

Cambridge University Press
052181409X - Multivariate Analysis of Ecological Data using CANOCO
Jan Leps and Petr Smilauer
Frontmatter
[More information](#)

Multivariate Analysis of Ecological Data using CANOCO

This book is primarily written for ecologists needing to analyse data resulting from field observations and experiments. It will be particularly useful for students and researchers dealing with complex ecological problems, such as the variation of biotic communities with environmental conditions or the response of biotic communities to experimental manipulation. Following a simple introduction to ordination methods, the text focuses on constrained ordination methods (RDA, CCA) and the use of permutation tests of statistical hypotheses about multivariate data. An overview of classification methods, or modern regression methods (GLM, GAM, loess), is provided and guidance on the correct interpretation of ordination diagrams is given. Seven case studies of varying difficulty help to illustrate the suggested analytical methods, using Canoco for Windows software. The case studies utilize both the descriptive and manipulative approaches, and they are supported by data sets and project files available from the book website.

JAN LEPS is Professor of Ecology in the Department of Botany, at the University of South Bohemia, and in the Institute of Entomology at the Czech Academy of Sciences, Czech Republic.

PETR ŠMILAUER is Lecturer in Multivariate Statistics at the University of South Bohemia, Czech Republic.

Cambridge University Press
052181409X - Multivariate Analysis of Ecological Data using CANOCO
Jan Leps and Petr Smilauer
Frontmatter
[More information](#)

Multivariate Analysis of Ecological Data using CANOCO

JAN LEPS
University of South Bohemia, and
Czech Academy of Sciences,
Czech Republic

PETR ŠMILAUER
University of South Bohemia,
Czech Republic



Cambridge University Press
052181409X - Multivariate Analysis of Ecological Data using CANOCO
Jan Leps and Petr Šmilauer
Frontmatter
[More information](#)

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Jan Lepš and Petr Šmilauer 2003

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2003

Printed in the United Kingdom at the University Press, Cambridge

Typefaces Lexicon No. 2 10/14 pt and Lexicon No. 1 System L^AT_EX 2_ε [TB]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Lepš, Jan 1953–
Multivariate analysis of ecological data using CANOCO / Jan Lepš & Petr Šmilauer.
p. cm.
Includes bibliographical references (p.).
ISBN 0 521 81409 X (hb.) ISBN 0 521 89108 6 (pb.)
1. Ecology – Statistical methods. 2. Multivariate analysis. I. Šmilauer, Petr, 1967-II.
Title.

QH541.15.S72L47 2003
577'.07'27–dc21 2002034801

ISBN 0 521 81409 X hardback
ISBN 0 521 89108 6 paperback

Contents

Preface page ix

- 1. Introduction and data manipulation 1
 - 1.1. Why ordination? 1
 - 1.2. Terminology 4
 - 1.3. Types of analyses 6
 - 1.4. Response variables 7
 - 1.5. Explanatory variables 7
 - 1.6. Handling missing values in data 9
 - 1.7. Importing data from spreadsheets – WCanImp program 11
 - 1.8. Transformation of species data 13
 - 1.9. Transformation of explanatory variables 15
- 2. Experimental design 16
 - 2.1. Completely randomized design 16
 - 2.2. Randomized complete blocks 17
 - 2.3. Latin square design 18
 - 2.4. Most frequent errors – pseudoreplications 19
 - 2.5. Combining more than one factor 19
 - 2.6. Following the development of objects in time – repeated observations 21
 - 2.7. Experimental and observational data 23
- 3. Basics of gradient analysis 25
 - 3.1. Techniques of gradient analysis 26
 - 3.2. Models of species response to environmental gradients 26
 - 3.3. Estimating species optima by the weighted averaging method 29
 - 3.4. Calibration 32

vi Contents

- 3.5. Ordination 33
- 3.6. Constrained ordination 36
- 3.7. Basic ordination techniques 37
- 3.8. Ordination diagrams 38
- 3.9. Two approaches 38
- 3.10. Testing significance of the relation with environmental variables 40
- 3.11. Monte Carlo permutation tests for the significance of regression 41

