
Contents

Preface to the Second Edition	xvii
Preface to the First Edition	xix
1 Overview of spatial data problems	1
1.1 Introduction to spatial data and models	1
1.1.1 Point-level models	5
1.1.2 Areal models	5
1.1.3 Point process models	6
1.1.4 Software and datasets	7
1.2 Fundamentals of cartography	8
1.2.1 Map projections	8
1.2.2 Calculating distances on the earth's surface	13
1.3 Maps and geodesics in R	16
1.4 Exercises	20
2 Basics of point-referenced data models	23
2.1 Elements of point-referenced modeling	23
2.1.1 Stationarity	23
2.1.2 Variograms	24
2.1.3 Isotropy	25
2.1.4 Variogram model fitting	30
2.2 Anisotropy	31
2.2.1 Geometric anisotropy	31
2.2.2 Other notions of anisotropy	32
2.3 Exploratory approaches for point-referenced data	32
2.3.1 Basic techniques	32
2.3.2 Assessing anisotropy	36
2.3.2.1 Directional semivariograms and rose diagrams	37
2.3.2.2 Empirical semivariogram contour (ESC) plots	38
2.4 Classical spatial prediction	40
2.4.0.3 Noiseless kriging	43
2.5 Computer tutorials	44
2.5.1 EDA and spatial data visualization in R	44
2.5.2 Variogram analysis in R	46
2.6 Exercises	50
3 ★ Some theory for point-referenced data models	53
3.1 Formal modeling theory for spatial processes	53
3.1.1 Some basic stochastic process theory for spatial processes	55
3.1.2 Covariance functions and spectra	57
3.1.2.1 More general isotropic correlation functions	60

3.1.3	Constructing valid covariance functions	60
3.1.4	Smoothness of process realizations	61
3.1.5	Directional derivative processes	63
3.2	Nonstationary spatial process models *	63
3.2.1	Deformation	64
3.2.2	Nonstationarity through kernel mixing of process variables	65
3.2.3	Mixing of process distributions	69
3.3	Exercises	70
4	Basics of areal data models	73
4.1	Exploratory approaches for areal data	74
4.1.1	Measures of spatial association	75
4.1.2	Spatial smoothers	77
4.2	Brook's Lemma and Markov random fields	78
4.3	Conditionally autoregressive (CAR) models	80
4.3.1	The Gaussian case	81
4.3.2	The non-Gaussian case	84
4.4	Simultaneous autoregressive (SAR) models	85
4.4.1	CAR versus SAR models	87
4.4.2	STAR models	87
4.5	Computer tutorials	88
4.5.1	Adjacency matrices from maps using <code>spdep</code>	89
4.5.2	Moran's I and Geary's C in <code>spdep</code>	90
4.5.3	SAR and CAR model fitting using <code>spdep</code> in R	90
4.6	Exercises	95
5	Basics of Bayesian inference	97
5.1	Introduction to hierarchical modeling and Bayes' Theorem	97
5.1.0.1	Illustrations of Bayes' Theorem	98
5.2	Bayesian inference	100
5.2.1	Point estimation	100
5.2.2	Interval estimation	101
5.2.3	Hypothesis testing and model choice	101
5.2.3.1	Bayes factors	102
5.2.3.2	The DIC criterion	103
5.2.3.3	Posterior predictive loss criterion	105
5.2.3.4	Model assessment using hold out data	106
5.3	Bayesian computation	107
5.3.1	The Gibbs sampler	108
5.3.2	The Metropolis-Hastings algorithm	109
5.3.3	Slice sampling	111
5.3.4	Convergence diagnosis	112
5.3.5	Variance estimation	113
5.4	Computer tutorials	115
5.4.1	Basic Bayesian modeling in R	115
5.4.2	Advanced Bayesian modeling in WinBUGS	116
5.5	Exercises	119

