On Distance-Based Permutation Tests for Between-Group Comparisons

Philip T. Reiss,[1,2,*] M. Henry H. Stevens,[3] Zarrar Shehzad,[1]
Eva Petkova,[1,2] and Michael P. Milham[1]

[1]Department of Child and Adolescent Psychiatry, New York University, New York 10016, U.S.A.
[2]Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York 10962, U.S.A.
[3]Department of Botany, Miami University, Oxford, Ohio 45056, U.S.A.
[*]*email:* phil.reiss@nyumc.org

SUMMARY. Permutation tests based ondistances among multivariate observations have found many applications in the biological sciences. Two major testing frameworks of this kind are multiresponse permutation procedures and pseudo-F tests arising from a distance-based extension of multivariate analysis of variance. In this article, we derive conditions under which these two frameworks are equivalent. The methods and equivalence results are illustrated by reanalyzing an ecological data set and by a novel application to functional magnetic resonance imaging data.

KEY WORDS: Analysis of distance; Bray–Curtis distance; Functional magnetic resonance imaging; Multiresponse permutation procedure; Multivariate analysis of variance; Pseudo-F test.

1. Introduction

In a variety of scientific settings, researchers are interested in comparing two or more groups in terms of multivariate outcomes for which standard multivariate analysis of variance (MANOVA) methodology is inappropriate. Possible reasons for the inapplicability of MANOVA include nonnormal observations, heterogeneous dispersion matrices, missing data, or a greater number of variables than observations (Gower and Krzanowski, 1999). One way to proceed in such settings is to deploy methods whose starting point is not the observations themselves, but the distances among them.

In this article, we focus on two such approaches, both of which proceed by constructing a test statistic based on the interpoint distances and comparing it to a permutation distribution. The first approach, multiresponse permutation procedures (MRPP) (Mielke, Berry, and Johnson, 1976; Mielke and Berry, 2007) uses a (weighted) average within-group distance as the test statistic (cf. Good, 1982; Smith, Pontasch, and Cairns, 1990). The second approach, originating in the field of ecology and sometimes called permutational MANOVA, was introduced by McArdle and Anderson (2001) and Anderson (2001, 2005). They described a method for partitioning variation inherent in dissimilarity or distance matrices, and defined a "pseudo-F" statistic analogous to the usual univariate F statistic for testing individual terms in ANOVA or more general linear models. This inferential approach has been applied recently in the genetics literature as well (Wessel and Schork, 2006; Zapala and Schork, 2006; Nievergelt, Libiger, and Schork, 2007). Such a partitioning is formulated elsewhere as "analysis of distance" (Gower and Krzanowski,

1999; Krzanowski, 2002, 2006), although the latter work aims primarily to develop graphical and other descriptive methodology as opposed to significance tests.

Permutation tests based on both MRPP and the pseudo-F statistic are implemented in the R package vegan (Oksanen et al., 2008) for quantitative ecology. Each approach has found many applications in that discipline. However, the theoretical relationship between the MRPP and pseudo-F statistics is not well understood. This article seeks to provide some insight into this relationship. We derive conditions, pertaining to the distance function and/or to the design, under which the two test procedures are in fact equivalent.

This article makes three main contributions. First, we present an algebraic formulation of the pseudo-F statistic that encompasses testing a full model against either a null model (as in McArdle and Anderson, 2001) or an intermediate model. Second, we derive results showing the equivalence of this statistic and the MRPP statistic in certain fundamental cases. Third, we illustrate the versatility of these methods both by re-examining a well-studied ecological data set and by presenting a novel approach to a high-dimensional, low-sample-size set of neuroimaging data. An overarching goal of this article is to unify different lines of work on distance-based permutation tests, and thereby to help foster the continued growth of this area of biometric methodology.

2. A Motivating Application: Functional Connectivity of the Human Brain

Functional MRI records the blood oxygen level dependent (BOLD) time series, an index of brain activity, at each of
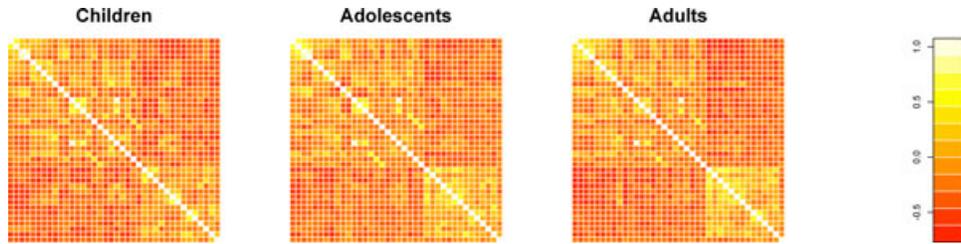
Figure 1. Mean correlation matrices among 38 brain regions of interest, for the three age groups indicated. The first 24 and last 14 regions belong to two proposed anticorrelated systems known, respectively, as the task-positive and task-negative networks. The differentiation of these networks can be seen to progress from childhood to adulthood. This figure appears in color in the electronic version of this article.

a grid of locations in the subject's brain. Traditional fMRI studies seek to identify brain regions whose BOLD signal indicates a response to some stimulus. A growing body of work has instead examined the brain in its "resting state." By scanning subjects while they attend to no stimulus in particular, one can investigate which brain regions' activity levels tend to vary in tandem, and how the brain is organized into functional networks. A key finding was the identification by Fox et al. (2005) of two widely distributed brain networks playing opposite roles: a "task-positive" network showing higher activation during overt task performance, and a "task-negative" or "default mode" network more active during rest.

