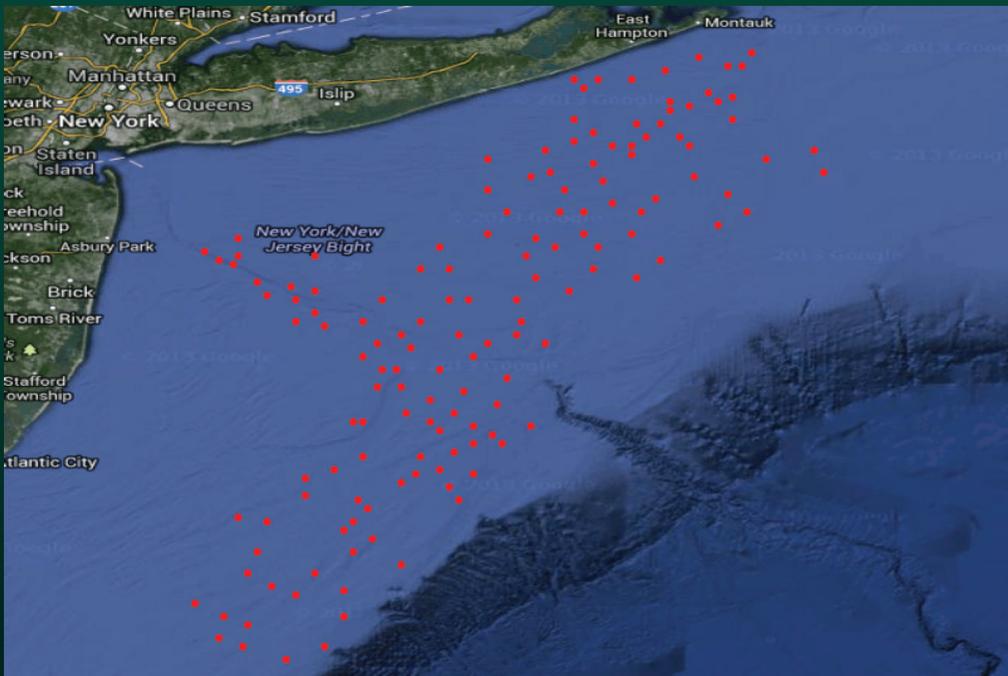


# Hierarchical Modeling and Analysis for Spatial Data

## Second Edition



**Sudipto Banerjee**  
**Bradley P. Carlin**  
**Alan E. Gelfand**



**Hierarchical Modeling  
and Analysis for  
Spatial Data**  
Second Edition

# MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

General Editors

**F. Bunea, V. Isham, N. Keiding, T. Louis, R. L. Smith, and H. Tong**

1. Stochastic Population Models in Ecology and Epidemiology *M.S. Barlett* (1960)
2. Queues *D.R. Cox and W.L. Smith* (1961)
3. Monte Carlo Methods *J.M. Hammersley and D.C. Handscomb* (1964)
4. The Statistical Analysis of Series of Events *D.R. Cox and P.A.W. Lewis* (1966)
5. Population Genetics *W.J. Ewens* (1969)
6. Probability, Statistics and Time *M.S. Barlett* (1975)
7. Statistical Inference *S.D. Silvey* (1975)
8. The Analysis of Contingency Tables *B.S. Everitt* (1977)
9. Multivariate Analysis in Behavioural Research *A.E. Maxwell* (1977)
10. Stochastic Abundance Models *S. Engen* (1978)
11. Some Basic Theory for Statistical Inference *E.J.G. Pitman* (1979)
12. Point Processes *D.R. Cox and V. Isham* (1980)
13. Identification of Outliers *D.M. Hawkins* (1980)
14. Optimal Design *S.D. Silvey* (1980)
15. Finite Mixture Distributions *B.S. Everitt and D.J. Hand* (1981)
16. Classification *A.D. Gordon* (1981)
17. Distribution-Free Statistical Methods, 2nd edition *J.S. Maritz* (1995)
18. Residuals and Influence in Regression *R.D. Cook and S. Weisberg* (1982)
19. Applications of Queuing Theory, 2nd edition *G.F. Newell* (1982)
20. Risk Theory, 3rd edition *R.E. Beard, T. Pentikäinen and E. Pesonen* (1984)
21. Analysis of Survival Data *D.R. Cox and D. Oakes* (1984)
22. An Introduction to Latent Variable Models *B.S. Everitt* (1984)
23. Bandit Problems *D.A. Berry and B. Fristedt* (1985)
24. Stochastic Modelling and Control *M.H.A. Davis and R. Vinter* (1985)
25. The Statistical Analysis of Composition Data *J. Aitchison* (1986)
26. Density Estimation for Statistics and Data Analysis *B.W. Silverman* (1986)
27. Regression Analysis with Applications *G.B. Wetherill* (1986)
28. Sequential Methods in Statistics, 3rd edition *G.B. Wetherill and K.D. Glazebrook* (1986)
29. Tensor Methods in Statistics *P. McCullagh* (1987)
30. Transformation and Weighting in Regression *R.J. Carroll and D. Ruppert* (1988)
31. Asymptotic Techniques for Use in Statistics *O.E. Bandorff-Nielsen and D.R. Cox* (1989)
32. Analysis of Binary Data, 2nd edition *D.R. Cox and E.J. Snell* (1989)
33. Analysis of Infectious Disease Data *N.G. Becker* (1989)
34. Design and Analysis of Cross-Over Trials *B. Jones and M.G. Kenward* (1989)
35. Empirical Bayes Methods, 2nd edition *J.S. Maritz and T. Lwin* (1989)
36. Symmetric Multivariate and Related Distributions *K.T. Fang, S. Kotz and K.W. Ng* (1990)
37. Generalized Linear Models, 2nd edition *P. McCullagh and J.A. Nelder* (1989)
38. Cyclic and Computer Generated Designs, 2nd edition *J.A. John and E.R. Williams* (1995)
39. Analog Estimation Methods in Econometrics *C.F. Manski* (1988)
40. Subset Selection in Regression *A.J. Miller* (1990)
41. Analysis of Repeated Measures *M.J. Crowder and D.J. Hand* (1990)
42. Statistical Reasoning with Imprecise Probabilities *P. Walley* (1991)
43. Generalized Additive Models *T.J. Hastie and R.J. Tibshirani* (1990)
44. Inspection Errors for Attributes in Quality Control *N.L. Johnson, S. Kotz and X. Wu* (1991)
45. The Analysis of Contingency Tables, 2nd edition *B.S. Everitt* (1992)

46. The Analysis of Quantal Response Data *B.J.T. Morgan* (1992)
47. Longitudinal Data with Serial Correlation—A State-Space Approach *R.H. Jones* (1993)
48. Differential Geometry and Statistics *M.K. Murray and J.W. Rice* (1993)
49. Markov Models and Optimization *M.H.A. Davis* (1993)
50. Networks and Chaos—Statistical and Probabilistic Aspects  
*O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall* (1993)
51. Number-Theoretic Methods in Statistics *K.-T. Fang and Y. Wang* (1994)
52. Inference and Asymptotics *O.E. Barndorff-Nielsen and D.R. Cox* (1994)
53. Practical Risk Theory for Actuaries *C.D. Daykin, T. Pentikäinen and M. Pesonen* (1994)
54. Biplots *J.C. Gower and D.J. Hand* (1996)
55. Predictive Inference—An Introduction *S. Geisser* (1993)
56. Model-Free Curve Estimation *M.E. Tarter and M.D. Lock* (1993)
57. An Introduction to the Bootstrap *B. Efron and R.J. Tibshirani* (1993)
58. Nonparametric Regression and Generalized Linear Models *P.J. Green and B.W. Silverman* (1994)
59. Multidimensional Scaling *T.F. Cox and M.A.A. Cox* (1994)
60. Kernel Smoothing *M.P. Wand and M.C. Jones* (1995)
61. Statistics for Long Memory Processes *J. Beran* (1995)
62. Nonlinear Models for Repeated Measurement Data *M. Davidian and D.M. Giltinan* (1995)
63. Measurement Error in Nonlinear Models *R.J. Carroll, D. Rupert and L.A. Stefanski* (1995)
64. Analyzing and Modeling Rank Data *J.J. Marden* (1995)
65. Time Series Models—In Econometrics, Finance and Other Fields  
*D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen* (1996)
66. Local Polynomial Modeling and its Applications *J. Fan and I. Gijbels* (1996)
67. Multivariate Dependencies—Models, Analysis and Interpretation *D.R. Cox and N. Wermuth* (1996)
68. Statistical Inference—Based on the Likelihood *A. Azzalini* (1996)
69. Bayes and Empirical Bayes Methods for Data Analysis *B.P. Carlin and T.A. Louis* (1996)
70. Hidden Markov and Other Models for Discrete-Valued Time Series *I.L. MacDonald and W. Zucchini* (1997)
71. Statistical Evidence—A Likelihood Paradigm *R. Royall* (1997)
72. Analysis of Incomplete Multivariate Data *J.L. Schafer* (1997)
73. Multivariate Models and Dependence Concepts *H. Joe* (1997)
74. Theory of Sample Surveys *M.E. Thompson* (1997)
75. Retrial Queues *G. Falin and J.G.C. Templeton* (1997)
76. Theory of Dispersion Models *B. Jørgensen* (1997)
77. Mixed Poisson Processes *J. Grandell* (1997)
78. Variance Components Estimation—Mixed Models, Methodologies and Applications *P.S.R.S. Rao* (1997)
79. Bayesian Methods for Finite Population Sampling *G. Meeden and M. Ghosh* (1997)
80. Stochastic Geometry—Likelihood and computation  
*O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout* (1998)
81. Computer-Assisted Analysis of Mixtures and Applications—Meta-Analysis, Disease Mapping and Others  
*D. Böhning* (1999)
82. Classification, 2nd edition *A.D. Gordon* (1999)
83. Semimartingales and their Statistical Inference *B.L.S. Prakasa Rao* (1999)
84. Statistical Aspects of BSE and vCJD—Models for Epidemics *C.A. Donnelly and N.M. Ferguson* (1999)
85. Set-Indexed Martingales *G. Ivanoff and E. Merzbach* (2000)
86. The Theory of the Design of Experiments *D.R. Cox and N. Reid* (2000)
87. Complex Stochastic Systems *O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg* (2001)
88. Multidimensional Scaling, 2nd edition *T.F. Cox and M.A.A. Cox* (2001)
89. Algebraic Statistics—Computational Commutative Algebra in Statistics  
*G. Pistone, E. Riccomagno and H.P. Wynn* (2001)
90. Analysis of Time Series Structure—SSA and Related Techniques  
*N. Golyandina, V. Nekrutkin and A.A. Zhigljavsky* (2001)
91. Subjective Probability Models for Lifetimes *Fabio Spizzichino* (2001)
92. Empirical Likelihood *Art B. Owen* (2001)
93. Statistics in the 21st Century *Adrian E. Raftery, Martin A. Tanner, and Martin T. Wells* (2001)

94. Accelerated Life Models: Modeling and Statistical Analysis  
*Vilijandas Bagdonavicius and Mikhail Nikulin* (2001)
95. Subset Selection in Regression, Second Edition *Alan Miller* (2002)
96. Topics in Modelling of Clustered Data *Marc Aerts, Helena Geys, Geert Molenberghs, and Louise M. Ryan* (2002)
97. Components of Variance *D.R. Cox and P.J. Solomon* (2002)
98. Design and Analysis of Cross-Over Trials, 2nd Edition *Byron Jones and Michael G. Kenward* (2003)
99. Extreme Values in Finance, Telecommunications, and the Environment  
*Bärbel Finkenstädt and Holger Rootzén* (2003)
100. Statistical Inference and Simulation for Spatial Point Processes  
*Jesper Møller and Rasmus Plenge Waagepetersen* (2004)
101. Hierarchical Modeling and Analysis for Spatial Data  
*Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand* (2004)
102. Diagnostic Checks in Time Series *Wai Keung Li* (2004)
103. Stereology for Statisticians *Adrian Baddeley and Eva B. Vedel Jensen* (2004)
104. Gaussian Markov Random Fields: Theory and Applications *Håvard Rue and Leonhard Held* (2005)
105. Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition  
*Raymond J. Carroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu* (2006)
106. Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood  
*Youngjo Lee, John A. Nelder, and Yudi Pawitan* (2006)
107. Statistical Methods for Spatio-Temporal Systems  
*Bärbel Finkenstädt, Leonhard Held, and Valerie Isham* (2007)
108. Nonlinear Time Series: Semiparametric and Nonparametric Methods *Jiti Gao* (2007)
109. Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis  
*Michael J. Daniels and Joseph W. Hogan* (2008)
110. Hidden Markov Models for Time Series: An Introduction Using R  
*Walter Zucchini and Iain L. MacDonald* (2009)
111. ROC Curves for Continuous Data *Wojtek J. Krzanowski and David J. Hand* (2009)
112. Antedependence Models for Longitudinal Data *Dale L. Zimmerman and Vicente A. Núñez-Antón* (2009)
113. Mixed Effects Models for Complex Data *Lang Wu* (2010)
114. Introduction to Time Series Modeling *Genshiro Kitagawa* (2010)
115. Expansions and Asymptotics for Statistics *Christopher G. Small* (2010)
116. Statistical Inference: An Integrated Bayesian/Likelihood Approach *Murray Aitkin* (2010)
117. Circular and Linear Regression: Fitting Circles and Lines by Least Squares *Nikolai Chernov* (2010)
118. Simultaneous Inference in Regression *Wei Liu* (2010)
119. Robust Nonparametric Statistical Methods, Second Edition  
*Thomas P. Hetmansperger and Joseph W. McKean* (2011)
120. Statistical Inference: The Minimum Distance Approach  
*Ayanendranath Basu, Hiroyuki Shioya, and Chanseok Park* (2011)
121. Smoothing Splines: Methods and Applications *Yuedong Wang* (2011)
122. Extreme Value Methods with Applications to Finance *Serguei Y. Novak* (2012)
123. Dynamic Prediction in Clinical Survival Analysis *Hans C. van Houwelingen and Hein Putter* (2012)
124. Statistical Methods for Stochastic Differential Equations  
*Mathieu Kessler, Alexander Lindner, and Michael Sørensen* (2012)
125. Maximum Likelihood Estimation for Sample Surveys  
*R. L. Chambers, D. G. Steel, Suojin Wang, and A. H. Welsh* (2012)
126. Mean Field Simulation for Monte Carlo Integration *Pierre Del Moral* (2013)
127. Analysis of Variance for Functional Data *Jin-Ting Zhang* (2013)
128. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition *Peter J. Diggle* (2013)
129. Constrained Principal Component Analysis and Related Techniques *Yoshio Takane* (2014)
130. Randomised Response-Adaptive Designs in Clinical Trials *Anthony C. Atkinson and Atanu Biswas* (2014)
131. Theory of Factorial Design: Single- and Multi-Stratum Experiments *Ching-Shui Cheng* (2014)
132. Quasi-Least Squares Regression *Justine Shults and Joseph M. Hilbe* (2014)
133. Data Analysis and Approximate Models: Model Choice, Location-Scale, Analysis of Variance, Nonparametric Regression and Image Analysis *Laurie Davies* (2014)
134. Dependence Modeling with Copulas *Harry Joe* (2014)
135. Hierarchical Modeling and Analysis for Spatial Data, Second Edition *Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand* (2014)

Monographs on Statistics and Applied Probability 135

# **Hierarchical Modeling and Analysis for Spatial Data**

## **Second Edition**

**Sudipto Banerjee**

Division of Biostatistics, School of Public Health  
University of Minnesota, Minneapolis, USA

**Bradley P. Carlin**

Division of Biostatistics, School of Public Health  
University of Minnesota, Minneapolis, USA

**Alan E. Gelfand**

Department of Statistical Science  
Duke University, Durham, North Carolina, USA



**CRC Press**

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

CRC Press  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC  
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works  
Version Date: 20140527

International Standard Book Number-13: 978-1-4398-1918-0 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Visit the Taylor & Francis Web site at**  
**<http://www.taylorandfrancis.com>**

**and the CRC Press Web site at**  
**<http://www.crcpress.com>**

TO SHARBANI, CAROLINE, AND MARIASUN



---

# Contents

---

<b>Preface to the Second Edition</b>	<b>xvii</b>
<b>Preface to the First Edition</b>	<b>xix</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	4
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	19
<b>2 Basics of point-referenced data models</b>	<b>23</b>
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	37
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	47
2.6 Exercises	50
<b>3 Some theory for point-referenced data models</b>	<b>53</b>
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 $\star$	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
<b>4</b>	<b>Basics of areal data models</b>	<b>73</b>
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
<b>5</b>	<b>Basics of Bayesian inference</b>	<b>97</b>
5.1	Introduction to hierarchical modeling and Bayes' Theorem	97
5.1.1	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	118

