

Contents

Contributors.....	xix
1 Introduction	1
1.1 Part 1: Applied statistical theory.....	1
1.2 Part 2: The case studies.....	3
1.3 Data, software and flowcharts.....	6
2 Data management and software.....	7
2.1 Introduction.....	7
2.2 Data management	8
2.3 Data preparation.....	9
2.4 Statistical software.....	13
3 Advice for teachers.....	17
3.1 Introduction.....	17
4 Exploration.....	23
4.1 The first steps.....	24
4.2 Outliers, transformations and standardisations	38
4.3 A final thought on data exploration	47
5 Linear regression.....	49
5.1 Bivariate linear regression	49
5.2 Multiple linear regression	67
5.3 Partial linear regression	73
6 Generalised linear modelling.....	79
6.1 Poisson regression.....	79
6.2 Logistic regression.....	88
7 Additive and generalised additive modelling	97
7.1 Introduction.....	97
7.2 The additive model	101
7.3 Example of an additive model	102
7.4 Estimate the smoother and amount of smoothing	104
7.5 Additive models with multiple explanatory variables.....	108

7.6	Choosing the amount of smoothing	112
7.7	Model selection and validation	115
7.8	Generalised additive modelling	120
7.9	Where to go from here	124
8	Introduction to mixed modelling.....	125
8.1	Introduction	125
8.2	The random intercept and slope model.....	128
8.3	Model selection and validation	130
8.4	A bit of theory.....	135
8.5	Another mixed modelling example.....	137
8.6	Additive mixed modelling	140
9	Univariate tree models	143
9.1	Introduction	143
9.2	Pruning the tree.....	149
9.3	Classification trees.....	152
9.4	A detailed example: Ditch data.....	152
10	Measures of association.....	163
10.1	Introduction	163
10.2	Association between sites: Q analysis	164
10.3	Association among species: R analysis.....	171
10.4	Q and R analysis: concluding remarks.....	176
10.5	Hypothesis testing with measures of association	179
11	Ordination — First encounter.....	189
11.1	Bray–Curtis ordination	189
12	Principal component analysis and redundancy analysis.....	193
12.1	The underlying principle of PCA.....	193
12.2	PCA: Two easy explanations	194
12.3	PCA: Two technical explanations.....	196
12.4	Example of PCA	197
12.5	The biplot.....	200
12.6	General remarks.....	205
12.7	Chord and Hellinger transformations.....	206
12.8	Explanatory variables	208
12.9	Redundancy analysis	210
12.10	Partial RDA and variance partitioning.....	219
12.11	PCA regression to deal with collinearity	221
13	Correspondence analysis and canonical correspondence analysis.....	225
13.1	Gaussian regression and extensions.....	225
13.2	Three rationales for correspondence analysis	231
13.3	From RGR to CCA	238
13.4	Understanding the CCA triplot.....	240

13.5 When to use PCA, CA, RDA or CCA	242
13.6 Problems with CA and CCA.....	243
14 Introduction to discriminant analysis.....	245
14.1 Introduction.....	245
14.2 Assumptions	248
14.3 Example	250
14.4 The mathematics	254
14.5 The numerical output for the sparrow data	255
15 Principal coordinate analysis and non-metric multidimensional scaling 259	
15.1 Principal coordinate analysis	259
15.2 Non-metric multidimensional scaling.....	261
16 Time series analysis — Introduction.....	265
16.1 Using what we have already seen before	265
16.2 Auto-regressive integrated moving average models with exogenous variables.....	281
17 Common trends and sudden changes	289
17.1 Repeated LOESS smoothing.....	289
17.2 Identifying the seasonal component.....	293
17.3 Common trends: MAFA	299
17.4 Common trends: Dynamic factor analysis	303
17.5 Sudden changes: Chronological clustering	315
18 Analysis and modelling of lattice data	321
18.1 Lattice data.....	321
18.2 Numerical representation of the lattice structure	323
18.3 Spatial correlation	327
18.4 Modelling lattice data	331
18.5 More exotic models	334
18.6 Summary.....	338
19 Spatially continuous data analysis and modelling	341
19.1 Spatially continuous data	341
19.2 Geostatistical functions and assumptions	342
19.3 Exploratory variography analysis	346
19.4 Geostatistical modelling: Kriging	359
19.5 A full spatial analysis of the bird radar data	363
20 Univariate methods to analyse abundance of decapod larvae	373
20.1 Introduction.....	373
20.2 The data	374
20.3 Data exploration.....	377
20.4 Linear regression results	379

