

# Multivariate Analysis of Ecological Data

**MICHAEL GREENACRE**

Professor of Statistics at the Pompeu Fabra University in Barcelona, Spain

**RAUL PRIMICERIO**

Associate Professor of Ecology, Evolutionary Biology and Epidemiology  
at the University of Tromsø, Norway

---

## Appendix B Offprint

## Bibliography and Web Resources

First published: December 2013

ISBN: 978-84-92937-50-9

Supporting websites:

[www.fbbva.es](http://www.fbbva.es)

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

© the authors, 2013

© Fundación BBVA, 2013

Fundación **BBVA**



## Bibliography and Web Resources

This appendix lists various bibliographical resources, with short annotations, for further reading. In addition, some web resources are given for supporting information and material such as R software and tutorials.

### Contents

Annotated bibliography .....	293
Web resources .....	301

#### *Study design and data analysis*

#### Annotated bibliography

ANDERSON D.R., K.P. BURNHAM, W.R. GOULD, and S. CHERRY. “Concerns about finding effects that are actually spurious”. *Wildlife Society Bulletin* 29 (2001): 311-316.

*(The authors discuss the characteristics of research studies that are more exposed to the risk of finding spurious effects, and propose various ways to cope with the problem and avoid spurious results).*

BENINGER P.G., I. BOLDINA, and S. KATSAKENEVAKIS. “Strengthening statistical usage in marine ecology”. *Journal of Experimental Marine Biology and Ecology* 426-427 (2012): 97-108.

*(A review of common statistical fallacies in the ecological literature and how to avoid them).*

COTTINGHAM K.L., J.T. LENNON, and B.L. BROWN. “Knowing when to draw the line: designing more informative ecological experiments”. *Frontiers in Ecology and the Environment* 3 (2005): 145-152.

*(Review of experimental design options for ANOVA and regression types of ecological studies).*

DAY R.W., and G.P. QUINN. “Comparisons of treatments after an analysis of variance in ecology”. *Ecological Monographs* 59 (1989): 433-463.

*(A review of approaches for comparisons of treatments following an ANOVA, including parametric and nonparametric tests, with discussion of pitfalls and solutions when dealing with hypothesis testing under unplanned multiple comparisons).*

GRAHAM M.H., and M.S. EDWARDS. “Statistical significance versus fit: estimating the importance of individual factors in ecological analysis of variance”. *Oikos* 93 (2001): 503-515.

*(The importance of effect size estimation and the available tools for variance decomposition in the context of complex ANOVA designs are presented clearly and succinctly).*

MAINDONALD J. *The design of research studies – A statistical perspective. Part I: planning and reporting*, 2000, 120 p.

[https://digitalcollections.anu.edu.au/bitstream/1885/41533/2/GS00\\_2.pdf](https://digitalcollections.anu.edu.au/bitstream/1885/41533/2/GS00_2.pdf)

*(A very informative introduction to the design of experimental and observational studies).*

NAKAGAWA S., and I.C. CUTHILL. “Effect size, confidence interval and statistical significance: a practical guide for biologists”. *Biological Reviews* 82 (2007): 591-605.

*(An indispensable review of effect size estimation. R code that allows to perform the analyses discussed in the paper is available and can be downloaded at:*

*<http://www.bristol.ac.uk/biology/research/staff/cuthill.i>).*

NAKAGAWA S., and R.P. FRECKLETON. “Missing inaction: the dangers of ignoring missing data”. *Trends in Ecology and Evolution* 23: 592-596.

*(The authors warn against deletion of cases with missing observations due to the ensuing reduced statistical power and increased estimation bias, and provide a compact review of how to deal properly with missing data).*

PARKHURST D.F. “Statistical significance tests: equivalence and reverse tests should reduce misinterpretation”. *Bioscience* 51 (2001): 1051-1057.

*(A gentle introduction to the concepts and methods of equivalence and reverse testing to help avoid pitfalls of results interpretation in classical null statistical hypothesis testing).*

QUINN G., and M. KEOUGH. *Experimental Design and Data Analysis for Biologists*. Cambridge, UK: Cambridge University Press, 2002.

*(Introductory textbook to study design and data analysis. Popular in courses at undergraduate and graduate level in experimental study design and ecological statistics. Useful background material to refresh ideas while reading Multivariate Analysis of Ecological Data).*

REGAN H.M., M. COLIVAN, and M.A. BURGMAN. "A taxonomy and treatment of uncertainty for ecology and conservation biology". *Ecological Applications* 12 (2002): 618-628.  
(A thorough discussion of sources of uncertainty in ecology and how to deal with them).

