Contents

Preface xvii

List of Contributors xix

Part I Introduction 1

1 A Short History of Compositional Data Analysis 3
John Bacon-Shone

1.1 Introduction 3
1.2 Spurious Correlation 3
1.3 Log and Log-Ratio Transforms 4
1.4 Subcompositional Dependence 5
1.5 alr, clr, ilr: Which Transformation to Choose? 5
1.6 Principles, Perturbations and Back to the Simplex 6
1.7 Biplots and Singular Value Decompositions 7
1.8 Mixtures 7
1.9 Discrete Compositions 8
1.10 Compositional Processes 8
1.11 Structural, Counting and Rounded Zeros 8
1.12 Conclusion 9
Acknowledgement 9
References 9

2 Basic Concepts and Procedures 12
Juan José Egozcue and Vera Pawlowsky-Glahn

2.1 Introduction 12
2.2 Election Data and Raw Analysis 13
2.3 The Compositional Alternative 15
2.3.1 Scale Invariance: Vectors with Proportional Positive Components Represent the Same Composition 15
2.3.2 Subcompositional Coherence: Analyses Concerning a Subset of Parts Must Not Depend on Other Non-Involved Parts 16

2.3.3 Permutation Invariance: The Conclusions of a Compositional Analysis Should Not Depend on the Order of the Parts 17
2.4 Geometric Settings 17
2.5 Centre and Variability 22
2.6 Conclusion 27
Acknowledgements 27
References 27

Part II Theory – Statistical Modelling 29

3 The Principle of Working on Coordinates 31
Glòria Mateu-Figueras, Vera Pawlowsky-Glahn and Juan José Egozcue
3.1 Introduction 31
3.2 The Role of Coordinates in Statistics 32
3.3 The Simplex 33
3.3.1 Basis of the Simplex 34
3.3.2 Working on Orthonormal Coordinates 35
3.4 *Move* or *Stay* in the Simplex 38
3.5 Conclusions 40
Acknowledgements 41
References 41

4 Dealing with Zeros 43
Josep Antoni Martín-Fernández, Javier Palarea-Albaladejo and Ricardo Antonio Olea
4.1 Introduction 43
4.2 Rounded Zeros 44
4.2.1 Non-Parametric Replacement of Rounded Zeros 45
4.2.2 Parametric Modified EM Algorithm for Rounded Zeros 47
4.3 Count Zeros 50
4.4 Essential Zeros 53
4.5 Difficulties, Troubles and Challenges 55
Acknowledgements 57
References 57

5 Robust Statistical Analysis 59
Peter Filzmoser and Karel Hron
5.1 Introduction 59
5.2 Elements of Robust Statistics from a Compositional Point of View 60
5.3 Robust Methods for Compositional Data 63
5.3.1 Multivariate Outlier Detection 64
5.3.2 Principal Component Analysis 64
5.3.3 Discriminant Analysis 65
5.4 Case Studies 66
5.4.1 Multivariate Outlier Detection 66
5.4.2 Principal Component Analysis 68
5.4.3 Discriminant Analysis 68
5.5 Summary 70

Acknowledgement 71
References 71

6 Geostatistics for Compositions 73
Raimon Tolosana-Delgado, Karl Gerald van den Boogaart and Vera Pawlowsky-Glahn

6.1 Introduction 73
6.2 A Brief Summary of Geostatistics 74
6.3 Cokriging of Regionalised Compositions 76
6.4 Structural Analysis of Regionalised Composition 76
6.5 Dealing with Zeros: Replacement Strategies and Simplicial Indicator Cokriging 78
6.6 Application 79
6.6.1 Delimiting the Body: Simplicial Indicator Kriging 81
6.6.2 Interpolating the Oil–Brine–Solid Content 82
6.7 Conclusions 84
Acknowledgements 84
References 84

