

# CAP

## Canonical Analysis of Principal coordinates

A computer program  
by Marti J. Anderson



Department of Statistics  
University of Auckland  
(2002)

## DISCLAIMER

This FORTRAN program is provided without any explicit or implicit warranty of correct functioning. It has been developed as part of a university-based research program. If, however, you should encounter problems with this program, the author will be happy to help solve them. Researchers can use this program for scientific and research purposes, but intellectual property and copyright for the source code and program remains the property of Marti Jane Anderson. This program relies on several canned FORTRAN routines from other sources, including:

- (a) routines “tred2” and “tqli” from Numerical Recipes (Press et al. 1986) are used to finding the eigenvalues of a real symmetric distance matrix, using householder reduction;
- (b) routine “gaussj” from Numerical Recipes (Press et al. 1986) is used to find the solution to linear equations (i.e. the inverse of a matrix) by Gauss-Jordan elimination;
- (c) routine “cholde” from Numerical Recipes (Press et al. 1986) is used to find the Cholesky decomposition of a positive-definite symmetric matrix;
- (d) routines “svd” and “pythag” from a translation of the Algol procedure (Wilkinson and Reinsch 1971) were used to find the singular value decomposition of a matrix.

Research publications that use this method should cite the following papers.

Anderson, M.J. and Robinson, J. (in review). Generalised discriminant analysis based on distances. *Australian and New Zealand Journal of Statistics*.

Anderson, M.J. and Willis, T.J. (in press). Canonical analysis of principal coordinates: an ecologically meaningful approach for constrained ordination. *Ecology*.

Users of the computer program may also refer to the present user’s guide as follows:

Anderson, M.J. (2002). CAP: a FORTRAN computer program for canonical analysis of principal coordinates. Department of Statistics, University of Auckland.

### *Author’s contact details:*

Dr Marti J. Anderson  
 Department of Statistics  
 Tamaki Campus  
 University of Auckland  
 Private Bag 92019  
 Auckland, New Zealand  
 Tel: 64-9-373-7599 ext 5052  
 Fax: 64-9-373-7000  
 Email: [mja@stat.auckland.ac.nz](mailto:mja@stat.auckland.ac.nz)  
 Website: <http://www.stat.auckland.ac.nz/~mja>

## I. Description

CAP is a computer program that calculates a canonical analysis on the principal coordinates based on any symmetric distance matrix, including a test by permutation, as described by Anderson and Willis (in press) and Anderson and Robinson (in review). Consider an  $(N \times p)$  matrix of response variables  $\mathbf{Y}$  ( $N$  = the total number of observational units and  $p$  = the number of variables). Consider also an  $(N \times q)$  matrix,  $\mathbf{X}$ , which is of interest for a multivariate hypothesis. The purpose of CAP is to perform a non-parametric canonical analysis for the effect of  $\mathbf{X}$ , if any, on  $\mathbf{Y}$  on the basis of any distance measure of choice, using permutations of the observations. Note that  $\mathbf{X}$  may contain the codes of an ANOVA model (a design matrix), yielding a generalised discriminant analysis, or it may contain one or more explanatory (predictor) variables of interest (e.g. environmental variables), yielding a generalised canonical correlation analysis.

To do the analysis, the first step is to let  $\mathbf{D} = (d_{ij})$  be an  $(N \times N)$  distance matrix calculated from observation units of  $\mathbf{Y}$ , using some chosen appropriate distance measure. Let  $\mathbf{A} = (a_{ij}) = (-\frac{1}{2}d_{ij}^2)$ , then calculate Gower's (1966) centered matrix ( $\mathbf{G}$ ) by centering the elements of  $\mathbf{A}$ , i.e.

$$\mathbf{G} = \left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}'\right)\mathbf{A}\left(\mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}'\right)$$

where  $\mathbf{1}$  is a column of 1's of length  $N$  and  $\mathbf{I}$  is an  $(N \times N)$  identity matrix. Matrix  $\mathbf{G}$  is then decomposed into its component eigenvalues and corresponding orthonormal eigenvectors  $\mathbf{Q}$ . We then choose a subset of these eigenvectors (say  $m$  of them), as a matrix for the ensuing canonical analysis that we will call  $\mathbf{Q}_m$ .