SCHEINER S.M., and J. GUREVITCH. *Design and Analysis of Ecological Experiments*. Oxford: Oxford University Press, 2001.

(A valuable collection of chapters by several authors dealing with study design, statistical modelling, spatial data analysis and meta-analysis).

WARTON, D.I., and F.K.C. HUI. "The arcsine is asinine: the analysis of proportions in ecology". *Ecology* 92 (2011): 3-10.

(A brief review of useful transformations for proportions with some warnings against established traditions when dealing with this type of data).

#### Statistical modelling

BOLKER B.M. *Ecological Models and Data in R*. Princeton, New Jersey: Princeton University Press, 2008.

(A gentle introduction to ecological modelling with clear and well structured coverage of maximum likelihood models and estimation).

BOLKER B.M., M.E. BROOKS, C.J. CLARK, S.W. GEANGE, J.R. POULSEN, M.H. STEVENS, and J.S. WHITE. "Generalized linear mixed models: a practical guide for ecology and evolution". *Trends in Ecology and Evolution* 24 (2009): 127-135.

(The paper reviews how to deal with nonnormal data that include random effects with the help of generalized linear mixed models).

CLARK J.S. *Models for Ecological Data: an Introduction*. Princeton, New Jersey: Princeton University Press, 2007.

(Rigorous and rich introduction to statistical modelling, including approaches to temporal and spatial data. A companion lab manual provides examples using R).

GRUEBER C.E., S. NAKAGAWA, R.L. LAWS, and I.G. JAMIESON. "Multimodel inference in ecology and evolution: challenges and solutions". *Journal of Evolutionary Biology* 24 (2011): 699-711.

(A comprehensive review of model selection and multimodel inference introducing basic concepts and approaches in a clear and balanced way).

HILBORN R., and C. MANGEL. *The Ecological Detective: Confronting Models with Data*. Princeton, New Jersey: Princeton University Press, 1997.

*(The book provides a very good introduction to theoretical and statistical modelling in ecology, explaining concepts, principles and protocols to the uninitiated).*

HOBBS N.T., and R. HILBORN. "Alternatives to statistical hypothesis testing in ecology: a guide to self teaching". *Ecological Applications* 16 (2006): 5-19.

*(A clear and concise introduction to statistical modelling, maximum likelihood estimation, model selection, Bayesian analysis and meta-analysis).*

STEPHENS P.A., S.W. BUSKIRK, and C.M. DEL RIO. "Inference in ecology and evolution". *Trends in Ecology and Evolution* 22 (2007): 192-197.

*(The authors discuss the limitations of traditional null hypothesis significance tests and suggest to rely on more useful approaches, briefly reviewed, such as effect size estimation and model selection).*

#### *Multivariate analysis*

ANDERSON M.J. "Permutation tests for univariate or multivariate analysis of variance and regression". *Canadian Journal of Fisheries and Aquatic Sciences* 58 (2001): 626-639.

*(An informative and concise review of rationale and applications of permutation tests in experimental and observational studies with complex designs).*

ANDERSON M.J., and T.J. WILLIS. "Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology". *Ecology* 84 (2003): 511-525.

*(A flexible method for constrained ordination capable of accommodating any distance or dissimilarity matrix).*

BEALS E.W. "Bray-Curtis ordination: an effective strategy for analysis of multivariate ecological data". *Advances in Ecological Research* 14 (1984): 1-55.

*(A good introduction to Bray-Curtis (or polar) ordination, also covering other methods).*

BORCARD D., F. GILLET, and P. LEGENDRE. *Numerical Ecology with R*. New York: Springer, 2011.

*(Compact introduction to multivariate statistics, including multivariate analysis of spatial and temporal data. It is the R companion to Numerical Ecology by Legendre and Legendre, 2012).*

GAUCH H. G. Jr. *Multivariate Analysis in Community Ecology*. Cambridge, United Kingdom: Cambridge University Press, 1982.

*(A thorough introduction to gradient analysis, it relates ecological theory and statistical methods clarifying the rationale behind the approach).*

GREENACRE M.J. *Correspondence Analysis in Practice, 2nd Edition*. London: Chapman & Hall/CRC, 2007. Free download of the Spanish edition, published by the BBVA Foundation, 2008, at [www.multivariatestatistics.org](http://www.multivariatestatistics.org).

*(Comprehensive introduction to correspondence analysis, multiple correspondence analysis, subset correspondence analysis and canonical correspondence analysis).*

GREENACRE M.J. 2010. *Biplots in Practice*. Madrid: BBVA Foundation, 2010. Free download from [www.multivariatestatistics.org](http://www.multivariatestatistics.org).

