Contents

Preface xiii

Part 1. Integrated Analysis in Geography: The Way to Cloud Computing xix

Introduction to Part 1 xxi
Dominique LAFFLY

Chapter 1. Geographical Information and Landscape, Elements of Formalization 1
Dominique LAFFLY

Chapter 2. Sampling Strategies 7
Dominique LAFFLY

2.1. References 18

Chapter 3. Characterization of the Spatial Structure 19
Dominique LAFFLY

Chapter 4. Thematic Information Structures 27
Dominique LAFFLY

Chapter 5. From the Point to the Surface, How to Link Endogenous and Exogenous Data 35
Dominique LAFFLY

5.1. References 44

Chapter 6. Big Data in Geography 45
Dominique LAFFLY

Conclusion to Part 1 55
Dominique LAFFLY

Part 2. Basic Mathematical, Statistical and Computational Tools 59

Chapter 7. An Introduction to Machine Learning 61
Hichem SAHLI

7.1. Predictive modeling: introduction 61
7.2. Bayesian modeling. 61
7.2.1. Basic probability theory 62
7.2.2. Bayes rule 63
7.2.3. Parameter estimation 63
7.2.4. Learning Gaussians 64
7.3. Generative versus discriminative models 66
7.4. Classification. 67
7.4.1. Naïve Bayes 68
7.4.2. Support vector machines 69
7.5. Evaluation metrics for classification evaluation. 71
7.5.1. Confusion matrix-based measures. 71
7.5.2. Area under the ROC curve (AUC) 73
7.6. Cross-validation and over-fitting 73
7.7. References 74

Chapter 8. Multivariate Data Analysis 75
Astrid JOURDAN and Dominique LAFFLY

8.1. Introduction 75
8.2. Principal component analysis. 77
8.2.1. How to measure the information. 78
8.2.2. Scalar product and orthogonal variables 80
8.2.3. Construction of the principal axes. 81
8.2.4. Analysis of the principal axes 84
8.2.5. Analysis of the data points 86
8.3. Multiple correspondence analysis 88
8.3.1. Indicator matrix 89
8.3.2. Cloud of data points 90
8.3.3. Cloud of levels 92
8.3.4. MCA or PCA? 94

8.4. Clustering . . . 96
8.4.1. Distance between data points . . . 97
8.4.2. Dissimilarity criteria between clusters . . . 98
8.4.3. Variance (inertia) decomposition . . . 99
8.4.4. k-means method . . . 101
8.4.5. Agglomerative hierarchical clustering . . . 104
8.5. References . . . 105

Chapter 9. Sensitivity Analysis . . . 107
Astrid JOURDAN and Peio LOUBIÈRE

9.1. Generalities . . . 107
9.2. Methods based on linear regression . . . 109
9.2.1. Presentation . . . 109
9.2.2. R practice . . . 111
9.3. Morris' method . . . 114
9.3.1. Elementary effects method (Morris' method) . . . 114
9.3.2. R practice . . . 117
9.4. Methods based on variance analysis . . . 119
9.4.1. Sobol' indices . . . 120
9.4.2. Estimation of the Sobol' indices . . . 122
9.4.3. R practice . . . 123
9.5. Conclusion . . . 126
9.6. References . . . 127

Chapter 10. Using R for Multivariate Analysis . . . 129
Astrid JOURDAN

10.1. Introduction . . . 129
10.1.1. The dataset . . . 131
10.1.2. The variables . . . 134
10.2. Principal component analysis . . . 136
10.2.1. Eigenvalues . . . 137
10.2.2. Data points (Individuals) . . . 139
10.2.3. Supplementary variables . . . 143
10.2.4. Other representations . . . 143
10.3. Multiple correspondence analysis . . . 144
10.4. Clustering . . . 145
10.4.1. k-means algorithm . . . 145
10.5. References . . . 151

Part 3. Computer Science 153

Chapter 11. High Performance and Distributed Computing 155
Sebastiano Fabio SCHIFANO, Eleonora LUPPI, Didin Agustian PERMADI, Thi Kim Oanh NGUYEN, Nhat Ha Chi NGUYEN and Luca TOMASSETTI

