Research Article

Statistical Analysis of Twin Populations Using Dissimilarity Measurements in Hippocampus Shape Space

Youngser Park,[1] Carey E. Priebe,[1] Michael I. Miller,[1] Nikhil R. Mohan,[1] and Kelly N. Botteron[2]

[1] *Center for Imaging Science, Johns Hopkins University, Baltimore, MD 21218, USA*
[2] *Department of Psychiatry, Washington University, St. Louis, MO 63110, USA*

Correspondence should be addressed to Youngser Park, youngser@jhu.edu

By analyzing interpoint comparisons, we obtain significant results describing the relationship in "hippocampus shape space" of clinically depressed, high-risk, and control populations. In particular, our analysis demonstrates that the high-risk population is closer in shape space to the control population than to the clinically depressed population.

1. INTRODUCTION

Major depressive disorder (MDD) is a mental disorder affecting about 16% of the US adult population, and is a major cause for concern not only in the United States but the world over. It is a disorder characterized by depressed mood, diminished interest or pleasure, significant weight loss, feelings of guilt or low self-worth, insomnia or hypersomnia, fatigue, poor concentration, or recurrent thoughts of death. The symptoms are widespread, and tend to be quite stable. In 2000, the World Health Organization (WHO) estimated depression to be the leading cause of disability as measured by years lived with disability (YLD) and the fourth leading contributor to the global burden of disease. See [1].

Over the years, a significant amount of research has been dedicated to finding physiological causes of MDD. One such study involved the catecholamine hypothesis [2] that suggested that MDD is caused by decreased levels of the neurotransmitters norepinephrine and serotonin. This finding led to most modern day medication for MDD, which works by preventing the reuptake of these neurotransmitters. Neuroimaging research has also shown that enlarged ventricles, sulci, reduced volume of the frontal lobe and basal ganglia are also associated with depressive episodes [2].

The studies aforementioned involved studying the brain once MDD had already set in. The physiological changes are associated with the symptoms themselves. What about physiological *predictors* for MDD? Such predictors would facilitate the diagnosis of the disorder well before the onset of the symptoms, perhaps allowing measures to prevent the symptoms from ever appearing.

A vast amount of research is being conducted in order to find biological predispositions to MDD. There is evidence correlating shape differences of the hippocampus to depression [3] and schizophrenia [4]. In this manuscript we analyze interpoint comparisons [5] to investigate the relationship in "hippocampus shape space" of three populations among twins. The subjects are categorized into three categories: the affected subjects (clinically depressed, or MDD), the nonaffected cotwin of the MDD subjects (high-risk, or HR), and the nonaffected twin pair (Control, or CTRL). The dataset includes both monozygotic (MZ) and dizygotic (DZ) twin subjects.

According to established literature, the concordance rate for monozygotic (MZ) HR subjects is 40%, and for dizygotic (DZ) HR subjects 11% [6]. This demonstrates that the subjects labeled HR (due to the fact that their twin is MDD) are in fact high risk—they develop MDD at a higher rate than the general population.

2. DATA

Our data set includes $n = 114$ subjects (57 twin pairs): 29 CTRL-CTRL pairs, 22 HR-MDD pairs, and 6 MDD-MDD pairs. The subjects are young female twins recruited through an epidemiological sample based on Missouri birth records.

To ensure that hippocampus shape space is the only independent variable, other factors had to be controlled; all of the subjects were right handed and were screened for factors that may cause structural changes of the brain such as loss of consciousness greater than 5 minutes, chronic medical or neurological illnesses, or pregnancy.

To obtain images of the hippocampus, very high resolution magnetic resonance imaging (MRI) scans were required. The Siemens Vision/Sonata 1.5T scanner was used to acquire three MPRAGE scans [7] (160 slices at $256x256$ FoV, 1.0 mm^3 isotropic voxels). Using Analyze [8], the images were registered and averaged, converted to 8-bits while optimizing the intensity range, and interpolated to 0.5 mm isotropic voxels. The image protocol implemented above allows for optimal comparative analysis.

For each of left and right hemispheres separately, 22 three-dimensional landmarks were identified for each hippocampus and were used to generate and align hippocampal subcubes to a standardized orientation. It is these landmark data that we employ herein.

3. SHAPE

Using the landmark data, for each pair of subjects and for each of left and right hippocampus, we produce an interpoint shape comparison, as described below.

For two subjects x and y (for the left hemisphere, say), let $x_1, \ldots, x_N$ and $y_1, \ldots, y_N$ be the corresponding landmarks, where $N = 22$.