The general idea is to include as much of the relevant information in matrix  $\mathbf{Q}$  (and thus  $\mathbf{G}$ ) as is reasonable for further analysis. It is important to keep  $m$  relatively small compared to  $N$ , the total number of observations. The user can either choose  $m$  manually or choose to allow the computer program to choose  $m$  on a non-arbitrary basis using diagnostic information on the appropriate dimensionality for the canonical analysis.

Next, we calculate the "hat" or projection matrix  $\mathbf{H} = \mathbf{X}[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'$  (e.g. Johnson & Wichern 1992). Then, the canonical test statistic is the trace:

$$tr(\mathbf{Q}_m'\mathbf{H}\mathbf{Q}_m).$$

The eigenvalues of matrix  $\mathbf{Q}^* = \mathbf{Q}_m'\mathbf{H}\mathbf{Q}_m$  are the squared canonical correlations  $\delta_1^2, \dots, \delta_s^2$  (where  $s = \min(m, q)$ ). In addition to the trace statistic, the program calculates the greatest root statistic  $\delta_1^2$ . A  $P$ -value for each of the trace and greatest root statistics is then obtained by recalculating each of them for a large number of random re-orderings of the observations (i.e. the rows and columns of  $\mathbf{Q}_m$ ), while keeping  $\mathbf{X}$  (and  $\mathbf{H}$ ) constant. There will be  $s$  canonical axes, for plotting, which are given by  $\mathbf{Q}^*$ , standardized by the appropriate canonical correlation,  $\delta_i$ ,  $i = 1, \dots, s$ .

Output from the program includes:

- (a) Eigenvalues and eigenvectors from the principal coordinate analysis. The latter are the PCO axes that can be used to plot an unconstrained (metric MDS) of the data.
- (b) Canonical correlations and squared canonical correlations
- (c) Canonical axes scores (position of multivariate points on the canonical axes to be used for plotting).
- (d) Correlations of each of the original variables with each of the canonical axes.
- (e) Diagnostics used to determine the appropriate value for the choice of  $m$ . The criterion used is either the value of  $m$  resulting in the minimum misclassification error (in the case of groups) or the minimum residual sum of squares (in the case of  $\mathbf{X}$  containing one or more quantitative variables). Also,  $m$  must not exceed  $p$  or  $N$  and is chosen so that the proportion of the variability explained by the first  $m$  PCO axes is more than 60% and less than 100% of the total variability in the original dissimilarity matrix.

- (f) In the case of groups, a table of results for the “leave-one-out” classification of individual observations to groups is given, along with the misclassification error for the choice of  $m$  used.
- (g) If requested, the results of a permutation test using the two different test statistics, (trace and largest root).

## II. What is the difference between CAP and NPMANOVA?

An important point is that the analysis described here takes into account the correlation structure among the variables. Although the test statistic described for the programs NPMANOVA (Anderson 2000, Anderson 2001) or DISTLM (McArdle and Anderson 2001) test a similar multivariate hypothesis for a linear model, they do not take into account the correlation structure among the variables. CAP does this by essentially following the more traditional multivariate discriminant (or canonical) analysis, but does this on the principal coordinates from the distance matrix. One might expect to get similar results using CAP and the permutation test given by NPMANOVA. Both have exact type I error for the same multivariate null hypothesis of no differences among *a priori* groups. However, there are important situations in which CAP will tend to be more powerful.

For example, in ecology, there are often situations where there are several highly abundant species that do not change across treatment groups and are strongly correlated with one another. However, there may be less abundant or patchy species that are not correlated with the abundant species, and do indeed differ significantly across treatment groups being considered. In this case, CAP will find these real and significant differences among the assemblages more often than NPMANOVA (Anderson and Robinson, in review).

