

Springer Series in Statistics

Advisors:

P. Bickel, P. Diggle, S. Fienberg, U. Gather,
I. Olkin, S. Zeger

Springer Series in Statistics

- Alho/Spencer*: Statistical Demography and Forecasting.
Andersen/Borgan/Gill/Keiding: Statistical Models Based on Counting Processes.
Atkinson/Riani: Robust Diagnostic Regression Analysis.
Atkinson/Riani/Cerioni: Exploring Multivariate Data with the Forward Search.
Berger: Statistical Decision Theory and Bayesian Analysis, 2nd edition.
Borg/Groenen: Modern Multidimensional Scaling: Theory and Applications, 2nd edition.
Brockwell/Davis: Time Series: Theory and Methods, 2nd edition.
Bucklew: Introduction to Rare Event Simulation.
Cappé/Moulines/Rydén: Inference in Hidden Markov Models.
Chan/Tong: Chaos: A Statistical Perspective.
Chen/Shao/Ibrahim: Monte Carlo Methods in Bayesian Computation.
Coles: An Introduction to Statistical Modeling of Extreme Values.
David/Edwards: Annotated Readings in the History of Statistics.
Devroye/Lugosi: Combinatorial Methods in Density Estimation.
Efromovich: Nonparametric Curve Estimation: Methods, Theory, and Applications.
Eggermont/LaRiccia: Maximum Penalized Likelihood Estimation, Volume I: Density Estimation.
Fahrmeir/Tutz: Multivariate Statistical Modelling Based on Generalized Linear Models, 2nd edition.
Fan/Yao: Nonlinear Time Series: Nonparametric and Parametric Methods.
Farebrother: Fitting Linear Relationships: A History of the Calculus of Observations 1750-1900.
Federer: Statistical Design and Analysis for Intercropping Experiments, Volume I: Two Crops.
Federer: Statistical Design and Analysis for Intercropping Experiments, Volume II: Three or More Crops.
Ferraty/View: Nonparametric Functional Data Analysis: Models, Theory, Applications, and Implementation
Ghosh/Ramamoorthi: Bayesian Nonparametrics.
Glaz/Naus/Wallenstein: Scan Statistics.
Good: Permutation Tests: Parametric and Bootstrap Tests of Hypotheses, 3rd edition.
Gouriéroux: ARCH Models and Financial Applications.
Gu: Smoothing Spline ANOVA Models.
Györfil/Kohler/Krzyżak/Walk: A Distribution-Free Theory of Nonparametric Regression.
Haberman: Advanced Statistics, Volume I: Description of Populations.
Hall: The Bootstrap and Edgeworth Expansion.
Härdle: Smoothing Techniques: With Implementation in S.
Harrell: Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.
Hart: Nonparametric Smoothing and Lack-of-Fit Tests.
Hastie/Tibshirani/Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction.
Hedayat/Sloane/Stufken: Orthogonal Arrays: Theory and Applications.
Heyde: Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation.

(continued after index)

Nhu D. Le James V. Zidek

Statistical Analysis of Environmental Space-Time Processes

 Springer

Nhu D. Le
British Columbia Cancer
Research Center
675 West 10th Avenue
Vancouver V5Z 1L3
Canada
nle@bccrc.ca

James V. Zidek
Department of Statistics
University of British Columbia
333-6356 Agricultural Road
Vancouver V6T 1Z2
Canada
jim@stat.ubc.ca

Library of Congress Control Number: 2005939015

ISBN-10: 0-387-26209-1

ISBN-13: 978-0387-26209-3

© 2006 Springer Science+Business Media, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Springer Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (MVY)

9 8 7 6 5 4 3 2 1

springer.com

To Lynne, Hilda, Adrian, and Megan

Preface

This book presents knowledge gained by the authors along with methods they developed, over more than 30 years of experience measuring, modeling, and mapping environmental space–time fields. That experience embraces both large (continentwide) spatial domains and small. In part it comes from their research, working with students as well as coinvestigators. But much was gained from all sorts of interactions with many individuals who have had to contend with the challenges these fields present. They include statistical as well as subject area scientists, in areas as diverse as analytical chemistry, air sampling, atmospheric science, environmental epidemiology, environmental risk management, and occupational health among others. We have collaborated and consulted with government scientists as well as policy-makers, in all, a large group of individuals from whom we have learned a lot and to whom we are indebted. We hope all in these diverse groups will find something of value in this book. We believe it will also benefit graduate students, both in statistics and subject areas who must deal with the analysis of environmental fields.