3.1. Landmark matching

Finding the shape comparison involves a landmark matching (LM) transformation. The transformation is nonparametric, and this flexibility implies that overfitting must be guarded against via regularization. LM finds a diffeomorphism φ that minimizes an error criterion which includes both landmark mismatch and transformation complexity. That is,

$$\varphi^* = \arg \inf_{\varphi} \ \sigma \, d(\varphi_{\mathrm{identity}}, \varphi)^2 + \sum_{i=1}^{N} \|\varphi(x_i) - y_i\|^2, \quad (1)$$

where d is a geodesic distance in a group of diffeomorphisms [9] and $\sigma > 0$ is a regularization parameter which controls the relative contribution of transformation complexity versus landmark mismatch to the optimization objective. The algorithm solves the nonlinear Euler equation by a Newton method combined with a shooting procedure [10].

We use $\mathrm{LM}(x, y) = \|\varphi^*\|$, the energy of the minimizing diffeomorphism, as the shape comparison between two subjects x and y (for the left hemisphere, say).

3.2. Interpoint comparison matrices

Applying LM to the left or right hippocampus data for each pair of subjects yields an interpoint comparison matrix $\widetilde{D}$. However, $\widetilde{D}$ is $n \times n$, hollow (zeros on the diagonal) and is asymmetric. That is, we obtain matrices $\widetilde{D}_{\mathrm{LM\text{-}Left}}$ and $\widetilde{D}_{\mathrm{LM\text{-}Right}}$.
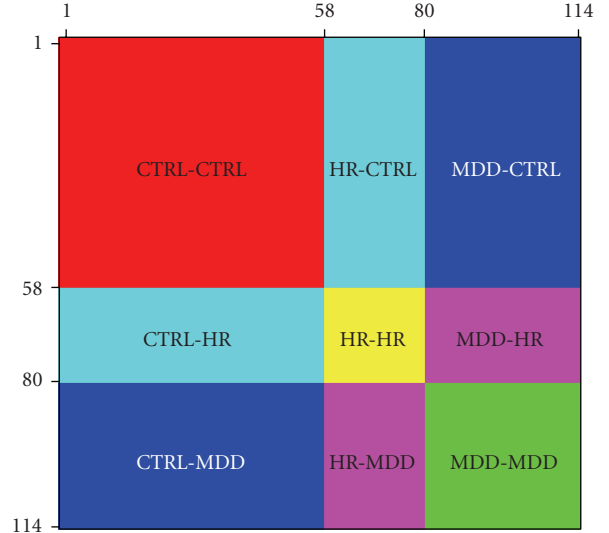

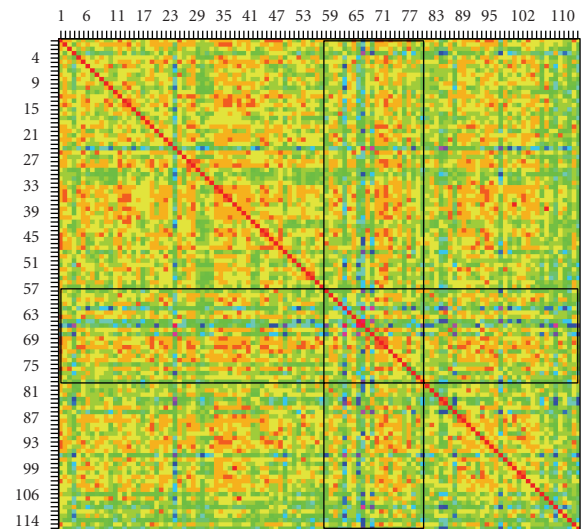FIGURE 1: Structure of the interpoint comparison matrices D for the 114 subjects.

FIGURE 2: The interpoint comparison matrix $D_{\mathrm{LM\text{-}Left}}$, after symmetrization, for the 114 subjects. The comparison values are color-coded, with red representing zero (e.g., the diagonal entries) to green representing large values.

The nature of the hippocampus shape space is such that under ideal conditions, it should yield a symmetric distance matrix. The asymmetry of the matrix $\widetilde{D}$ does not reflect the true nature of the hippocampus shape space, and is in fact a result of the limitations in the LM matching method. Hence, before further investigation, $\widetilde{D}$ must be symmetrized to D, using an appropriate symmetrization technique [11]. In this work we symmetrize via $d_{ij} = \min\{\tilde{d}_{ij}, \tilde{d}_{ji}\}$.

