

Multivariate Analysis of Ecological Data with ade4

Jean Thioulouse • Stéphane Dray
Anne-Béatrice Dufour • Aurélie Siberchicot
Thibaut Jombart • Sandrine Pavoine

Multivariate Analysis of Ecological Data with ade4

 Springer

Jean Thioulouse
Laboratoire de Biométrie
et Biologie Evolutive
CNRS UMR 5558 – Université de Lyon
Villeurbanne, France

Stéphane Dray
Laboratoire de Biométrie
et Biologie Evolutive
CNRS UMR 5558 – Université de Lyon
Villeurbanne, France

Anne-Béatrice Dufour
Laboratoire de Biométrie
et Biologie Evolutive
CNRS UMR 5558 – Université de Lyon
Villeurbanne, France

Aurélie Siberchicot
Laboratoire de Biométrie
et Biologie Evolutive
CNRS UMR 5558 – Université de Lyon
Villeurbanne, France

Thibaut Jombart
Department of Infectious
Disease Epidemiology
London School of Hygiene
and Tropical Medicine
London, UK

Sandrine Pavoine
Centre d'Ecologie et des Sciences
de la Conservation (CESCO)
Muséum national d'Histoire naturelle,
CNRS, Sorbonne Université
Paris, France

ISBN 978-1-4939-8848-8 ISBN 978-1-4939-8850-1 (eBook)
<https://doi.org/10.1007/978-1-4939-8850-1>

Library of Congress Control Number: 2018952909

© Springer Science+Business Media, LLC, part of Springer Nature 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Science+Business Media, LLC part of Springer Nature.

The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Foreword

« Les programmes sont des objets scientifiques bizarres : les uns y cachent la compréhension mathématique des modèles qui les supportent, les autres en font des objets expérimentaux. Lieu par excellence d'échanges et de conflits, d'appropriation souhaitable ou abusive, produit sans auteur présumé pour les camelots de la démonstration (lesquels programment rarement) ou objet largement surestimé, sa valeur dépend du moment et de l'environnement. Il faut concilier deux logiques, celle de l'utilisateur et celle du statisticien. Notons à ce propos qu'on peut militer pour la libre circulation des programmes ou (exclusif) des données : il faut rassurer tout le monde. Image d'une méthode pour celui qui l'écrit, le programme change de nature pour celui qui l'emploie, image d'une problématique pour celui qui l'acquiert, les données changent de nature quand elles servent d'illustration. La libre circulation des données et des programmes est un facteur décisif du développement : une seule chose est inconcevable, c'est qu'il n'y ait qu'un seul point de vue sur ces objets. »

Daniel Chessel, 1992

Contents

| | | |
|----------|---|----|
| 1 | Introduction | 1 |
| 1.1 | Intended Readership | 1 |
| 1.2 | Evolutions of ade4 | 2 |
| 1.2.1 | The ade4 Add-On Package for R | 2 |
| 1.2.2 | Previous Versions of ade4 | 3 |
| 1.3 | Using ade4 | 4 |
| 1.3.1 | Computer Hardware | 4 |
| 1.3.2 | Installing R | 4 |
| 1.3.3 | Installing ade4 | 4 |
| 1.3.4 | Dependencies | 5 |
| 1.3.5 | Packages of the ade4 Family | 5 |
| 1.3.6 | Version of the Packages Used in This Book | 7 |
| 1.3.7 | The adelist Forum | 8 |
| 1.3.8 | Using the Help System | 9 |
| 1.4 | Interactive Code Snippets | 9 |
| 1.5 | Ecological Data Sets | 9 |
| 2 | Useful R Functions and Data Structures | 13 |
| 2.1 | Introduction | 13 |
| 2.2 | Basic Data Import and Export Functions | 14 |
| 2.2.1 | read.table | 14 |
| 2.2.2 | write.table | 15 |
| 2.2.3 | Data | 16 |
| 2.3 | Vectors | 16 |
| 2.4 | Data Frames | 17 |
| 2.4.1 | Dimensions | 18 |
| 2.4.2 | Row and Column Names | 18 |
| 2.4.3 | Accessing Data Frame Elements | 19 |
| 2.4.4 | Row and Column Sums and Means | 20 |
| 2.4.5 | Row and Column Selection | 21 |