Some recent work (e.g., Fair et al., 2008; Kelly et al., 2009) has examined how functional networks of this kind mature from childhood to adulthood. Consider Figure 1, which displays resting-state correlation matrices among 38 regions of interest (ROIs) from Toro, Fox, and Paus (2008), averaged, respectively, over 13 child subjects, 13 adolescents, and 26 adults. Due to the high dimension of the correlation matrices, we employed the shrinkage methodology of Schäfer and Strimmer (2005) to estimate them. These plots suggest a progressive differentiation from childhood to adulthood between the first 24 ROIs, which belong to the task-positive network, and the last 14, belonging to the default mode network. But how can we formally test this subjective impression?

One class of approaches is referred to in the neuroimaging literature as "mass-univariate": essentially one tests for among-group differences separately at each of the $38 \times \frac{38-1}{2} = 703$ distinct connections between ROIs (using, say, a one-way ANOVA F-test), and then use some multiple comparisons method to infer which of these tests yield significant results. For instance, Fair et al. (2008) employ the false discovery rate; Section 7.2 below describes a mass-univariate permutation test procedure. A possible limitation of such methods is that they can detect only particular connections displaying unusually large differences among groups. We were interested in a test sensitive to differences in the overall pattern of connections, which might not be reflected in such large differences for any particular connection. We therefore turned to the distance-based procedures considered in this article.

3. Overview of the Two Approaches

The methods discussed here take as their starting point an $n \times n$ symmetric matrix $\boldsymbol{D} = (d_{ij})_{1 \leqslant i, j \leqslant n}$ representing nonnegative distances among n observations. It is assumed that $d_{ii} = 0$ for each i and that $d_{ij} > 0$ for some $i \neq j$. The n obser-

vations are divided into some sort of a priori groups that, according to the null hypothesis, are indistinguishable from each other. This hypothesis is tested by permutation methods. Let π denote a permutation of the numbers $\{1, \ldots, n\}$, i.e., a one-to-one function from $\{1, \ldots, n\}$ to itself, and let Π be a group of permutations such that, under the null hypothesis, $d_{\pi(i)\pi(j)}$ has the same distribution as d_{ij} $(1 \leqslant i, j \leqslant n)$ for each $\pi \in \Pi$. The null hypothesis can thus be tested by computing a test statistic t that depends on the d_{ij}'s, and referring t to the null distribution of t_π, the corresponding statistic calculated from the $d_{\pi(i)\pi(j)}$'s, for $\pi \in \Pi$. In practice, this permutation distribution is usually approximated by Monte Carlo simulation. In some settings, it may be preferable to permute residuals rather than raw distances (see Legendre and Anderson, 1999; Anderson and ter Braak, 2003), but (other than the test of an interaction term in Section 7.2) permutation of residuals lies beyond the scope of this article.

3.1 *MRPP*

Suppose the n observations are divided among g a priori groups $\mathcal{G}_1, \ldots, \mathcal{G}_g$ of size $n_1, \ldots, n_g$. (More generally one may include a $(g + 1)$th "excess" group consisting of "unclassified" observations, but we will assume this group to be empty.) The MRPP statistic is given by

$$\delta = \sum_{k=1}^{g} C_k \frac{2}{n_k(n_k - 1)} \sum_{i < j, \ i, j \in \mathcal{G}_k} \Delta_{ij}, \qquad (1)$$

where Δ_{ij} denotes a measure of dissimilarity between the ith and jth observations and $C_1, \ldots, C_g$ are weights summing to 1. In Mielke and Berry's work, Δ_{ij} typically equals d_{ij}^v for some $v > 0$, most often d_{ij} or d_{ij}^2, where d_{ij} denotes a metric such as Euclidean distance. In general these authors advocate taking Δ_{ij} to be a metric rather than a squared metric, based on robustness considerations and a notion of congruence between the data and analysis spaces (Mielke, 1986). (Recall that a distance or dissimilarity function is a metric if, in addition to the assumptions given at the beginning of Section 3, all distinct pairs of points have positive distance, and the triangle inequality holds.) The choice of weights

$$C_k = \frac{n_k - 1}{n - g} \qquad (2)$$

is efficient in the sense that it minimizes the asymptotic order of the permutation distribution's variance (Mielke and Berry, 2007). The alternative weights

$$C_k = \frac{n_k(n_k - 1)}{\sum_{m=1}^{g} n_m(n_m - 1)} \quad (3)$$

reduce (1) to $\delta = \frac{2}{\sum_{k=1}^{g} n_k(n_k-1)} \sum_{k=1}^{g} \sum_{i<j, \ i,j \in \mathcal{G}_k} \Delta_{ij}$, which is simply the mean of the within-group differences (cf. Mantel and Valand, 1970). However, this choice is not efficient in the above sense.

A point of nomenclature: the terms "distance" and "dissimilarity" are ordinarily treated as synonyms in multivariate analysis, but here, to avoid possible confusion between d_{ij} and Δ_{ij}, we will refer to d_{ij} as a distance and to Δ_{ij} as a dissimilarity.

