

## HYBRID PERMUTATION TEST WITH APPLICATION TO SURFACE SHAPE ANALYSIS

Chunxiao Zhou and Yongmei Michelle Wang

*University of Illinois at Urbana-Champaign*

*Abstract:* This paper presents a new statistical surface analysis framework that aims to accurately and efficiently localize regionally specific shape changes between groups of 3D surfaces. With unknown distribution and small sample size of the data, existing shape morphometry analysis involves testing thousands of hypotheses for statistically significant effects through permutation. In this work, we develop a novel hybrid permutation test approach to improve the system's efficiency by approximating the permutation distribution of the test statistic with a Pearson distribution series that involves the calculation of the first four moments of the permutation distribution. We propose to derive these moments theoretically and analytically without any permutation. Detailed derivations and experimental results using two different test statistics are demonstrated using simulated data and brain data for shape morphometry analysis. Furthermore, an adaptive procedure is utilized to control the False Discovery Rate (FDR) for increased power of finding significance.

*Key words and phrases:* FDR, MRI, Pearson distribution, permutation test, ROI, shape analysis, surface morphometry.

### 1. Introduction

Permutation tests are among the most powerful nonparametric tests that can be applied when parametric tests do not work. For example, when the data distribution is unknown and sample size is not large enough, permutation tests can obtain *p-values* from permutation distributions of a test statistic, rather than from parametric distributions. In addition, permutation tests require few assumptions concerning statistical distributions but exchangeability. They belong to the nonparametric “distribution-free” category of hypothesis testing and have been used successfully in both structural MR image analysis by Nichols and Holmes (2001), and functional MR image analysis by Pantazis, Leahy, Nichols and Styner (2004). There are three major approaches to constructing the permutation distribution (Hubert (1987), Mielke and Berry (1994), Pesarin (2001) and Good (2005)). First, exact permutation enumerates all possible arrangements. The second approach is an approximate permutation distribution based on random sampling from all possible permutations. Third, permutation distribution

approximation uses the analytical moments of the exact permutation distribution under the null hypothesis. The computational cost is the main disadvantage of the exact permutation, due to the factorial increase in the number of permutations with the increasing number of subjects. The second technique has the problem of replication, and causes more type I errors. When a large number of repeated tests are needed, the random permutation strategy is also computationally expensive to achieve satisfactory *p-value* accuracy. Finally, the moments of the exact permutation distribution may not exist. In addition, if they do, it may be difficult to obtain them.

In this paper, we propose a hybrid strategy to take advantage of nonparametric permutation tests and a parametric Pearson distribution approximation for both accuracy/flexibility and efficiency. Specifically, we employ a general theoretical method to derive the moments of the permutation distribution for any linear test statistic on multivariate data. The key idea is to separate the moments of the permutation distribution into two parts: a permutation of test statistic coefficients and functions of the data. We can then obtain the moments without any permutations since the permutation of test statistic coefficients can be derived theoretically. Regarding nonlinear test statistics, one strategy is to use a monotonic linear test statistic to replace the original nonlinear one. Since only the order of test statistics of all permutations matters for *p-values*, the monotonic linear test statistic shares the same *p-value* with the original nonlinear test statistic. Given the first four moments, the permutation distribution can be well-fitted by Pearson distribution series. The *p-values* are then estimated without any real permutation. For multiple comparison of a two-group differences, given the sample sizes  $n_1 = 21$  and  $n_2 = 21$ , with the number of tests  $M = 2,000$ , we need to conduct  $M \times (n_1 + n_2)!/n_1!/n_2! \approx 1.1 \times 10^{15}$  permutations for the exact permutation test. Even for 20,000 random permutations per test, we still need  $M \times 20,000 \approx 4 \times 10^7$  permutations. Alternatively, our hybrid permutation method using a Pearson distribution approximation only involves the calculation of analytically derived first four moments of the exact permutation distributions while achieving good robustness and high accuracy in addition to its efficiency (see Section 7). Instead of the calculating test statistics in factorial scale with exact permutations, our hybrid method only requires  $O(n)$  or  $O(n^2)$  computational cost depending on the choices of test statistics, where  $n = n_1 + n_2$ .

Accurate and efficient brain morphometry analysis is of great importance in detecting morphological changes in structures of interest for neuroscience research, medical diagnosis, and treatment. Different techniques have been studied in Chung, Worsley, Paus, Cherif, Giedd, Rapoport and Evans (2001) and Shen and Davatzikos (2003). The aim is to identify and characterize localized structural differences of the human brain within and across groups of healthy individuals and/or persons with various diseases (Ashburner and Friston (2004)). In

neuroimaging research, there is increasing evidence that surface shape analysis of brain structures provides new information that is not available with conventional analysis Thompson, Hayashi, De, Zubicaray, Janke, Rose, Semple, Hong, Herman, Gravano, Doddrell and Toga (2004) and Styner, Lieberman, McClure, Weingberger, Jones and Gerig (2005)). A critical issue in surface morphometry is the shape description and representation. Various strategies have been investigated recently in the literature, such as Brechbuhler, Gerig and Kubler (1995), Dryden and Mardia (1998), Wang and Staib (2000), Thompson, Hayashi, De, Zubicaray, Janke, Rose, Semple, Hong, Herman, Gravano, Doddrell and Toga (2004), Yushkevich, Zhang and Gee (2006) and Zhou, Park, Styner and Wang (2007). The SPHARM approach using spherical harmonics as basis functions for a parametric surface description was proposed in Brechbuhler, Gerig and Kubler (1995). The correspondence across different surfaces is established by aligning the parameterizations via the first order ellipsoid. The present work employs the SPHARM-PDM shape description (Styner, Oguz, Xu, Brechbuehler, Pantazis, Levitt, Shenton and Gerig (2006)), which leads to corresponding location vectors across all surfaces for our subsequent statistical analysis of surface shape.

