

Climatology is, to a large degree, the study of the statistics of our climate. The powerful tools of mathematical statistics therefore find wide application in climatological research, ranging from simple methods for determining the uncertainty of a climatological mean to sophisticated techniques which reveal the dynamics of the climate system.

The purpose of this book is to help the climatologist understand the basic precepts of the statistician's art and to provide some of the background needed to apply statistical methodology correctly and usefully. The book is self contained: introductory material, standard advanced techniques, and the specialized techniques used specifically by climatologists are all contained within this one source. There is a wealth of real-world examples drawn from the climate literature to demonstrate the need, power and pitfalls of statistical analysis in climate research.

This book is suitable as a main text for graduate courses on statistics for climatic, atmospheric and oceanic science. It will also be valuable as a reference source for researchers in climatology, meteorology, atmospheric science, and oceanography.

Hans von Storch is Director of the Institute of Hydrophysics of the GKSS Research Centre in Geesthacht, Germany and a Professor at the Meteorological Institute of the University of Hamburg.

Francis W. Zwiers is Chief of the Canadian Centre for Climate Modelling and Analysis, Atmospheric Environment Service, Victoria, Canada, and an Adjunct Professor of the Department of Mathematics and Statistics of the University of Victoria.

Cambridge University Press
0521450713 - Statistical Analysis in Climate Research
Hans von Storch and Francis W. Zwiers
Frontmatter
[More information](#)

Statistical Analysis in Climate Research

Hans von Storch
and Francis W. Zwiers



Cambridge University Press
0521450713 - Statistical Analysis in Climate Research
Hans von Storch and Francis W. Zwiers
Frontmatter
[More information](#)

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK <http://www.cup.cam.ac.uk>
40 West 20th Street, New York, NY 10011-4211, USA <http://www.cup.org>
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1999

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1999

Printed in the United Kingdom at the University Press, Cambridge

Typeset in Times 10/12pt [DBD]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Storch, H. V. (Hans von), 1949–
Statistical analysis in climate research / Hans von Storch and
Francis W. Zwiers.
p. cm.
Includes index.
ISBN 0 521 45071 3
1. Climatology – Statistical methods. I. Title.
QC981.S735 1998
551.5'072–dc21 98-17416 CIP

ISBN 0 521 45071 3 hardback

Contents

Preface	ix
Thanks	x
1 Introduction	1
1.1 The Statistical Description	1
1.2 Some Typical Problems and Concepts	2
I Fundamentals	17
2 Probability Theory	19
2.1 Introduction	19
2.2 Probability	20
2.3 Discrete Random Variables	21
2.4 Examples of Discrete Random Variables	23
2.5 Discrete Multivariate Distributions	26
2.6 Continuous Random Variables	29
2.7 Example of Continuous Random Variables	33
2.8 Random Vectors	38
2.9 Extreme Value Distributions	45
3 Distributions of Climate Variables	51
3.1 Atmospheric Variables	52
3.2 Some Other Climate Variables	63
4 Concepts in Statistical Inference	69
4.1 General	69
4.2 Random Samples	74
4.3 Statistics and Sampling Distributions	76
5 Estimation	79
5.1 General	79
5.2 Examples of Estimators	80
5.3 Properties of Estimators	84
5.4 Interval Estimators	90
5.5 Bootstrapping	93
II Confirmation and Analysis	95
Overview	97

6	The Statistical Test of a Hypothesis	99
6.1	The Concept of Statistical Tests	99
6.2	The Structure and Terminology of a Test	100
6.3	Monte Carlo Simulation	104
6.4	On Establishing Statistical Significance	106
6.5	Multivariate Problems	108
6.6	Tests of the Mean	111
6.7	Test of Variances	118
6.8	Field Significance Tests	121
6.9	Univariate Recurrence Analysis	122
6.10	Multivariate Recurrence Analysis	126
7	Analysis of Atmospheric Circulation Problems	129
7.1	Validating a General Circulation Model	129
7.2	Analysis of a GCM Sensitivity Experiment	131
7.3	Identification of a Signal in Observed Data	133
7.4	Detecting the 'CO ₂ Signal'	136
III	Fitting Statistical Models	141
	Overview	143
8	Regression	145
8.1	Introduction	145
8.2	Correlation	146
8.3	Fitting and Diagnosing Simple Regression Models	150
8.4	Multiple Regression	160
8.5	Model Selection	166
8.6	Some Other Topics	168
9	Analysis of Variance	171
9.1	Introduction	171
9.2	One Way Analysis of Variance	173
9.3	Two Way Analysis of Variance	181
9.4	Two Way ANOVA with Mixed Effects	184
9.5	Tuning a Basin Scale Ocean Model	191
IV	Time Series	193
	Overview	195
10	Time Series and Stochastic Processes	197
10.1	General Discussion	197
10.2	Basic Definitions and Examples	199
10.3	Auto-regressive Processes	203
10.4	Stochastic Climate Models	211
10.5	Moving Average Processes	213
11	Parameters of Univariate and Bivariate Time Series	217
11.1	The Auto-covariance Function	217
11.2	The Spectrum	222
11.3	The Cross-covariance Function	228
11.4	The Cross-spectrum	234
11.5	Frequency–Wavenumber Analysis	241