In fact we have given a successful statistics graduate course based on it. The book (and course) reflect our conviction about the need for statistical scientists to learn about the phenomena they purport to explain. To the extent feasible, we have covered important nonstatistical issues involved in dealing with environmental processes. Thus in writing the book we have tried to strike a balance between important qualitative and quantitative aspects of the subject. Much of the most technical statistical-mathematical material has been placed in the starred sections, chapters, and appendices. These could well be skipped, at least on first reading. In fact the simplest path to that technical material would be through Chapter 14; it contains a more-or-less self-contained tutorial on methods developed by the authors. That tutorial relies on R software that can be downloaded by the interested reader.

When we started analyzing environmental processes, we soon came to know some of the inadequacies of geostatistical methods. These purely spatial methods had been around for a long time and proven very successful in

geostatistical application. Thanks to the SIMS group at Stanford they had even been appropriated in the 1970s for use in analyzing ozone space–time fields. However, the acid rain fields that were the initial focus of our study involved multivariate responses with up to a dozen chemical species measured at a large number of sites over a broad spatial domain. Moreover, it became clear that while these responses could be transformed to have an approximately normal distribution, their spatial covariances were far from stationary, a condition of fundamental importance in classical geostatistics. The failure of that assumption led Paul Sampson and Peter Guttorp to their discovery of an elegant route around that assumption (Chapter 6). The need to handle multivariate responses and reflect our considerable uncertainty about the spatial covariance matrix led us to our hierarchical Bayes theory, the subject of Chapters 9 and 10. Chapter 9, the simplified version, conveys the basic elements of our theory.

Chapter 10 presents the fully general (multivariate) theory. It incorporates enhancements made over time to contend with difficult situations encountered in applications. The last published extension appeared in 2002. Additional theory was developed for the book. To avoid excessive technicality, we have given much of the detail in the Appendices.

The theory in that chapter really provides the “engine” that drives our model and applications in Chapters 11–13. Chapter 11 uses that engine to drive a theory for designing networks for monitoring environmental processes, one of the most difficult challenges facing environmental scientists. Other challenges are seen in Chapter 12 where the important topic of environmental process extremes is visited. In spite of their immense importance in environmental risk analysis, this topic has received relatively little emphasis in environmental statistics. In contrast, the topic of Chapter 13, environmental risk, has been heavily studied. Our contributions to it, in particular, to environmental health risk analysis appear there.

The novelty of the methods emphasized in this book has necessitated the development of software for implementation. Sampson and Guttorp developed theirs for covariance modeling and we have incorporated a version of it in ours. Although our research group developed the code needed to implement our multivariate theory, that code has been greatly refined thanks to the substantial contributions of our colleague and sometime research partner, Rick White.

Although the book features a lot of our own methods and approaches, we try to give a reasonably comprehensive review of the many other, often ingenious approaches that have been developed by others over the years. In all cases we try to indicate strengths and limitations. An extensive bibliography should enable interested readers to find out more about the alternatives.