6	Hierarchical modeling for univariate spatial data	123
6.1	Stationary spatial process models	123
6.1.1	Isotropic models	124
6.1.1.1	Prior specification	124
6.1.2	Bayesian kriging in WinBUGS	127
6.1.3	More general isotropic correlation functions, revisited	129
6.1.4	Modeling geometric anisotropy	133
6.2	Generalized linear spatial process modeling	136
6.3	Fitting hierarchical models for point-referenced data in spBayes	139
6.3.1	Gaussian spatial regression models	139
6.3.1.1	Prediction	143
6.3.1.2	Model selection	145
6.3.2	Non-Gaussian spatial GLM	147
6.4	Areal data models	151
6.4.1	Disease mapping	151
6.4.2	Traditional models and frequentist methods	152
6.4.3	Hierarchical Bayesian methods	152
6.4.3.1	Poisson-gamma model	153
6.4.3.2	Poisson-lognormal models	154
6.4.3.3	CAR models and their difficulties	155
6.4.4	Extending the CAR model	159
6.5	General linear areal data modeling	160
6.6	Comparison of point-referenced and areal data models	161
6.7	Exercises	162
7	Spatial misalignment	165
7.1	Point-level modeling	166
7.1.1	Gaussian process models	166
7.1.1.1	Motivating data set	166
7.1.1.2	Model Assumptions and Analytic Goals	167
7.1.2	Methodology for the point-level realignment	168
7.2	Nested block-level modeling	172
7.2.0.1	Methodology for nested block-level realignment	174
7.2.0.2	Individual block group estimation	177
7.2.0.3	Aggregate estimation: Block groups near the Ithaca, NY, waste site	178
7.3	Nonnested block-level modeling	180
7.3.1	Motivating data set	180
7.3.2	Methodology for nonnested block-level realignment	182
7.3.2.1	Total population interpolation model	186
7.3.2.2	Age and sex effects	189
7.4	A data assimilation example	190
7.5	Misaligned regression modeling	191
7.6	Exercises	195
8	Modeling and Analysis for Point Patterns	199
8.1	Introduction	199
8.2	Theoretical development	203
8.2.1	Counting measure	204
8.2.2	Moment measures	205
8.3	Diagnostic tools	207
8.3.1	Exploratory data analysis; investigating complete spatial randomness	208

8.3.2	G and F functions	208
8.3.3	The K function	210
8.3.4	Empirical estimates of the intensity	213
8.4	Modeling point patterns; NHPP's and Cox processes	214
8.4.1	Parametric specifications	215
8.4.2	Nonparametric specifications	216
8.4.3	Bayesian modeling details	217
8.4.3.1	The "poor man's" version; revisiting the ecological fallacy	218
8.4.4	Examples	218
8.5	Generating point patterns	219
8.6	More general point pattern models	221
8.6.1	Cluster processes	221
8.6.1.1	Neyman-Scott processes	222
8.6.2	Shot noise processes	223
8.6.3	Gibbs processes	224
8.6.4	Further Bayesian model fitting and inference	225
8.6.5	Implementing fully Bayesian inference	226
8.6.6	An example	227
8.7	Marked point processes	228
8.7.1	Model specifications	228
8.7.2	Bayesian model fitting for marked point processes	229
8.7.3	Modeling clarification	230
8.7.4	Enriching intensities	232
8.7.4.1	Introducing non-spatial covariate information	233
8.7.4.2	Results of the analysis	235
8.8	Space-time point patterns	237
8.8.1	Space-time Poisson process models	238
8.8.2	Dynamic models for discrete time data	238
8.8.3	Space-time Cox process models using stochastic PDE's	239
8.8.3.1	Modeling the House Construction Data for Irving, TX	240
8.8.3.2	Results of the data analysis	242
8.9	Additional topics	242
8.9.1	Measurement error in point patterns	242
8.9.1.1	Modeling details	244
8.9.2	Presence-only data application	246
8.9.2.1	Probability model for presence locations	248
8.9.3	Scan statistics	250
8.9.4	Preferential sampling	252
8.10	Exercises	254
9	Multivariate spatial modeling for point-referenced data	257
9.1	Joint modeling in classical multivariate Geostatistics	257
9.1.1	Co-kriging	259
9.1.2	Intrinsic Multivariate Correlation and Nested Models	260
9.2	Some Theory for Cross Covariance Functions	261
9.3	Separable models	263
9.4	Spatial prediction, interpolation, and regression	264
9.4.1	Regression in the Gaussian case	266
9.4.2	Avoiding the symmetry of the cross-covariance matrix	268
9.4.3	Regression in a probit model	268
9.4.4	Examples	269
9.4.5	Conditional modeling	273

