Contents

List of Figures xi

List of Tables xiv

Preface to the first edition xvi

Preface to the second edition xviii

Preface to the third edition xx

1 Introduction 1
- 1.1 Epistemology 1
- 1.2 Paradigms ruling analysis 4

2 Patterns in vegetation ecology 7
- 2.1 Pattern recognition 7
- 2.2 Multivariate pattern analysis 11
- 2.3 Sampling for pattern recognition 13
  - 2.3.1 Getting a sample 13
  - 2.3.2 Organizing the data 15
- 2.4 Pattern recognition in $\mathcal{R}$ 18

3 Transformation 23
- 3.1 Data types 23
- 3.2 Scalar transformation and the species enigma 27
- 3.3 Vector transformation 29
- 3.4 Example: Transformation of plant cover data 33
- 3.5 Which transformation? 35

4 Multivariate comparison 37
- 4.1 Resemblance in multivariate space 37
- 4.2 Geometric approach 38
- 4.3 Contingency measures 43
- 4.4 Product moments 45

4.5 The resemblance matrix 49
4.6 Assessing the quality of classifications 50
4.7 Which resemblance function? 52

5 Classification **54**
5.1 The legacy of vegetation classification 54
5.2 Group structures 55
5.3 Agglomerative clustering 58
5.3.1 Linkage clustering 58
5.3.2 Average linkage clustering revisited 60
5.3.3 Minimum-variance clustering 62
5.4 Divisive clustering 64
5.5 Forming groups 66
5.6 Silhouette plot and fuzzy representation 69
5.7 Revising classifications 72
5.8 Which classification method? 75

6 Ordination **79**
6.1 Why ordination? 79
6.2 Principal component analysis 81
6.2.1 Operational steps 81
6.2.2 Interpretation by example 84
6.3 Principal coordinates analysis 89
6.4 Correspondence analysis 94
6.5 Heuristic ordination 97
6.5.1 The horseshoe or arch effect 97
6.5.2 Flexible shortest path adjustment 99
6.5.3 Nonmetric multidimensional scaling 101
6.5.4 Detrended correspondence analysis 103
6.6 How to interpret ordinations 104
6.7 Ranking by orthogonal components 108
6.7.1 RANK method 108
6.7.2 A sampling design based on RANK (example) 112
6.8 Which ordination method? 115

7 Ecological patterns **119**
7.1 Pattern and ecological response 119
7.2 Evaluating groups 120
7.2.1 Variance testing 121
7.2.2 Variance ranking 124
7.2.3 Ranking by indicator values 127
7.2.4 Analysis of concentration 129
7.3 Correlating spaces 133
7.3.1 The Mantel test 133

7.3.2 Correlograms 135
7.3.3 More trends: 'Schlaenggli' data revisited 138
7.4 Constrained ordination 142
7.5 Nonparametric multiple analysis of variance 148
7.5.1 Method and example 148
7.5.2 Data transformation revisited 152
7.5.3 Clustering revisited 153
7.6 Synoptic vegetation tables 154
7.6.1 The aim of ordering tables 154
7.6.2 Steps involved in sorting tables 155
7.6.3 Example: ordering Ellenberg's data 159

8 Traits and indicators 162
8.1 Vegetation beyond the species concept 162
8.2 Analytical framework 164
8.3 Matrix operations in a nutshell 166
8.4 Schlaenggli data example 171
8.4.1 Preparing data matrices 171
8.4.2 Deriving and projecting new data spaces 174
8.4.3 Measuring convergence 177
8.5 Rebuilding community ecology? 180

9 Static predictive modelling 183
9.1 Predictive or explanatory? 183
9.2 Evaluating environmental predictors 184
9.3 Generalized linear models 187
9.4 Generalized additive models 192
9.5 Classification and regression trees 194
9.6 Testing and building scenarios 197
9.7 Modelling vegetation types 200
9.8 Expected wetland vegetation (example) 204

10 Vegetation change in time 211
10.1 Coping with time 211
10.2 Temporal autocorrelation 212
10.3 Detecting trend 215
10.4 Rate of change 217
10.5 Early succession: Vraconnaz revisited 219
10.6 Markov models 222
10.6.1 Method an example 222
10.6.2 Limitations and practice 228
10.7 Space-for-time substitution 229
10.7.1 Principle and method 229
10.7.2 Swiss National Park succession (example) 232

10.8 Dynamics in pollen diagrams 236

11 Dynamic modelling **241**
11.1 Principles of systems 241
11.2 Simulating exponential growth 243
11.3 Logistic growth 244
11.4 The Lotka–Volterra model 247
11.5 Simulating space processes 251
11.6 Processes in the Swiss National Park 252
11.6.1 The temporal model 252
11.6.2 The spatial model 256

