Increasing DBM Reliability using Distribution Independent Tests and Information Fusion Techniques

Vidya Rajagopalan

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Dr. Chris. L. Wyatt    Chair
Dr. Lamine Mili
Dr. Yue Wang
Dr. Ge Wang
Dr. Inyoung Kim

Oct 6, 2009
Blacksburg, Virginia

Keywords: deformation based morphometry, MRI, image registration, distribution independent tests, information fusion
Copyright 2009, Vidya Rajagopalan
Abstract

Increasing DBM Reliability using Distribution Independent Tests and Information Fusion Techniques

Vidya Rajagopalan

In deformation based morphometry (DBM) group-wise differences in brain structure are measured using deformable registration and some form of statistical test. However, it is known that DBM results are sensitive to both the registration method and statistical test used. Given the lack of an objective model of group variation it has been difficult to determine the extent of the influence of registration implementation or contraints on DBM analysis. In this thesis, we use registration methods with varying levels of theoretic similarity to study the influence of registration mechanics on DBM results. We show that because of the extent of the influence of registration mechanics on DBM results, analysis of changes should always be made with a thorough understanding of the registration method used. We also show that minor variations in registration methods can lead to large changes in DBM results. When using DBM, it would be imprudent to use only one registration method to draw any conclusions about the variations being studied. In order to provide a more complete representation of inter-group changes, we propose a method for combining multiple registration methods using Dempster-Shafer evidence theory to produce belief maps of categorical changes between groups. We show that the Dempster-Shafer combination produces a unique and easy to interpret belief map of regional changes between and within groups without the complications associated with hypothesis testing.

Another, often confounding, element of DBM is the parametric hypothesis test used to specify voxels undergoing significant change between the two groups. The accuracy and reliability of these tests are contingent on a number of fundamental assumptions made about the distribution of the data used in the tests. Many DBM studies often overlook these assumptions and fail to verify their validity for the data being tested. This raises many doubts about the credibility of the results from such tests. In this thesis, we propose to perform statistical analysis on DBM data using non-parametric, distribution independent hypothesis tests. With no data distributional assumptions, these tests provide both increased flexibility and reliability of DBM statistical analysis.
Acknowledgements

I am very grateful to my advisor Dr. Chris Wyatt for the wonderful opportunity of working with him, for the opportunity of working on the non-human primate project which gave us the very famous nickname of “the drunk monkeys group” and also for steering me towards neuroimaging statistics. I have immensely enjoyed my work with him and also the numerous discussion on image processing, biology, evolution and statistics. I am also very grateful to my committee members for the valuable suggestions throughout my PhD.

I am also grateful to Dr. Wyatt for bringing together a wonderful group of people at the Bioimaging Systems Lab (BSL). I am very grateful to all my friends at BSL (past and current): Jung, Bonsong, Jenny, Yuan, Ginger, Shaza, Stephen, Micah and Matt. I have immensely enjoyed all the good conversations both technical and otherwise. I have so many wonderful memories of my time with them which I shall cherish forever. I am particularly thankful to Jenny without whom my quest to vanquish the code-bugs would have lasted too long.

To my other friends at Blacksburg: life would have not been the same without you. Ramya, Amrita, Anu, Ajit, Siva, Rajesh, Arun and Parthi have been my family away from home. To Sunitha and SreeVi for showing me what the word BFF actually means.

I finally thank the most important people in my life: mom, dad and Goutham. All of this would have never happened but for my parents who taught me that education is much bigger than grades, transcripts and diplomas. I would like to thank Goutham for being patient with me through these years, introducing me to Cilk, disagreeing with my views on frequentist statistics and most of all for just being him.
# Contents

1 Influence of Registration Mathematics on Deformation Based Morphometry 1
   1.1 Introduction .................................................... 1
   1.2 Background ..................................................... 3
      1.2.1 Registration Methods ....................................... 3
      1.2.2 Log-Euclidean Strain Vectors (LeSV) ...................... 4
      1.2.3 Multivariate Statistics ...................................... 4
   1.3 Methods ......................................................... 5
      1.3.1 Registration ................................................ 5
   1.4 Results and Discussion ......................................... 6
      1.4.1 Aging Dataset ............................................... 6
      1.4.2 Oasis Dataset ............................................... 8
      1.4.3 Comparison of properties across datasets ................. 10
   1.5 Conclusion .................................................... 10

2 Combining Morphometric Evidence from Multiple
   Registration Methods using Dempster-Shafer Theory 11
   2.1 Introduction .................................................... 11
   2.2 Background – Dempster-Shafer Theory ...................... 12
   2.3 Methods ......................................................... 12
2.3.1 Combining Morphometric Information using DST .......................... 13
2.4 Results and Discussion ..................................................................... 15
2.5 Conclusion ....................................................................................... 17

