362–369.

Polanyi M (1958) *Personal Knowledge: Towards a Post-Critical Philosophy.* University of Chicago Press, Chicago, 428 pp.

Politis DN (2003) The impact of bootstrap methods on time series analysis. *Statistical Science* 18(2): 219–230.

Politis DN, Romano JP (1992a) A circular block-resampling procedure for stationary data. In: LePage R, Billard L (Eds) *Exploring the Limits of Bootstrap.* Wiley, New York, pp 263–270.

Politis DN, Romano JP (1992b) A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *The Annals of Statistics* 20(4): 1985–2007.

Politis DN, Romano JP (1994) The stationary bootstrap. *Journal of the American Statistical Association* 89(428): 1303–1313.

Politis DN, Romano JP, Lai T-L (1992) Bootstrap confidence bands for spectra and cross-spectra. *IEEE Transactions on Signal Processing* 40(5): 1206–1215.

Politis DN, Romano JP, Wolf M (1999) *Subsampling.* Springer, New York, 347 pp.

Politis DN, White H (2004) Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23(1): 53–70.

Popper K (1935) *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft.* Julius Springer, Wien, 248 pp.

Powell JL (1986) Censored regression quantiles. *Journal of Econometrics* 32(1): 143–155.

Prais SJ, Winsten CB (1954) *Trend Estimators and Serial Correlation.* Cowles Commission, Yale University, New Haven, CT, 26 pp. [Discussion Paper No. 383]

Preisendorfer RW (1988) *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, Amsterdam, 425 pp.

Prell WL, Imbrie J, Martinson DG, Morley JJ, Pisias NG, Shackleton NJ, Streeter HF (1986) Graphic correlation of oxygen isotope stratigraphy application to the late Quaternary. *Paleoceanography* 1(2): 137–162.

Prescott P, Walden AT (1980) Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* 67(3): 723–724.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) *Numerical Recipes in Fortran 77: The Art of Scientific Computing.* Second edition. Cambridge University Press, Cambridge, 933 pp.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1996) *Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing.* Second edition. Cambridge University Press, Cambridge, pp 935–1486.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes: The Art of Scientific Computing.* Third edition. Cambridge University Press, Cambridge, 1235 pp. [C++ code]

Prichard D, Theiler J (1995) Generalized redundancies for time series analysis. *Physica D* 84(3–4): 476–493.

Priestley MB (1962a) The analysis of stationary processes with mixed spectra—I. *Journal of the Royal Statistical Society, Series B* 24(1): 215–233.

Priestley MB (1962b) Analysis of stationary processes with mixed spectra—II. *Journal of the Royal Statistical Society, Series B* 24(2): 511–529.

Priestley MB (1981) *Spectral Analysis and Time Series.* Academic Press, London, 890 pp.

Priestley MB (1988) *Non-linear and Non-stationary Time Series Analysis.* Academic Press, London, 237 pp.

Priestley MB (1996) Wavelets and time-dependent spectral analysis. *Journal of Time Series Analysis* 17(1): 85–103.

Priestley MB (1997) Detection of periodicities. In: Subba Rao T, Priestley MB, Lessi O (Eds) *Applications of Time Series Analysis in Astronomy and Meteorology.* Chapman and Hall, London, pp 65–88.

Priestley MB, Chao MT (1972) Non-parametric function fitting. *Journal of the Royal Statistical Society, Series B* 34(3): 385–392.

Prieto GA, Parker RL, Vernon III FL (2009) A Fortran 90 library for multitaper spectrum analysis. *Computers and Geosciences* 35(8): 1701–1710.

Prieto GA, Thomson DJ, Vernon FL, Shearer PM, Parker RL (2007) Confidence intervals for earthquake source parameters. *Geophysical Journal International* 168(3): 1227–1234.

Prokopenko AA, Hinnov LA, Williams DF, Kuzmin MI (2006) Orbital forcing of continental climate during the Pleistocene: A complete astronomically tuned climatic record from Lake Baikal, SE Siberia. *Quaternary Science Reviews* 25(23–24): 3431–3457.

Prueher LM, Rea DK (2001) Volcanic triggering of late Pliocene glaciation: Evidence from the flux of volcanic glass and ice-rafted debris to the North Pacific Ocean. *Palaeogeography, Palaeoclimatology, Palaeoecology* 173(3–4): 215–230.

Pujol N, Neppel L, Sabatier R (2007) Regional tests for trend detection in maximum precipitation series in the French Mediterranean region. *Hydrological Sciences Journal* 52(5): 956–973.

Pyper BJ, Peterman RM (1998) Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Canadian Journal of Fisheries and Aquatic Sciences* 55(9): 2127–2140. [Corrigendum: 1998 Vol. 55(12): 2710]

Quinn BG (1989) Estimating the number of terms in a sinusoidal regression. *Journal of Time Series Analysis* 10(1): 71–75.

Quinn BG, Hannan EJ (2001) *The Estimation and Tracking of Frequency*. Cambridge University Press, Cambridge, 266 pp.

Rahmstorf S (2003) Timing of abrupt climate change: A precise clock. *Geophysical Research Letters* 30(10): 1510. [doi:10.1029/2003GL017115]

Ramesh NI, Davison AC (2002) Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology* 256(1–2): 106–119.

Ramsey CB (2008) Deposition models for chronological records. *Quaternary Science Reviews* 27(1–2): 42–60.

Randall DA, Wood RA, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Stouffer RJ, Sumi A, Taylor KE (2007) Climate models and their evaluation. In: Solomon S, Qin D, Manning M, Marquis M, Averyt K, Tignor MMB, Miller Jr HL, Chen Z (Eds) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, pp 589–662.

Rao AR, Hamed KH (2000) *Flood Frequency Analysis*. CRC Press, Boca Raton, FL, 350 pp.

Raymo ME (1997) The timing of major climate terminations. *Paleoceanography* 12(4): 577–585.

Raymo ME, Huybers P (2008) Unlocking the mysteries of the ice ages. *Nature* 451(7176): 284–285.

Raynaud D, Jouzel J, Barnola JM, Chappellaz J, Delmas RJ, Lorius C (1993) The ice record of greenhouse gases. *Science* 259(5097): 926–934.

Reed BC (1989) Linear least-squares fits with errors in both coordinates. *American Journal of Physics* 57(7): 642–646. [Corrigendum: 1990 Vol. 58(2): 189]

Reed BC (1992) Linear least-squares fits with errors in both coordinates. II: Comments on parameter variances. *American Journal of Physics* 60(1): 59–62.

Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW, Bertrand CJH, Blackwell PG, Buck CE, Burr GS, Cutler KB, Damon PE, Edwards RL, Fairbanks RG, Friedrich M, Guilderson TP, Hogg AG, Hughen KA, Kromer B, McCormac G, Manning S, Ramsey CB, Reimer RW, Remmele S, Southon JR, Stuiver M, Talamo S, Taylor FW, van der Plicht J, Weyhenmeyer CE (2004) INTCAL04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3): 1029–1058.

Reinsel GC (2002) Trend analysis of upper stratospheric Umkehr ozone data for evidence of turnaround. *Geophysical Research Letters* 29(10): 1451. [doi:10.1029/2002GL014716]

Reinsel GC, Miller AJ, Weatherhead EC, Flynn LE, Nagatani RM, Tiao GC, Wuebbles DJ (2005) Trend analysis of total ozone data for turnaround and dynamical contributions. *Journal of Geophysical Research* 110(D16): D16306. [doi:10.1029/2004JD004662]

Reinsel GC, Weatherhead EC, Tiao GC, Miller AJ, Nagatani RM, Wuebbles DJ, Flynn LE (2002) On detection of turnaround and recovery in trend for ozone. *Journal of Geophysical Research* 107(D10): 4078. [doi:10.1029/2001JD000500]

Reis Jr DS, Stedinger JR (2005) Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology* 313(1–2): 97–116.

Reiss R-D, Thomas M (1997) *Statistical Analysis of Extreme Values.* Birkhäuser, Basel, 316 pp.

Resnick SI (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York, 320 pp.

Rimbu N, Lohmann G, Lorenz SJ, Kim JH, Schneider RR (2004) Holocene climate variability as derived from alkenone sea surface temperature and coupled ocean–atmosphere model experiments. *Climate Dynamics* 23(2): 215–227.

Rind D (2002) The Sun's role in climate variations. *Science* 296(5568): 673–677.

Ripley BD, Thompson M (1987) Regression techniques for the detection of analytical bias. *Analyst* 112(4): 377–383.

Ritson D (2004) Comment on "Global climate models violate scaling of the observed atmospheric variability." *Physical Review Letters* 92(15): 159803. [doi:10.1103/PhysRevLett.92.159803]

Roberts DH, Lehár J, Dreher JW (1987) Time series analysis with CLEAN. I. Derivation of a spectrum. *The Astronomical Journal* 93(4): 968–989.

Robinson PM (1977) Estimation of a time series model from unequally spaced data. *Stochastic Processes and their Applications* 6(1): 9–24.

Robinson PM (Ed) (2003) *Time Series with Long Memory.* Oxford University Press, Oxford, 382 pp.

Robock A (2000) Volcanic eruptions and climate. *Reviews of Geophysics* 38(2): 191–219.

Rodionov SN (2004) A sequential algorithm for testing climate regime shifts. *Geophysical Research Letters* 31(9): L09204. [doi:10.1029/2004GL019448]

Rodionov SN (2006) Use of prewhitening in climate regime shift detection. *Geophysical Research Letters* 33(12): L12707. [doi:10.1029/2006GL025904]

Rodó X, Baert E, Comín FA (1997) Variations in seasonal rainfall in southern Europe during the present century: Relationships with the North Atlantic Oscillation and the El Niño–Southern Oscillation. *Climate Dynamics* 13(4): 275–284.

Rodriguez RN (1982) Correlation. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 2. Wiley, New York, pp 193–204.

Rodriguez-Iturbe I, Rinaldo A (1997) *Fractal River Basins: Chance and Self-Organization.* Cambridge University Press, Cambridge, 547 pp.

Roe GH, Steig EJ (2004) Characterization of millennial-scale climate variability. *Journal of Climate* 17(10): 1929–1944.

Rohling EJ, Pälike H (2005) Centennial-scale climate cooling with a sudden cold event around 8,200 years ago. *Nature* 434(7036): 975–979.

Röthlisberger R, Bigler M, Hutterli M, Sommer S, Stauffer B, Junghans HG, Wagenbach D (2000) Technique for continuous high-resolution analysis of trace substances in firn and ice cores. *Environmental Science & Technology* 34(2): 338–342.

Röthlisberger R, Mudelsee M, Bigler M, de Angelis M, Fischer H, Hansson M, Lambert F, Masson-Delmotte V, Sime L, Udisti R, Wolff EW (2008) The southern hemisphere at glacial terminations: Insights from the Dome C ice core. *Climate of the Past* 4(4): 345–356.

Rothman DH (2001) Global biodiversity and the ancient carbon cycle. *Proceedings of the National Academy of Sciences of the United States of America* 98(8): 4305–4310.

Rothman DH (2002) Atmospheric carbon dioxide levels for the last 500 million years. *Proceedings of the National Academy of Sciences of the United States of America* 99(7): 4167–4171.

Rousseeuw PJ, Leroy AM (1987) *Robust Regression and Outlier Detection.* Wiley, New York, 329 pp.

Rubin DB (1976) Inference and missing data (with discussion). *Biometrika* 63(3): 581–592.

Ruddiman WF, Raymo ME (2003) A methane-based time scale for Vostok ice. *Quaternary Science Reviews* 22(2–4): 141–155.

Ruelle D (1990) Deterministic chaos: The science and the fiction. *Proceedings of the Royal Society of London, Series A* 427(1873): 241–248.

Ruiz NE, Vargas WM (1998) 500 hPa vorticity analyses over Argentina: Their climatology and capacity to distinguish synoptic-scale precipitation. *Theoretical and Applied Climatology* 60(1–4): 77–92.

Ruppert D, Carroll RJ (1980) Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75(372): 828–838.

Rust HW, Maraun D, Osborn TJ (2009) Modelling seasonality in extreme precipitation: A UK case study. *European Physical Journal Special Topics* 174(1): 99–111.

Rust HW, Mestre O, Venema VKC (2008) Fewer jumps, less memory: Homogenized temperature records and long memory. *Journal of Geophysical Research* 113(D19): D19110. [doi:10.1029/2008JD009919]

Rutherford S, D'Hondt S (2000) Early onset and tropical forcing of 100,000-year Pleistocene glacial cycles. *Nature* 408(6808): 72–75.

Rützel E (1976) Zur Ausgleichsrechnung: Die Unbrauchbarkeit von Linearisierungsmethoden beim Anpassen von Potenz- und Exponentialfunktionen. *Archiv für Psychologie* 128(3–4): 316–322.