In this paper, we focus on the surface shape analysis for two groups, though our method can be extended to the multi-group case. As shown in Figure 1, at each corresponding position on the surfaces, we test whether there is significant mean vector difference between location vectors of two groups. If a hypothesis test leads to a p-value lower than the pre-chosen  $\alpha$ -level, we reject the null hypothesis and conclude that a significant shape difference exists at this surface location.

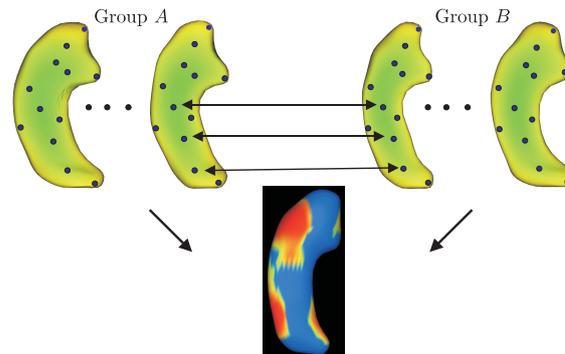


Figure 1. Top: two groups of hippocampus surfaces with corresponding location vectors; Bottom: statistical analysis of surface shape (hot color coded: red denotes significantly different; blue denotes no difference).

The distribution of the location vectors is unknown, there is a limited number of subject samples, and the same tests are repeated on thousands of locations.

Yet without any permutation, we can accurately estimate the  $p$ -value of the permutation distribution of the test statistic at each surface locations using our hybrid approach to surface shape analysis.

The analysis of the location vectors involves testing thousands of hypotheses. False positives must be controlled over all tests. Common techniques to handle multiple testing problems in imaging include cluster-size thresholding, Random Field Theory (Taylor and Worsley (2007)), controlling the Family-Wise Error Rate (FWER), and the False Discovery Rate (FDR) (Benjamini and Hochberg (1995)). Compared with the FDR, FWER control is too conservative for many applications. FDR control is becoming more popular due to its power of finding real discoveries through the control of the fraction of false discoveries over total discoveries (Benjamini and Hochberg (1995), Storey (2002) and Benjamini, Krieger and Yekutieli (2006)). The FDR method was initially proposed by Benjamini and Hochberg (1995), so called BH's FDR. In this work, we develop a Region of Interest (ROI) constrained adaptive FDR to enhance the power of finding true discoveries.

## 2. Shape Descriptor and Surface Alignment

We use the SPHARM-PDM software developed by Styner, Oguz, Xu, Brechbuehler, Pantazis, Levitt, Shenton, and Gerig (2006) to establish surface correspondence and align the surface location vectors across all subjects. The sampled SPHARM-PDM is a smooth, accurate, fine-scale shape representation. The spherical parameterization is computed via optimizing an equal area mapping of the 3D voxel mesh onto the sphere and minimizing angular distortions. The basis functions of the parameterized surface are spherical harmonics. Each individual SPHARM description is composed of a set of coefficients, weighting the basis functions. Truncating the spherical harmonic series at different degrees would result in object representations at different levels of detail. The surfaces of different subjects also get aligned via the first order ellipsoid from the spherical harmonic coefficients. The surface shapes of different objects are thus represented by the same number of location vectors (with each location vector consisting of the spatial  $x$ ,  $y$ , and  $z$  coordinates of the corresponding vertex on the SPHARM-PDM surface) for our subsequent statistical analysis of surface shape.

## 3. Hypothesis and Test Statistics

Given registered location vectors across all subjects, surface shape morphometry analysis becomes a two-sample test for equality of means at each surface location. The hypothesis is typically constructed as:

$$H_0 : \mu_A = \mu_B \quad \text{vs.} \quad H_a : \mu_A \neq \mu_B, \quad (3.1)$$

where  $\underline{\mu}_A = [\mu_A^{(x)} \ \mu_A^{(y)} \ \mu_A^{(z)}]^T$  and  $\underline{\mu}_B = [\mu_B^{(x)} \ \mu_B^{(y)} \ \mu_B^{(z)}]^T$  are the three dimensional mean vectors of group  $A$  and group  $B$ .

**3.1. Univariate mean difference test statistics**

To test the equality of two sample mean vectors, we can choose univariate group mean differences  $\bar{x}_A - \bar{x}_B, \bar{y}_A - \bar{y}_B, \bar{z}_A - \bar{z}_B$  as test statistics for three partial tests, where  $\bar{x}_A, \bar{x}_B, \bar{y}_A, \bar{y}_B, \bar{z}_A, \bar{z}_B$  are the univariate sample means of group  $A$  and group  $B$  in  $x, y$  and  $z$  components, respectively. Then we determine the existence of shape difference at a location if any of the three partial tests gives a p-value less than the pre-chosen significance level.

**3.2. Modified Hotelling’s  $T^2$  test statistics**

Hotelling’s  $T^2$  is an optimal multivariate test statistic for mean vector difference testing under multivariate normality (Johnson and Wichern (2002) and Liao (2002)). It is defined as

$$T = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_A - \bar{X}_B)^T S_{pooled}^{-1} (\bar{X}_A - \bar{X}_B),$$

where the pooled variance-covariance matrix  $S_{pooled} = [(n_1)S_A + (n_2 - 1)S_B] / (n_1 + n_2 - 2)$  is a combination of sample variance-covariance matrices of group  $A$  and of group  $B$ , and  $\bar{X}_A$  and  $\bar{X}_B$  are mean vectors of group  $A$  and group  $B$ . However, it is impossible to compute the pooled covariance matrix without real permutations. We resolve this by replacing the pooled covariance matrix  $S_{pooled}$  in Hotelling’s  $T^2$  with the sample variance-covariance matrix  $S$  over all the subjects of the two groups. This leads to a modified  $T^2$  as

$$T^* = (\bar{X}_A - \bar{X}_B)^T S^{-1} (\bar{X}_A - \bar{X}_B).$$

Since it has been proved in Wald and Wolfowitz (1944) that the modified  $T^2$  is a monotonic function of Hotelling’s  $T^2$ , the two statistics share the same p-value in permutation tests.