CAP provides a constrained ordination diagram, which is not obtained by these other methods. In situations such as that described above, an unconstrained ordination (such as metric or non-metric MDS or principal coordinate analysis) will not show any clear separation of groups, even though group differences do occur. This is because the group differences occur along another dimension. This dimension will be drawn by the canonical plot (output by CAP). In fact, the canonical analysis finds the axis (or axes) in the principal coordinate space that is best at discriminating among the *a priori* groups.

A potential disadvantage of using CAP is that it will ignore some of the variability contained in matrix  $\mathbf{G}$ . This will probably be irrelevant noise, but the choice of  $m$  is crucial to the analysis. Also,  $m$  must not be too large, or the canonical analysis may give results that are not meaningful (see Anderson and Willis, in press, for more details). No such decision needs to be made when using NPMANOVA or DISTLM, as either of these approaches will simply use all of the information in the distance matrix. In addition, NPMANOVA (or DISTLM) will provide a more powerful test if the original variables are not strongly correlated with one another (Anderson and Robinson, unpublished results).

## III. Input file(s)

The program allows the user to input a distance matrix  $\mathbf{D}$  directly or a raw multivariate data matrix of response variables  $\mathbf{Y}$ . In either case, the file should be saved as ASCII \*.txt, with no column or row headings. This is the first input file for the program. If a file with a raw data matrix is input, either the rows or the columns may correspond to the variables for the analysis (the user will be given the option to choose one of these). Then, the user has several options for transformation, standardization and choice of distance measure.

If the hypothesis of interest involves differences among groups, then the user should choose “Discriminant Analysis” as the type of analysis to be done. The program will then ask the user relevant questions regarding the number of groups and the number of observations per group (see section IV below). Thus, no other data files are needed.

On the other hand, if the hypothesis of interest is the relationship of the species data with some other quantitative variables (e.g. environmental data), then the user should choose “Canonical Correlation

Analysis” and follow the relevant instructions. In the case, another input file is required, consisting of an **X** matrix containing one or more variables of interest for canonical analysis. For the **X** input matrix, variables must be columns and observation units must be rows. The file must be saved in ASCII \*.txt format with no headers or labels of any kind. The user will specify the number of columns, while the number of rows must be equal to the number of observational units already specified for **Y** (or **D**).

To avoid dealing with long file names and paths to locate files, place the relevant input file(s) in the same location on your computer (i.e. the same directory) as the CAP.exe file, for use with the program. Double-click on the “CAP.exe” file to run the program.

The program uses dynamic memory allocation, and so (theoretically) does not have any limits in the number of parameters (number of rows or columns) that may be used for the input files. However, if you have a large input file and cannot seem to get the program to work, then please contact the author.

## IV. Questions Asked by the Program

The questions asked by the program are best demonstrated by an example. The data analysed here are from an underwater visual census of fishes at three different times at the Poor Knights Islands, New Zealand (courtesy of Mr Trevor Willis). There were 47 fish species variables recorded as divers swam along 25m transects. Counts from nine such transects at the location were pooled to constitute a single observational unit. Data were collected in September 1998 ( $n_1 = 15$ ), March 1999 ( $n_2 = 21$ ) and September 1999 ( $n_3 = 20$ ). Thus,  $N = 56$  and  $p = 47$ . The hypothesis of interest (examined below) was to test for a significant difference in the assemblages of fishes recorded at the Poor Knights at each of the three times. For more details, see Willis and Denny (2000), Anderson and Robinson (in review) and Anderson and Willis (in press).

The response variable matrix **Y** (56 rows x 47 columns) was contained in an ASCII text file called “PK.txt”. The file containing the matrix of codes for dummy variables indicating the three times (56 rows x 2 columns) was called “Time.txt.”

The questions asked by the program are given in Courier font, while responses given for this data set are in ***Times bold italics***.