Rybski D, Bunde A, Havlin S, von Storch H (2006) Long-term persistence in climate and the detection problem. *Geophysical Research Letters* 33(6): L06718. [doi:10.1029/2005GL025591]

Saltzman B (2002) *Dynamical Paleoclimatology: Generalized Theory of Global Climate Change.* Academic Press, San Diego, 354 pp.

Saltzman B, Verbitsky MY (1993) Multiple instabilities and modes of glacial rhythmicity in the Plio–Pleistocene: A general theory of late Cenozoic climatic change. *Climate Dynamics* 9(1): 1–15.

Sankarasubramanian A, Lall U (2003) Flood quantiles in a changing climate: Seasonal forecasts and causal relations. *Water Resources Research* 39(5): 1134. [doi:10.1029/2002WR001593]

Scafetta N (2008) Comment on "Heat capacity, time constant, and sensitivity of Earth's climate system" by S. E. Schwartz. *Journal of Geophysical Research* 113(D15): D15104. [doi:10.1029/2007JD009586]

Scafetta N, West BJ (2007) Phenomenological reconstructions of the solar signature in the northern hemisphere surface temperature records since 1600. *Journal of Geophysical Research* 112(D24): D24S03. [doi:10.1029/2007JD008437]

Scargle JD (1982) Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal* 263(2): 835–853.

Scargle JD (1989) Studies in astronomical time series analysis. III. Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. *The Astrophysical Journal* 343(2): 874–887.

Scargle JD (1997) Wavelet methods in astronomical time series analysis. In: Subba Rao T, Priestley MB, Lessi O (Eds) *Applications of Time Series Analysis in Astronomy and Meteorology.* Chapman and Hall, London, pp 226–248.

Schiffelbein P (1984) Effect of benthic mixing on the information content of deep-sea stratigraphical signals. *Nature* 311(5987): 651–653.

Schiffelbein P (1985) Extracting the benthic mixing impulse response function: A constrained deconvolution technique. *Marine Geology* 64(3–4): 313–336.

Schrage L (1979) A more portable Fortran random number generator. *ACM Transactions on Mathematical Software* 5(2): 132–138.

Schreiber T, Schmitz A (2000) Surrogate time series. *Physica D* 142(3–4): 346–382.

Schulz M (1996) *SPECTRUM und ENVELOPE: Computerprogramme zur Spektralanalyse nicht äquidistanter paläoklimatischer Zeitreihen.* Sonderforschungsbereich 313, University of Kiel, Kiel, 131 pp. [Report No. 65]

Schulz M (2002) On the 1470-year pacing of Dansgaard–Oeschger warm events. *Paleoceanography* 17(2): 1014. [doi:10.1029/2000PA000571]

Schulz M, Berger WH, Sarnthein M, Grootes PM (1999) Amplitude variations of 1470-year climate oscillations during the last 100,000 years linked to fluctuations of continental ice mass. *Geophysical Research Letters* 26(22): 3385–3388.

Schulz M, Mudelsee M (2002) REDFIT: Estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. *Computers and Geosciences* 28(3): 421–426.

Schulz M, Paul A (2002) Holocene climate variability on centennial-to-millennial time scales: 1. Climate records from the North-Atlantic realm. In: Wefer G, Berger W, Behre K-E, Jansen E (Eds) *Climate Development and History of the North Atlantic Realm.* Springer, Berlin, pp 41–54.

Schulz M, Stattegger K (1997) SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series. *Computers and Geosciences* 23(9): 929–945.

Schulze U (1987) *Mehrphasenregression.* Akademie-Verlag, Berlin, 178 pp.

Schuster A (1898) On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism* 3(1): 13–41.

Schuster A (1906) On the periodicities of sunspots. *Philosophical Transactions of the Royal Society of London, Series A* 206: 69–100.

Schwartz SE (2007) Heat capacity, time constant, and sensitivity of Earth's climate system. *Journal of Geophysical Research* 112(D24): D24S05. [doi:10.1029/2007JD008746]

Schwartz SE (2008) Reply to comments by G. Foster et al., R. Knutti et al., and N. Scafetta on "Heat capacity, time constant, and sensitivity of Earth's climate system." *Journal of Geophysical Research* 113(D15): D15105. [doi:10.1029/2008JD009872]

Schwarzacher W (1964) An application of statistical time-series analysis of a limestone–shale sequence. *Journal of Geology* 72(2): 195–213.

Schwarzacher W (1975) *Sedimentation Models and Quantitative Stratigraphy.* Elsevier, Amsterdam, 382 pp.

Schwarzacher W (1991) Milankovitch cycles and the measurement of time. In: Einsele G, Ricken W, Seilacher A (Eds) *Cycles and Events in Stratigraphy.* Springer, Berlin, pp 855–863.

Schwarzacher W (1993) *Cyclostratigraphy and the Milankovitch Theory.* Elsevier, Amsterdam, 225 pp.

Schwarzacher W (1994) Searching for long cycles in short sections. *Mathematical Geology* 26(7): 759–768.

Schweingruber FH (1988) *Tree Rings: Basics and Applications of Dendrochronology.* Kluwer, Dordrecht, 276 pp.

Scott DW (1979) On optimal and data-based histograms. *Biometrika* 66(3): 605–610.

Seber GAF, Wild CJ (1989) *Nonlinear Regression.* Wiley, New York, 768 pp.

Seibold E, Berger WH (1993) *The Sea Floor.* Second edition. Springer, Berlin, 356 pp.

Seidel DJ, Lanzante JR (2004) An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes. *Journal of Geophysical Research* 109(D14): D14108. [doi:10.1029/2003JD004414]

Seleshi Y, Demarée GR, Delleur JW (1994) Sunspot numbers as a possible indicator of annual rainfall at Addis Ababa, Ethiopia. *International Journal of Climatology* 14(8): 911–923.

Selley RC, Cocks LRM, Plimer IR (Eds) (2005) *Encyclopedia of Geology*, volume 1–5. Elsevier, Amsterdam, 3297 pp.

Sen A, Srivastava M (1990) *Regression Analysis: Theory, Methods, and Applications.* Springer, New York, 347 pp.

Sercl P, Stehlik J (2003) The August 2002 flood in the Czech Republic. *Geophysical Research Abstracts* 5: 12404.

Shackleton N (1967) Oxygen isotope analyses and Pleistocene temperatures reassessed. *Nature* 215(5096): 15–17.