12 Revising classifications **261**
12.1 Beyond statistical analysis 261
12.2 Wetland data 263
12.3 Preprocessing data 264
12.3.1 Suppressing outliers 264
12.3.2 Selecting groups 265
12.4 Evaluating classification revisions 267
12.5 Carry-over nomenclature? 270
12.6 Step by step in $\mathcal{R}$ 273
12.7 Revising classification - or data? 276

13 Swiss forests: a case study **278**
13.1 Aim of the study 278
13.2 Structure of the data set 279
13.3 Selected questions 281
13.3.1 Is the similarity pattern discrete or continuous? 281
13.3.2 Is there a scale effect from plot size? 286
13.3.3 Which factors reflect vegetation pattern? 289
13.3.4 Is tree species distribution man-made? 293
13.3.5 Is the tree species pattern expected to change? 299
13.4 Conclusions 300

14 Back to the roots? **302**

Bibliography **307**

Appendix A: Functions in package dave **325**

Appendix B: Data sets used **326**

Index **327**

List of Figures

1.1 Principles of the filter model 6

2.1 Picture of a flower of a water lily (*Nymphaea alba*) 8
2.2 Vegetation mapping as a method for assessing a pattern . . . 9
2.3 Ordination of a typical horseshoe-shaped vegetation gradient 10
2.4 Spatial vegetation patterns 11
2.5 Impacts causing patterns 12
2.6 The elements of sampling design 15
2.7 Sampling plan of the 'Schlaenggli' data set 16

3.1 Scalar transformation of population size 28
3.2 Scalar transformations of the coordinates of a graph 29
3.3 Overlap of two species with Gaussian response 30

4.1 Presentation of data in the Euclidean space 39
4.2 Three ways of measuring distance 39
4.3 The correlation of vector j with vector k. 46
4.4 Similarities within and between the forest types of Switzerland 51

5.1 Two-dimensional group structures 56
5.2 A dendrogram from agglomerative hierarchical clustering . . 58
5.3 Comparing different methods of linkage clustering 59
5.4 Variance within and between groups 62
5.5 Cutting dendrograms derived by different methods 67
5.6 Silhouette plot example 69
5.7 Silhouette plot of four clustering solutions 70
5.8 Comparison of classification methods 76

6.1 Three-dimensional representation of similarity relationships . 80
6.2 Operations in PCA ordination 82
6.3 Interpretation of PCA results using real world data 85
6.4 Projection of five-dimensional PCA ordination 88
6.5 PCOA ordination using the 'Schlaenggli' data set 91
6.6 PCOA ordinations with six different resemblance measures . 93

6.7 Comparison of CA and PCA 96
6.8 Origin of the arch effect 98
6.9 Comparing PCOA and FSPA. 100
6.10 Comparison of PCOA and NMDS. 103
6.11 Comparison of CA and DCA. 104
6.12 Interpretations of CA. 105
6.13 Surface fitting to interpret ordinations. 106
6.14 Relevés chosen by RANK for permanent investigation 113
6.15 Ordinations of the **sveg** vegetation data set 116

7.1 Distinctness of group structure 121
7.2 Ordination of group structure in data set 'nveg' 132
7.3 Biplot and correlogram of 10 pH measurements 137
7.4 Projecting distances in different directions 139
7.5 Evaluating the direction of the main floristic gradient 139
7.6 Correlograms of site factors with vegetation 141
7.7 Comparison of RDA and CCA 146
7.8 Using distance matrices in NP-MANOVA 150
7.9 Performance of clustering methods 154
7.10 Graphical display of vegetation tables 158
7.11 Structuring the meadow data set of Ellenberg 161

8.1 Processing species-, traits- and indicator-based data 165
8.2 Histograms of 20 site factors in `ssit` 172
8.3 Species-, traits-, indicator- and site-based data as ordinations 174
8.4 Convergence of four spaces 178
8.5 PCA ordinations with trend surfaces 181

9.1 Pairwise plot of selected site variables 186
9.2 Linear and logistic regression of pH and *Sphagnum recurvum* 188
9.3 Occurrence of *Spagnum recurvum* and prediction by GLM . . 191
9.4 Prediction of *Spagnum recurvum* by GAM 193
9.5 Regression tree to predict *Spagnum recurvum* by pH 194
9.6 Predicting *Spagnum recurvum* by classification tree 195
9.7 Testing a model using independent variables 197
9.8 Scenarios for predicting *Spagnum recurvum* occurrence 199
9.9 Multivariate logistic regression 201
9.10 Simulated wetland vegetation 206
9.11 Occurrence probability of three species 208
9.12 Steps of computation in multinomial logistic regression 209

10.1 Type of environmental study needed to assess change 212
10.2 Temporal arrangement of measurements (pH) 213
10.3 Ordination of data from plots in the Swiss National Park . . 216
10.4 Measuring rate of change in time series of multi-state systems 218

