# STATISTICAL SURFACE-BASED MORPHOMETRY USING A NON-PARAMETRIC APPROACH

*Dimitrios Pantazis*[1], *Richard M. Leahy*[1], *Thomas E. Nichols*[2], *Martin Styner*[3]

[1]Signal & Image Processing Inst., University of Southern California, Los Angeles, CA 90089-2564
[2]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, USA
[3]M.E.Müller Research Center STB, University of Bern, 3001 Bern, Switzerland

## ABSTRACT

We present a novel method of statistical surface-based morphometry based on the use of non-parametric permutation tests. In order to evaluate morphological differences of brain structures, we compare anatomical structures acquired at different times and/or from different subjects. Registration to a common coordinate system establishes corresponding locations and the differences between such locations are modeled as a displacement vector field (DVF). The analysis of DVFs involves testing thousands of hypothesis for signs of statistically significant effects. We randomly permute the surface data among two groups to determine thresholds that control the familywise (type 1) error rate. These thresholds are based on the maximum distribution of the amplitude of the vector fields, which implicitly accounts for spatial correlation of the fields. We propose two normalization schemes for achieving uniform spatial sensitivity. We demonstrate their application in a shape similarity study of the lateral ventricles of monozygotic twins and non-related subjects.

## 1. INTRODUCTION

The advancement of MRI has given us an invaluable tool for extracting structural brain information. The study of brain morphology has emerged as a new field of computational neuroanatomy and can provide great insights into brain function and development, as well as investigate the effects of many pathological diseases. In order to evaluate morphological differences of brain structures we need to compare structures extracted from many images acquired at different times and from different subjects. A statistical analysis of these structures may rely on global features, as in classical MRI-based volumetry, which measures and compares the volumes of homologous regions. Statistical analysis of shape based features has also been proposed. In shape analysis, anatomically corresponding locations between different images are computed and a mathematical transformation between such locations, called deformation, is evaluated. This deformation can be mathematically modeled as a displacement vector field (DVF) and the study of this field is called deformation-based morphometry. Recent work includes computation of DVF's using deformable registration schemes on images [1, 2] and using structural correspondence establishing methods on boundary and medial shape descriptions [3, 4].

Analysis of the DVFs involves testing from a few tens to many thousands of hypothesis (one per surface element) for statistically significant effects. It is important to control for the multiple testing problem, and the most common measure of multiple false positives is the familywise error rate (FWER).

The multiple testing problem has been an active area of research in the functional neuroimaging community. The most widely used methods in the analysis of neuroimaging data use random field theory [5] [6] and make inferences based on the maximum distribution. In this framework, a closed form approximation for the tail of the maximum distribution is available, based on the expected value of the Euler characteristic of the thresholded image [6]. However, random field theory relies on many assumptions such as the same parametric distribution at each spatial location, a point spread function with two derivatives at the origin, sufficient smoothness to justify the application of the continuous random field theory, and a sufficiently high threshold for the asymptotic results to be accurate.

In contrast, non-parametric methods rely on minimal assumptions, deal with the multiple comparisons problem and can be applied when the assumptions of the parametric approach are untenable. Non-parametric permutation tests are exact, distribution free and adaptive to underlying correlation patterns in the data. Further, they are conceptually straightforward and, with recent improvements in computing power, are computationally tractable. They have been

This work was supported in part by grants from NIBIB (R01 EB002010) and NCRR (P41 RR013642). All permutation experiments were carried out by Martin Styner. Douglas Jones and Daniel Weinberger at NIMH, Clinical Brain Disorder Branch (Bethesda, MD), kindly provided the original MRI datasets of the ventricle data. Jeffrey Lieberman and the neuro-image analysis lab at UNC Chapel Hill, Department of Psychiatry, provided the segmentations of the ventricle data.