<b>6</b>	<b>Hierarchical modeling for univariate spatial data</b>	<b>123</b>
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	146
6.4	Areal data models	150
6.4.1	Disease mapping	150
6.4.2	Traditional models and frequentist methods	151
6.4.3	Hierarchical Bayesian methods	152
6.4.3.1	Poisson-gamma model	152
6.4.3.2	Poisson-lognormal models	153
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	160
6.7	Exercises	161
<b>7</b>	<b>Spatial misalignment</b>	<b>165</b>
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	173
7.2.1	Methodology for nested block-level realignment	173
7.2.2	Individual block group estimation	177
7.2.3	Aggregate estimation: Block groups near the Ithaca, NY, waste site	179
7.3	Nonnested block-level modeling	179
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	188
7.4	A data assimilation example	189
7.5	Misaligned regression modeling	190
7.6	Exercises	195
<b>8</b>	<b>Modeling and Analysis for Point Patterns</b>	<b>199</b>
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	213
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	220
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	226
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	231
	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	247
8.9.3	Scan statistics	251
8.9.4	Preferential sampling	252
8.10	Exercises	253
<b>9</b>	<b>Multivariate spatial modeling for point-referenced data</b>	<b>257</b>
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	272
9.4.6	Spatial regression with kernel averaged predictors	275
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
<b>10</b>	<b>Models for multivariate areal data</b>	<b>305</b>
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	311
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	319
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 three cancers from Minnesota	324
10.9	Exercises	327
<b>11</b>	<b>Spatiotemporal modeling</b>	<b>329</b>
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	348
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	374
11.8.2	Hierarchical modeling	376
11.9	Exercises	378
<b>12</b>	<b>Modeling large spatial and spatiotemporal datasets</b>	<b>381</b>
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	411
<b>13</b>	<b>Spatial gradients and wombling</b>	<b>413</b>
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
13.10.1	Review of existing methods	433
13.10.2	Simple MRF-based areal wombling	434
13.10.2.1	Adding covariates	437
13.10.3	Joint site-edge areal wombling	438

13.10.3.1	Edge smoothing and random neighborhood structure	439
13.10.3.2	Two-level CAR model	439
13.10.3.3	Site-edge (SE) models	440
13.10.4	FDR-based areal wombling	444
13.11	Wombling with point process data	445
13.12	Concluding remarks	445
<b>14</b>	<b>Spatial survival models</b>	<b>447</b>
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
14.3.1	Results for the full model	459
14.3.2	Bayesian model choice	460
14.4	Multivariate models $\star$	462
14.4.1	Static spatial survival data with multiple causes of death	462
14.4.2	MCAR specification, simplification, and computing	462
14.4.3	Spatiotemporal survival data	463
14.5	Spatial cure rate models $\star$	466
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	471
14.5.2	Spatial frailties in cure rate models	471
14.5.3	Model comparison	472
14.6	Exercises	475
<b>15</b>	<b>Special topics in spatial process modeling</b>	<b>479</b>
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	Spatiotemporal 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 CDF's	492
15.3.1	Basic definitions and motivating data sets	492
15.3.2	Derived-process spatial CDF's	495
15.3.2.1	Point- versus block-level spatial CDF's	495
15.3.2.2	Covariate weighted SCDF's for misaligned data	496
15.3.3	Randomly weighted SCDF's	496

<b>Appendices</b>	<b>501</b>
<b>A Spatial computing methods</b>	<b>503</b>
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
A.2.3 Mean structure process with nugget	509
A.3 Structured MCMC sampling for areal model fitting	509
A.3.1 SMCMC algorithm basics	510
A.3.2 Applying structured MCMC to areal data	510
A.3.3 Algorithmic schemes	512
A.4 spBayes: Under the hood	513
<b>B Answers to selected exercises</b>	<b>515</b>
<b>Bibliography</b>	<b>529</b>

---

# 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 being restricted 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 on 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 marked 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 postdoctoral 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

---

# Preface to the First Edition

---

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 4 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 5 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, 4, and 5 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 6 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 7 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 9. Spatiotemporal models are considered in Chapter 11, while Chapter 14 presents an extended application of areal unit data modeling in the context of survival analysis methods. Chapter 15 considers novel methodology associated with spatial process modeling, including spatial directional derivatives, spatially varying coefficient models, and spatial cumulative distribution functions (SCDF's). 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 5, and Section 6.1 to Section 3.2. If areal data models are of greater interest, we suggest beginning with Chapter 1, Chapter 4, Chapter 5, Section 6.4, and Section 6.5. 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



# Overview of spatial data problems

---

## 1.1 Introduction to spatial data and models

Researchers in diverse areas such as climatology, ecology, environmental health, and real estate marketing are increasingly faced with the task of analyzing data that are:

- highly multivariate, with many important predictors and response variables,
- geographically referenced, and often presented as maps, and
- temporally correlated, as in longitudinal or other time series structures.

For example, for an epidemiological investigation, we might wish to analyze lung, breast, colorectal, and cervical cancer rates by county and year in a particular state, with smoking, mammography, and other important screening and staging information also available at some level. Public health professionals who collect such data are charged not only with surveillance, but also statistical *inference* tasks, such as *modeling* of trends and correlation structures, *estimation* of underlying model parameters, *hypothesis testing* (or comparison of competing models), and *prediction* of observations at unobserved times or locations.

In this text we seek to present a practical, self-contained treatment of hierarchical modeling and data analysis for complex spatial (and spatiotemporal) datasets. Spatial statistics methods have been around for some time, with the landmark work by Cressie (1993) providing arguably the only comprehensive book in the area. However, recent developments in Markov chain Monte Carlo (MCMC) computing now allow fully Bayesian analyses of sophisticated multilevel models for complex geographically referenced data. This approach also offers full inference for non-Gaussian spatial data, multivariate spatial data, spatiotemporal data, and, for the first time, solutions to problems such as geographic and temporal misalignment of spatial data layers.

This book does not attempt to be fully comprehensive, but does attempt to present a fairly thorough treatment of hierarchical Bayesian approaches for handling all of these problems. The book's mathematical level is roughly comparable to that of Carlin and Louis (2000). That is, we sometimes state results rather formally, but spend little time on theorems and proofs. For more mathematical treatments of spatial statistics (at least on the geostatistical side), the reader is referred to Cressie (1993), Wackernagel (1998), Chiles and Delfiner (1999), and Stein (1999a). For more descriptive presentations the reader might consult Bailey and Gattrell (1995), Fotheringham and Rogerson (1994), or Haining (1990). Our primary focus is on the issues of *modeling* (where we offer rich, flexible classes of hierarchical structures to accommodate both static and dynamic spatial data), *computing* (both in terms of MCMC algorithms and methods for handling very large matrices), and *data analysis* (to illustrate the first two items in terms of inferential summaries and graphical displays). Reviews of both traditional spatial methods (Chapters 2, 3 and 4) and Bayesian methods (Chapter 5) attempt to ensure that previous exposure to either of these two areas is not required (though it will of course be helpful if available).

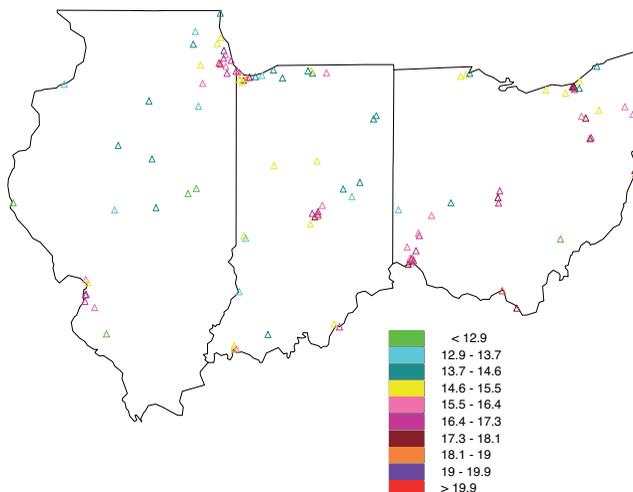


Figure 1.1 Map of  $PM_{2.5}$  sampling sites over three midwestern U.S. states; plotting character indicates range of average monitored  $PM_{2.5}$  level over the year 2001.

Following convention, we classify spatial data sets into one of three basic types:

- *point-referenced data*, where  $Y(\mathbf{s})$  is a random vector at a location  $\mathbf{s} \in \mathbb{R}^r$ , where  $\mathbf{s}$  varies *continuously* over  $D$ , a fixed subset of  $\mathbb{R}^r$  that contains an  $r$ -dimensional rectangle of positive volume;
- *areal data*, where  $D$  is again a fixed subset (of regular or irregular shape), but now partitioned into a finite number of areal units with well-defined boundaries;
- *point pattern data*, where now  $D$  is itself random; its index set gives the locations of random events that are the spatial point pattern.  $Y(\mathbf{s})$  itself can simply equal 1 for all  $\mathbf{s} \in D$  (indicating occurrence of the event), or possibly give some additional covariate information (producing a *marked point pattern process*).

The first case is often referred to as *geocoded* or *geostatistical* data, names apparently arising from the long history of these types of problems in mining and other geological sciences. Figure 1.1 offers an example of this case, showing the locations of 114 air-pollution monitoring sites in three midwestern U.S. states (Illinois, Indiana, and Ohio). The plotting character indicates the 2001 annual average  $PM_{2.5}$  level (measured in ppb) at each site.  $PM_{2.5}$  stands for particulate matter less than 2.5 microns in diameter, and is a measure of the density of very small particles that can travel through the nose and windpipe and into the lungs, potentially damaging a person's health. Here we might be interested in a model of the geographic distribution of these levels that account for spatial correlation and perhaps underlying covariates (regional industrialization, traffic density, and the like). The use of colors makes it somewhat easier to read, since the color allows the categories to be ordered more naturally, and helps sharpen the contrast between the urban and rural areas. Again, traditional analysis methods for point level data like this are described in Chapter 2, while Chapter 6 introduces the corresponding hierarchical modeling approach.

The second case above (areal data) is often referred to as *lattice* data, a term we find misleading since it connotes observations corresponding to “corners” of a checkerboard-like grid. Of course, there *are* data sets of this type; for example, as arising from agricultural field trials (where the plots cultivated form a regular lattice) or image restoration (where the data correspond to pixels on a screen, again in a regular lattice). However, in practice most areal data are summaries over an *irregular* lattice, like a collection of county or other

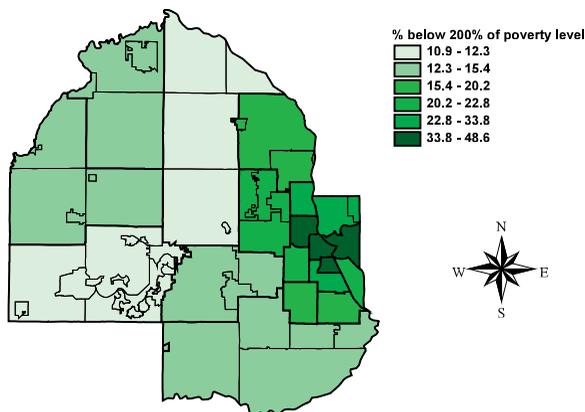


Figure 1.2 ArcView map of percent of surveyed population with household income below 200% of the federal poverty limit, regional survey units in Hennepin County, MN.

regional boundaries, as in Figure 1.2. Here we have information on the percent of a surveyed population with household income falling below 200% of the federal poverty limit for a collection of regions comprising Hennepin County, MN. Note that we have no information on any single household in the study area, only regional summaries for each region. Figure 1.2 is an example of a *choropleth map*, meaning that it uses shades of color (or greyscale) to classify values into a few broad classes (six in this case), like a histogram (bar chart) for nonspatial data. Choropleth maps are visually appealing (and therefore, also common), but of course provide a rather crude summary of the data, and one that can be easily altered simply by manipulating the class cutoffs.

As with any map of the areal units, choropleth maps *do* show reasonably precise *boundaries* between the regions (i.e., a series of exact spatial coordinates that when connected in the proper order will trace out each region), and thus we also know which regions are adjacent to (touch) which other regions. Thus the “sites”  $s \in D$  in this case are actually the regions (or *blocks*) themselves, which in this text we will denote not by  $s_i$  but by  $B_i$ ,  $i = 1, \dots, n$ , to avoid confusion between points  $s_i$  and blocks  $B_i$ . It may also be illuminating to think of the county centroids as forming the vertices of an irregular lattice, with two lattice points being connected if and only if the counties are “neighbors” in the spatial map, with physical adjacency being the most obvious (but not the only) way to define a region’s neighbors.

Some spatial data sets feature *both* point- and areal-level data, and require their simultaneous display and analysis. Figure 1.3 offers an example of this case. The first component of this data set is a collection of eight-hour maximum ozone levels at 10 monitoring sites in the greater Atlanta, GA, area for a particular day in July 1995. Like the observations in Figure 1.1, these were made at fixed monitoring stations for which exact spatial coordinates (say, latitude and longitude) are known. (That is, we assume the  $Y(s_i)$ ,  $i = 1, \dots, 10$  are random, but the  $s_i$  are not.) The second component of this data set is the number of children in the area’s zip codes (shown using the irregular subboundaries on the map) that reported at local emergency rooms (ERs) with acute asthma symptoms on the following day; confidentiality of health records precludes us from learning the precise address of any of the children. These are areal summaries that could be indicated by shading the zip codes, as in Figure 1.2. An obvious question here is whether we can establish a connection between high ozone and subsequent high pediatric ER asthma visits. Since the data are misaligned (point-level ozone but block-level ER counts), a formal statistical investigation of this question requires a preliminary *realignment* of the data; this is the subject of Chapter 7.

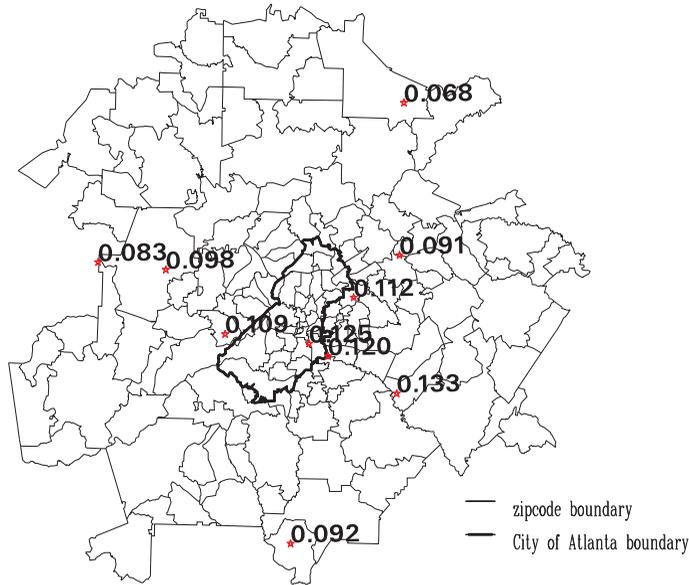


Figure 1.3 Zip code boundaries in the Atlanta metropolitan area and 8-hour maximum ozone levels (ppm) at 10 monitoring sites for July 15, 1995.

The third case above (spatial point pattern data) could be exemplified by residences of persons suffering from a particular disease, or by locations of a certain species of tree in a forest. Here the response  $Y$  is often fixed (occurrence of the event), and only the locations  $s_i$  are thought of as random. In some cases this information might be supplemented by age or other covariate information, producing a *marked* point pattern). Such data are often of interest in studies of event *clustering*, where the goal is to determine whether an observed spatial point pattern is an example of a clustered process (where points tend to be spatially close to other points), or merely the result of a random event process operating independently and homogeneously over space. Note that in contrast to areal data, where no individual points in the data set could be identified, here (and in point-referenced data as well) precise locations are known, and so must often be protected to protect the privacy of the persons in the set.

In the remainder of this initial section, we give a brief outline of the basic models most often used for each of these three data types. Here we only intend to give a flavor of the models and techniques to be fully described in the remainder of this book.

Even though our preferred inferential outlook is Bayesian, the statistical inference tools discussed in Chapters 2 through 4 are entirely classical. While all subsequent chapters adopt the Bayesian point of view, our objective here is to acquaint the reader with the classical techniques first, since they are more often implemented in standard software packages. Moreover, as in other fields of data analysis, classical methods can be easier to compute, and produce perfectly acceptable results in relatively simple settings. Classical methods often have interpretations as limiting cases of Bayesian methods under increasingly vague prior assumptions. Finally, classical methods can provide insight for formulating and fitting hierarchical models.

### 1.1.1 Point-level models

In the case of point-level data, the location index  $s$  varies *continuously* over  $D$ , a fixed subset of  $\mathbb{R}^d$ . Suppose we assume that the covariance between the random variables at two

locations depends on the *distance* between the locations. One frequently used association specification is the exponential model. Here the covariance between measurements at two locations is an exponential function of the interlocation distance, i.e.,  $Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_{i'})) \equiv C(d_{ii'}) = \sigma^2 e^{-\phi d_{ii'}}$  for  $i \neq i'$ , where  $d_{ii'}$  is the distance between sites  $s_i$  and  $s_{i'}$ , and  $\sigma^2$  and  $\phi$  are positive parameters called the *partial sill* and the *decay parameter*, respectively ( $1/\phi$  is called the *range parameter*). A plot of the covariance versus distance is called the *covariogram*. When  $i = i'$ ,  $d_{ii'}$  is of course 0, and  $C(d_{ii'}) = Var(Y(\mathbf{s}_i))$  is often expanded to  $\tau^2 + \sigma^2$ , where  $\tau^2 > 0$  is called a *nugget effect*, and  $\tau^2 + \sigma^2$  is called the *sill*. Of course, while the exponential model is convenient and has some desirable properties, many other parametric models are commonly used; see Section 2.1 for further discussion of these and their relative merits.

