0.1 Introduction

Third edition

The third edition of this unified framework for non-parametric analysis of multivariate data, underlying the PRIMER software package, has the same form and similar chapter headings to its predecessor (with an additional chapter). However, the text has been much expanded to include full cover of methods that were implemented in PRIMER v6 but only described in the PRIMER v6 User Manual, and also the entire range of new methods contained in PRIMER v7.

Whilst text has been altered throughout, PRIMER v6 users familiar with the 2nd edition, who just want to locate the new material, will find it below:

Table 0.1. Manual pages primarily covering new material

Image not found or type unknown **Topics**	**Pages**
Additions to the framework	1.7
Missing data and variable weightings	2.6
Similarity profile tests (SIMPROF) of clusters on sample dendrograms	3.5
Unconstrained binary divisive (UNCTREE) and fixed group (k-R) clustering	3.6, 3.7
More nMDS diagnostics (MST, similarity joins, 3-d cluster on MDS, scree plots)	5.3, 5.7
Metric MDS (mMDS), threshold MDS	5.8
Combined MDS ('fix collapse' by nMDS + mMDS, composite biotic/abiotic nMDS)	5.9, 5.10
ANOSIM for ordered factors	6.10 to 6.13
3-way ANOSIM designs	6.14 to 6.17
Species Analyses (new chapter, in effect): *SIMPROF on species (coherent curves)* *Shade plots (+dendrograms, axes orders)* *Bubble plots (for groups, segmented)*	7.1 to 7.3 7.4 to 7.7 7.10
Testing curves (dominance/particle/growth)	8.5, 8.6
Analysing multiple diversity indices	8.7
Dispersion weighting	9.5, 9.6

Image not found or type unknown **Topics**	**Pages**
Vector plots in PCA and MDS	11.2, 11.3
Global BEST test (allowing for selection) *and constrained BEST analyses*	11.4
Linkage trees: binary clusters, constrained by abiotic 'explanations' (LINKTREE)	11.6
Model matrices, RELATE tests of seriation and cyclicity, constrained RELATE	15.5, 15.6
Second-stage analysis (2STAGE) *Zero-adjusted Bray-Curtis for sparse data* *Defining and comparing resemblances* *Second-stage 'interaction' plots*	16.6 16.6 16.7
Taxonomic (relatedness-based) dissimilarity	17.11, 17.12
Means plots & 'bootstrap average' regions	18.1 to 18.5

Attribution (and responsibility for queries)

These new sections have all been authored by KRC but build heavily on collaborations, joint publications and novel algorithmic and computer coding work with/by PJS and RNG. In the retained material from the 2nd edition (authored by KRC and RMW), KRC was largely responsible for Chapters 1-7, 9, 11 and 16 and RMW for 10 and 12-14, with the responsibility for Chapters 8, 15 and 17 shared between them.

Purpose

This manual accompanies the computer software package PRIMER (Plymouth Routines In Multivariate Ecological Research), obtainable from PRIMER-e, (see www.primer-e.com). Its scope is the analysis of data arising in community ecology and environmental science which is multivariate in character (many species, multiple environmental variables), and it is intended for use by ecologists with no more than a minimal background in statistics. As such, this *methods* manual complements the PRIMER *user* manual, by giving the background to the statistical techniques employed by the analysis programs (Table 0.2), at a level of detail which should allow the scientist to understand the output from the programs, be able to describe the results in a non-technical way to others and have confidence that the right methods are being used for the right problem.

This may seem a tall order, in an area of statistics (primarily multivariate analysis) which has a reputation as esoteric and mathematically complex! However, whilst it is true that the computational details of some of the core techniques described here (for example, non-metric multidimensional scaling) are decidedly non- trivial, we maintain that all of the methods that have been adopted or developed within PRIMER are so *conceptually* straightforward as to be amenable to simple explanation and transparent interpretation. In fact, the adoption of non-parametric and permutation approaches for display and testing of multivariate data requires, paradoxically, a lower

level of statistical sophistication on the part of the user than does a satisfactory exposition of classic (parametric) hypothesis testing in the *univariate* case.

