Appendix A. Description of statistical tests and related methods.

Let \( Y \) be an \( N \times p \) matrix of \( i = 1, ..., N \) multivariate observations (rows) by \( k = 1, ..., p \) variables (columns). Let \( D = \{d_{ij}\} \) be a square symmetric \( N \times N \) matrix of distances (or dissimilarities) between all pairs of observations \( i = 1, ..., N \) and \( j = 1, ..., N \) with diagonal elements \( d_{ij} = 0 \) \( \forall \ i = j \). For example, the Euclidean distance is:

\[
d_{ij}^{(E)} = \sqrt{\sum_{k=1}^{p} (y_{ik} - y_{jk})^2}
\]  

(A.1)

In ecology, typically, some other resemblance may be calculated for non-negative count data, such as the Bray-Curtis measure:

\[
d_{ij}^{(BC)} = \frac{\sum_{k=1}^{p} |y_{ik} - y_{jk}|}{\sum_{k=1}^{p} (y_{ik} + y_{jk})}
\]  

(A.2)

which ranges from 0 to 1, and is also often expressed as a percent similarity:

\[
s_{ij}^{(BC)} = 100 \times (1 - d_{ij}^{(BC)})
\]

Another commonly-used measure, calculated on presence-absence data and directly interpretable as the proportion of unshared species, is the Jaccard measure:

\[
d_{ij}^{(J)} = \frac{\sum_{k=1}^{p} \psi(y_{ik}) - \psi(y_{jk})}{\left\{ \sum_{k=1}^{p} \psi(y_{ik}) - \psi(y_{jk}) + \psi(y_{ik} \cdot y_{jk}) \right\}}
\]  

(A.3)

where \( \psi(\cdot) \) is an indicator function such that
Appendix A, Anderson and Walsh

ψ(x) = \begin{cases} 
0 & \text{if } x = 0 \\
1 & \text{if } x > 0 
\end{cases}

Next, suppose the observations belong \emph{a priori} to \( \ell = 1, \ldots, g \) groups, with sample sizes \( n_1, n_2, \ldots, n_g \) and \( N = \sum_{\ell=1}^{g} n_\ell \). Let \( X \) be an \( N \times (g - 1) \) matrix of full rank containing orthogonal contrasts among the \( g \) groups. We can construct an \( N \times N \) projection matrix for this group structure according to the classical linear least-squares solutions to the Gauss-Markov normal equations (e.g., Plackett 1949) as:

\[ H = X[X'X]^{-1}X' \quad (A.4) \]

As outlined in McArdle and Anderson (2001), this can be used to obtain a partitioning of the multivariate variability inherent in matrix \( D \), by relying on the following transformation, due to Gower (1966) and highlighted for this purpose originally by McArdle (1991). Let matrix \( A \) consist of elements \( a_{ij} = -\frac{1}{2}d_{ij}^2 \), which, after centering on its rows and columns, gives a matrix directly interpretable as sums of squares and cross products (SSCP). Namely, matrix \( G \) of elements:

\[ g_{ij} = a_{ij} - \bar{a}_i - \bar{a}_j + \bar{a} \quad (A.5) \]

where \( \bar{a}_i = \frac{1}{N} \sum_{j=1}^{N} a_{ij} \), \( \bar{a}_j = \frac{1}{N} \sum_{i=1}^{N} a_{ij} \) and \( \bar{a} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \). Indeed, supposing each variable is centered on its mean to yield the centered data matrix \( Y_c \) with elements \{\( y_{ij}^{(C)} \)\} (namely, where for each column variable \( j \), we have \( y_{ij}^{(C)} = (y_{ij} - \bar{y}_j) \) and \( \bar{y}_j = \frac{1}{N} \sum_{i=1}^{N} y_{ij} \) and we have used Euclidean distances \( d^{(E)} \), then matrix \( G \) is equivalent to the outer product

\[ G = Y_c Y_c' \quad (A.6) \]

The total sum of squares is obtained as the trace (sum of diagonal elements) of this matrix. If Euclidean distance is used, this is equivalent to the trace of the inner product SSCP for \( Y \); i.e.
\[ tr[G] = tr[Y_cY_c'] = tr[Y_cY_c'] \]  
(A.7)

where “\(tr\)” indicates the trace of a matrix. If \(d^{(BC)}\), \(d^{(J)}\) or some other resemblance measure is used to construct \(D\), however, then the relationship between \(G\) and \(Y\) is not so straightforward.