Adding a joint distributional model to these variance and covariance assumptions then enables likelihood inference in the usual way. The most convenient approach would be to assume a multivariate *normal* (or *Gaussian*) distribution for the data. That is, suppose we are given observations  $\mathbf{Y} \equiv \{Y(\mathbf{s}_i)\}$  at known locations  $\mathbf{s}_i$ ,  $i = 1, \dots, n$ . We then assume that

$$\mathbf{Y} \mid \mu, \boldsymbol{\theta} \sim N_n(\mu \mathbf{1}, \Sigma(\boldsymbol{\theta})), \quad (1.1)$$

where  $N_n$  denotes the  $n$ -dimensional normal distribution,  $\mu$  is the (constant) mean level,  $\mathbf{1}$  is a vector of ones, and  $(\Sigma(\boldsymbol{\theta}))_{ii'}$  gives the covariance between  $Y(\mathbf{s}_i)$  and  $Y(\mathbf{s}_{i'})$ . For the variance-covariance specification of the previous paragraph, we have  $\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi)^T$ , since the covariance matrix depends on the nugget, sill, and range.

In fact, the simplest choices for  $\Sigma$  are those corresponding to *isotropic* covariance functions, where we assume that the spatial correlation is a function solely of the distance  $d_{ii'}$  between  $\mathbf{s}_i$  and  $\mathbf{s}_{i'}$ . As mentioned above, exponential forms are particularly intuitive examples. Here,

$$(\Sigma(\boldsymbol{\theta}))_{ii'} = \sigma^2 \exp(-\phi d_{ii'}) + \tau^2 I(i = i'), \quad \sigma^2 > 0, \phi > 0, \tau^2 > 0, \quad (1.2)$$

where  $I$  denotes the indicator function (i.e.,  $I(i = i') = 1$  if  $i = i'$ , and 0 otherwise). Many other choices are possible for  $Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_{i'}))$ , including for example the powered exponential,

$$(\Sigma(\boldsymbol{\theta}))_{ii'} = \sigma^2 \exp(-\phi d_{ii'}^\kappa) + \tau^2 I(i = i'), \quad \sigma^2 > 0, \phi > 0, \tau^2 > 0, \kappa \in (0, 2],$$

the spherical, the Gaussian, and the Matérn (see Subsection 2.1.3 for a full discussion). In particular, while the latter requires calculation of a modified Bessel function, Stein (1999a, p. 51) illustrates its ability to capture a broader range of local correlation behavior despite having no more parameters than the powered exponential. We shall say much more about point-level spatial methods and models in Chapters 2, 3 and 6 and also provide illustrations using freely available statistical software.

### 1.1.2 Areal models

In models for areal data, the geographic regions or *blocks* (zip codes, counties, etc.) are denoted by  $B_i$ , and the data are typically sums or averages of variables over these blocks. To introduce spatial association, we define a *neighborhood* structure based on the arrangement of the blocks in the map. Once the neighborhood structure is defined, models resembling autoregressive time series models are considered. Two very popular models that incorporate such neighborhood information are the *simultaneously* and *conditionally autoregressive* models (abbreviated SAR and CAR), originally developed by Whittle (1954) and Besag (1974), respectively. The SAR model is computationally convenient for use with likelihood methods. By contrast, the CAR model is computationally convenient for Gibbs sampling used

in conjunction with Bayesian model fitting, and in this regard is often used to incorporate spatial correlation through a vector of spatially varying random effects  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^T$ . For example, writing  $Y_i \equiv Y(B_i)$ , we might assume  $Y_i \stackrel{ind}{\sim} N(\phi_i, \sigma^2)$ , and then impose the CAR model

$$\phi_i | \boldsymbol{\phi}_{(-i)} \sim N \left( \mu + \sum_{j=1}^n a_{ij}(\phi_j - \mu), \tau_i^2 \right), \quad (1.3)$$

where  $\boldsymbol{\phi}_{(-i)} = \{\phi_j : j \neq i\}$ ,  $\tau_i^2$  is the conditional variance, and the  $a_{ij}$  are known or unknown constants such that  $a_{ii} = 0$  for  $i = 1, \dots, n$ . Letting  $A = (a_{ij})$  and  $M = \text{Diag}(\tau_1^2, \dots, \tau_n^2)$ , by Brook's Lemma (c.f. Section 4.2), we can show that

$$p(\boldsymbol{\phi}) \propto \exp\{-(\boldsymbol{\phi} - \mu \mathbf{1})^T M^{-1} (I - A)(\boldsymbol{\phi} - \mu \mathbf{1})/2\}, \quad (1.4)$$

where  $\mathbf{1}$  is an  $n$ -vector of 1's, and  $I$  is a  $n \times n$  identity matrix.

A common way to construct  $A$  and  $M$  is to let  $A = \rho \text{Diag}(1/w_{i+})W$  and  $M^{-1} = \tau^{-2} \text{Diag}(w_{i+})$ . Here  $\rho$  is referred to as the *spatial correlation* parameter, and  $W = (w_{ij})$  is a neighborhood matrix for the areal units, which can be defined as

$$w_{ij} = \begin{cases} 1 & \text{if subregions } i \text{ and } j \text{ share a common boundary, } i \neq j \\ 0 & \text{otherwise} \end{cases}. \quad (1.5)$$

Thus  $\text{Diag}(w_{i+})$  is a diagonal matrix with  $(i, i)$  entry equal to  $w_{i+} = \sum_j w_{ij}$ . Letting  $\boldsymbol{\alpha} \equiv (\rho, \tau^2)$ , the covariance matrix of  $\boldsymbol{\phi}$  then becomes  $C(\boldsymbol{\alpha}) = \tau^2 [\text{Diag}(w_{i+}) - \rho W]^{-1}$ , where the inverse exists for an appropriate range of  $\rho$  values; see Subsection 4.3.1.

In the context of Bayesian hierarchical areal modeling, when choosing a prior distribution  $\pi(\boldsymbol{\phi})$  for a vector of spatial random effects  $\boldsymbol{\phi}$ , the CAR distribution (1.3) is often used with the 0–1 *weight* (or *adjacency*) *matrix*  $W$  in (1.5) and  $\rho = 1$ . While this results in an *improper* (nonintegrable) prior distribution, this problem is remedied by imposing a sum-to-zero constraint on the  $\phi_i$  (which turns out to be easy to implement numerically using Gibbs sampling). In this case the more general conditional form (1.3) is replaced by

$$\phi_i | \boldsymbol{\phi}_{(-i)} \sim N(\bar{\phi}_i, \tau^2/m_i), \quad (1.6)$$

where  $\bar{\phi}_i$  is the average of the  $\phi_{j \neq i}$  that are adjacent to  $\phi_i$ , and  $m_i$  is the number of these adjacencies (see, e.g., Besag, York, and Mollié, 1991). We discuss areal models in greater detail in Chapters 4 and 6.

### 1.1.3 Point process models

In the point process model, the spatial domain  $D$  is itself random, so that the elements of the index set  $D$  are the locations of random events that constitute the spatial point pattern.  $Y(\mathbf{s})$  then normally equals the constant 1 for all  $\mathbf{s} \in D$  (indicating occurrence of the event), but it may also provide additional covariate information, in which case the data constitute a marked point process.

Questions of interest with data of this sort typically center on whether the data are *clustered* more or less than would be expected if the locations were determined completely by chance. Stochastically, such uniformity is often described through a *homogeneous Poisson process*, which implies that the expected number of occurrences in region  $A$  is  $\lambda|A|$ , where  $\lambda$  is the *intensity* parameter of the process and  $|A|$  is the area of  $A$ . To investigate this in practice, plots of the data are typically a good place to start, but the tendency of the human eye to see clustering or other structure in virtually every point pattern renders a strictly graphical approach unreliable. Instead, statistics that measure clustering, and

perhaps even associated significance tests, are often used. The most common of these is *Ripley's K function*, given by

$$K(d) = \frac{1}{\lambda} E[\text{number of points within } d \text{ of an arbitrary point}], \quad (1.7)$$

where again  $\lambda$  is the intensity of the process, i.e., the mean number of points per unit area.

The theoretical value of  $K$  is known for certain spatial point process models. For instance, for point processes that have no spatial dependence at all, we would have  $K(d) = \pi d^2$ , since in this case the number of points within  $d$  of an arbitrary point should be proportional to the area of a circle of radius  $d$ ; the  $K$  function then divides out the average intensity  $\lambda$ . However, if the data are clustered we might expect  $K(d) > \pi d^2$ , while if the points follow some regularly spaced pattern we would expect  $K(d) < \pi d^2$ . This suggests a potential inferential use for  $K$ ; namely, comparing an estimate of it from a data set to some theoretical quantities, which in turn suggests whether clustering is present, and if so, which model might be most plausible. The usual estimator for  $K$  is given by

$$\hat{K}(d) = n^{-2} |A| \sum_{i \neq j} p_{ij}^{-1} I_d(d_{ij}), \quad (1.8)$$

where  $n$  is the number of points in  $A$ ,  $d_{ij}$  is the distance between points  $i$  and  $j$ ,  $p_{ij}$  is the proportion of the circle with center  $i$  and passing through  $j$  that lies within  $A$ , and  $I_d(d_{ij})$  equals 1 if  $d_{ij} < d$ , and 0 otherwise.

We provide an extensive account for point processes in Chapter 8. Other useful texts focusing primarily upon point processes and patterns include Diggle (2003), Lawson and Denison (2002), and Møller and Waagepetersen (2004) for treatments of spatial point processes and related methods in spatial cluster detection and modeling.

#### 1.1.4 Software and datasets

This text extensively uses the R ([www.r-project.org](http://www.r-project.org)) software programming language and environment for statistical computing and graphics. R is released under the GNU open-source license and can be downloaded for free from the Comprehensive R Archive Network (CRAN), which can be accessed from <http://cran.us.r-project.org/>. The capabilities of R are easily extended through “libraries” or “packages” that perform more specialized tasks. These packages are also available from CRAN and can be downloaded and installed from within the R software environment.

There are a variety of spatial packages in R that perform modeling and analysis for the different types of spatial data. For example, the `gstat` and `geoR` packages provide functions to perform traditional (classical) analysis for point-level data; the latter also offers simpler Bayesian models. The packages `spBayes` and `sptimer` have much more elaborate Bayesian functions, the latter focusing primarily upon space-time data. We will provide illustrations using some of these R packages in Chapters 2 and 6.

The `spdep` package in R provides several functions for analyzing areal-level data, including basic descriptive statistics for areal data as well as fitting areal models using classical likelihood methods. For Bayesian analysis, the BUGS language and the WinBUGS software is still perhaps the most widely used engine to fit areal models. We will discuss areal models in greater detail in Chapters 4 and 6.

Turning to point-process models, a popular spatial R package, `spatstat`, allows computation of  $K$  for any data set, as well as the approximate 95% intervals for it so the significance of departure from some theoretical model may be judged. However, full inference likely requires use of the R package `Splancs`, or perhaps a fully Bayesian approach with user-specific coding (also see Wakefield and Morris, 2001). We provide some examples of R packages for point-process models in Chapter 8.

We will use a number of spatial and spatiotemporal datasets for illustrating the modeling and software implementation. While some of these datasets are included in the R packages we will be using, others are available from [www.biostat.umn.edu/~brad/data2.html](http://www.biostat.umn.edu/~brad/data2.html). We remark that the number of R packages performing spatial analysis is already too large to be discussed in this text. We refer the reader to the CRAN Task View <http://cran.r-project.org/web/views/Spatial.html> for an exhaustive list of such packages and brief descriptions regarding their capabilities.

## 1.2 Fundamentals of cartography

In this section we provide a brief introduction to how geographers and spatial statisticians understand the geometry of (and determine distances on) the surface of the earth. This requires a bit of thought regarding cartography (mapmaking), especially map projections, and the meaning of latitude and longitude, which are often understood informally (but incorrectly) as being equivalent to Cartesian  $x$  and  $y$  coordinates.

### 1.2.1 Map projections

A map projection is a systematic representation of all or part of the surface of the earth on a plane. This typically comprises lines delineating meridians (longitudes) and parallels (latitudes), as required by some definitions of the projection. A well-known fact from topology is that it is impossible to prepare a distortion-free flat map of a surface curving in all directions. Thus, the cartographer must choose the characteristic (or characteristics) that are to be shown accurately in the map. In fact, it cannot be said that there is a “best” projection for mapping. The purpose of the projection and the application at hand lead to projections that are appropriate. Even for a single application, there may be several appropriate projections, and choosing the “best” projection can be subjective. Indeed there are an infinite number of projections that can be devised, and several hundred have been published.

Since the sphere cannot be flattened onto a plane without distortion, the general strategy for map projections is to use an intermediate surface that can be flattened. This intermediate surface is called a *developable surface* and the sphere is first projected onto the this surface, which is then laid out as a plane. The three most commonly used surfaces are the cylinder, the cone and the plane itself. Using different orientations of these surfaces leads to different classes of map projections. Some examples are given in Figure 1.4. The points on the globe are projected onto the wrapping (or tangential) surface, which is then laid out to form the map. These projections may be performed in several ways, giving rise to different projections.

Before the availability of computers, the above orientations were used by cartographers in the physical construction of maps. With computational advances and digitizing of cartography, analytical formulae for projections were desired. Here we briefly outline the underlying theory for equal-area and conformal (locally shape-preserving) maps. A much more detailed and rigorous treatment may be found in Pearson (1990).

The basic idea behind deriving equations for map projections is to consider a sphere with the geographical coordinate system  $(\lambda, \phi)$  for longitude and latitude and to construct an appropriate (rectangular or polar) coordinate system  $(x, y)$  so that

$$x = f(\lambda, \phi), \quad y = g(\lambda, \phi),$$

where  $f$  and  $g$  are appropriate functions to be determined, based upon the properties we want our map to possess. We will study map projections using differential geometry concepts, looking at infinitesimal patches on the sphere (so that curvature may be neglected

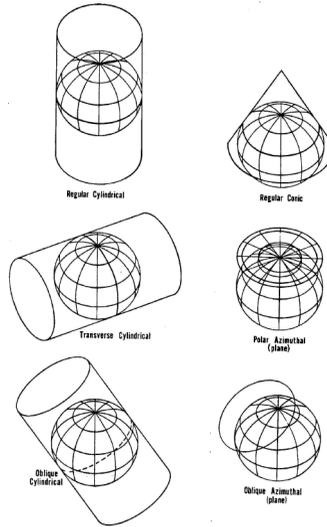


Figure 1.4 *The geometric constructions of projections using developable surfaces (figure courtesy of the U.S. Geological Survey).*

and the patches are closely approximated by planes) and deriving a set of (partial) differential equations whose solution will yield  $f$  and  $g$ . Suitable initial conditions are set to create projections with desired geometric properties.

Thus, consider a small patch on the sphere formed by the infinitesimal quadrilateral,  $ABCD$ , given by the vertices,

$$A = (\lambda, \phi), \quad B = (\lambda, \phi + d\phi), \quad C = (\lambda + d\lambda, \phi), \quad D = (\lambda + d\lambda, \phi + d\phi).$$

So, with  $R$  being the radius of the earth, the horizontal differential component along an arc of latitude is given by  $|AC| = (R \cos \phi)d\lambda$  and the vertical component along a great circle of longitude is given by  $|AB| = Rd\phi$ . Note that since  $AC$  and  $AB$  are arcs along the latitude and longitude of the globe, they intersect each other at right angles. Therefore, the area of the patch  $ABCD$  is given by  $|AC||AB|$ . Let  $A'B'C'D'$  be the (infinitesimal) image of the patch  $ABCD$  on the map. Then, we see that

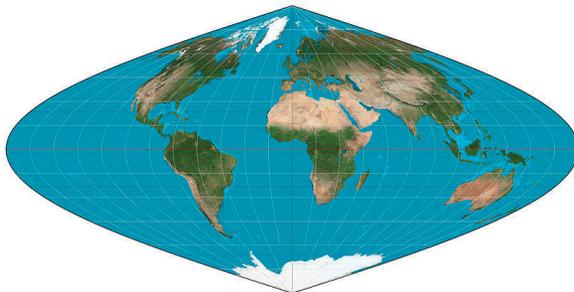
$$\begin{aligned} A' &= (f(\lambda, \phi), g(\lambda, \phi)), \\ C' &= (f(\lambda + d\lambda, \phi), g(\lambda + d\lambda, \phi)), \\ B' &= (f(\lambda, \phi + d\phi), g(\lambda, \phi + d\phi)), \\ \text{and } D' &= (f(\lambda + d\lambda, \phi + d\phi), g(\lambda + d\lambda, \phi + d\phi)). \end{aligned}$$

This in turn implies that

$$\overrightarrow{A'C'} = \left( \frac{\partial f}{\partial \lambda}, \frac{\partial g}{\partial \lambda} \right) d\lambda \quad \text{and} \quad \overrightarrow{A'B'} = \left( \frac{\partial f}{\partial \phi}, \frac{\partial g}{\partial \phi} \right) d\phi.$$

If we desire an equal-area projection we need to equate the area of the patches  $ABCD$  and  $A'B'C'D'$ . But note that the area of  $A'B'C'D'$  is given by the area of parallelogram formed by vectors  $\overrightarrow{A'C'}$  and  $\overrightarrow{A'B'}$ . Treating them as vectors in the  $xy$  plane of an  $xyz$  system, we see that the area of  $A'B'C'D'$  is the cross-product,

$$(\overrightarrow{A'C'}, 0) \times (\overrightarrow{A'B'}, 0) = \left( \frac{\partial f}{\partial \lambda} \frac{\partial g}{\partial \phi} - \frac{\partial f}{\partial \phi} \frac{\partial g}{\partial \lambda} \right) d\lambda d\phi.$$

Figure 1.5 *The sinusoidal projection.*

Therefore, we equate the above to  $|AC||AB|$ , leading to the following partial differential equation in  $f$  and  $g$ :

$$\left( \frac{\partial f}{\partial \lambda} \frac{\partial g}{\partial \phi} - \frac{\partial f}{\partial \phi} \frac{\partial g}{\partial \lambda} \right) = R^2 \cos \phi .$$

Note that this is the equation that must be satisfied by any equal-area projection. It is an underdetermined system, and further conditions need to be imposed (that ensure other specific properties of the projection) to arrive at  $f$  and  $g$ .