Shackleton NJ (2000) The 100,000-year ice-age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* 289(5486): 1897–1902.

Shackleton NJ, Backman J, Zimmerman H, Kent DV, Hall MA, Roberts DG, Schnitker D, Baldauf JG, Desprairies A, Homrighausen R, Huddlestun P, Keene JB, Kaltenback AJ, Krumsiek KAO, Morton AC, Murray JW, Westberg-Smith J (1984) Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature* 307(5952): 620–623.

Shackleton NJ, Berger AL, Peltier WR (1990) An alternative astronomical calibration of the lower Pleistocene timescale based on ODP Site 677. *Transactions of the Royal Society of Edinburgh, Earth Sciences* 81(4): 251–261.

Shackleton NJ, Crowhurst S, Hagelberg T, Pisias NG, Schneider DA (1995a) A new late Neogene time scale: Application to Leg 138 sites. In: Pisias NG, Mayer LA, Janecek TR, Palmer-Julson A, van Andel TH (Eds) *Proc. ODP, Sci. Results*, volume 138. Ocean Drilling Program, College Station, TX, pp 73–101.

Shackleton NJ, Fairbanks RG, Chiu T-c, Parrenin F (2004) Absolute calibration of the Greenland time scale: Implications for Antarctic time scales and for Δ^{14}C. *Quaternary Science Reviews* 23(14–15): 1513–1522.

Shackleton NJ, Hall MA (1984) Oxygen and carbon isotope stratigraphy of Deep Sea Drilling Project hole 552a: Plio–Pleistocene glacial history. In: Roberts DG, Schnitker D, Backman J, Baldauf JG, Desprairies A, Homrighausen R, Huddlestun P, Kaltenback AJ, Krumsiek KAO, Morton AC, Murray JW, Westberg-Smith J, Zimmerman HB (Eds) *Init. Repts. DSDP*, volume 81. U.S. Govt. Printing Office, Washington, DC, pp 599–609.

Shackleton NJ, Hall MA, Pate D (1995b) Pliocene stable isotope stratigraphy of Site 846. In: Pisias NG, Mayer LA, Janecek TR, Palmer-Julson A, van Andel TH (Eds) *Proc. ODP, Sci. Results*, volume 138. Ocean Drilling Program, College Station, TX, pp 337–355.

Shaman P, Stine RA (1988) The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association* 83(403): 842–848.

Shapiro HS, Silverman RA (1960) Alias-free sampling of random noise. *Journal of the Society for Industrial and Applied Mathematics* 8(2): 225–248.

Shenton LR, Johnson WL (1965) Moments of a serial correlation coefficient. *Journal of the Royal Statistical Society, Series B* 27(2): 308–320.

Sherman M, Speed Jr FM, Speed FM (1998) Analysis of tidal data via the blockwise bootstrap. *Journal of Applied Statistics* 25(3): 333–340.

Shumway RH, Stoffer DS (2006) *Time Series Analysis and Its Applications: With R Examples.* Second edition. Springer, New York, 575 pp.

Siegel AF (1980) Testing for periodicity in a time series. *Journal of the American Statistical Association* 75(370): 345–348.

Siegenthaler U, Stocker TF, Monnin E, Lüthi D, Schwander J, Stauffer B, Raynaud D, Barnola J-M, Fischer H, Masson-Delmotte V, Jouzel J (2005) Stable carbon cycle–climate relationship during the late Pleistocene. *Science* 310(5752): 1313–1317.

Sievers W (1996) Standard and bootstrap confidence intervals for the correlation coefficient. *British Journal of Mathematical and Statistical Psychology* 49(2): 381–396.

Silverman BW (1982) Kernel density estimation using the Fast Fourier Transform. *Applied Statistics* 31(1): 93–99.

Silverman BW (1986) *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, 175 pp.

Silverman BW (1999) Wavelets in statistics: Beyond the standard assumptions. *Philosophical Transactions of the Royal Society of London, Series A* 357(1760): 2459–2473.

Silverman BW, Young GA (1987) The bootstrap: To smooth or not to smooth? *Biometrika* 74(3): 469–479.

Simonoff JS (1996) *Smoothing Methods in Statistics.* Springer, New York, 338 pp.

Singer BS, Pringle MS (1996) Age and duration of the Matuyama–Brunhes geomagnetic polarity reversal from ^{40}Ar/^{39}Ar incremental heating analyses of lavas. *Earth and Planetary Science Letters* 139(1–2): 47–61.

Singh K (1981) On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics* 9(6): 1187–1195.

Slepian D (1978) Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case. *Bell System Technical Journal* 57(5): 1371–1430.

Smith AFM (1975) A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* 62(2): 407–416.

Smith RL (1985) Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 72(1): 67–90.

Smith RL (1987) Estimating tails of probability distributions. *The Annals of Statistics* 15(3): 1174–1207.

Smith RL (1989) Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statistical Science* 4(4): 367–393.

Smith RL (2004) Statistics of extremes, with applications in environment, insurance, and finance. In: Finkenstädt B, Rootzén H (Eds) *Extreme Values in Finance, Telecommunications, and the Environment.* Chapman and Hall, Boca Raton, FL, pp 1–78.

Smith RL, Shively TS (1994) *A Point Process Approach to Modeling Trends in Tropospheric Ozone Based on Exceedances of a High Threshold.* National Institute of Statistical Sciences, Research Triangle Park, NC, 20 pp. [Technical Report Number 16]

Smith RL, Shively TS (1995) Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold. *Atmospheric Environment* 29(23): 3489–3499.

Smith RL, Tawn JA, Coles SG (1997) Markov chain models for threshold exceedances. *Biometrika* 84(2): 249–268.

Smith RL, Tebaldi C, Nychka D, Mearns LO (2009) Bayesian modeling of uncertainty in ensembles of climate models. *Journal of the American Statistical Association* 104(485): 97–116.

Sokal A, Bricmont J (1998) *Intellectual Impostures*. Profile Books, London, 274 pp.

Solanki SK, Usoskin IG, Kromer B, Schüssler M, Beer J (2004) Unusual activity of the Sun during recent decades compared to the previous 11,000 years. *Nature* 431(7012): 1084–1087.

Solomon S, Qin D, Manning M, Marquis M, Averyt K, Tignor MMB, Miller Jr HL, Chen Z (Eds) (2007) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 996 pp.

Solow AR (1987) Testing for climate change: An application of the two-phase regression model. *Journal of Climate and Applied Meteorology* 26(10): 1401–1405.