<i>CONTENTS</i>	vii
12 Estimating Covariance Functions and Spectra	251
12.1 Non-parametric Estimation of the Auto-correlation Function	252
12.2 Identifying and Fitting Auto-regressive Models	255
12.3 Estimating the Spectrum	263
12.4 Estimating the Cross-correlation Function	281
12.5 Estimating the Cross-spectrum	282
V Eigen Techniques	289
Overview	291
13 Empirical Orthogonal Functions	293
13.1 Definition of Empirical Orthogonal Functions	294
13.2 Estimation of Empirical Orthogonal Functions	299
13.3 Inference	301
13.4 Examples	304
13.5 Rotation of EOFs	305
13.6 Singular Systems Analysis	312
14 Canonical Correlation Analysis	317
14.1 Definition of Canonical Correlation Patterns	317
14.2 Estimating Canonical Correlation Patterns	322
14.3 Examples	323
14.4 Redundancy Analysis	327
15 POP Analysis	335
15.1 Principal Oscillation Patterns	335
15.2 Examples	339
15.3 POPs as a Predictive Tool	345
15.4 Cyclo-stationary POP Analysis	346
15.5 State Space Models	350
16 Complex Eigentechniques	353
16.1 Introduction	353
16.2 Hilbert Transform	353
16.3 Complex and Hilbert EOFs	357
VI Other Topics	367
Overview	369
17 Specific Statistical Concepts in Climate Research	371
17.1 The Decorrelation Time	371
17.2 Potential Predictability	374
17.3 Composites and Associated Correlation Patterns	378
17.4 Teleconnections	382
17.5 Time Filters	384
18 Forecast Quality Evaluation	391
18.1 The Skill of Categorical Forecasts	392
18.2 The Skill of Quantitative Forecasts	395
18.3 The Murphy–Epstein Decomposition	399
18.4 Issues in the Evaluation of Forecast Skill	402
18.5 Cross-validation	405

VII Appendices	407
A Notation	409
B Elements of Linear Analysis	413
C Fourier Analysis and Fourier Transform	416
D Normal Density and Cumulative Distribution Function	419
E The χ^2 Distribution	421
F Student's t Distribution	423
G The F Distribution	424
H Table-Look-Up Test	431
I Critical Values for the Mann–Whitney Test	437
J Quantiles of the Squared-ranks Test Statistic	443
K Quantiles of the Spearman Rank Correlation Coefficient	446
L Correlations and Probability Statements	447
M Some Proofs of Theorems and Equations	451
References	455

Preface

The tools of mathematical statistics find wide application in climatological research. Indeed, climatology is, to a large degree, the study of the statistics of our climate. Mathematical statistics provides powerful tools which are invaluable for this pursuit. Applications range from simple uses of sampling distributions to provide estimates of the uncertainty of a climatological mean to sophisticated statistical methodologies that form the basis of diagnostic calculations designed to reveal the dynamics of the climate system. However, even the simplest of statistical tools has limitations and pitfalls that may cause the climatologist to draw false conclusions from valid data if the tools are used inappropriately and without a proper understanding of their conceptual foundations. The purpose of this book is to help the climatologist understand the basic precepts of the statistician's art and to provide some of the background needed to apply statistical methodology correctly and usefully.

We do not claim that this volume is in any way an exhaustive or comprehensive guide to the use of statistics in climatology, nor do we claim that the methodology described here is a current reflection of the art of applied statistics as it is conducted by statisticians. Statistics as it is applied in climatology is far removed from the cutting edge of methodological development. This is partly because statistical research has not come yet to grips with many of the problems encountered by climatologists and partly because climatologists have not yet made very deep excursions into the world of mathematical statistics. Instead, this book presents a subjectively chosen discourse on the tools we have found useful in our own research on climate diagnostics.

We will discuss a variety of statistical concepts and tools which are useful for solving problems in climatological research, including the following.

- The concept of a sample.
- The notions of exploratory and confirmatory statistics.

