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 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, Czech Republic.

Petr Šmilauer is Lecturer in Multivariate Statistics at the University of South Bohemia, Czech Republic.

Multivariate Analysis of Ecological Data using CANOCO

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

PETR ŠMILAUER
University of South Bohemia,
Czech Republic

CAMBRIDGE
UNIVERSITY PRESS

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

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* LaTeX 2ε [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 ecolological 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

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

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.

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.