Now the one-way PERMANOVA test-statistic (McArdle and Anderson 2001) is easily obtained through a partitioning of the \(G\) matrix to yield a pseudo-\(F\) statistic:

\[
F_{\text{pseudo}} = \frac{tr[HG]/v_1}{tr[(I-H)G]/v_2}
\]

(A.8)

where \(v_1 = (g-1)\), \(v_2 = (N-g)\) and \(I\) is an \(N \times N\) identity matrix with ones along the diagonal and zeros elsewhere. Note that this equation is equivalent to equation (4) in McArdle and Anderson (2001) because of the idempotency of matrix \(H\) (i.e., \(HH = H\)) and the fact that \(tr[HGH] = tr[HHG]\). Now, with just one variable \((p = 1)\) and Euclidean distances, the value of pseudo-\(F\) is precisely equal to the original univariate \(F\) ratio (Snedecor 1934) used in classical analysis of variance.

For the one-way case, a \(p\) value is calculated for PERMANOVA by a random re-ordering (permutation or randomization) of the \(N\) observation rows of \(Y\) relative to the fixed ordered list of \(n_1 + n_2 + \ldots, n_r\) labels for the groups (Edgington 1995, Manly 2006). This is equivalent to a random simultaneous re-ordering of the rows and columns of matrix \(D\), which maintains the inter-point structure in the multivariate space, but changes the group label with which each point is associated (Anderson 2001b). If the design is balanced, then all observations have an equal chance of falling into any particular group. If the design is unbalanced, then this is not true; however, the structure of the existing imbalance in the number of replicates per group is maintained under randomization and all re-orderings of the \(N\) observations relative to this structure are equally likely. The test-statistic is re-calculated for each randomization \((F^{(\pi)}, \text{say})\).
and a distribution of $F^{(\pi)}$ is thereby generated under a null hypothesis of ‘no differences among the groups’, conditional on the observed data. A random subset of all possible re-orderings can be used for accurate inference (Hope 1968). A $p$ value is calculated as the proportion of $F^{(\pi)}$ obtained under randomization that are greater than or equal to the observed value of pseudo-$F$.

Note also that (A.8) can be calculated directly from sums of squared distances (or dissimilarities) in matrix $D$ as described in Anderson (2001a); namely,

$$F_{\text{pseudo}} = \frac{(SS_T - SS_W)/v_1}{SS_W/v_2}$$  \hspace{1cm} (A.9)

where $SS_T$ is the sum of squared inter-point dissimilarities divided by the number of points:

$$SS_T = \left[ \sum_{i=1}^{(N-1)} \sum_{j=(i+1)}^{N} d_{ij}^2 \right] / N$$  \hspace{1cm} (A.10)

and $SS_W$ is the sum of squared inter-point dissimilarities within each group divided by the number of observations within that group, and then summed across all groups:

$$SS_W = \sum_{i=1}^{k} \left[ \sum_{i=1}^{(N-1)} \sum_{j=(i+1)}^{N} \zeta_{ij} \cdot d_{ij}^2 \right] / n_i$$  \hspace{1cm} (A.11)