*(A practical introduction to biplots, the concept underlying many multivariate methods that reduce dimensionality in large data sets, and visualize the results).*

GREENACRE M.J. "Correspondence analysis of raw data". *Ecology* 91 (2010): 958-963.

*(Alternative approach to analysing abundance or biomass matrices where the data are not expressed relative to the row and column margins, in contrast to regular CA and CCA where relative amounts are analysed).*

GREENACRE M.J. "The contributions of rare objects in correspondence analysis". *Ecology* 94 (2013): 241-249.

*(Shows that CA and CCA are not unduly affected by the presence of rare species in an ecological data set, contrary to a popular misconception that these analyses are over-sensitive to species that occur sparsely and in low abundance).*

GREENACRE M.J. "Fuzzy coding in constrained ordinations". *Ecology* 94 (2013): 280-286.

*(The use of fuzzy coding for explanatory variables in the CCA context, demonstrating the benefits and also how to choose the number of fuzzy categories).*

GREENACRE M.J. "Contribution biplots". *Journal of Computational and Graphical Statistics* 22 (2013): 107-122.

*(An alternative scaling of the results of ordination methods such as PCA, CA, LRA, CCA and RDA, where the variables that contribute most to the solution are immediately detectable in the ordination).*

GREENACRE, M.J., and P.J. LEWIS. "Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio-scale measurements". *Journal of Classification* 26 (2009): 29-54.

*(Demonstrates clearly the advantage of weighting of variables in log-ratio analysis, as is done in regular CA, as well as the ability of log-ratio biplots to diagnose multiplicative models when variables line up in the ordinations).*

JACKSON D.A. “Stopping rules in principal component analysis: a comparison of heuristical and statistical approaches”. *Ecology* 74 (1993): 2204-2214.

(*A short introduction to some available options for evaluating the significance of principal components*).

JAMES F.C. “Multivariate analysis in ecology and systematics: panacea or pandora’s box?”. *Annual Review in Ecology and Systematics* 21 (1990): 129–166.

(*An early review of multivariate statistical applications in ecology, at a time when increased computer and software availability made these methods available to all ecologists*).

JOHNSON, R.A., and D.W. WICHERN. *Applied Multivariate Statistical Analysis, 6th edition*. New Jersey: Prentice Hall, 2007.

(*Widely read book reviewing applications of multivariate methods for biologists, physicists and sociologists*).

JONGMAN R.H.G., C.J.F. TER BRAAK, and O.F.R. VAN TONGEREN. *Data Analysis in Community and Landscape Ecology*. Cambridge, United Kingdom: Cambridge University Press, 1995.

(*A balanced treatment of ecological data analysis including extensive treatment of multivariate methods by a group of statisticians and ecologists with a strong quantitative background*).

LEGENDRE P., and M.J. ANDERSON. “Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments”. *Ecological Monographs* 69 (1999): 1-24.

(*Introduction to the flexible tool of distance-based redundancy analysis by the authors that eventually developed and generalized further this useful numerical approach*).

LEGENDRE P., and E.D. GALLAGHER. “Ecologically meaningful transformations for ordination of species data”. *Oecologia* 129 (2001): 271-280.

(*Review of effective transformation for species data that allow to extract relevant information when subject to ordination analysis*).

LEGENDRE P., and L. LEGENDRE. *Numerical Ecology, 3rd English edition*. Amsterdam: Elsevier, 2012, 853 p.

(*Classic introduction to statistics for ecologists with very good coverage of ecological data and multivariate methods, including an exhaustive treatment of distance and dissimilarity measures. It has a R companion [Borcard et al. 2011]*).

LEPŠ J., and P. ŠMILAUER. *Multivariate Analysis of Ecological Data using Canoco*. Cambridge United Kingdom: Cambridge University Press, 2003.



*(Much more than a handbook for Canoco applications, the book is an informative review of multivariate methods with many inspiring ecological examples).*

MAINDONALD J., and J. BRAUN. *Data Analysis and Graphics with R. An Example Based Approach, 3rd edition*. Cambridge, United Kingdom: Cambridge University Press, 2011.

*(The book provides a comprehensive overview of data analysis including parametric and nonparametric methods, statistical modelling and multivariate methods with R examples).*

MANLY B.F.J. *Multivariate Statistical Methods: a Primer, 3rd edition*. London: Chapman and Hall, 2004.

*(A gentle introduction to multivariate methods blessed by the clear expository style of a distinguished and successful author).*

MANLY B.F.J. *Randomization, Bootstrap and Monte Carlo Methods in Biology, 3rd edition*. London: Chapman and Hall, 2007.