---

```

CAP
-----
Canonical Analysis of Principal coordinates

A program for
generalised discriminant analysis or
canonical correlation analysis
on the basis of any distance measure

by M.J. Anderson
Department of Statistics
University of Auckland (2002)

Type the name of the input file containing your data
(e.g. species variables or a distance matrix).
PK.txt

Type a name for the output file of results (*.txt)
Results.txt

Nature of the data in the input file:
1) raw data (n x p)
2) distance matrix (n x n)
1

```

Structure of the input file:

- 1) rows are samples and columns are variables
- 2) columns are samples and rows are variables

**1**

How many variables (columns) are there?

**47**

How many observations (rows) are there?

**56**

Choice of transformation for Y data:

- 1) none
- 2) square-root
- 3) fourth-root
- 4)  $\ln(x)$
- 5)  $\ln(x+1)$
- 6)  $\log_{10}(x)$
- 7)  $\log_{10}(x+1)$
- 8) presence/absence

**5**

Choice of standardisation:

- 1) none
- 2) standardise by row (sample) sums
- 3) standardise by column (variable) sums
- 4) double standardise by row and column sums
- 5) standardise each variable to z-scores (normalise)
- 6) standardise each variable by dividing by its range

**1**

Choice of distance measure:

- 1) Bray-Curtis dissimilarity
- 2) square root of Bray-Curtis
- 3) Euclidean distance
- 4) Orloci's Chord distance
- 5) Chi-square metric
- 6) Chi-square distance
- 7) Hellinger distance
- 8) Gower dissimilarity
- 9) Canberra distance
- 10) square root of Canberra distance
- 11) CY dissimilarity
- 12) Deviance based on the binomial
- 13) Deviance per observation (scale invariant)
- 14) Kulczynski dissimilarity

**1**

Now doing the principal coordinate analysis...

Choose one of the following types of analysis:

- 1) Discriminant Analysis  
(i.e. find axes that maximise separation among groups)
- 2) Canonical Correlation Analysis  
(i.e. find axes that maximise correlation with a set of  
quantitative variables in a second matrix, X)

**1**

How many groups are there?

**3**

Type the sample sizes per group, separated by spaces.

**15 21 20**

Choose whether you wish to:

1) choose  $m$  manually

2) let the computer program choose  $m$

Note:  $m$  is the number of PCO axes to be used in the canonical analysis.

**2**

Note: not knowing how many axes would be appropriate a priori, it is usually the case that one will let the computer program choose  $m$ , at least in the first instance.

Do you wish to do a test by permutation?

1) yes

2) no

**1**

Type the number of random permutations for the test.

(e.g. 99, 499, 999, 4999, 9999, etc.)

**9999**

Type an integer to be used as the seed  
for the random permutations

**12**

Note: Any random integer seed will do here. If you wish to re-calculate the exact same permutation test at a later date (i.e. with the exact same set of re-orderings), make a note of this seed and use it again.

Please wait while I do some diagnostic calculations...

Note: Here, the program sequentially chooses increasing values for  $m$  and gets the diagnostic information (proportion of variability explained by the first  $m$  PCO axes, residual sum of squares, squared canonical correlations, and misclassification error in the case of groups) all to be output later. If you manually choose an integer value for  $m$  yourself, then you will also have the option of skipping these diagnostic calculations. It is advised that you choose to do the diagnostic calculations in any event, as they are useful for making a decision about the appropriate dimensionality of the problem.

Now doing the canonical analysis...

Please wait while I do the permutation test...

Results have been sent to the output file.

End of the program.

Press q to quit.