Here and in what follows, $\zeta_{ij}$ is an indicator such that $\zeta_{ij} = 1$ if sample units $i$ and $j$ are in the same group, or else $\zeta_{ij} = 0$. Note also that $\text{tr}[G] = SS_T$. Legendre and Anderson (1999, see Theorem 1 in Appendix B therein) have shown the equivalence of (A.11) with the sum-of-squared distances to group centroids in the case of Euclidean distances. A geometric $F$ statistic constructed using sums of squared Euclidean distances to centroids within and between groups was described as a possible multivariate randomization test by Edgington (1995, pp. 188–191). Pillar and Orlóci (1996) had also suggested the use of a related test-statistic, $Q_B = (SS_T - SS_W)$, which, in the specific case of a one-way ANOVA model only, is monotonic on the pseudo-$F$.
statistic given in (A.9) under permutation, as the degrees of freedom ($v_1$ and $v_2$), and also $SS_T$ will all remain constant for any random re-ordering of the data, so identical $p$ values will be obtained. It may be noted here that the PERMANOVA test-statistic has the advantage of being constructed as a pivotal test-statistic (i.e., $F_{pseudo}$ calculated from a Euclidean distance matrix for 1 variable is equivalent to the classical univariate $F$ statistic), so should not be affected adversely by the presence of nuisance parameters and can be easily extended to multi-way designs. It is also clearly not restricted to the use of the Euclidean distance. Note also, however, that the construction of the test effectively relies holistically on sums of squared distances within (and between) groups, without any regard whatsoever for the particular direction of those distances within the multivariate space, which distinguishes it from the classical MANOVA test statistics.

Next, the ANOSIM statistic of Clarke (1993) is easily described as a function of the ranks of matrix $D$. There will be $M = N(N-1)/2$ inter-point distance values $d_{ij}$ within the upper-triangular (or, equivalently, the lower-triangular) portion of matrix $D$ (excluding the diagonal); namely, for $i = 1, ..., (N-1)$ and $j = (i+1), ..., N$. Let the values $d_{ij}$ be replaced by the rank order of their values, $r_{ij}$, where the lowest value of $d_{ij}$ is given a value of $r_{ij} = 1$ and the highest value of $d_{ij}$ is given a value of $r_{ij} = M$. The ANOSIM test-statistic (Clarke 1993) is then given by:

$$R = \frac{(\bar{r}_g - \bar{r}_w)}{M/2}$$  \hspace{1cm} (A.12)

where $\bar{r}_w$ is the average of the ranked dissimilarities between observations within the same group:
A $p$ value is obtained for the one-way case in the same way for ANOSIM as for PERMANOVA, using random re-orderings of the observations relative to the group structure and calculating a distribution of $R^{(X)}$ against which the value of $R$ for the original ordering is then compared to provide a $p$ value for the test.

The Mantel test was first described as a test of association between two resemblance matrices (Mantel 1967, Mantel and Valand 1970). For a given set of $N$ observations, suppose there are two resemblance matrices; for example, the first might be dissimilarities based on species data while the second might be geographic distances. A cross-product (or Pearson or Spearman correlation coefficient) is calculated between the matched paired values in the two matrices and this is compared with the distribution of the same under random re-ordering of the original $N$ observations for one of the two matrices.

The Mantel test may also be used for a goodness-of-fit test between a matrix of resemblances and a model matrix (Legendre and Legendre 1998, see pp. 555–557). For example, to model the group structure as in ANOVA, the $N \times N$ model matrix may have zeros in place of the between-group distances and ones in place of the within-group distances. In other words, the model matrix consists of the indicators $\zeta_{ij}$, as defined in equation (A.11) above. A cross-
product between the sub-diagonal elements (as these matrices are symmetric) then simply gives the sum of the within-group dissimilarities,

\[
    z_{(1,0)} = \sum_{i=1}^{N+1} \sum_{j=(i+1)}^{N} \zeta_{ij} \cdot d_{ij}
\]  

(A.15)

Note that the value of \( z_{(1,0)} \) will decrease with increasing degree of “clumping” within groups, so the \( p \)-value for the test using this statistic must be calculated as the proportion of values of \( z_{(1,0)}^{(\pi)} \) that are less than or equal to the observed value of \( z_{(1,0)} \).