**Example 1.1** Equal-area maps are used for statistical displays of areal-referenced data. An easily derived equal-area projection is the sinusoidal projection, shown in Figure 1.5. This is obtained by specifying  $\partial g / \partial \phi = R$ , which yields equally spaced straight lines for the parallels, and results in the following analytical expressions for  $f$  and  $g$  (with the 0 degree meridian as the central meridian):

$$f(\lambda, \phi) = R\lambda \cos \phi; g(\lambda, \phi) = R\phi .$$

Another popular equal-area projection (with equally spaced straight lines for the meridians) is the Lambert cylindrical projection given by

$$f(\lambda, \phi) = R\lambda; g(\lambda, \phi) = R \sin \phi .$$

For conformal (angle-preserving) projections we set the angle  $\angle(AC, AB)$  equal to  $\angle(A'C', A'B')$ . Since  $\angle(AC, AB) = \pi/2$ ,  $\cos(\angle(AC, AB)) = 0$ , leading to

$$\frac{\partial f}{\partial \lambda} \frac{\partial f}{\partial \phi} + \frac{\partial g}{\partial \lambda} \frac{\partial g}{\partial \phi} = 0$$

or, equivalently, the Cauchy-Riemann equations of complex analysis,

$$\left( \frac{\partial f}{\partial \lambda} + i \frac{\partial g}{\partial \lambda} \right) \left( \frac{\partial f}{\partial \phi} - i \frac{\partial g}{\partial \phi} \right) = 0 .$$

A sufficient partial differential equation system for conformal mappings of the Cauchy-Riemann equations that is simpler to use is

$$\frac{\partial f}{\partial \lambda} = \frac{\partial g}{\partial \phi} \cos \phi; \frac{\partial g}{\partial \lambda} = \frac{\partial f}{\partial \phi} \cos \phi .$$

**Example 1.2** The Mercator projection shown in Figure 1.6 is a classical example of a conformal projection. It has the interesting property that rhumb lines (curves that intersect the meridians at a constant angle) are shown as straight lines on the map. This is particularly useful for navigation purposes. The Mercator projection is derived by letting  $\partial g / \partial \phi =$



Figure 1.6 *The Mercator projection.*

$R \sec \phi$ . After suitable integration, this leads to the analytical equations (with the 0 degree meridian as the central meridian),

$$f(\lambda, \phi) = R\lambda; g(\lambda, \phi) = R \ln \tan \left( \frac{\pi}{4} + \frac{\phi}{2} \right).$$

As is seen above, even the simplest map projections lead to complex transcendental equations relating latitude and longitude to positions of points on a given map. Therefore, rectangular grids have been developed for use by surveyors. In this way, each point may be designated merely by its distance from two perpendicular axes on a flat map. The  $y$ -axis usually coincides with a chosen central meridian,  $y$  increasing north, and the  $x$ -axis is perpendicular to the  $y$ -axis at a latitude of origin on the central meridian, with  $x$  increasing east. Frequently, the  $x$  and  $y$  coordinates are called “eastings” and “northings,” respectively, and to avoid negative coordinates, may have “false eastings” and “false northings” added to them. The grid lines usually do not coincide with any meridians and parallels except for the central meridian and the equator.

One such popular grid, adopted by The National Imagery and Mapping Agency (NIMA) (formerly known as the Defense Mapping Agency), and used especially for military use throughout the world, is the Universal Transverse Mercator (UTM) grid; see Figure 1.7. The UTM divides the world into 60 north-south zones, each of width six degrees longitude. Starting with Zone 1 (between 180 degrees and 174 degrees west longitude), these are numbered consecutively as they progress eastward to Zone 60, between 174 degrees and 180 degrees east longitude. Within each zone, coordinates are measured north and east in meters, with northing values being measured continuously from zero at the equator, in a northerly direction. Negative numbers for locations south of the equator are avoided by assigning an arbitrary false northing value of 10,000,000 meters (as done by NIMA’s cartographers). A central meridian cutting through the center of each 6 degree zone is assigned an easting value of 500,000 meters, so that values to the west of the central meridian are less than 500,000 while those to the east are greater than 500,000. In particular, the conterminous 48 states of the United States are covered by 10 zones, from Zone 10 on the west coast through Zone 19 in New England.

In practice, the UTM is used by overlaying a transparent grid on the map, allowing distances to be measured in meters at the map scale between any map point and the nearest grid lines to the south and west. The northing of the point is calculated as the sum

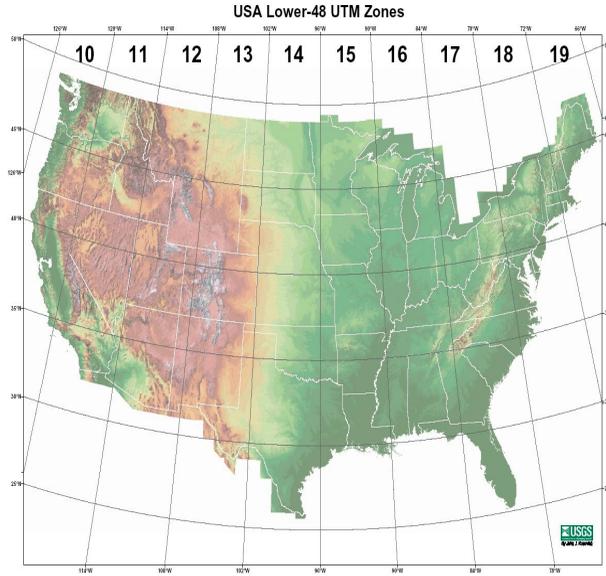


Figure 1.7 *Example of a UTM grid over the United States (figure courtesy of the U.S. Geological Survey).*

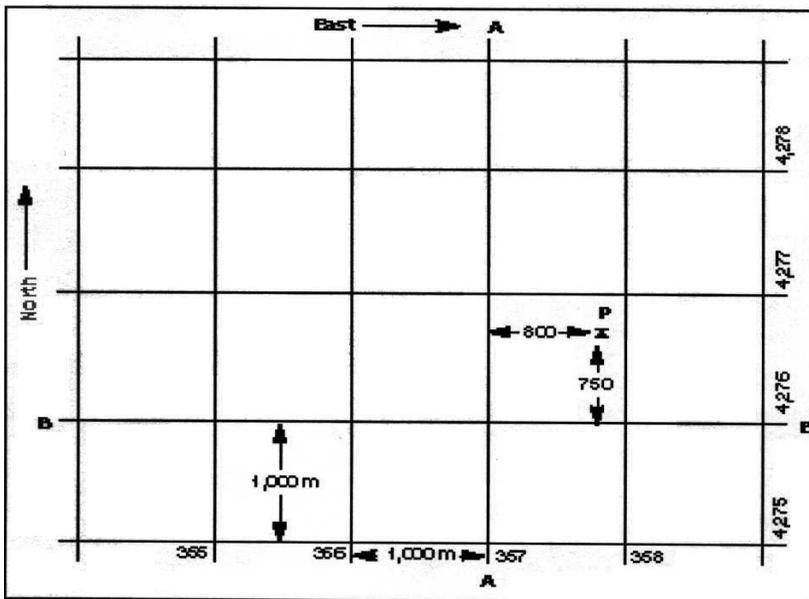


Figure 1.8 *Finding the easting and northing of a point in a UTM projection (figure courtesy of the U.S. Geological Survey).*

of the value of the nearest grid line south of it and its distance north of that line. Similarly, its easting is the value of the nearest grid line west of it added to its distance east of that line. For instance, in Figure 1.8, the grid value of line A-A is 357,000 meters east, while that of line B-B is 4,276,000 meters north. Point P is 800 meters east and 750 meters north of the grid lines resulting in the grid coordinates of point P as north 4,276,750 and east 357,800.

Finally, since spatial modeling of point-level data often requires computing distances between points on the earth's surface, one might wonder about a *planar* map projection, which would preserve distances between points. Unfortunately, the existence of such a map is precluded by Gauss' Theorema Eggregium in differential geometry (see, e.g., Guggenheimer, 1977, pp. 240–242). Thus, while we have seen projections that preserve area and shapes, distances are always distorted. The *gnomonic* projection (Snyder, 1987, pp. 164–168) gives the correct distance from a single reference point, but is less useful for the practicing spatial analyst who needs to obtain complete intersite distance matrices (since this would require not one but many such maps). Banerjee (2005) explores different strategies for computing distances on the earth and their impact on statistical inference. We present a brief summary below.

### 1.2.2 Calculating distances on the earth's surface

Distance computations are indispensable in spatial analysis. Precise inter-site distance computations are used in variogram analysis to assess the strength of spatial association. They help in setting starting values for the non-linear least squares algorithms in classical analysis (more in Chapter 2) and in specifying priors on the range parameter in Bayesian modeling (more in Chapter 5), making them crucial for correct interpretation of spatial range and the convergence of statistical algorithms. For data sets covering relatively small spatial domains, ordinary Euclidean distance offers an adequate approximation. However, for larger domains (say, the entire continental U.S.), the curvature of the earth causes distortions because of the difference in differentials in longitude and latitude (a unit increment in degree longitude is not the same length as a unit increment in degree latitude except at the equator).

Suppose we have two points on the surface of the earth,  $P_1 = (\theta_1, \lambda_1)$  and  $P_2 = (\theta_2, \lambda_2)$ . We assume both points are represented in terms of latitude and longitude. That is, let  $\theta_1$  and  $\lambda_1$  be the latitude and longitude, respectively, of the point  $P_1$ , while  $\theta_2$  and  $\lambda_2$  are those for the point  $P_2$ . The main problem is to find the shortest distance (*geodesic*) between the points. The solution is obtained via the following formulae:

$$D = R\phi$$

where  $R$  is the radius of the earth and  $\phi$  is an angle (measured in *radians*) satisfying

$$\cos \phi = \sin \theta_1 \sin \theta_2 + \cos \theta_1 \cos \theta_2 \cos (\lambda_1 - \lambda_2) . \quad (1.9)$$

These formulae are derived as follows. The geodesic is actually the arc of the great circle joining the two points. Thus, the distance will be the length of the arc of a *great circle* (i.e., a circle with radius equal to the radius of the earth). Recall that the length of the arc of a circle equals the angle subtended by the arc at the center multiplied by the radius of the circle. Therefore it suffices to find the angle subtended by the arc; denote this angle by  $\phi$ .

Let us form a three-dimensional Cartesian coordinate system  $(x, y, z)$ , with the origin at the center of the earth, the  $z$ -axis along the North and South Poles, and the  $x$ -axis on the plane of the equator joining the center of the earth and the Greenwich meridian. Using the left panel of Figure 1.9 as a guide, elementary trigonometry provides the following relationships between  $(x, y, z)$  and the latitude-longitude  $(\theta, \lambda)$ :

$$\begin{aligned} x &= R \cos \theta \cos \lambda, \\ y &= R \cos \theta \sin \lambda, \\ \text{and } z &= R \sin \theta . \end{aligned}$$

Now form the vectors  $\mathbf{u}_1 = (x_1, y_1, z_1)$  and  $\mathbf{u}_2 = (x_2, y_2, z_2)$  as the Cartesian coordinates corresponding to points  $P_1$  and  $P_2$ . Hence  $\phi$  is the angle between  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . From

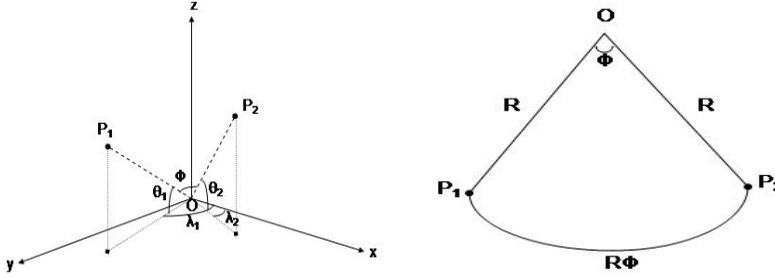


Figure 1.9 *Diagrams illustrating the geometry underlying the calculation of great circle (geodesic) distance.*

standard analytic geometry, the easiest way to find this angle is therefore to use the following relationship between the cosine of this angle and the dot product of  $\mathbf{u}_1$  and  $\mathbf{u}_2$ :

$$\cos \phi = \frac{\langle \mathbf{u}_1, \mathbf{u}_2 \rangle}{\|\mathbf{u}_1\| \|\mathbf{u}_2\|}.$$

We then compute  $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle$  as

$$\begin{aligned} R^2 [\cos \theta_1 \cos \lambda_1 \cos \theta_2 \cos \lambda_2 + \cos \theta_1 \sin \lambda_1 \cos \theta_2 \sin \lambda_2 + \sin \theta_1 \sin \theta_2] \\ = R^2 [\cos \theta_1 \cos \theta_2 \cos (\lambda_1 - \lambda_2) + \sin \theta_1 \sin \theta_2]. \end{aligned}$$

But  $\|\mathbf{u}_1\| = \|\mathbf{u}_2\| = R$ , so the result in (1.9) follows. Looking at the right panel of Figure 1.9, our final answer is thus

$$D = R\phi = R \arccos[\sin \theta_1 \sin \theta_2 + \cos \theta_1 \cos \theta_2 \cos (\lambda_1 - \lambda_2)]. \quad (1.10)$$

While calculating (1.10) is straightforward, Euclidean metrics are popular due to their simplicity and easier interpretability. More crucially, statistical modeling of spatial correlations proceed from *correlation functions* that are often valid only with Euclidean metrics. For example, using (1.10) to calculate the distances in general covariance functions may not result in a positive definite  $\Sigma(\boldsymbol{\theta})$  in (1.1). We consider a few different approaches for computing distances on the earth using Euclidean metrics, classifying them as those arising from the classical spherical coordinates, and those arising from planar projections.

Equation (1.10) clearly reveals that the relationship between the Euclidean distances and the geodetic distances is not just a matter of scaling. We cannot multiply one by a constant number to obtain the other. A simple scaling of the geographical coordinates results in a “naive Euclidean” metric obtained directly in degree units, and converted to kilometer units as:  $\|P_1 - P_2\| \pi R / 180$ . This metric performs well on small domains but always overestimates the geodetic distance, *flattening out* the meridians and parallels, and stretching the curved domain onto a plane, thereby stretching distances as well. As the domain increases, the estimation deteriorates.

Banerjee (2005) also explores a more natural metric, which is along the “chord” joining the two points. This is simply the Euclidean metric  $\|\mathbf{u}_2 - \mathbf{u}_1\|$ , yielding a “burrowed through the earth” distance — the chordal length between  $P_1$  and  $P_2$ . The slight underestimation of the geodetic distance is expected, since the chord “penetrates” the domain, producing a straight line approximation to the geodetic arc.

The first three rows of Table 1.1 compare the geodetic distance with the “naive Euclidean” and chordal metrics. The next three rows show distances computed by using three planar projections: the Mercator, the sinusoidal and a centroid-based data projection, which is developed in Exercise 10. The first column corresponds to the distance between the farthest

Methods	Colorado	Chicago-Minneapolis	New York-New Orleans
geodetic	741.7	562.0	1897.2
naive Euclidean	933.8	706.0	2172.4
chord	741.3	561.8	1890.2
Mercator	951.8	773.7	2336.5
sinusoidal	742.7	562.1	1897.7
centroid-based	738.7	562.2	1901.5

Table 1.1 *Comparison of different methods of computing distances (in kms). For Colorado, the distance reported is the maximum inter-site distance for a set of 50 locations.*

points in a spatially referenced data set comprising 50 locations in Colorado (we will revisit this dataset later in Chapter 11), while the next two present results for two differently spaced pairs of cities. The overestimation and underestimation of the “naive Euclidean” and “chordal” metrics respectively is clear, although the chordal metric excels even for distances over 2000 kms (New York and New Orleans). We find that the sinusoidal and centroid-based projections seem to be distorting distances much less than the Mercator, which performs even worse than the naive Euclidean.

