TESTS FOR DISPERSION AMONG MACROFUNGAL SPECIES ASSEMBLAGES

David A. Ratkowsky

School of Agricultural Science and School of Plant Science, University of Tasmania, Private Bags 54 and 55, Hobart, Tasmania 7001, Australia

Abstract

Two previous papers in this series described recent advances in the generalisation of canonical discriminant analysis and canonical correlations analysis, respectively, and applied them to data consisting of lists of macrofungal species obtained from surveys in the Warra long-term ecological research (LTER) site in southern Tasmania. The key feature of the new advances is that they can use the presence/absence, or the abundance, of each species in what often is a long list of species, data that are clearly non-normal (non-Gaussian) in distribution. The earlier paper on canonical discriminant analysis focussed on tests of the equality of multivariate locations (i.e., the positions of the group mean vectors). The present paper, like the former one, uses permutation tests to obtain exact probability values, but deals with multivariate dispersion rather than location. Because differences in dispersion may be indicative of differences in the biodiversity of assemblages, a focus of this study was to examine whether dispersion can serve as a surrogate for species richness. It was found that high dispersion did not necessarily reflect greater fungal biodiversity, but may also result from small numbers of seasonally occurring species. As before, the examples are drawn from macrofungal surveys conducted at the Warra LTER site.

Introduction

Multivariate analysis of variance (MANOVA) and canonical variate analysis (CVA) are statistical procedures that test differences between the locations, i.e. the positions of the group mean vectors, of predefined groups of experimental units. As the multivariate counterpart of analysis of variance (ANOVA), both of these techniques assume that the variables follow multivariate normal (i.e. Gaussian) distributions, an assumption that is untenable for ecological data arising from inventory sampling, where the raw data may involve the presence/absence, or abundance, of each species in a very long list of species. Other restrictions also apply to the traditional methods, for example, the number of variables (species) must be less than the number of observations (samples), whereas in ecological surveys the species lists may be much longer than the number of samples (i.e., the total number of visits made to the sampling units in the examples treated here).

In response to the need for a flexible method of constrained ordination that could be applied to any distance or dissimilarity measure, Anderson and Willis (2003) coupled metric principal coordinate analysis (PCoA), which allows any definition of distance or dissimilarity, with CVA to produce canonical analysis of principal coordinates (CAP). This procedure, and the associated computer program by which the technique can be put into practice, was the subject of two earlier papers in this series that gave applications to data from macrofungal surveys. The first (Ratkowsky 2007) was devoted to the option of CAP that tests hypotheses concerning groups, and the other (Ratkowsky and Gates 2008) was confined to the option that tests hypotheses regarding relationships with additional sets of quantitative predictor variables.

Circumstances present themselves where the focus of attention is upon the dispersion (i.e. the spread) of points in a group rather than

upon their location. For example, the power of the test for equality of the group mean vectors in canonical discriminant analysis, whether it is the traditional procedure that assumes multivariate normality or a generalisation such as CAP, depends upon the equality of the group dispersions, the power increasing as the dispersions became more and more equal. This is the multivariate counterpart of the requirement of homogeneity of variance in ANOVA. Note that even where the test for location is based upon permutation or randomisation, it is sensitive to inequality of dispersion among groups. Tests for dispersion may be of interest in their own right, particularly in ecology, where a larger spread of points in one or more of the groups compared to that in the other groups may be interpreted as a symptom of environmental stress (Warwick and Clarke 1993).

Because ecological data sets often have more variables (i.e., species) than observations (i.e., samples), may be highly skewed in distribution, involve counts rather than measures, and have a resulting data matrix that is dominated by zeroes, the traditional tests for homogeneity of variances, such as Bartlett's test and Levene's test (Levene 1960), are not accurate. Improvements to Levene's test were made by Brown and Forsythe (1974), who substituted medians for means in Levene's test, by Van Valen (1978), who proposed a multivariate analogue using group centroids, by O'Brien (1992) and Manly (1994), who substituted multivariate medians for group centroids, and more recently, by Anderson (2006), whose dissimilarity-based multivariate generalisation of Levene's test has two test statistics, one based on the ANOVA F-statistic that uses distances to centroids, and the other that uses distances to spatial medians. The centroid of a group of points is the point that minimises the sum of squared distances to the points within the group, whereas the spatial median is the point that minimises the sum of distances to the points within the group (i.e., without squaring the distances). For both statistics, one can use F-tables (or the mathematical formulae by which the critical values in the F-tables are calculated) to determine the P-value corresponding to the F-statistic or one can use a permutation procedure. The output is therefore four different statistics by which the null hypothesis may be judged. When using the distances to centroids, an appropriate permutation procedure is to permute the least-

square residuals (Anderson and Robinson, 2001), whereas when using deviations from spatial medians, an appropriate procedure is to permute the least-absolute-deviation (LAD) residuals, these being the distances to the point that results in the largest absolute deviation being a minimum (Cade and Richards, 1996).

