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/Cerioli: 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/Vieu: 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.
Györfi/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

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 environmentricians. 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 environmetrics. 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 — *Nhu D Le*
March 2006 — *James V Zidek*

Contents