- 4. **Using the Canoco for Windows 4.5 package** 43
 - 4.1. Overview of the package 43
 - 4.2. Typical flow-chart of data analysis with Canoco for Windows 48
 - 4.3. Deciding on the ordination method: unimodal or linear? 50
 - 4.4. PCA or RDA ordination: centring and standardizing 51
 - 4.5. DCA ordination: detrending 52
 - 4.6. Scaling of ordination scores 54
 - 4.7. Running CanoDraw for Windows 4.0 57
 - 4.8. New analyses providing new views of our data sets 58

- 5. **Constrained ordination and permutation tests** 60
 - 5.1. Linear multiple regression model 60
 - 5.2. Constrained ordination model 61
 - 5.3. RDA: constrained PCA 62
 - 5.4. Monte Carlo permutation test: an introduction 64
 - 5.5. Null hypothesis model 65
 - 5.6. Test statistics 65
 - 5.7. Spatial and temporal constraints 68
 - 5.8. Split-plot constraints 69
 - 5.9. Stepwise selection of the model 70
 - 5.10. Variance partitioning procedure 73

- 6. **Similarity measures** 76
 - 6.1. Similarity measures for presence–absence data 77
 - 6.2. Similarity measures for quantitative data 80
 - 6.3. Similarity of samples versus similarity of communities 85
 - 6.4. Principal coordinates analysis 86
 - 6.5. Non-metric multidimensional scaling 89
 - 6.6. Constrained principal coordinates analysis (db-RDA) 91
 - 6.7. Mantel test 92

- 7. **Classification methods** 96
 - 7.1. Sample data set 96
 - 7.2. Non-hierarchical classification (K-means clustering) 97
 - 7.3. Hierarchical classifications 101
 - 7.4. TWINSpan 108

- 8. **Regression methods** 117
 - 8.1. Regression models in general 117
 - 8.2. General linear model: terms 119
 - 8.3. Generalized linear models (GLM) 121
 - 8.4. Loess smoother 123
 - 8.5. Generalized additive models (GAM) 124
 - 8.6. Classification and regression trees 125
 - 8.7. Modelling species response curves with
CanoDraw 126

- 9. **Advanced use of ordination** 140
 - 9.1. Testing the significance of individual constrained ordination
axes 140
 - 9.2. Hierarchical analysis of community variation 141
 - 9.3. Principal response curves (PRC) method 144
 - 9.4. Linear discriminant analysis 147

- 10. **Visualizing multivariate data** 149
 - 10.1. What we can infer from ordination diagrams: linear
methods 149
 - 10.2. What we can infer from ordination diagrams: unimodal
methods 159
 - 10.3. Visualizing ordination results with statistical models 162
 - 10.4. Ordination diagnostics 163
 - 10.5. t-value biplot interpretation 166

- 11. **Case study 1: Variation in forest bird assemblages** 168
 - 11.1. Data manipulation 168
 - 11.2. Deciding between linear and unimodal ordination 169
 - 11.3. Indirect analysis: portraying variation in bird community 171
 - 11.4. Direct gradient analysis: effect of altitude 177
 - 11.5. Direct gradient analysis: additional effect of other habitat
characteristics 180

- 12. **Case study 2: Search for community composition patterns and
their environmental correlates: vegetation of spring
meadows** 183
 - 12.1. The unconstrained ordination 184
 - 12.2. Constrained ordinations 187
 - 12.3. Classification 192
 - 12.4. Suggestions for additional analyses 193

viii Contents

- 13. **Case study 3: Separating the effects of explanatory variables** 196
 - 13.1. Introduction 196
 - 13.2. Data 197
 - 13.3. Data analysis 197

- 14. **Case study 4: Evaluation of experiments in randomized complete blocks** 206
 - 14.1. Introduction 206
 - 14.2. Data 207
 - 14.3. Data analysis 207

- 15. **Case study 5: Analysis of repeated observations of species composition from a factorial experiment** 215
 - 15.1. Introduction 215
 - 15.2. Experimental design 216
 - 15.3. Sampling 216
 - 15.4. Data analysis 216
 - 15.5. Univariate analyses 218
 - 15.6. Constrained ordinations 218
 - 15.7. Further use of ordination results 222
 - 15.8. Principal response curves 224

- 16. **Case study 6: Hierarchical analysis of crayfish community variation** 236
 - 16.1. Data and design 236
 - 16.2. Differences among sampling locations 237
 - 16.3. Hierarchical decomposition of community variation 239