**4. Hybrid Permutation**

**4.1 Pearson distribution series**

The Pearson distribution series (Pearson I ~ VII) is a family of probability distributions that are more general than the normal distribution (Hubert (1987)). As indicated in Figure 2 (Hahn and Shapiro (1967)), it covers all distributions in the  $(\beta_1, \beta_2)$  plane and includes the normal, beta, gamma, log-normal, etc., where distribution shape parameters  $\beta_1, \beta_2$  are, respectively, the square of the standardized measure of skewness and of peakness. Given the first four moments,

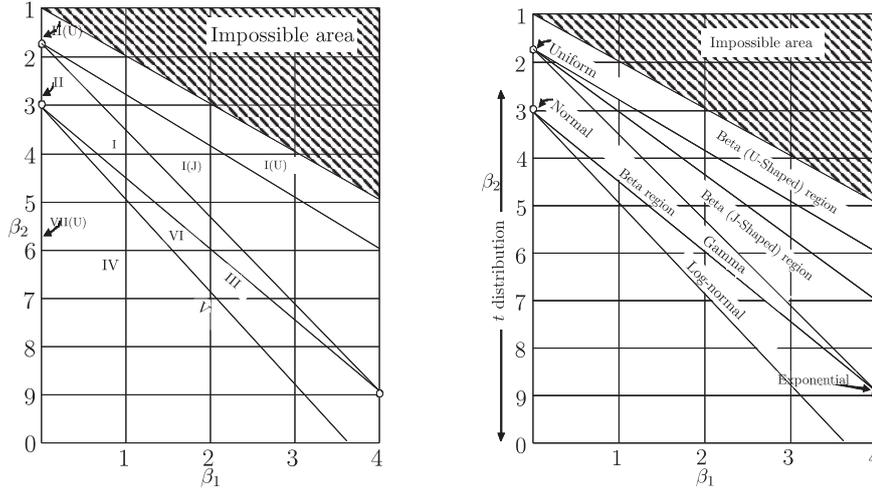


Figure 2. Left: Regions in  $(\beta_1, \beta_2)$  plane for various types of Pearson distribution series; Right: Regions in  $(\beta_1, \beta_2)$  plane for various types of parametric distributions (Hahn and Shapiro (1967)).

the Pearson distribution series can be utilized to approximate the permutation distribution of the test statistic.

**4.2. Theoretical general derivation of moments**

In order to approximate the permutation distribution with a Pearson distribution, the moments of the exact permutation distribution need to be computed. In this section, we first describe our general method regarding this derivation for linear test statistics on one-dimensional and on multivariate data. Then we show our specific mathematical formulations for mean difference and Hotelling’s  $T^2$  test statistics used in the surface morphometry analysis.

Let  $X = [x_1 \cdots x_n]^T$  be the one-dimensional data, and  $T = C^T P X$  denote the linear test statistic for the permutation test with  $C = [c_1 \cdots c_n]^T$  the linear test statistic vector. The permutation matrix  $P$  is a matrix that has one entry of 1 in each row and each column, and 0’s elsewhere; it permutes the rows of an identity matrix according to some permutation of the numbers 1 to  $n$ .

For example, if  $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ ,  $X = [x_1 \ x_2 \ x_3]^T$  and  $C = [c_1 \ c_2 \ c_3]^T$ , then

$PX = [x_3 \ x_1 \ x_2]^T$  and  $T = C^T P X = c_2 x_1 + c_3 x_2 + c_1 x_3 = \sum_i c_{\pi(i)} x_i$ , where  $(\pi(1), \pi(2), \pi(3)) = (2, 3, 1)$  denotes a permutation of vector data by row. Then the linear test statistic can be denoted as  $T = C^T P X = \sum_i c_{\pi(i)} x_i$ . The  $r$ th

moment of  $T$  is

$$\begin{aligned}
 E(T^r(X, \pi)|X) &= E\left((C^T P X)^r\right) = \frac{1}{n!} \sum_{\pi} \left(\sum_i c_{\pi(i)} x_i\right)^r = \frac{1}{n!} \sum_{\pi} \sum_{i_1, \dots, i_r} \left(\prod_{k=1}^r c_{\pi(i_k)} x_{i_k}\right) \\
 &= \frac{1}{n!} \sum_{\pi} \sum_{i_1, \dots, i_r} \left(\prod_{k=1}^r x_{i_k} \prod_{k=1}^r c_{\pi(i_k)}\right) = \sum_{i_1, \dots, i_r} \left(\prod_{k=1}^r x_{i_k} \frac{1}{n!} \sum_{\pi} \prod_{k=1}^r c_{\pi(i_k)}\right). \tag{4.1}
 \end{aligned}$$

