

Cambridge University Press

978-1-107-69440-8 - Multivariate Analysis of Ecological Data using Canoco 5: Second Edition

Petr Šmilauer and Jan Lepš

Frontmatter

[More information](#)

## Multivariate Analysis of Ecological Data using Canoco 5

This revised and updated edition focuses on constrained ordination (RDA, CCA), variation partitioning and the use of permutation tests of statistical hypotheses about multivariate data. Both classification and modern regression methods (GLM, GAM, loess) are reviewed and species functional traits and spatial structures are analysed.

Nine case studies of varying difficulty help to illustrate the suggested analytical methods, using the latest version of Canoco 5. All studies utilise descriptive and manipulative approaches, and are supported by data sets and project files available from the book website: <http://regent.prf.jcu.cz/maed2/>.

Written primarily for community ecologists needing to analyse data resulting from field observations and experiments, this book is a valuable resource for students and researchers dealing with both simple and complex ecological problems, such as the variation of biotic communities with environmental conditions or their response to experimental manipulation.

**Petr Šmilauer** is Associate Professor of Ecology in the Department of Ecosystem Biology, at the University of South Bohemia. His main research interests are: multivariate statistical analysis, modern regression methods, as well as the role of arbuscular mycorrhizal symbiosis in plant communities. He is co-author of the multivariate analysis software Canoco 5, CANOCO for Windows 4.5, CanoDraw, and TWINSPAN for Windows.

**Jan Lepš** 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. His main research interests include: plant community biology, statistical analysis in the field of ecology, as well as the studies of species diversity, and the role of functional traits in plant community ecology and ecology of hemiparasitic plants. Together with P. Šmilauer, he regularly offers international courses on multivariate statistics.

Cambridge University Press

978-1-107-69440-8 - Multivariate Analysis of Ecological Data using Canoco 5: Second Edition

Petr Šmilauer and Jan Lepš

Frontmatter

[More information](#)

---

Cambridge University Press

978-1-107-69440-8 - Multivariate Analysis of Ecological Data using Canoco 5: Second Edition

Petr Šmilauer and Jan Lepš

Frontmatter

[More information](#)

# Multivariate Analysis of Ecological Data using Canoco 5

Second Edition

PETR ŠMILAUER

University of South Bohemia, Czech Republic

JAN LEPŠ

University of South Bohemia, Czech Republic



CAMBRIDGE  
UNIVERSITY PRESS

## CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Published in the United States of America by Cambridge University Press, New York

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781107694408](http://www.cambridge.org/9781107694408)

© Cambridge University Press 2014

This publication 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

Second Edition 2014

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

*A catalogue record for this publication is available from the British Library*

*Library of Congress Cataloguing in Publication data*

Šmilauer, Petr, 1967–

Multivariate analysis of ecological data using CANOCO 5 / Petr Šmilauer, University of South Bohemia, Czech Republic, Jan Lepš, University of South Bohemia, Czech Republic. – Second edition.

pages cm

Includes bibliographical references and index.

ISBN 978-1-107-69440-8 (pbk.)

1. Ecology – Statistical methods. 2. Multivariate analysis. 3. Ecology – Statistical methods – Case studies. 4. Multivariate analysis – Case studies. I. Lepš, Jan, 1953– II. Title.

QH541.15.S72S63 2014

577.01'5195 – dc23 2013048954

ISBN 978-1-107-69440-8 Paperback

Additional resources for this publication at <http://regent.prf.jcu.cz/maed2/>

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

## Contents

	<i>page x</i>
<i>Preface</i>	
<b>1 Introduction and data types</b>	<b>1</b>
1.1 Why ordination?	1
1.2 Data types	4
1.3 Data transformation and standardisation	7
1.4 Missing values	11
1.5 Types of analyses	12
<b>2 Using Canoco 5</b>	<b>15</b>
2.1 Philosophy of Canoco 5	15
2.2 Data import and editing	17
2.3 Defining analyses	24
2.4 Visualising results	33
2.5 Beware, CANOCO 4.x users!	36
<b>3 Experimental design</b>	<b>39</b>
3.1 Completely randomised design	39
3.2 Randomised complete blocks	40
3.3 Latin square design	41
3.4 Pseudoreplicates	42
3.5 Combining more than one factor	44
3.6 Following the development of objects in time: repeated observations	45
3.7 Experimental and observational data	48
<b>4 Basics of gradient analysis</b>	<b>50</b>
4.1 Techniques of gradient analysis	51
4.2 Models of response to gradients	51
4.3 Estimating species optima by weighted averaging	53
4.4 Calibration	56
4.5 Unconstrained ordination	57