- The concept of the statistical model. Such a model is implicit in every statistical analysis technique and has substantial implications for the conclusions drawn from the analysis.
- The differences between parametric and non-parametric approaches to statistical analysis.
- The estimation of 'parameters' that describe the properties of the geophysical process being studied. Examples of these 'parameters' include means and variances, temporal and spatial power spectra, correlation coefficients, empirical orthogonal functions and Principal Oscillation Patterns. The concept of parameter estimation includes not only point estimation (estimation of the specific value of a parameter) but also interval estimation which account for uncertainty.
- The concepts of hypothesis testing, significance, and power.

We do *not* deal with:

- *Bayesian statistics*, which is philosophically quite different from the more common *frequentist* approach to statistics we use in this book. Bayesians, as they are known, incorporate *a priori* beliefs into a statistical analysis of a sample in a rational manner (see Epstein [114], Casella [77], or Gelman et al. [139]).
- *Geostatistics*, which is widely used in geology and related fields. This approach deals with the analysis of spatial fields sampled at a relatively small number of locations. The most prominent technique is called *kriging* (see Journel and Huijbregts [207], Journel [206], or Wackernagel [406]), which is related to the *data assimilation* techniques used in atmospheric and oceanic science (see, e.g., Daley [98] and Lorenc [258]).

A collection of applications of many statistical techniques has been compiled by von Storch and Navarra [395]; we recommend this collection as complementary reading to this book and refer to

its contributions throughout. This collection does not cover the field systematically; instead it offers examples of the exploitation of statistical methods in the analysis of climatic data and numerical experiments.

Cookbook recipes for a variety of standard statistical situations are not offered by this book because they are dangerous for anyone who does not understand the basic concepts of statistics. Therefore, we offer a course in the concepts and discuss cases we have encountered in our work. Some of these examples refer to standard situations, and others to more exotic cases. Only the understanding of the principles and concepts prevents the scientist from falling into the many pitfalls specific to our field, such as multiplicity in statistical tests, the serial dependence within samples, or the enormous size of the climate's phase space. If these dangers are not understood, then the use of simple recipes will often lead to erroneous conclusions. Literature describes many cases, both famous and infamous, in which this has occurred.

We have tried to use a consistent notation throughout the book, a summary of which is offered in Appendix A. Some elements of linear algebra are available in Appendix B, and some aspects of Fourier analysis and transform are listed in Appendix C. Proofs of statements, which we do not consider essential for the overall understanding, are in Appendix M.

Thanks

We are deeply indebted to a very large number of people for their generous assistance with this project. We have tried to acknowledge all who contributed, but we will inevitably have overlooked some. We apologize sincerely for these oversights.

- Thanks for her excellent editorial assistance: Robin Taylor.

- Thanks for discussion, review, advice and useful comments: Gerd Bürger, Bill Burrows, Ulrich Callies, Susan Chen, Christian Eckert, Claude Frankignoul, Marco Giorgetta, Silvio Gualdi, Stefan Güß, Klaus Hasselmann, Gabi Hegerl, Patrick Heimbach, Andreas Hense, Hauke Heyen, Martina Junge, Thomas Kaminski, Frank Kauker, Dennis Lettenmaier, Bob Livezey, Ute Luksch, Katrin Maak, Rol Madden, Ernst Maier-Reimer, Peter Müller, Dörthe Müller-Navarra, Matthias Münnich, Allan Murphy, Antonio Navarra, Peter Rayner, Mark Saunders, Reiner Schnur, Dennis Shea, Achim Stössel, Sylvia Venegas, Stefan Venzke, Koos Verbeek, Jin-Song von Storch, Hans Wackernagel, Xiaolan Wang, Chris Wickle, Arne Winguth, Eduardo Zorita.
- Thanks for making diagrams available to us: Howard Barker, Anthony Barnston, Grant Branstator, Gerd Bürger, Bill Burrows, Klaus Fraedrich, Claude Frankignoul, Eugenia Kalnay, Viacheslav Kharin, Kees Korevaar, Steve Lambert, Dennis Lettenmaier, Bob Livezey, Katrin Maak, Allan Murphy, Hisashi Nakamura, Reiner Schnur, Lucy Vincent, Jin-Song von Storch, Mike Wallace, Peter Wright, Eduardo Zorita.
- Thanks for preparing diagrams: Marion Grunert, Doris Lewandowski, Katrin Maak, Norbert Noreiks, and Hinrich Reichardt, who helped also to create some of the tables in the Appendices. For help with the \LaTeX -system: Jörg Wegner. For help with the Hamburg computer network: Dierk Schriever. For help with the Canadian Centre for Climate Modelling and Analysis computer network in Victoria: Mike Berkley. For scanning diagrams: Mike Berkley, Jutta Bernlöhr, and Marion Grunert.