Table of Contents

Preface xi

Chapter 1. Advanced Mapping of Environmental Data: Introduction 1
M. KANEVSKI

1.1. Introduction 1
1.2. Environmental data analysis: problems and methodology 3
1.2.1. Spatial data analysis: typical problems 3
1.2.2. Spatial data analysis: methodology 5
1.2.3. Model assessment and model selection 8
1.3. Resources 12
1.3.1. Books, tutorials 12
1.3.2. Software 12
1.4. Conclusion 14
1.5. References 15

Chapter 2. Environmental Monitoring Network Characterization and Clustering 19
D. TUIA and M. KANEVSKI

2.1. Introduction 19
2.2. Spatial clustering and its consequences 20
2.2.1. Global parameters 21
2.2.2. Spatial predictions 22
2.3. Monitoring network quantification 23
2.3.1. Topological quantification 23
2.3.2. Global measures of clustering 23
2.3.2.1. Topological indices 23
2.3.2.2. Statistical indices 24
2.3.3. Dimensional resolution: fractal measures of clustering 26
2.3.3.1. Sandbox method 27

2.3.3.2. Box-counting method 30
2.3.3.3. Lacunarity 33
2.4. Validity domains 34
2.5. Indoor radon in Switzerland: an example of a real monitoring network . . 36
2.5.1. Validity domains. 37
2.5.2. Topological index 37
2.5.3. Statistical indices 38
2.5.3.1. Morisita index 38
2.5.3.2. *K*-function 39
2.5.4. Fractal dimension 40
2.5.4.1. Sandbox and box-counting fractal dimension 40
2.5.4.2. Lacunarity 42
2.6. Conclusion 43
2.7. References 44

Chapter 3. Geostatistics: Spatial Predictions and Simulations 47
E. SAVELIEVA, V. DEMYANOV and M. MAIGNAN

3.1. Assumptions of geostatistics 47
3.2. Family of kriging models 49
3.2.1. Simple kriging 50
3.2.2. Ordinary kriging 50
3.2.3. Basic features of kriging estimation 51
3.2.4. Universal kriging (kriging with trend) 56
3.2.5. Lognormal kriging. 56
3.3. Family of co-kriging models 58
3.3.1. Kriging with linear regression 58
3.3.2. Kriging with external drift 58
3.3.3. Co-kriging 59
3.3.4. Collocated co-kriging 60
3.3.5. Co-kriging application example 61
3.4. Probability mapping with indicator kriging. 64
3.4.1. Indicator coding 64
3.4.2. Indicator kriging 66
3.4.3. Indicator kriging applications 69
3.4.3.1. Indicator kriging for ^{241}Am analysis 69
3.4.3.2. Indicator kriging for aquifer layer zonation 71
3.4.3.3. Indicator kriging for localization of crab crowds 74
3.5. Description of spatial uncertainty with conditional stochastic simulations 76
3.5.1. Simulation vs. estimation. 76
3.5.2. Stochastic simulation algorithms 77
3.5.3. Sequential Gaussian simulation 81
3.5.4. Sequential indicator simulations 84

3.5.5. Co-simulations of correlated variables 88
3.6. References 92

Chapter 4. Spatial Data Analysis and Mapping Using Machine Learning Algorithms 95
F. RATLE, A. POZDNOUKHOV, V. DEMYANOV, V. TIMONIN and E. SAVELIEVA

4.1. Introduction 95
4.2. Machine learning: an overview 96
4.2.1. The three learning problems 96
4.2.2. Approaches to learning from data 100
4.2.3. Feature selection 101
4.2.4. Model selection 103
4.2.5. Dealing with uncertainties 107
4.3. Nearest neighbor methods 108
4.4. Artificial neural network algorithms 109
4.4.1. Multi-layer perceptron neural network 109
4.4.2. General Regression Neural Networks 119
4.4.3. Probabilistic Neural Networks 122
4.4.4. Self-organizing (Kohonen) maps 124
4.5. Statistical learning theory for spatial data: concepts and examples 131
4.5.1. VC dimension and structural risk minimization 131
4.5.2. Kernels 132
4.5.3. Support vector machines 133
4.5.4. Support vector regression 137
4.5.5. Unsupervised techniques 141
4.5.5.1. Clustering 142
4.5.5.2. Nonlinear dimensionality reduction 144
4.6. Conclusion 146
4.7. References 146

Chapter 5. Advanced Mapping of Environmental Spatial Data: Case Studies 149
L. FORESTI, A. POZDNOUKHOV, M. KANEVSKI, V. TIMONIN, E. SAVELIEVA, C. KAISER, R. TAPIA and R. PURVES

