PRIMER v7 and PERMANOVA+ WORKSHOPS
Long Beach, California, USA
Multivariate Analysis in Ecology (& other Sciences)

Presenters: **Dr. Paul Somerfield** *and* **Dr. Adam N. H. Smith**
Venue: **California State University, Long Beach, California, USA**
Dates: **Week 1: January 7th - 11th, 2019** – **PRIMER v7** (Paul Somerfield)
Week 2: January 14th - 18th, 2019 – **PERMANOVA+** (Adam N. H. Smith)

PROVISIONAL SCHEDULE

WEEK 1

Monday, January 7th

08:15-08:30	Introduction
08:30-10:15	Lecture: *Measures of resemblance (similarity/dissimilarity/distance) in multivariate structure for assemblage and environmental data, including shade plots used to assess the effects of pre-treatment options (e.g. standardisation, transformation, normalisation, dispersion weighting), and guidelines for different coefficient choices for different data types, and experimental purpose and protocols*
10:15-10:30	**Coffee break**
10:30-12:15	Lecture: *Clustering of samples by hierarchical agglomerative (CLUSTER) and unconstrained divisive (UNCTREE) methods; also, non-hierarchical ('flat') clustering for specified numbers of groups. Includes discussion of a global test for the presence of any multivariate structure in a priori unstructured biotic or abiotic samples, using similarity profiles (Type 1 SIMPROF tests on samples)*
12:15-13:15	**Lunch break**
13:15-13:45	Introduction to PRIMER 7 software
13:45-15:15	Lab session on transforms, similarity options, alternative clustering methods and SIMPROF tests
15:15-15:30	**Coffee break**
15:30-16:30	Lecture: *Ordination (for environmental data) by Principal Components Analysis (PCA)*
16:30-17:15	Lab session on ensuring participants can structure their own data to take into PRIMER[1]

Tuesday, January 8th

08:15-09:15	Lab session on ordination by PCA
09:15-11:15	Lecture: *Ordination (of assemblage data) by non-metric Multi-Dimensional Scaling (nMDS,) and MDS diagnostics (e.g. stress, MST, cluster overlay) for adequacy of low-d representation. Also, how this relates, through the Shepard diagram, to metric MDS (mMDS), useful for abiotic data (Euclidean distances) and for means plots from biotic data with few points* **(includes coffee break c. 10:30-10:45**)
11:15-12:15	Lab session on ordination by nMDS (and mMDS), animating the convergence, examining Shepard plots and other diagnostics, and display tools on MDS plots
12:15-13:15	**Lunch break**
13:15-14:15	Continuation of the lab session on MDS
14:15-15:15	Lecture: *Global hypothesis tests of no agreement between two resemblance matrices (RELATE), comparing assemblage (or environmental) structure with linear or cyclic models in space and time*

---

[1] Throughout, participants will be given real data sets to analyse, but they may also wish to bring their own data. These should be in numeric, rectangular arrays, with variables (e.g. species) as rows, samples as columns, or vice-versa, in an Excel spreadsheet or text file. Non-numeric information (factors) on each sample are placed below (or to the side of) this table, separated by a blank row (or blank column). There is also a 3-column format (sample label, variable label, non-zero entry) suitable for entry from large record-type databases. Opportunity should be taken to discuss your data with the lecturer during the labs and breaks, prior to the Friday afternoon as well as in that final session.

15:15-15:30	**Coffee break**
15:30-17:15	Lab session on RELATE tests for simple seriation without replication and a cyclic model with replication, including 'fixing' MDS collapse and dispersion weighting

Wednesday, January 9th

08:15-09:15	Lecture: *Multivariate tests for differences among <u>a priori</u> groups of samples (1-way ANOSIM and pairwise). Multivariate means plots where global test has established differences. Brief discussion of approximate region estimates for means, in mMDS plots, using bootstrap averaging.*
09:15-10:15	Lab session on 1-way ANOSIM, means plots and mMDS of region plots for means
10:15-10:30	**Coffee break**
10:30-11:30	Lecture: *Ordered ANOSIM tests, 2- and 3-way ANOSIM designs and testing (unordered) factors in crossed ANOSIM designs without replication, by comparing patterns*
11:30-12:15	Lab session on ordered and multi-way ANOSIM
12:15-13:15	**Lunch break**
13:15-14:30	Lecture: *Linking potential environmental drivers to an observed assemblage pattern, via the matching of multivariate structures (the BEST procedure). Test of no evidence for a biota-environment link, allowing for selection effects in finding an optimum match (global BEST test)*
14:30-15:15	Lab session on PCA and the BEST routine for linking to environmental variables; global BEST test
15:15-15:30	**Coffee break**
15:30-16:30	Lecture: *Linkage trees – a further technique for 'explaining' assemblage patterns by environmental variables (LINKTREE), and its relation to unconstrained divisive clustering (UNCTREE)*
16:30-17:15	Lab session on link to abiotic variables (LINKTREE) and comparison to unconstrained cluster