To compute  $(1/n!) \sum_{\pi} \prod_{k=1}^r c_{\pi(i_k)}$ , partition the index space  $U = \{1 \cdots n\}^r$  into  $\bigcup_{(\lambda_1, \dots, \lambda_q) \in L} U^{(\lambda_1, \dots, \lambda_q)}$ , where  $L = \{(\lambda_1, \dots, \lambda_q) : \lambda_1, \dots, \lambda_q \in \mathbb{Z}^+; \lambda_1 \leq \dots \leq \lambda_q; \lambda_1 + \dots + \lambda_q = r\}$ .  $U^{(\lambda_1, \dots, \lambda_q)}$  means that all  $r$  indices are permuted into  $q$  different numbers, and each number corresponds to  $\lambda_i$  indices. When  $r = 3$ ,  $U = U^{(1,1,1)} \cup U^{(1,2)} \cup U^{(3)}$ , where  $U^{(1,1,1)}$  is the set of  $\{i_1 \neq i_2 \text{ and } i_1 \neq i_3 \text{ and } i_2 \neq i_3\}$  with  $q = 3$  and  $\lambda_1 = \lambda_2 = \lambda_3 = 1$ ,  $U^{(1,2)}$  is the set of  $\{i_1 = i_2 \neq i_3 \text{ or } i_1 = i_3 \neq i_2 \text{ or } i_2 = i_3 \neq i_1\}$  with  $q = 2$  and  $\lambda_1 = 1, \lambda_2 = 2$ , and  $U^{(3)}$  is the set of  $\{i_1 = i_2 = i_3\}$  with  $q = 1$  and  $\lambda_1 = 3$ . Since permutation is equally related to all  $r$  indices,  $(1/n!) \sum_{\pi} \prod_{k=1}^r c_{\pi(i_k)}$  is invariant in each category, i.e.,  $(1/n!) \sum_{\pi} \prod_{k=1}^r c_{\pi(i_k)} = a_{(\lambda_1, \dots, \lambda_q)}$  if  $(i_1 \cdots i_r) \in U^{(\lambda_1, \dots, \lambda_q)}$ . Eventually, the  $r$ th moment is:

$$E(T^r(X, \pi)|X) = \sum_{\lambda_1, \dots, \lambda_q} a_{(\lambda_1, \dots, \lambda_q)} \sum_{(i_1, \dots, i_r) \in U^{(\lambda_1, \dots, \lambda_q)}} \left(\prod_{k=1}^r x_{i_k}\right). \tag{4.2}$$

Note that (4.2) separates the permutation from the data. To get the moments, we only need to derive the permutation of the coefficients of pre-chosen test statistics and calculate the summation terms of data. Due to the simple pattern of the coefficients of test statistics, which is the same for repeated tests, we can derive the moments of the permutation distribution without permuting the data. Note that the computational cost of data summation terms for the  $r$ th moment in each index subspace can be reduced to  $O(n)$ , not  $O(n^r)$ .

The discussed approach can be easily extended to the multivariate case. Consider the linear test statistic  $T = \sum_{i,j} e_{\pi(i),j} x_{i,j}$ , where  $X$  is a  $n \times m$  matrix with  $n$  observations on  $m$  variables,  $E$  is a  $n \times m$  linear test statistic matrix, and  $\pi(i)$  is a row permutation that permutes the row index  $i$  from 1 to  $n$ . The moments for multivariate data are then derived similarly as in Eq. (4.1).

$$E(T^r(X, \pi)|X) = \frac{1}{n!} \sum_{\pi} \left(\sum_{i,j} e_{\pi(i),j} x_{i,j}\right)^r = \frac{1}{n!} \sum_{\pi} \sum_{\substack{i_1, \dots, i_r \\ j_1, \dots, j_r}} \left(\prod_{k=1}^r e_{\pi(i_k),j_k} x_{i_k,j_k}\right)$$

$$\begin{aligned}
 &= \frac{1}{n!} \sum_{\pi} \sum_{\substack{i_1, \dots, i_r \\ j_1, \dots, j_r}} \left( \prod_{k=1}^r x_{i_k, j_k} \prod_{k=1}^r e_{\pi(i_k, j_k)} \right) \\
 &= \sum_{\substack{i_1, \dots, i_r \\ j_1, \dots, j_r}} \left( \prod_{k=1}^r x_{i_k, j_k} \frac{1}{n!} \sum_{\pi} \prod_{k=1}^r e_{\pi(i_k, j_k)} \right). \tag{4.3}
 \end{aligned}$$

Similarly, we can partition the row index space  $U = \{1, \dots, n\}^r$  into  $\bigcup_{(\lambda_1, \dots, \lambda_q) \in L} U^{(\lambda_1, \dots, \lambda_q)}$ . Thus the  $r$ th moment for multivariate case is:

$$E\left(T^r(X, \pi) | X\right) = \sum_{j_1, \dots, j_r} \sum_{\lambda_1, \dots, \lambda_q} a_{(\lambda_1, \dots, \lambda_q), j_1, \dots, j_r} \sum_{(i_1, \dots, i_r) \in U^{(\lambda_1, \dots, \lambda_q)}} \left( \prod_{k=1}^r x_{i_k, j_k} \right). \tag{4.4}$$

Alternatively, all  $a$ 's can be calculated by computer simulation without analytical derivation. Note that the computational cost of data summation terms for the  $r$ th moment in each index subspace can be reduced to  $O(m^r n)$ , not  $O(m^r n^r)$ .

### 4.3. Derivation of moments for specific test statistics

In surface morphometry analysis, suppose we have two groups of location vectors for each corresponding surface position, and that the numbers of subjects in the two groups are  $n_1$  and  $n_2$ . For each location on the surface, let  $X = [x_1 \cdots x_{n_1} \ x_{n_1+1} \cdots x_{n_1+n_2}]^T$ ,  $Y = [y_1 \cdots y_{n_1} \ y_{n_1+1} \cdots y_{n_1+n_2}]^T$  and  $Z = [z_1 \cdots z_{n_1} \ z_{n_1+1} \cdots z_{n_1+n_2}]^T$  denote the  $x, y, z$  components of the location vectors across all subjects. To analyze the surface shape difference between two groups, we test the null hypothesis (3.1) at each location one by one and independently, and then detect the locations with significant shape changes. Note that mean vector difference is used as the test statistic. For the hypothesis (3.1),  $T = [T_X \ T_Y \ T_Z]^T$ , where  $T_X = \sum_{i=1}^{n_1} x_i/n_1 - \sum_{i=n_1+1}^{n_1+n_2} x_i/n_2 = CPX$ ,  $T_Y = CPY$ , and  $T_Z = CPZ$ . The mean difference vector  $C = [(1/n_1)1_{1 \times n_1}, (-1/n_2)1_{1 \times n_2}]$ . Since  $T_X, T_Y$  and  $T_Z$  have the same  $C$ , their moments coefficients  $a$ 's are the same. Following the discussed method, we demonstrate below how to derive the moments coefficients for the first four moments of  $T_X$ . The derivation of moments coefficients for  $T_Y$  and  $T_Z$  can be calculated in the same way.

