

Statistics for Spatial Data

Statistics for Spatial Data

Revised Edition

NOEL A. C. CRESSIE

Iowa State University



A Wiley-Interscience Publication

JOHN WILEY & SONS, INC.

New York • Chichester • Toronto • Brisbane • Singapore

This text is printed on acid-free paper.

Copyright © 1993 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008.

Library of Congress Cataloging in Publication Data:

Cressie, Noel A. C.

Statistics for spatial data revised edition / Noel A. C. Cressie.

p. cm.—(Wiley series in probability and mathematical statistics. Applied probability and statistics section)

“A Wiley-Interscience publication.”

Includes bibliographical references and index.

1. Spatial analysis (Statistics) I. Title. II. Series.

QA278.2.C75 1993

519.5—dc20

93-775

CIP

ISBN 0-471-00255-0

Printed in the United States of America

30 29 28 27 26 25 24 23 22 21

To Yoko, Amie, and Sean

Contents

Preface	xv
Acknowledgments	xix
1. Statistics for Spatial Data	1
1.1 Spatial Data and Spatial Models, 7	
1.2 Introductory Examples, 10	
1.2.1 Geostatistical Data, 10	
1.2.2 Lattice Data, 11	
1.2.3 Point Patterns, 12	
1.3 Statistics for Spatial Data: Why?, 13	
 PART I GEOSTATISTICAL DATA	
2. Geostatistics	29
2.1 Continuous Spatial Index, 29	
2.2 [†] Spatial Data Analysis of Coal Ash in Pennsylvania, 30	
2.2.1 [†] Intrinsic Stationarity, 40	
2.2.2 [†] Square-Root-Differences Cloud, 41	
2.2.3 [†] The Pocket Plot, 42	
2.2.4 [†] Decomposing the Data into Large- and Small-Scale Variation, 46	
2.2.5 [†] Analysis of Residuals, 48	
2.2.6 [†] Variogram of Residuals from Median Polish, 50	
2.3* Stationary Processes, 52	
2.3.1 Variogram, 58	
2.3.2 Covariogram and Correlogram, 67	

* Theory

[†] Applications

- 2.4 Estimation of the Variogram, 69
 - 2.4.1 Comparison of Variogram and Covariogram Estimation, 70
 - 2.4.2 Exact Distribution Theory for the Variogram Estimator, 73
 - 2.4.3 Robust Estimation of the Variogram, 74
- 2.5* Spectral Representations, 83
 - 2.5.1* Valid Covariograms, 84
 - 2.5.2* Valid Variograms, 86
- 2.6 Variogram Model Fitting, 90
 - 2.6.1 Criteria for Fitting a Variogram Model, 91
 - 2.6.2 Least Squares, 94
 - 2.6.3 Properties of Variogram-Parameter Estimators, 99
 - 2.6.4 Cross-Validating the Fitted Variogram, 101

3. Spatial Prediction and Kriging

105

- 3.1 Scale of Variation, 112
- 3.2 Ordinary Kriging, 119
 - 3.2.1 Effect of Variogram Parameters on Kriging, 127
 - 3.2.2 Lognormal and Trans-Gaussian Kriging, 135
 - 3.2.3 Cokriging, 138
 - 3.2.4 Some Final Remarks, 142
- 3.3 Robust Kriging, 144
- 3.4 Universal Kriging, 151
 - 3.4.1 Universal Kriging of Coal-Ash Data, 157
 - 3.4.2 Trend-Surface Prediction, 162
 - 3.4.3 Estimating the Variogram for Universal Kriging, 165
 - 3.4.4 Bayesian Kriging, 170
 - 3.4.5 Kriging Revisited, 172
- 3.5 Median-Polish Kriging, 183
 - 3.5.1 Gridded Data, 184
 - 3.5.2 Nongridded Data, 193
 - 3.5.3 Median Polishing Spatial Data: Inference Results, 194
 - 3.5.4 Median-Based Covariogram Estimators are Less Biased, 196
- 3.6 Geostatistical Data, Simulated and Real, 200
 - 3.6.1 Simulation of Spatial Processes, 201
 - 3.6.2 Conditional Simulation, 207
 - 3.6.3 Geostatistical Data, 209