Thursday, January 10th

08:15-09:15	Lecture: *Species contributions to sample patterns: stepwise form of BEST to identify minimal-sized species subsets reconstructing the full assemblage pattern (a whole pattern approach), and species contributions to similarities (SIMPER, a pairwise approach for statistically established groups)*
09:15-10:15	Lab session on various methods for identifying species contributions (Matrix display, stepwise BEST, bubble plots and SIMPER – the latter on groups from 'flat' clustering)
10:15-10:30	**Coffee break**
10:30-11:15	Lecture: *Direct analysis of species (or other variables) through species resemblances: techniques for identifying coherent groups of species (or other variables) in their response across samples*
11:15-12:15	Lab session on coherent variable sets
12:15-13:15	**Lunch break**
13:15-14:15	Lecture: *Diversity measures (DIVERSE) and multivariate treatment of multiple indices; dominance plots and testing differences between sets of curves (e.g. DOMDIS)*
14:15-15:15	Lab session on DIVERSE, dominance plots, and multivariate analyses of multiple diversity indices
15:15-15:30	**Coffee break**
15:30-16:15	Lecture: *Taxonomic (or phylogenetic) diversity and distinctness for quantitative data, or simple species lists, as valid biodiversity measures (DIVERSE) over broad spatial and temporal scales; sampling properties and testing structures (TAXDTEST)*
16:15-17:15	Lab session on TAXDTEST

Friday, January 11th

08:15-09:15	Lecture: *Second-stage analysis (2STAGE) to compare taxonomic levels and transformation etc; also for a possible testing framework in some repeated measures designs*
09:15-10:15	Lab session on 2STAGE
10:15-10:30	**Coffee break**
10:30-12:15	Lecture: *Any methods that have not arisen in earlier discussion (e.g. further resemblance options: modifying Bray-Curtis for denuded samples; resemblance calculations when some data are missing; perhaps dissimilarity measures based on taxonomic distinctness etc)*
12:15-12:15	Wrap up of Week 1; prologue to Week 2 – brief introduction to PERMANOVA+
12:15-13:15	**Lunch break**
13:15-17:15	Main lab session on analysing own data using PRIMER

Monday, January 14th

08:15-08:45	Introduction
08:45-09:45	Lecture: *Introduction to permutational multivariate analysis of variance (PERMANOVA); motivation and comparison with MANOVA and ANOSIM; distance-based partitioning of sums of squares (SS) for univariate and multivariate data*
09:45-10:15	Practical 1: Introduction to the PERMANOVA+ add-on for PRIMER; one-way design and comparison with ANOSIM results
10:15-10:30	**Coffee break**
10:30-11:45	Lecture: *Testing and interpreting two-way interactions for multivariate data; logical choices for pair-wise comparisons after tests for interaction; permutation methods for testing main effects and interaction terms; random unrestricted permutation, restricted permutation, permutation of residuals under a reduced model or under a full model*
11:45-12:15	Practical 2: Two-way analyses with PERMANOVA; testing and interpreting multivariate interactions; pair-wise comparisons; constructing specific a priori contrasts
12:15-13:15	**Lunch**
13:15-14:15	Lecture: *Consideration of dissimilarity measures and their properties: simple matching coefficient, Jaccard, Sørensen, Euclidean, chi-squared, Bray-Curtis, modified Gower*
14:15-15:15	Practical 3: Effects of different dissimilarity measures on perceptions of multivariate patterns. PERMDISP and tests of dispersion.
15:15-15:30	**Coffee break**
15:30-16:15	Lecture: *Permutational tests of homogeneity of multivariate dispersions (PERMDISP); permutation of residuals; tests to examine beta diversity; the effects of different dissimilarity measures on perceptions of relative within-group dispersions for count data and sparse data*
16:15-17:15	Practical 3, continued: PERMDISP on ecological data and tests of dispersions; analyses of beta diversity; the use of PERMDISP and PERMANOVA together for interpreting differences among groups in dispersion and/or location for two-way designs

