



Data Analysis in Vegetation Ecology

SECOND
EDITION

Otto
Wildi

 WILEY-BLACKWELL



Data Analysis in Vegetation Ecology

Data Analysis in Vegetation Ecology

Second Edition

Otto Wildi

*WSL Swiss Federal Institute for Forest, Snow and Landscape Research
Birmensdorf, Switzerland*



WILEY-BLACKWELL

A John Wiley & Sons, Ltd., Publication

This edition first published 2013
© 2013 by John Wiley & Sons, Ltd

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

Registered office: John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial offices: 9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK
111 River Street, Hoboken, NJ 07030-5774, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this work has been asserted in accordance with the UK Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author(s) have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Wildi, Otto.

Data analysis in vegetation ecology / Otto Wildi.
pages cm

Includes bibliographical references and index.

ISBN 978-1-118-38404-6 (cloth) – ISBN 978-1-118-38403-9 (pbk.) 1. Plant communities–Data processing. 2. Plant communities–Mathematical models. 3. Plant ecology–Data processing. 4. Plant ecology–Mathematical models. I.

Title.

QK911.W523 2013

581.70285 – dc23

2012047729

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

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Cover image: Image supplied by Author

Cover design by Steve Thompson

Typeset in 10.5/13 Times by Laserwords Private Limited, Chennai, India

First Impression 2013

*Plants are so unlike people that it's very difficult
for us to appreciate fully their complexity and sophistication.*

Michael Pollan, *The Botany of Desire*

Contents

Preface to the second edition	xi
Preface to the first edition	xv
List of figures	xix
List of tables	xxv
About the companion website	xxvii
1 Introduction	1
2 Patterns in vegetation ecology	5
2.1 Pattern recognition	5
2.2 Interpretation of patterns	9
2.3 Sampling for pattern recognition	12
2.3.1 Getting a sample	12
2.3.2 Organizing the data	14
2.4 Pattern recognition in $\mathcal{R}$	17
3 Transformation	23
3.1 Data types	23
3.2 Scalar transformation and the species enigma	26
3.3 Vector transformation	30
3.4 Example: Transformation of plant cover data	33
4 Multivariate comparison	37
4.1 Resemblance in multivariate space	37
4.2 Geometric approach	38
4.3 Contingency measures	43
4.4 Product moments	45
4.5 The resemblance matrix	48
4.6 Assessing the quality of classifications	50

5	Classification	53
5.1	Group structures	53
5.2	Linkage clustering	56
5.3	Average linkage clustering	59
5.4	Minimum-variance clustering	61
5.5	Forming groups	63
5.6	Silhouette plot and fuzzy representation	66
6	Ordination	71
6.1	Why ordination?	71
6.2	Principal component analysis	75
6.3	Principal coordinates analysis	82
6.4	Correspondence analysis	86
6.5	Heuristic ordination	89
6.5.1	The horseshoe or arch effect	89
6.5.2	Flexible shortest path adjustment	91
6.5.3	Nonmetric multidimensional scaling	93
6.5.4	Detrended correspondence analysis	95
6.6	How to interpret ordinations	96
6.7	Ranking by orthogonal components	100
6.7.1	RANK method	100
6.7.2	A sampling design based on RANK (example)	104
7	Ecological patterns	109
7.1	Pattern and ecological response	109
7.2	Evaluating groups	111
7.2.1	Variance testing	111
7.2.2	Variance ranking	115
7.2.3	Ranking by indicator values	117
7.2.4	Contingency tables	120
7.3	Correlating spaces	124
7.3.1	The Mantel test	124
7.3.2	Correlograms	127
7.3.3	More trends: ‘Schlaenggli’ data revisited	130
7.4	Multivariate linear models	134
7.4.1	Constrained ordination	134
7.4.2	Nonparametric multiple analysis of variance	141

7.5	Synoptic vegetation tables	146
7.5.1	The aim of ordering tables	146
7.5.2	Steps involved in sorting tables	147
7.5.3	Example: ordering Ellenberg's data	151
8	Static predictive modelling	155
8.1	Predictive or explanatory?	155
8.2	Evaluating environmental predictors	156
8.3	Generalized linear models	159
8.4	Generalized additive models	164
8.5	Classification and regression trees	166
8.6	Building scenarios	169
8.7	Modelling vegetation types	171
8.8	Expected wetland vegetation (example)	176
9	Vegetation change in time	185
9.1	Coping with time	185
9.2	Temporal autocorrelation	186
9.3	Rate of change and trend	188
9.4	Markov models	192
9.5	Space-for-time substitution	199
9.5.1	Principle and method	199
9.5.2	The Swiss National Park succession (example)	203
9.6	Dynamics in pollen diagrams (example)	207
10	Dynamic modelling	213
10.1	Simulating time processes	214
10.2	Simulating space processes	222
10.3	Processes in the Swiss National Park	223
10.3.1	The temporal model	223
10.3.2	The spatial model	228
11	Large data sets: wetland patterns	233
11.1	Large data sets differ	233
11.2	Phytosociology revisited	235
11.3	Suppressing outliers	239
11.4	Replacing species with new attributes	241
11.5	Large synoptic tables?	245