Table 0.2. Chapters in this manual in which the methods underlying specific PRIMER routines are principally found.¶

Image not found or type unknown **Routines**	**Chapters**
Resemblance	
(Dis)similarity/distance for samples	2
Association index for species	2, 7
Dummy variables (zero-adjusted coefficient)	16
Taxonomic dissimilarities	17
Cluster	
CLUSTER (hierarchical: agglomerative)	3
LINKTREE (" :constrained divisive)	11
UNCTREE (" :unconstrained divisive)	3
kRCLUSTER (non-hierarchical)	3
Clustering variables (species)	7
SIMPROF	
tests for sample groups from Cluster	3, 11
tests for species groups	7
Coherence plots (Line plots)	7
PCA (+ Vector plot)	4, 11
MDS	
Non-metric, Metric, Combined, Fix collapse	5
Shepard diagram, Scree plot	5
Overlay clusters, trajectory, MST, join pairs	5
Vector plot	7, 11
Bubble plots (groups, multiple)	6, 7, 11
ANOSIM (1/2/3-way, crossed/nested, ordered)	6
SIMPER	7
Shade Plot (Matrix display)	7
Diversity indices	
DIVERSE	8, 17
CASWELL, Geometric Class Plots	8
Dominance Plots, DOMDIS	8
Species Accumulation Plots	17
TAXDTEST, histogram/funnel/ellipse plots	17

Image not found or type unknown **Routines**	**Chapters**
Pre-treatment *Transform, Standardise* *Normalise Variables* *Cumulate Samples* *Dispersion Weighting, Variability Weighting*	2, 7, 9 2, 4, 11 8 9
Aggregate	10, 16
BEST *BIO-ENV, Draftsman Plot* *Constrained BEST (Within factor levels)* *BVSTEP, Global BEST test*	11 11 11, 16
MVDISP	15
RELATE (Seriation, Cyclicity, Model Matrix)	15
2STAGE (Single and Multiple matrices)	16
Bootstrap Averages	18

¶PRIMER has a range of other data manipulation and plotting routines: Select, Edit, Summary stats, Average, Sum, Transpose, Rank, Merge, Missing data and Bar/Box/Means/Scatter/Surface/ Histogram Plots, etc – see the PRIMER User Manual/Tutorial.*

One primary aim of this manual is therefore to describe a coherent strategy for the interpretation of data on community structure, namely values of abundance, biomass, % cover, presence/absence etc. for a set of 'species' variables and one or more replicate samples which are taken:

a) at a number of sites at one time (spatial analysis);

b) at the same site at a number of times (temporal analysis);

c) for a community subject to different uncontrolled or controlled manipulative 'treatments';

or some combination of these.

These species-by-samples arrays are typically quite large, and usually involve many variables (p species, say) so that the total number (n) of observed samples can be considered to be n points in high-dimensional (p-dimensional) space. Classical statistical methods, based on multivariate normality are often impossible to reconcile with abundance values which are predominantly zero for many species in most samples, making their distributions highly right-skewed. Even worse, classic methods require that n is much larger than p in order to have any hope of estimating the parameters (unknown constants, such as means and variances for each species, and correlations between species) on which such parametric models are based.

Statistical testing therefore requires methods which can represent high-dimensional relationships among samples through similarity measures between them, and test hypotheses without such model assumptions (non-parametrically within PRIMER by permutation). A key feature is that testing must be carried out on the similarities, which represent the true relationships among samples (in the high-d space), rather than on some lower-dimensional approximation to this high-d space, such as a 2- or 3-d 'ordination'.