To conclude, we would like to express our deepest appreciation to all who have helped us gain the knowledge reflected in this book. Our gratitude also goes to those who helped implement that knowledge and develop the tools we needed to handle space–time fields. That includes our many co-authors,

including former students. A special thanks goes to Bill Caselton who first stimulated the second author's interest in environmental processes, and to our long time research compatriots, Peter Guttorp and Paul Sampson for a long and fruitful collaboration as well as for generously allowing us to use their software. John Kimmel, Springer's Executive Editor–Statistics, and several anonymous reviewers have provided numerous thoughtful comments and suggestions that have undoubtedly improved the book's presentation. The Copy-Editors, Valerie Greco and Natacha Menar were superb. Part of the book is based on work done while the second author was on leave at the University of Bath and later at the Statistical and Applied Mathematical Science Institute; both generously provided facilities and support. The Natural Sciences and Engineering Research Council of Canada (NSERC) has been a constant source of funding, partially supporting our research developments described in this book. Finally, we thank our wives, Hilda and Lynne for their support and patience throughout this book's long gestation period. Without that this book would certainly not have been written!

Vancouver, British Columbia
March 2006

Nhu D Le
James V Zidek

Contents

Preface

Part I: Environmental Processes

1	First Encounters	3
1.1	Environmental Fields	3
1.1.1	Examples	8
1.2	Modeling Foundations	10
1.2.1	Space–Time Domains	11
1.2.2	Procedure Performance Paradigms	11
1.2.3	Bayesian Paradigm	12
1.2.4	Space–time Fields	13
1.3	Wrapup	13
2	Case Study	15
2.1	The Data	15
2.2	Preliminaries	16
2.3	Space–time Process Modeling	19
2.4	Results!	19
2.5	Wrapup	24
3	Uncertainty	27
3.1	Probability: “The Language of Uncertainty”	27
3.2	Probability and Uncertainty	28
3.3	Uncertainty Versus Information	30
3.3.1	Variance	31
3.3.2	Entropy	32
3.4	Wrapup	33

4	Measurement	35
4.1	Spatial Sampling	36
4.1.1	Acid Precipitation	36
4.1.2	The Problem of Design Objectives	39
4.1.3	A Probability-Based Design Solution	40
4.1.4	Pervasive Principles	41
4.2	Sampling Techniques	42
4.2.1	Measurement: The Illusion!	42
4.2.2	Air Pollution	42
4.2.3	Acid Precipitation Again	43
4.2.4	Toxicology and Biomarkers	44
4.3	Data Quality	45
4.3.1	Cost Versus Precision	45
4.3.2	Interlaboratory and Measurement Issues	45
4.4	Measurement Error	46
4.4.1	A Taxonomy of Types	47
4.5	Effects	49
4.5.1	Subtleties	50
4.6	Wrapup	51
5	Modeling	53
5.1	Why Model?	53
5.2	What Makes a Model Good?	56
5.3	Approaches to Modeling***	57
5.3.1	Modeling with Marginals	59
5.3.2	Modeling by Conditioning	59
5.3.3	Single Timepoints	60
5.3.4	Hierarchical Bayesian Modeling	61
5.3.5	Dynamic state-space Models	62
5.3.6	Orthogonal Series	63
5.3.7	Computer Graphical Models	66
5.3.8	Markov Random Fields	68
5.3.9	Latent Variable Methods	70
5.3.10	Physical–Statistical Models	71
5.4	Gaussian Fields	74
5.5	Log Gaussian Processes	77
5.6	Wrapup	78