12 Swiss forests: a case study	255
12.1 Aim of the study	255
12.2 Structure of the data set	256
12.3 Selected questions	258
12.3.1 Is the similarity pattern discrete or continuous?	258
12.3.2 Is there a scale effect from plot size?	262
12.3.3 Does the vegetation pattern reflect environmental conditions?	266
12.3.4 Is tree species distribution man-made?	270
12.3.5 Is the tree species pattern expected to change?	276
12.4 Conclusions	278
Bibliography	281
Appendix A Functions in package dave	293
Appendix B Data sets used	295
Index	297

Preface to the second edition

Successful attempts to include instructions in $\mathcal{R}$ motivated me to prepare a second edition of the book while keeping it basically unchanged in style and content. Hence, I hoped to circumvent yet another introduction to a software environment as done earlier for MULVA-5 (Wildi and Orlóci 1996), which I previously used in many of my examples. I found the syntax of $\mathcal{R}$ to be close to ordinary mathematical notation allowing technical instructions to be minimized. Finally, this book is not an introduction to $\mathcal{R}$. There are many others providing this, such as Crawley (2005), Venables and Ripley (2010), or for advanced users Borcard *et al.* (2011), all highly recommended and referenced. The instructions I included in this second edition are aimed to serve the inexperienced in $\mathcal{R}$, getting technical help from colleagues or experts in installing and initializing $\mathcal{R}$ and loading some packages and functions, including the one I specifically provide for this book (package `dave`). Unintendedly, doing the examples explained in this second edition may even act as a beginners course in $\mathcal{R}$, hopefully with minimum effort.

Writing this second edition was a delicate task too. First, various results had to be reproduced by an entirely different or newly developed software. Only after revising the very last chapter was it clear that all this could be done in $\mathcal{R}$. It is well known that many scientists using $\mathcal{R}$ love it, those who avoid it, fear it. My objective is to encourage newcomers to do the examples and I put every effort into most parsimonious solutions. The instructions and functions I prepared for the book look and hopefully feel simple, hiding the tremendous complexity of the $\mathcal{R}$ environment. In this context I thank my colleagues who gave me technical advice, Thomas Dalang, Dirk Schmatz, Meinrad Küchler and Alan Haynes. The attendees of a course held with

an early version of the book, namely Angéline Bedolla, Elizabeth Feldmeyer, Ulrich Graf, Julia Haas, Alan Haynes, Caroline Heiri, Martina Hobi, Christine Keller, Meinrad Küchler, Helen Küchler, Mathieu Lévesque, Anna Pedretti, Kathrin Priewasser, Anita C. Risch, Marcus Schaub, Martin Schütz, Anna Schweiger, Andreas Schwyzer, Bastian Ullrich and Sonja Wipf, helped me to identify bugs and traps. Again, Anita C. Risch and Martin Schütz were willing to read the whole text critically.

All examples in the book are derived in $\mathcal{R}$ version 2.15.2 (R Development Core Team 2012). Whenever a specific method was missing I wrote a new function to avoid overloading readers with cumbersome code. On the downside every new function represents yet another black box. In the current state the reader will find solutions for all methods presented in the book, although figures may appear a little different: for the book I adapted these to layout requirements using an extended set of plot parameters explained in $\mathcal{R}$ when typing `?plot.default` and further screening for `par`. In the end I devise an $\mathcal{R}$ package for this book: `dave`, the name composed of the initials of the book title (Appendix A). An integrated part of `dave` consists of the many data sets listed in Appendix B. I would like to express my thanks to all authors cited there for giving the right to access these, as far as yet unpublished. Many are real world examples, although, with respect to ongoing research, fairly aged.

While elaborating this second edition I got trapped by the temptation to extend the panoply of methods where functions of other packages were ready to use. This concerns, for example, resemblance measures, classification techniques and ordination methods. In the modelling part I replaced my old fashioned heuristic approach by the now widely used logistic regression techniques including instructions for scenario building, considered important in the time of global change. For newcomers in $\mathcal{R}$ I highly recommend following the instructions quite carefully: $\mathcal{R}$ is very much like a programming language and for the average human brain it is extremely difficult to exactly remember all the details to get the examples running. To support proper use I extended the index considerably to facilitate quick access to all major methods covered in this book. The later will work only when all packages required are loaded, namely `dave`, `labdsv`, `tree` and `vegan` and all considered ‘related’ upon downloading from a CERAN repository found on the Internet.