4. Applications of Geostatistics	211
4.1 [†] Wolfcamp-Aquifer Data, 212	
4.1.1 [†] Intrinsic-Stationarity Assumption, 213	
4.1.2 [†] Nonconstant-Mean Assumption, 217	
4.2 [†] Soil–Water Tension Data, 224	
4.3 [†] Soil–Water-Infiltration Data, 230	
4.3.1 [†] Estimating and Modeling the Spatial Dependence, 232	
4.3.2 [†] Inference on Mean Effects (Spatial Analysis of Variance), 238	
4.4 [†] Sudden-Infant-Death-Syndrome Data, 244	
4.5 [†] Wheat-Yield Data, 248	
4.5.1 [†] Presence of Trend in the Data, 250	
4.5.2 [†] Intrinsic Stationarity, 251	
4.5.3 [†] Median-Polish (Robust) Kriging, 255	
4.6 [†] Acid-Deposition Data, 259	
4.6.1 [†] Spatial Modeling and Prediction, 260	
4.6.2 [†] Sampling Design, 268	
4.7 Space–Time Geostatistical Data, 273	
5. Special Topics in Statistics for Spatial Data	277
5.1* Nonlinear Geostatistics, 278	
5.2 Change of Support, 284	
5.3 Stability of the Geostatistical Method, 289	
5.3.1 Estimation of Spatial-Dependence Parameters, 291	
5.3.2 Stability of the Kriging Predictor, 292	
5.3.3 Stability of the Kriging Variance, 296	
5.4 Intrinsic Random Functions of Order k , 299	
5.5* Applications of the Theory of Random Processes, 309	
5.6 Spatial Design, 313	
5.6.1 Spatial Sampling Design, 314	
5.6.2* Spatial Experimental Design, 324	
5.7 Field Trials, 338	
5.7.1 Nearest-Neighbor Analyses, 338	
5.7.2 Analyses Based on Spatial Modeling, 344	
5.8 Infill Asymptotics, 350	
5.9 The Many Faces of Spatial Prediction, 356	
5.9.1 Stochastic Methods of Spatial Prediction, 357	
5.9.2 Nonstochastic Methods of Spatial Prediction, 370	
5.9.3 Comparisons and Some Final Remarks, 378	

PART II LATTICE DATA

6. Spatial Models on Lattices	383
6.1 Lattices, 383	
6.2 [†] Spatial Data Analysis of Sudden Infant Deaths in North Carolina, 385	
6.2.1 [†] Nonspatial Data Analysis, 391	
6.2.2 [†] Spatial Data Analysis, 393	
6.2.3 [†] Trend Removal, 396	
6.2.4 [†] Some Final Remarks, 401	
6.3 Conditionally and Simultaneously Specified Spatial Gaussian Models, 402	
6.3.1 Simultaneously Specified Spatial Gaussian Models, 405	
6.3.2 Conditionally Specified Spatial Gaussian Models, 407	
6.3.3 Comparison, 408	
6.4* Markov Random Fields, 410	
6.4.1* Neighbors, Cliques, and the Negpotential Function Q , 414	
6.4.2* Pairwise-Only Dependence and Conditional Exponential Distributions, 419	
6.4.3* Some Final Remarks, 422	
6.5 Conditionally Specified Spatial Models for Discrete Data, 423	
6.5.1 Binary Data, 423	
6.5.2 Counts Data, 427	
6.6 Conditionally Specified Spatial Models for Continuous Data, 433	
6.7 Simultaneously Specified and Other Spatial Models, 440	
6.7.1 Simultaneously Specified Spatial Models, 440	
6.7.2 Other Spatial Models, 447	
6.8 Space–Time Models, 449	
7. Inference for Lattice Models	453
7.1 [†] Inference for the Mercer and Hall Wheat-Yield Data, 453	
7.1.1 [†] Data Description, 454	
7.1.2 [†] Spatial Lattice Models, 456	
7.2 Parameter Estimation for Lattice Models, 458	
7.2.1 Estimation Criteria, 458	
7.2.2 Gaussian Maximum Likelihood Estimation, 465	
7.2.3 Some Computational Details, 472	