Consider next a randomized block design, in which the n observations are arranged in g groups and b blocks, with one observation in each group/block pair. For this setting Mielke and Iyer (1982) propose a modification of MRPP, the multivariate randomized block permutation procedure (MRBP) statistic

$$\delta = \left[g \binom{b}{2} \right]^{-1} \sum_{i=1}^{bg} \sum_{j>i, j \sim_b i} \Delta_{ij}, \quad (4)$$

where $j \sim_b i$ means observations j, i are in the same block. Although the MRPP and MRBP statistics look somewhat different, they are motivated by the same underlying idea: both statistics represent average within-group differences that, under the alternative hypothesis, should lie in the left tail of the permutation distribution. Mielke and Berry (2007) extend the MRPP methodology to more general linear models, but these are not considered here.

3.2 *Pseudo-F Tests*

Let $\mathbf{A} = (-\frac{1}{2} d_{ij}^2)_{1 \leqslant i,j \leqslant n}$, and let $\mathbf{G} = (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\mathbf{A}(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)$, where $\mathbf{1}$ is a vector of n 1's. $\mathbf{G}$ is the centered matrix used in Gower's (1966) development of principal coordinate analysis. Consider three "partial" design matrices $\mathbf{X}_0 = \mathbf{1}, \mathbf{X}_1$, and $\mathbf{X}_2$, such that for $k = 0, 1, 2$, $\mathbf{X}_k$ is an $n \times m_k$ matrix of rank $m_k(<n)$. Without loss of generality, for distinct $k, l \in \{0, 1, 2\}$, $\mathbf{X}_k^T \mathbf{X}_l = \mathbf{0}_{m_k \times m_l}$. (Otherwise we can use the Gram–Schmidt process to obtain modified design matrices for which this orthogonality does hold.) We can then consider a nested sequence of design matrices, representing a null (intercept-only) model, an intermediate model, and a full model:

$$\mathbf{1}, \ (\mathbf{1} \quad \mathbf{X}_1), \ \mathbf{X} = (\mathbf{1} \quad \mathbf{X}_1 \quad \mathbf{X}_2).$$

By definition $m_0 = 1$; m_1 may be zero (in which case $\mathbf{X}_1$ is null), whereas m_2 is assumed positive. The dimension of $\mathbf{X}$ is $n \times m$, where $m = 1 + m_1 + m_2$. For $k = 0, 1, 2$, let $\mathbf{H}_k = \mathbf{X}_k(\mathbf{X}_k^T \mathbf{X}_k)^{-1} \mathbf{X}_k^T$ be the hat matrix associated with $\mathbf{X}_k$, i.e., the matrix of projection onto the column space of $\mathbf{X}_k$; thus $\mathbf{H}_0 = \mathbf{1}\mathbf{1}^T/n$. Similarly let

$$\mathbf{H} = (h_{ij})_{1 \leqslant i,j \leqslant n} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \mathbf{H}_0 + \mathbf{H}_1 + \mathbf{H}_2.$$

If $m_1 = 0$ then take $\mathbf{H}_1 = 0$. The pseudo-F statistic is then given by

$$F^* = \frac{\mathrm{tr}(\mathbf{H}_2 \mathbf{G} \mathbf{H}_2)/m_2}{\mathrm{tr}[(\mathbf{I} - \mathbf{H})\mathbf{G}(\mathbf{I} - \mathbf{H})]/(n - m)}. \quad (5)$$

See Web Appendix A for further details. F^* can be presented in an analysis of distance table (analogous to an ANOVA table) as in McArdle and Anderson (2001) or Anderson (2001). When $m_1 = 0$, i.e., when testing the full model against the intercept-only model, (5) reduces to the original expression of McArdle and Anderson (2001),

$$F^* = \frac{\mathrm{tr}(\mathbf{HGH})/(m-1)}{\mathrm{tr}[(\mathbf{I} - \mathbf{H})\mathbf{G}(\mathbf{I} - \mathbf{H})]/(n - m)}. \quad (6)$$

The pseudo-F statistic does not in general have an F-distribution under the null hypothesis; rather, like the MRPP and MRBP statistics, it has an unknown null distribution, so its significance is assessed with a permutation test.

The motivation for the pseudo-F statistic (6), and hence for (5), can be summarized as follows. Suppose first that $\mathbf{D}$ represents the Euclidean distances among univariate outcomes $y_1, \ldots, y_n$. Then F^* reduces to the usual F statistic for testing the model regressing these values on a design matrix $\mathbf{X}$ versus the null model, as can be inferred from an elementary identity relating the sum of squared residuals to the sum of squared distances (e.g., Legendre and Legendre, 1998, equations (8.5) and (8.6)). More generally, suppose that $\mathbf{D}$ represents the Euclidean distances among $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathcal{R}^p$, for some p that may be larger than n. Even if $\mathbf{D}$ was not originally formed from the distances among such a set of points, it can be shown that such points do exist (i.e., $\mathbf{D}$ "has the Euclidean property") if and only if $\mathbf{G}$ is positive definite (Mardia, Kent, and Bibby, 1979); indeed, the principal coordinates of Gower (1966) provide such a set of points. In this case, if we performed separate regressions of each component of the $\mathbf{y}_i$'s with design matrix $\mathbf{X}$, and if A_k, B_k were the numerator and denominator of the F statistic for the kth component, then we would have $F^* = \sum_{k=1}^{p} A_k / \sum_{k=1}^{n} B_k$ (Legendre and Anderson, 1999). Finally, observe that F^* can be calculated even if $\mathbf{G}$ is not positive definite ($\mathbf{D}$ is not Euclidean). In short, the pseudo-F statistic is a generalization of the classical F statistic that can be calculated directly from the distance matrix $\mathbf{D}$, whether or not $\mathbf{D}$ is Euclidean. This last property is especially helpful in the field of ecology, where analyses often employ non-Euclidean distance measures such that defined by Bray and Curtis (1957):

