

---

# Contents

---

<b>Preface</b>	<b>xv</b>
<b>1 Overview of spatial data problems</b>	<b>1</b>
1.1 Introduction to spatial data and models	1
1.1.1 Point-level models	6
1.1.2 Areal models	7
1.1.3 Point process models	8
1.2 Fundamentals of cartography	10
1.2.1 Map projections	10
1.2.2 Calculating distance on the earth's surface	17
1.3 Exercises	18
<b>2 Basics of point-referenced data models</b>	<b>21</b>
2.1 Elements of point-referenced modeling	22
2.1.1 Stationarity	22
2.1.2 Variograms	22
2.1.3 Isotropy	24
2.1.4 Variogram model fitting	29
2.2 Spatial process models $\star$	30
2.2.1 Formal modeling theory for spatial processes	30
2.2.2 Covariance functions and spectra	32
2.2.3 Smoothness of process realizations	36
2.2.4 Directional derivative processes	37
2.2.5 Anisotropy	37
2.3 Exploratory approaches for point-referenced data	39
2.3.1 Basic techniques	39
2.3.2 Assessing anisotropy	44
2.4 Classical spatial prediction	48
2.5 Computer tutorials	52
2.5.1 EDA and variogram fitting in <code>S+SpatialStats</code>	52
2.5.2 Kriging in <code>S+SpatialStats</code>	56
2.5.3 EDA, variograms, and kriging in <code>geOR</code>	61

2.6	Exercises	65
<b>3</b>	<b>Basics of areal data models</b>	<b>69</b>
3.1	Exploratory approaches for areal data	70
3.1.1	Measures of spatial association	71
3.1.2	Spatial smoothers	75
3.2	Brook's Lemma and Markov random fields	76
3.3	Conditionally autoregressive (CAR) models	79
3.3.1	The Gaussian case	79
3.3.2	The non-Gaussian case	83
3.4	Simultaneous autoregressive (SAR) models	84
3.5	Computer tutorials	88
3.5.1	Adjacency matrix construction in <code>S+SpatialStats</code>	88
3.5.2	SAR and CAR model fitting in <code>S+SpatialStats</code>	90
3.5.3	Choropleth mapping using the <code>maps</code> library in <code>S-plus</code>	92
3.6	Exercises	95
<b>4</b>	<b>Basics of Bayesian inference</b>	<b>99</b>
4.1	Introduction to hierarchical modeling and Bayes' Theorem	99
4.2	Bayesian inference	103
4.2.1	Point estimation	103
4.2.2	Interval estimation	104
4.2.3	Hypothesis testing and model choice	105
4.3	Bayesian computation	110
4.3.1	The Gibbs sampler	111
4.3.2	The Metropolis-Hastings algorithm	113
4.3.3	Slice sampling	116
4.3.4	Convergence diagnosis	116
4.3.5	Variance estimation	119
4.4	Computer tutorials	120
4.4.1	Basic Bayesian modeling in <code>R</code> or <code>S-plus</code>	120
4.4.2	Advanced Bayesian modeling in <code>WinBUGS</code>	122
4.5	Exercises	125
<b>5</b>	<b>Hierarchical modeling for univariate spatial data</b>	<b>129</b>
5.1	Stationary spatial process models	129
5.1.1	Isotropic models	130
5.1.2	Bayesian kriging in <code>WinBUGS</code>	134
5.1.3	More general isotropic correlation functions	135
5.1.4	Modeling geometric anisotropy	141
5.2	Generalized linear spatial process modeling	146
5.3	Nonstationary spatial process models $\star$	149
5.3.1	Deformation	151
5.3.2	Kernel mixing of process variables	152