I would again like to thank the publications team of Wiley-Blackwell for all the encouragement and support I have received throughout this revision. We agreed that the new edition shall serve users not only in theory but

now also in practice, a combination adding to the complexity of publication. Finally, I express my thanks to my host institution, the Swiss Federal Institute for Forest, Snow and Landscape Research WSL, for providing access to its computer network and literature databases needed to complete this work.

Birmensdorf, 1 October 2012

Preface to the first edition

When starting to rearrange my lecture notes I had a 'short introduction to multivariate vegetation analysis' in mind. It ended up as a 'not so short introduction'. The book now summarizes some of the well-known methods used in vegetation ecology. The matter presented is but a small selection of what is available to date. By focussing on methodological issues I try to explain what plant ecologists do, and why they measure and analyse data. Rather than just generating numbers and pretty graphs, the models and methods I discuss are a contribution to the understanding of the state and functioning of the ecosystems analysed. But because researchers are usually driven by their curiosity about the functioning of the systems I successively began to integrate examples encountered in my work. These now occupy a considerable portion of this book. I am convinced that the fascination of research lies in the perception of the real world and its amalgamation in the form of high-quality data with hidden content processed by a variety of methods reflecting our model view of the world. Neither my results nor my conclusions are final. Hoping that the reader will like some of my ideas and perspectives, I encourage them to use and to improve on them. There is a considerable potential for innovation left.

The examples presented in this book all come from Central Europe. While this was not intended originally, I became convinced the topics they cover are of general relevance, as similar investigations exist almost everywhere in the world. An example is the pollen data set: pollen profiles offer the unique chance to study vegetation change over millennia. This is the time scale of processes such as climate change and the expansion of the human population. Another, much shorter time series than that of pollen data is found in permanent plot data originating from the Swiss National Park that I had the opportunity to look at. The unique feature of this is that it dates back to the year 1917, when Josias Braun-Banquet personally installed the first wooden poles, which are still in place. Records of the full set of species

have been collected ever since in five-year steps. A totally different data set comes from the Swiss Forest Inventory, presented in the last chapter of this book. Whereas many vegetation surveys are merely preferential collections of plot data, this data set is an example of systematic sampling on a grid encompassing huge environmental gradients. It helps to assess which patterns really exist, and whether some of those described in papers or textbooks are real or merely reflect the imagination or preference of researchers scanning the landscape for nice locations. In this case the data set available for answering the question is still moderate in size, but handling of large data sets will eventually be needed in similar contexts. I used the Swiss wetland data set as an example for handling data of much larger size, in this case with $n = 17\,608$ relevés. Although this is outnumbered by others, it resides on a statistical sampling design.

Some basic knowledge of vegetation ecology might be needed to understand the examples presented in this book. Readers wishing to acquire this are advised to refer, for example, to the comprehensive volumes *Vegetation Ecology* by van der Maarel (2005) and *Aims and Methods of Vegetation Ecology* by Mueller-Dombois and Ellenberg (1974), presently available as a reprint. The structure of my book is influenced by Orlóci's (1978) *Multivariate Analysis in Vegetation Research*, which I explored the first time when proofreading it in 1977. Various applications are found in the books of Gauch (1982), Pielou (1984) and Digby and Kempton (1987) and many multivariate methods used in vegetation ecology are introduced in Jongman *et al.* (1995). To study statistical methods used in this book in more detail, I strongly recommend the probably most comprehensive textbook existing today, the second edition of *Numerical Ecology* by Legendre and Legendre (1998). Several books provide an introduction to the use of statistical packages, which are referred to in the appendix. For many reasons I decided to omit the software issue in the main text; upon the request of several reviewers I added a section to the appendix where I reveal how I calculated my examples and mention programs, program packages and databases.

I would like to express my thanks to all individuals that have contributed to the success of this book. First of all Rachel Wade from Wiley-Blackwell, who strongly supported the efforts to print the manuscript in time and organized all the technical work. I thank Tim West for careful copy-editing, and Robert Hambrook for managing the production process. My colleagues Anita C. Risch and Martin Schütz revised the entire text, providing corrections and suggestions. Meinrad Küchler helped in the computation of several examples. André F. Lotter provided the pollen data set. I cannot remember all the people who had an influence on the point of view presented here:

many ideas came from László Orlóci through our long lasting collaboration, others from Madhur Anand, Enrico Féoli, Valério de Patta Pillar, Janos Podani and Helene Wagner. I particularly thank my family for encouraging me to tackle this work and for their tolerance when I was working at night and on weekends to get it completed.