The present paper illustrates the application of the computer program PERMDISP2, developed by Prof. M.J. Anderson of the University of Auckland to enable tests for the homogeneity of multivariate dispersions, the subject of a paper by Anderson (2006), to be carried out. The program is available for downloading without charge from her website, http://www.stat.auckland.ac.nz/~mja. PERMDISP2 allows unequal replication, unlike her program PERMDISP, which is confined to equally replicated groups. Instructions for the use of PERMDISP, which applies equally well for PERMDISP2 with only minor adaptation, can also be downloaded from the same website. The present paper serves two purposes, one of which is to illustrate the use of those tests for data derived from macrofungal surveys. The other purpose is to examine whether dispersion can serve as a surrogate for fungal species richness. Anderson (2006) found that greater dispersion strongly reflected biological factors in each of the three examples she presented there. One of these examples, involving spatial variation in New Zealand fish, had eight sites within each of four locations. Underwater divers swimming along transects within each of these sites made lists of abundances of individual fish species, the lists being pooled at the level of site for statistical analysis. As sampling was repeated in each of four years, there were 32 lists per location (although one location later had one more site added to it in two years), enabling a test of whether the greater biodiversity of the fish assemblages at the northern locations that are more exposed to the influences of the East Auckland current was reflected in greater multivariate dispersion. Irrespective of whether the traditional F-distribution or permutation tests were used, the tests suggested that there were important differences in the biodiversity of fish assemblages as measured by multivariate spread.

Materials and Methods

The examples for which PERMDISP2 will be illustrated are drawn from two surveys

conducted at the Warra long-term ecological research (LTER) site in southern Tasmania. These are the same two data sets that were used in an earlier paper in this series (Ratkowsky 2007). The first survey compared the macrofungal species in a regenerating coupe to the species in a nearby mature forest (Gates *et al.* 2005) and the second survey compared the macrofungal species found in the harvested and unharvested portions of an aggregated retention coupe with the species in a mature forest (Gates and Ratkowsky 2006). Readers interested in more details about the forest type, the size and location of the coupes, and the complete list of fungal species in both coupes, should consult the papers cited. Pseudoreplication was provided by multiple visits made during each of the four seasons; the number of visits is given in Table 1 of Ratkowsky (2007) for each survey. In both surveys, records of the fungal species recorded during each visit to each coupe (or in the case of the second example, to the

separate parts of the aggregated retention coupe as well as to the mature forest) were converted to presence/absence data. The dissimilarity measured used was Bray-Curtis, without transformation or standardisation.

Results

First Warra survey:

Fig. 1 displays the first three coordinate axes of a PCoA for the first Warra survey, plotted as principal coordinate axis 2 (PC 2) vs. PC 1, and as PC 3 vs. PC 1. Although the separation between the harvested coupe and the mature forest is clearly visualised and was shown previously (see Ratkowsky 2007) to be significantly different using canonical analysis of principal coordinates (CAP), it is unclear whether there is any difference in the dispersion of points within each of the two coupes. The results from using PERMDISP2 are summarised by the P-values in Table 1 and the mean distances in Table 2.
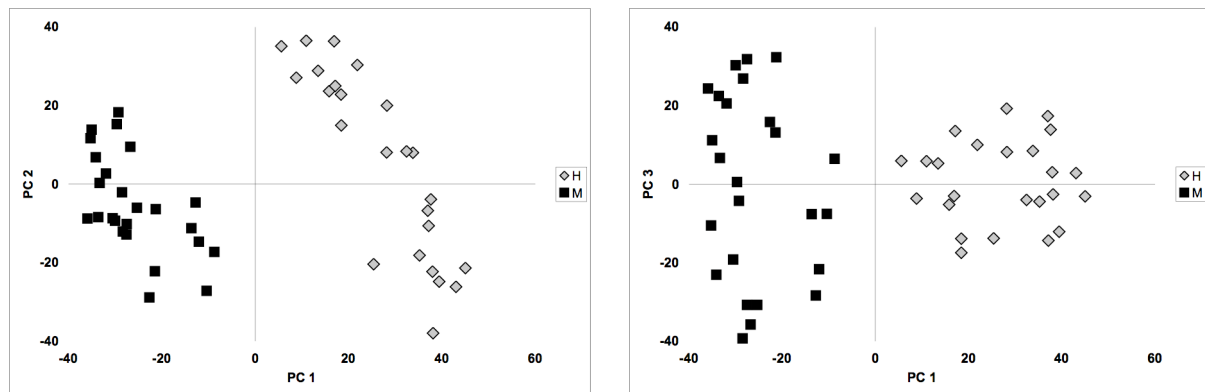

Figure 1. The first three principal coordinates of the PCoA analysis, first Warra survey (H=harvested coupe; M=mature forest).