5.1. Introduction 149
5.2. Air temperature modeling with machine learning algorithms and geostatistics 150
5.2.1. Mean monthly temperature 151
5.2.1.1. Data description 151
5.2.1.2. Variography 152
5.2.1.3. Step-by-step modeling using a neural network 153

5.2.1.4. Overfitting and undertraining 154
5.2.1.5. Mean monthly air temperature prediction mapping 156
5.2.2. Instant temperatures with regionalized linear dependencies 159
5.2.2.1. The Föhn phenomenon 159
5.2.2.2. Modeling of instant air temperature influenced by Föhn 160
5.2.3. Instant temperatures with nonlinear dependencies 163
5.2.3.1. Temperature inversion phenomenon 163
5.2.3.2. Terrain feature extraction using Support Vector Machines 164
5.2.3.3. Temperature inversion modeling with MLP 165
5.3. Modeling of precipitation with machine learning and geostatistics. 168
5.3.1. Mean monthly precipitation 169
5.3.1.1. Data description 169
5.3.1.2. Precipitation modeling with MLP 171
5.3.2. Modeling daily precipitation with MLP 173
5.3.2.1. Data description 173
5.3.2.2. Practical issues of MLP modeling 174
5.3.2.3. The use of elevation and analysis of the results 177
5.3.3. Hybrid models: NNRK and NNRS 179
5.3.3.1. Neural network residual kriging 179
5.3.3.2. Neural network residual simulations 182
5.3.4. Conclusions. 184
5.4. Automatic mapping and classification of spatial data using machine learning 185
5.4.1. k-nearest neighbor algorithm 185
5.4.1.1. Number of neighbors with cross-validation 187
5.4.2. Automatic mapping of spatial data 187
5.4.2.1. KNN modeling 188
5.4.2.2. GRNN modeling 190
5.4.3. Automatic classification of spatial data 192
5.4.3.1. KNN classification 193
5.4.3.2. PNN classification 194
5.4.3.3. Indicator kriging classification 197
5.4.4. Automatic mapping – conclusions 199
5.5. Self-organizing maps for spatial data – case studies 200
5.5.1. SOM analysis of sediment contamination 200
5.5.2. Mapping of socio-economic data with SOM 204
5.6. Indicator kriging and sequential Gaussian simulations for probability mapping. Indoor radon case study 209
5.6.1. Indoor radon measurements 209
5.6.2. Probability mapping. 211
5.6.3. Exploratory data analysis. 212
5.6.4. Radon data variography 216
5.6.4.1. Variogram for indicators 216

5.6.4.2. Variogram for Nscores 217
5.6.5. Neighborhood parameters 218
5.6.6. Prediction and probability maps. 219
5.6.6.1. Probability maps with IK 219
5.6.6.2. Probability maps with SGS 220
5.6.7. Analysis and validation of results. 221
5.6.7.1. Influence of the simulation net and the number of neighbors. . . 221
5.6.7.2. Decision maps and validation of results 222
5.6.8. Conclusions. 225
5.7. Natural hazards forecasting with support vector machines – case study: snow avalanches 225
5.7.1. Decision support systems for natural hazards. 227
5.7.2. Reminder on support vector machines 228
5.7.2.1. Probabilistic interpretation of SVM. 229
5.7.3. Implementing an SVM for avalanche forecasting 230
5.7.4. Temporal forecasts 230
5.7.4.1. Feature selection 231
5.7.4.2. Training the SVM classifier 232
5.7.4.3. Adapting SVM forecasts for decision support. 233
5.7.5. Extending the SVM to spatial avalanche predictions 237
5.7.5.1. Data preparation 237
5.7.5.2. Spatial avalanche forecasting. 239
5.7.6. Conclusions. 241
5.8. Conclusion 241
5.9. References 242

Chapter 6. Bayesian Maximum Entropy – BME 247
G. CHRISTAKOS

6.1. Conceptual framework 247
6.2. Technical review of BME 251
6.2.1. The spatiotemporal continuum 251
6.2.2. Separable metric structures 253
6.2.3. Composite metric structures 255
6.2.4. Fractal metric structures 256
6.3. Spatiotemporal random field theory 257
6.3.1. Pragmatic S/TRF tools 258
6.3.2. Space-time lag dependence: ordinary S/TRF 260
6.3.3. Fractal S/TRF 262
6.3.4. Space-time heterogenous dependence: generalized S/TRF 264
6.4. About BME 267
6.4.1. The fundamental equations 267
6.4.2. A methodological outline 273

6.4.3. Implementation of BME: the SEKS-GUI 275
6.5. A brief review of applications 281
6.5.1. Earth and atmospheric sciences 282
6.5.2. Health, human exposure and epidemiology 291
6.6. References 299

List of Authors 307

Index 309