Birmensdorf, 1 December 2009

List of figures

2.1	Portrait of Abraham Lincoln.	6
2.2	Vegetation mapping as a method for assessing a pattern.	7
2.3	Ordination of a typical horseshoe-shaped vegetation gradient.	8
2.4	A natural and a man-made event.	10
2.5	Primary production of the vegetation of Europe.	11
2.6	Distribution pattern of oak haplotypes in Switzerland.	11
2.7	The elements of sampling design.	14
2.8	Organization of vegetation and site data in $\mathcal{R}$.	16
2.9	Window view of data frame <code>nsit</code> .	20
3.1	An example of three data types.	24
3.2	Scalar transformation of population size.	27
3.3	Scalar transformation of the coordinates of a graph.	28
3.4	Overlap of two species with Gaussian response.	29
4.1	Presentation of data in the Euclidean space.	38
4.2	Three ways of measuring distance.	39
4.3	The correlation of vector j with vector k .	46
4.4	The average distance as a measure for homogeneity.	49
4.5	Similarities within and between the forest types of Switzerland.	51

5.1	Two-dimensional group structures.	54
5.2	A dendrogram from agglomerative hierarchical clustering.	56
5.3	Comparing different methods of linkage clustering.	57
5.4	Variance within and between groups.	61
5.5	Cutting dendrograms derived by different methods.	64
5.6	Silhouette plot example.	66
5.7	Silhouette plot of four clustering solutions.	68
6.1	Three-dimensional representation of similarity relationships.	72
6.2	Common operations in ordination.	73
6.3	Projecting data into ordination space in PCA.	75
6.4	Numerical example of PCA.	76
6.5	Main results of a PCA using real world data.	78
6.6	Projection of five-dimensional PCA ordination.	82
6.7	PCOA ordination using the 'Schlaengli' data set.	83
6.8	PCOA ordinations with six different resemblance measures.	85
6.9	Comparison of CA and PCA.	89
6.10	Origin of the arch effect.	90
6.11	Comparing PCOA and FSPA.	92
6.12	Comparison of PCOA and NMDS.	94
6.13	Comparison of CA and DCA.	96
6.14	Interpretations of CA.	98
6.15	Surface fitting to interpret ordinations.	99
6.16	Relevés chosen by RANK for permanent investigation.	106
7.1	Distinctness of group structure.	113
7.2	Ordination of group structure in data set 'nveg'.	123
7.3	Biplot and correlogram of 10 pH measurements.	128

7.4	Projecting distances in different directions.	130
7.5	Evaluating the direction of the main floristic gradient.	131
7.6	Correlograms of site factors with vegetation.	133
7.7	Comparison of RDA and CCA.	139
7.8	Using distance matrices in NP-MANOVA.	143
7.9	Graphical display of vegetation tables.	149
7.10	Structuring the meadow data set of Ellenberg.	153
8.1	Pairwise plot of selected site variables.	158
8.2	Linear and logistic regression of pH and <i>Sphagnum recurvum</i> .	160
8.3	Occurrence of <i>Sphagnum recurvum</i> and prediction by GLM.	163
8.4	Prediction of <i>Sphagnum recurvum</i> by GAM.	165
8.5	Regression tree to predict <i>Sphagnum recurvum</i> by pH.	167
8.6	Predicting <i>Sphagnum recurvum</i> occurrence by classification tree.	169
8.7	Scenarios for predicting <i>Sphagnum recurvum</i> occurrence.	171
8.8	Multivariate logistic regression.	173
8.9	Simulated wetland vegetation.	178
8.10	Occurrence probability of three species.	181
8.11	Steps of computation in multinomial logistic regression.	182
9.1	Type of environmental study needed to assess change.	186
9.2	Temporal arrangement of measurements (pH).	186
9.3	Measuring rate of change in time series of multistate systems.	189
9.4	Ordination of data from plots in the Swiss National Park.	190
9.5	Rate of change in plot Tr6, Swiss National Park.	192
9.6	A Markov model of the Lippe <i>et al.</i> (1985) data set.	197
9.7	PCA ordination of the Lippe succession data.	198