Data visualisation, however, makes good use of such low-dimensional ordinations to view the approximate biological relationships among samples, in the form of a 'map' in 2- or 3-d. Patterns of distance between sample points in that map should then reflect, as closely as possible, the patterns of biological dissimilarity among samples. Testing and visualisation are therefore used in conjunction to identify and characterise changes in community structure in time or space, and in relation to changing environmental or experimental conditions.

Scope of techniques

It should be made clear at the outset that the title 'Change in Marine Communities' does not in any way reflect a restriction in the scope of the techniques in the PRIMER package to the marine environment. The first edition of this manual *was* intended primarily for a marine audience and, given that the examples and rationale are still largely set around the literature of marine ecology, and some of the original chapters in this context have been retained, it seems sensible to retain the historic continuity of title. However, it will soon be evident to the reader that there is rather little in the *methods* of the following pages that is exclusively marine or even confined to ecology. In fact, the PRIMER package is now not only used in over 120 countries world-wide (and in all US states) for a wide range of marine community surveys and experiments, of benthic fauna, algae, fish, plankton, corals, dietary data etc, but is also commonly found in freshwater & terrestrial ecology, palaeontology, agriculture, vegetation & soil science, forestry, bio-informatics and genetics, microbiology, physical (remote sensing, sedimentary, hydrological) and chemical/biochemical studies, geology, biogeography and even in epidemiology, medicine, environmental economics, social sciences (questionnaire returns), on ecosystem box model outputs, archaeology, and so ons.

Indeed, it is relevant to any context in which multiple measurement variables are recorded from each sample unit (the definition of *multivariate* data) and classical multivariate statistics is unavailable, i.e. especially (as intimated above) where there are a large number of variables in relation to the number of samples (and in microbial/genetic studies there can be many thousands of bands with intensities measured, from each sample), or characterised by a presence/absence structure in which the information is contained at least partly in pattern of the presences of non-zero readings, as well as their actual values (in other words, data for which zero is a 'special' number).

As a result of the authors' own research interests and the widespread use of community data in pollution monitoring, a major thrust of the manual is the biological effects of contaminants but, again, most of the methods are much more generally applicable. This is reflected in a range of more fundamental ecological studies among the real data sets exemplified here.

The literature contains a large array of sophisticated statistical techniques for handling species-by-samples matrices, ranging from their reduction to simple diversity indices, through curvilinear or

distributional representations of richness, dominance, evenness etc., to a plethora of multivariate approaches involving clustering or ordination methods. This manual does *not* attempt to give an overview of all the options. Instead it presents a strategy which has evolved over decades within the Community Ecology/Biodiversity groups at Plymouth Marine Laboratory (PML), and subsequently within the 'spin-out' PRIMER-E Ltd company, and which has now been tested for ease of understanding and relevance to analysis requirements at well over 100 practical 1-week training workshops.

The workshop content has continued to evolve, in line with development of the software, and the utility of the methods in interpreting a range of community data can be seen from the references listed under Clarke, Warwick, Somerfield or Gorley in Appendix 3, which between them have amassed a total of >20,000 citations in SCI journals. The analyses and displays in these papers, and certainly in this manual, have very largely been accomplished with routines available in PRIMER (though in many cases annotations etc have been edited by simply copying and pasting into graphics presentation software such as Microsoft Powerpoint).

Note also that, whilst other software packages will not encompass this specific combination of routines, several of the individual techniques (though by no means all) can be found elsewhere. For example, the core clustering and ordination methods described here are available in many mainstream statistical packages, and there are at least two other specialised statistical programs (CANOCO and PC-ORD) which tackle essentially similar problems, though usually employing different techniques and strategies; other authors have produced freely-downloadable routines in the R statistical framework, covering some of these methods.