Figure 1 depicts the structure of the interpoint comparison matrices for the 114 subjects. Figure 2 depicts the actual interpoint comparison matrix $D_{\mathrm{LM\text{-}Left}}$ (after symmetrization) for the 114 subjects.
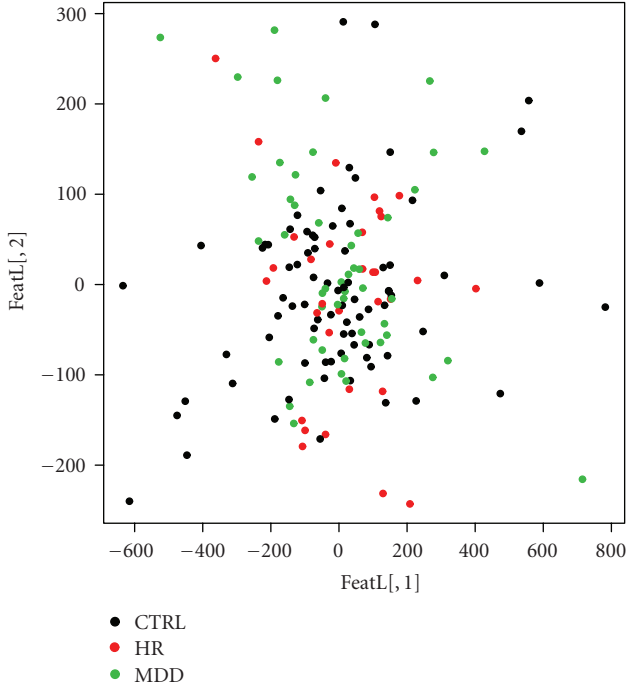
FIGURE 3: A multidimensional scaling scatter plot of $D_{\text{LM-Left}}$ mapped into $\mathbb{R}^2$. Little can be discerned from this plot regarding the relationship in hippocampus shape space of the three populations (MDD, HR, CTRL)—no class-conditional differentiation is apparent.

4. STATISTICAL ANALYSIS

Our task is to begin describing the relationship of the three populations (MDD, HR, CTRL) amongst one another in the hippocampus shape space elicited by the LM interpoint comparisons. First, we present a multidimensional scaling (MDS) [12] scatter plot; unfortunately, we see in Figure 3 that no significant relationship can be discerned from this plot. Employing linear discriminant analysis (LDA) after MDS for all possible MDS target dimensionalities—analysis via LDA ∘ MDS ∘ LM (·) *a la* Miller et al. [13]—yields no classification capabilities statistically significantly superior to chance. Nevertheless, we will see in Figures 4 and 5 a suggestion that perhaps progress can be made on our task, given a sufficiently clever methodology.

Figure 4 depicts kernel probability density estimates [14] for the LM-Left comparisons to show that the entries of the interpoint comparisons matrix $D_{\text{LM-Left}}$ that correspond to comparisons between HR and CTRL (the solid line in Figure 4) are, overall, *smaller* than the entries which correspond to comparisons between HR and MDD. That is, Figure 4 suggests a *stochastic ordering* relationship [15]: $d(\text{HR}, \text{CTRL}) <^{\text{st}} d(\text{HR}, \text{MDD})$. Such a result is precisely what we seek. Again, dependencies amongst the entries of D make it difficult to assess the statistical significance of the result depicted in Figure 4.

Each row of the interpoint comparisons matrix D, corresponding to a single HR subject, gives rise to two samples: $\{d(\text{HR}, \text{CTRL}_i)\}$ and $\{d(\text{HR}, \text{MDD}_j)\}$. That is, we have
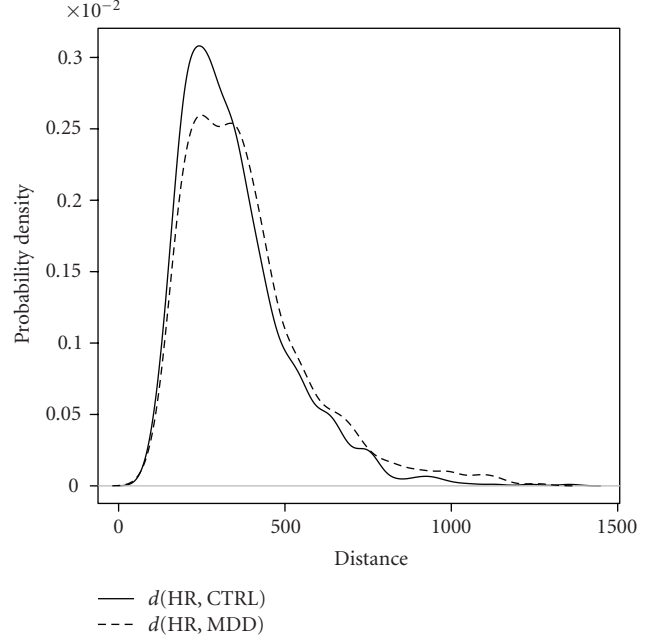