Tuesday, January 15th

08:15-09:15	Lecture: *Fixed vs random factors; nested vs crossed relationships among factors; consequences for the expected mean squares (EMS) and construction of appropriate pseudo-F ratios; consequences for the hypotheses being tested and the extent of the inferences; multivariate analogues in PERMANOVA follow the univariate results*
09:15-10:15	Practical 4: Crossed designs; nested designs; fixed, random and mixed models; degrees of freedom and potential power; testing the design; choosing appropriate pair-wise comparisons to do after fitting and analysing the full model
10:15-10:30	**Coffee break**
10:30-12:15	Practical 5: Estimating components of variation; exchangeable units for permutation dictated by denominator mean squares; pooling or removing terms from a model; linear combinations of mean squares; how to tackle higher-way multi-factorial designs
12:15-13:15	**Lunch**
13:15-14:15	Lecture: *Principal coordinate analysis (PCO) for direct analyses of dissimilarity matrices, its uses, limitations and its relationship with PCA and metric/non-metric MDS*
14:15-15:15	Practical 6: Comparisons of PCA, PCO and MDS; negative eigenvalues and properties of distance measures; vector overlays (pros and cons)
15:15-15:30	**Coffee break**
15:30-17:15	Practical 7: Distances among centroids; choice of relevant ordination graphics to accompany and interpret multi-factorial analyses; Monte Carlo P-values; testing dispersions for a nested design

Wednesday, January 16th

08:15-09:15	Lecture: *Unbalanced designs; non-independence of terms in the model; Types of Sums of Squares; consequences for expectations of mean squares; designs that include covariates; interactions between covariates and ANOVA factors; tests, interpretations and inferences*

09:15-10:15	Practical 8: Examples of unbalanced designs and designs with covariates; consequences of different choices of Types of SS
10:15-10:30	**Coffee break**
10:30-11:45	Lecture: *Designs for detecting environmental impact; BACI and Beyond BACI; designs that lack replication; asymmetrical designs*
11:45-12:15	Practical 9: Randomized blocks and repeated measures; an asymmetrical 'ACI' design
12:15-13:15	**Lunch**
13:15-14:15	Lecture: *Continuous predictor variables; regression; least-squares solutions to the normal equations for linear models; multiple regression; marginal and sequential (conditional) tests*
14:15-15:15	Practical 10: Simple and multiple linear regression; choosing the order of terms; DISTLM dialogue
15:15-15:30	**Coffee break**
15:30-16:45	Lecture: *Multivariate multiple regression: Redundancy Analysis (RDA); analysing the relationship between species data and continuous (e.g. environmental) variables; duality leading to the distance-based linear model; procedures for model fitting (forward selection, backward elimination, step-wise selection and a 'best' procedure); model selection criteria (R^2, adjusted R^2, AIC, AICc, BIC).*
16:45-17:15	Practical 11: Diagnostics on predictor variables; fitting multivariate regression models and model selection using DISTLM

Thursday, January 17th

08:15-09:15	Lecture: *Visualizing regression models in a constrained ordination; distance-based redundancy analysis (dbRDA); interpretation of dbRDA axes; biplot vectors and their interpretation; constrained and unconstrained ordination; comparison of dbRDA with PCO*
09:15-10:15	Practical 12: DISTLM and dbRDA; comparison with BIOENV results
10:15-10:30	**Coffee break**
10:30-12:15	Practical 12, continued: Analysing variables in sets; categorical predictor variables
12:15-13:15	**Lunch**
13:15-14:15	Lecture: *Canonical analysis of principal coordinates (CAP); generalized discriminant analysis based on distances; finding axes through the cloud of points that best discriminate among groups; leave-one-out allocation success*
14:15-15:15	Practical 13: CAP and PCO for data with a priori groupings; understanding the diagnostics; interpreting the results of a CAP analysis; CAP as a predictive model; allocation of new (unknown or validation) samples to existing groups
15:15-15:30	**Coffee break**
15:30-16:15	Lecture: *Canonical analysis of gradients using CAP; leave-one-out residual SS; obtaining models of community change along environmental gradients; models of 'ecosystem health' and monitoring studies; prediction and placement of new points onto gradients using CAP models; canonical correlation and multiple X variables*
16:15-17:15	Practical 14: Predictive models of environmental pollution gradients using CAP; canonical correlation analysis