*(The book provides a comprehensive overview of resampling and permutation methods with many relevant biological example applications).*

MCGARIGAL K., S. CUSHMAN, and S. STAFFORD. *Multivariate Statistics for Wildlife and Ecology Research*. New York: Springer, 2000.

*(Introduction to multivariate statistics in ecology and wildlife management, focusing on practical applications).*

PALMER M.W. "Putting things in even better order: the advantages of canonical correspondence analysis". *Ecology* 74 (1993): 2215-2230.

*(A clear exposition of the advantages of canonical correspondence analysis (CCA) applied to ecological data).*

PERES-NETO P.R. "How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test". *Oecologia* 129 (2001): 169-178.

*(The use and value of Procrustean superimposition to compare (match) multivariate data sets).*

PERES-NETO P.R., and D.A. JACKSON. "The importance of scaling of multivariate analysis in ecological studies". *Ecoscience* 8 (2001): 522-526.

*(A clear survey of the role of scaling in multivariate ecological data analysis, making use of intuitive graphical presentations to stress the important concepts and their implications).*

PERES-NETO P.R., D.A. JACKSON, and K.M. SOMERS. “Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis”. *Ecology* 84 (2003): 2347-2363.

*(The authors compare a variety of approaches for assessing the significance of eigenvector coefficients in terms of type I error rates and power).*

PIELOU, E.C. *The Interpretation of Ecological Data: a Primer on Classification and Ordination*. New York: John Wiley & Sons, Inc., 1984.

*(An early review of ecological data analysis linking ecological and statistical concepts to ease the interpretation of results).*

TER BRAAK, C.J.F. “Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis”. *Ecology* 67 (1986): 1167-1179.

*(Another citation classic for an author that has made important contributions to the field, ensuring a wide availability of the new methods via software development (Canoco, in collaboration with Šmilauer – see Lepš and Šmilauer 2003)).*

TER BRAAK, C.J.F., and I.C. PRENTICE. “A theory of gradient analysis”. *Advances in Ecological Research* 18 (1988): 271-313.

*(A classic introduction to gradient analysis theory and its ecological applications).*

WIEDMANN, M., M. ASCHAN, G. CERTAIN, A. DOLGOV, M. GREENACRE, E. JOHANNESSEN, B. PLANQUE, and R. PRIMICERIO. “Functional diversity of the Barents Sea fish community”. *Marine Ecology Progress Series*, in press, doi: 10.3354/meps10558.

*(A more extensive analysis of functional traits of Barents Sea fish species, compared to our case study of Chapter 20, using more fish species and a more recent time series).*

YEE T.W. “Constrained additive ordination”. *Ecology* 87 (2006): 203-213.

*(The paper introduces constrained additive ordination (CAO) models, described as “loosely speaking, [...] generalized additive models fitted to a very small number of latent variables”. The paper provides the R code to implement the CAO methodology with some clear example applications).*

ZUUR A.F., E.N. IENO, and C.S. ELPHICK. “A protocol for data exploration to avoid common statistical problems”. *Methods in Ecology and Evolution* 1 (2010): 3-14.

*(The authors provide a protocol for data exploration, discussing “current tools to detect outliers, heterogeneity of variance, collinearity, dependence of observations, problems with interactions, double zeros in multivariate analysis, zero inflation in generalized linear modelling, and the correct type of relationships between dependent and independent variables; and []*

*provide advice on how to address these problems when they arise“. The paper also provides R code to implement the protocol).*

ZUUR A.F., E.N. IENO, N.J. WALKER, A.A. SAVELIEV, and G.M. SMITH. *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer, 2009.

*(A very popular introduction to mixed effects models rich with relevant ecological examples based on the kind of “messy” data ecologists need to cope with).*

<http://www.multivariatestatistics.org>

Web resources

GREENACRE M.J. *Multivariate books, data sets and R scripts*.

*(This web site supports three books published by the BBVA Foundation: La Práctica del Análisis de Correspondencias (Spanish translation of Correspondence Analysis in Practice, Second Edition), Biplots in Practice and the present book. Full text, data sets and R scripts available for free download).*

<http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf>

OKSANEN J. *Multivariate analyses of ecological communities in R: vegan tutorial*. 43 p.

*(This compact tutorial to the R package vegan is a very useful, brief introduction to the application of multivariate statistics in community ecology).*

<http://ordination.okstate.edu>

PALMER M.W. *Ordination methods for ecologists*

*(The web site is a very useful overview of ordination methods, providing links, references, and guides for self study, including a much needed glossary).*