

Contents

Preface

1 Introduction

1.1	Motivating examples	1
1.2	Terminology and notation	9
1.2.1	Support	9
1.2.2	Multivariate responses and explanatory variables	10
1.2.3	Sampling design	12
1.3	Scientific objectives	12
1.4	Generalised linear geostatistical models	13
1.5	What is in this book?	15
1.5.1	Organisation of the book	16
1.5.2	Statistical pre-requisites	17
1.6	Computation	17
1.6.1	Elevation data	17
1.6.2	More on the <code>geodata</code> object	20
1.6.3	Rongelap data	22
1.6.4	The Gambia malaria data	24
1.6.5	The soil data	24
1.7	Exercises	26

2 An overview of model-based geostatistics

2.1	Design	27
2.2	Model formulation	28
2.3	Exploratory data analysis	30
2.3.1	Non-spatial exploratory analysis	30

2.3.2	Spatial exploratory analysis	31
2.4	The distinction between parameter estimation and spatial prediction	35
2.5	Parameter estimation	36
2.6	Spatial prediction	37
2.7	Definitions of distance	39
2.8	Computation	40
2.9	Exercises	45
3	Gaussian models for geostatistical data	46
3.1	Covariance functions and the variogram	46
3.2	Regularisation	48
3.3	Continuity and differentiability of stochastic processes	49
3.4	Families of covariance functions and their properties	51
3.4.1	The Matérn family	51
3.4.2	The powered exponential family	53
3.4.3	Other families	54
3.5	The nugget effect	56
3.6	Spatial trends	57
3.7	Directional effects	58
3.8	Transformed Gaussian models	60
3.9	Intrinsic models	63
3.10	Unconditional and conditional simulation	66
3.11	Low-rank models	68
3.12	Multivariate models	69
3.12.1	Cross-covariance, cross-correlation and cross-variogram	70
3.12.2	Bivariate signal and noise	71
3.12.3	Some simple constructions	72
3.13	Computation	74
3.14	Exercises	77
4	Generalized linear models for geostatistical data	79
4.1	General formulation	79
4.2	The approximate covariance function and variogram	81
4.3	Examples of generalised linear geostatistical models	82
4.3.1	The Poisson log-linear model	82
4.3.2	The binomial logistic-linear model	83
4.3.3	Spatial survival analysis	84
4.4	Point process models and geostatistics	86
4.4.1	Cox processes	87
4.4.2	Preferential sampling	89
4.5	Some examples of other model constructions	93
4.5.1	Scan processes	93
4.5.2	Random sets	94
4.6	Computation	94
4.6.1	Simulating from the generalised linear model	94
4.6.2	Preferential sampling	96

4.7	Exercises	97
5	Classical parameter estimation	99
5.1	Trend estimation	100
5.2	Variograms	100
5.2.1	The theoretical variogram	100
5.2.2	The empirical variogram	102
5.2.3	Smoothing the empirical variogram	102
5.2.4	Exploring directional effects	104
5.2.5	The interplay between trend and covariance structure	105
5.3	Curve-fitting methods for estimating covariance structure	107
5.3.1	Ordinary least squares	108
5.3.2	Weighted least squares	108
5.3.3	Comments on curve-fitting methods	110
5.4	Maximum likelihood estimation	112
5.4.1	General ideas	112
5.4.2	Gaussian models	112
5.4.3	Profile likelihood	114
5.4.4	Application to the surface elevation data	114
5.4.5	Restricted maximum likelihood estimation for the Gaussian linear model	116
5.4.6	Trans-Gaussian models	117
5.4.7	Analysis of Swiss rainfall data	118
5.4.8	Analysis of soil calcium data	121
5.5	Parameter estimation for generalized linear geostatistical models	123
5.5.1	Monte Carlo maximum likelihood	124
5.5.2	Hierarchical likelihood	125
5.5.3	Generalized estimating equations	125
5.6	Computation	126
5.6.1	Variogram calculations	126
5.6.2	Parameter estimation	130
5.7	Exercises	132
6	Spatial prediction	134
6.1	Minimum mean square error prediction	134
6.2	Minimum mean square error prediction for the stationary Gaussian model	136
6.2.1	Prediction of the signal at a point	136
6.2.2	Simple and ordinary kriging	137
6.2.3	Prediction of linear targets	138
6.2.4	Prediction of non-linear targets	138
6.3	Prediction with a nugget effect	139
6.4	What does kriging actually do to the data?	140
6.4.1	The prediction weights	141
6.4.2	Varying the correlation parameter	144
6.4.3	Varying the noise-to-signal ratio	146