FIGURE 4: This figure shows kernel probability density estimates for $D_{\text{LM-Left}}$. The solid line depicts $d(\text{HR}, \text{CTRL})$ and the dashed line depicts $d(\text{HR}, \text{MDD})$.
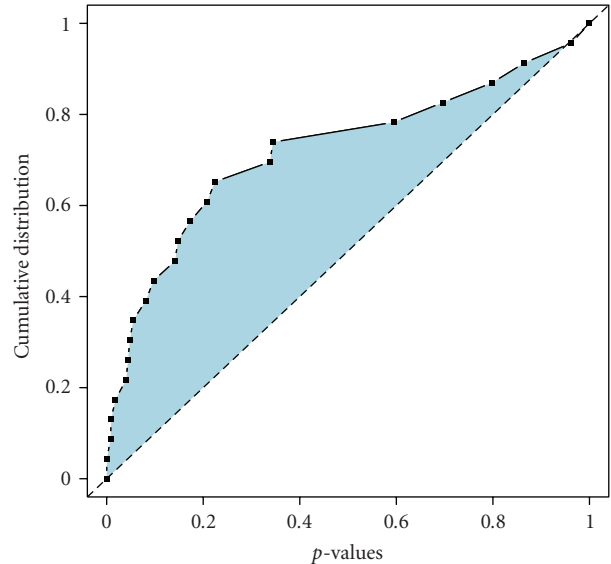

FIGURE 5: This figure shows the quantile-quantile plot for $D_{\text{LM-Left}}$. Depicted are the individual P-values for a Wilcoxon-Mann-Whitney test of each HR subject, in turn, based on the two samples $\{d(\text{HR}, \text{CTRL}_i)\}$ and $\{d(\text{HR}, \text{MDD}_i)\}$.

the vector of comparisons from that HR subject to every CTRL subject, and we have the vector of comparisons from that HR subject to every MDD subject. (We do not include in these vectors the twin of the particular HR subject under consideration; ignoring twinnedness in the analysis proves beneficial that we eliminate bias in similarity status between a subject and her twin that is not due to

condition (MDD,HR,CTRL).) For this individual HR subject's two sample data, a Wilcoxon-Mann-Whitney test [16] of the null hypothesis that the distribution of comparisons $d(\mathrm{HR}, \mathrm{CTRL})$ is the same as the distribution of comparisons $d(\mathrm{HR}, \mathrm{MDD})$, against the alternative of *stochastic ordering*, yields a P-value. Figure 5 provides a quantile-quantile plot of these P-values for $D_{\mathrm{LM\text{-}Left}}$. Under the null hypothesis, these P-values would be expected to be distributed approximately uniform(0,1). The plot demonstrates a clear deviation from a uniform distribution, again suggesting a *stochastic ordering* relationship—$d(\mathrm{HR}, \mathrm{CTRL}) <^{\mathrm{st}} d(\mathrm{HR}, \mathrm{MDD})$. Again, dependencies amongst the entries of D make it difficult to assess the statistical significance of the result depicted in Figure 5.

The quantile-quantile plot independently reiterates the suggestion of a stochastic ordering relationship that was first seen using the kernel probability density estimates. Thus, while Figures 4 and 5 give an inkling of the type of information that can be gleaned regarding the relationship in hippocampus shape space of the three populations (MDD, HR, CTRL) amongst one another, it remains henceforth to accurately assess the Figures' suggestion.

5. CLASSIFICATION

To further uncover the characteristics of hippocampus shape space, we consider the task of *classifying* each HR subject as either MDD or CTRL.

As before, we consider the two samples, $\{d(\mathrm{HR}, \mathrm{CTRL}_i)\}$ and $\{d(\mathrm{HR}, \mathrm{MDD}_j)\}$, associated with each individual HR subject. We classify the HR subject as belonging to MDD or CTRL based on the Wilcoxon-Mann-Whitney test statistic P-value, as described in [17]; (see also [15, page 183]).