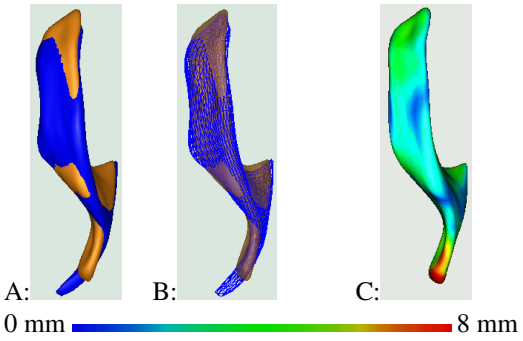


**Fig. 1**. Visualization of the distance map between the right lateral ventricles of a twin-pair (superior view). A: The two ventricles after alignment. B: Same as A, one ventricle shown transparent and the other as grid-mesh. C: Distance map with color-coded distance at each boundary-point.

applied in a wide range of functional imaging applications [7] [8]. Sowell et al. [9] has used permutation tests to validate statistical results of a random field study on brain morphology changes between childhood, adolescence and adulthood. In contrast to that work, the method presented here produces a local threshold map applied directly to the DVF.

## 2. METHOD

The objective of this study is to localize regions of brain structures that exhibit statistically significant morphological variation among two population groups while controlling the risk of any false positives. We find local thresholds that control the FWER and at the same time achieve uniform sensitivity among all surface elements.

### 2.1. Permutation Scheme

We assume we have two groups of DVFs, group A and group B. Each DVF measures the difference between a brain structure and a template or a designated comparison object. For example, group A may be composed of lateral ventricles of twins and each DVF models the morphological difference for each pair of twins (see Fig. 1). Group B may have similarly constructed DVFs but from pairs of unrelated individuals of similar age and same gender. We want to test the two groups for difference in the means at each spatial location (see Fig. 2). Permutation tests are a valid and tractable approach for such an application, as described in the introduction. Our null hypothesis is that the distribution of the DVF at each spatial element is the same for every subject regardless of the group. Permutations among the two groups satisfy the exchangeability condition, i.e. they leave the distribution of the statistic of interest unaltered under the null hypothesis. Given $n_1$ members of the first group $a_k$, $k = 1 \dots n_1$ and $n_2$ members of the second

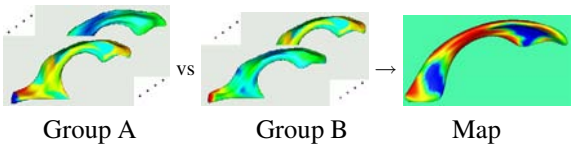


**Fig. 2**. Concept of statistical significance maps: For two groups of objects, DVF maps are compared locally resulting in a map of significant differences between groups.

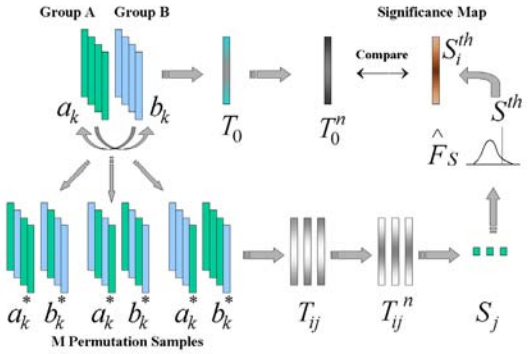


**Fig. 3**. Illustration of the permutation scheme. We create $M$ permutation samples from the original data and form the statistic $T_{ij}$. The statistic is normalized into $T_{ij}^n$ and the data are summarized in space to create $S_j$. The empirical distribution of $S_j$, called $\hat{F}_S$ is used to define a global threshold $S^{th}$ which corresponds to spatially varying thresholds $S_i^{th}$.

group $b_k$, $k = 1 \dots n_2$, we can create $M \leq \binom{n_1+n_2}{n_2}$ permutation samples. A value of M from 1000 and up should yield results that are negligibly different from using all permutations [10].