This manual does *not* cover the PERMANOVA+ routines, which are available as an add-on to the PRIMER package. The PERMANOVA+ software has been further developed and fully coded by PRIMER-E (in the Microsoft Windows '.Net' framework of all recent PRIMER versions) in very close collaboration with their instigator, Prof Marti Anderson (Massey University, NZ). These methods complement those in PRIMER, utilising the same graphical/data-handling environment, moving the emphasis away from non-parametric to semi-parametric (but still permutation based and thus distribution-free) techniques, which are able to extend hypothesis testing for data with more complex, higher-way designs (allowing, for example, for concepts of fixed *vs* random effects, and factor partitioning into main effect and interaction terms). This, and several other analyses which more closely parallel those available in classical univariate contexts, but are handled by permutation testing, are fully described in the combined Methods and User manual for PERMANOVA+, Anderson, Gorley & Clarke (2008) .

Example data sets

Throughout the manual, extensive use is made of data sets from the published literature to illustrate the techniques. Appendix 1 gives the original literature source for each of these 40 data sets and an index to all the pages on which they are analysed. Each data set is allocated a single letter designation (upper or lower case) and, to avoid confusion, referred to in the text of the manual by that letter, placed in curly brackets (e.g. *{A}* = Amoco-Cadiz oil spill, macrofauna; *{B}* = Bristol Channel, zooplankton; *{C}* = Celtic Sea, zooplankton, *{c}* = Creran Loch, macrobenthos etc). Many of these data sets (though not all) are made available automatically with the PRIMER software.

Literature citation

Appendix 2 lists some background papers appropriate to each chapter, including the source of analyses and figures, and a full listing of references cited is given in Appendix 3. Since this manual is effectively a book, not accessible within the refereed literature, referral to the methods it describes should probably be by citing the primary papers for these methods (this will not always be possible, however, since some of the new routines in PRIMER v7 are being described here for the first time). Summaries of the early core methods in PRIMER for multivariate and univariate/graphical analyses are given respectively in Clarke (1993) and Warwick (1993) . Some primary techniques papers are: Field, Clarke & Warwick (1982) for clustering, MDS; Warwick (1986) and Clarke (1990) for ABC and dominance plots; Clarke & Green (1988) for 1-way ANOSIM, transformation; Warwick (1988b) and Olsgard, Somerfield & Carr (1997) for aggregation; Clarke & Ainsworth (1993) for BEST/ Bio-Env; Clarke (1993) and Clarke & Warwick (1994) for 2-way ANOSIM with and without replicates, similarity percentages; Clarke, Warwick & Brown (1993) for seriation; Warwick & Clarke (1993b) for multivariate dispersion; Clarke & Warwick (1998a) for structural redundancy, BEST/BVStep; Somerfield & Clarke (1995) and Clarke, Somerfield, Airoldi *et al.* (2006) for second-stage analyses; Warwick & Clarke (1995b) , Warwick (1988a) , Warwick & Clarke (2001) , Clarke & Warwick (1998b) , Clarke & Warwick (2001) for taxonomic distinctness; Clarke, Chapman, Somerfield *et al.* (2006) for dispersion weighting; Clarke, Somerfield & Chapman (2006) for resemblances and sparsity; Clarke, Somerfield & Gorley (2008) for similarity profiles and linkage trees; Clarke, Tweedley & Valesini (2014) for shade plots; and Somerfield & Clarke (2013) for coherent species curves.

---

§*The list seems endless: the most recent attempt to look at which papers have cited at least one of the PRIMER manuals, or a highly cited paper (Clarke (1993)) which lays out the philosophy and some core methods in the PRIMER approach, was in August 2012, and resulted in 8370 citations in refereed journals (SCI-listed), from 773(!) different journal titles. Of course, there is no guarantee that a paper citing the PRIMER manuals has used PRIMER – though most will have – but, equally, there are several score of PRIMER methods papers that may have been cited in place of the manuals, especially for the many PRIMER developments that have taken place since the Clarke (1993) paper, so the above citation total is likely to be a significant underestimate.*

Revision #88
Created 9 February 2022 08:32:10 by Arden
Updated 16 October 2024 01:50:33 by Marti