Once we have classified each of the HR subjects in this way, we assess the relative similarity of HR to CTRL versus MDD based on the classifier's performance—based on the collection of HR subjects' classifier-assigned class labels, taken as a whole.

This procedure can be employed with LM interpoint comparisons obtained on Left, Right, or both Left and Right hippocampuses, and with any of the three populations (HR, CTRL, MDD) as the population of interest—the role of HR in the description above.

6. RESULTS

Classifying the 22 HR subjects as either MDD or CTRL using $D_{\mathrm{LM\text{-}Left}}$ results in 19 classified as CTRL versus 3 classified as MDD. The probability of obtaining a result this extreme or more extreme (the P-value) under the least favorable null hypothesis H_0 : *HR are equally likely to be classified as MDD as CTRL* is $P < .01$ against each one-sided alternative. LM-Right yields 16 classified as CTRL versus 6 classified as MDD—classification performance not statistically significantly distinguishable from chance. Combining left and right, the shape comparisons LM (LM-Left and Right) yields 20 classified as CTRL versus 2 classified as MDD—$P < .0005$ for each one-sided alternative and strong statistical evidence that HR is more like CTRL than MDD in hippocampus shape space.

HR   CTRL   MDD

FIGURE 6: Artist's rendition of what hippocampus shape space might look like were it one-dimensional—if the population shapes could be accurately represented in $\mathbb{R}^1$. Our results suggest that the joint relationship of the three populations, in terms of shape, puts the CTRL population *between* the HR and MDD populations. The relationship depicted here holds for both LM-Left and LM-Right, although our results suggest that for the left hippocampus CTRL is shifted closer to HR while for the right hippocampus the CTRL is shifted closer to MDD.

An analogous analysis—classifying the 33 MDD subjects as either HR or CTRL using LM-Left—shows that MDD is more like CTRL than it is like HR ($P < .00002$ for each one-sided alternative), and that the left carries more information than does the right—the P-values are smaller indicating that the signal is stronger.

The results obtained from classifying the 59 CTRL subjects as either HR or MDD are more nuanced: in this case, using LM-Left indicates that CTRL is more like HR than it is like MDD ($P < .0005$) while using LM-Right indicates that CTRL is more like MDD than it is like HR ($P < .0005$). This hemispherical ambiguity provides further insight into hippocampus shape space.

Finally, we note that in the last two columns of Table 1 we consider classifying the 22 HR subjects (via leave-one-out crossvalidation) as HR or MDD and as HR or CTRL. These results are consistent with our other findings—HR is more difficult to distinguish from CTRL than from MDD, and the information extracted via LM-Left is more powerful for this task than is LM-Right.

7. CONCLUSIONS/DISCUSSION

Our analysis indicates that HR is more like CTRL than it is like MDD, MDD is more like CTRL than it is like HR, and CTRL is not obviously more like one or the other. Also, we discern that the left hippocampus carries more information than does the right.

If hippocampus shape space were one-dimensional—if the population shapes could be accurately represented in $\mathbb{R}^1$—then the joint relationship described by these three results could be depicted as in Figure 6, with the CTRL population *between* the HR and MDD populations in terms of shape. However, it must be noted that this depiction (Figure 6) offers only a simplified view of the true infinite dimensional nature of the shape space configuration, as suggested by the fact that the 2-dimensional MDS embedding depicted in Figure 3 presents little or no class separation.

7.1. On Populations

Our stated task is in terms of *populations*—to begin describing the relationship in hippocampus shape space of the three populations (MDD, HR, CTRL) amongst one another. However, our results are *conditional*—using LM-Left we classify, for example, the 22 HR subjects representing the HR

TABLE 1: Output of classifier based on the Wilcoxon-Mann-Whitney test statistic. For example, the first numerical column, "H : CvM," gives the number of HR classified as CTRL versus MDD and the second numerical column, "H : MvC," gives the number of HR classified as MDD versus CTRL. Thus, we find that combining left and right, the shape comparisons LM (LM-Left and Right) yields 20 HR classified as CTRL versus 2 HR classified as MDD—strong statistical evidence that HR is more like CTRL than MDD in hippocampus shape space. (This analysis is based on 22 HR subjects, 33 MDD subjects, and 59 CTRL subjects. Thus, e.g., the two HR numbers, H : CvM and H : MvC, should sum to 22. Discrepancies are due to situations in which the classifier makes "no decision" as described in [17]; (see also [15, page 183]).