7 Compositional VARIMA Time Series 87
Carles Barceló-Vidal, Lucía Aguilar and Josep Antoni Martín-Fernández

7.1 Introduction 87
7.2 The Simplex $\mathcal{S}^D$ as a Compositional Space 89
7.2.1 Basic Concepts and Notation 89
7.2.2 The Covariance Structure on the Simplex 90
7.3 Compositional Time Series Models 91
7.3.1 $\mathcal{C}$-Stationary Processes 92
7.3.2 $\mathcal{C}$-VARIMA Processes 93
7.4 CTS Modelling: An Example 94
7.4.1 Expenditure Shares in the UK 94
7.4.2 Model Selection 95
7.4.3 Estimation of Parameters 96
7.4.4 Interpretation and Comparison 96
7.5 Discussion 99
Acknowledgements 99
References 100
Appendix 102

8 Compositional Data and Correspondence Analysis 104
Michael Greenacre

8.1 Introduction 104
8.2 Comparative Technical Definitions 105
8.3 Properties and Interpretation of LRA and CA 107
8.4 Application to Fatty Acid Compositional Data 107
8.5 Discussion and Conclusions 111
Acknowledgements 112
References 112

9 Use of Survey Weights for the Analysis of Compositional Data 114
Monique Graf

9.1 Introduction 114
9.2 Elements of Survey Design 115
9.2.1 Randomization 115
9.2.2 Design-Based Estimation 118
9.3 Application to Compositional Data 122
9.3.1 Weighted Arithmetic and Geometric Means 123
9.3.2 Closed Arithmetic Mean of Amounts 123
9.3.3 Centred Log-Ratio of the Geometric Mean Composition 124
9.3.4 Closed Geometric Mean Composition 124
9.3.5 Example: Swiss Earnings Structure Survey (SESS) 125
9.4 Discussion 126
References 126

10 Notes on the Scaled Dirichlet Distribution 128
Gianna Serafina Monti, Glòria Mateu-Figueras and Vera Pawlowsky-Glahn

10.1 Introduction 128
10.2 Genesis of the Scaled Dirichlet Distribution 129
10.3 Properties of the Scaled Dirichlet Distribution 131
10.3.1 Graphical Comparison 131
10.3.2 Membership in the Exponential Family 133
10.3.3 Measures of Location and Variability 134
10.4 Conclusions 136
Acknowledgements 137
References 137

Part III Theory – Algebra and Calculus 139

11 Elements of Simplicial Linear Algebra and Geometry 141
Juan José Egozcue, Carles Barceló-Vidal, Josep Antoni Martín-Fernández, Eusebi Jarauta-Bragulat, José Luis Díaz-Barrero and Glòria Mateu-Figueras

11.1 Introduction 141
11.2 Elements of Simplicial Geometry 142
11.2.1 n-Part Simplex 142
11.2.2 Vector Space 143
11.2.3 Centred Log-Ratio Representation 146
11.2.4 Metrics 147
11.2.5 Orthonormal Basis and Coordinates 149
11.3 Linear Functions 151
11.3.1 Linear Functions Defined on the Simplex 152
11.3.2 Simplicial Linear Function Defined on a Real Space 153
11.3.3 Simplicial Linear Function Defined on the Simplex 154
11.4 Conclusions 156
Acknowledgements 156
References 156

12 Calculus of Simplex-Valued Functions 158
Juan José Egozcue, Eusebi Jarauta-Bragulat and José Luis Díaz-Barrero

12.1 Introduction 158
12.2 Limits, Continuity and Differentiability 161
12.2.1 Limits and Continuity 161
12.2.2 Differentiability 163
12.2.3 Higher Order Derivatives 169
12.3 Integration 171
12.3.1 Antiderivatives. Indefinite Integral 171
12.3.2 Integration of Continuous SV Functions 172
12.4 Conclusions 174
Acknowledgements 175
References 175

13 Compositional Differential Calculus on the Simplex 176
Carles Barceló-Vidal, Josep Antoni Martín-Fernández and Glòria Mateu-Figueras

