Contents

Preface xi

1 What are compositional data, and why are they special? 1

1.1 Examples of compositional data 1
1.2 Why are compositional data different from other types of data? 2
1.3 Basic terminology and notation in compositional data analysis 3
1.4 Basic principles of compositional data analysis 5
1.5 Ratios and logratios 6

2 Geometry and visualization of compositional data 9

2.1 Simple graphics 9
2.2 Geometry in a simplex 12
2.3 Moving out of the simplex 14
2.4 Distances between points in logratio space 15

3 Logratio transformations 17

3.1 Additive logratio transformations 17
3.2 Centred logratio transformations 18
3.3 Logratios incorporating amalgamations 19
3.4 Isometric logratio transformations 19
3.5 Comparison of logratios in practice 21
3.6 Practical interpretation of logratios 23

4 Properties and distributions of logratios 25

4.1 Lognormal distribution 25
4.2 Logit function 27
4.3 Additive logistic normal distribution 27
4.4 Logratio variances and covariances 28

4.5 Testing for multivariate normality 30
4.6 When logratios are not normal 31

5 Regression models involving compositional data 33

5.1 Visualizing ratios as a graph 33
5.2 Using simple logratios as predictors 34
5.3 Compositions as responses – total logratio variance 37
5.4 Redundancy analysis 39

6 Dimension reduction using logratio analysis 41

6.1 Weighted principal component analysis 41
6.2 Logratio analysis 42
6.3 Different biplot scaling options 44
6.4 Constrained compositional biplots 46

7 Clustering of compositional data 49

7.1 Logratio distances between rows and between columns 49
7.2 Clustering based on logratio distances 50
7.3 Weighted Ward clustering 50
7.4 Isometric logratio versus amalgamation balances 53

8 Problem of zeros, with some solutions 57

8.1 Zero replacement 57
8.2 Sensitivity to zero replacement 58
8.3 Subcompositional incoherence 59
8.4 Correspondence analysis alternative 60

9 Simplifying the task: variable selection 65

9.1 Explaining total logratio variance 65
9.2 Stepwise selection of logratios 66
9.3 Parsimonious variable selection 68
9.4 Amalgamation logratios as variables for selection 70
9.5 Signal and noise in compositional data 70

10 Case study: Fatty acids of marine amphipods 73

10.1 Introduction 73
10.2 Material and methods 74
10.3 Results 75
10.4 Discussion and conclusion 80

A Appendix: Theory of compositional data analysis 81

A.1 Basic notation 81
A.2 Ratios and logratios 82
A.3 Logratio distance 85
A.4 Logratio variance 85
A.5 Logratio analysis (LRA) 86
A.6 Principal component analysis (PCA) 87
A.7 Procrustes analysis 88
A.8 Constrained logratio analysis and redundancy analysis 89
A.9 Permutation tests 89
A.10 Weighted Ward clustering 90

B Appendix: Bibliography of compositional data analysis 91

B.1 Books 91
B.2 Articles 92
B.3 Web resources 95

C Appendix: Computation of compositional data analysis 97

C.1 Simple graphics for compositional data 97
C.2 Logratio transformations 98
C.3 Compositional data modelling 102
C.4 Compositional data analytics 103

D Appendix: Glossary of terms 111

E Appendix: Epilogue 115

Index 119