9.4.6	Spatial regression with kernel averaged predictors	274
9.5	Coregionalization models ★	278
9.5.1	Coregionalization models and their properties	278
9.5.2	Unconditional and conditional Bayesian specifications	281
9.5.2.1	Equivalence of likelihoods	281
9.5.2.2	Equivalence of prior specifications	282
9.6	Spatially varying coefficient models	283
9.6.1	Approach for a single covariate	285
9.6.2	Multivariate spatially varying coefficient models	286
9.6.3	Spatially varying coregionalization models	288
9.6.4	Model-fitting issues	288
9.6.4.1	Fitting the joint model	288
9.6.4.2	Fitting the conditional model	289
9.7	Other constructive approaches ★	293
9.7.1	Generalized linear model setting	296
9.8	Illustrating multivariate spatial modeling with <code>spBayes</code>	297
9.9	Exercises	301
10	Models for multivariate areal data	305
10.1	The multivariate CAR (MCAR) distribution	306
10.2	Modeling with a proper, non-separable MCAR distribution	308
10.3	Conditionally specified Generalized MCAR (GMCAR) distributions	312
10.4	Modeling using the GMCAR distribution	314
10.5	Illustration: Fitting conditional GMCAR to Minnesota cancer data	315
10.6	Coregionalized MCAR distributions	319
10.6.1	Case 1: Independent and identical latent CAR variables	320
10.6.2	Case 2: Independent but not identical latent CAR variables	320
10.6.3	Case 3: Dependent and not identical latent CAR variables	321
10.7	Modeling with coregionalized MCAR's	322
10.8	Illustrating coregionalized MCAR models with 3 cancers from Minnesota	324
10.9	Exercises	327
11	Spatiotemporal modeling	329
11.1	General modeling formulation	330
11.1.1	Preliminary analysis	330
11.1.2	Model formulation	331
11.1.3	Associated distributional results	333
11.1.4	Prediction and forecasting	335
11.2	Point-level modeling with continuous time	339
11.3	Nonseparable spatiotemporal models ★	343
11.4	Dynamic spatiotemporal models ★	344
11.4.1	Brief review of dynamic linear models	345
11.4.2	Formulation for spatiotemporal models	345
11.4.3	Spatiotemporal data	349
11.5	Fitting dynamic spatiotemporal models using <code>spbayes</code>	352
11.6	Geostatistical space-time modeling driven by differential equations	355
11.7	Areal unit space-time modeling	361
11.7.1	Aligned data	361
11.7.2	Misalignment across years	365
11.7.3	Nested misalignment both within and across years	367
11.7.4	Nonnested misalignment and regression	370
11.8	Areal-level continuous time modeling	373

11.8.1	Areally referenced temporal processes	373
11.8.2	Hierarchical modeling	376
11.9	Exercises	378
12	Modeling large spatial and spatiotemporal datasets	381
12.1	Introduction	381
12.2	Approximate likelihood approaches	381
12.2.1	Spectral methods	381
12.2.2	Lattice and conditional independence methods	382
12.2.3	INLA	383
12.2.4	Approximate likelihood	384
12.2.5	Variational Bayes algorithm for spatial models	384
12.2.6	Covariance tapering	386
12.3	Models for large spatial data: low rank models	386
12.3.1	Kernel-based dimension reduction	387
12.3.2	The Karhunen-Loève representation of Gaussian processes	388
12.4	Predictive process models	390
12.4.1	The predictive process	390
12.4.2	Properties of the predictive process	392
12.4.3	Biases in low-rank models and the bias-adjusted modified predictive process	393
12.4.4	Selection of knots	395
12.4.5	A simulation example using the two step analysis	397
12.4.6	Non-Gaussian first stage models	397
12.4.7	Spatiotemporal versions	398
12.4.8	Multivariate predictive process models	399
12.5	Modeling with the predictive process	400
12.6	Fitting a predictive process model in <code>spBayes</code>	404
12.7	Exercises	410
13	Spatial gradients and wombling	413
13.1	Introduction	413
13.2	Process smoothness revisited \star	415
13.3	Directional finite difference and derivative processes	417
13.4	Distribution theory for finite differences and directional gradients	418
13.5	Directional derivative processes in modeling	420
13.6	Illustration: Inference for differences and gradients	422
13.7	Curvilinear gradients and wombling	424
13.7.1	Gradients along curves	424
13.7.2	Wombling boundary	426
13.8	Distribution theory for curvilinear gradients	427
13.9	Illustration: Spatial boundaries for invasive plant species	429
13.10	Areal wombling	432
13.10.0.1	Review of existing methods	433
13.10.1	Simple MRF-based areal wombling	434
13.10.1.1	Adding covariates	437
13.10.2	Joint site-edge areal wombling	438
13.10.2.1	Edge smoothing and random neighborhood structure	439
13.10.2.2	Two-level CAR model	439
13.10.2.3	Site-edge (SE) models	440
13.10.3	FDR-based areal wombling	444
13.11	Wombling with point process data	445

