ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/299981435

Spatial Data Analysis: Theory and Practice

Book · April 2003

DOI: 10.1017/CBO9780511754944

CITATIONS
969

READS
6,624

1 author:

Robert Haining
University of Cambridge
194 PUBLICATIONS **6,086** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Modelling the geographical distribution of cases of outdoor rape in Stockholm (with Vania Ceccato and Philip Li); View project

Developing a spatial decision support system (SDSS) for responding to outbreaks of malaria in Kalaburagi, India View project

All content following this page was uploaded by Robert Haining on 02 January 2014.

The user has requested enhancement of the downloaded file.

Spatial Data Analysis

Theory and Practice

ROBERT HAINING
University of Cambridge

CAMBRIDGE
UNIVERSITY PRESS

Contents

Preface xv
Acknowledgements xvii

Introduction 1
0.1 About the book 1
0.2 What is spatial data analysis? 4
0.3 Motivation for the book 5
0.4 Organization 8
0.5 The spatial data matrix 10

Part A **The context for spatial data analysis**

1 Spatial data analysis: scientific and policy context 15
1.1 Spatial data analysis in science 15
1.1.1 Generic issues of place, context and space in scientific explanation 16
(a) Location as place and context 16
(b) Location and spatial relationships 18
1.1.2 Spatial processes 21
1.2 Place and space in specific areas of scientific explanation 22
1.2.1 Defining spatial subdisciplines 22
1.2.2 Examples: selected research areas 24
(a) Environmental criminology 24
(b) Geographical and environmental (spatial) epidemiology 26
(c) Regional economics and the new economic geography 29

(d) Urban studies 31
(e) Environmental sciences 32
1.2.3 Spatial data analysis in problem solving 33
1.3 Spatial data analysis in the policy area 36
1.4 Some examples of problems that arise in analysing spatial data 40
1.4.1 Description and map interpretation 40
1.4.2 Information redundancy 41
1.4.3 Modelling 41
1.5 Concluding remarks 41

2 The nature of spatial data 43
2.1 The spatial data matrix: conceptualization and representation issues 44
2.1.1 Geographic space: objects, fields and geometric representations 44
2.1.2 Geographic space: spatial dependence in attribute values 46
2.1.3 Variables 47
(a) Classifying variables 48
(b) Levels of measurement 50
2.1.4 Sample or population? 51
2.2 The spatial data matrix: its form 54
2.3 The spatial data matrix: its quality 57
2.3.1 Model quality 58
(a) Attribute representation 59
(b) Spatial representation: general considerations 59
(c) Spatial representation: resolution and aggregation 61
2.3.2 Data quality 61
(a) Accuracy 63
(b) Resolution 67
(c) Consistency 70
(d) Completeness 71
2.4 Quantifying spatial dependence 74
(a) Fields: data from two-dimensional continuous space 74
(b) Objects: data from two-dimensional discrete space 79
2.5 Concluding remarks 87

Part B **Spatial data: obtaining data and quality issues**

3 Obtaining spatial data through sampling 91
3.1 Sources of spatial data 91
3.2 Spatial sampling 93
3.2.1 The purpose and conduct of spatial sampling 93
3.2.2 Design- and model-based approaches to spatial sampling 96
(a) Design-based approach to sampling 96
(b) Model-based approach to sampling 98
(c) Comparative comments 99
3.2.3 Sampling plans 100
3.2.4 Selected sampling problems 103
(a) Design-based estimation of the population mean 103
(b) Model-based estimation of means 106
(c) Spatial prediction 107
(d) Sampling to identify extreme values or detect rare events 108
3.3 Maps through simulation 113

4 Data quality: implications for spatial data analysis 116
4.1 Errors in data and spatial data analysis 116
4.1.1 Models for measurement error 116
(a) Independent error models 117
(b) Spatially correlated error models 118
4.1.2 Gross errors 119
(a) Distributional outliers 119
(b) Spatial outliers 122
(c) Testing for outliers in large data sets 123
4.1.3 Error propagation 124
4.2 Data resolution and spatial data analysis 127
4.2.1 Variable precision and tests of significance 128
4.2.2 The change of support problem 129
(a) Change of support in geostatistics 129
(b) Areal interpolation 131
4.2.3 Analysing relationships using aggregate data 138
(a) Ecological inference: parameter estimation 141
(b) Ecological inference in environmental epidemiology: identifying valid hypotheses 147
(c) The modifiable areal units problem (MAUP) 150

4.3 Data consistency and spatial data analysis 151
4.4 Data completeness and spatial data analysis 152
4.4.1 The missing-data problem 154
(a) Approaches to analysis when data are missing 156
(b) Approaches to analysis when spatial data are missing 159
4.4.2 Spatial interpolation, spatial prediction 164
4.4.3 Boundaries, weights matrices and data completeness 174
4.5 Concluding remarks 177