Solow AR (1991) An exploratory analysis of the occurrence of explosive volcanism in the northern hemisphere, 1851–1985. *Journal of the American Statistical Association* 86(413): 49–54.

Spall JC (Ed) (1988) *Bayesian Analysis of Time Series and Dynamic Models*. Marcel Dekker, New York, 536 pp.

Spearman C (1904) The proof and measurement of association between two things. *American Journal of Psychology* 15(1): 72–101.

Spearman C (1906) 'Footrule' for measuring correlation. *British Journal of Psychology* 2(1): 89–108.

Spötl C, Mangini A, Richards DA (2006) Chronology and paleoenvironment of Marine Isotope Stage 3 from two high-elevation speleothems, Austrian Alps. *Quaternary Science Reviews* 25(9–10): 1127–1136.

Squire PT (1990) Comment on "Linear least-squares fits with errors in both coordinates," by B. C. Reed [Am. J. Phys. 57, 642–646 (1989)]. *American Journal of Physics* 58(12): 1209.

Stainforth DA, Allen MR, Tredger ER, Smith LA (2007) Confidence, uncertainty and decision-support relevance in climate predictions. *Philosophical Transactions of the Royal Society of London, Series A* 365(1857): 2145–2161.

Stanley SM (1989) *Earth and life through time*. Second edition. Freeman, New York, 689 pp.

Stattegger K (1986) Die Beziehungen zwischen Sediment und Hinterland: Mathematisch-statistische Modelle aus Schwermineraldaten rezenter fluviatiler und fossiler Sedimente. *Jahrbuch der Geologischen Bundesanstalt* 128(3–4): 449–512.

Stedinger JR, Crainiceanu CM (2001) Climate variability and flood-risk analysis. In: Haimes YY, Moser DA, Stakhiv EZ (Eds) *Risk-Based Decision Making in Water Resources IX*. American Society of Civil Engineers, Reston, VA, pp 77–86.

Steele JH, Thorpe SA, Turekian KK (Eds) (2001) *Encyclopedia of Ocean Sciences*, volume 1–6. Academic Press, San Diego, 3399 pp.

Steffensen JP, Andersen KK, Bigler M, Clausen HB, Dahl-Jensen D, Fischer H, Goto-Azuma K, Hansson M, Johnsen SJ, Jouzel J, Masson-Delmotte V, Popp T, Rasmussen SO, Röthlisberger R, Ruth U, Stauffer B, Siggaard-Andersen M-L, Sveinbjörnsdóttir ÁE, Svensson A, White JWC (2008) High-resolution Greenland ice core data show abrupt climate change happens in few years. *Science* 321(5889): 680–684.

Stensrud DJ (2007) *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*. Cambridge University Press, Cambridge, 459 pp.

Stephenson DB, Pavan V, Bojariu R (2000) Is the North Atlantic Oscillation a random walk? *International Journal of Climatology* 20(1): 1–18.

Stern DI, Kaufmann RK (1999) Econometric analysis of global climate change. *Environmental Modelling and Software* 14(6): 597–605.

Stern DI, Kaufmann RK (2000) Detecting a global warming signal in hemispheric temperature series: A structural time series analysis. *Climatic Change* 47(4): 411–438.

Stine RA (1987) Estimating properties of autoregressive forecasts. *Journal of the American statistical association* 82(400): 1072–1078.

Stine RA (1997) Nonlinear time series. In: Kotz S, Read CB, Banks DL (Eds) *Encyclopedia of statistical sciences*, volume U1. Wiley, New York, pp 430–437.

Storey JD (2007) The optimal discovery procedure: A new approach to simultaneous significance testing. *Journal of the Royal Statistical Society, Series B* 69(3): 347–368.

Stott PA, Tett SFB, Jones GS, Allen MR, Mitchell JFB, Jenkins GJ (2000) External control of 20th century temperature by natural and anthropogenic forcings. *Science* 290(5499): 2133–2137.

Strupczewski WG, Kaczmarek Z (2001) Non-stationary approach to at-site flood frequency modelling II. Weighed least squares estimation. *Journal of Hydrology* 248(1–4): 143–151.

Strupczewski WG, Singh VP, Feluch W (2001a) Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. *Journal of Hydrology* 248(1-4): 123–142.

Strupczewski WG, Singh VP, Mitosek HT (2001b) Non-stationary approach to at-site flood frequency modelling. III. Flood analysis of Polish rivers. *Journal of Hydrology* 248(1–4): 152–167.

Stuart A (1983) Kendall's tau. In: Kotz S, Johnson NL, Read CB (Eds) *Encyclopedia of statistical sciences*, volume 4. Wiley, New York, pp 367–369.

Stuiver M, Braziunas TF (1993) Sun, ocean, climate and atmospheric $^{14}CO_2$: An evaluation of causal and spectral relationships. *The Holocene* 3(4): 289–305.

Stuiver M, Reimer PJ, Bard E, Beck JW, Burr GS, Hughen KA, Kromer B, McCormac G, van der Plicht J, Spurk M (1998) INTCAL98 radiocarbon age calibration, 24,000–0 cal BP. *Radiocarbon* 40(3): 1041–1083.

Subba Rao T, Gabr MM (1984) *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer, New York, 280 pp.

Suess HE (1965) Secular variations of the cosmic-ray-produced carbon 14 in the atmosphere and their interpretations. *Journal of Geophysical Research* 70(23): 5937–5952.

Suess HE, Linick TW (1990) The ^{14}C record in bristlecone pine wood of the last 8000 years based on the dendrochronology of the late C. W. Ferguson. *Philosophical Transactions of the Royal Society of London, Series A* 330(1615): 403–412.

Sura P, Newman M, Penland C, Sardeshmukh P (2005) Multiplicative noise and non-Gaussianity: A paradigm for atmospheric regimes? *Journal of the Atmospheric Sciences* 62(5): 1391–1409.

Svensmark H, Friis-Christensen E (1997) Variation of cosmic ray flux and global cloud coverage—a missing link in solar–climate relationships. *Journal of Atmospheric and Solar-Terrestrial Physics* 59(11): 1225–1232.

Sweldens W, Schröder P (2000) Building your own wavelets at home. In: Klees R, Haagmans R (Eds) *Wavelets in the Geosciences*. Springer, Berlin, pp 72–130.