20.5 Additive modelling results.....	381
20.6 How many samples to take?	383
20.7 Discussion.....	385
21 Analysing presence and absence data for flatfish distribution in the Tagus estuary, Portugal	389
21.1 Introduction	389
21.2 Data and materials	390
21.3 Data exploration	392
21.4 Classification trees.....	395
21.5 Generalised additive modelling	397
21.6 Generalised linear modelling	398
21.7 Discussion.....	401
22 Crop pollination by honeybees in Argentina using additive mixed modelling.....	403
22.1 Introduction	403
22.2 Experimental setup	404
22.3 Abstracting the information	404
22.4 First steps of the analyses: Data exploration.....	407
22.5 Additive mixed modelling	408
22.6 Discussion and conclusions	414
23 Investigating the effects of rice farming on aquatic birds with mixed modelling.....	417
23.1 Introduction	417
23.2 The data	419
23.3 Getting familiar with the data: Exploration	420
23.4 Building a mixed model.....	424
23.5 The optimal model in terms of random components	427
23.6 Validating the optimal linear mixed model	430
23.7 More numerical output for the optimal model.....	431
23.8 Discussion.....	433
24 Classification trees and radar detection of birds for North Sea wind farms.....	435
24.1 Introduction	435
24.2 From radars to data	437
24.3 Classification trees.....	439
24.4 A tree for the birds	440
24.5 A tree for birds, clutter and more clutter.....	445
24.6 Discussion and conclusions	447
25 Fish stock identification through neural network analysis of parasite fauna.....	449
25.1 Introduction	449
25.2 Horse mackerel in the northeast Atlantic.....	450

25.3 Neural networks.....	452
25.4 Collection of data.....	455
25.5 Data exploration.....	456
25.6 Neural network results	457
25.7 Discussion.....	460
26 Monitoring for change: Using generalised least squares, non-metric multidimensional scaling, and the Mantel test on western Montana grasslands.....	463
26.1 Introduction.....	463
26.2 The data	464
26.3 Data exploration.....	467
26.4 Linear regression results	472
26.5 Generalised least squares results.....	476
26.6 Multivariate analysis results	479
26.7 Discussion.....	483
27 Univariate and multivariate analysis applied on a Dutch sandy beach community.....	485
27.1 Introduction.....	485
27.2 The variables.....	486
27.3 Analysing the data using univariate methods.....	487
27.4 Analysing the data using multivariate methods	494
27.5 Discussion and conclusions	499
28 Multivariate analyses of South-American zoobenthic species — spoil for choice	503
28.1 Introduction and the underlying questions.....	503
28.2 Study site and sample collection.....	504
28.3 Data exploration.....	506
28.4 The Mantel test approach.....	509
28.5 The transformation plus RDA approach	512
28.6 Discussion and conclusions	512
29 Principal component analysis applied to harbour porpoise fatty acid data	515
29.1 Introduction.....	515
29.2 The data	515
29.3 Principal component analysis	517
29.4 Data exploration.....	518
29.5 Principal component analysis results	518
29.6 Simpler alternatives to PCA.....	524
29.7 Discussion.....	526
30 Multivariate analyses of morphometric turtle data — size and shape	529
30.1 Introduction.....	529
30.2 The turtle data.....	530

30.3 Data exploration	531
30.4 Overview of classic approaches related to PCA	534
30.5 Applying PCA to the original turtle data	536
30.6 Classic morphometric data analysis approaches	537
30.7 A geometric morphometric approach	542
31 Redundancy analysis and additive modelling applied on savanna tree data	547
31.1 Introduction	547
31.2 Study area	548
31.3 Methods	548
31.4 Results	551
31.5 Discussion.....	559
32 Canonical correspondence analysis of lowland pasture vegetation in the humid tropics of Mexico	561
32.1 Introduction	561
32.2 The study area.....	562
32.3 The data	563
32.4 Data exploration	565
32.5 Canonical correspondence analysis results	568
32.6 African star grass	571
32.7 Discussion and conclusion.....	573
33 Estimating common trends in Portuguese fisheries landings	575
33.1 Introduction	575
33.2 The time series data	576
33.3 MAFA and DFA	579
33.4 MAFA results	580
33.5 DFA results	582
33.6 Discussion.....	587
34 Common trends in demersal communities on the Newfoundland-Labrador Shelf	589
34.1 Introduction	589
34.2 Data.....	590
34.3 Time series analysis	591
34.4 Discussion.....	598
35 Sea level change and salt marshes in the Wadden Sea: A time series analysis	601
35.1 Interaction between hydrodynamical and biological factors	601
35.2 The data	603
35.3 Data exploration	605
35.4 Additive mixed modelling	607
35.5 Additive mixed modelling results	610
35.6 Discussion.....	613

36 Time series analysis of Hawaiian waterbirds	615
36.1 Introduction.....	615
36.2 Endangered Hawaiian waterbirds	616
36.3 Data exploration.....	617
36.4 Three ways to estimate trends.....	619
36.5 Additive mixed modelling	626
36.6 Sudden breakpoints.....	630
36.7 Discussion.....	631
37 Spatial modelling of forest community features in the Volzhsko-Kamsky reserve	633
37.1 Introduction.....	633
37.2 Study area	635
37.3 Data exploration.....	636
37.4 Models of boreality without spatial auto-correlation	638
37.5 Models of boreality with spatial auto-correlation	640
37.6 Conclusion	646
References	649
Index	667