3 Distribution Independent Statistical Testing for Deformation Based Morphometry 24
   3.1 Introduction .................................................................................. 24
   3.2 Background .................................................................................. 26
       3.2.1 Distance from Identity transform ........................................... 26
       3.2.2 Test of Normality ................................................................ 27
       3.2.3 Cramer von Mises Test ......................................................... 28
   3.3 Methods ...................................................................................... 31
   3.4 Results and Discussion .................................................................. 32
       3.4.1 Normality Tests .................................................................. 32
   3.5 Conclusion ................................................................................... 34

4 Distribution Independent Multivariate Statistical Testing for DBM 39
   4.1 Introduction .................................................................................. 39
   4.2 Background .................................................................................. 40
       4.2.1 An illustrative example for the JMTS test .............................. 41
   4.3 Methods ....................................................................................... 44
   4.4 Results and Discussion .................................................................. 46
   4.5 Conclusion ................................................................................... 47
References ............................................................................................. 48
List of Figures

1.1 p-Value map overlayed on structural image - Aging Data .............................. 8
1.2 p-Value map overlayed on structural image - Oasis Data .............................. 9

2.1 DBM results using combined morphometric evidence – belief maps overlayed on a structural image. None = No Change. Exp = Expansion. Cont = Contraction. ........ 16
2.2 DBM results using diffeomorphic demons – belief maps overlayed on a structural image. None = No Change. Exp = Expansion. Cont = Contraction. ............ 17
2.3 DBM results using curvature registration method – belief maps overlayed on a structural image. None = No Change. Exp = Expansion. Cont = Contraction. ............ 18
2.4 DBM results using HDW – belief maps overlayed on a structural image. None = No Change. Exp = Expansion. Cont = Contraction. ............................... 19
2.5 DBM results using combined morphometric evidence for Oasis dataset – belief maps overlayed on a structural image. None = No Change. Exp = Expansion. Cont = Contraction. ............................... 20
2.6 DBM results using DD for Oasis dataset – belief maps overlayed on a structural image. None = No Change. Exp = Expansion. Cont = Contraction. ............... 21
2.7 DBM results using curvature registration for Oasis dataset – belief maps overlayed on a structural image. None = No Change. Exp = Expansion. Cont = Contraction. .......... 22
2.8 DBM results using HDW for Oasis dataset – belief maps overlayed on a structural image. None = No Change. Exp = Expansion. Cont = Contraction. ............... 23

3.1 Power comparison for Hotelling’s $T^2$ (blue line) and CvM (purple line) tests .......... 30
3.2 Results of Hotelling’s $T^2$ test – Aging Data set ................................. 35
3.3 Results of CvM test – Aging Data set .................................................. 36
3.4 Results of Hotelling’s $T^2$ test – Oasis Data set .................................. 37
3.5 Results of CvM test – Oasis Data set .................................................. 38

4.1 Simplexes formed with 2-D data ......................................................... 42
4.2 Data used for simulation study of JMTS $T$ statistic ........................... 44
4.3 Results of JMTS $T$ test – Aging Data set ......................................... 46
Chapter 1

Influence of Registration Mathematics on Deformation Based Morphometry

1.1 Introduction

Computational anatomy (CA) involves a quantitative, non-invasive study of brain anatomy involving individual or populations of subjects. The method uses high resolution imaging techniques such as MRI or CT scans to obtain structural or functional images of the brains being studied. The changes in these images across time or across subjects is quantified and analyzed using complex mathematical algorithms.

CA allows us to perform non-invasive study of the human brain which was previously possible only with histological studies. One of the most popular applications of CA is the study of population wide structural brain changes caused by disease or treatment. The time scale at which structural changes occur makes them robust indicators of effects of disease or treatment [1]. Such population wide studies allows us to better understand the effect of disease, drugs, etc. on brain structure.

Deformation based morphometry (DBM), is a CA technique, used to study voxel-wise, whole-brain structural changes between groups of subjects [2]. In a typical DBM study, high resolution MR images of subject brains are registered to a template using an affine transformation followed by a non-linear registration method. The local structural changes between the subjects and the template is captured in their deformation vector fields (DVF). The resulting DVF or some derived scalar/vector morphometric metrics are then analyzed to deduce local, structural changes in the brain between the groups being studied.

A number of studies that use DBM to study structural brain changes, caused by various factors, are
being reported. These studies use a wide variety of non-linear registration methods and a slew of statistical tools to quantify changes in brain structure. Cardenas et al. \[3\] and Rohlfing et al. \[4\] have studied alcohol related brain changes using DBM with a normalized mutual information (NMI) based B-Spline free form registration method (proposed by Studholme et al. \[5\]). Boardman et al. \[6\] have studied changes in pre-term infant brains using the same registration method. Registration methods included in the SPM toolkit is another popular choice for DBM studies. Examples of those are: a schizophrenia study by Volz et al \[7\] and a study on brain atrophy caused by frontotemporal dementia by Brambati et al. \[8\]. In a study to quantify age related changes in the brain, Pieperhoff et al. \[9\] have used an in-house implementation of an elastic registration method (Homke et al.) and Riddle et al. \[10\] have used demons based registration whose implementation is freely available through the Insight Toolkit.