For one-way designs, other arbitrary contrast coefficients can be used in the indicator model matrix to distinguish the within-group versus the between-group dissimilarities, yet would yield the same result. For example, consider the use of \((-1,+1)\) rather than \((0,1)\) to give:

\[
    z_{(-1,+1)} = \sum_{i=1}^{N+1} \sum_{j=(i+1)}^{N} (1 - \zeta_{ij}) \cdot d_{ij} - \sum_{i=1}^{N+1} \sum_{j=(i+1)}^{N} \zeta_{ij} \cdot d_{ij}
\]  

(A.16)

As the sum of all the dissimilarities in the sub-diagonal matrix of \( D \) is a constant, (A.16) will yield a cross-product that is monotonic with (A.15) under permutation, so will result in equivalent \( p \) values for the Mantel test. Furthermore, Legendre and Legendre (1998, p. 562) demonstrated the clear relationship between the Mantel test and ANOSIM. Specifically, in the model matrix, let the code for within-group resemblances be:

\[
    c_w = \frac{-\left(\sum_{i=1}^{g} n_i (n_i - 1)/2\right)}{(M/2)}
\]  

(A.17)

and the code for the between-group resemblances be:

\[
    c_b = \frac{\left(M - \sum_{i=1}^{g} n_i (n_i - 1)/2\right)}{(M/2)}
\]  

(A.18)
then the Mantel cross-product statistic $z$ yields a test statistic with an equivalent structure to the $R$-statistic of ANOSIM, but it is calculated on the averages of the between-group and within-group dissimilarity values themselves, rather than on the averages of their ranks, namely:

$$z_{(c_w,c_x)} = \frac{(\bar{d}_b - \bar{d}_w)}{M/2}$$  \hspace{1cm} (A.19)

The use of (A.19) will yield an equivalent $p$-value for the one-way model as the use of either (A.15) or (A.16). It will not, however, yield the same results as the ANOSIM $R$ statistic in (A.12), which is based on ranks. The form of the Mantel test-statistic given in (A.19) was the one we used in our simulations.

To draw further parallels, the Mantel test also has a clear and close kinship with the resemblance-based permutation test statistic described by Good (1982) and Smith et al. (1990), namely $\frac{\bar{d}_b}{\bar{d}_w}$. This would also be monotonic under permutation with any of (A.15), (A.16) or (A.19), and thus would yield identical $p$ values to the Mantel test for these one-way model simulations. (Although originally described in terms of average similarity, $s$, rather than dissimilarity, $d$, generally one can easily write a simple inverse function $d = 1 - s$, so the result still holds).

Other important parallels can be drawn between the methods we have included in our simulations and the multi-response permutation procedure (MRPP, Mielke et al. 1981, Mielke and Berry 2001). The general formulation of the MRPP statistic is given by

$$\delta = \sum_{\ell=1}^{g} C_{\ell} \xi_{\ell}$$  \hspace{1cm} (A.20)

where $C_{\ell} > 0$ is a group weight, $\sum_{\ell=1}^{g} C_{\ell} = 1$, and

$$\xi_{\ell} = \frac{1}{n_{\ell}(n_{\ell}-1)/2} \sum_{i=1}^{N_{\ell}} \sum_{j=i+1}^{N_{\ell}} \xi_{ij}^{(\ell)} \Delta_{ij}$$  \hspace{1cm} (A.21)
Appendix A, Anderson and Walsh

is the average of pairwise distance function values $\Delta_{ij}$ within each group $\ell$ ($\ell = 1, \ldots, g$), where $\varphi^{(\ell)}_{ij}$ is an indicator such that $\varphi^{(\ell)}_{ij} = 1$ if sample units $i$ and $j$ are both within group $\ell$. The test statistic gets smaller with increased “clumping” of observations within groups, so the $p$-value for the MRPP test is calculated as the proportion of values of $\delta^{(x)}$ under permutation that are less than or equal to the observed value of $\delta$. If we let $\Delta_{ij} = d_{ij}$ and assign the weights $C_{\ell}$ to be proportional to the group sample sizes, i.e., $C_{\ell} = n_{\ell} / N$, then we have the direct result that

\[ \delta = z_{(c_w', 0)} , \quad \text{where} \quad c_w' = 2 / [N(n_{\ell} - 1)] . \]