Tachikawa K, Vidal L, Sonzogni C, Bard E (2009) Glacial/interglacial sea surface temperature changes in the southwest Pacific over the past 360 ka. *Quaternary Science Reviews* 28(13–14): 1160–1170.

Talkner P, Weber RO (2000) Power spectrum and detrended fluctuation analysis: Application to daily temperatures. *Physical Review E* 62(1): 150–160.

Tate RF (1954) Correlation between a discrete and a continuous variable. Point-biserial correlation. *Annals of Mathematical Statistics* 25(3): 603–607.

Taylor RE (1987) *Radiocarbon Dating: An Archaeological Perspective.* Academic Press, Orlando, FL, 212 pp.

Tebaldi C, Sansó B (2009) Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayesian approach. *Journal of the Royal Statistical Society, Series A* 172(1): 83–106.

Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD (1992) Testing for nonlinearity in time series: The method of surrogate data. *Physica D* 58(1–4): 77–94.

Thiébaux HJ, Zwiers FW (1984) The interpretation and estimation of effective sample size. *Journal of Climate and Applied Meteorology* 23(5): 800–811.

Thompson DWJ, Kennedy JJ, Wallace JM, Jones PD (2008) A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. *Nature* 453(7195): 646–649.

Thomson DJ (1982) Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70(9): 1055–1096.

Thomson DJ (1990a) Quadratic-inverse spectrum estimates: Applications to palaeoclimatology. *Philosophical Transactions of the Royal Society of London, Series A* 332(1627): 539–597.

Thomson DJ (1990b) Time series analysis of Holocene climate data. *Philosophical Transactions of the Royal Society of London, Series A* 330(1615): 601–616.

Thomson DJ (1997) Dependence of global temperatures on atmospheric CO_2 and solar irradiance. *Proceedings of the National Academy of Sciences of the United States of America* 94(16): 8370–8377.

Thomson DJ, Chave AD (1991) Jackknifed error estimates for spectra, coherences, and transfer functions. In: Haykin S (Ed) *Advances in Spectrum Analysis and Array Processing*, volume 1. Prentice-Hall, Englewood Cliffs, NJ, pp 58–113.

Thomson J, Cook GT, Anderson R, MacKenzie AB, Harkness DD, McCave IN (1995) Radiocarbon age offsets in different-sized carbonate components of deep-sea sediments. *Radiocarbon* 37(2): 91–101.

Thomson PJ, Robinson PM (1996) Estimation of second-order properties from jittered time series. *Annals of the Institute of Statistical Mathematics* 48(1): 29–48.

Thywissen K (2006) *Components of Risk: A Comparative Glossary.* United Nations University, Institute for Environment and Human Security, Bonn, 48 pp. [Studies of the University: Research, Counsel, Education No. 2]

Tjøstheim D, Paulsen J (1983) Bias of some commonly-used time series estimates. *Biometrika* 70(2): 389–399.

Tol RSJ, de Vos AF (1993) Greenhouse statistics—time series analysis. *Theoretical and Applied Climatology* 48(2–3): 63–74.

Tol RSJ, de Vos AF (1998) A Bayesian statistical analysis of the enhanced greenhouse effect. *Climatic Change* 38(1): 87–112.

Tomé AR, Miranda PMA (2004) Piecewise linear fitting and trend changing points of climate parameters. *Geophysical Research Letters* 31(2): L02207. [doi:10.1029/2003GL019100]

Tomé AR, Miranda PMA (2005) Continuous partial trends and low-frequency oscillations of time series. *Nonlinear Processes in Geophysics* 12(4): 451–460.

Tong H (1990) *Non-linear Time Series.* Clarendon Press, Oxford, 564 pp.

Tong H (1992) Some comments on a bridge between nonlinear dynamicists and statisticians. *Physica D* 58(1–4): 299–303.

Tong H (1995) A personal overview of non-linear time series analysis from a chaos perspective (with discussion). *Scandinavian Journal of Statistics* 22(4): 399–445.

Tong H, Lim KS (1980) Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society, Series B* 42(3): 245–292.

Tong H, Yeung I (1991) Threshold autoregressive modelling in continuous time. *Statistica Sinica* 1(2): 411–430.

Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* 79(1): 61–78.

Trauth MH (1998) TURBO: A dynamic-probabilistic simulation to study the effects of bioturbation on paleoceanographic time series. *Computers and Geosciences* 24(5): 433–441.

Trauth MH (2007) *MATLAB® Recipes for Earth Sciences.* Second edition. Springer, Berlin, 288 pp.

Trauth MH, Larrasoaña JC, Mudelsee M (2009) Trends, rhythms and events in Plio–Pleistocene African climate. *Quaternary Science Reviews* 28(5–6): 399–411.

Traverse A (2007) *Paleopalynology.* Second edition. Springer, Dordrecht, 813 pp.

Trenberth KE (1984a) Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelations. *Monthly Weather Review* 112(12): 2359–2368.

Trenberth KE (1984b) Some effects of finite sample size and persistence on meteorological statistics. Part II: Potential predictability. *Monthly Weather Review* 112(12): 2369–2379.

Triacca U (2001) On the use of Granger causality to investigate the human influence on climate. *Theoretical and Applied Climatology* 69(3–4): 137–138.

Triacca U (2007) Granger causality and contiguity between stochastic processes. *Physics Letters A* 362(4): 252–255.

Tsay RS (1988) Outliers, level shifts, and variance changes in time series. *Journal of Forecasting* 7(1): 1–20.

Tsonis AA, Elsner JB (1995) Testing for scaling in natural forms and observables. *Journal of Statistical Physics* 81(5–6): 869–880.

Tsonis AA, Elsner JB (Eds) (2007) *Nonlinear Dynamics in Geosciences.* Springer, New York, 604 pp.

Tukey JW (1977) *Exploratory Data Analysis.* Addison-Wesley, Reading, MA, 688 pp.

Udelhofen PM, Cess RD (2001) Cloud cover variations over the United States: An influence of cosmic rays or solar variability? *Geophysical Research Letters* 28(13): 2617–2620.

Ulbrich U, Brücher T, Fink AH, Leckebusch GC, Krüger A, Pinto JG (2003a) The central European floods of August 2002: Part 1 – Rainfall periods and flood development. *Weather* 58(10): 371–377.

Ulbrich U, Brücher T, Fink AH, Leckebusch GC, Krüger A, Pinto JG (2003b) The central European floods of August 2002: Part 2 – Synoptic causes and considerations with respect to climatic change. *Weather* 58(11): 434–442.