- 7.3 Properties of Estimators, 477
 - 7.3.1* Increasing-Domain Asymptotics, 480
 - 7.3.2 The Jackknife and Bootstrap for Spatial Lattice Data, 489
 - 7.3.3 Cross-Validation and Model Selection, 497
- 7.4 Statistical Image Analysis and Remote Sensing, 499
 - 7.4.1 Remote Sensing, 501
 - 7.4.2 Ordinary Discriminant Analysis, 502
 - 7.4.3* Markov-Random-Field Models, 509
 - 7.4.4* Edge Processes, 521
 - 7.4.5* Textured Images, 525
 - 7.4.6* Single Photon Emission Tomography, 525
 - 7.4.7* Least Squares and Image Regularization, 528
 - 7.4.8* Method of Sieves, 532
 - 7.4.9* Mathematical Morphology, 534
- 7.5 Regional Mapping, Scotland Lip-Cancer Data, 535
 - 7.5.1[†] Exploratory Regional Mapping, 537
 - 7.5.2 Parametric Empirical Bayes Mapping, 544
- 7.6 Sudden-Infant-Death-Syndrome Data, 548
 - 7.6.1[†] Exploratory Spatial Data Analysis, 549
 - 7.6.2 Auto-Poisson Model, 553
 - 7.6.3 Auto-Gaussian Model, 555
- 7.7 Lattice Data, Simulated and Real, 568
 - 7.7.1 Simulation of Lattice Processes, 569
 - 7.7.2 Lattice Data, 572

PART III SPATIAL PATTERNS

8. Spatial Point Patterns

577

- 8.1 Random Spatial Index, 578
- 8.2 Spatial Data Analysis of Longleaf Pines (*Pinus palustris*), 579
 - 8.2.1[†] Data Description, 579
 - 8.2.2 Complete Spatial Randomness, Regularity, and Clustering, 580
 - 8.2.3[†] Quadrat Methods, 588
 - 8.2.4[†] Kernel Estimators of the Intensity Function, 597
 - 8.2.5[†] Distance Methods, 602
 - 8.2.6[†] Nearest-Neighbor Distribution Functions and the K Function, 613
 - 8.2.7[†] Some Final Remarks, 618

8.3*	Point Process Theory, 619	
8.3.1*	Moment Measures, 622	
8.3.2*	Generating Functionals, 624	
8.3.3*	Stationary and Isotropic Point Processes, 628	
8.3.4*	Palm Distributions, 630	
8.3.5*	Reduced Second Moment Measure, 631	
8.4	Complete Spatial Randomness, Distance Functions, and Second Moment Measures, 633	
8.4.1	Complete Spatial Randomness, 633	
8.4.2	Distance Functions, 636	
8.4.3	K Functions, 639	
8.4.4 [†]	Animal-Behavior Data, 644	
8.4.5	Some Final Remarks, 649	
8.5	Models and Model Fitting, 650	
8.5.1*	Inhomogeneous Poisson Process, 650	
8.5.2*	Cox Process, 657	
8.5.3*	Poisson Cluster Process, 661	
8.5.4*	Simple Inhibition Point Processes, 669	
8.5.5*	Markov Point Process, 673	
8.5.6*	Thinned and Related Point Processes, 689	
8.5.7*	Other Models, 693	
8.5.8*	Some Final Remarks, 694	
8.6*	Multivariate Spatial Point Processes, 696	
8.6.1*	Theoretical Considerations, 696	
8.6.2*	Estimation of the Cross K Function, 698	
8.6.3*	Bivariate Spatial–Point-Process Models, 699	
8.7*	Marked Spatial Point Processes, 707	
8.7.1*	Theoretical Considerations, 707	
8.7.2*	Estimation of Moment Measures, 714	
8.7.3*	Marked Spatial–Point-Process Models, 716	
8.8	Space–Time Point Patterns, 719	
8.9	Spatial Point Patterns, Simulated and Real, 722	
8.9.1	Simulation of Spatial Point Patterns, 722	
8.9.2	Spatial Point Patterns, 723	
9.	Modeling Objects	725
9.1	Set Models, 727	
9.1.1	Fractal Sets, 727	
9.1.2	Fuzzy Sets, 731	
9.1.3	Random Closed Sets: An Example, 736	

9.2 [†]	Random Parallelograms in \mathbb{R}^2 , 739	
9.3*	Random Closed Sets and Mathematical Morphology, 742	
9.3.1*	Theory and Methods, 745	
9.3.2*	Inference on Random Closed Sets, 750	
9.4	The Boolean Model, 753	
9.4.1*	Main Properties, 755	
9.4.2*	Generalizations of the Boolean Model, 756	
9.5	Methods of Boolean-Model Parameter Estimation, 759	
9.5.1	Analysis of Random-Parallelograms Data, 761	
9.5.2	Analysis of Heather-Incidence Data, 763	
9.5.3*	Intensity Estimation in the Boolean Model, 765	
9.6	Inference for the Boolean Model, 770	
9.7	Modeling Growth with Random Sets, 776	
9.7.1	Random-Set Growth Models, 777	
9.7.2	Tumor-Growth Data, 783	
9.7.3	Fitting the Tumor-Growth Parameters, 794	
References		803
Author Index		873
Subject Index		887

Preface

The purpose of this book is to present Statistics for spatial data to scientists and engineers. (Notice that Statistics is capitalized to distinguish it from its other meaning: a collection of numbers that summarize a complex phenomenon—such as baseball or cricket.) In the last 10 years, much interest has been generated in the area, but its exposure in the literature has been uneven. This book attempts to take that literature and extend it, correct it, and unify it. What appears to be a gathering of unconnected subject areas can be annealed into a cohesive approach to the analysis of spatial data. Chapter 1 provides an overview of the approach and of the enormous diversity of problems involving spatial data, from the microscopic to the astronomic.