- 17. **Case study 7: Differentiating two species and their hybrids with discriminant analysis** 245
 - 17.1. Data 245
 - 17.2. Stepwise selection of discriminating variables 246
 - 17.3. Adjusting the discriminating variables 249
 - 17.4. Displaying results 250

- Appendix A: Sample datasets and projects* 253
- Appendix B: Vocabulary* 254
- Appendix C: Overview of available software* 258
- References* 262
- Index* 267

Preface

The multidimensional data on community composition, properties of individual populations, or properties of environment are the bread and butter of an ecologist's life. They need to be analysed with taking their multidimensionality into account. A reductionist approach of looking at the properties of each variable separately does not work in most cases. The methods for statistical analysis of such data sets fit under the umbrella of 'multivariate statistical methods'.

In this book, we present a hopefully consistent set of approaches to answering many of the questions that an ecologist might have about the studied systems. Nevertheless, we admit that our views are biased to some extent, and we pay limited attention to other less parametric methods, such as the family of non-metric multidimensional scaling (NMDS) algorithms or the group of methods similar to the Mantel test or the ANOSIM method. We do not want to fuel the controversy between proponents of various approaches to analysing multivariate data. We simply claim that the solutions presented are not the only ones possible, but they work for us, as well as many others.

We also give greater emphasis to ordination methods compared to classification approaches, but we do not imply that the classification methods are not useful. Our description of multivariate methods is extended by a short overview of regression analysis, including some of the more recent developments such as generalized additive models.

Our intention is to provide the reader with both the basic understanding of principles of multivariate methods and the skills needed to use those methods in his/her own work. Consequently, all the methods are illustrated by examples. For all of them, we provide the data on our web page (see Appendix A), and for all the analyses carried out by the CANOCO program, we also provide the CANOCO project files containing all the options needed for particular analysis. The seven case studies that conclude the book contain tutorials, where the

x Preface

analysis options are explained and the software use is described. The individual case studies differ intentionally in the depth of explanation of the necessary steps. In the first case study, the tutorial is in a ‘cookbook’ form, whereas a detailed description of individual steps in the subsequent case studies is only provided for the more complicated and advanced methods that are not described in the preceding tutorial chapters.

The methods discussed in this book are widely used among plant, animal and soil biologists, as well as in the hydrobiology. The slant towards plant community ecology is an inevitable consequence of the research background of both authors.

This handbook provides study materials for the participants of a course regularly taught at our university called Multivariate Analysis of Ecological Data. We hope that this book can also be used for other similar courses, as well as by individual students seeking improvement in their ability to analyse collected data.

We hope that this book provides an easy-to-read supplement to the more exact and detailed publications such as the collection of Cajo Ter Braak’s papers and the Canoco for Windows 4.5 manual. In addition to the scope of those publications, this book adds information on classification methods of multivariate data analysis and introduces modern regression methods, which we have found most useful in ecological research.

In some case studies, we needed to compare multivariate methods with their univariate counterparts. The univariate methods are demonstrated using the Statistica for Windows package (version 5.5). We have also used this package to demonstrate multivariate methods not included in the CANOCO program, such as non-metric multidimensional scaling or the methods of cluster analysis. However, all those methods are available in other statistical packages so the readers can hopefully use their favourite statistical package, if different from Statistica. Please note that we have omitted the trademark and registered trademark symbols when referring to commercial software products.

We would like to thank John Birks, Robert Pillsbury and Samara Hamzé for correcting the English used in this textbook. We are grateful to all who read drafts of the manuscript and gave us many useful comments: Cajo Ter Braak, John Birks, Mike Palmer and Marek Rejmánek. Additional useful comments on the text and the language were provided by the students of Oklahoma State University: Jerad Linneman, Jerry Husak, Kris Karsten, Raelene Crandall and Krysten Schuler. Sarah Price did a great job as our copy-editor, and improved the text in countless places.

Camille Flinders, Tomáš Hájek and Milan Štech kindly provided data sets, respectively, for case studies 6, 2 and 7.

Cambridge University Press
052181409X - Multivariate Analysis of Ecological Data using CANOCO
Jan Leps and Petr Smilauer
Frontmatter
[More information](#)

P.Š. wants to thank his wife Marie and daughters Marie and Tereza for their continuous support and patience with him.

J.L. insisted on stating that the ordering of authorship is based purely on the alphabetical order of their names. He wants to thank his parents for support and his daughters Anna and Tereza for patience.