Urban FE, Cole JE, Overpeck JT (2000) Influence of mean climate change on climate variability from a 155-year tropical Pacific coral record. *Nature* 407(6807): 989–993.

Usoskin IG, Marsh N, Kovaltsov GA, Mursula K, Gladysheva OG (2004) Latitudinal dependence of low cloud amount on cosmic ray induced ionization. *Geophysical Research Letters* 31(16): L16109. [doi:10.1029/2004GL019507]

van der Linden P, Mitchell JFB (Eds) (2009) *ENSEMBLES: Climate change and its impacts at seasonal, decadal and centennial timescales.* Met Office Hadley Centre, Exeter, 160 pp.

van de Wiel MA, Di Bucchianico A (2001) Fast computation of the exact null distribution of Spearman's ρ and Page's L statistic for samples with and without ties. *Journal of Statistical Planning and Inference* 92(1–2): 133–145.

VanDongen HPA, Olofsen E, VanHartevelt JH, Kruyt EW (1997) *Periodogram analysis of unequally spaced data: The Lomb method.* Leiden University, Leiden, 66 pp. [ISBN 9080385115]

Van Dongen HPA, Olofsen E, VanHartevelt JH, Kruyt EW (1999) A procedure of multiple period searching in unequally spaced time-series with the Lomb–Scargle method. *Biological Rhythm Research* 30(2): 149–177.

Van Montfort MAJ, Witter JV (1985) Testing exponentiality against generalised Pareto distribution. *Journal of Hydrology* 78(3–4): 305–315.

Vecchi GA, Knutson TR (2008) On estimates of historical North Atlantic tropical cyclone activity. *Journal of Climate* 21(14): 3580–3600.

Verdes PF (2005) Assessing causality from multivariate time series. *Physical Review E* 72(3): 026222. [doi:10.1103/PhysRevE.72.026222]

Vidakovic B (1999) *Statistical Modeling by Wavelets.* Wiley, New York, 382 pp.

von Storch H, Zwiers FW (1999) *Statistical Analysis in Climate Research.* Cambridge University Press, Cambridge, 484 pp.

von Weizsäcker CF (1985) *Aufbau der Physik.* Deutscher Taschenbuch Verlag, Munich, 662 pp.

Vyushin D, Bunde A, Brenner S, Havlin S, Govindan RB, Schellnhuber H-J (2004) Vjushin et al. reply. *Physical Review Letters* 92(15): 159804. [doi:10.1103/PhysRevLett.92.159804]

WAFO group (2000) *WAFO: A Matlab Toolbox for Analysis of Random Waves and Loads.* Lund Institute of Technology, Lund University, Lund, 111 pp.

Wagenbach D (1989) Environmental records in Alpine glaciers. In: Oeschger H, Langway Jr CC (Eds) *The Environmental Record in Glaciers and Ice Sheets.* Wiley, Chichester, pp 69–83.

Wagenbach D, Preunkert S, Schäfer J, Jung W, Tomadin L (1996) Northward transport of Saharan dust recorded in a deep Alpine ice core. In: Guerzoni S, Chester R (Eds) *The Impact of Desert Dust Across the Mediterranean.* Kluwer, Dordrecht, pp 291–300.

Wald A (1940) The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11(3): 284–300.

Walden AT (1992) Asymptotic percentage points for Siegel's test statistic for compound periodicities. *Biometrika* 79(2): 438–440.

Walker GT (1914) Correlation in seasonal variations of weather, III. On the criterion for the reality of relationships or periodicities. *Memoirs of the Indian Meteorological Department* 21(9): 13–15.

Walker M (2005) *Quaternary Dating Methods.* Wiley, Chichester, 286 pp.

Wand MP, Jones MC (1995) *Kernel Smoothing.* Chapman and Hall, London, 212 pp.

Wanner H, Beer J, Bütikofer J, Crowley TJ, Cubasch U, Flückiger J, Goosse H, Grosjean M, Joos F, Kaplan JO, Küttel M, Müller SA, Prentice IC, Solomina O, Stocker TF, Tarasov P, Wagner M, Widmann M (2008) Mid- to late Holocene climate change: An overview. *Quaternary Science Reviews* 27(19–20): 1791–1828.

Wasserman L (2004) *All of Statistics: A Concise Course in Statistical Inference.* Springer, New York, 442 pp.

Wasserman L (2006) *All of Nonparametric Statistics.* Springer, New York, 268 pp.

Weedon GP (2003) *Time-Series Analysis and Cyclostratigraphy.* Cambridge University Press, Cambridge, 259 pp.

Weikinn C (1958) *Quellentexte zur Witterungsgeschichte Europas von der Zeitwende bis zum Jahre 1850: Hydrographie, Teil 1 (Zeitwende–1500).* Akademie-Verlag, Berlin, 531 pp.

Weikinn C (1960) *Quellentexte zur Witterungsgeschichte Europas von der Zeitwende bis zum Jahre 1850: Hydrographie, Teil 2 (1501–1600).* Akademie-Verlag, Berlin, 486 pp.

Weikinn C (1961) *Quellentexte zur Witterungsgeschichte Europas von der Zeitwende bis zum Jahre 1850: Hydrographie, Teil 3 (1601–1700).* Akademie-Verlag, Berlin, 586 pp.

Weikinn C (1963) *Quellentexte zur Witterungsgeschichte Europas von der Zeitwende bis zum Jahre 1850: Hydrographie, Teil 4 (1701–1750).* Akademie-Verlag, Berlin, 381 pp.

Weikinn C (2000) *Quellentexte zur Witterungsgeschichte Europas von der Zeitwende bis zum Jahr 1850: Hydrographie, Teil 5 (1751–1800).* Gebrüder Borntraeger, Berlin, 674 pp. [Börngen M, Tetzlaff G (Eds)]

Weikinn C (2002) *Quellentexte zur Witterungsgeschichte Europas von der Zeitwende bis zum Jahr 1850: Hydrographie, Teil 6 (1801–1850).* Gebrüder Borntraeger, Berlin, 728 pp. [Börngen M, Tetzlaff G (Eds)]

Welch PD (1967) The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* 15(2): 70–73.

White JS (1961) Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika* 48(1–2): 85–94.

Whittle P (1952) The simultaneous estimation of a time series harmonic components and covariance structure. *Trabajos de Estadística* 3(1–2): 43–57.

Wigley TML, Santer BD, Taylor KE (2000) Correlation approaches to detection. *Geophysical Research Letters* 27(18): 2973–2976.