| | | |
|----------|---|-----------|
| 2.4.6 | Changing Values | 21 |
| 2.4.7 | Missing Values in Data Frames | 22 |
| 2.4.8 | Data Transformation | 24 |
| 2.4.9 | Apply | 24 |
| 2.4.10 | Summary | 24 |
| 2.4.11 | Other Functions | 25 |
| 2.5 | Factors | 25 |
| 2.5.1 | Using Factors | 26 |
| 2.5.2 | Generating Factors | 27 |
| 2.5.3 | Re-ordering Levels | 28 |
| 3 | The <code>dudi</code> Class | 29 |
| 3.1 | Introduction | 29 |
| 3.2 | Principles of Multivariate Analysis | 30 |
| 3.3 | Structure | 33 |
| 3.4 | Functions | 36 |
| 3.5 | Elements of <code>dudi</code> Objects | 38 |
| 3.5.1 | <code>pca1\$tab</code> | 38 |
| 3.5.2 | <code>pca1\$cw</code> and <code>pca1\$lw</code> | 39 |
| 3.5.3 | <code>pca1\$eig</code> , <code>pca1\$rank</code> and <code>pca1\$nf</code> | 40 |
| 3.5.4 | <code>pca1\$c1</code> , <code>pca1\$l1</code> , <code>pca1\$co</code> and <code>pca1\$li</code> | 40 |
| 3.5.5 | <code>pca1\$cent</code> and <code>pca1\$norm</code> | 42 |
| 3.6 | Using <code>dudi</code> Objects | 42 |
| 3.6.1 | The <code>print</code> and <code>summary</code> Functions | 43 |
| 3.6.2 | The <code>scatter</code> and <code>biplot</code> Functions | 44 |
| 3.6.3 | The <code>score</code> Function | 44 |
| 3.6.4 | The <code>s.label</code> and <code>plot</code> Functions | 46 |
| 3.6.5 | The <code>inertia</code> Function | 47 |
| 3.6.6 | Other Graphical Functions | 50 |
| 3.7 | Exporting <code>dudi</code> Elements | 51 |
| 4 | Multivariate Analysis Graphs | 53 |
| 4.1 | Introduction | 53 |
| 4.2 | Basics of adegraphics | 54 |
| 4.2.1 | <code>ADEg</code> and <code>ADEgS</code> Objects | 54 |
| 4.2.2 | Graphical Parameters | 56 |
| 4.2.3 | Main Functions and Methods | 59 |
| 4.2.4 | (Big) Data Storage | 60 |
| 4.3 | Simple Examples | 61 |
| 4.4 | Spatial Representations | 65 |
| 4.5 | Automatic Graph Collections | 66 |
| 4.5.1 | Splitting Individuals with the <code>facets</code> Argument | 66 |
| 4.5.2 | Multiple Variables | 66 |
| 4.5.3 | Outputs of Multivariate Methods | 67 |

- 4.6 Step-by-Step Creation of an ADEgS 68
 - 4.6.1 Graphical Representations of One Axis 69
 - 4.6.2 Graphical Representations of Two Axes 72
- 4.7 Conclusion 75
- 5 Description of Environmental Variables Structures 77**
 - 5.1 Introduction 77
 - 5.2 Standardised Principal Component Analysis (PCA) 79
 - 5.3 Multiple Correspondence Analysis (MCA) 86
 - 5.4 Hill and Smith Analysis (HSA) 93
 - 5.5 Other Simple Methods 96
- 6 Description of Species Structures 97**
 - 6.1 Introduction 97
 - 6.2 Correspondence Analysis (CA) 100
 - 6.3 Centred PCA (cPCA) 109
 - 6.4 Standardised and Non-centred PCA 110
 - 6.5 Principal Coordinate Analysis (PCoA) 112
 - 6.6 Other Simple Methods 116
- 7 Taking into Account Groups of Sites 119**
 - 7.1 Introduction 119
 - 7.2 An Environmental Situation 122
 - 7.3 Between-Class Analysis: Analysing Differences Between Groups 125
 - 7.3.1 Between-Site Analysis 127
 - 7.3.2 Between-Season Analysis 131
 - 7.4 Within-Class Analysis: Removing Differences Between Groups of Sites 131
 - 7.5 Discriminant Analysis 135
 - 7.6 Conclusion 140
- 8 Description of Species-Environment Relationships 141**
 - 8.1 Introduction 141
 - 8.2 Indirect Ordination 141
 - 8.3 Coinertia Analysis 144
 - 8.4 Analysis on Instrumental Variables 150
 - 8.4.1 Redundancy Analysis 152
 - 8.4.2 Canonical Correspondence Analysis 158
 - 8.4.3 Related Software and Methods 162
- 9 Analysing Changes in Structures 167**
 - 9.1 Introduction 167
 - 9.2 *K*-table Management in **ade4** 168
 - 9.2.1 *K*-table Examples 168
 - 9.2.2 Building and Using a *K*-table 169
 - 9.2.3 Separate Analyses of a *K*-table 172