10.5 Rate of change in plot Tr6, Swiss National Park 218
10.6 Ordination of data from bare peat plots of Vraconnaz 220
10.7 Boxplots of changes in vegetation in plots with bare peat . . 220
10.8 A Markov model of the Lippe *et al.* (1985) data set 226
10.9 PCA ordination of the Lippe succession data 227
10.10 Markov model of plot 15 of the Vraconnaz data 229
10.11 The principle of space-for-time substitution 230
10.12 The similarity of time series 231
10.13 *Pinus mugo* on a former pasture in the Swiss National Park 233
10.14 Minimum spanning tree (Swiss National Park) 234
10.15 Order of 59 time series from the Swiss National Park 235
10.16 Succession in pastures of the Swiss National Park 235
10.17 Tree species in a pollen diagram (Lotter 1999) 237
10.18 Velocity profile of the Soppensee pollen diagram 237
10.19 Time trajectory of the Soppensee pollen diagram 238
10.20 Velocity profiles of quantitative and qualitative content . . . 240

11.1 Attempt to get a dynamic model under control (Wildi 1976). 242
11.2 Numerical integration of the exponential growth equation . . 244
11.3 Simulating logistic growth 246
11.4 Abundance of lynx and snowshoe hare 247
11.5 Response of the Lotka–Volterra model 249
11.6 The mechanism of spatial exchange 252
11.7 Overgrowth of a plot by a new guild 253
11.8 Temporal succession in the Swiss National Park 255
11.9 Spatial design of SNP model 257
11.10 Spatial simulation of succession, Alp Stabelchod. 258

12.1 Frequency distribution of nearest-neighbour distances of relevés 264
12.2 Group sizes in the sample of wetland data 266
12.3 Comparison of the full versus the reduced wetland sample . . 266
12.4 F-values in environmental models based on classifications . . 269
12.5 Response of three classifications to elevation 270
12.6 Summary tables of three alternative classifications 271

13.1 Two ordinations of the Swiss forest data set 282
13.2 Vegetation map of Swiss forests (eight groups) 283
13.3 Boxplot of Swiss forest types (eight groups) 285
13.4 The effect of different plot size on similarity pattern. 289
13.5 Vegetation probability map (eight groups). 291
13.6 Observed and potential distribution of four tree species . . . 295
13.7 Ordination of forest stands. Four selected tree species marked 296
13.8 Ecograms of forest stands. Four selected tree species marked 297
13.9 Tree and herb layers of three species in ecological space . . . 299

List of Tables

2.1 Terms used in sampling design 14
2.2 Organization of vegetation and site data in $\mathcal{R}$ 17

3.1 Selected data types used for plant description 25
3.2 Effects of different vector transformations 30
3.3 Numerical examples of vector transformation 31
3.4 Transformation of cover-abundance values in phytosociology . 34

4.1 Notations in contingency tables 43
4.2 Resemblance measures using the notations in Table 4.1 44
4.3 Product moments 46
4.4 The average distance as a measure for homogeneity 49

5.1 Properties of four average linkage clustering methods 61
5.2 Reassigning plots to splinter groups in divisive clustering . . . 65
5.3 Data set illustrating the k-means algorithm. 74

6.1 Data set and results illustrating the RANK algorithm. 109
6.2 Ranking relevés of the 'Schlaenggli' data set. 113
6.3 Ranking species of the 'Schlaenggli' data set. 114

7.1 Synoptic table of `nveg` and `snit` 123
7.2 Variance ranking of species 125
7.3 Variance ranking of site factors 126
7.4 Ranking of species by indicator values 128
7.5 Mantel correlogram 138
7.6 Mantel test of the site factors 140
7.7 Storage location of parameters from functions `rda()` and `cca()` 147
7.8 Evaluating data transformation distance function 152
7.9 Steps involved in sorting synoptic tables 156
7.10 Frequency table of structured synoptic vegetation table . . . 159

8.1 Variable definitions (traits, indicators, site factors) 173
8.2 Mean values of site factors for plant traits 176
8.3 Mean values of indicator values for plant traits 176

9.1 Input and output data of multivariate logistic regression . . . 201
9.2 Group means and standard deviations of pH and average water level 206

10.1 Temporal autocorrelation in a time series 213
10.2 Markov process, measured and modelled data 223

11.1 The effect of time step length in numerical integration 244
11.2 Parameters of the Lotka–Volterra model 248
11.3 Initial values in the temporal model SNP 254
11.4 Six discrete vegetation states used as initial conditions 257

12.1 Site factors in data frame `wetsit` 264
12.2 Evaluation of classifications 268
12.3 Jancey's ranking applied to three classifications 272

13.1 Data sets used in Chapter 13 281
13.2 Composition of eight vegetation types. 282
13.3 Frequencies of tree species in data sets of different scale . . . 286
13.4 F-values of site factors based on eight forest vegetation types. 290
13.5 Multinomial models with different relevé plot size 293
13.6 Tree species frequencies in different vegetation layers 298