Part C **The exploratory analysis of spatial data**

5 Exploratory spatial data analysis: conceptual models 181
5.1 EDA and ESDA 181
5.2 Conceptual models of spatial variation 183
(a) The regional model 183
(b) Spatial 'rough' and 'smooth' 184
(c) Scales of spatial variation 185

6 Exploratory spatial data analysis: visualization methods 188
6.1 Data visualization and exploratory data analysis 188
6.1.1 Data visualization: approaches and tasks 189
6.1.2 Data visualization: developments through computers 192
6.1.3 Data visualization: selected techniques 193
6.2 Visualizing spatial data 194
6.2.1 Data preparation issues for aggregated data: variable values 194
6.2.2 Data preparation issues for aggregated data: the spatial framework 199
(a) Non-spatial approaches to region building 200
(b) Spatial approaches to region building 201
(c) Design criteria for region building 203
6.2.3 Special issues in the visualization of spatial data 206
6.3 Data visualization and exploratory spatial data analysis 210
6.3.1 Spatial data visualization: selected techniques for univariate data 211
(a) Methods for data associated with point or area objects 211
(b) Methods for data from a continuous surface 215
6.3.2 Spatial data visualization: selected techniques for bi- and multi-variate data 218

6.3.3 Uptake of breast cancer screening in Sheffield 219
6.4 Concluding remarks 225

7 Exploratory spatial data analysis: numerical methods 226
7.1 Smoothing methods 227
7.1.1 Resistant smoothing of graph plots 227
7.1.2 Resistant description of spatial dependencies 228
7.1.3 Map smoothing 228
(a) Simple mean and median smoothers 230
(b) Introducing distance weighting 230
(c) Smoothing rates 232
(d) Non-linear smoothing: headbanging 234
(e) Non-linear smoothing: median polishing 236
(f) Some comparative examples 237
7.2 The exploratory identification of global map properties: overall clustering 237
7.2.1 Clustering in area data 242
7.2.2 Clustering in a marked point pattern 247
7.3 The exploratory identification of local map properties 250
7.3.1 Cluster detection 251
(a) Area data 251
(b) Inhomogeneous point data 259
7.3.2 Focused tests 263
7.4 Map comparison 265
(a) Bivariate association 265
(b) Spatial association 268

Part D **Hypothesis testing and spatial autocorrelation**

8 Hypothesis testing in the presence of spatial dependence 273
8.1 Spatial autocorrelation and testing the mean of a spatial data set 275
8.2 Spatial autocorrelation and tests of bivariate association 278
8.2.1 Pearson's product moment correlation coefficient 278
8.2.2 Chi-square tests for contingency tables 283

Part E **Modelling spatial data**
9 Models for the statistical analysis of spatial data 289
9.1 Descriptive models 292
9.1.1 Models for large-scale spatial variation 293

9.1.2 Models for small-scale spatial variation 293
(a) Models for data from a surface 293
(b) Models for continuous-valued area data 297
(c) Models for discrete-valued area data 304
9.1.3 Models with several scales of spatial variation 306
9.1.4 Hierarchical Bayesian models 307
9.2 Explanatory models 312
9.2.1 Models for continuous-valued response variables: normal regression models 312
9.2.2 Models for discrete-valued area data: generalized linear models 316
9.2.3 Hierarchical models
(a) Adding covariates to hierarchical Bayesian models 320
(b) Modelling spatial context: multi-level models 321

10 Statistical modelling of spatial variation: descriptive modelling 325
10.1 Models for representing spatial variation 325
10.1.1 Models for continuous-valued variables 326
(a) Trend surface models with independent errors 326
(b) Semi-variogram and covariance models 327
(c) Trend surface models with spatially correlated errors 331
10.1.2 Models for discrete-valued variables 334
10.2 Some general problems in modelling spatial variation 338
10.3 Hierarchical Bayesian models 339

11 Statistical modelling of spatial variation: explanatory modelling 350
11.1 Methodologies for spatial data modelling 350
11.1.1 The 'classical' approach 350
11.1.2 The econometric approach 353
(a) A general spatial specification 355
(b) Two models of spatial pricing 356
11.1.3 A 'data-driven' methodology 358
11.2 Some applications of linear modelling of spatial data 358
11.2.1 Testing for regional income convergence 359
11.2.2 Models for binary responses 361
(a) A logistic model with spatial lags on the covariates 361
(b) Autologistic models with covariates 364
11.2.3 Multi-level modelling 365

11.2.4 Bayesian modelling of burglaries in Sheffield 367
11.2.5 Bayesian modelling of children excluded from school 376
11.3 Concluding comments 378

Appendix I Software 379

Appendix II Cambridgeshire lung cancer data 381

Appendix III Sheffield burglary data 385

Appendix IV Children excluded from school: Sheffield 391

References 394
Index 424