First moment,  $U = U^{(1)}$ :

$$a_{(1)} = \frac{1}{(n_1 + n_2)!} \sum_{\pi} c_{\pi(i_1)} = \frac{\binom{n_1}{1} \frac{1}{n_1} - \binom{n_2}{1} \frac{1}{n_2}}{\binom{n_1 + n_2}{1}} = 0, \quad E(T_X(X, \pi) | X) = 0.$$

Second moment,  $U = U^{(1,1)} \cup U^{(2)}$ :

$$\begin{aligned}
 a_{(1,1)} &= \frac{1}{(n_1 + n_2)!} \sum_{\substack{\pi \\ i_1 \neq i_2}} c_{\pi(i_1)} c_{\pi(i_2)} = \frac{\binom{n_1}{2} \frac{1}{n_1^2} + \binom{n_2}{2} \frac{1}{n_2^2} - \binom{n_1}{1} \binom{n_2}{1} \frac{1}{n_1 n_2}}{\binom{n_1 + n_2}{2}} \\
 &= -\frac{1}{n_1 n_2 (n_1 + n_2 - 1)}, \\
 a_{(2)} &= \frac{1}{(n_1 + n_2)!} \sum_{\substack{\pi \\ i_1 \neq i_2}} c_{\pi(i_1)} c_{\pi(i_2)} = \frac{1}{(n_1 + n_2)!} \sum_{\pi} c_{\pi(i)}^2 \\
 &= \frac{\binom{n_1}{1} \frac{1}{n_1^2} + \binom{n_2}{1} \frac{1}{n_2^2}}{\binom{n_1 + n_2}{2}} = \frac{2}{n_1 n_2 (n_1 + n_2 - 1)},
 \end{aligned}$$

$$\begin{aligned}
 E(T_X^2(X, \pi) | X) &= a_{(1,1)} \sum_{(i_1, i_2) \in U^{(1,1)}} x_{i_1} x_{i_2} + a_{(2)} \sum_{(i_1, i_2) \in U^{(2)}} x_{i_1} x_{i_2} \\
 &= a_{(1,1)} \sum_{i_1 \neq i_2} x_{i_1} x_{i_2} + a_{(2)} \sum_{i_1} x_{i_1}^2.
 \end{aligned}$$

The third and fourth moments are given in the online supplement (<http://www3.stat.sinica.edu.tw/statistica>). Note that the third moment is zero in the balanced design ( $n_1 = n_2$ ).

For the modified Hotelling's  $T^2$  test statistic, let

$$D = \begin{bmatrix} x_1 \cdots x_{n_1} & x_{n_1+1} \cdots x_{n_1+n_2} \\ y_1 \cdots y_{n_1} & y_{n_1+1} \cdots y_{n_1+n_2} \\ z_1 \cdots z_{n_1} & z_{n_1+1} \cdots z_{n_1+n_2} \end{bmatrix}^T$$

denote the location vector matrix over all subjects in both groups, then  $T^* = E((C^T P V) S^{-1} (C^T P V)^T) = E((C^T P X) (C^T P X)^T)$ , where  $V = D S^{-1/2}$  is used for simplicity. For the  $r$ th moment,

$$E(T^{*r}) = E(((C^T P X) (C^T P X)^T)^r) = E(((C^T P V V^T P^T C)^r) = E(((C^T P W P^T C)^r).$$

Since  $W = V V^T$  is a square matrix, there is no difference between the permutation of row and column indices, and

$$E(T^{*r}(W, \pi) | W) = \sum_{\lambda_1, \dots, \lambda_q} a_{(\lambda_1, \dots, \lambda_q)} \sum_{(i_1, \dots, i_r; i_{r+1}, \dots, i_{2r}) \in U^{(\lambda_1, \dots, \lambda_q)}} \left( \prod_{k=1}^r w_{i_k, i_{r+k}} \right).$$

The equation above also shows that the derivation of moments for the modified Hotelling's  $T^2$  is separable since here  $\pi(i, j) = (\pi(i), \pi(j))$ .

First moment,  $U = U^{(1,1)} \cup U^{(2)}$ :

$$a_{(1,1)} = -\frac{1}{n_1 n_2 (n_1 + n_2 - 1)}, \quad a_{(2)} = \frac{1}{n_1 n_2},$$

$$\begin{aligned} E(T^*(W, \pi) | W) &= a_{(1,1)} \sum_{(i_1, i_2) \in U^{(1,1)}} w_{i_1, i_2} + a_{(2)} \sum_{(i_1, i_2) \in U^{(2)}} w_{i_1, i_2} \\ &= a_{(1,1)} \sum_{i_1 \neq i_2} w_{i_1, i_2} + a_{(2)} \sum_{i_1} w_{i_1, i_1}. \end{aligned}$$