$$d(\mathbf{x}, \mathbf{y}) = \frac{\sum_{k=1}^{p} |x_k - y_k|}{\sum_{k=1}^{p} (x_k + y_k)}, \quad (7)$$

where $\mathbf{x}$, $\mathbf{y}$ are p-dimensional vectors of nonnegative numbers.

The quantity

$$\frac{\mathrm{tr}(\mathbf{H}_k \mathbf{G} \mathbf{H}_k)}{\mathrm{tr}(\mathbf{G})} \quad (8)$$

can be viewed as the proportion of variation explained by $\mathbf{X}_k$ (Legendre and Anderson, 1999; McArdle and Anderson, 2001). This expression, like the pseudo-F statistic, can be motivated as a generalization of the corresponding expression for ordinary ANOVA, by appealing to a principal coordinates argument (Gower and Krzanowski, 1999).

Expression (5) is sufficiently general to encompass the two simple designs for which we present equivalence results below: the one-way design (see Section 5) and the randomized complete block design (Section 6). Modifications would be required for more complex designs involving nested or random effects or interactions. `PERMANOVA+` software (Anderson, Gorley, and Clarke, 2008), an add-on to the `PRIMER` statistical package for ecologists (Clarke and Gorley, 2006), implements pseudo-F tests for ANOVA and other regression models with general multiway balanced or unbalanced designs.

In order to derive the results below regarding permutation tests, it will be helpful to define explicit notation for permutation of distances. If we denote by e_i the n-dimensional vector with 1 in the ith position and 0 elsewhere, then applying permutation π to the distance matrix D means replacing the latter with

$$D_\pi \equiv (d_{\pi(i)\pi(j)})_{1\leqslant i,j\leqslant n} = E_\pi D E_\pi^T, \qquad (9)$$

where $E_\pi = (e_{\pi(1)} \ldots e_{\pi(n)})^T$. We can similarly define A_π and G_π by replacing D with D_π in the above, from which we obtain $A_\pi = E_\pi A E_\pi^T$ and (using the fact that $I - 11^T/n$ commutes with E_π)

$$G_\pi = E_\pi G E_\pi^T. \qquad (10)$$

4. Reduction of the Pseudo-F to a Simpler Test Statistic

In this article, we refer to two permutation tests as equivalent if, whenever both are conducted using the same set of permuted distance matrices (9), the two tests yield the same p-value. The following result provides conditions under which F^* is equivalent in this sense to a simpler test statistic. This equivalence may be of some interest in its own right, and will be used in Sections 5 and 6 to connect pseudo-F tests with MRPP and MRBP. Proofs of this result and of those in the sequel are given in Web Appendix B.

THEOREM 1: *If*

there exists a constant $K < \dfrac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}d_{ij}^2$

such that $tr(H_1 A_\pi) = K$ *for all* $\pi \in \Pi$, $\qquad (11)$

and if

$$tr(-HA_\pi) \geqslant 0 \text{ for all } \pi \in \Pi, \qquad (12)$$

then a permutation test based on F^* *(with rejection region in the right tail of the permutation distribution) is equivalent to a test based on* $tr(-HA)$ *(with rejection region in the left tail).*

For the $m_1 = 0$ case, we can identify two more transparent conditions—one pertaining to the design, one to the distance function—either of which implies the conclusion of Theorem 1.

PROPOSITION 1: *If* $m_1 = 0$ *then the equivalence of Theorem 1 holds*

(i) *if* $h_{ij} \geqslant 0$ *for all* $i, j \in \{1, \ldots, n\}$, *or*
(ii) *if* G *is positive semidefinite (i.e.,* D *is Euclidean).*

The test statistic $tr(-HA) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}h_{ij}d_{ij}^2$ of Theorem 1 is a weighted sum of the squared distances, where the weights depend on the design. Thus, as the next two sections will show, this test statistic serves as a bridge from the pseudo-F statistic to the MRPP and MRBP statistics, which are particular weighted sums of dissimilarities for two simple designs. More generally, the equivalence of the pseudo-F statistic and $tr(-HA)$ (under the conditions of Theorem 1) provides some insight into what sort of models the pseudo-F test treats as "good" models (i.e., models in favor of which we would reject the null model)—namely, those for which $tr(-HA)$ is low, or equivalently, for which h_{ij} tends to be large when d_{ij} is small. In the context of a linear model with univariate outcomes $y_1, \ldots, y_n$, h_{ij} can be interpreted as $\frac{\partial \hat{y}_i}{\partial y_j}$ or the effect of y_j on $\hat{y}_i$. Thus, the above heuristic criterion for a good model stipulates that observations that are closer to the ith observation should have greater impact on the ith fitted value. This is precisely the idea that underlies local regression methodology (e.g., Fan and Gijbels, 1996).