Part II: Space–Time Modeling

6	Covariances	83
6.1	Moments and Variograms	84
6.1.1	Finite-Dimensional Distributions	84
6.2	Stationarity	86

6.3	Variogram Models for Stationary Processes	88
6.3.1	Characteristics of Covariance Functions	88
6.4	Isotropic Semi-Variogram Models	89
6.5	Correlation Models for Nonstationary Processes	93
6.5.1	The Sampson–Guttorp Method	93
6.5.2	The Higdon, Swall, and Kern Method	97
6.5.3	The Fuentes Method	98
6.6	Wrapup	99
7	Spatial Prediction: Classical Approaches	101
7.1	Ordinary Kriging	104
7.2	Universal Kriging	107
7.3	Cokriging	111
7.4	Disjunctive Kriging	113
7.5	Wrapup	116
8	Bayesian Kriging	119
8.1	The Kitanidis Framework***	121
8.1.1	Model Specification	121
8.1.2	Prior Distribution	122
8.1.3	Predictive Distribution	123
8.1.4	Remarks	123
8.2	The Handcock and Stein Method***	124
8.3	The Bayesian Transformed Gaussian Approach	126
8.3.1	The BTG Model	127
8.3.2	Prior Distribution	128
8.3.3	Predictive Distribution	128
8.3.4	Numerical Integration Algorithm	129
8.4	Remarks	130
9	Hierarchical Bayesian Kriging	131
9.1	Univariate Setting	134
9.1.1	Model Specification	135
9.1.2	Predictive Distribution	136
9.2	Missing Data	141
9.3	Staircase Pattern of Missing Data	142
9.3.1	Notation	143
9.3.2	Staircase Model Specification	145
9.3.3	The GIW Distribution	146
9.3.4	Predictive Distributions	146
9.4	Wrapup	148

Part III: Design and Risk Assessment

10	Multivariate Modeling***	153
10.1	General Staircase	155
10.1.1	Notation	155
10.2	Model Specification	158
10.3	Predictive Distributions	159
10.4	Posterior Distributions	162
10.5	Posterior Expectations	165
10.6	Hyperparameter Estimation	167
10.6.1	Two-Step Estimation Procedure	167
10.6.2	Spatial Covariance Separability	168
10.6.3	Estimating Gauged Site Hyperparameters	171
10.6.4	Estimating Ungauged Site Hyperparameters	177
10.7	Systematically Missing Data	178
10.8	Credibility Ellipsoids	181
10.9	Wrapup	183
11	Environmental Network Design	185
11.1	Design Strategies	187
11.2	Entropy-Based Designs	191
11.3	Entropy	191
11.4	Entropy in Environmental Network Design	194
11.5	Entropy Criteria	196
11.6	Predictive Distribution	196
11.7	Criteria	198
11.8	Incorporating Cost	199
11.9	Computation***	200
11.10	Case Study	202
11.11	Pervasive Issues***	206
11.12	Wrapup	213
12	Extremes	215
12.1	Fields of Extremes	216
12.1.1	Theory of Extremes	216
12.2	Hierarchical Bayesian Model	220
12.2.1	Empirical Assessment	221
12.3	Designer Challenges	222
12.3.1	Loss of Spatial Dependence	222
12.3.2	Uncertain Design Objectives	227
12.4	Entropy Designs for Monitoring Extremes	239
12.5	Wrapup	241

Part IV: Implementation

13 Risk Assessment 245

13.1 Environmental Risk Model 245

13.2 Environmental Risk 246

13.3 Risk in Postnormal Science 249

13.4 Environmental Epidemiology*** 252

13.4.1 Impact Assessment*** 253

13.5 Case Study 263

13.6 Wrapup 268

14 R Tutorial 271

14.1 Exploratory Analysis of the Data 272

14.2 Spatial Predictive Distribution and Parameter Estimation 278

14.2.1 Parameter Estimation: Gauged Sites Through the
EM-algorithm 279

14.2.2 Parameter Estimation: The Sampson–Guttorp Method 282

14.2.3 Parameter Estimation: Ungauged Sites 290

14.3 Spatial Interpolation 290

14.4 Monitoring Network Extension 291

Appendices 297

15.1 Probabilistic Distributions 297

15.1.1 Multivariate and Matrix Normal Distribution 297

15.1.2 Multivariate and Matric-*t* Distribution 298

15.1.3 Wishart and Inverted Wishart Distribution 299

15.1.4 Generalized Inverted Wishart Distribution 300

15.2 Bartlett Decomposition 302

15.2.1 Two-Block Decomposition 302

15.2.2 Recursive Bartlett Decomposition for Multiple Blocks 302

15.3 Useful Matrix Properties 303

15.4 Proofs for Chapter 10 307

References 313

Author Index 327

Subject Index 331