6.5	Trans-Gaussian kriging	147
6.5.1	Analysis of Swiss rainfall data (continued)	149
6.6	Kriging with non-constant mean	151
6.6.1	Analysis of soil calcium data (continued)	151
6.7	Computation	151
6.8	Exercises	155
7	Bayesian inference	157
7.1	The Bayesian paradigm: a unified treatment of estimation and prediction	157
7.1.1	Prediction using plug-in estimates	157
7.1.2	Bayesian prediction	158
7.1.3	Obstacles to practical Bayesian prediction	160
7.2	Bayesian estimation and prediction for the Gaussian linear model	160
7.2.1	Estimation	161
7.2.2	Prediction when correlation parameters are known	163
7.2.3	Uncertainty in the correlation parameters	164
7.2.4	Prediction of targets which depend on both the signal and the spatial trend	165
7.3	Trans-Gaussian models	166
7.4	Case studies	167
7.4.1	Surface elevations	167
7.4.2	Analysis of Swiss rainfall data (continued)	169
7.5	Bayesian estimation and prediction for generalized linear geostatistical models	172
7.5.1	Markov chain Monte Carlo	172
7.5.2	Estimation	173
7.5.3	Prediction	176
7.5.4	Some possible improvements to the MCMC algorithm	177
7.6	Case studies in generalized linear geostatistical modelling	179
7.6.1	Simulated data	179
7.6.2	Rongelap island	181
7.6.3	Childhood malaria in The Gambia	185
7.6.4	<i>Loa loa</i> prevalence in equatorial Africa	187
7.7	Computation	193
7.7.1	Gaussian models	193
7.7.2	Non-Gaussian models	196
7.8	Exercises	196
8	Geostatistical design	199
8.1	Choosing the study region	201
8.2	Choosing the sample locations: uniform designs	202
8.3	Designing for efficient prediction	203
8.4	Designing for efficient parameter estimation	204
8.5	A Bayesian design criterion	206
8.5.1	Retrospective design	206

8.5.2	Prospective design	209
8.6	Exercises	211
A	Statistical background	213
A.1	Statistical models	213
A.2	Classical inference	213
A.3	Bayesian inference	215
A.4	Prediction	216
	References	218
	Index	227

Motivating examples

The term *spatial statistics* is used to describe a wide range of statistical models and methods intended for the analysis of spatially-referenced data. Cressie (1991) provides a general overview. Within spatial statistics, the term *geostatistics* refers to models and methods for data with the following characteristics. Firstly, values Y_i , $i = 1, \dots, n$ are observed at a discrete set of sampling locations x_i within some spatial region A . Secondly, each observed value Y_i is a direct measurement of, or is statistically related to, the value of an underlying continuous spatial phenomenon, $S(z)$, at the corresponding sampling location x_i . This rather abstract formulation can be translated to a variety of imaginable scientific settings, as the following examples demonstrate.

1.1. Surface elevations

The data for this example are taken from Davis (1972). They give the measured surface elevations y_i at each of 52 locations x_i within a square, A , with side length 1 unit. The unit of distance is 50 feet (≈ 15.24 meters), whereas one unit of elevation represents 10 feet (≈ 3.05 meters) of elevation.

Figure 1.1 is a *circle plot* of the data. Each datum (x_i, y_i) is represented by a small circle with centre at x_i and radius proportional to y_i . The observed elevations range between 690 and 960 units. For the plot, we have subtracted 600 from each observed elevation, to heighten the visual contrast between low and high values. Note in particular the cluster of low values near the top-centre of the