This approximation of the chordal metric has an important theoretical implication for the spatial modeler. A troublesome aspect of geodetic distances is that they are *not* necessarily valid arguments for correlation functions defined on Euclidean spaces (see Chapter 2 for more general forms of correlation functions). However, the excellent approximation of the chordal metric (which is Euclidean) ensures that in most practical settings valid correlation functions in  $\mathbb{R}^3$  such as the Matérn and exponential yield positive definite correlation matrices with geodetic distances and enable proper convergence of the statistical estimation algorithms.

Schoenberg (1942) develops a necessary and sufficient representation for valid positive-definite functions on spheres in terms of normalized Legendre polynomials  $P_k$  of the form:

$$\psi(t) = \sum_{k=0}^{\infty} a_k P_k(\cos t),$$

where  $a_k$ 's are positive constants such that  $\sum_{k=0}^{\infty} a_k$  converges. An example is given by

$$\psi(t) = \frac{1}{\sqrt{1 + \alpha^2 - 2\alpha \cos t}}, \quad \alpha \in (0, 1),$$

which can be easily shown to have the Legendre polynomial expansion  $\sum_{k=0}^{\infty} \alpha^k P_k(\cos t)$ . The chordal metric also provides a simpler way to construct valid correlation functions over the sphere using a sinusoidal composition of any valid correlation function on Euclidean space. To see this, consider a unit sphere ( $R = 1$ ) and note that

$$\|\mathbf{u}_1 - \mathbf{u}_2\| = \sqrt{2 - 2\langle \mathbf{u}_1, \mathbf{u}_2 \rangle} = 2 \sin(\phi/2).$$

Therefore, a correlation function  $\rho(d)$  (suppressing the range and smoothness parameters) on the Euclidean space transforms to  $\rho(2 \sin(\phi/2))$  on the sphere, thereby *inducing* a valid correlation function on the sphere. This has some advantages over the Legendre polynomial approach of Schoenberg: (1) we retain the interpretation of the smoothness and decay parameters, (2) it is simpler to construct and compute, and (3) it builds upon a rich legacy of investigations (both theoretical and practical) of correlation functions on Euclidean spaces (again, see Chapter 2 for different correlation functions).

### 1.3 Maps and geodesics in R

The R statistical software environment today offers excellent interfaces with Geographical Information Systems (GIS) through a number of libraries (also known as packages). At the core of R's GIS capabilities is the `maps` library originally described by Becker and Wilks (1993). This `maps` library contains the geographic boundary files for several maps, including county boundaries for every state in the U.S. For example, creating a map of the state of Minnesota with its county boundaries is as simple as the following line of code:

```
> library(maps)
> mn.map <- map(database="county", region="minnesota")
```

If we do not want the county boundaries, we simply write

```
> mn.map <- map("state", "minnesota"),
```

which produces a map of Minnesota with only the state boundary. The above code uses the boundaries from R's own `maps` database. However, other important regional boundary types (say, zip codes) and features (rivers, major roads, and railroads) are generally not available, although topographic features and an enhanced GIS interface is available through the library `RgoogleMaps`. While in some respects R is perhaps not nearly as versatile as `ArcView` or other purely GIS packages, it does offer a rare combination of GIS and statistical analysis capabilities.

It is possible to import shapefiles from other GIS software (e.g. `ArcView`) into R using the `maptools` package. We invoke the `readShapePoly` function in the `maptools` package to read the external shapefile and store the output in `minnesota.shp`. To produce the map, we apply `plot` to this output.

```
> library(maptools)
> minnesota.shp <- readShapePoly("minnesota.shp",
+                               proj4string=CRS("+proj=longlat"))
> plot(minnesota.shp).
```

For the above to work, you will need three files with extensions `.shp`, `.shx` and `.dbf`. They must have the same name and differ only in the extension. The `minnesota.shp` file contains the geometry data, the `minnesota.shx` file contains the spatial index, and the `minnesota.dbf` file contains the attribute data. These are read using the `readShapePoly()` function to produce a spatial polygon object.

The above is an example of how to draw bare maps of a state within the USA using either R's own database or an external shapefile. We can also draw maps of other countries using the `mapdata` package, which has some world map data, in conjunction with `maps`. For example, to draw a map of Canada, we write

```
> library(maps)
> library(mapdata)
> map("worldHires", "Canada",
+     xlim=c(-141, -53), ylim=c(40, 85),
+     col="gray90", fill=TRUE)
```

We leave the reader to experiment further with these examples.

In practice, we are not interested in bare maps but would want to plot spatially referenced data on the map. Let us return to the counties in Minnesota. Consider a new file `newdata.csv` that includes information on the population of each county of Minnesota along with the number of influenza A (H1N1) cases from each county. We first merge our new dataset with the `minnesota.shp` object already created using the county names.

```
> newdata <- read.csv("newdata.csv")
> minnesota.shp@data <- merge(minnesota.shp@data, newdata,
+                             by="NAME").
```

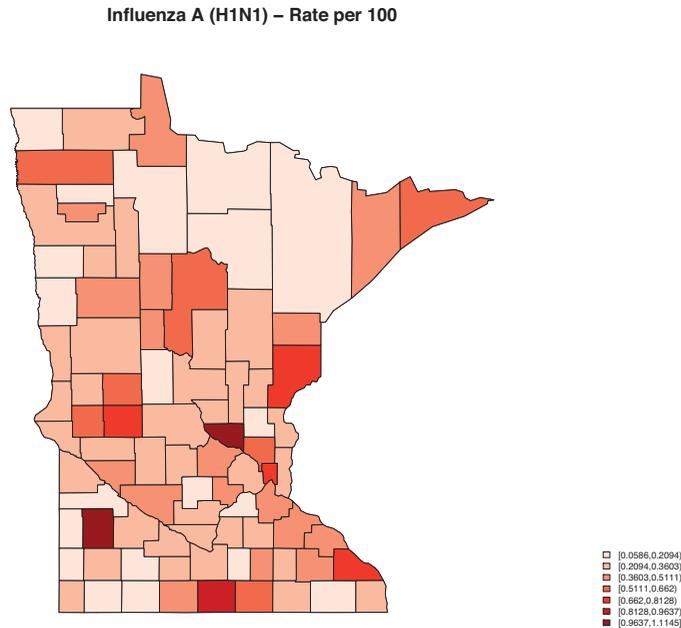


Figure 1.10 Map showing Influenza A (H1N1) rates (cases/population) $\times 100$  in different counties of Minnesota for 1999.

To plot the data in a visually appealing way, we use two additional packages: **RColorBrewer**, which creates nice color schemes, and **classInt**, which facilitates classifying the data into classes which will correspond to the plotting colors. Below, we present the code to plot H1N1 influenza rates per 100 after classifying the data into equal intervals.

```

> library(RColorBrewer)
> library(classInt)
> CASES <- minnesota.shp@data$cases
> POP1999 <- minnesota.shp@data$POP1999
> var <- (CASES/POP1999)*100
> nclr <- 7
> plotclr <- brewer.pal(nclr,"Reds")
> class <- classIntervals(var, nclr, style="equal", dataPrecision=4)
> colcode <- findColours(class, plotclr)
> plot(minnesota.shp)
> plot(minnesota.shp, col=colcode, add=T)
> title(main="Influenza A (H1N1) - Rate per 100")
> legend("bottomright", legend=names(attr(colcode, "table")),
+   fill=attr(colcode, "palette"), cex=0.6, bty="n").

```

The resulting map is shown in Figure 1.10 and is called a *choropleth map*.

Map projections in R can be performed using the **mapproj** package. For example, we can create a variety of cartographic map projections of any part of the world using **xlim** and **ylim** to specify the bounds (using longitudes and latitudes) of the plotting region and use **projection** to specify the type of projection we desire. In fact, the **maps** package uses **mapproj** for its map projections. Therefore, loading **maps** automatically invokes **mapproj**.

```

> library(maps) ## automatically loads mapproj
> sinusoidal.proj = map(database= "world", ylim=c(45,90), xlim=c(-160,-50),
+   col="grey80", fill=TRUE, plot=FALSE, projection="sinusoidal")

```

```
> map(sinusoidal.proj)
```

produces a Sinusoidal map projection (Example 1.1) between latitudes 45 and 90 degrees and longitudes  $-160$  and  $-150$  degrees. (Run the above code and test your geography to see what part of the world this is!) Repeating the above with `projection` set to "mercator" will result in a Mercator projection (Example 1.2). For higher resolution maps, one can load the `mapdata` package and use "worldHires" instead of "world" in the database.

For distance computations using a map projection, as was done for the Mercator and sinusoidal in Table 1.1, it is convenient to use the `mapproject` function in package `mapproj`. This simply converts latitude and longitude to rectangular coordinates. Suppose `LON` and `LAT` are two vectors containing longitudes and latitudes on the earth's surface to be projected. Then, a simple command such as

```
> xy.sinusoidal <- mapproject(LON, LAT, projection="sinusoidal")
```

produces the projected coordinates in Euclidean space. The coordinates are accessed by `xy.sinusoidal$x` and `xy.sinusoidal$y`, respectively. For distance computations that will approximate the great circle distance, these coordinates need to be multiplied by the radius of the earth.

While `mapproject` does not offer the UTM projections, the `rgdal` package can be used to construct UTM projections. This is particularly useful when plotting point-level data through the `RgoogleMaps` package, another exciting GIS interface offered by R. It is especially useful for plotting points such as GPS (Global Positioning System) locations on maps. For example, the Colorado data used in Table 1.1 is the file `ColoradoS-T.dat` and can be downloaded from [www.biostat.umn.edu/~brad/data2.html](http://www.biostat.umn.edu/~brad/data2.html). We can read the file in R as

```
> coloradoST <- read.table("ColoradoS-T.dat", header = TRUE).
```

Next, we use the `GetMap.bbox` function in the `RgoogleMaps` package to set the region within which our coordinates will be plotted. A "mptype" argument provides some options for the type of background map we want. The region can also be plotted using the `PlotOnStaticMap` function. This is sometimes useful to obtain an idea whether the underlying map will be useful with reference to our plotting coordinates.

```
> library(RgoogleMaps)
> MyMap <- GetMap.bbox(lonR = range(coloradoST$Longitude),
+                       latR = range(coloradoST$Latitude),
+                       size=c(640,640), mptype = "hybrid")
> PlotOnStaticMap(MyMap).
```

We now convert the longitude and latitude to the same coordinate system as in the `MyMap` object we created above.

```
> convert_points <- LatLon2XY.centered(MyMap,
+                                       coloradoST$Latitude,
+                                       coloradoST$Longitude)
> points(convert_points$newX, convert_points$newY,
+        col = 'red', pch=19).
```

Finally, we convert our points to UTM coordinates using the `rgdal` package. The code below performs the conversion from latitude-longitude to UTM coordinates and then plots these coordinates on the map created above. This will be achieved in two steps. First, we convert out longitudes and latitudes in the Colorado dataset into a special data type known as `SpatialPoints`. We store this in `SP_longlat`, which is converted by `spTransform` to UTM coordinates. The `zone` parameter specifies the zone of the UTM needs to be supplied explicitly.

```
> library(sp)
> library(rgdal)
```

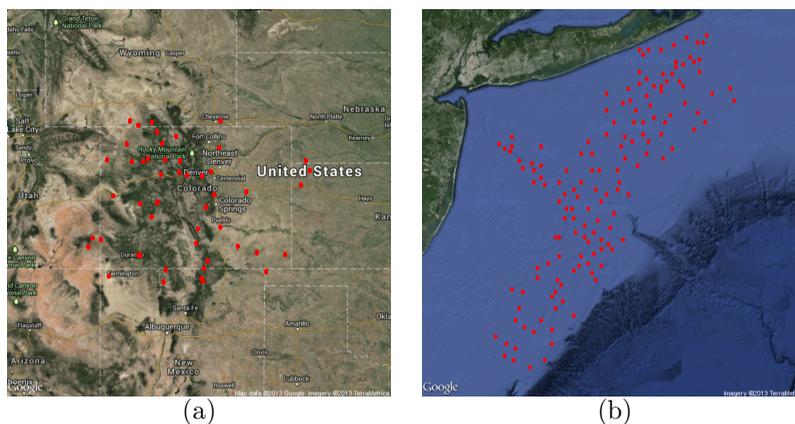


Figure 1.11 (a) A plot of the 50 locations in Colorado; (b) a plot of locations in the New York/New Jersey Bight reporting scallop catches.

```
> SP_longlat <- SpatialPoints(coords =
+   cbind(coloradoST$Longitude,
+         coloradoST$Latitude),
+   proj4string = CRS("+proj=longlat +ellps=WGS84"))
> SP_utm <- spTransform(SP_longlat,
+   CRS("+proj=utm +zone=13 +datum=WGS84"))
> plot(SP_utm).
```

We repeat the above exercise with another well known spatial data set involving the catches of scallops in the New York/New Jersey Bight. This data set also has coordinates in terms of latitude and longitude and can be downloaded from [www.biostat.umn.edu/~brad/data/scallops.txt](http://www.biostat.umn.edu/~brad/data/scallops.txt). The resulting plots are presented in Figure 1.11. The map features, including the background, can be altered by changing the parameter “`mptype`” in the function `GetMap.bbox`. The options are “`roadmap`”, “`mobile`”, “`satellite`”, “`terrain`” and “`hybrid`”.

Finally, we mention the `fields` package in R, which offers several useful functions for spatial analysis. In particular, it includes two functions `rdist` and `rdist.earth` that conveniently compute inter-site distances. Let  $X1$  and  $X2$  be two matrices representing two different sets of locations. Then,

```
> library(fields)
> euclidean.dist = rdist(X1, X2)
> spherical.dist = rdist.earth(X1,X2)
```

computes the inter-site distance matrices between the locations in  $X1$  and  $X2$ . The function `rdist` uses the Euclidean distance, while `rdist.earth` uses the spherical or geodetic distance. The latter should be used only when  $X1$  and  $X2$  contain latitude-longitude coordinates.

## 1.4 Exercises

1. What sorts of areal unit variables can you envision that could be viewed as arising from point-referenced variables? What sorts of areal unit variables can you envision whose mean could be viewed as arising from a point-referenced surface? What sorts of areal unit variables fit neither of these scenarios?

2. What sorts of sensible properties should characterize association between point-referenced measurements? What sorts of sensible properties should characterize association between areal unit measurements?
3. Suggest some regional-level covariates that might help explain the spatial pattern evident in Figure 1.2. (*Hint:* The roughly rectangular group of regions located on the map's eastern side is the city of Minneapolis, MN.)
- 4.(a) Suppose you recorded elevation and average daily temperature on a particular day for a sample of locations in a region. If you were given the elevation at a new location, how would you make a plausible estimate of the average daily temperature for that location?
- (b) Why might you expect spatial association between selling prices of single-family homes in this region to be weaker than that between the observed temperature measurements?
5. For what sorts of point-referenced spatial data would you expect measurements across time to be essentially independent? For what sorts of point-referenced data would you expect measurements across time to be strongly dependent?
6. For point-referenced data, suppose the means of the variables are spatially associated. Would you expect the association between the variables themselves to be weaker than, stronger than, or the same as the association between the means?
- 7.(a) Write your own R function that will compute the distance between two points  $P_1$  and  $P_2$  on the surface of the earth. The function should take the latitude and longitude of the  $P_i$  as input, and output the geodesic distance  $D$  given in (1.10). Use  $R = 6371$  km.
- (b) Use your program to obtain the geodesic distance between Chicago (87.63W, 41.88N) and Minneapolis (93.22W, 44.89N), and between New York (73.97W, 40.78N) and New Orleans (90.25W, 29.98N).
8. A “naive Euclidean” distance may be computed between two points by simply applying the Euclidean distance formula to the longitude-latitude coordinates, and then multiplying by  $(R\pi/180)$  to convert to kilometers. Find the naive Euclidean distance between Chicago and Minneapolis, and between New York and New Orleans, comparing your results to the geodesic ones in the previous problem.
9. The *chordal* (“burrowing through the earth”) distance separating two points is given by the Euclidean distance applied to the Cartesian spherical coordinate system given in Subsection 1.2.2. Find the chordal distance between Chicago and Minneapolis, and between New York and New Orleans, comparing your results to the geodesic and naive Euclidean ones above.
10. A two-dimensional projection, often used to approximate geodesic distances by applying Euclidean metrics, sets up rectangular axes along the centroid of the observed locations, and scales the points according to these axes. Thus, with  $N$  locations having geographical coordinates  $(\lambda_i, \theta_i)_{i=1}^N$ , we first compute the centroid  $(\bar{\lambda}, \bar{\theta})$  (the mean longitude and latitude). Next, two distances are computed. The first,  $d_X$ , is the geodesic distance (computed using (1.10) between  $(\bar{\lambda}, \theta_{\min})$  and  $(\bar{\lambda}, \theta_{\max})$ , where  $\theta_{\min}$  and  $\theta_{\max}$  are the minimum and maximum of the observed latitudes. Analogously,  $d_Y$  is the geodesic distance computed between  $(\lambda_{\min}, \bar{\theta})$  and  $(\lambda_{\max}, \bar{\theta})$ . These actually scale the axes in terms of true geodesic distances. The projection is then given by

$$x = \frac{\lambda - \bar{\lambda}}{\lambda_{\max} - \lambda_{\min}} d_X; \text{ and } y = \frac{\theta - \bar{\theta}}{\theta_{\max} - \theta_{\min}} d_Y .$$

Applying the Euclidean metric to the projected coordinates yields a good approximation to the inter-site geodesic distances. This projection is useful for entering coordinates in spatial statistics software packages that require two-dimensional coordinate input and uses Euclidean metrics to compute distances (e.g., the variogram functions in `geoR`, the `spatial.exp` function in `WinBUGS`, etc.).