**q**

---

When the program has completed the analysis, it will print the results to the named output file. In this case, the output file looks like this:

```

CAP
-----
Canonical Analysis of Principal coordinates
-----

A program for generalised
discriminant analysis or
canonical correlation analysis
on the basis of any distance measure

by M.J. Anderson
Department of Statistics
University of Auckland (2002)

Input file name for Y matrix: PK.txt
No. of variables in Y      = 47
No. of observations       = 56
No. of groups             = 3
Sample sizes per group = 15 21 20
Choice of m               = 7
Data were transformed to ln(x+1)
No standardisation
Analysis based on Bray-Curtis dissimilarities

-----
Results:  PRINCIPAL COORDINATE ANALYSIS
-----

Percentage of variation explained by axes
            individual    cumulative
Axis  1      20.718%      20.718%
Axis  2      12.370%      33.088%
Axis  3      10.803%      43.891%
Axis  4       8.123%      52.014%
Axis  5       7.566%      59.580%
Axis  6       6.650%      66.230%
Axis  7       5.149%      71.379%

Principal Coordinate Axes (unconstrained)
      Axes
Sample  1      2      3      4      5      6      7
1      0.1104   0.0586   0.0331  -0.0248  -0.0561   0.0557   0.0854
2     -0.1242   0.1344  -0.1197   0.1823  -0.0136   0.1633  -0.0245
3      0.1430  -0.0724  -0.0300  -0.0399   0.0639   0.0947   0.0283
4      0.1564  -0.0332  -0.0771  -0.0998  -0.0027  -0.0765   0.0019
5     -0.1594  -0.1224  -0.0095  -0.0702   0.0857   0.0255  -0.0305
6      0.1349  -0.1624  -0.1126  -0.0679   0.0956   0.0647   0.0596
7      0.0844  -0.1275  -0.0345   0.0617  -0.0093   0.1569  -0.0162
8     -0.0075  -0.0866  -0.0685   0.0818  -0.0357  -0.0357   0.0683
9      0.0887  -0.2787  -0.0103   0.1601   0.0198  -0.0167   0.0804
10     -0.1084  -0.0483  -0.1631  -0.0537  -0.0168   0.0134   0.0746
11      0.1321  -0.1106  -0.0622   0.0265   0.0801   0.0443   0.0457
12     -0.1659  -0.1058   0.1832  -0.1624   0.1487   0.1027   0.1362
13     -0.0966  -0.2130   0.0043   0.0144   0.0381   0.0738   0.0213
14     -0.2774  -0.2218   0.0524   0.1005   0.1331   0.0626   0.0392
15      0.0714  -0.0416  -0.1357   0.0642  -0.0979  -0.0125   0.0004
16      0.2198   0.1283   0.0740   0.1223   0.2173  -0.0276   0.0389
17     -0.0062   0.2719  -0.2205  -0.0261   0.0629  -0.0388   0.0338
18      0.1580   0.1192   0.1257  -0.0022  -0.0164   0.0282  -0.0198
19      0.1517   0.0630   0.0706  -0.0624  -0.0224  -0.1328  -0.0768
20     -0.1600   0.0392   0.0282  -0.1608   0.0749  -0.0519  -0.0001
21      0.0561  -0.1468  -0.0701   0.0211  -0.0724  -0.1189  -0.0444
22     -0.0376  -0.0145   0.1287   0.0274   0.0161  -0.1968  -0.0260
23      0.1291  -0.0415  -0.0082   0.1975   0.0914  -0.0750   0.0284
24      0.0481  -0.0450   0.1548  -0.1642  -0.0136   0.0186   0.0921
25      0.0503  -0.0932  -0.0893   0.0297  -0.1242  -0.1182   0.0027
26     -0.2067   0.0111  -0.0661   0.0343  -0.0186  -0.1494   0.0154
27      0.0964   0.0597   0.0594  -0.1435  -0.0458  -0.0761   0.0466