The book attempts to give a somewhat complete coverage of each of three parts, dealing with *geostatistical data*, *lattice data*, and *point patterns*. Thus, the subject areas are classified according to the type of observations encountered, reflecting my belief that the roots of statistical science are in data. Statistical models, then, try to make sense out of the data, albeit imperfectly. Design, inference, and diagnostics are natural consequences of the data-model symbiosis, and all play an important role in Statistics for spatial data.

This book grew from lecture notes for a one-semester, 3-credit Statistics graduate course that I conduct at Iowa State University. In 45 lectures, each of 50-minutes duration, I cover the following topics:

Part I (Geostatistical Data): All of Chapter 2 except Section 2.5. All of Chapter 3 except Sections 3.3 and 3.6.

Part II (Lattice Data): Chapter 6. Chapter 7, Sections 7.2, 7.3, and 7.6.

Part III (Point Patterns): Chapter 8, Sections 8.1, 8.2, 8.4, 8.5.1, 8.5.2, 8.5.3. Chapter 9, Section 9.1.

Prerequisites for the Statistics graduate course are one semester of Masters-level statistical inference and one semester of Masters-level linear models. While giving the course and preparing the book, I have benefitted from useful reference books in the area. Reading lists for the course have included Matern (1960), Bartlett (1975), Journel and Huijbregts (1978), Cliff and Ord (1981), Ripley (1981), and Upton and Fingleton (1985).

It is my hope that this book can be read in varying depths by people with varying mathematical and statistical backgrounds. However, there are sections in the book, on random processes, point processes, and random sets, that are beyond a Masters-level student and closer to the frontiers of theoretical research (these sections are denoted by an asterisk). Equally, there are sections that concentrate purely on an application and do not add to the theoretical development of the subject (these sections are denoted by a dagger). These applications-oriented sections of the book should appeal to a large number of scientists and engineers, with at least the background of a service course in Statistics (or equivalent). For this reason, most chapters begin with an application, which is meant to be an invitation to read on. The emphasis on applications has led to considerable use of graphs and illustrations.

There are some features of the book that I believe will enhance its value as a textbook and as a reference book. An attempt has been made at completeness, in terms of the topics covered, and uniformity, in terms of the depth to which they are covered, except for Chapters 5 and 9. These two chapters contain material that is either of personal interest or is speculative in nature. The reader will notice frequent referencing to a diverse literature; one of the interesting features of Statistics for spatial data is that a large proportion of it is appearing *outside* Statistics journals. The referencing allows me to pick up apparently different streams of thought and tie them together. Equally importantly, I have tried to give credit where it is due. The linear reader will also notice a certain amount of repetition between chapters (and, to a lesser extent, between sections). This is deliberate and is meant to help the sporadic reader who wants to understand the essence of a topic but who has not read all the previous pages.

We should not forget our roots. All data sets for the various spatial statistical analyses are given in the book, as well as some background to the problems being studied. Further, each of the three parts has a section devoted to sources of spatial data, both real and simulated.

No exercises are given at the end of sections; the depth of coverage within a section should allow practice exercises to suggest themselves to an instructor teaching from the book. Software is not given. (A geostatistics package, *Toolkit*, by Geostokos, London, was used for the kriging presented in Chapter 3.) Currently, those of us who work in the area tend to custom-build our own software, which is usually not very portable. Statistics for spatial data will truly realize its enormous potential when a comprehensive software package is developed.

This is a big book. I had thought of splitting it into smaller volumes; however, the present format emphasizes the subject's unity. This may be the last time spatial Statistics will be squeezed between two covers. A healthy exponential growth of the literature is apparent from the bibliography.

The future of the subject is in solving problems for spatiotemporal data; some sections are devoted to it, but a proper treatment needs another book. Statistics for spatial and temporal data would provide dynamic models for phenomena distributed through space and evolving in time. Onward into the next decade!