5. Application to One-Way Design

Suppose we wish to test for differences among groups $\mathcal{G}_1, \ldots, \mathcal{G}_g$ of sizes $n_1, \ldots, n_g$ respectively, and that the $n = n_1 + \cdots + n_g$ rows and columns of H are arranged by these groups. Then

$$H = \begin{pmatrix} \mathbf{1}_{n_1}\mathbf{1}_{n_1}^T/n_1 & 0 & \ldots & 0 \\ 0 & \mathbf{1}_{n_2}\mathbf{1}_{n_2}^T/n_2 & \ldots & 0 \\ \vdots & . & \ddots & \vdots \\ 0 & \ldots & 0 & \mathbf{1}_{n_g}\mathbf{1}_{n_g}^T/n_g \end{pmatrix},$$

and thus

$$tr(-HA) = \sum_{k=1}^{g}\frac{1}{2n_k}\sum_{i,j\in\mathcal{G}_k}d_{ij}^2 = \sum_{k=1}^{g}\frac{1}{n_k}\sum_{i<j,\ i,j\in\mathcal{G}_k}d_{ij}^2.$$

On the other hand, if we take dissimilarity $\Delta_{ij} = d_{ij}^2$ and weights (2) then

$$tr(-HA) = \frac{n-g}{2}\delta. \qquad (13)$$

Since condition (i) of Proposition 1 applies, (13) leads to the following result.

PROPOSITION 2: *A pseudo-F test of the group effect, with distance* d_{ij}, *is equivalent to an MRPP test with dissimilarity* $\Delta_{ij} = d_{ij}^2$ *and weights* $C_k = \frac{n_k-1}{n-g}$.

In Section 7.1, we illustrate how the MRPP with dissimilarity $\Delta_{ij} \neq d_{ij}^2$ (e.g., $\Delta_{ij} = d_{ij}$) and/or weights other than those in Proposition 2 can yield different conclusions than the pseudo-F test.

We remark that Proposition 2 can be proved without reference to Proposition 1, by combining (13) with Lemma 1 of

Web Appendix B to express δ as a strictly decreasing function of F^*:

$$\delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2}{n\left[n - g + (g-1)F^*\right]}.$$

This equality directly generalizes the relationship between δ and the ordinary F statistic in the setting of one-way ANOVA (Mielke and Berry, 2007, Section 2.9).

Given a distance d_{ij} such as the Bray–Curtis distance (7), Proposition 1 reveals that the choice between an MRPP with $\Delta_{ij} = d_{ij}$ (and appropriate weights) and a pseudo-F test is not really a choice between the two methods—since the former is the same as a pseudo-F with distance $\sqrt{d_{ij}}$, while the latter is tantamount to an MRPP with squared dissimilarity d_{ij}^2. Rather, the question reduces to which distance to use for the pseudo-F (d_{ij} or $\sqrt{d_{ij}}$) or equivalently, which dissimilarity for the MRPP (d_{ij}^2 or d_{ij}). The above-cited congruence concept (Mielke, 1986) would seem to favor the second of either pair of choices.

6. Application to Randomized Block Design

Recall the notation $\sim_b$ of Section 3.1, and similarly write $i \sim_g j$ if observations i, j belong to the same group. For the randomized block design with $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ representing block and group effects, respectively, the generic elements of $\boldsymbol{H}_1$, $\boldsymbol{H}_2$, and $\boldsymbol{H}$ are

$$h_{ij}^{(1)} = \frac{I(i \sim_b j)}{g} - \frac{1}{bg}, \tag{14}$$

$$h_{ij}^{(2)} = \frac{I(i \sim_g j)}{b} - \frac{1}{bg}, \text{ and} \tag{15}$$

$$h_{ij} = \frac{I(i \sim_b j)}{g} + \frac{I(i \sim_g j)}{b} - \frac{1}{bg}, \tag{16}$$

respectively. These formulas enable us to prove the following result.

PROPOSITION 3: *If the squared distance function d_{ij}^2 is a metric, then the pseudo-F test is equivalent to the MRBP test based on squared distances, i.e., with $\Delta_{ij} = d_{ij}^2$.*

If the squared distance is a metric then so is the raw distance, but not conversely. Web Appendix C presents an example of a metric distance function whose square is not a metric and for which the two tests are not equivalent. Proposition 3 could, perhaps, be formulated more naturally by expressing the distance in terms of the dissimilarity, rather than vice versa. The result could then be restated as: an MRBP test for which the dissimilarity Δ_{ij} is a metric equivalent to the pseudo-F test with distance $d_{ij} = \sqrt{\Delta_{ij}}$.

7. Real Data Illustrations

In this section, we first revisit a previously studied ecological data set in light of our equivalence results. We then return to our motivating application in neuroscience, and show how distance-based permutation tests offer a novel inferential approach.