5.3.3	Mixing of process distributions	156
5.4	Areal data models	158
5.4.1	Disease mapping	158
5.4.2	Traditional models and frequentist methods	158
5.4.3	Hierarchical Bayesian methods	159
5.5	General linear areal data modeling	167
5.6	Comparison of point-referenced and areal data models	168
5.7	Exercises	170
<b>6</b>	<b>Spatial misalignment</b>	<b>175</b>
6.1	Point-level modeling	176
6.1.1	Gaussian process models	176
6.1.2	Methodology for the point-level realignment	178
6.2	Nested block-level modeling	182
6.3	Nonnested block-level modeling	192
6.3.1	Motivating data set	193
6.3.2	Methodology for nonnested block-level realignment	195
6.4	Misaligned regression modeling	206
6.5	Exercises	212
<b>7</b>	<b>Multivariate spatial modeling</b>	<b>217</b>
7.1	Separable models	218
7.1.1	Spatial prediction, interpolation, and regression	220
7.1.2	Regression in the Gaussian case	222
7.1.3	Avoiding the symmetry of the cross-covariance matrix	224
7.1.4	Regression in a probit model	225
7.2	Coregionalization models $\star$	230
7.2.1	Coregionalization models and their properties	230
7.2.2	Unconditional and conditional Bayesian specifications	233
7.2.3	Spatially varying coregionalization models	237
7.2.4	Model-fitting issues	237
7.3	Other constructive approaches $\star$	243
7.4	Multivariate models for areal data	245
7.4.1	Motivating data set	246
7.4.2	Multivariate CAR (MCAR) theory	247
7.4.3	Modeling issues	249
7.5	Exercises	253
<b>8</b>	<b>Spatiotemporal modeling</b>	<b>255</b>
8.1	General modeling formulation	257
8.1.1	Preliminary analysis	257
8.1.2	Model formulation	258
8.1.3	Associated distributional results	261
8.1.4	Prediction and forecasting	263

8.2	Point-level modeling with continuous time	267
8.3	Nonseparable spatiotemporal models ★	274
8.4	Dynamic spatiotemporal models ★	275
	8.4.1 Brief review of dynamic linear models	276
	8.4.2 Formulation for spatiotemporal models	277
8.5	Block-level modeling	282
	8.5.1 Aligned data	282
	8.5.2 Misalignment across years	287
	8.5.3 Nested misalignment both within and across years	290
	8.5.4 Nonnested misalignment and regression	293
8.6	Exercises	298
<b>9</b>	<b>Spatial survival models</b>	<b>301</b>
9.1	Parametric models	303
	9.1.1 Univariate spatial frailty modeling	303
	9.1.2 Spatial frailty versus logistic regression models	308
9.2	Semiparametric models	311
	9.2.1 Beta mixture approach	311
	9.2.2 Counting process approach	313
9.3	Spatiotemporal models	315
9.4	Multivariate models ★	320
9.5	Spatial cure rate models ★	328
	9.5.1 Models for right- and interval-censored data	329
	9.5.2 Spatial frailties in cure rate models	333
	9.5.3 Model comparison	334
9.6	Exercises	339
<b>10</b>	<b>Special topics in spatial process modeling</b>	<b>343</b>
10.1	Process smoothness revisited ★	344
	10.1.1 Smoothness of a univariate spatial process	344
	10.1.2 Directional finite difference and derivative processes	346
	10.1.3 Distribution theory	348
	10.1.4 Directional derivative processes in modeling	350
10.2	Spatially varying coefficient models	356
	10.2.1 Approach for a single covariate	357
	10.2.2 Multivariate spatially varying coefficient models	359
	10.2.3 Spatiotemporal data	361
	10.2.4 Generalized linear model setting	366
10.3	Spatial CDFs	367
	10.3.1 Basic definitions and motivating data sets	367
	10.3.2 Derived-process spatial CDFs	370
	10.3.3 Randomly weighted SCDFs	372
	<b>Appendices</b>	<b>377</b>