Our modeling proceeds by forming a vector of statistics for each permutation sample $j$, called $\boldsymbol{T}_j$, with elements:

$$T_{ij} = \left| \frac{\sum_{k=1}^{n_1} \|a_{ki}^*\|}{n_1} - \frac{\sum_{k=1}^{n_2} \|b_{ki}^*\|}{n_2} \right| \qquad (1)$$

where $i$ is the spatial index, $j$ the permutation index, and $k$ the group member index. Further, $\|a_{ki}^*\|$ ($\|b_{ki}^*\|$) is the amplitude of the DVF at the $i_{th}$ spatial location of the $k_{th}$ member of group A (group B). The symbol $(^*)$ indicates that the values $a_{ki}^*$ and $b_{ki}^*$ are created by permutation (Fig. 3). In essence, $T_{ij}$ captures the difference of the amplitudes of the DVFs among the two groups. The goal is to use this statistic to estimate the maximum distribution of the vector amplitude differences over all surface elements $i$. We may do this by summarizing in space using a maximum statistic $S_j = max_i(T_{ij})$ and use the empirical distribution to extract thresholds that control the false positives to a desired level. However, the method requires some modifications to achieve uniform sensitivity across space.

### 2.2. Uniform Sensitivity

Permutation tests are always valid given the assumption of exchangeability under the null hypothesis. However, we may face the problem of uneven sensitivity in the spatial dimension if the null distribution varies across space. For example, with a maximum statistic over space, surface elements for which the morphological structure varies significantly will contribute more to the maximum distribution than others with smaller morphological variability. The impact is a relatively generous threshold for the highly varying locations and a stringent threshold for the other locations. We can overcome this problem by including some form of normalization in the statistic $T_{ij}$. In this paper we propose two different normalization schemes.

The first normalization scheme is based on computing the standard deviation $\sigma_i$ of the statistic $T_{ij}$ across permutations for each surface element $i$ and normalizing $T_{ij}$ with this value: $T^{\sigma}_{ij} = T_{ij}/\sigma_i$. The normalized statistic $T^{\sigma}_{ij}$ has a new empirical distribution over $j$ which, under specific assumptions, is invariant across all surface elements $i$. We assume that, under the null hypothesis, the x,y and z vector components of the DVFs are zero mean Gaussian random variables with possibly different variances $\sigma_x, \sigma_y$ and $\sigma_z$. It is very hard to extract the theoretical distribution of $T_{ij}$ based on eq 1, so we resorted to Monte Carlo simulations. For 1000 times we simulated 100 samples of $T_{ij}$ (excluding the outside absolute value operator) with $n_1 = n_2 = 10$ and tested how well they match a Gaussian distribution using a Lilliefors test. The 0.05 Lilliefors threshold is 0.0889 and only 51 out of 1000 samples exceeded this threshold, so $T_{ij}$ pass the Gaussianity test. Similar experiments indicated that the Gaussianity assumption is reasonable even for $n_1 = n_2 \geq 5$. Finally, other simulations demonstrated that $T_{ij}$ has the same skewed distribution for all elements $i$, as long as $\sigma_x = \sigma_y = \sigma_z$. Under the above conditions, the $T^{\sigma}_{ij}$ normalization scheme guarantees uniform sensitivity.

Alternatively, we can also normalize based on p-values, i.e. at each spatial location we compute the empirical distribution across permutations and then replace the statistic $T_{ij}$ for each permutation sample with its p-value. The p-value at surface element $i$ for permutation $j$, called $T^{p}_{ij}$, is defined by $T^{p}_{ij} = p_i(T_{ij})$, where:

$$p_i(t) = \frac{1}{M}\sum_j H(T_{ij} - t), \quad H(x) = \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases} \quad (2)$$

We now guarantee that $T^{p}_{ij}$ has a uniform distribution under $H_o$ for each $i$.

### 2.3. Local Threshold Map

We can use the information contained in the normalized data to define a local threshold map that controls the FWER to a desired level, say 5%, when applied to the original data. For this it is necessary to estimate the maximum distibution of the vector amplitude differences over all surface elements $i$. We achieve this by summarizing in space using an extremal statistic:

$$S^{\sigma}_j = max_i\{T^{\sigma}_{ij}\},\ S^{p}_j = min_i\{T^{p}_{ij}\} \quad (3)$$