7.1 *Ecological Data*

For the one-way design, Proposition 2 asserts that the pseudo-F test with distance d_{ij} and the MRPP with dissimilarity Δ_{ij} are equivalent, when $\Delta_{ij} = d_{ij}^2$ and the MRPP employs weights (2). We examine here the effect of removing either of these conditions, using the Dutch dune meadow vegetation data set from Jongman, ter Braak, and van Tongeren (1995) included in the `vegan` package. The data consist of cover class values for 30 species, along with several environmental variables such as moisture and land use, for each of 20 sites. We assessed whether the Bray–Curtis distances (7) among the meadows betray a significant effect of the amount of manure applied. This is actually an ordinal variable, with levels from 0 to 4, but the tests treat it as categorical. The nonmetric multidimensional scaling (NMDS; Kruskal, 1964) plot in Figure 2 indicates a marked difference between the meadows with no manure applied and those at levels 1–4.

Our analyses were performed in `R` version 2.7.0 (`R` Development Core Team, 2008) with code based on the `vegan` (version 1.15–0) functions `mrpp` and `adonis`, but modified to use the same set of 9999 permutations for the two types of tests. We confirmed that the MRPP with weights (2) and squared Bray–Curtis measure produced exactly the same significance level as the pseudo-F test: $p = 0.0141$. Changing the power to which (7) was raised had little effect: for instance, the MRPP with raw Bray–Curtis measure yielded $p = 0.0147$. However, changing the weights to (3) raised the p-value to 0.0763. A simple explanation for this loss of significance is that (3) places a higher proportion of the weight on dissimilarities within large groups than does (2). Thus, the rather high dissimilarities among observations in the largest group, level 0, render δ not as small relative to the permutation distribution.

Changing the MRPP weights has not had a dramatic effect in most examples we have studied. But in this case, weights (2) produced results seemingly more in line with the NMDS plot than did weights (3). As noted in Section 3.1, Mielke and Berry (2007) favor weights (2) over (3) for asymptotic efficiency reasons. To this argument one might add that, by Proposition 2, the former choice brings MRPP into line with the pseudo-F approach.

7.2 *Functional Connectivity Data*

We now return to the functional connectivity data introduced in Section 2. For comparison with the distance-based tests of this article, we applied a mass-univariate permutation procedure similar to that used by Church et al. (2009) for data of the same type: for each of 499 permutations of the age-group labels, we calculated F statistics for the age effect at each of the 703 connections, and obtained the permutation distribution of the *maximum* of the 703 F-statistics, to which the real-data F-statistics were referred. Only one connection was found significant even at the 0.4 level: that between the subgenual anterior cingulate cortex and posterior cingulate cortex, for which we obtained $p = 0.038$. The observed mean correlations between these two regions were 0.05 for children, 0.25 for adolescents, and 0.40 for adults. These two regions are the core nodes of the default mode network, and this finding suggests that the strengthening of their connection may be an important component of brain maturation, in line with
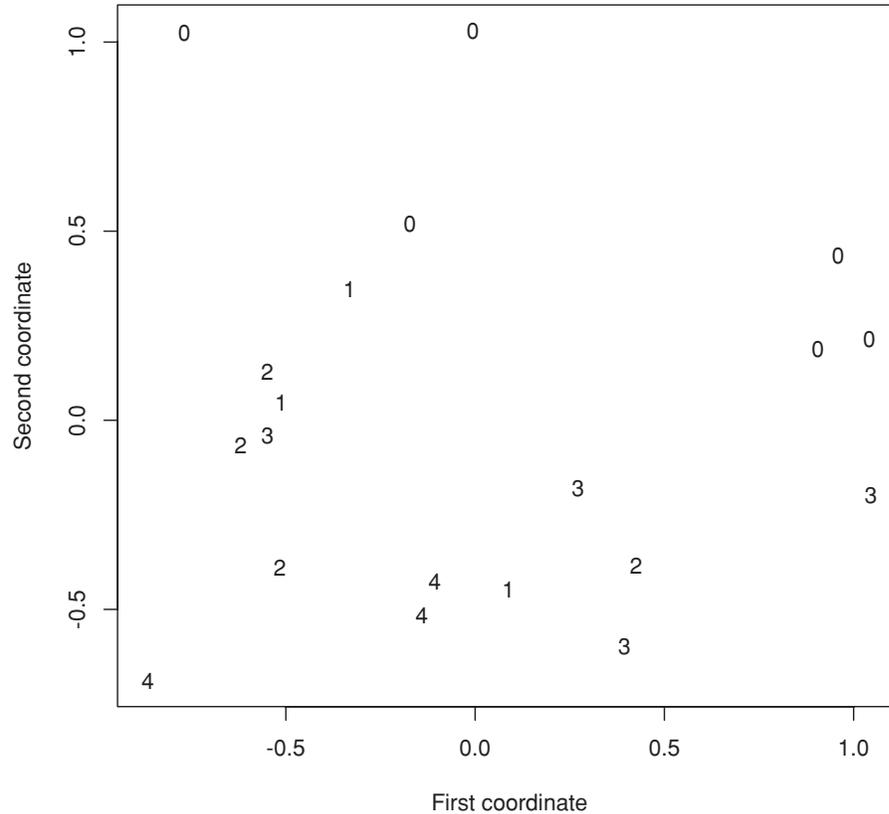
Figure 2. Nonmetric multidimensional scaling plot of the 20 Dutch meadows, indicating the level of manure applied on a scale from 0 to 4.