9.8	Markov model of the time series of the Swiss National Park.	200
9.9	The principle of space-for-time substitution.	201
9.10	The similarity of time series.	202
9.11	<i>Pinus mugo</i> on a former pasture in the Swiss National Park.	204
9.12	Minimum spanning tree (Swiss National Park).	204
9.13	Ordering of 59 time series from the Swiss National Park.	205
9.14	Succession in pastures of the Swiss National Park.	206
9.15	Tree species in a pollen diagram (Lotter 1999).	207
9.16	Velocity profile of the Soppensee pollen diagram.	208
9.17	Time trajectory of the Soppensee pollen diagram.	209
9.18	Velocity profiles from quantitative towards qualitative content.	210
9.19	Time acceleration trajectory of the Soppensee pollen diagram.	211
10.1	Attempt to get a dynamic model under control (Wildi 1976).	214
10.2	Numerical integration of the exponential growth equation.	216
10.3	Logistic growth of two populations, model 1.	218
10.4	Logistic growth of two populations, model 2.	219
10.5	Logistic growth of two populations, model 3.	220
10.6	The mechanism of spatial exchange.	223
10.7	Overgrowth of a plot by a new guild.	224
10.8	Original and simulated temporal succession.	227
10.9	Spatial design of SNP model.	229
10.10	Spatial simulation of succession, Alp Stabelchod.	230

11.1	Alliances represented in a wetland vegetation sample.	237
11.2	Frequency distribution of nearest-neighbor pairs of relevés.	239
11.3	Ordination of mire vegetation with and without outliers.	240
11.4	Projecting a given sample into a new resemblance space.	242
11.5	Ordination of the wetland sample in the indicator space.	243
11.6	Indicator values superimposed on ordination.	244
11.7	Similarity matrices of 12 vegetation types.	247
11.8	Synoptic table of mire vegetation data, outliers removed.	249
11.9	Synoptic table of mire vegetation data, outliers not removed.	250
12.1	Two ordinations of the Swiss forest data set.	259
12.2	Vegetation map of Swiss forests (eight groups).	262
12.3	The effect of different plot size on similarity pattern.	266
12.4	Vegetation probability map (eight groups).	269
12.5	Observed and potential distribution of four tree species.	272
12.6	Ordination of forest stands. Four selected tree species marked.	274
12.7	Ecograms of forest stands. Four selected tree species marked.	275
12.8	Tree- and herb layers of three species in ecological space.	278

List of tables

2.1	Terms used in sampling design (International Statistical Institute 2009).	13
3.1	Effects of different vector transformations.	30
3.2	Numerical example of vector transformation.	31
3.3	Transformation of cover-abundance values in phytosociology.	34
4.1	Notations in contingency tables.	43
4.2	Resemblance measures using the notations in Table 4.1.	44
4.3	Product moments.	46
5.1	Properties of four average linkage clustering methods.	60
6.1	Data set and results illustrating the RANK algorithm.	101
6.2	Ranking relevés of the ‘Schlaenggli’ data set.	105
6.3	Ranking species of the ‘Schlaenggli’ data set.	107
7.1	Synoptic table of <code>nveg</code> and <code>snit</code> .	112
7.2	Variance ranking of species.	116
7.3	Variance ranking of site factors.	118
7.4	Ranking of species by indicator values.	119
7.5	Mantel correlogram.	129
7.6	Mantel test of the site factors.	132

7.7	Storage location of parameters from functions <code>rda()</code> and <code>cca()</code> .	141
7.8	Choosing data transformation and distance function.	145
7.9	Steps involved in sorting synoptic tables.	148
7.10	Frequency table of structured synoptic vegetation table.	152
8.1	Input and output data of multivariate logistic regression.	173
8.2	Group means and standard deviations of pH and water level.	177
9.1	Temporal autocorrelation in a time series.	187
9.2	Markov process, measured and modeled data.	194
10.1	The effect of time step length in numerical integration.	216
10.2	Initial values in the temporal model SNP.	226
10.3	Six discrete vegetation states used as initial conditions.	229
11.1	Numbers and names of alliances.	238
11.2	Frequency table of data as displayed in Figure 11.9.	251
12.1	Data sets used in Chapter 12.	259
12.2	Composition of eight vegetation types.	260
12.3	Frequencies of tree species in data sets of different scale.	263
12.4	<i>F</i> -values of site factors based on eight forest vegetation types.	267
12.5	Multinomial models with different relevé plot size.	270
12.6	Tree species frequencies in different vegetation layers.	277
A.1	Main functions in the R package <code>dave</code> .	293
B.1	Data sets included in the R package <code>dave</code> .	295

About the companion website

This book is accompanied by a companion website:

www.wiley.com/go/wildi/dataanalysis

The website includes:

- Powerpoints of all figures from the book for downloading
- PDFs of tables from the book
- Links to the associated statistical package