Friday, January 18th

08:15-17:15	Practical*: Overview and synthesis of the methods in PERMANOVA+ followed by an all-day lab session and analysis of participants' own data using PERMANOVA+ and PRIMER

_____

* Participants will be given real data sets to analyse in the lab sessions, to illustrate the main points but it is expected they will bring their own data to the workshop, to analyse during this session and throughout the week, whilst the lecturer is on hand to give advice. As for the PRIMER week, data should be in numeric, rectangular arrays, with variables (e.g. species) as rows, samples as columns, or vice-versa, in Excel or a text file. Non-numeric sets of information (factors) on each sample are placed below (or to the side of) this table, separated by a blank row (or blank column). There is also a 3-column format (sample label, variable label, non-zero entry) suitable for very large arrays.

Some key and well-cited papers on PRIMER and PERMANOVA+ methodology

PRIMER

Clarke KR (1990) Comparisons of dominance curves. *J Exp Mar Biol Ecol* 138: 143-157

Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* 18: 117-143

Clarke KR (1999) Non-metric multivariate analysis in community-level ecotoxicology. *Environ Toxicol Chem* 18: 118-127

Clarke KR, Ainsworth M (1993) A method of linking multivariate community structure to environmental variables. *Mar Ecol Prog Ser* 92: 205-219

Clarke KR, Chapman MG, Somerfield PJ, Needham HR (2006) Dispersion-based weighting of species counts in assemblage analyses. *Mar Ecol Prog Ser* 320: 11-27

Clarke KR, Gorley RN (2001, 2006, 2015) *PRIMER v5, v6, v7: User manual/tutorial.* PRIMER-E, Plymouth, UK, 91pp, 192pp, 296pp

Clarke KR, Green RH (1988) Statistical design and analysis for a 'biological effects' study. *Mar Ecol Prog Ser* 46: 213-226

Clarke KR, Somerfield PJ, Airoldi L, Warwick RM (2006) Exploring interactions by second-stage community analyses. *J Exp Mar Biol Ecol* 338: 179-192

Clarke KR, Somerfield PJ, Chapman MG (2006) On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages. *J Exp Mar Biol Ecol* 330: 55-80

Clarke KR, Somerfield PJ, Gorley RN (2008). Exploratory null hypothesis testing for community data: similarity profiles and biota-environment linkage. *J Exp Mar Biol Ecol* 366: 56-69

Clarke KR, Somerfield PJ, Gorley RN (2016) Clustering in non-parametric multivariate analyses. *J Exp Mar Biol Ecol* 483: 147-155.

Clarke KR, Tweedley JR, Valesini FJ (2014) Simple shade plots aid better long-term choices of data pre-treatment in multivariate assemblage studies. *J Mar Biol Assoc UK* 94: 1-16

Clarke KR, Warwick RM (1994, 2001, 2014) *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation.* PRIMER-E, Plymouth, UK. *1st ed:* 144pp; *2nd ed:* 172pp. *3rd ed:* (authors: Clarke KR, Gorley RN, Somerfield PJ, Warwick RM) 260pp

Clarke KR, Warwick RM (1998) Quantifying structural redundancy in ecological communities. *Oecologia* 113: 278-289

Clarke KR, Warwick RM (1998) A taxonomic distinctness index and its statistical properties. *J Appl Ecol* 35: 523-531

Clarke KR, Warwick RM (2001) A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Mar Ecol Prog Ser* 216: 265-278

Field JG, Clarke KR, Warwick RM (1982) A practical strategy for analysing multispecies distribution patterns. *Mar Ecol Prog Ser* 8: 37-52

Somerfield PJ, Clarke KR (1995) Taxonomic levels, in marine community studies, revisited. *Mar Ecol Prog Ser* 127: 113-119

Somerfield PJ, Clarke KR (2013) Inverse analysis in non-parametric multivariate analyses: distinguishing groups of associated species which covary coherently across samples. *J Exp Mar Biol Ecol* 449: 261-273