the findings of Kelly et al. (2009). Still, this test procedure has several limitations. The finding of a single significantly different connection offers no confirmation of the broader differentiation of task-positive and task-negative networks, from childhood to adulthood, as suggested by Figure 1. Indeed, had we not ordered the rows and columns in that figure as task-positive followed by task-negative based on a priori considerations, the mass-univariate finding would not have led us to suspect such a broad pattern of differential connectivity among the age groups. Furthermore, due to the need for separate tests at each connection, the relatively modest number of permutations required over 11 minutes on a MacBook Pro with a 2.16 GHz Intel Core Duo processor; applications to not-much-larger data sets could be excessively time-consuming.

In view of these limitations, the distance-based permutation tests implemented in the `vegan` package offer a useful complement to the mass-univariate analysis. We applied an MRPP to the three age groups, after applying the Fisher (1921) z transformation to the correlations, as is commonly done for functional connectivity data. The dissimilarity used was the square of the Frobenius metric (e.g., Bickel and Levina, 2008) between pairs of correlation matrices; this norm can be defined, for two matrices A and B, as the Euclidean distance between $\text{vec}(A)$ and $\text{vec}(B)$. It took just 3.8 seconds to run 9999 permutations using the same machine as above, resulting in a p-value of 0.0163; see Figure 3. (Without squaring the Frobenius norm, we obtained $p = 0.0212$.) An equivalent permutational MANOVA was slower (9.7 seconds),

but contributed the additional information that age group explains about 4.9% of variation, in the sense of (8). Moreover, the latter method can be extended to include other predictors. Adding gender to the model resulted in the analysis of distance table shown in Table 1. Gender appears not to have much effect, whereas, as we would expect, the p-value for age group agrees very closely with the MRPP result. Subsequent tests comparing each pair of age groups found a clear difference between children and adults ($p = 0.0041$) but no significant differences between the adolescents and the other two groups, supporting the intuition that adolescent functional connectivity patterns lie between those of children and adults.

Although tests of interactions lie beyond the scope of our main development, they can be performed within either the MRPP or the pseudo-F framework (Mielke and Berry, 2007; Legendre and Anderson, 1999). We carried out a pseudo-F test as described in Web Appendix A, and found no significant age group-by-gender interaction ($p = 0.215$, based on 9999 permutations).

8. Discussion

Mielke and Berry (2007) describe applications of MRPP in a wide array of disciplines from archeology to meteorology to psychometrics. Pseudo-F tests have a shorter history, with applications first in ecology, then in genetics, and more recently in computer graphics (Čadík et al., 2008). Thus, specialists in different fields may be conducting distance-based permutation tests under these two paradigms, with at best limited
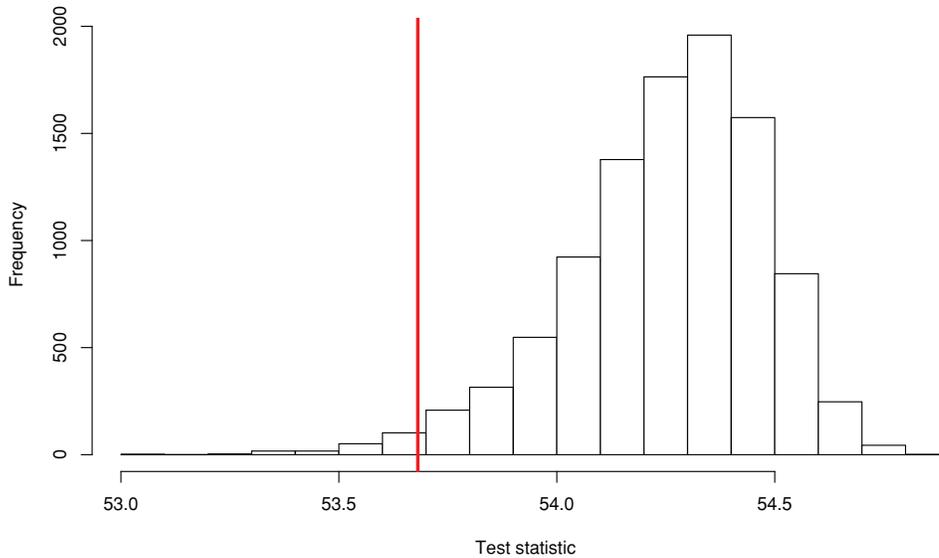
Figure 3. Permutation distribution of average within-group dissimilarity among the fMRI correlation matrices, with a vertical line indicating the observed value. This figure appears in color in the electronic version of this article.

Table 1
Analysis of distance table for the functional connectivity data

	df	SS	MS	F	R^2	$\Pr(>F)$
Age group	2	68.3076	34.1538	1.2735	0.0494	0.0164
Gender	1	27.8779	27.8779	1.0395	0.0202	0.3309
Residuals	48	1287.3076	26.8189		0.9305	
Total	51	1383.4932			1	

awareness of the close relationship between them. Such a situation contributes to a lack of understanding across disciplines. It is our hope that the equivalence results presented here will help to reduce this mutual incomprehension. In addition, we hope that our neuroimaging application may inspire a few researchers to consider distance-based permutation tests as a tool for meeting the challenges posed, with increasing frequency, by high-dimensional data analyses.