NOEL CRESSIE

Ames, Iowa
December 1, 1990

Acknowledgments

As a boy, knowing “why,” “how,” and “when” was not enough for me. With encouragement from my parents, a keen interest in the “where” question eventually led me to summer jobs with mining companies in Western Australia (while an undergraduate) and three years of doctoral work at Princeton University. There, I had the good fortune to be taught by Geof Watson, John Tukey, and, in my last year, Julian Besag. Geof gave me an appreciation for all things spatial and geometric, John showed me how to make my data speak (sometimes, even sing) to me, and Julian introduced me to the mysteries of Markov random fields (and English hockey). More recently, in his role as a series editor for Wiley and as an aficionado of Statistics for spatial data, Geof Watson has been of immense help with his comments and his support.

The influence of Georges Matheron of the École Nationale Supérieure des Mines de Paris in Part I and Chapter 9 is obvious; his work has been truly pioneering. I was fortunate to spend a post-doctoral period of five months at his center in Fontainebleau, France, in 1975.

My coauthors on spatial articles provided valuable impetus to my research in the subject and my colleagues throughout the world have (through their letters, their telephone calls, their questions at seminars, their anonymous referees’ reports, and their comments in hallways) helped shape my current opinions on Statistics in general and on Statistics for spatial data in particular. Along with those mentioned in the paragraphs above, I would like to thank Peter Diggle for early suggestions on topics that a course on spatial statistics might cover, Dale Zimmerman for comments on Chapter 5, Subhash Lele for comments on parts of Chapters 6 and 7, and Daryl Daley for comments on Chapter 8.

The task of taking a diverse and uneven literature on spatial Statistics and extending it, correcting it, and unifying it has not been an easy one. The editors at Wiley have been very understanding of my desire for a complete and uniformly comprehensive product. Bea Shube provided ideas and encouragement in the early stages of this project. In the last three years, helpful advice has come from Kate Roach, and, for a brief period before her, from Maggie Irwin. My department head at Iowa State University, Dean Isaacson,

has been equally understanding and in various ways helped to make an impossible task possible.

The writing of this book started in the second half of 1985 while I was an ASA/NSF/Census Fellow at the U.S. Bureau of the Census. Based on an experimental course called Spatial Statistics that I conducted at Iowa State University in 1984, I gave a series of seven seminars at the Census Bureau and distributed material that later became Chapter 6. The rest of the book (over 90%) was written at Iowa State, partially supported by the Department of Statistics and by the National Science Foundation. A special mention should be made of the superb journal and book collection of the Parks Library at Iowa State; it is one of the great resources of the University, and has helped me achieve a coverage that would otherwise have been impossible.

We are very fortunate, in our Department of Statistics, to be surrounded by intelligent and motivated graduate students. I have now offered my course on Spatial Statistics four times; it has helped me attract good students with a keen interest in the subject. As research assistants and as candidates for graduate degrees, they have been involved in various aspects of this book. The contributions of three of them deserve special mention and thanks. Stephen Rathbun wrote a preliminary draft of Sections 8.3, 8.5 through 8.9, and 7.4, and commented on subsequent drafts of Chapter 8. Carol Gotway wrote a preliminary draft of Sections 3.6 and 5.1, Martín Grondona wrote a preliminary draft of Sections 5.6 and 5.7, and Gotway and Grondona gave comments on Part I. All three were involved in the production of the figures. My thanks also go to Renkuan Guo, Jeff Helterbrand, Fred Hulting, and Jay Ver Hoef, who contributed in various ways to the preliminary through penultimate and final drafts.

This manuscript was turned into immaculate type by Sharon Shepard; by comparing what I gave her and what she gave me, it is clear that Sharon was able to work near miracles. Jeanette LaGrange, Rose Ann Anderson, and Jan Franklin also provided valuable secretarial assistance.

A trace of my space-time line would show a trip to Tokyo, Japan, in 1987. There I met Yoko. She could rightly claim that Western marriage vows say nothing about writing a book. Yet somehow she understood. We are fortunate to have been able to create two new space-time lines in our household, Amie and Sean. It was difficult for them to understand, and even harder for me to persevere. My deepest gratitude goes to all of them for their love and patience.

I am glad I started exploring Statistics for spatial data, because I have learned a great deal. Some of the territory is now well charted. Other parts are only passable by those who are sure-footed, and by not looking down. Much more of it I can only glimpse and tell you what I see. It is an exciting area that deserves a place in every statistical scientist's repertoire; I hope you will agree.

N. A. C. C.

Statistics for Spatial Data