13.12 Concluding remarks	445
14 Spatial survival models	447
14.1 Parametric models	448
14.1.1 Univariate spatial frailty modeling	448
14.1.1.1 Bayesian implementation	449
14.1.2 Spatial frailty versus logistic regression models	453
14.2 Semiparametric models	454
14.2.1 Beta mixture approach	455
14.2.2 Counting process approach	456
14.3 Spatiotemporal models	457
14.3.0.1 Results for the full model	459
14.3.0.2 Bayesian model choice	460
14.4 Multivariate models ★	461
14.4.0.3 Static spatial survival data with multiple causes of death	462
14.4.0.4 MCAR specification, simplification, and computing	462
14.4.0.5 Spatiotemporal survival data	463
14.5 Spatial cure rate models ★	467
14.5.1 Models for right- and interval-censored data	468
14.5.1.1 Right-censored data	468
14.5.1.2 Interval-censored data	470
14.5.2 Spatial frailties in cure rate models	471
14.5.3 Model comparison	471
14.6 Exercises	475
15 Special topics in spatial process modeling	479
15.1 Data assimilation	479
15.1.1 Algorithmic and pseudo-statistical approaches in weather prediction	479
15.1.2 Fusion modeling using stochastic integration	480
15.1.3 The downscaler	482
15.1.4 Spatio-temporal versions	484
15.1.5 An Illustration	485
15.2 Space-time modeling for extremes	486
15.2.1 Possibilities for modeling maxima	487
15.2.2 Review of Extreme Value Theory	488
15.2.3 A Continuous spatial process model	489
15.2.4 Using copulas	490
15.2.5 Hierarchical modeling for spatial extreme values	491
15.3 Spatial CDFs	492
15.3.1 Basic definitions and motivating data sets	492
15.3.2 Derived-process spatial CDFs	495
15.3.2.1 Point- versus block-level spatial CDFs	495
15.3.2.2 Covariate weighted SCDFs for misaligned data	496
15.3.3 Randomly weighted SCDFs	497
Appendices	501
A Spatial computing methods	503
A.1 Fast Fourier transforms	503
A.2 Slice Gibbs sampling for spatial process model fitting	504
A.2.1 Constant mean process with nugget	507
A.2.2 Mean structure process with no pure error component	508

xvi		
	A.2.3 Mean structure process with nugget	509
	A.3 Structured MCMC sampling for areal model fitting	509
	A.3.0.1 SMCMC algorithm basics	510
	A.3.1 Applying structured MCMC to areal data	510
	A.3.2 Algorithmic schemes	512
	A.4 spBayes: Under the hood.	513
	B Answers to selected exercises	515
	Bibliography	529
	Index	559

Preface to the Second Edition

In the ten years that have passed since the first edition of this book, we believe the statistical landscape has changed substantially, even more so for analyzing space and space-time data. Apart from the remarkable growth in data collection, with datasets now of enormous size, the fields of statistics and biostatistics are also witnessing a change toward examination of observational data, rather than restricting to carefully-collected experimentally designed data. We are witnessing an increased examination of complex systems using such data, requiring synthesis of multiple sources of information (empirical, theoretical, physical, etc.), necessitating the development of multi-level models. We are seeing repeated exemplification of the hierarchical framework $[data|process, parameters][process|parameters][parameters]$. The role of the statistician is evolving in this landscape to that of an integral participant in team-based research: a participant in the framing of the questions to be investigated, the determination of data needs to investigate these questions, the development of models to examine these questions, the development of strategies to fit these models, and the analysis and summarization of the resultant inference under these specifications. It is an exciting new world for modern statistics, and spatial analysis is a particularly important player in this new world due to the increased appreciation of the information carried in spatial locations, perhaps across temporal scales, in learning about these complex processes. Applications abound, particularly in the environmental sciences but also in public health, real estate, and many other fields.