Notice that the minimum p-value plays the same role as the maximum statistic in FWER. After summarizing in space we can use the empirical distribution of $S_j$ to define a threshold $S^{th}$ that controls the FWER. If $\hat{F}_{S^{\sigma}}$ and $\hat{F}_{S^{p}}$ are the empirical cdfs of $S^{\sigma}$ and $S^{p}$, then the appropriate global thresholds for a level $\alpha$ test would be $\hat{F}^{-1}_{S^{\sigma}}(1-\alpha)$ and $\hat{F}^{-1}_{S^{p}}(\alpha)$ respectively (Table 1). For example, if we choose a threshold that leaves 5% of the area of the empirical distribution on the right side for $S^{\sigma}_j$, resp. left side for $S^{p}_j$, then we have 5% probability of one or more false positives throughout the entire surface. The threshold $S^{th}$ cannot be directly applied to $T_{i0}$ (the statistic formed by the original data with permutation index $j = 0$). Since the statistic $T_{ij}$ was normalized separately for each surface element $i$, the same $S^{th}$ will correspond to different values of local thresholds at different surface elements. These local thresholds are found with the inverse normalization transformation, where surface element $i$ shows significant variation if $T_{i0} \geq S^{th}\sigma_i$, resp. $T_{i0} \geq p_i^{-1}(S^{th})$.

| | $T_{ij}$ Normalized $T^{n}_{ij}$ | Space-Summarizing $S_j$ | Global Threshold $S^{th}$ | Local Threshold $S^{th}_i$ |
|---|---|---|---|---|
| Method 1 | $T^{\sigma}_{ij}$ | $max_i\{T^{\sigma}_{ij}\}$ | $\hat{F}^{-1}_{S^{\sigma}}(1-\alpha)$ | $S^{th}\cdot\sigma_i$ |
| Method 2 | $T^{p}_{ij}$ | $min_i\{T^{p}_{ij}\}$ | $\hat{F}^{-1}_{S^{p}}(\alpha)$ | $p_i^{-1}(S^{th})$ |

**Table 1**. Normalization schemes, summary statistics and thresholds for the two permutation methods.

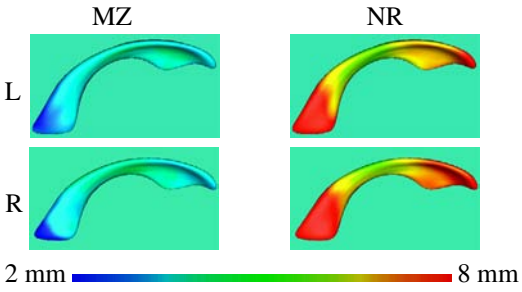


**Fig. 4**. Color-coded average distance maps visualize the absolute distances between pairs, averaged over the group. The displays show a smaller average distance between MZ twins compared to NR subjects.

## 3. RESULTS

We applied our method to a shape similarity study of lateral ventricles, a fluid-filled space in the human brain. We

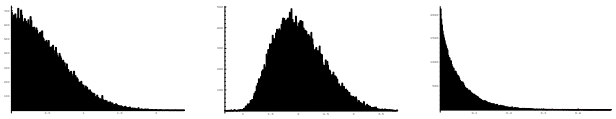

**Fig. 5**. Left: Example empirical distribution of the statistic $T_{ij}$ across $j$ at a surface element $i$. Middle: Empirical distribution of maximum statistic $S_j^\sigma$. Right: Empirical distribution of minimum statistic $S_j^p$.

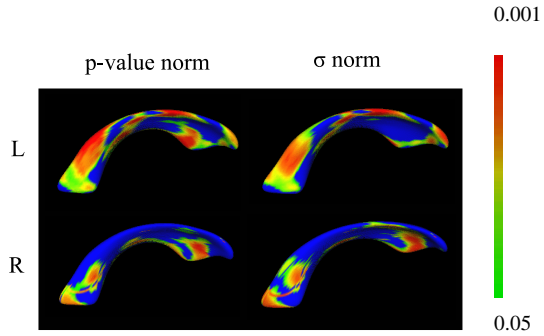


**Fig. 6**. Significance maps for the two normalization methods. The maps indicate regions where unrelated individuals have more variability that twins. The methods perform similarly on both the left and right lateral ventricle.

used 2 groups of subjects each consisting of 20 subjects (age and gender matched): a monozygotic (MZ) twin group and a non-related subject group (NR). In this study we are interested in the difference in similarity of the lateral ventricles between these groups. The ventricles were segmented from single gradient-echo MRI via automatic brain tissue classification and connectivity based postprocessing. Using the SPHARM surface description [3], each object was described by 4002 surface points with known correspondence. The difference between MZ twins and NR subjects was computed as shown in Figure 1. We generated $M = 50,000$ permutation samples and derived significance maps according to the methods described in Table 1.