Table 1. First Warra survey, showing the P-values obtained using PERMDISP2 for the test of differences between dispersions within the two coupes. There are two criteria, viz. deviations from centroids and deviations from spatial medians, respectively, and within each of those, values obtained from ANOVA tables and values obtained by permuting residuals, the residuals from the use of centroids being least-squares (LS) residuals and those from the use of spatial medians being least absolute deviation (LAD) residuals.

Deviations from centroids:	P-values
ANOVA tables	0.00074
Permutation of LS residuals	0.00267 (using 2999 permutations)
Deviations from spatial medians:	
ANOVA tables	0.00279
Permutation of LAD residuals	0.00300 (using 2999 permutations)

Table 2. First Warra survey, showing the mean distances from the group centroids or spatial medians and their standard errors.

Group	Mean dist. from centroid (± SE)	Mean dist. from spatial median (± SE)
Harvested	48.55 ± 1.11	48.49 ± 1.33
Mature	55.31 ± 1.46	55.25 ± 1.63

The four P-values are in agreement, suggesting that there is a significant difference in dispersion between the species lists of the 24 visits to the harvested coupe where non-null lists were obtained and the non-null species lists recorded during 25 visits to the mature forest, with $P \leq 0.003$ in each case. The mean distances in Table 2 show that the dispersion is significantly higher in the mature forest than in the harvested CBS coupe. The clear difference between the dispersions of the fungal lists obtained from the two coupes mirrors the reduced richness (131 species) of the harvested coupe compared to the richness of the mature forest (248 species). Not only is species richness lower in the harvested coupe, so are other measures of diversity. For example, very few mycorrhizal species were present in the harvested coupe. Although the lower richness and diversity of the harvested coupe matches the lower dispersion found for that coupe, the situation is less clear when one considers seasons as well as coupes, and applies PERMDISP2 to the eight unequal-sized groups formed from the combination of the two coupes with the four seasons in which visits were made. The results of doing this are given in Tables 3 and 4.

Table 3. First Warra survey, showing the P-values obtained using PERMDISP2 for the test of differences between dispersions within the eight groups. See caption to Table 1 for the explanation about the two different types of deviations and residuals.

Deviations from centroids:	P-values
ANOVA tables	0.2746
Permutation of LS residuals	0.5466 (using 4999 permutations)
Deviations from spatial medians:	
ANOVA tables	0.6239
Permutation of LAD residuals	0.6398 (using 4999 permutations)

Table 4. First Warra survey, showing the mean distances from the group centroids or spatial medians and their standard errors. Group identifications - first letter: H=harvested coupe, M=mature forest; second letter: A=autumn, S=summer, V=spring, W=winter.

Group	Mean dist. from centroid (± SE)	Mean dist. from spatial median (± SE)
HA	39.83 ± 4.24	39.58 ± 4.86
HS	44.53 ± 3.47	44.19 ± 5.99
HV	39.94 ± 1.91	39.92 ± 2.13
HW	40.94 ± 3.95	40.66 ± 4.87
MA	44.21 ± 2.56	44.13 ± 2.97
MS	51.82 ± 4.91	51.33 ± 7.08
MV	46.82 ± 3.17	46.59 ± 4.51
MW	40.74 ± 2.94	40.59 ± 3.57

From Table 3, one can conclude that there are no significant differences in dispersion between the coupes, irrespective of which statistic is employed. This is reflected by the much higher standard errors in Table 4 compared to those in Table 2. In Table 5, the mean distances from the centroids are listed again, this time in ascending order by group, along with the overall species richness for the group (i.e. the number of different species found in that group over all visits). It is obvious that there are major differences in group richness between the warmer, drier months of spring/summer and the cooler, wetter months of autumn/winter within each coupe, but that these differences are not reflected by differences in dispersion.

Table 5. First Warra survey, showing the mean distance from the centroid and the overall species richness for each group. Groups are listed in order of increasing mean dispersion and group identifications are the same as in Table 4.