- (a) Compute the above projection for Chicago and Minneapolis ( $N = 2$ ) and find the Euclidean distance between the projected coordinates. Compare with the geodesic distance. Repeat this exercise for New York and New Orleans.
  - (b) When will the above projection fail to work?
11. Use the `sp`, `rgdal` and `RgoogleMaps` packages to create an UTM projection for the locations in the scallops data and produce the picture in Figure 1.11(b).
  12. Use the `fields` package to produce the inter-site distance matrix for the locations in the scallops data. Compute this matrix using the `rdist.earth` function, which yields the geodetic distances. Next project the data to UTM coordinates and use the `rdist` function to compute the inter-site Euclidean distance matrix. Draw histograms of the inter-site distances and comment on any notable discrepancies resulting from the map projection.



# Basics of point-referenced data models

---

In this chapter we present the essential elements of spatial models and classical analysis for point-referenced data. As mentioned in Chapter 1, the fundamental concept underlying the theory is a stochastic process  $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ , where  $D$  is a fixed subset of  $r$ -dimensional Euclidean space. Note that such stochastic processes have a rich presence in the time series literature, where  $r = 1$ . In the spatial context, usually we encounter  $r$  to be 2 (say, northings and eastings) or 3 (e.g., northings, eastings, and altitude above sea level). For situations where  $r > 1$ , the process is often referred to as a *spatial process*. For example,  $Y(\mathbf{s})$  may represent the level of a pollutant at site  $\mathbf{s}$ . While it is conceptually sensible to assume the existence of a pollutant level at all possible sites in the domain, in practice the data will be a partial realization of that spatial process. That is, it will consist of measurements at a finite set of locations, say  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ , where there are monitoring stations. The problem facing the statistician is inference about the spatial process  $Y(\mathbf{s})$  and prediction at new locations, based upon this partial realization. The remarkable feature of the models we employ here is that, despite only seeing the process, equivalently, the spatial surface at a finite set of locations, we can infer about the surface at an uncountable number of locations. The reason is that we specify association through *structured* dependence which enables this broad interpolation.

This chapter is organized as follows. We begin with a survey of the building blocks of point-level data modeling, including stationarity, isotropy, and variograms (and their fitting via traditional moment-matching methods). We defer theoretical discussion of the spatial (typically Gaussian) process modeling that enables likelihood (and Bayesian) inference in these settings to Chapters 3 and 6. Then, we illustrate helpful exploratory data analysis tools, as well as traditional classical methods, especially kriging (point-level spatial prediction). We view all of these activities in an exploratory fashion, i.e., as a prelude to fully model-based inference under a hierarchical model. We close with some short tutorials in R using easy to use and widely available point-level spatial statistical analysis packages.

The material we cover in this chapter is traditionally known as *geostatistics*, and could easily fill many more pages than we devote to it here. While we prefer the more descriptive term “point-level spatial modeling,” we will at times still use “geostatistics” for brevity and perhaps consistency when referencing the literature.

## 2.1 Elements of point-referenced modeling

### 2.1.1 Stationarity

For our discussion we assume that our spatial process has a mean, say  $\mu(\mathbf{s}) = E(Y(\mathbf{s}))$ , associated with it and that the variance of  $Y(\mathbf{s})$  exists for all  $\mathbf{s} \in D$ . The process  $Y(\mathbf{s})$  is said to be *Gaussian* if, for any  $n \geq 1$  and any set of sites  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ ,  $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$  has a multivariate normal distribution. The process is said to be *strictly stationary* (sometimes *strong stationarity*) if, for any given  $n \geq 1$ , any set of  $n$  sites  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  and any  $\mathbf{h} \in \mathfrak{R}^r$ ,

the distribution of  $(Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))$  is the same as that of  $(Y(\mathbf{s}_1 + \mathbf{h}), \dots, Y(\mathbf{s}_n + \mathbf{h}))$ . Here  $D$  is envisioned as  $\mathfrak{R}^r$  as well.

A less restrictive condition is given by *weak stationarity* (also called second-order stationarity). A spatial process is called weakly stationary if  $\mu(\mathbf{s}) \equiv \mu$  (i.e., it has a constant mean) and  $\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C(\mathbf{h})$  for all  $\mathbf{h} \in \mathfrak{R}^r$  such that  $\mathbf{s}$  and  $\mathbf{s} + \mathbf{h}$  both lie within  $D$ . In fact, for stationarity as a second-order property we will need only the second property;  $E(Y(\mathbf{s}))$  need not equal  $E(Y(\mathbf{s} + \mathbf{h}))$ . But since we will apply the definition only to a mean 0 spatial residual process, this distinction is not important for us. Weak stationarity implies that the covariance relationship between the values of the process at any two locations can be summarized by a covariance function  $C(\mathbf{h})$ , and this function depends only on the separation vector  $\mathbf{h}$ . Note that with all variances assumed to exist, strong stationarity implies weak stationarity. The converse is not true in general, but it *does* hold for Gaussian processes; see Exercise 4.

We offer a simple illustration of a weakly stationary process that is not strictly stationary. It is easy to see in the one-dimensional case. Suppose the process  $Y_t, t = 1, 2, \dots$  consists of a sequence of independent variables such that for  $t$  odd,  $Y_t$  is a binary variable taking the values 1 and  $-1$  each with probability .5 while for  $t$  even,  $Y_t$  is normal with mean 0 and variance 1. We have weak stationarity since  $\text{cov}(Y(t), Y(t')) = 0, t \neq t', t = 1, t = t'$ . That is, we only need to know the value of  $t - t'$  to specify the covariance. However, clearly  $Y_1, Y_3$  does not have the same distribution as  $Y_2, Y_4$ .

### 2.1.2 Variograms

There is a third type of stationarity called *intrinsic* stationarity. Here we assume  $E[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})] = 0$  and define

$$E[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})]^2 = \text{Var}(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})) = 2\gamma(\mathbf{h}) . \quad (2.1)$$

Equation (2.1) makes sense only if the left-hand side depends *solely* on  $\mathbf{h}$  (so that the right-hand side can be written at all), and not the particular choice of  $\mathbf{s}$ . If this is the case, we say the process is *intrinsically stationary*. The function  $2\gamma(\mathbf{h})$  is then called the *variogram*, and  $\gamma(\mathbf{h})$  is called the *semivariogram*. We can offer some intuition behind the variogram but it really arose simply as a result of its appearance in traditional kriging where one seeks the best linear unbiased predictor, as we clarify below. Behaviorally, at short distances (small  $\|\mathbf{h}\|$ ), we would expect  $Y(\mathbf{s} + \mathbf{h})$  and  $Y(\mathbf{s})$  to be very similar, that is,  $(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s}))^2$  to be small. As  $\|\mathbf{h}\|$  grows larger, we expect less similarity between  $Y(\mathbf{s} + \mathbf{h})$  and  $Y(\mathbf{s})$ , i.e., we expect  $(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s}))^2$  to be larger. So, a plot of  $\gamma(\mathbf{h})$  would be expected to increase with  $\|\mathbf{h}\|$ , providing some insight into spatial behavior. (The covariance function  $C(\mathbf{h})$  is sometimes referred to as the *covariogram*, especially when plotted graphically.) Note that intrinsic stationarity defines only the first and second moments of the differences  $Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})$ . It says nothing about the joint distribution of a collection of variables  $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)$ , and thus provides no likelihood.

In fact, it says nothing about the moments of the  $Y(\mathbf{s})$ 's, much less their distribution. It only describes the behavior of differences rather than the behavior of the data that we observe, clearly unsatisfying from the perspective of data analysis. In fact, the  $Y(\mathbf{s})$ 's need not have any moments. For example, we might have  $Y(\mathbf{s}) = W(\mathbf{s}) + V$  where  $W(\mathbf{s})$  is a collection of i.i.d. normal variables and, independently,  $V$  is a Cauchy random variable. Then  $Y(\mathbf{s})$  is intrinsically stationary but the  $Y(\mathbf{s})$ 's have no moments. Even more disconcerting, the distribution of  $Y(\mathbf{s}) - Y(\mathbf{s}')$  may be proper while the distribution of  $Y(\mathbf{s})$  and of  $Y(\mathbf{s}')$  may be improper. For instance, suppose the joint distribution,  $f(Y(\mathbf{s}), Y(\mathbf{s}')) \propto e^{-(Y(\mathbf{s}) - Y(\mathbf{s}'))^2/2}$ . Then,  $Y(\mathbf{s}) - Y(\mathbf{s}') \sim N(0, 1)$  and, in fact,  $f(Y(\mathbf{s})|Y(\mathbf{s}'))$  and  $f(Y(\mathbf{s}')|Y(\mathbf{s}))$  are proper,  $N(0, 1)$  but the joint distribution is improper. How could we employ a probability specification that could not possibly have generated the data we observe?

It is easy to see the relationship between the variogram and the covariance function:

$$\begin{aligned}
 2\gamma(\mathbf{h}) &= \text{Var}(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})) \\
 &= \text{Var}(Y(\mathbf{s} + \mathbf{h})) + \text{Var}(Y(\mathbf{s})) - 2\text{Cov}(Y(\mathbf{s} + \mathbf{h}), Y(\mathbf{s})) \\
 &= C(\mathbf{0}) + C(\mathbf{0}) - 2C(\mathbf{h}) \\
 &= 2[C(\mathbf{0}) - C(\mathbf{h})] .
 \end{aligned}$$

Thus,

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) . \quad (2.2)$$

From (2.2) we see that given  $C$ , we are able to recover  $\gamma$  easily. But what about the converse; in general, can we recover  $C$  from  $\gamma$ ? Here it turns out we need to assume a bit more: if the spatial process is *ergodic*, then  $C(\mathbf{h}) \rightarrow 0$  as  $\|\mathbf{h}\| \rightarrow \infty$ , where  $\|\mathbf{h}\|$  denotes the length of the  $\mathbf{h}$  vector. This is an intuitively sensible condition, since it means that the covariance between the values at two points vanishes as the points become further separated in space. But taking the limit of both sides of (2.2) as  $\|\mathbf{h}\| \rightarrow \infty$ , we then have that  $\lim_{\|\mathbf{h}\| \rightarrow \infty} \gamma(\mathbf{h}) = C(\mathbf{0})$ . Thus, using the dummy variable  $\mathbf{u}$  to avoid confusion, we have

$$C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h}) = \lim_{\|\mathbf{u}\| \rightarrow \infty} \gamma(\mathbf{u}) - \gamma(\mathbf{h}) . \quad (2.3)$$

In general, the limit on the right-hand side need not exist, but if it does, then the process is weakly (second-order) stationary with  $C(\mathbf{h})$  as given in (2.3). We therefore have a way to determine the covariance function  $C$  from the semivariogram  $\gamma$ . Thus weak stationarity implies intrinsic stationarity, but the converse is not true; indeed, the next section offers examples of processes that are intrinsically stationary but not weakly stationary.

A valid variogram necessarily satisfies a negative definiteness condition. In fact, for any set of locations  $\mathbf{s}_1, \dots, \mathbf{s}_n$  and any set of constants  $a_1, \dots, a_n$  such that  $\sum_i a_i = 0$ , if  $\gamma(\mathbf{h})$  is valid, then

$$\sum_i \sum_j a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0 . \quad (2.4)$$

To see this, note that

$$\begin{aligned}
 \sum_i \sum_j a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) &= \frac{1}{2} E \sum_i \sum_j a_i a_j (Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2 \\
 &= -E \sum_i \sum_j a_i a_j Y(\mathbf{s}_i) Y(\mathbf{s}_j) \\
 &= -E \left[ \sum_i a_i Y(\mathbf{s}_i) \right]^2 \leq 0 .
 \end{aligned}$$

We remark that, despite the suggestion of expression (2.2), there is no relationship between this result and the positive definiteness condition for covariance functions (see Subsection 3.1.2). Cressie (1993) discusses further necessary conditions for a valid variogram. Lastly, the condition (2.4) emerges naturally in ordinary kriging (see Section 2.4).

### 2.1.3 Isotropy

Another important related concept is that of isotropy. If the semivariogram function  $\gamma(\mathbf{h})$  depends upon the separation vector only through its length  $\|\mathbf{h}\|$ , then we say that the variogram is *isotropic*; that is, if  $\gamma(\mathbf{h})$  is a real-valued function of a univariate argument, and can be written as  $\gamma(\|\mathbf{h}\|)$ . If not, we say it is *anisotropic*. Because of the foregoing issues with intrinsic stationarity, in the absence of a full probabilistic specification (as detailed in

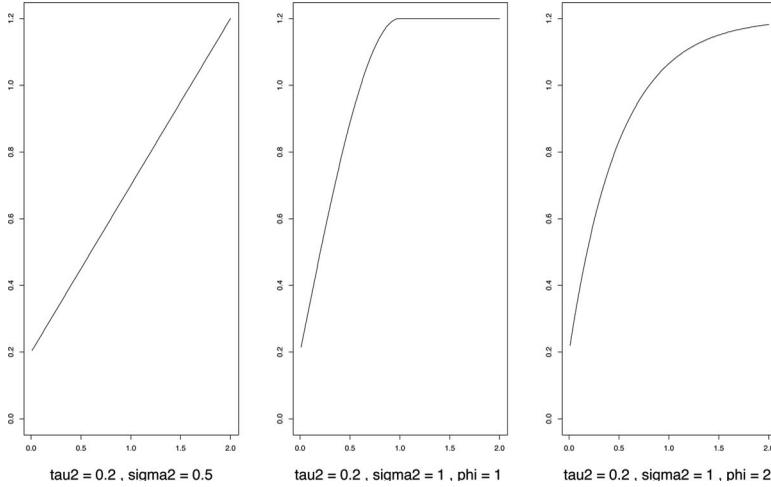


Figure 2.1 *Theoretical semivariograms for three models: (a) linear, (b) spherical, and (c) exponential.*

Chapter 3), we are reluctant to associate an isotropic variogram with a stochastic process. Nonetheless, in the literature, we find terminology stating that, if a process is intrinsically stationary and isotropic, it is also called *homogeneous*.

Isotropic variograms are popular because of their simplicity, interpretability, and, in particular, because a number of relatively simple parametric forms are available as candidates for the semivariogram. Denoting  $\|\mathbf{h}\|$  by  $d$  for notational simplicity, we now consider a few of the more important such forms.

1. *Linear:*

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d & \text{if } d > 0, \tau^2 > 0, \sigma^2 > 0 \\ 0 & \text{otherwise} \end{cases} .$$

Note that  $\gamma(d) \rightarrow \infty$  as  $d \rightarrow \infty$ , and so this semivariogram does not correspond to a weakly stationary process (although it is intrinsically stationary). This semivariogram is plotted in Figure 2.1(a) using the parameter values  $\tau^2 = 0.2$  and  $\sigma^2 = 0.5$ .

2. *Spherical:*

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 & \text{if } d \geq 1/\phi, \\ \tau^2 + \sigma^2 \left\{ \frac{3\phi d}{2} - \frac{1}{2} (\phi d)^3 \right\} & \text{if } 0 < d \leq 1/\phi, \\ 0 & \text{otherwise} \end{cases} .$$

The spherical semivariogram is valid in  $r = 1, 2$ , or 3 dimensions, but for  $r \geq 4$  it fails to correspond to a spatial variance matrix that is positive definite (as required to specify a valid joint probability distribution). This variogram owes its popularity largely to the fact that it offers clear illustrations of the *nugget*, *sill*, and *range*, three characteristics traditionally associated with variograms. Specifically, consider Figure 2.1(b), which plots the spherical semivariogram using the parameter values  $\tau^2 = 0.2$ ,  $\sigma^2 = 1$ , and  $\phi = 1$ . While  $\gamma(0) = 0$  by definition,  $\gamma(0^+) \equiv \lim_{d \rightarrow 0^+} \gamma(d) = \tau^2$ ; this quantity is the *nugget*. Next,  $\lim_{d \rightarrow \infty} \gamma(d) = \tau^2 + \sigma^2$ ; this asymptotic value of the semivariogram is called the *sill*. (The sill minus the nugget, which is simply  $\sigma^2$  in this case, is called the *partial sill*.) Finally, the value  $d = 1/\phi$  at which  $\gamma(d)$  first reaches its ultimate level (the sill) is called the *range*. It is for this reason that many of the variogram models of this subsection are often parametrized through  $R \equiv 1/\phi$ . Confusingly, both  $R$  and  $\phi$  are sometimes referred to as the *range* parameter, although  $\phi$  is often more accurately referred to as the *decay* parameter.