<b>A Matrix theory and spatial computing methods</b>	<b>379</b>
A.1 Gaussian elimination and LU decomposition	379
A.2 Inverses and determinants	383
A.3 Cholesky decomposition	384
A.4 Fast Fourier transforms	385
A.5 Strategies for large spatial and spatiotemporal data sets	387
A.5.1 Subsampling	387
A.5.2 Spectral methods	388
A.5.3 Lattice methods	389
A.5.4 Dimension reduction	390
A.5.5 Coarse-fine coupling	391
A.6 Slice Gibbs sampling for spatial process model fitting	392
A.6.1 Constant mean process with nugget	395
A.6.2 Mean structure process with no pure error component	396
A.6.3 Mean structure process with nugget	397
A.7 Structured MCMC sampling for areal model fitting	398
A.7.1 Applying structured MCMC to areal data	400
A.7.2 Algorithmic schemes	402
<b>B Answers to selected exercises</b>	<b>405</b>
<b>References</b>	<b>423</b>
<b>Author index</b>	<b>443</b>
<b>Subject index</b>	<b>449</b>

---

# Preface

---

As recently as two decades ago, the impact of hierarchical Bayesian methods outside of a small group of theoretical probabilists and statisticians was minimal at best. Realistic models for challenging data sets were easy enough to write down, but the computations associated with these models required integrations over hundreds or even thousands of unknown parameters, far too complex for existing computing technology. Suddenly, around 1990, the “Markov chain Monte Carlo (MCMC) revolution” in Bayesian computing took place. Methods like the Gibbs sampler and the Metropolis algorithm, when coupled with ever-faster workstations and personal computers, enabled evaluation of the integrals that had long thwarted applied Bayesians. Almost overnight, Bayesian methods became not only feasible, but the method of choice for almost any model involving multiple levels incorporating random effects or complicated dependence structures. The growth in applications has also been phenomenal, with a particularly interesting recent example being a Bayesian program to delete spam from your incoming email (see [popfile.sourceforge.net](http://popfile.sourceforge.net)).

Our purpose in writing this book is to describe hierarchical Bayesian methods for one class of applications in which they can pay substantial dividends: spatial (and spatiotemporal) statistics. While all three of us have been working in this area for some time, our motivation for writing the book really came from our experiences teaching courses on the subject (two of us at the University of Minnesota, and the other at the University of Connecticut). In teaching we naturally began with the textbook by Cressie (1993), long considered the standard as both text and reference in the field. But we found the book somewhat uneven in its presentation, and written at a mathematical level that is perhaps a bit high, especially for the many epidemiologists, environmental health researchers, foresters, computer scientists, GIS experts, and other users of spatial methods who lacked significant background in mathematical statistics. Now a decade old, the book also lacks a current view of hierarchical modeling approaches for spatial data.

But the problem with the traditional teaching approach went beyond the mere need for a less formal presentation. Time and again, as we presented

the traditional material, we found it wanting in terms of its flexibility to deal with realistic assumptions. Traditional Gaussian kriging is obviously the most important method of point-to-point spatial interpolation, but extending the paradigm beyond this was awkward. For areal (block-level) data, the problem seemed even more acute: CAR models should most naturally appear as priors for the parameters in a model, not as a model for the observations themselves.

This book, then, attempts to remedy the situation by providing a fully Bayesian treatment of spatial methods. We begin in Chapter 1 by outlining and providing illustrative examples of the three types of spatial data: point-level (geostatistical), areal (lattice), and spatial point process. We also provide a brief introduction to map projection and the proper calculation of distance on the earth's surface (which, since the earth is round, can differ markedly from answers obtained using the familiar notion of Euclidean distance). Our statistical presentation begins in earnest in Chapter 2, where we describe both exploratory data analysis tools and traditional modeling approaches for point-referenced data. Modeling approaches from traditional geostatistics (variogram fitting, kriging, and so forth) are covered here. Chapter 3 offers a similar presentation for areal data models, again starting with choropleth maps and other displays and progressing toward more formal statistical models. This chapter also presents Brook's Lemma and Markov random fields, topics that underlie the conditional, intrinsic, and simultaneous autoregressive (CAR, IAR, and SAR) models so often used in areal data settings.