4.6	Constrained ordination	60
4.7	Basic ordination techniques	61
4.8	Ordination axes as optimal predictors	62
4.9	Ordination diagrams	64
4.10	Two approaches	66
4.11	Testing significance of the relation with explanatory variables	66
4.12	Monte Carlo permutation tests for the significance of regression	67
4.13	Relating two biotic communities	68
4.14	Community composition as a cause: using reverse analysis	69
<b>5</b>	<b>Permutation tests and variation partitioning</b>	<b>71</b>
5.1	Permutation tests: the philosophy	71
5.2	Pseudo-F statistics and significance	72
5.3	Testing individual constrained axes	74
5.4	Tests with spatial or temporal constraints	75
5.5	Tests with hierarchical constraints	79
5.6	Simple versus conditional effects and stepwise selection	83
5.7	Variation partitioning	88
5.8	Significance adjustment for multiple tests	91
<b>6</b>	<b>Similarity measures and distance-based methods</b>	<b>92</b>
6.1	Similarity measures for presence–absence data	93
6.2	Similarity measures for quantitative data	96
6.3	Similarity of cases versus similarity of communities	101
6.4	Similarity between species in trait values	102
6.5	Principal coordinates analysis	103
6.6	Constrained principal coordinates analysis (db–RDA)	106
6.7	Non-metric multidimensional scaling	107
6.8	Mantel test	108
<b>7</b>	<b>Classification methods</b>	<b>112</b>
7.1	Example data set properties	112
7.2	Non-hierarchical classification (K-means clustering)	113
7.3	Hierarchical classification	116
7.4	TWINSPAN	121
<b>8</b>	<b>Regression methods</b>	<b>129</b>
8.1	Regression models in general	129
8.2	General linear model: terms	131
8.3	Generalized linear models (GLM)	133
8.4	Loess smoother	135

8.5	Generalized additive models (GAM)	136
8.6	Mixed-effect models (LMM, GLMM and GAMM)	137
8.7	Classification and regression trees (CART)	139
8.8	Modelling species response curves with Canoco	140
<b>9</b>	<b>Interpreting community composition with functional traits</b>	<b>151</b>
9.1	Required data	152
9.2	Two approaches in traits – environment studies	154
9.3	Community-based approach	158
9.4	Species-based approach	162
<b>10</b>	<b>Advanced use of ordination</b>	<b>167</b>
10.1	Principal response curves (PRC)	167
10.2	Separating spatial variation	169
10.3	Linear discriminant analysis	173
10.4	Hierarchical analysis of community variation	174
10.5	Partitioning diversity indices into alpha and beta components	177
10.6	Predicting community composition	182
<b>11</b>	<b>Visualising multivariate data</b>	<b>184</b>
11.1	Reading ordination diagrams of linear methods	186
11.2	Reading ordination diagrams of unimodal methods	195
11.3	Attribute plots	199
11.4	Visualising classification, groups, and sequences	202
11.5	T-value biplot	205
<b>12</b>	<b>Case study 1: Variation in forest bird assemblages</b>	<b>208</b>
12.1	Unconstrained ordination: portraying variation in bird community	209
12.2	Simple constrained ordination: the effect of altitude on bird community	215
12.3	Partial constrained ordination: additional effect of other habitat characteristics	218
12.4	Separating and testing alpha and beta diversity	221
<b>13</b>	<b>Case study 2: Search for community composition patterns and their environmental correlates: vegetation of spring meadows</b>	<b>226</b>
13.1	Unconstrained ordination	227
13.2	Constrained ordination	230
13.3	Classification	237
13.4	Suggestions for additional analyses	238
13.5	Comparing two communities	239

<b>14</b>	<b>Case study 3: Separating the effects of explanatory variables</b>	246
	14.1 Introduction	246
	14.2 Data	247
	14.3 Changes in species richness and composition	247
	14.4 Changes in species traits	255
<b>15</b>	<b>Case study 4: Evaluation of experiments in randomised complete blocks</b>	258
	15.1 Introduction	258
	15.2 Data	258
	15.3 Analysis	259
	15.4 Calculating ANOVA using constrained ordination	265
<b>16</b>	<b>Case study 5: Analysis of repeated observations of species composition from a factorial experiment</b>	267
	16.1 Introduction	267
	16.2 Experimental design	267
	16.3 Data coding and use	268
	16.4 Univariate analyses	270
	16.5 Constrained ordinations	270
	16.6 Principal response curves	275
	16.7 Temporal changes across treatments	280
	16.8 Changes in composition of functional traits	285
<b>17</b>	<b>Case study 6: Hierarchical analysis of crayfish community variation</b>	301
	17.1 Data and design	301
	17.2 Differences among sampling locations	302
	17.3 Hierarchical decomposition of community variation	305
<b>18</b>	<b>Case study 7: Analysis of taxonomic data with discriminant analysis and distance-based ordination</b>	309
	18.1 Data	309
	18.2 Summarising morphological data with PCA	310
	18.3 Linear discriminant analysis of morphological data	313
	18.4 Principal coordinates analysis of AFLP data	317
	18.5 Testing taxon differences in AFLP data using db-RDA	320
	18.6 Taking populations into account	322
<b>19</b>	<b>Case study 8: Separating effects of space and environment on oribatid community with PCNM</b>	324
	19.1 Ignoring the space	324
	19.2 Detecting spatial trends	326