Figure 4 shows the average distance maps, which clearly display a higher degree of similarity of the MZ group as compared to the NR group. Figure 5 shows the empirical distributions of $S_j^\sigma$ and $S_j^p$. The right tail of the maximum statistic $S_j^\sigma$ is well behaved and the threshold value $\hat{F}^{-1}_{S^\sigma}(0.95)$ can thus be reliably computed. This is not the case for the minimum statistic $S_j^p$, as the left tail is bounded by zero and, due to discreteness of the emperical distribution, the threshold $\hat{F}^{-1}_{S^p}(0.05)$ required many permutations ($M = 50,000$) for an accurate computation. The significance maps produced by the two normalization schemes are quite similar (Fig. 6), which demonstrates that the ventricle data reasonably satisfy the assumptions of the first normalization scheme.

## 4. CONCLUSION

We have presented a novel method for the statistical analysis of morphological differences of brain structures. We have used permutation tests to extract local thresholds that control the number of false positives at a specified level $\alpha$ over all surface elements and our approach implicitly accounts for spatial correlations of the data. Further, we have proposed two normalization schemes to achieve uniform detection sensitivity.

If the data are Gaussian under the null hypothesis, the normalization scheme based on the standard deviation $\sigma_i$ may be used. Otherwise, we resort to the p-value normalization scheme which requires more permutations but makes no distributional assumptions.

## 5. REFERENCES


[1] C. Davatzikos, M. Vaillant, S. Resnick, J.L Prince, S. Letovsky, and R.N. Bryan, "A computerized method for morphological analysis of the corpus callosum," *Journal of Computer Assisted Tomography*, vol. 20, pp. 88–97, Jan./Feb 1996.

[2] J.G. Csernansky, S. Joshi, L.E. Wang, J. Haller, M. Gado, J.P. Miller, U. Grenander, and M.I. Miller, "Hippocampal morphometry in schizophrenia via high dimensional brain mapping," *Proc. Natl. Acad. Sci. USA*, vol. 95, pp. 11406–11411, September 1998.

[3] G. Gerig, M. Styner, D. Jones, D. Weinberger, and J. Lieberman, "Shape analysis of brain ventricles using spharm," in *Mathematical Methods in Biomedical Image Analysis*. 2001, pp. 171–178, IEEE press.

[4] M. Styner, G. Gerig, J. Lieberman, D. Jones, and D. Weinberger, "Statistical shape analysis of neuroanatomical structures based on medial models," *Medical Image Analysis*, vol. 7, no. 3, pp. 207–220, 2003.

[5] M. K. Chung, *Statistical Morphometry in Neuroanatomy*, Ph.D. thesis, McGill University, Montreal, 2001.

[6] K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans, "A unified statistical approach for determining significant signals in images of cerebral activation," *Human Brain Mapping*, vol. 4, pp. 58–73, 1996.

[7] T. E. Nichols and A. P. Holmes, "Nonparametric permutation tests for functional neuroimaging: A primer with examples," *Human Brain Mapping*, vol. 15, pp. 1–25, 2001.

[8] D. Pantazis, T. E. Nichols, S. Baillet, and R. M. Leahy, "Spatiotemporal localization of significant activation in meg using permutation tests," in *Information Processing in Medical Imaging*, 2003, pp. 512–523.

[9] E. R. Sowell, P. M. Thompson, C. J. Holmes, T. L. Jernigan, and A. W. Toga, "In vivo evidence for post-adolescent frontal and striatal maturation," *Nature Neuroscience*, Sept./ Oct. 1999.

[10] E. S. Edgington, Ed., *Randomization Tests*, Academic Press, 1995.