```



28	-0.1622	0.0038	0.0695	0.0851	0.2305	-0.0356	-0.1173
29	0.1177	0.0246	-0.0749	0.0675	0.0386	-0.1425	0.0030
30	-0.3484	0.1034	0.0423	-0.0593	0.0466	-0.1721	0.0348
31	-0.2282	0.1543	0.0078	0.0681	-0.1802	0.0665	0.1852
32	0.1369	0.1287	0.2050	0.1389	-0.0177	0.1264	0.0219
33	-0.2232	0.1391	0.0187	0.0291	0.0367	-0.0132	-0.0004
34	0.1943	0.0593	0.0141	0.0635	0.1240	-0.0222	-0.0634
35	-0.0979	-0.0315	0.1062	0.0585	-0.0813	-0.0878	0.0361
36	0.0818	0.1088	0.1492	0.1599	-0.0895	-0.0478	0.0587
37	0.1262	0.1361	0.0570	-0.0067	0.0052	-0.0096	-0.0051
38	-0.0341	0.0459	-0.2414	0.0309	-0.0726	0.0725	-0.0278
39	0.2336	-0.0905	0.0634	-0.1007	-0.0903	0.0270	-0.1454
40	0.0851	0.0589	0.1298	-0.1100	-0.1635	0.1448	-0.0468
41	0.0621	-0.1935	-0.1337	-0.0644	0.0231	0.0429	0.0258
42	-0.0453	-0.0126	0.0290	0.0320	-0.0901	0.0636	0.0192
43	-0.1732	0.0100	-0.0638	-0.1501	0.0255	0.0343	-0.2024
44	0.0557	0.0363	-0.0702	-0.1976	0.0567	-0.0485	0.0189
45	0.0089	-0.0357	-0.1174	-0.0332	-0.0418	-0.0176	-0.0997
46	-0.1834	0.1052	-0.1289	-0.0160	-0.0280	0.0047	0.0564
47	0.1817	0.1806	0.0126	-0.0526	0.1249	-0.0274	-0.0399
48	0.2002	-0.0038	-0.0818	0.0455	0.0109	-0.0327	-0.1657
49	0.1480	-0.0338	0.0325	-0.1539	-0.1112	0.0012	0.0288
50	-0.1713	-0.0536	0.2713	0.0321	0.0186	0.0603	-0.1435
51	-0.0906	-0.1491	0.1620	0.0712	-0.2018	-0.1061	-0.0846
52	-0.1290	-0.0058	0.0268	-0.0379	-0.0456	0.0183	0.0600
53	-0.2760	0.0910	-0.1102	0.0192	-0.0224	0.0599	-0.1055
54	0.0714	0.1846	0.0052	-0.0215	0.0817	0.1315	0.0126
55	-0.1072	0.0297	0.0850	0.0134	-0.0852	0.0532	-0.1166
56	-0.0054	0.0895	-0.1019	0.0430	-0.0841	0.0786	-0.0121

-----  
Results: CANONICAL ANALYSIS  
-----

Eigenvalues (Correlations)  
0.78101 0.69142

Squared Correlations (delta^2)  
0.60998 0.47807

Canonical Axes (constrained)  
Axes

Sample	1	2
1	-0.0093	0.0152
2	-0.0821	-0.0383
3	-0.1435	0.0296
4	0.0085	-0.0119
5	-0.1058	0.0082
6	-0.2029	0.0897
7	-0.2120	-0.0094
8	-0.0320	0.1058
9	-0.1303	0.2121
10	-0.1042	0.0462
11	-0.1250	0.0993
12	-0.1205	0.1361
13	-0.1839	0.0744
14	-0.1560	0.1736
15	-0.0465	-0.0106
16	0.1258	0.1548
17	0.1043	-0.0106
18	0.0866	-0.0608
19	0.1648	0.0581
20	0.0534	-0.0193
21	0.0049	-0.0012
22	0.2120	0.0531
23	0.0593	0.1688
24	-0.0138	0.0270
25	0.0292	0.0088
26	0.1192	0.0563
27	0.1016	-0.0244
28	0.0637	0.0389
29	0.1195	0.0741
30	0.2127	0.0570
31	0.0404	0.0633

32	0.0497	0.0032
33	0.0991	0.0001
34	0.0630	0.0107
35	0.1050	0.0533
36	0.1761	0.0631
37	0.1036	-0.0347
38	-0.1114	-0.0828
39	-0.0559	-0.1983
40	-0.0540	-0.2074
41	-0.2032	0.0409
42	-0.0457	-0.0245
43	-0.0570	-0.2420
44	0.0098	-0.0287
45	-0.0428	-0.1127
46	0.0082	0.0058
47	0.1193	-0.0378
48	0.0111	-0.1255
49	-0.0258	-0.0734
50	0.0202	-0.1064
51	0.0863	-0.0728
52	-0.0155	0.0194
53	-0.0297	-0.1376
54	-0.0160	-0.0396
55	0.0105	-0.1585
56	-0.0432	-0.0778