Group	Mean distance from centroid (± SE)	Overall species richness
HA	39.8	83
HV	39.9	40
MW	40.7	151
HW	40.9	95
MA	44.2	146
HS	44.5	21
MV	46.8	66
MS	51.8	71

Second Warra survey:

Fig. 2 displays the first three coordinate axes of a PCoA for the second Warra survey, plotted as principal coordinate axis 2 (PC 2) vs. PC 1, and as PC 3 vs. PC 1. The separation of the group means of the harvested portion of the aggregated retention coupe (H) from the
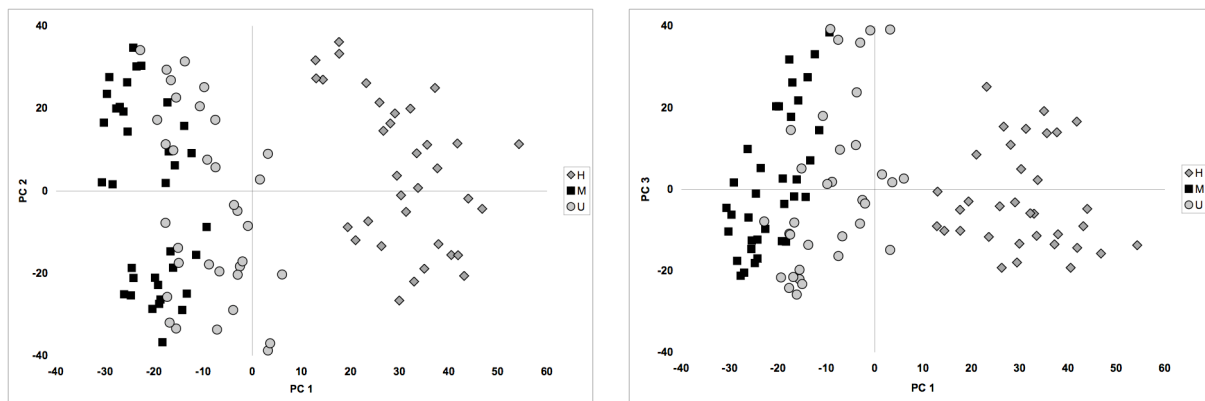

Figure 2. The first three principal coordinates of the PCoA analysis, second Warra survey (H=harvested portion of the aggregated retention coupe; U= unharvested portion of that coupe; M=mature forest).

unharvested portion (U) and from the mature forest (M) is clearly visualised in Fig. 2, a fact that was demonstrated previously for these data (Ratkowsky 2007) using CAP. The question here is whether there is any difference in the dispersion of points within the three experimental units. The results from PERMDISP2 are summarised by the P-values in Table 6 and the mean distances in Table 7.

Table 6. Second Warra survey, showing the P-values obtained using PERMDISP2 for the test of differences between dispersions within the three experimental units. See caption to Table 1 for the explanation about the two different types of deviations and residuals.

Deviations from centroids:	P-values
ANOVA tables	0.00000
Permutation of LS residuals	0.00033 (using 2999 permutations)
Deviations from spatial medians:	
ANOVA tables	0.00000
Permutation of LAD residuals	0.00033 (using 2999 permutations)

Table 7. Second Warra survey, showing the mean distances from the group centroids or spatial medians and their standard errors.

Group	Mean dist. from centroid (± SE)	Mean dist. from spatial median (± SE)
Harvested	52.24 ± 1.31	52.20 ± 1.42
Unharvested	61.51 ± 1.03	61.49 ± 1.11
Mature	57.15 ± 0.74	57.14 ± 0.81

The four P-values all indicate highly significant differences in dispersion between the species lists obtained during the visits to the three experimental units, with the low standard errors attached to the mean distances in Table 7 resulting in all pairwise differences being significant. As was the case with the results from the first Warra survey in Table 2, the mean distances in Table 7 show that the dispersion is significantly higher in mature or unharvested forest than in harvested areas. We now examine whether there is a relationship between species richness and dispersion.

Table 8, whose structure parallels that of Table 5, gives the mean distance from the centroid and the overall species richness for each group, sorted by increasing distance from the centroid. As was the case for the results in Table 5, there is no correlation between richness and dispersion, the main influence upon richness being the contrast between the warmer, drier months (spring and summer) and the cooler, wetter months (autumn and winter). This suggests that the interpretation of differences in dispersion, as revealed by the use of PERMDISP2, is more complex than simply equating high dispersion with high richness.