Second moment,  $U = U^{(1,1,1,1)} \cup U^{(1,1,2)} \cup U^{(1,3)} \cup U^{(2,2)} \cup U^{(4)}$ :

$$\begin{aligned} a_{(1,1,1,1)} &= \left[ \binom{n_1}{4} \frac{1}{n_1^4} - \binom{n_1}{3} \binom{n_2}{1} \frac{1}{n_1^3 n_2} + \binom{n_1}{2} \binom{n_2}{2} \frac{1}{n_1^2 n_2^2} \right. \\ &\quad \left. - \binom{n_1}{1} \binom{n_2}{3} \frac{1}{n_1 n_2^3} + \binom{n_2}{4} \frac{1}{n_2^4} \right] / \binom{n_1 + n_2}{4}, \\ a_{(1,1,2)} &= \left[ \binom{n_1}{3} \frac{1}{n_1^3} - \frac{2}{3} \binom{n_1}{2} \binom{n_2}{1} \frac{1}{n_1^2 n_2} + \frac{1}{3} \binom{n_1}{2} \binom{n_2}{1} \frac{1}{n_1^2 n_2^2} \right. \\ &\quad \left. + \frac{1}{3} \binom{n_1}{1} \binom{n_2}{2} \frac{1}{n_1 n_2^2} - \frac{2}{3} \binom{n_2}{1} \binom{n_2}{2} \frac{1}{n_1 n_2^3} + \binom{n_2}{3} \frac{1}{n_2^3} \right] / \binom{n_1 + n_2}{3}, \\ a_{(1,3)} &= \left[ \binom{n_1}{2} \frac{1}{n_1^2} - \frac{1}{2} \binom{n_1}{1} \binom{n_2}{1} \frac{1}{n_1 n_2} - \frac{1}{2} \binom{n_1}{1} \binom{n_2}{1} \frac{1}{n_1 n_2^3} \right. \\ &\quad \left. + \binom{n_2}{2} \frac{1}{n_2^2} \right] / \binom{n_1 + n_2}{2}, \\ a_{(2,2)} &= \left[ \binom{n_1}{2} \frac{1}{n_1^2} + \binom{n_1}{1} \binom{n_2}{1} \frac{1}{n_1 n_2} + \binom{n_2}{2} \frac{1}{n_2^2} \right] / \binom{n_1 + n_2}{2}, \\ a_{(4)} &= \left[ \binom{n_1}{1} \frac{1}{n_1} + \binom{n_2}{1} \frac{1}{n_2} \right] / \binom{n_1 + n_2}{1}. \end{aligned}$$

The derivations of the third and fourth moments are described in the online supplement (<http://www3.stat.sinica.edu.tw/statistica>).

## 5. Multiple Comparison via Adaptive ROI Constrained FDR

Determining whether a location on the brain surface has significant group shape difference or not corresponds to performing a hypothesis test of (3.1) at that position. Clearly, the location-wise p-values are spatially dependent. The significance rule, applied in the conventional False Discovery Rate (FDR) approach is defined as the expected proportion of false positives among the declared

significant results. It is more powerful and less stringent than the Family-Wise-Error-Rate (FWER) approach. In this work, we adopt the adaptive concept of the FDR (Benjamini, Krieger and Yekutieli (2006)) and develop a Region of Interest (ROI) constrained adaptive FDR (Zhou, Park, Styner and Wang (2007)), with detailed procedures as follows.

- Step 1: Given an observed statistic  $T^*$  (i.e., modified  $T^2$ ), the corresponding critical value at the pre-chosen significance level  $\alpha_{ROI}$  is calculated based on the constructed Pearson distribution. Then, the voxel is included in the ROI if its  $T^*$  value is larger than the critical value.
- Step 2: For  $\{p_i | location\ i \in ROI\}$ , order the set of observed p-values as  $p_{(1)} \leq \dots \leq p_{(m_r)}$ , where  $m_r$  is the number of locations within ROI.
- Step 3: Starting with  $p_{(m_r)}$  and in decreasing order, find the largest  $k$  for which  $p_{(k)} \leq k[q/((1+q)M)]$ , where  $q$  is the preset FDR error rate and  $m$  is the total number of location vectors on the surface.
- Step 4: Estimate the number of null locations by using  $\hat{m}_0 = (1+q)(M-k)$ .
- Step 5: Calculate the “BH-FDR-adjusted p-values” for locations within the ROI through  $p_{(i)}^{BH} = \min\{p_{(j)}\hat{m}_0/j | j \geq i\}$ .
- Step 6: Set  $p_i^{BH} = 1$ , if location  $i \notin ROI$ .
- Step 7: Compare all  $p_i^{BH}$  with desired level of FDR  $q$ ; all  $p_i^{BH} \leq q$  are declared significant.

This adaptive FDR control is more powerful than the conventional one. It can find more significant areas while preserving the same desired FDR rate. The ROI constrained adaptive scheme in this work is more efficient since only the p-values within the ROI need to be sorted. We set the p-values outside the ROI to 1; because it is unlikely to change the ranking of the p-values of true significances, therefore, this procedure won't change the significance detection result with the adaptive FDR strategy.

## 6. Experiments and Results

### 6.1. Simulated data for mean difference test statistics

To evaluate the accuracy and efficiency of our hybrid permutation tests, we consider six simulated cases in the first experiment for testing the difference between two groups  $A$  and  $B$ . For group  $A$ ,  $n_1$  observations are generated independently from Normal(0, 1) in Cases 1–2, from Gamma(3, 3) in Cases 3–4, and from Beta(0.8, 0.8) in Cases 5–6. For group  $B$ ,  $n_2$  independent observations are generated from Normal(1, 0.5) in Cases 1–2, from Gamma (3, 2) in Cases 3–4, and from Beta(0.1, 0.1) in Cases 5–6. The design is balanced in Cases 1,

Table 1. Comparison of computational costs and p-values of three permutation methods: hybrid permutation (HP), random permutation (RP), and exact permutation (EP). The  $t_{HP}$ ,  $t_{RP}$ , and  $t_{EP}$  denote the computation time (in seconds), and  $p_{HP}$ ,  $p_{RP}$  and  $p_{EP}$  are the p-values.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
$t_{HP}$	0.01130	0.01133	0.0123	0.01378	0.01723	0.00181
$t_{RP}$	1.15841	1.14383	1.13693	1.12501	1.12620	1.13843
$t_{EP}$	4.43899	4.27950	4.29839	4.32407	4.13205	4.29486
$p_{HP}$	0.04991	0.13144	0.00101	0.02497	0.09083	0.08052
$p_{RP}$	0.04955	0.12695	0.0012	0.0242	0.0889	0.0818
$p_{EP}$	0.04980	0.13016	0.00101	0.02508	0.09250	0.08033

3 and 5 with  $n_1 = n_2 = 10$ , and unbalanced in Cases 2, 4 and 6 with  $n_1 = 6$ ,  $n_2 = 18$ .