Correlations of Canonical Axes (Q\*) with Original Variables (Y)

Var	Axes	
	1	2
1	0.3138	0.1391
2	0.1998	0.2430
3	0.3118	0.1140
4	0.2105	0.1220
5	-0.1435	-0.0779
6	0.0078	0.3033
7	-0.0165	-0.0276
8	-0.1477	0.1338
9	0.1226	-0.1656
10	0.0136	-0.2090
11	-0.1864	-0.0266
12	0.1099	-0.1331
13	0.1121	0.1241
14	0.1147	0.3073
15	0.0540	0.0888
16	0.4184	0.0796
17	0.4596	0.2383
18	0.3106	0.2265
19	0.3842	0.2452
20	0.2515	0.1743
21	0.1348	0.2432
22	-0.1296	0.0610
23	-0.3677	0.3113
24	-0.1892	-0.1290
25	-0.0554	-0.2402
26	0.2537	0.0658
27	0.0268	0.0739
28	0.3219	-0.1067
29	0.1052	-0.0409
30	0.0965	0.3064
31	-0.2959	-0.1358
32	0.1800	-0.0208
33	-0.2777	0.5117
34	0.7451	0.1841
35	-0.0323	-0.0749
36	-0.2252	0.2998
37	0.0234	0.0650
38	0.1913	0.1019
39	0.5656	-0.3508
40	-0.1299	0.3767
41	-0.1716	0.3172
42	0.1497	-0.1477
43	0.1939	-0.2596
44	0.3698	-0.0142
45	0.0012	-0.3048

```

46      0.1047   -0.1829
47      0.0494    0.0535

```

---

Results: DIAGNOSTICS

---

m	prop.G	ssres	d_1^2	d_2^2	%correct
1	0.20718	2.09078	0.00015	0.00000	8.929%
2	0.33088	1.85268	0.27131	0.00014	55.357%
3	0.43891	1.87664	0.33280	0.01279	51.786%
4	0.52014	1.92431	0.33289	0.08603	46.429%
5	0.59580	2.00565	0.35139	0.15689	44.643%
6	0.66230	1.55688	0.60976	0.22268	64.286%
7	0.71379	1.31294	0.60998	0.47807	71.429%
8	0.76010	1.24198	0.61529	0.50993	67.857%
9	0.80200	1.29525	0.61672	0.51686	66.071%
10	0.83996	1.44111	0.62080	0.53536	64.286%
11	0.87360	1.36885	0.68756	0.53851	64.286%
12	0.90610	1.35568	0.68781	0.54218	62.500%
13	0.93365	1.26236	0.73063	0.55217	62.500%
14	0.96002	1.30508	0.73149	0.55271	60.714%
15	0.98338	1.38981	0.73313	0.56195	60.714%

Note: m cannot include axes that cause the total SS to exceed the original total SS.

---

Results: CROSS-VALIDATION

---

Leave-one-out Allocation of Observations to Groups  
(for the choice of m = 7)

		Classified into groups			Total	%correct
		1	2	3		
Original						
Group	1	9	0	6	15	60.000%
Group	2	0	15	6	21	71.429%
Group	3	1	3	16	20	80.000%

Total correct = 40/ 56 = 71.429%  
Mis-classification error = 28.571%

---

Results: PERMUTATION TEST

---

trace statistic = (tr(Q\_m'HQ\_m))  
first squared canonical correlation = (delta\_1^2)