11.1. High performance computing 155
11.2. Systems based on multi-core CPUs 157
11.2.1. Systems based on GPUs 159

Chapter 12. Introduction to Distributed Computing 163
Eleonora LUPPI

12.1. Introduction 163
12.1.1. A brief history 163
12.1.2. Design requirements 165
12.1.3. Models 168
12.1.4. Grid computing 171
12.2. References 176

Chapter 13. Towards Cloud Computing 179
Peio LOUBIÈRE and Luca TOMASSETTI

13.1. Introduction 179
13.1.1. Generalities 179
13.1.2. Benefits and drawbacks 180
13.2. Service model 180
13.2.1. Software as a Service 181
13.2.2. Platform as a Service 182
13.2.3. Infrastructure as a Service 182
13.2.4. And many more: XaaS 182
13.3. Deployment model 183
13.3.1. Public cloud 183
13.3.2. Private cloud 183
13.3.3. Hybrid cloud 184
13.4. Behind the hood, a technological overview 184
13.4.1. Structure 184
13.4.2. Virtualization 185
13.4.3. Scalability 186
13.4.4. Web-Oriented Architecture 187
13.5. Conclusion 187
13.6. References 188

Chapter 14. Web-Oriented Architecture – How to design a RESTFull API 191
Florent DEVIN

14.1. Introduction 191
14.2. Web services 192
14.2.1. Introduction 192
14.2.2. SOAP web services 193
14.2.3. REST web services 195
14.3. Web-Oriented Applications – Microservice applications 198
14.3.1. Stateless and scalabilty 199
14.3.2. API 200
14.3.3. HTTP Methods 201
14.3.4. Example of an API 202
14.4. WSDL example. 203
14.5. Conclusion 205
14.6. References 205

Chapter 15. SCALA – Functional Programming 207
Florent DEVIN

15.1. Introduction 207
15.1.1. Programming languages 208
15.1.2. Paradigm 208
15.2. Functional programming. 212
15.2.1. Introduction 212
15.2.2. Why now? 212
15.2.3. High order function 213
15.2.4. Basic functional blocks 215
15.3. Scala 217
15.3.1. Types systems 218
15.3.2. Basic manipulation of collection 222
15.4. Rational 224
15.5. Why immutability matters? 224
15.6. Conclusion 226
15.7. References 227

Chapter 16. Spark and Machine Learning Library 229
Yannick LE NIR

16.1. Introduction 229
16.2. Spark 230
16.2.1. Spark introduction 230

16.2.2. RDD presentation 230
16.2.3. RDD lifecycle 231
16.2.4. Operations on RDD 232
16.2.5. Exercises for environmental sciences 236
16.3. Spark machine learning library 237
16.3.1. Local vectors 237
16.3.2. Labeled points 237
16.3.3. Learning dataset 238
16.3.4. Classification and regression algorithms in Spark 238
16.3.5. Exercises for environmental sciences 239
16.4. Conclusion 242

Chapter 17. Database for Cloud Computing 245
Peio LOUBIÈRE

17.1. Introduction 245
17.2. From myGlsrdbms to NoSQL 245
17.2.1. CAP theorem 246
17.2.2. From ACID to BASE 247
17.3. NoSQL database storage paradigms 248
17.3.1. Column-family oriented storage 249
17.3.2. Key/value-oriented storage 249
17.3.3. Document-oriented storage 250
17.3.4. Graph-oriented storage 251
17.4. SQL versus NoSQL, the war will not take place 251
17.5. Example: a dive into MongoDB 252
17.5.1. Presentation 253
17.5.2. First steps 254
17.5.3. Database level commands 254
17.5.4. Data types 255
17.5.5. Modifying data 255
17.6. Conclusion 273
17.7. References 273

Chapter 18. WRF Performance Analysis and Scalability on Multicore High Performance Computing Systems 275
Didin Agustian PERMADI, Sebastiano Fabio SCHIFANO, Thi Kim Oanh NGUYEN, Nhat Ha Chi NGUYEN, Eleonora LUPPI and Luca TOMASSETTI

18.1. Introduction 276
18.2. The weather research and forecast model and experimental set-up 276
18.2.1. Model architecture 276

18.3. Architecture of multicore HPC system 282
18.4. Results. 283
18.4.1. Results of experiment E1 283
18.4.2. Results of experiment E2 286
18.5. Conclusion 288
18.6. References 288

List of Authors 291

Index. 293

Summaries of other volumes 295