| | | |
|-----------|--|------------|
| 9.3 | Strategies of K -table Methods | 174 |
| 9.4 | Partial Triadic Analysis | 176 |
| 9.5 | Foucart COA | 181 |
| 9.6 | STATIS on Operators | 183 |
| 9.7 | Multiple Factor Analysis | 193 |
| 9.8 | Multiple Coinertia Analysis | 198 |
| 9.9 | Conclusion | 201 |
| 10 | Analysing Changes in Co-structures | 205 |
| 10.1 | Introduction | 205 |
| 10.2 | BGCOIA | 207 |
| 10.3 | STATICO | 209 |
| 10.4 | COSTATIS | 216 |
| 10.5 | Conclusion | 220 |
| 11 | Relating Species Traits to Environment | 223 |
| 11.1 | Introduction | 223 |
| 11.2 | RLQ Analysis | 224 |
| 11.3 | Fourth-Corner Analysis | 230 |
| 11.4 | Combining Both Approaches | 234 |
| 11.5 | Extensions | 237 |
| 12 | Analysing Spatial Structures | 239 |
| 12.1 | Introduction | 239 |
| 12.2 | Managing Spatial Data | 239 |
| 12.3 | From Spatial Data to Spatial Weights | 242 |
| 12.4 | Spatial Autocorrelation | 246 |
| 12.5 | Detecting Spatial Multivariate Structures | 248 |
| | 12.5.1 Moran's Eigenvector Maps (MEMs) | 250 |
| | 12.5.2 MULTISPATI Analysis | 253 |
| 13 | Analysing Phylogenetic Structures | 261 |
| 13.1 | Introduction | 261 |
| 13.2 | Managing Phylogenetic Comparative Data | 262 |
| 13.3 | Computing Phylogenetic Proximities | 264 |
| 13.4 | Detecting Phylogenetic Structures | 266 |
| | 13.4.1 Moran's I | 266 |
| | 13.4.2 Abouheif's Test | 268 |
| 13.5 | Describing the Phylogenetic Signal | 269 |
| | 13.5.1 Orthonormal Bases | 269 |
| | 13.5.2 Phylogenetic Decomposition with the Orthogram | 271 |
| | 13.5.3 Removing Phylogenetic Autocorrelation | 273 |
| 13.6 | Phylogenetic Principal Component Analysis (pPCA) | 277 |
| 14 | Analysing Patterns of Biodiversity | 281 |
| 14.1 | Introduction | 281 |
| 14.2 | Ordination of the Faunistic Table | 282 |

- 14.3 From Trait Data to Dissimilarities 283
- 14.4 Double Principal Coordinate Analysis (DPCoA) 285
- 14.5 DPCoA and Diversity..... 291
- 14.6 Conclusions 294

- A A Euclidean Viewpoint on Statistics 295**
 - A.1 Inner and Dot Products 295
 - A.2 Length, Projection, Angle and Distance 296
 - A.3 Mean and Variance..... 297
 - A.4 Weighted Mean and Variance 298
 - A.5 Covariance and Correlation 299
 - A.6 Linear Regression..... 300
 - A.7 Categorical Variables 302
 - A.8 Weighted Multiple Regression 303

- B Graphical User Interface 307**
 - B.1 Introduction 307
 - B.2 Overview of the **ade4TkGUI** Package 308
 - B.3 Conclusion 311

- C Index of Boxes..... 313**
 - C.1 Chapter 3: The `dudi` Class 313
 - C.2 Chapter 4: Multivariate Analysis Graphs 313
 - C.3 Chapter 5: Description of Environmental Variables Structures.... 313
 - C.4 Chapter 6: Description of Species Structures..... 313
 - C.5 Chapter 7: Taking into Account Groups of Sites 314
 - C.6 Chapter 8: Description of Species-Environment Relationships ... 314
 - C.7 Chapter 9: Analysing Changes in Structures 314
 - C.8 Chapter 10: Analysing Changes in Co-structures 315
 - C.9 Chapter 11: Relating Species Traits to Environment..... 315
 - C.10 Chapter 12: Analysing Spatial Structures..... 315
 - C.11 Chapter 13: Analysing Phylogenetic Structures..... 315
 - C.12 Chapter 14: Analysing Patterns of Biodiversity 315

- References..... 317**

- Index..... 325**