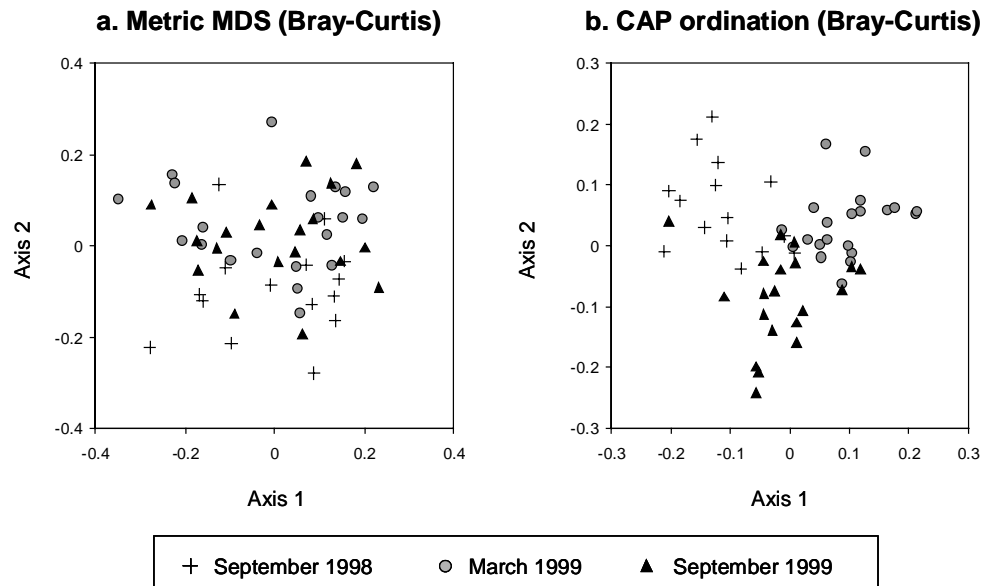
---

tr(Q_m'HQ_m) =	1.088049	P =	0.000100
delta_1^2 =	0.609984	P =	0.000100

---

No. of permutations used = 9999  
Integer for the random seed = 12

These results indicate that there is a significant difference in the composition and relative abundance of fish species at the three different times of observation. Below are two ordination plots. The first is the (unconstrained) metric multi-dimensional scaling plot of the Bray-Curtis distances (produced using the first two principal coordinate axes from the output above). The second is the plot of the first two canonical axes produced by CAP (the canonical axes in the output above). The group differences are not apparent on the principal coordinate plot, while the canonical plot shows the differences quite clearly. This is because the axis of real group differences that occurred in multivariate space was not in the same direction as the maximum variation, which is why the differences do not appear in the unconstrained plot. Clearly, it is useful to have both the unconstrained and the constrained ordination to examine multivariate patterns in this data set.



## V. References

- Anderson, M.J. (2000). NPMANOVA: a FORTRAN computer program for non-parametric multivariate analysis of variance (for any two-factor ANOVA design) using permutation tests. Department of Statistics, University of Auckland.
- Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26: 32-46.
- Anderson, M.J. and Robinson. (in review). Generalised discriminant analysis based on distances. *Australian and New Zealand Journal of Statistics*.
- Anderson, M.J. and Willis, T.J. (in press). Canonical analysis of principal coordinates: an ecologically meaningful approach for constrained ordination. *Ecology*.
- Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53: 325-338.
- Johnson, R.A. and Wichern, D.W. (1992). *Applied multivariate statistical analysis*, 3<sup>rd</sup> edition. Prentice-Hall, Englewood Cliffs, New Jersey.
- McCordle, B.H. and Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82(1): 290-297.
- Press, W.H., Teuklosky, S.A., Vetterling, W.T. and Flannery, B.P. (1992). *Numerical Recipes in FORTRAN*, 2<sup>nd</sup> edition. Cambridge University Press, Cambridge.
- Wilkinson, J.H. and Reinsch, C. (1971). Linear algebra. Pages 134-151 in *Handbook for automatic computation, Volume 2*. (F. L. Bauer, chief editor). Springer-Verlag, Berlin.
- Willis, T.J. & Denny, C.M. (2000). Effects of the Poor Knights Islands Marine Reserve on demersal fish populations. Report to the Department of Conservation, Science and Research Grant No. 2519, Leigh Marine Laboratory, University of Auckland, New Zealand.