Chapter 4 provides a review of the hierarchical Bayesian approach in a fairly generic setting, for readers previously unfamiliar with these methods and related computing and software. (The penultimate sections of Chapters 2, 3, and 4 offer tutorials in several popular software packages.) This chapter is not intended as a replacement for a full course in Bayesian methods (as covered, for example, by Carlin and Louis, 2000, or Gelman et al., 2004), but should be sufficient for readers having at least some familiarity with the ideas. In Chapter 5 then we are ready to cover hierarchical modeling for univariate spatial response data, including Bayesian kriging and lattice modeling. The issue of nonstationarity (and how to model it) also arises here.

Chapter 6 considers the problem of spatially misaligned data. Here, Bayesian methods are particularly well suited to sorting out complex interrelationships and constraints and providing a coherent answer that properly accounts for all spatial correlation and uncertainty. Methods for handling multivariate spatial responses (for both point- and block-level data) are discussed in Chapter 7. Spatiotemporal models are considered in Chapter 8, while Chapter 9 presents an extended application of areal unit data modeling in the context of survival analysis methods. Chapter 10 considers novel methodology associated with spatial process modeling, including spa-

tial directional derivatives, spatially varying coefficient models, and spatial cumulative distribution functions (SCDFs). Finally, the book also features two useful appendices. Appendix A reviews elements of matrix theory and important related computational techniques, while Appendix B contains solutions to several of the exercises in each of the book's chapters.

Our book is intended as a research monograph, presenting the “state of the art” in hierarchical modeling for spatial data, and as such we hope readers will find it useful as a desk reference. However, we also hope it will be of benefit to instructors (or self-directed students) wishing to use it as a textbook. Here we see several options. Students wanting an introduction to methods for point-referenced data (traditional geostatistics and its extensions) may begin with Chapter 1, Chapter 2, Chapter 4, and Section 5.1 to Section 5.3. If areal data models are of greater interest, we suggest beginning with Chapter 1, Chapter 3, Chapter 4, Section 5.4, and Section 5.5. In addition, for students wishing to minimize the mathematical presentation, we have also marked sections containing more advanced material with a star ( $\star$ ). These sections may be skipped (at least initially) at little cost to the intelligibility of the subsequent narrative. In our course in the Division of Biostatistics at the University of Minnesota, we are able to cover much of the book in a 3-credit-hour, single-semester (15-week) course. We encourage the reader to check <http://www.biostat.umn.edu/~brad/> on the web for many of our data sets and other teaching-related information.

We owe a debt of gratitude to those who helped us make this book a reality. Kirsty Stroud and Bob Stern took us to lunch and said encouraging things (and more importantly, picked up the check) whenever we needed it. Cathy Brown, Alex Zirpoli, and Desdamona Racheli prepared significant portions of the text and figures. Many of our current and former graduate and postdoctoral students, including Yue Cui, Xu Guo, Murali Haran, Xiaoping Jin, Andy Mugglin, Margaret Short, Amy Xia, and Li Zhu at Minnesota, and Deepak Agarwal, Mark Ecker, Sujit Ghosh, Hyon-Jung Kim, Ananda Majumdar, Alexandra Schmidt, and Shanshan Wu at the University of Connecticut, played a big role. We are also grateful to the Spring 2003 *Spatial Biostatistics* class in the School of Public Health at the University of Minnesota for taking our draft for a serious “test drive.” Colleagues Jarrett Barber, Nicky Best, Montserrat Fuentes, David Higdon, Jim Hodges, Oli Schabenberger, John Silander, Jon Wakefield, Melanie Wall, Lance Waller, and many others provided valuable input and assistance. Finally, we thank our families, whose ongoing love and support made all of this possible.

SUDIPTO BANERJEE  
BRADLEY P. CARLIN  
ALAN E. GELFAND

Minneapolis, Minnesota  
Durham, North Carolina  
October 2003