We believe this new edition moves forward in this spirit. The first edition was intended as a research monograph, presenting a state-of-the-art treatment of hierarchical modeling for spatial data. It has been a delightful success, far exceeding our expectations in terms of sales and reception by the community. However, reflecting the decade that has passed, we have made consequential changes from the first edition. Not surprisingly, the new volume is more than 50% bigger, reflecting the major growth in spatial statistics as a research area and as an area of application.

Rather than describing the contents, chapter by chapter, we note the following major changes. First, we have added a much needed chapter on spatial point patterns. This is a subfield that is finding increased importance but, in terms of application, has lagged behind the use of point-referenced and areal unit data. We offer roughly 80 new pages here, developed primarily from a modeling perspective, introducing as much current hierarchical and Bayesian flavor as we could. Second, reflecting the ubiquitous increases in the sizes of datasets, we have developed a “big data” chapter. Here, we focus on the predictive process in its various forms, as an attractive tool for handling reasonably large datasets. Third, near the end of the book we have added a new chapter on spatial and spatiotemporal gradient modeling, with associated developments by us and others in spatial boundary analysis and wombling. As elsewhere in the book, we divide our descriptions here into those appropriate for point-referenced data (where underlying spatial processes guarantee the existence of spatial derivatives) and areal data (where processes are not possible but boundaries can still be determined based on alternate ways of hierarchically smoothing the areal map. Fourth, since geostatistical (point-referenced) modeling is still the most prevalent setting for spatial analysis, we have chosen to present this material in two separate chapters. The first (Chapter 2) is a basic introduction, presented for the reader who is more focused on the

practical side of things. In addition, we have developed a more theoretical chapter (Chapter 3) which provides much more insight into the scope of issues that arise in the geostatistical setting and how we deal with them formally. The presentation of this material is still gentle compared with that in many stochastic processes texts, and we hope it provides valuable model-building insight. At the same time, we recognize that Chapter 3 may be somewhat advanced for more introductory courses, so we mark it as a starred chapter. In addition to these four new chapters, we have greatly revised and expanded the multivariate and spatio-temporal chapters, again in response to the growth of work in these areas. We have also added two new special topics sections, one on data fusion/assimilation, and one on spatial analysis for data on extremes. We have roughly doubled the number of exercises in the book, and also include many more color figures, now integrated appropriately into the text. Finally, we have updated the computational aspects of the book. Specially, we work with the newest version of `WinBUGS`, the new flexible `spBayes` software, and we introduce other suitable `R` packages as needed, especially for exploratory data analysis.

In addition to those to whom we expressed our gratitude in the preface to the first edition, we now extend this list to record (in alphabetical order) the following colleagues, current and former post doctoral researchers and students: Dipankar Bandyopadhyay, Veronica Berrocal, Avishek Chakraborty, Jim Clark, Jason (Jun) Duan, David Dunson, Andrew Finley, Souparno Ghosh, Simone Gray, Rajarshi Guhaniyogi, Michele Guindani, Xiaoping Jin, Giovanna Jona Lasinio, Matt Heaton, Dave Holland, Thanasis Kottas, Andrew Latimer, Tommy Leininger, Pei Li, Shengde Liang, Haolan Lu, Kristian Lum, Haijun Ma, Marshall McBean, Marie Lynn Miranda, Joao Vitor Monteiro, XuanLong Nguyen, Lucia Paci, Sonia Petrone, Gavino Puggioni, Harrison Quick, Cavan Reilly, Qian Ren, Abel Rodriguez, Huiyan Sang, Sujit Sahu, Maria Terres, Beth Virnig, Fangpo Wang, Adam Wilson, Gangqiang Xia, and Kai Zhu. In addition, we much appreciate the continuing support of CRC/Chapman and Hall in helping to bring this new edition to fruition, in particular the encouragement of the steadfast and indefatigable Rob Calver.

SUDIPTO BANERJEE
BRADLEY P. CARLIN
ALAN E. GELFAND

Minneapolis, Minnesota
Durham, North Carolina
July 2013