Note that for the linear semivariogram, the nugget is  $\tau^2$  but the sill and range are both infinite. For other variograms (such as the next one we consider), the sill is finite, but only reached asymptotically.

### 3. Exponential:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\phi d)) & \text{if } d > 0, \\ 0 & \text{otherwise} \end{cases} .$$

The exponential has an advantage over the spherical in that it is simpler in functional form while still being a valid variogram in all dimensions (and without the spherical's finite range requirement). However, note from Figure 2.1(c), which plots this semivariogram assuming  $\tau^2 = 0.2$ ,  $\sigma^2 = 1$ , and  $\phi = 2$ , that the sill is only reached asymptotically; strictly speaking, the range  $R = 1/\phi$  is infinite. In cases like this, the notion of an *effective range* is often used, i.e., the distance at which there is essentially no lingering spatial correlation. To make this notion precise, we must convert from  $\gamma$  scale to  $C$  scale (possible here since  $\lim_{d \rightarrow \infty} \gamma(d)$  exists; the exponential is not only intrinsically but also weakly stationary). From (2.3) we have

$$\begin{aligned} C(d) &= \lim_{u \rightarrow \infty} \gamma(u) - \gamma(d) \\ &= \tau^2 + \sigma^2 - [\tau^2 + \sigma^2(1 - \exp(-\phi d))] \\ &= \sigma^2 \exp(-\phi d) . \end{aligned}$$

Hence

$$C(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } d = 0 \\ \sigma^2 \exp(-\phi d) & \text{if } d > 0 \end{cases} . \quad (2.5)$$

If the nugget  $\tau^2 = 0$ , then this expression reveals that the correlation between two points  $d$  units apart is  $\exp(-\phi d)$ ; note that  $\exp(-\phi d) = 1^-$  for  $d = 0^+$  and  $\exp(-\phi d) = 0$  for  $d = \infty$ , both in concert with this interpretation.

A common definition of the *effective range*,  $d_0$ , is the distance at which this correlation is *negligible*, customarily taken as having dropped to only 0.05. Setting  $\exp(-\phi d_0)$  equal to this value we obtain  $t_0 \approx 3/\phi$ , since  $\log(0.05) \approx -3$ . The range will be discussed in more detail in Subsection 3.1.2.

Finally, the form of (2.5) gives a clear example of why the nugget ( $\tau^2$  in this case) is often viewed as a “nonspatial effect variance,” and the partial sill ( $\sigma^2$ ) is viewed as a “spatial effect variance.” That is, we have two variance components. Along with  $\phi$ , a statistician would likely view fitting this model to a spatial data set as an exercise in estimating these three parameters. We shall return to variogram model fitting in Subsection 2.1.4.

### 4. Gaussian:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\phi^2 d^2)) & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases} . \quad (2.6)$$

The Gaussian variogram is an analytic function and yields very smooth realizations of the spatial process. We shall say much more about process smoothness in Subsection 3.1.4.

### 5. Powered exponential:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-|\phi d|^p)) & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases} . \quad (2.7)$$

Here  $0 < p \leq 2$  yields a family of valid variograms. Note that both the Gaussian and the exponential forms are special cases.

### 6. Rational quadratic:

$$\gamma(d) = \begin{cases} \tau^2 + \frac{\sigma^2 d^2}{(\phi + d^2)} & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases} .$$

Model	Covariance function, $C(d)$
Linear	$C(d)$ does not exist
Spherical	$C(d) = \begin{cases} 0 & \text{if } d \geq 1/\phi \\ \frac{\sigma^2 \left[1 - \frac{3}{2}\phi d + \frac{1}{2}(\phi d)^3\right]}{\tau^2 + \sigma^2} & \text{if } 0 < d \leq 1/\phi \\ \sigma^2 & d = 0 \end{cases}$
Exponential	$C(d) = \begin{cases} \sigma^2 \exp(-\phi d) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Powered exponential	$C(d) = \begin{cases} \sigma^2 \exp(- \phi d ^p) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Gaussian	$C(d) = \begin{cases} \sigma^2 \exp(-\phi^2 d^2) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Rational quadratic	$C(d) = \begin{cases} \sigma^2 \left(1 - \frac{d^2}{(\phi + d^2)}\right) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Wave	$C(d) = \begin{cases} \frac{\sigma^2 \sin(\phi d)}{\phi d} & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Power law	$C(d)$ does not exist
Matérn	$C(d) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\phi d)^\nu K_\nu(\phi d) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Matérn at $\nu = 3/2$	$C(d) = \begin{cases} \sigma^2 (1 + \phi d) \exp(-\phi d) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$

Table 2.1 Summary of common isotropic parametric covariance functions (covariograms).

7. *Wave*:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{\sin(\phi d)}{\phi d}\right) & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Note this is an example of a variogram that is not monotonically increasing. The associated covariance function is  $C(d) = \sigma^2 \sin(\phi d)/(\phi d)$ . Bessel functions of the first kind include the wave covariance function and are discussed in detail in Subsections 3.1.2 and 6.1.3.

8. *Power law*:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d^\lambda & \text{of } d > 0 \\ 0 & d = 0 \end{cases}.$$

This generalizes the linear case and produces valid intrinsic (albeit not weakly) stationary semivariograms provided  $0 \leq \lambda < 2$ .

9. *Matérn*: The variogram for the Matérn class is given by

$$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu}d\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}d\phi)\right] & \text{if } d > 0 \\ 0 & d = 0 \end{cases}. \quad (2.8)$$

This class was originally suggested by Matérn (1960, 1986). Interest in it was revived by Handcock and Stein (1993) and Handcock and Wallis (1994), who demonstrated attractive interpretations for  $\nu$  as well as  $\phi$ . In particular,  $\nu > 0$  is a parameter controlling the smoothness of the realized random field (see Subsection 3.1.4) while  $\phi$  is a spatial decay parameter. The function  $\Gamma(\cdot)$  is the usual gamma function while  $K_\nu$  is the modified Bessel function of order  $\nu$  (see, e.g., Abramowitz and Stegun, 1965, Chapter 9). Implementations of this function are available in several C/C++ libraries and also in R packages such as `spBayes` and `geoR`. Note that special cases of the above are the exponential ( $\nu = 1/2$ ) and the Gaussian ( $\nu \rightarrow \infty$ ). At  $\nu = 3/2$  we obtain a closed form as well, namely  $\gamma(d) =$

Model	Variogram, $\gamma(d)$
Linear	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Spherical	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 & \text{if } d \geq 1/\phi \\ \tau^2 + \sigma^2 \left[ \frac{3}{2}\phi d - \frac{1}{2}(\phi d)^3 \right] & \text{if } 0 < d \leq 1/\phi \\ 0 & d = 0 \end{cases}$
Exponential	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi d)) & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Powered exponential	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(- \phi d ^p)) & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Gaussian	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 d^2)) & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Rational quadratic	$\gamma(d) = \begin{cases} \tau^2 + \frac{\sigma^2 d^2}{(\phi + d^2)} & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Wave	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{\sin(\phi d)}{\phi d}\right) & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Power law	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d^\lambda & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Matérn	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(\phi d)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(\phi d)\right] & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Matérn at $\nu = 3/2$	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 [1 - (1 + \phi d) \exp(-\phi d)] & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$

Table 2.2 Summary of common parametric isotropic variograms.

$\tau^2 + \sigma^2 [1 - (1 + \phi d) \exp(-\phi d)]$  for  $t > 0$ . In fact, we obtain a polynomial times exponential form for the Matérn for all  $\nu$  of the form,  $\nu = k + \frac{1}{2}$  with  $k$  a non-negative integer. The Matérn covariance function is often reparametrized to  $\alpha = 2\sqrt{\nu}\phi$  along with  $\eta = \sigma^2\phi^{2\nu}$  and  $\nu$ . This transformation is helpful in providing better behaved model fitting, particularly using Markov chain Monte Carlo (see Chapter 6 for further discussion).

The covariance functions and variograms we have described in this subsection are conveniently summarized in Tables 2.1 and 2.2, respectively. An important point which the reader may wonder about is the fact that every presented covariance function and variogram has a discontinuity at 0. That is, the limit as  $d \rightarrow 0$  for the covariance function is  $\sigma^2$  not  $\sigma^2 + \tau^2$  and for the variogram is  $\tau^2$ , not 0. Evidently, this is not a typo! What is the reason? We elaborate this in greater detail in Chapter 3. Here, we offer a simple explanation. Consider the form of the residual for a spatial model. We might write it as  $r(\mathbf{s}) = w(\mathbf{s}) + \epsilon(\mathbf{s})$  where  $w(\mathbf{s})$  is a spatial process, say a stationary Gaussian process with mean 0 and covariance function  $\sigma^2\rho(\mathbf{s} - \mathbf{s}')$  and  $\epsilon(\mathbf{s})$  is a pure error process, i.e., the  $\epsilon(\mathbf{s})$ 's are i.i.d., say  $N(0, \tau^2)$  with the  $\epsilon$ 's independent of the  $w$ 's. Then, it is straightforward to compute  $\text{cov}(r(\mathbf{s}), r(\mathbf{s}')) = \sigma^2\rho(\mathbf{s} - \mathbf{s}')$  while  $\text{var}Y(\mathbf{s}) = \sigma^2 + \tau^2$ , whence the discontinuity at 0 emerges. And, in fact,  $\text{corr}(r(\mathbf{s}), r(\mathbf{s}'))$  is bounded by  $\sigma^2/(\sigma^2 + \tau^2)$ . The rationale for including the  $\epsilon(\mathbf{s})$ 's into the model is that we don't want to insist that all model error is spatial. Of course, we certainly want to include the  $w$ 's in order to be able to capture a spatial story. Possible explanations for the pure error contribution include: (i) measurement error associated with the data collection at a given location, (ii) replication error to express the possibility that repeated measurements at the same location might not provide identical observations, or (iii) micro-scale error to acknowledge that, though we never see

observations closer to each other than the minimum pairwise distance in our sample, there might be very fine scale structure represented as noise.

#### 2.1.4 Variogram model fitting

Having seen a fairly large selection of models for the variogram, one might well wonder how we choose one of them for a given data set, or whether the data can really distinguish them (see Subsection 6.1.3 in this latter regard). Historically, a variogram model is chosen by plotting the *empirical semivariogram* (Matheron, 1963), a simple nonparametric estimate of the semivariogram, and then comparing it to the various theoretical shapes available from the choices in the previous subsection. The customary empirical semivariogram is

$$\hat{\gamma}(d) = \frac{1}{2N(d)} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(d)} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2, \quad (2.9)$$

where  $N(d)$  is the set of pairs of points such that  $\|\mathbf{s}_i - \mathbf{s}_j\| = d$ , and  $|N(d)|$  is the number of pairs in this set. Notice that, unless the observations fall on a regular grid, the distances between the pairs will all be different, so this will not be a useful estimate as it stands. Instead we would “grid up” the  $d$ -space into intervals  $I_1 = (0, d_1)$ ,  $I_2 = (d_1, d_2)$ , and so forth, up to  $I_K = (d_{K-1}, d_K)$  for some (typically regular) grid  $0 < d_1 < \dots < d_K$ . Representing the  $d$  values in each interval by the midpoint of the interval, we then alter our definition of  $N(d)$  to

$$N(d_k) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in I_k\}, \quad k = 1, \dots, K.$$

Selection of an appropriate number of intervals  $K$  and location of the upper endpoint  $t_K$  is reminiscent of similar issues in histogram construction. Journel and Huijbregts (1979) recommend bins wide enough to capture at least 30 pairs per bin.

Clearly (2.9) is nothing but a method of moments (MOM) estimate, the semivariogram analogue of the usual sample variance estimate  $s^2$ . While very natural, there is reason to doubt that this is the best estimate of the semivariogram. Certainly it will be sensitive to outliers, and the sample average of the squared differences may be rather badly behaved since under a Gaussian distributional assumption for the  $Y(\mathbf{s}_i)$ , the squared differences will have a distribution that is a scale multiple of the heavily skewed  $\chi_1^2$  distribution. In this regard, Cressie and Hawkins (1980) proposed a robustified estimate that uses sample averages of  $|Y(\mathbf{s}_i) - Y(\mathbf{s}_j)|^{1/2}$ ; this estimate is available in several software packages (see Section 2.3 below). Perhaps more uncomfortable is the fact that (2.9) uses data differences, rather than the data itself. Also of concern is the fact that the components of the sum in (2.9) will be dependent within and across bins, and that  $N(d_k)$  will vary across bins.

In any case, an empirical semivariogram estimate can be plotted, viewed, and an appropriately shaped theoretical variogram model can be fitted to this “data.” Since any empirical estimate naturally carries with it a significant amount of noise in addition to its signal, this fitting of a theoretical model has traditionally been as much art as science: in any given real data setting, any number of different models (exponential, Gaussian, spherical, etc.) may seem equally appropriate. Indeed, fitting has historically been done “by eye,” or at best by using trial and error to choose values of nugget, sill, and range parameters that provide a good match to the empirical semivariogram (where the “goodness” can be judged visually or by using some least squares or similar criterion); again see Section 2.3. More formally, we could treat this as a statistical estimation problem, and use nonlinear maximization routines to find nugget, sill, and range parameters that minimize some goodness-of-fit criterion.

If we also have a distributional model for the data, we could use maximum likelihood (or restricted maximum likelihood, REML) to obtain sensible parameter estimates; see, e.g., Smith (2001) for details in the case of Gaussian data modeled with the various parametric

variogram families outlined in Subsection 2.1.3. In Chapter 5 and Chapter 6 we shall see that the hierarchical Bayesian approach is broadly similar to this latter method, although it will often be easier and more intuitive to work directly with the covariance model  $C(d)$ , rather than changing to a partial likelihood in order to introduce the semivariogram. In addition, we will gain full inference, e.g., posterior distributions for all unknowns of interest as well as more accurate assessment of uncertainty than appealing to arguably inappropriate asymptotics (see Chapter 5).

## 2.2 Anisotropy

Stationary correlation functions extend the class of correlation functions from isotropy where association only depends upon distance to association that depends upon the separation vector between locations. As a result, association depends upon direction. Here, we explore covariance functions that are stationary but not isotropic. (We defer discussion of nonstationary covariance functions to Chapter 3.) A simple example is the class where we *separate* the components. That is, suppose we write the components of  $\mathbf{s}$  as  $s_{lat}, s_{lon}$ , similarly for  $\mathbf{h}$ . Then, we can define  $\text{corr}(Y(\mathbf{s} + \mathbf{h}), Y(\mathbf{s})) = \rho_1(h_{lat})\rho_2(h_{lon})$  where  $\rho_1$  and  $\rho_2$  are valid correlation functions on  $\mathfrak{R}^1$ . Evidently, this correlation function is stationary but depends on direction. In particular, if we switch  $h_{lat}$  and  $h_{lon}$  we will get a different value for the correlation even though  $\|\mathbf{h}\|$  is unchanged. A common choice is  $e^{\phi_1(|h_{lat}|) + \phi_2(|h_{lon}|)}$ .

The separable correlation function is usually extended to a covariance function by introducing the multiplier  $\sigma^2$ , as we have done above. This covariance function tends to be used, for instance, with Gaussian processes in computer model settings (Santner, Williams and Notz, 2003; Rasmussen and Williams, 2006; Oakley and O’Hagan, 2004; Kennedy and O’Hagan, 2001). Here we seek a response surface over covariate space and, since the covariates live on their own spaces with their own scales, component-wise dependence seems appropriate. In the spatial setting, since lat and lon are on the same scale, it may be less suitable.

### 2.2.1 Geometric anisotropy

A commonly used class of stationary covariance functions is the geometric anisotropic covariance functions where we set

$$C(\mathbf{s} - \mathbf{s}') = \sigma^2 \rho((\mathbf{s} - \mathbf{s}')^T B (\mathbf{s} - \mathbf{s}')) . \quad (2.10)$$

In (2.10),  $B$  is positive definite with  $\rho$  a valid correlation function in  $\mathfrak{R}^r$  (say, from Table 2.1). We would omit the range/decay parameter since it can be incorporated into  $B$ . When  $r = 2$  we obtain a specification with three parameters rather than one. Contours of constant association arising from  $c$  in (2.10) are elliptical. In particular, the contour corresponding to  $\rho = .05$  provides the range in each spatial direction. Ecker and Gelfand (1997) provide the details for Bayesian modeling and inference incorporating (2.10); see also Subsection 6.1.4.