Somerfield PJ, Clarke KR, Olsgard F (2002) A comparison of the power of categorical and correlational tests applied to community ecology data from gradient studies. *J Anim Ecol* 71: 581-593

Warwick RM, Clarke KR (1991) A comparison of some methods for analysing changes in benthic community structure. *J Mar Biol Ass UK* 71: 225-244

Warwick RM, Clarke KR (1993) Increased variability as a symptom of stress in marine communities. *J Exp Mar Biol Ecol* 172: 215-226

Warwick RM, Clarke KR (1995) New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Mar Ecol Prog Ser* 129: 301-305

Warwick RM, Clarke KR (1998) Taxonomic distinctness and environmental assessment. *J appl Ecol* 35: 532-543

Warwick RM, Clarke KR (2001) Practical measures of marine biodiversity based on relatedness of species. *Oceanog Mar Biol Ann Rev* 39: 207-231

PERMANOVA+

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 26: 32-46

Anderson MJ (2001) Permutation tests for univariate or multivariate analysis of variance and regression. *Can J Fish Aquat Sci* 58: 626-639

Anderson MJ (2006) Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62: 245-253

Anderson MJ (2008) Animal-sediment relationships revisited: characterising species' distributions along an environmental gradient using canonical analysis and quantile regression splines. *J Exp Mar Biol Ecol* 366: 16-27

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, Sanders NJ, Cornell HV, Comita LS, Davies KF, Harrison SP, Kraft NJB, Stegen JC, Swenson NG (2011) Navigating the multiple meanings of b diversity: a roadmap for the practicing ecologist. *Ecol Lett* 14: 19-28.

Anderson MJ, Connell SD, Gillanders BM, Diebel CE, Blom WM, Landers TJ, Saunders JE (2005) Relationships between taxonomic resolution and spatial scales of multivariate variation in kelp holdfast assemblages. *J Anim Ecol* 74: 636-646

Anderson MJ, Diebel CE, Blom WM, Landers TJ (2005) Consistency and variation in kelp holdfast assemblages: spatial patterns of biodiversity for the major phyla at different taxonomic resolutions. *J Exp Mar Biol Ecol* 320: 35-56

Anderson MJ, Ellingsen KE, McArdle BH (2006) Multivariate dispersion as a measure of beta diversity. *Ecol Lett* 9: 683-693

Anderson MJ, Gorley RN, Clarke KR (2008) *PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods*. PRIMER-E: Plymouth, UK, 214pp

Anderson MJ, Gribble NA (1998) Partitioning the variation among spatial, temporal and environmental components in a multivariate data set. *Aust J Ecol* 23: 158-167

Anderson MJ, Legendre P (1999) An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J Statist Comput Sim* 62: 271-303

Anderson MJ, Millar RB (2004) Spatial variation and effects of habitat on temperate reef fish assemblages in northeastern New Zealand. *J Exp Mar Biol Ecol* 305(2): 191-221

Anderson MJ, Robinson J (2003) Generalized discriminant analysis based on distances. *Aust NZ J Stat* 45: 301-318

Anderson MJ, Robinson J (2001) Permutation tests for linear models. *Aust NZ J Stat* 43: 75-88

Anderson MJ, Santana-Garcon J (2015) Measures of precision for dissimilarity-based multivariate analysis of ecological communities. *Ecol Lett* 18: 66-73.

Anderson MJ, ter Braak CJF (2003) Permutation tests for multi-factorial analysis of variance. *J Statist Comput Sim* 73: 85-113

Anderson MJ, Walsh DCI (2013) What null hypothesis are you testing? PERMANOVA, ANOSIM and the Mantel test in the face of heterogeneous dispersions. *Ecol Monogr* 83: 557-574.

Anderson MJ, Walsh DCI, Clarke KR, Gorley RN, Guerra-Castro E (2017) Some solutions to the multivariate Behrens-Fisher problem for dissimilarity-based analyses. *Aust NZ J Stat* 59: 57-79.

Anderson MJ, Willis TJ (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84: 511-525

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol Monogr* 69: 1-24

McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82: 290-297

Paul WL, Anderson MJ (2013) Causal modelling with multivariate species data. *J Exp Mar Biol Ecol* 448: 72-84.