Thus, the MRPP test is equivalent to the Mantel test coded in this way for either balanced or unbalanced designs. Note, however, that under permutation the test statistic $z_{(c_w', 0)}$ will only be monotonic on the Mantel test statistics of $z_{(1, 0)}$, $z_{(+1, -1)}$ or $z_{(c_w', c_{\ell})}$ (as given in equations (A.15), (A.16), and (A.19) above, respectively) when there are equal numbers of replicate sample units per group. Thus, the MRPP test using $\delta$ (and with $C_{\ell} = n_{\ell} / N$) will yield equivalent permutation $p$-values to these more general implementations of the Mantel test only for balanced one-way designs.

Mielke and Berry (2001, p. 11) have also shown, for the one-way case, that MRPP, when based on squared Euclidean distances for a single variable, yields $p$ values equivalent to the univariate $F$ statistic under permutation. It is therefore easy to show here the relationship between MRPP and PERMANOVA more generally for one-way models. First, it is important that the distances be squared, i.e., let $\Delta_{ij} = d_{ij}^2$. Then, let the weights be $C_{\ell} = (n_{\ell} - 1)/(N - g)$, and the relationship between the PERMANOVA statistic of (A.9) and the MRPP statistic of (A.20) is

\[ F_{\text{pseudo}} = \frac{2SS_T - v_i \delta}{v_i \delta} . \quad \text{(A.22)} \]
As the values of $SS_T$, $v_1$ and $v_2$ are all constant under permutation, $\delta$ based on squared dissimilarities with this choice of weights will yield a $p$ value for MRPP that is equivalent to PERMANOVA. This relationship holds for either balanced or unbalanced one-way designs.

In this study, the resemblance-based permutation tests (PERMANOVA, ANOSIM and Mantel) were compared with one another and with the classical MANOVA test statistic described by Pillai (1955). Given that the $SSCP$ matrix for the within-group variation is $W = Y'_c(I - H)Y_c$ and the $SSCP$ matrix for the between-group variation is $B = Y'_cHY_c$, then Pillai’s trace is defined as $V^{(s)} = tr[B(W + B)^{-1}]$. To obtain a $p$-value, the following $F$-approximation (Pillai 1955) was used:

$$F_{\text{pillai}} = \frac{(2t + s + 1)V^{(s)}}{(2q + s + 1)(s - V^{(s)})}$$

(A.23)

with $s(2q + s + 1)$ and $s(2t + s + 1)$ degrees of freedom, where, $s = \min(v_1, p)$, $q = \frac{1}{2}(\frac{v_1}{2} - 1)$ and $t = \frac{1}{2}(\frac{v_2}{2} - p - 1)$. Note that we must have $v_2 \geq p$. Also note that for Euclidean distances only, we can write the PERMANOVA pseudo-$F$ as:

$$F_{\text{pseudo}} = \frac{tr[B]/v_1}{tr[W]/v_2}$$

(A.24)

which highlights how it differs from Pillai’s trace. Pseudo-$F$ is a ratio of two traces, each of these being a pure sum of individual sums of squares, thus ignoring all off-diagonal cross-products and hence correlation structure. For Pillai’s trace, in contrast, the off-diagonal cross-product terms will play a role through the calculation of an inverse followed by the matrix multiplication, both of which occur prior to taking the trace.
LITERATURE CITED

Anderson, M. J. 2001b. Permutation tests for univariate or multivariate analysis of variance and  
regression. Canadian Journal of Fisheries and Aquatic Sciences 58:626–639.
Gower, J. C. 1966. Some distance properties of latent root and vector methods used in  
Statistical Society, Series B 30:582–598.
Legendre, P., and M. J. Anderson. 1999. Distance-based redundancy analysis: testing  
multispecies responses in multifactorial ecological experiments. Ecological Monographs  
69:1–24.
Amsterdam, The Netherlands.
Manly, B. F. J. 2006. Randomization, bootstrap and Monte Carlo methods in biology, 3rd  
Mantel, N. 1967. The detection of disease clustering and a generalized regression approach.  