13.1 Introduction 176
13.2 Vector-Valued Functions on the Simplex 177
13.2.1 Scale-Invariant Vector-Valued Functions on $\mathbb{R}^n_+$ 177
13.2.2 Vector-Valued Functions on $\mathcal{S}^n$ 178
13.3 $\mathcal{C}$-Derivatives on the Simplex 178
13.3.1 Derivative of a Scale-Invariant Vector-Valued Function on $\mathbb{R}^n_+$ 178
13.3.2 Directional $\mathcal{C}$-Derivatives 180
13.3.3 $\mathcal{C}$-Derivative 182
13.3.4 $\mathcal{C}$-Gradient 184
13.3.5 Critical Points of a $\mathcal{C}$-Differentiable Real-Valued Function on $\mathcal{S}^n$ 184
13.4 Example: Experiments with Mixtures 185
13.4.1 Polynomial of Degree One 185
13.4.2 Polynomial of Degree Two 186
13.4.3 Polynomial of Degree One in Logarithms 187
13.4.4 A numerical Example 188
13.5 Discussion 189
Acknowledgements 190
References 190

Part IV Applications 191

14 Proportions, Percentages, PPM: Do the Molecular Biosciences Treat Compositional Data Right? 193
David Lovell, Warren Müller, Jen Taylor, Alec Zwart and Chris Helliwell

14.1 Introduction 193
14.2 The Omics Imp and Two Bioscience Experiment Paradigms 194
14.3 The Impact of Compositional Constraints in the Omics 197
14.3.1 Univariate Impact of Compositional Constraints 197
14.3.2 Impact of Compositional Constraints on Multivariate Distance Metrics 199

14.4 Impact of Compositional Constraints on Correlation and Covariance 201
14.4.1 Compositional Constraints, Covariance, Correlation and Log-Transformed Data 202
14.4.2 A Simulation Approach to Understanding the Impact of Closure 202
14.5 Implications 204
14.5.1 Gathering Information to Infer Absolute Abundance 204
14.5.2 Analysing Compositional Omics Data Appropriately 205
Acknowledgements 206
References 206

15 Hardy–Weinberg Equilibrium: A Nonparametric Compositional Approach 208
Jan Graffelman and Juan José Egozcue
15.1 Introduction 208
15.2 Genetic Data Sets 209
15.3 Classical Tests for HWE 210
15.4 A Compositional Approach 210
15.5 Example 214
15.6 Conclusion and Discussion 215
Acknowledgements 215
References 215

16 Compositional Analysis in Behavioural and Evolutionary Ecology 218
Michele Edoardo Raffaele Pierotti and Josep Antoni Martín-Fernández
16.1 Introduction 218
16.2 CODA in Population Genetics 219
16.3 CODA in Habitat Choice 222
16.4 Multiple Choice and Individual Variation in Preferences 224
16.5 Ecological Specialization 228
16.6 Time Budgets: More on Specialization 229
16.7 Conclusions 231
Acknowledgements 231
References 231

17 Flying in Compositional Morphospaces: Evolution of Limb Proportions in Flying Vertebrates 235
Luis Azevedo Rodrigues, Josep Daunis-i-Estadella, Glòria Mateu-Figueras and Santiago Thió-Henestrosa
17.1 Introduction 235
17.2 Flying Vertebrates – General Anatomical and Functional Characteristics 236
17.3 Materials 236
17.4 Methods 238
17.5 Aitchison Distance Disparity Metrics 239
17.5.1 Intragroup Aitchison Distance 239
17.5.2 Intergroup Aitchison Distance 240
17.6 Statistical Tests 243

17.7 Biplots 244
17.7.1 Chiroptera 244
17.7.2 Pterosauria 245
17.8 Balances 246
17.9 Size Effect 249
17.10 Final Remarks 249
17.10.1 All Groups 250
17.10.2 Aves 250
17.10.3 Pterosauria 250
17.10.4 Chiroptera 251
Acknowledgements 252
References 252

18 Natural Laws Governing the Distribution of the Elements in Geochemistry: The Role of the Log-Ratio Approach 255
Antonella Buccianti

18.1 Introduction 255
18.2 Geochemical Processes and Log-Ratio Approach 256
18.3 Log-Ratio Approach and Water Chemistry 258
18.4 Log-Ratio Approach and Volcanic Gas Chemistry 261
18.5 Log-Ratio Approach and Subducting Sediment Composition 263
18.6 Conclusions 265
Acknowledgements 265
References 265