9. Supplementary Materials

Web Appendix A, referenced in Sections 3.2 and 7.2; Web Appendix B, referenced in Sections 4 and 5; and Web Appendix C, referenced in Section 6, are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

References

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26,** 32–46.

Anderson, M. J. (2005). PERMANOVA: A FORTRAN computer program for permutational multivariate analysis of variance. Department of Statistics, University of Auckland, New Zealand.

Anderson, M. J. and ter Braak, C. J. F. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation* **73,** 85–113.

Anderson, M. J., Gorley, R. N., and Clarke, K. R. (2008). *PERMANOVA + for PRIMER: Guide to Software and Statistical Methods*. Plymouth, U.K.: PRIMER-E Ltd.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *Annals of Statistics* **36,** 2577–2604.

Bray, J. R. and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* **27,** 326–349.

Čadík, M., Wimmer, M., Neumann, L., and Artusi, A. (2008). Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers and Graphics* **32,** 330–349.

Church, J. A., Fair, D. A., Dosenbach, N. U. F., Cohen, A. L., Miezin, F. M., Petersen, S. E., and Schlaggar, B. L. (2009). Control networks in paediatric Tourette syndrome show immature and anomalous patterns of functional connectivity. *Brain* **132,** 225–238.

Clarke, K. R. and Gorley, R. N. (2006). *Primer v6: User Manual/Tutorial*. Plymouth, U.K.: PRIMER-E Ltd.

Fair, D. A., Cohen, A. L., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., Barch, D. M., Raichle, M. E., Petersen, S. E., and Schlaggar, B. L. (2008). The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences* **105,** 4028–4032.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron* **1,** 3–32.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic anticorrelated functional networks. *Proceedings of the National Academy of Sciences* **102,** 9673–9678.

Good, I. J. (1982). An index of separateness of clusters and a permutation test for its significance. *Journal of Statistical Computation and Simulation* **15,** 81–84.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53,** 325–338.

Gower, J. C. and Krzanowski, W. J. (1999). Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. *Applied Statistics* **48,** 505–519.

Jongman, R. H. G, ter Braak, C. J. F., and van Tongeren, O. F. R., eds (1995). *Data Analysis in Community and Landscape Ecology*, new edition. Cambridge, U.K.: Cambridge University Press.

Kelly, A. M. C., Di Martino, A., Uddin, L. Q., Shehzad, Z., Gee, D. G., Reiss, P. T., Margulies, D. S., Castellanos, F. X., and Milham, M. P. (2009). Development of anterior cingulate functional connectivity from late childhood to early adulthood. *Cerebral Cortex* **19,** 640–657.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29,** 115–129.

Krzanowski, W. J. (2002). Multifactorial analysis of distance in studies of ecological community structure. *Journal of Agricultural, Biological, and Environmental Statistics* **7,** 222–232.

Krzanowski, W. J. (2006). Sensitivity in metric scaling and analysis of distance. *Biometrics* **62,** 239–244.

Legendre, P. and Anderson, M. J. (1999). Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* **69,** 1–24.

Legendre, P. and Legendre, L. (1998). *Numerical Ecology*, 2nd English edition. Amsterdam: Elsevier.

Mantel, N. and Valand, R. S. (1970). A technique of nonparametric multivariate analysis. *Biometrics* **26,** 547–558.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

McArdle, B. H. and Anderson, M. J. (2001). Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* **82,** 290–297.

Mielke, P. W. (1986). Non-metric statistical analyses: Some metric alternatives. *Journal of Statistical Planning and Inference* **13,** 377–387.

Mielke, P. W. and Berry, K. J. (2007). *Permutation Methods: A Distance Function Approach*, 2nd edition. New York: Springer.

Mielke, P. W. and Iyer, H. K. (1982). Permutation techniques for analyzing multi-response data from randomized block experiments. *Communications in Statistics: Theory and Methods* **11,** 1427–1437.

Mielke, P. W., Berry, K. J., and Johnson, E. S. (1976). Multi-response permutation procedures for a priori classifications. *Communications in Statistics: Theory and Methods* **5,** 1409–1424.

Nievergelt, C. M., Libiger, O., and Schork, N. J. (2007). Generalized analysis of molecular variance. *PLoS Genetics* **3**(4), e51.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2008). *vegan*: Community ecology package. R package version 1.15–0. `http://cran.r-project.org/`, `http://vegan.r-forge.r-project.org/`.

`R` Development Core Team (2008). `R`: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org`.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology* **4**(1), article 32.

Smith, E. P., Pontasch, K. W., and Cairns, J. (1990). Community similarity and the analysis of multispecies environmental data: A unified statistical approach. *Water Research* **24,** 507–514.

Toro, R., Fox, P. T., and Paus, T. (2008). Functional coactivation map of the human brain. *Cerebral Cortex* **18,** 2553–2559.

Wessel, J. and Schork, N. J. (2006). Generalized genomic distance-based regression methodology for multilocus association analysis. *American Journal of Human Genetics* **79,** 792–806.

Zapala, M. A. and Schork, N. J. (2006). Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences* **103,** 19430–19435.