	H : CvM	H : MvC	M : HvC	M : CvH	C : HvM	C : MvH	H : HvM	H : HvC
LM-Left	19	3	5	28	48	11	16	7
LM-Right	16	6	9	24	22	33	13	8
LM-Left & **Right**	20	2	6	27	31	25	16	6

population as belonging to either the MDD or the CTRL class, conditionally on "training" data from MDD and CTRL. This, in fact, is the standard approach in probabilistic pattern recognition; see, for example, [18]. The difference between a focus on populations versus conditionals is indicative of a difference between "policy science" and "laboratory science" [19]. A justification for the conditional approach in "laboratory science" is given in [18] where it is claimed that the unconditional approach " ... would be unnatural, because in a given application, one has to live with the [training] data at hand." In "policy science", however, knowledge about the populations themselves may be the focus.

By performing our analysis thrice, for each of the three populations in turn conditionally on the "training" data from the other two, we obtain three conditionals. Letting n_j denote the class-conditional sample sizes for each of the three classes, we see that the joint distribution for our sample is $(n_1+n_2+n_3) \cdot d$-dimensional (where d is the presumed "shape space" dimensionality of each observation). Each conditional considered is $n_j \cdot d$-dimensional, with one population remaining. The overall joint distribution of interest—the three populations in "shape space"—is of course not simply the product of our three conditionals. However, *some* population inferences regarding stochastic ordering can be performed via the (multiple) conditionals, and in particular the conditional approach justifies the simplistic view of our three populations in "shape-space" given by Figure 6.

REFERENCES

[1] "Depression. National Institute of Mental Health," http://www .nimh.nih.gov/publicat/depression.cfm.

[2] L. B. Alloy, J. H. Riskind, and M. J. Manos, *Abnormal Psychology: Current Perspectives*, McGraw-Hill, New York, NY, USA, 9th edition, 2005.

[3] J. Posener, L. Wang, J. Price, et al., "High dimensional mapping of the hippocampus in depression," *American Journal of Psychiatry*, vol. 160, no. 1, pp. 83–89, 2003.

[4] J. G. Csernansky, S. Joshi, L. Wang, J. W. Haller, M. Gado, J. P. Miller, U. Grenander, and M. I. Miller, "Hippocampal morphometry in schizophrenia by high dimensional brain mapping," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 19, pp. 11406–11411, 1998.

[5] J.-F. Maa, D. K. Pearl, and R. Bartoszyński, "Reducing multidimensional two-sample data to one-dimensional interpoint comparisons," *Annals of Statistics*, vol. 24, no. 3, pp. 1069–1074, 1996.

[6] "Genetics and major psychiatric disorders:a program for genetic counselors. National Coalition for Health," http://www .nchpeg.org/cdrom/empiric.html.

[7] J. P. Mugler III and J. R. Brookeman, "Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE)," *Magnetic Resonance in Medicine*, vol. 15, no. 1, pp. 152–157, 1990.

[8] Analyze Software, "Mayo Clinic," http://www.mayo.edu/bir/ Software/Analyze/.

[9] M. I. Miller, A. Trouve, and L. Younes, "Geodesic shooting for computational anatomy," *Journal of Mathematical Imaging and Vision*, vol. 24, no. 2, pp. 209–228, 2006.

[10] S. Allassonnire, A. Trouve, and L. Younes, "Geodesic shooting and diffeomorphic matching via textures meshes," in *Proceedings of the 5th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR '05)*, pp. 365–381, Augustine, Fla, USA, November 2005.

[11] T. Saito and H. Yadohisa, *Data Analysis of Asymmetric Structures*, Marcel Dekker, New York, NY, USA, 2005.

[12] T. Cox and M. Cox, *Multidimensional Scaling*, Chapman & Hall, New York, NY, USA, 2nd edition, 2001.

[13] M. Miller, C. Priebe, and Y. Park, "Collaborative computational anatomy: the perfect storm for mri morphometry study of the human brain via diffeomophic metric mapping, multidimensional scaling and linear discriminant analysis," to appear in *Proceedings of the National Academy of Science*.

[14] B. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, New York, NY, USA, 1986.

[15] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, vol. 1, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 2005.

[16] J. Rice, *Mathematical Statistics and Data Analysis*, Addison-Wesley, Reading, Mass, USA, 2nd edition, 1995.

[17] C. E. Priebe, "Olfactory classification via interpoint distance analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 4, pp. 404–413, 2001.

[18] L. Devroye, L. Gyorfi, and G. Lugosi, "A Probabilistic Theory of Pattern Recognition," Number 31 in Applications of mathematics, 1996.

[19] B. Caffo, Personal communication, 2006.