Wilks DS (1995) *Statistical Methods in the Atmospheric Sciences.* Academic Press, San Diego, 467 pp.

Wilks DS (1997) Resampling hypothesis tests for autocorrelated fields. *Journal of Climate* 10(1): 65–82.

Wilks DS (2006) *Statistical Methods in the Atmospheric Sciences.* Second edition. Elsevier, Amsterdam, 627 pp.

Williams DA (1970) Discrimination between regression models to determine the pattern of enzyme synthesis in synchronous cell cultures. *Biometrics* 26(1): 23–32.

Willson RC, Hudson HS (1988) Solar luminosity variations in solar cycle 21. *Nature* 332(6167): 810–812.

Wilson RM (1997) Comment on "Downward trends in the frequency of intense Atlantic hurricanes during the past 5 decades" by C. W. Landsea et al. *Geophysical Research Letters* 24(17): 2203–2204.

Witt A, Schumann AY (2005) Holocene climate variability on millennial scales recorded in Greenland ice cores. *Nonlinear Processes in Geophysics* 12(3): 345–352.

Witte HJL, Coope GR, Lemdahl G, Lowe JJ (1998) Regression coefficients of thermal gradients in northwestern Europe during the last glacial–Holocene transition using beetle MCR data. *Journal of Quaternary Science* 13(5): 435–445.

Wolff E, Kull C, Chappellaz J, Fischer H, Miller H, Stocker TF, Watson AJ, Flower B, Joos F, Köhler P, Matsumoto K, Monnin E, Mudelsee M, Paillard D, Shackleton N (2005) Modeling past atmospheric CO_2: Results of a challenge. *Eos, Transactions of the American Geophysical Union* 86(38): 341, 345.

Wolff EW, Fischer H, Röthlisberger R (2009) Glacial terminations as southern warmings without northern control. *Nature Geoscience* 2(3): 206–209.

Woods TN, Lean J (2007) Anticipating the next decade of Sun–Earth system variations. *Eos, Transactions of the American Geophysical Union* 88(44): 457–458.

Worsley KJ (1986) Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika* 73(1): 91–104.

Wu CFJ (1986) Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *The Annals of Statistics* 14(4): 1261–1350.

Wu P, Wood R, Stott P (2005) Human influence on increasing Arctic river discharges. *Geophysical Research Letters* 32(2): L02703. [doi:10.1029/2004GL021570]

Wu WB, Zhao Z (2007) Inference of trends in time series. *Journal of the Royal Statistical Society, Series B* 69(3): 391–410.

Wu Y (2005) *Inference for Change-Point and Post-Change Means After a CUSUM Test*. Springer, New York, 158 pp.

Wunsch C (2000) On sharp spectral lines in the climate record and the millennial peak. *Paleoceanography* 15(4): 417–424.

Wunsch C (2001) Reply. *Paleoceanography* 16(5): 548.

Wunsch C (2003) The spectral description of climate change including the 100 ky energy. *Climate Dynamics* 20(4): 353–363.

Wunsch C (2006) *Discrete Inverse and State Estimation Problems*. Cambridge University Press, Cambridge, 371 pp.

Wunsch C, Gunn DE (2003) A densely sampled core and climate variable aliasing. *Geo-Marine Letters* 23(1): 64–71.

Yamamoto R, Iwashima T, Sanga NK, Hoshiai M (1986) An analysis of climatic jump. *Journal of the Meteorological Society of Japan* 64(2): 273–281.

Yashchin E (1995) Estimating the current mean of a process subject to abrupt changes. *Technometrics* 37(3): 311–323.

Yee TW, Wild CJ (1996) Vector generalized additive models. *Journal of the Royal Statistical Society, Series B* 58(3): 481–493.

Yiou P, Ribereau P, Naveau P, Nogaj M, Brázdil R (2006) Statistical analysis of floods in Bohemia (Czech Republic) since 1825. *Hydrological Sciences Journal* 51(5): 930–945.

York D (1966) Least-squares fitting of a straight line. *Canadian Journal of Physics* 44(5): 1079–1086.

York D (1967) The best isochron. *Earth and Planetary Science Letters* 2(5): 479–482.

York D (1969) Least squares fitting of a straight line with correlated errors. *Earth and Planetary Science Letters* 5(5): 320–324.

Young GA (1988) A note on bootstrapping the correlation coefficient. *Biometrika* 75(2): 370–373.

Yu K, Lu Z, Stander J (2003) Quantile regression: Applications and current research areas. *The Statistician* 52(3): 331–350.

Yule GU (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London, Series A* 226: 267–298.

Zalasiewicz J, Williams M, Smith A, Barry TL, Coe AL, Brown PR, Brenchley P, Cantrill D, Gale A, Gibbard P, Gregory FJ, Hounslow MW, Kerr AC, Pearson P, Knox R, Powell J, Waters C, Marshall J, Oates M, Rawson P, Stone P (2008) Are we now living in the Anthropocene? *GSA Today* 18(2): 4–8.

Zar JH (1978) Approximations for the percentage points of the chi-squared distribution. *Applied Statistics* 27(3): 280–290.

Zeileis A, Leisch F, Hornik K, Kleiber C (2002) strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software* 7(2): 1–38.

Zhang X, Zwiers FW, Li G (2004) Monte Carlo experiments on the detection of trends in extreme values. *Journal of Climate* 17(10): 1945–1952.

Zheng Zg, Yang Y (1998) Cross-validation and median criterion. *Statistica Sinica* 8(3): 907–921.

Zielinski GA, Mayewski PA, Meeker LD, Whitlow S, Twickler MS (1996) A 110,000-yr record of explosive volcanism from the GISP2 (Greenland) ice core. *Quaternary Research* 45(2): 109–118.

Zielinski GA, Mayewski PA, Meeker LD, Whitlow S, Twickler MS, Morrison M, Meese DA, Gow AJ, Alley RB (1994) Record of volcanism since 7000 B.C. from the GISP2 Greenland ice core and implications for the volcano-climate system. *Science* 264(5161): 948–952.

Zolitschka B (Ed) (1999) *High-resolution records from European lakes*, volume 18(7) of *Quaternary Science Reviews*. [Special issue]

Zou GY (2007) Toward using confidence intervals to compare correlations. *Psychological Methods* 12(4): 399–413.

Zwiers FW, von Storch H (1995) Taking serial correlation into account in tests of the mean. *Journal of Climate* 8(2): 336–351.

Subject Index

D

Author Index