Table 1 and Figure 3 illustrate the high accuracy of our hybrid permutation technique. Furthermore, comparing with exact permutation or a random choice of 20,000 permutations, the hybrid permutation tests reduce more than 99% of the computational cost, and this efficiency gain increases with sample size. Table 1 shows the computation time and p-values of three permutation methods from one simulation. In order to demonstrate the robustness of our method, we

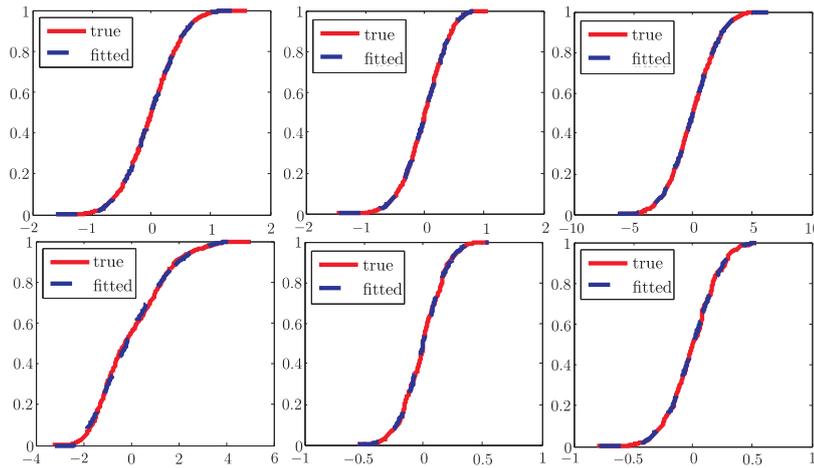


Figure 3. The Cumulative distribution functions (CDF) of exact permutation (in red) and Pearson distribution fitting (in blue) results. Left Column: for Case 1 (top) and Case 2 (bottom); Middle Column: for Case 3 (top) and Case 4 (bottom); Right Column: for Case 5 (top) and Case 6 (bottom).

Table 2. Robustness and accuracy comparison of hybrid permutation and random permutation across 10 simulations, considering the p-values of exact permutation as gold standard. Mean\_ABias\_HP and VAR\_ABias\_HP are the mean and variance of the absolute biases of p-values of hybrid permutation; Mean\_ABias\_RP and VAR\_ABias\_RP are the mean and variance of the absolute biases of p-values of random permutation, respectively.

Mean_ABias_HP	8.79e-5	8.97e-6	9.54e-5	2.16e-4	6.79e-4	4.53e-4
Mean_ABias_RP	2.82e-4	6.64e-5	2.14e-4	1.30e-3	2.78e-4	5.99e-4
VAR_ABias_HP	5.99e-8	1.34e-7	2.10e-6	3.66e-7	9.55e-7	9.78e-6
VAR_ABias_RP	1.98e-6	1.42e-7	1.41e-6	5.34e-6	1.05e-5	1.00e-5

repeated the simulation for 10 times in each case, and calculate the mean and variance of the absolute biases of p-values of both hybrid permutation and random permutation, treating the p-values of exact permutation as gold standard. In most cases, hybrid permutation is less biased and more stable than random permutation (Table 2), which demonstrates the robustness and accuracy of our method.

## 6.2. Data for modified Hotelling's $T^2$ test statistics

We applied the method to the MRI hippocampi that were semi-automatically segmented by human expert raters, and manually grouped into two groups with 21 subjects in group *A*, and 15 in group *B* (<http://www.ia.unc.edu/dev/download/shapeAnalysis/>). This dataset serves as a testing dataset for the methodology validation of the SPHARM-PDM software.

Evaluation of the standard hypothesis test using hybrid permutation with the modified Hotelling's  $T^2$  test statistics on the hippocampus dataset is shown in Figures 4 (a) and (b). It can be seen that the Pearson distribution approximation leads to little discrepancy with the raw  $p$ -value map from the permutation distribution. The false positive error control results are shown in Figure 4 (c)–(e). The raw  $p$ -map has the largest significance region, including numerous false positives. Since FWER control is conservative, it leads to the smallest size of significance areas. Our ROI constrained adaptive FDR method discovers more significant locations than the conventional BH's FDR method under the same FDR level (0.05).