Following the discussion in Subsection 3.1.2, we can extend geometric anisotropy to *product* geometric anisotropy. In the simplest case, we would set

$$C(\mathbf{s} - \mathbf{s}') = \sigma^2 \rho_1((\mathbf{s} - \mathbf{s}')^T B_1 (\mathbf{s} - \mathbf{s}')) \rho_2((\mathbf{s} - \mathbf{s}')^T B_2 (\mathbf{s} - \mathbf{s}')) , \quad (2.11)$$

noting that  $c$  is valid since it arises as a product of valid covariance functions. See Ecker and Gelfand (2003) for further details and examples. Evidently, we can extend to a product of more than two geometric anisotropy forms and we can create rich directional range behavior. However, a challenge with (2.11) is that it introduces 7 parameters into the covariance function and it will be difficult to identify and learn about all of them unless we have many, many locations.

### 2.2.2 Other notions of anisotropy

In a more general discussion, Zimmerman (1993) suggests three different notions of anisotropy: *sill* anisotropy, *nugget* anisotropy, and *range* anisotropy. More precisely, working with a variogram  $\gamma(\mathbf{h})$ , let  $\mathbf{h}$  be an arbitrary separation vector so that  $\mathbf{h}/\|\mathbf{h}\|$  is a unit vector in  $\mathbf{h}$ 's direction. Consider  $\gamma(a\mathbf{h}/\|\mathbf{h}\|)$ . Let  $a \rightarrow \infty$  and suppose  $\lim_{a \rightarrow \infty} \gamma(a\mathbf{h}/\|\mathbf{h}\|)$  depends upon  $\mathbf{h}$ . This situation is naturally referred to as sill anisotropy. If we work with the usual relationship  $\gamma(a\mathbf{h}/\|\mathbf{h}\|) = \tau^2 + \sigma^2 \left(1 - \rho \left(a \frac{\mathbf{h}}{\|\mathbf{h}\|}\right)\right)$ , then, in some directions,  $\rho$  must not go to 0 as  $a \rightarrow \infty$ . If this can be the case, then ergodicity assumptions (i.e., convergence assumptions associated with averaging) will be violated. If so, then perhaps the constant mean assumption, implicit for the variogram, does not hold. Alternatively, it is also possible that the constant nugget assumption fails.

Instead, let  $a \rightarrow 0$  and suppose  $\lim_{a \rightarrow 0} \gamma(a\mathbf{h}/\|\mathbf{h}\|)$  depends upon  $\mathbf{h}$ . This situation is referred to as nugget anisotropy. Since, by definition,  $\rho$  must go to 1 as  $a \rightarrow 0$ , this case says that the assumption of uncorrelated measurement errors with common variance may not be appropriate. In particular, a simple white noise process model with constant nugget for the nonspatial errors is not appropriate.

A third type of anisotropy is range anisotropy where the range depends upon direction. Zimmerman (1993) asserts that “this is the form most often seen in practice.” Geometric anisotropy and the more general product geometric anisotropy from the previous subsections are illustrative cases. However, given the various constructive strategies offered in Subsection 3.1.2 to create more general stationary covariance functions, we can envision nongeometric range anisotropy, implying general correlation function or variogram contours in  $\mathbb{R}^2$ . However, due to the positive definiteness restriction on the correlation function, the extent of possible contour shapes is still rather limited.

Lastly, motivated by directional variograms (see Subsection 2.3.2), some authors propose the idea of nested models (see Zimmerman, 1993, and the references therein). That is, for each separation vector there is an associated angle with, say, the  $x$ -axis, which by symmetry considerations can be restricted to  $[0, \pi)$ . Partitioning this interval into a set of angle classes, a different variogram model is assumed to operate for each class. In terms of correlations, this would imply a different covariance function is operating for each angle class. But evidently this does not define a valid process model: the resulting covariance matrix for an arbitrary set of locations need not be positive definite.

This can be seen with as few as three points and two angle classes. Let  $(\mathbf{s}_1, \mathbf{s}_2)$  belong to one angle class with  $(\mathbf{s}_1, \mathbf{s}_3)$  and  $(\mathbf{s}_2, \mathbf{s}_3)$  in the other. With exponential isotropic correlation functions in each class by choosing  $\phi_1$  and  $\phi_2$  appropriately we can make  $\rho(\mathbf{s}_1 - \mathbf{s}_2) \approx 0$  while  $\rho(\mathbf{s}_1 - \mathbf{s}_3) = \rho(\mathbf{s}_2 - \mathbf{s}_3) \approx 0.8$ . A quick calculation shows that the resulting  $3 \times 3$  covariance (correlation) matrix is not positive definite. So, in terms of being able to write proper joint distributions for the resulting data, nested models are inappropriate; they do not provide an extension of isotropy that allows for likelihood based inference.

## 2.3 Exploratory approaches for point-referenced data

### 2.3.1 Basic techniques

Exploratory data analysis (EDA) tools are routinely implemented in the process of analyzing one- and two-sample data sets, regression studies, generalized linear models, etc. (see, e.g., Chambers et al., 1983; Hoaglin, Mosteller, and Tukey, 1983, 1985; Aiktin et al., 1989). Similarly, such tools are appropriate for analyzing point-referenced spatial data.

For continuous spatial data, the starting point is the so-called “first law of geostatistics.” Figure 2.2 illustrates this “law” in a one-dimensional setting. The data is partitioned into a mean term and an error term. The mean corresponds to global (or *first-order*) behavior,

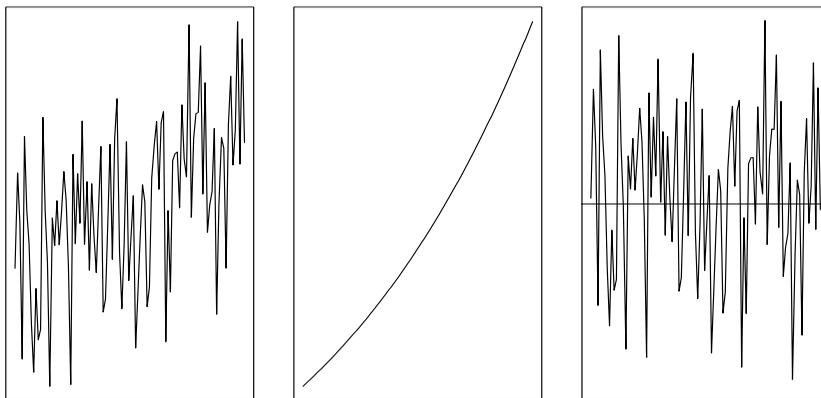


Figure 2.2 *Illustration of the first law of geostatistics.*

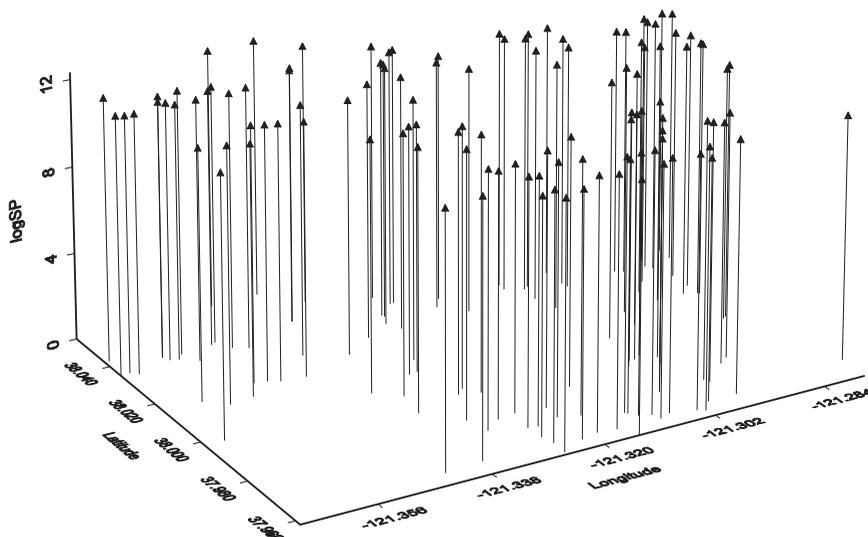


Figure 2.3 *Illustrative three-dimensional “drop line” scatterplot, Stockton data.*

while the error captures local (or *second-order*) behavior through a covariance function. EDA tools are available to examine both first- and second-order behavior.

The law also clarifies that spatial association in the data,  $Y(\mathbf{s})$ , need not resemble spatial association in the residuals,  $\epsilon(\mathbf{s})$ . That is, spatial association in the  $Y(\mathbf{s})$  corresponds to looking at  $E(Y(\mathbf{s}) - \mu)(Y(\mathbf{s}') - \mu)$ , while spatial structure in the  $\epsilon(\mathbf{s})$  corresponds to looking at  $E(Y(\mathbf{s}) - \mu(\mathbf{s}))(Y(\mathbf{s}') - \mu(\mathbf{s}'))$ . The difference between the former and the latter is  $(\mu - \mu(\mathbf{s}))(\mu - \mu(\mathbf{s}'))$ , which, if interest is in spatial regression, we would not expect to be negligible.

Certainly an initial exploratory display should be a simple map of the locations themselves. We need to assess how *regular* the arrangement of the points is and also whether there is a much larger maximum distance between points in some directions than in others. Next, some authors would recommend a stem-and-leaf display of the  $Y(\mathbf{s})$ . This plot is evidently nonspatial and is customarily for observations which are i.i.d. We expect both nonconstant mean and spatial dependence, but such a plot may at least suggest potential outliers. Next

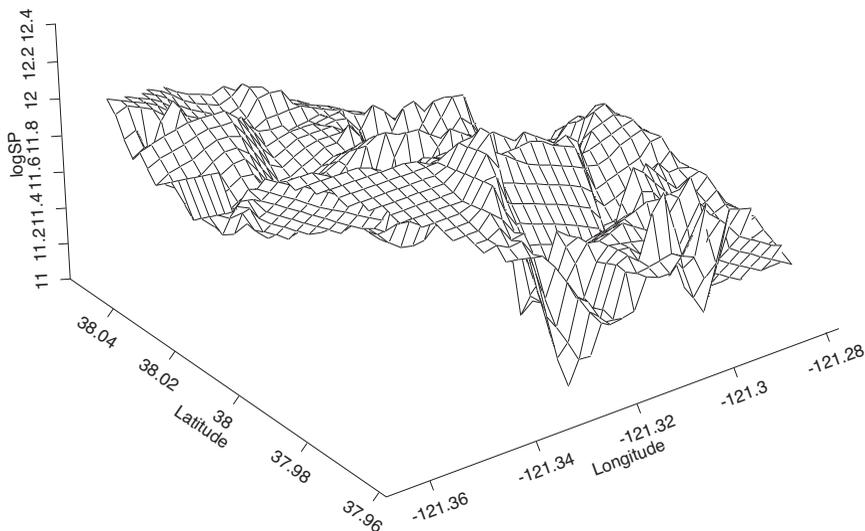


Figure 2.4 *Illustrative three-dimensional surface (“perspective”) plot, Stockton real estate data.*

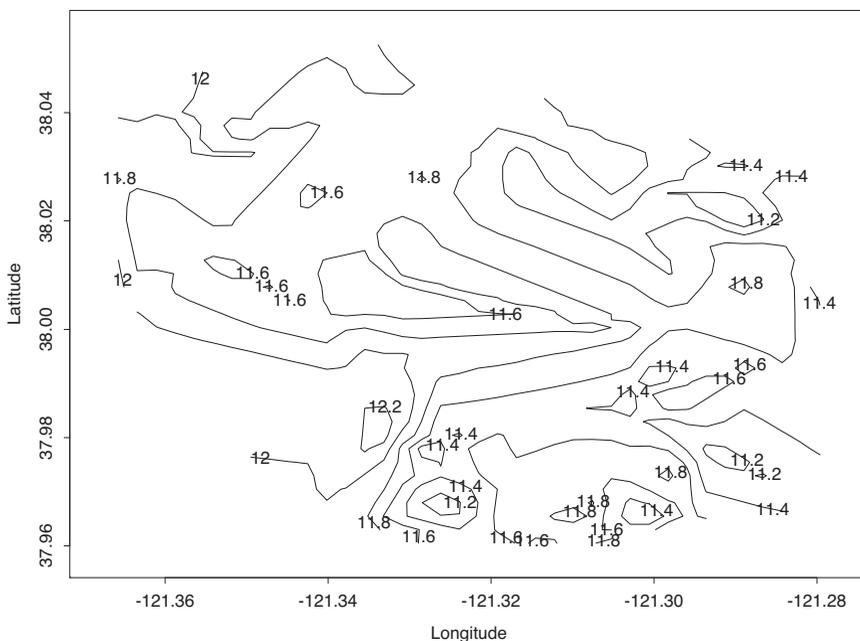


Figure 2.5 *Illustrative contour plot, Stockton real estate data.*

we might develop a three-dimensional “drop line” scatterplot of  $Y(\mathbf{s}_i)$  versus  $\mathbf{s}_i$ , alternatively, a three-dimensional surface plot or perhaps, a contour plot as a *smoothed* summary. Examples of these three plots are shown for a sample of 120 log-transformed home selling prices in Stockton, CA, in Figures 2.3, 2.4, and 2.5, respectively. Of these three, we find the contour plot to be the most effective in revealing the entire spatial surface. However, as the preceding paragraph clarifies, such displays may be deceiving. They may show spatial pattern that will disappear after  $\mu(\mathbf{s})$  is fitted, or perhaps vice versa. It seems more sensible to study spatial pattern in the residuals.

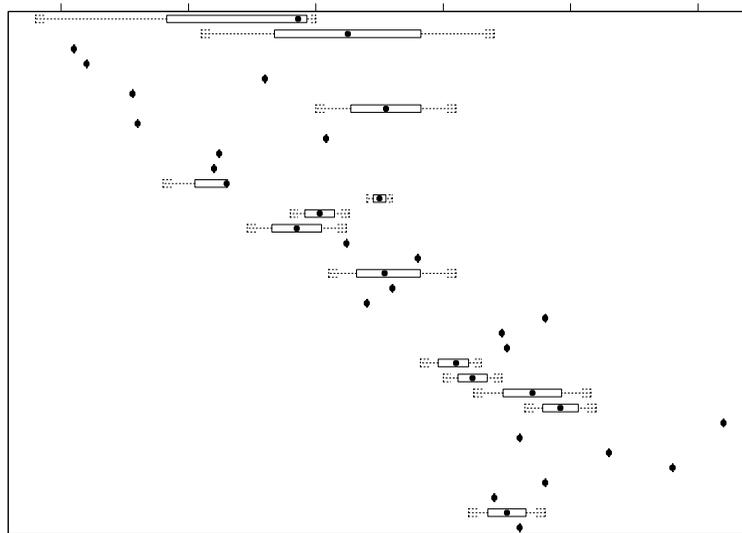


Figure 2.6 *Illustrative row box plots, Diggle and Ribeiro (2002) surface elevation data.*

In exploring  $\mu(\mathbf{s})$  we may have two types of information at location  $\mathbf{s}$ . One is the purely geographic information, i.e., the geocoded location expressed in latitude and longitude or as projected coordinates such as eastings and northings (Subsection 1.2.1 above). The other will be features relevant for explaining the  $Y(\mathbf{s})$  at  $\mathbf{s}$ . For instance, if  $Y(\mathbf{s})$  is a pollution concentration, then elevation, temperature, and wind information at  $\mathbf{s}$  could well be useful and important. If, instead,  $Y(\mathbf{s})$  is the selling price of a single-family home at  $\mathbf{s}$ , then characteristics of the home (square feet, age, number of bathrooms, etc.) would be useful.

When the mean is described purely through geographic information,  $\mu(\mathbf{s})$  is referred to as a *trend surface*. When  $\mathbf{s} \in \mathbb{R}^2$ , the surface is usually developed as a low-dimensional bivariate polynomial. For data that is roughly gridded (or can be assigned to row and column bins by overlaying a regular lattice on the points), we can make row and column boxplots looking for trend. Displaying these boxplots versus their center could clarify the existence and nature of such trend. In fact, median polishing (see, e.g., Hoaglin, Mosteller, and Tukey, 1985) could be used to extract row and column effects, and also to see if a multiplicative trend surface term is useful; see Cressie (1983, pp. 46–48) in this regard.

Figures 2.6 and 2.7 illustrate the row and column boxplot approach for a data set previously considered by Diggle and Ribeiro (2002). The response variable is the surface elevation (“height”) at 52 locations on a regular grid within a 310-foot square (and where the mesh of the grid is 50 feet). The plots reveals some evidence of spatial pattern as we move along the rows, but not down the columns of the regular grid.

To assess small-scale behavior, some authors recommend creating the *semivariogram cloud*, i.e., a plot of  $(Y(\mathbf{s}_i) - Y(\mathbf{s}_j))^2$  versus  $\|\mathbf{s}_i - \mathbf{s}_j\|$ . Usually this cloud is too “noisy” to reveal very much; see, e.g., Figure 6.1. The empirical semivariogram (2.9) is preferable in terms of reducing some of the noise, and can be a helpful tool in assessing the presence of spatial structure. Again, the caveat above suggests employing it for residuals (not the data itself) unless a constant mean is appropriate.