---

19.3	All-scale spatial variation of community and environment	328
19.4	Variation partitioning with spatial predictors	332
19.5	Visualising spatial variation	333
<b>20</b>	<b>Case study 9: Performing linear regression with redundancy analysis</b>	<b>337</b>
20.1	Data	337
20.2	Linear regression using program R	337
20.3	Linear regression with redundancy analysis	340
20.4	Fitting generalized linear models in Canoco	342
	<i>Appendix A Glossary</i>	343
	<i>Appendix B Sample data sets and projects</i>	346
	<i>Appendix C Access to Canoco and overview of other software</i>	347
	<i>Appendix D Working with R</i>	350
	<i>References</i>	351
	<i>Index to useful tasks in Canoco 5</i>	359
	<i>Subject index</i>	360

## 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. Such data need to be analysed taking into account their multidimensionality. 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 consistent set of approaches to answer many of the questions that an ecologist might have about the studied systems. Nevertheless, we happily admit that other quantitative ecologists may approach the same set of questions with a toolbox of methods (partly) different from those presented here. We pay only a 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. We do not want to fuel the sometimes seen controversy between proponents of various approaches to analysing multivariate data. We simply claim that the solutions presented here are not the only ones possible, but they worked for us, as well as for many other researchers.

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 the generalized additive models or CART models, because the regression models often complement the results of multivariate analyses.

We assume the reader has knowledge of testing statistical hypotheses, of linear regression and ANOVA in the range covered by introductory statistical courses for undergraduates and we make no attempt here to explain corresponding terms or principles.

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, the methods are illustrated by examples. For all of them, we provide the data on our web page (see Appendix B), and for all the analyses carried out by the Canoco<sup>1</sup> program, we also provide Canoco 5 project files, containing analyses with required settings and precomputed results.

<sup>1</sup> Although the CANOCO name was originally an acronym, it became a recognized entity over the years and this is reflected in the change from upper-case letters, as also made in the Canoco 5 manual.

The nine case studies which conclude the book contain tutorials, where the project/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 two case studies, the tutorial is more 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. You can work with offered case studies in any order, but if you are new to ordination methods, we recommend you to read at least Chapters 1 to 4 first, and if you are an experienced user of ordination methods (including constrained ones), but not yet friendly with Canoco software, read at least Chapter 2 that introduces the Canoco program<sup>2</sup> and then start with Case studies 1 and 2. The ‘Index to useful tasks in Canoco 5’, which is located before the standard Index, allows the reader to quickly find solutions to many common technical tasks in work with Canoco 5 software, which are described in this book.

In the second edition, we have tried to cater for both beginning and more advanced users. We have therefore put many of the more advanced comments or suggestions into an extensive set of footnotes. The main text of this book can be understood while ignoring the footnotes, but they provide greater insight for the advanced topics or explain technical details.

The methods discussed in this book are widely used among plant, animal and soil biologists, as well as in freshwater and marine biology or in landscape ecology. In most of these fields, the methods are now routinely applied also to the data sets obtained with molecular biology techniques.

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, the Canoco 5 manual, or the Legendre and Legendre (2012) textbook. The Reference manual and user’s guide to the new version of Canoco 5 is, in fact, so often referred to that instead of citing Ter Braak and Šmilauer (2012), we use ‘Canoco 5 manual’ throughout this book.

In some case studies, we needed to compare multivariate methods with their univariate counterparts. The univariate methods are demonstrated using the freely available R software (R Core Team 2013), which can be also used to work with multivariate methods described here, but not available in Canoco 5 software (cluster analysis, Mantel test). See Appendix D for a link to a brief tutorial on working with R (in the context of tasks present in this book). But these methods are also available in other statistical packages so the readers can hopefully use their favourite software, if different from R (see Appendix C with an overview of alternative software for multivariate statistical analysis).

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 our English in the first edition. We are grateful to all who read drafts of the first edition 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

<sup>2</sup> Reading Chapter 2 is recommended even to experienced users of CANOCO version 4.0 or 4.5.

Cambridge University Press

978-1-107-69440-8 - Multivariate Analysis of Ecological Data using Canoco 5: Second Edition

Petr Šmilauer and Jan Lepš

Frontmatter

[More information](#)

the students of Oklahoma State University: Jerad Linneman, Jerry Husak, Kris Karsten, Raelene Crandall, and Krysten Schuler.

We are much indebted to Mike Palmer for his meticulous reading of the whole manuscript of the second edition and many great suggestions concerning the language, presented theory and book usability, vastly improving the book quality. We are thankful to Cajo ter Braak for his numerous comments on the second edition. Francesco de Bello provided multiple suggestions on the text concerning the analysis of species traits.

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

PŠ wants to thank his wife Marie for her continuous support (and drawing Figure 10–3) and to his daughters Marie and Tereza for their patience with him. He also insisted on stating that the ordering of authorship is based purely on the alphabetical order of their names, this time reversed – for a change from the first edition.

JL wants to thank his parents and Olina for support, his daughters Anna and Tereza for patience and his grandchildren Anna, Eliška, and Matěj for keeping him in a good mood.