Table 8. Second Warra survey, showing the mean distance from the centroid and the overall species richness for each group. Groups are listed in order of increasing mean dispersion. Group identifications - first letter: H=harvested portion of the aggregated retention coupe; U=unharvested portion of that coupe; M=mature forest; second letter: A=autumn, S=summer, V=spring, W=winter.

Group	Mean distance from centroid (± SE)	Overall species richness
HW	39.7	92
HV	42.6	32
MW	46.0	134
MS	46.8	71
UW	47.8	101
MV	49.5	56
HA	50.6	85
MA	50.8	192
HS	51.0	23
UV	56.3	32
US	57.2	28
UA	58.0	105

Discussion

The two examples from macrofungal surveys at the silvicultural treatment trials in the Warra LTER site in southern Tasmania both lead to the conclusion that there are differences in dispersion between the experimental units. In both examples, the harvested areas had a significantly lower dispersion than the uncut areas. Since the species richness, as measured by the number of species present, was lower in the harvested areas than in the uncut areas, the results appear to support the findings of Anderson (2006), who, using data on spatial variation in fish, found that the biodiversity of fish assemblages was correlated with multivariate dispersion. In the present study, when seasonal effects were ignored, it did appear that differences in dispersion closely followed sampling unit differences, as harvested areas have a poorer mycota than uncut areas. Taking seasons into account reveals that the contrast between dry, warm months and wet, cool months is the most important factor determining richness, but these differences are not correlated with the dispersions listed in Tables 5 and 8. There are two major influences on species richness in the experimental units in the two examples studied here. The first is the contrast between harvested areas and unharvested ones, the harvested areas being open, unsheltered, drier areas containing a mycota that reflects disturbance, the species often being cosmopolitan ones, and largely saprotrophic. Mycorrhizal species are virtually absent. The unharvested areas, on the other hand, have a thorough mix of saprotrophic and mycorrhizal species. The other influence on species richness is the seasonality of fungal emergence. Although there are a few species that appear throughout the year when conditions are appropriate, usually after a rainfall event, the majority of species exhibit a strict seasonality, with those that emerge in spring or summer being different from those that appear in autumn and winter. Relatively small numbers of species occurring seasonally contribute towards differences in dispersion. The great advantage of having a program like PERMDISP2 is that it enables an accurate assessment to be made of differences in dispersion among experimental units. The present study demonstrated that there are differences in macrofungal species assemblages in experimental units subjected to different silvicultural treatments, and although it did not confirm a correlation between multivariate spread and fungal richness, it raised questions about why these differences exist. The search for the correct explanation is part of the stimulating scientific process of revealing the truth.

References

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

Anderson, M.J. & Robinson, J. (2001). Permutation tests for linear models. *Australian and New Zealand Journal of Statistics* **43**, 75-88.

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

Brown, M.B. & Forsythe, A.B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association* **69**, 364-376.

Cade, B.S. & Richards, J.D. (1996). Permutation tests for least absolute deviation regression. *Biometrics* **52**, 886-902.

Gates, G.M. & Ratkowsky, D.A. (2006). Macrofungal biodiversity in an aggregated retention coupe at the Warra LTER. A report prepared for Forestry Tasmania, Project F61857. 31pp.

Gates, G.M., Ratkowsky, D.A. & Grove, S.J. (2005). A comparison of macrofungi in young silvicultural regeneration and mature forest at the Warra LTER Site in the southern forests of Tasmania. *Tasforests* **16**, 127-152.

Levene, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics*, I. Olkin, S.G. Ghurye, W. Hoeffding, W.G. Madow, and H.B. Mann (eds), 278-292. Stanford University Press, Stanford, CA, U.S.A.

Manly, B.F.J. (1994). *Multivariate Statistical Methods: A Primer*. Second Edition. Chapman and Hall, Boca Raton, FL, U.S.A.

O'Brien, P.C. (1992). Robust procedures for testing equality of covariance matrices. *Biometrics* **48**, 819-827.

Ratkowsky, D.A. (2007). Visualising macrofungal species assemblage compositions using canonical discriminant analysis. *Australasian Mycologist* **26**: 75-85.

Ratkowsky, D.A. & Gates, G.M. (2008). Generalised canonical correlations analysis for explaining macrofungal species assemblages. *Australasian Mycologist* **27**: 33-40.

Van Valen, L. (1978). The statistics of variation. *Evolutionary Theory* **4**, 33-43, 202.

Warwick, R.M. & Clarke, K.R. (1993). Increased variability as a symptom of stress in marine communities. *Journal of Experimental Marine Biology and Ecology* **172**, 215-226.