19 Compositional Data Analysis in Planetology: The Surfaces of Mars and Mercury 267
Helmut Lammer, Peter Wurz, Josep Antoni Martín-Fernández and Herbert Iwo Maria Lichtenegger

19.1 Introduction 267
19.1.1 Mars 267
19.1.2 Mercury 269
19.1.3 Analysis of Surface Composition 270
19.2 Compositional Analysis of Mars' Surface 270
19.3 Compositional Analysis of Mercury's Surface 274
19.4 Conclusion 278
Acknowledgement 278
References 278

20 Spectral Analysis of Compositional Data in Cyclostratigraphy 282
Eulogio Pardo-Igúzquiza and Javier Heredia

20.1 Introduction 282
20.2 The Method 283
20.3 Case Study 285
20.4 Discussion 287
20.5 Conclusions 288
Acknowledgement 288
References 288

21 Multivariate Geochemical Data Analysis in Physical Geography 290
Jennifer McKinley and Christopher David Lloyd

21.1 Introduction 290
21.2 Context 291
21.3 Data 293
21.4 Analysis 295
21.5 Discussion 299
21.6 Conclusion 300
Acknowledgement 300
References 300

22 Combining Isotopic and Compositional Data: A Discrimination of Regions Prone to Nitrate Pollution 302
Roger Puig, Raimon Tolosana-Delgado, Neus Otero and Albert Folch

22.1 Introduction 302
22.2 Study Area 303
22.2.1 Maresme 304
22.2.2 Osona 305
22.2.3 Lluçanès 305
22.2.4 Empordà 306
22.2.5 Selva 306
22.3 Analytical Methods 306
22.4 Statistical Treatment 307
22.4.1 Data Scaling 307
22.4.2 Linear Discriminant Analysis 309
22.4.3 Discriminant Biplots 310
22.5 Results and Discussion 311
22.6 Conclusions 314
Acknowledgements 315
References 315

23 Applications in Economics 318
Tim Fry

23.1 Introduction 318
23.2 Consumer Demand Systems 319
23.3 Miscellaneous Applications 322
23.4 Compositional Time Series 323
23.5 New Directions 323
23.6 Conclusion 325
References 325

Part V Software 327

24 Exploratory Analysis Using CoDaPack 3D 329
Santiago Thió-Henestrosa and Josep Daunis-i-Estadella

24.1 CoDaPack 3D Description 329
24.2 Data Set Description 331

24.3 Exploratory Analysis 333
24.3.1 Numerical Analysis 333
24.3.2 Biplot 334
24.3.3 The Ternary Diagram 335
24.3.4 Principal Component Analysis 336
24.3.5 Balance-Dendrogram 336
24.3.6 By Groups Description 338
24.4 Summary and Conclusions 339
Acknowledgements 340
References 340

25 robCompositions: An R-package for Robust Statistical Analysis of Compositional Data 341
Matthias Templ, Karel Hron and Peter Filzmoser

25.1 General Information on the R-package `robCompositions` 341
25.1.1 Data Sets Included in the Package 342
25.1.2 Design Principles 343
25.2 Expressing Compositional Data in Coordinates 343
25.3 Multivariate Statistical Methods for Compositional Data Containing Outliers 345
25.3.1 Multivariate Outlier Detection 345
25.3.2 Principal Component Analysis and the Robust Compositional Biplot 347
25.3.3 Discriminant Analysis 350
25.4 Robust Imputation of Missing Values 351
25.5 Summary 354
References 354

26 Linear Models with Compositions in R 356
Raimon Tolosana-Delgado and Karl Gerald van den Boogaart

26.1 Introduction 356
26.2 The Illustration Data Set 357
26.2.1 The Data 357
26.2.2 Descriptive Analysis of Compositional Characteristics 358
26.3 Explanatory Binary Variable 360
26.4 Explanatory Categorical Variable 363
26.5 Explanatory Continuous Variable 365
26.6 Explanatory Composition 367
26.7 Conclusions 370
Acknowledgement 371
References 371

Index **373**