## 7. Conclusion

We present and develop a statistical surface morphometry analysis method by using novel hybrid permutation tests, where the permutation distributions are accurately approximated through Pearson distributions for considerably reduced computational cost. General and analytical formulations for moments

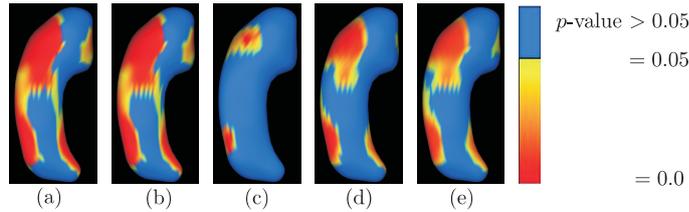


Figure 4. (a) and (b): Comparison of techniques in raw  $p$ -value measurement at  $\alpha = 0.05$  (without correction), through real permutation ((a); number of permutations=10,000), and using the present hybrid permutation (b). (c)–(e): Comparison of multiple testing techniques: FWER corrected  $p$ -map (c); BH's FDR corrected  $p$ -map (d); our ROI constrained adaptive FDR corrected  $p$ -map (e).

of permutation distribution are derived, for both univariate and multivariate test statistics. The proposed hybrid strategy takes advantage of nonparametric permutation tests and the parametric Pearson distribution approximation to achieve both accuracy/flexibility and efficiency. Note that the theoretical derivations described in this work are general and can be applied to any linear test statistics on multivariate data, and not limited to the test statistics used here. In addition, since the permutation of indices in our derivations is separable from the data, all the coefficients needed for moment calculations can also be estimated through computer simulations once the test statistic is chosen. Furthermore, a ROI constrained adaptive procedure is developed and employed to control the FDR for increased power for multiple testing. Experimental results demonstrate the effectiveness of our statistical analysis approach. The matlab codes of the simulated data for reproducibility are given in the online supplement (<http://www3.stat.sinica.edu.tw/statistica>).

### Acknowledgement

We would like to thank Professor Xiaofeng Shao, Professor John I. Marden and Dr. Maosheng Xiong for their invaluable comments and discussions. This work is part of Chunxiao Zhou's PH.D. thesis supported in part by a grant from Beckman Institute.

### References

- Ashburner, J. and Friston, K. (2004). Morphometry. In *Human Brain Function*, 2<sup>nd</sup> Edition. (Editor by R. Frackowiack, S. Zeki, J. Ashburner, K. Friston and C. Frith), 707-724. Elsevier, London.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289-300.

- Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006). Adaptive linear step-up false discovery rate controlling procedures. *Biometrika* **93**, 491-507.
- Brechtbuehler, C., Gerig, G. and Kubler, O. (1995). Parametrization of closed surfaces for 3-D shape description. *Comp. Vision, Graphics, and Image Proc.* **61**, 154-170.
- Chung, M. K., Worsley, K. J., Paus, T., Cherif, C., Giedd, J. N., Rapoport, J. L. and Evans, A. C. (2001). A unified statistical approach to deformation-based morphometry. *NeuroImage* **14**, 595-606.
- Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. John Wiley, Chichester.
- Good, P. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*. 3rd ed.. Springer, New York.
- Hahn, J. and Shapiro, S. S. (1967). *Statistical Models in Engineering*. John Wiley, New York.
- Hubert, L. (1987). *Assignment Methods in Combinatorial Data Analysis*. Marcel Dekker, Inc., New York.
- Johnson, R. A. and Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*, 5th Edition. Prentice Hall, Upper Saddle River, New Jersey.
- Liao, F. (2002). *Statistical Group Comparison*. John Wiley, New York.
- Mielke, P. Jr. and Berry, K. (1994). Permutation tests for common locations among samples with unequal variances. *J. Educ. and Behav. Stat.* **19**, 217-236.
- Nichols, T. E. and Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping* **15**, 1-25.
- Pantazis, D., Leahy, R. M., Nichols, T. E. and Styner, M. (2004). Statistical surface-based morphometry using a non-parametric approach. *IEEE Int. Sym. Biomedical Imaging*, 1283-1286.
- Pesarin, F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. John Wiley, Chichester.
- Shen, D. and Davatzikos, C. (2003). Very high resolution morphometry using mass-preserving deformations and HAMMER elastic registration. *Neuroimage* **18**, 28-41.
- Storey, J. D. (2002). A direct approach to false discovery rates. *J. Roy. Statist. Soc. Ser. B* **64**, 479-498.
- Styner, M., Lieberman, J. A., McClure, R. K., Weingberger, D. R., Jones, D. W. and Gerig, G. (2005). Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. *Proc. Nat. Acad. Sci.* **102**, 4872-4877.
- Styner, M., Oguz, I., Xu, S., Brechtbuehler, C., Pantazis, D., Levitt, J., Shenton, E. and Gerig, G. (2006). Framework for the statistical shape analysis of brain structures using SPHARM-PDM. *Open Science Workshop at MICCAI 2006, Insight Journal*, Dspace link <http://hdl.handle.net/1926/215>.
- Taylor, J. E. and Worsley, K. J. (2007). Random fields of multivariate test statistics, with applications to shape analysis and fMRI. *Ann. Statist.*, in press.
- Thompson, P. M., Hayashi, K. M., De, Zubizaray G. I., Janke, A. L., Rose, S. E., Semple, J., Hong, M. S., Herman, D. H., Gravano, D., Doddrell, D. M. and Toga, A. W. (2004). Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage* **22**, 1754-1766.
- Wald, A. and Wolfowitz, J. (1944). Statistical tests based on permutations of the observations. *Ann. Math. Statist.* **15**, 358-372.
- Wang, Y. and Staib, L. H. (2000). Boundary finding with prior shape and smoothness models. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 738-743.

Yushkevich, P. A., Zhang, H. and Gee, J. C. (2006). Continuous medial representation for anatomical structures. *IEEE Trans. on Medical Imaging* **25**, 1547-1564.

Zhou, C., Park, D. C., Styner, M. and Wang, Y. M. (2007). ROI constrained statistical surface morphometry. *IEEE International Symposium on Biomedical Imaging*, 1212-1215.

Departments of Statistics, Electrical & Computer Engineering, Beckman Institute, University of Illinois at Urbana-Champaign, IL, U.S.A.

E-mail: czhou4@illinois.edu

Departments of Statistics, Psychology, Bioengineering, Beckman Institute, University of Illinois at Urbana-Champaign, IL, U.S.A.

E-mail: ymw@illinois.edu

(Received April 2007; accepted March 2008))