With the lack of an established physical model for brain deformation, multiple solutions exist for any particular deformation problem. This also implies that a Gold standard for registration can not be established. Using a number of different theoretic models, constraints and properties, there are many comparable and consistent registration methods in published literature \[11\] \[12\] \[13\] \[14\] \[15\] \[16\] \[17\]. The availability of inexpensive computational power has resulted in higher dimensional DVF resulting in improved matching between the source and target images.

With the availability of a number of registration algorithms, researchers using DBM analysis now have a wide range of registration methods to choose from. However, variations in the theoretical framework of a registration method influences the ability of the method to accurately characterize the changes between the source and target images. The mathematical properties of a registration method can increase it sensitivity to certain types of changes while making it less sensitive to some other types of changes. This is compounded by the absence of ground truth and hence the lack of a systematic procedure for evaluating the suitability of any registration method for a particular DBM application. Discrepancy in the deformation vector fields is also propagated as variations in morphometric metrics computed from each registration method \[18\]. Hence it is expected that for the same dataset, the DBM results should vary with the choice of registration method. The lack of a physical model for brain deformation and ground truth for DBM results has made it infeasible study the effect of registration methods on DBM results using actual data. Yet, most published DBM studies have not justified their choice of registration method and there is sufficient cause to believe that choice of registration method is primarily driven by availability of implementation.

In this chapter, we hypothesize and prove that DBM results are very sensitive to DVM characterization and parameters. We compare and contrast DBM results among three registration methods.
Two of the three methods, diffeomorphic demons (DD) [13] and DARTEL [19], are theoretically similar. The availability of theoretically similar algorithms provides us with an unprecedented opportunity to study the effect of DVF properties on DBM results. All other factors (images, pre-processing, morphometric metric (log euclidean strain vectors) and statistical testing) being the same, it would be expected that DBM result from DD and DARTEL should be exceedingly similar if not identical. In spite of the theoretic similarity of the registration methods, the DBM results are largely inconsistent. Any discrepancy in this situation can be attributed to the effect of the variations in the DVF generated by the various registration methods. The third registration method, High dimension warping (HDW) [16], follows the same variational framework as the other two methods but with vastly assumptions and constraints in characterizing the DVF. The variations in the principles of the three methods is used to illustrate the conflict in DBM analysis induced only by changing the registration method. In the absence of a systematic method to choose one registration method over another, arbitrary choice of only one registration method presents an incomplete picture. Since the effect of DVF properties are vary based the nature of data being studied, we compare the performance of the three methods using two typical DBM applications: (i) detecting structural changes caused by healthy aging, and (ii) detecting structural changes caused by dementia.

1.2 Background

1.2.1 Registration Methods

DD and DARTEL

Both DD and DARTEL methods model registration as a diffusion process from the source to the target where the DVF are constrained to a space of diffeomorphisms. DARTEL is based on the LDDMM [20] registration framework and DD is considered to be a constrained optimization case of the Demons algorithm [12]. For the purposes of practicality, both methods constrain the DVF to be generated from the subgroup of one parameter stationary vector fields. Both methods use the framework with similar image matching metrics, similar properties and constraints on the DVF. Both registration methods start at the identity DVF and iteratively search for the optimal DVF in a space constrained to diffeomorphisms. The search for the update fields are performed as a levenberg-marquad optimization routine. In DARTEL the optimization is performed in the Hilbert space while DD performs optimization on Lie groups because the set diffeomorphic DVF do form a linear vector space. The methods include a regularization condition that enforces smoothness
constraints and ensures that the updated field at each iteration is a diffeomorphism. In DARTEL, the regularization constraint is a function of the update field generator and is related to the physical model imposed on the deformation process. In contrast, DD realizes regularization by imposing a Gaussian smoothing on the DVF and the update at each iteration. Another important difference between the two methods in their implementation of symmetric constraint. In DARTEL the symmetric constraint results in an invertible DVF between the target image and the warped source image. In DD, only the update DVF at each iteration is constrained to be symmetric and the final DVF matching the source to the target does not have to be symmetric. It has been shown that the two algorithms are theoretically similar, have similar image matching performance [21]. DD has a lower computational cost compared to DARTEL but yields less smooth DVF.

**HDW**

This registration method is an implementation of the inverse-consistent registration method proposed by Ashburner et al. [16]. HDW is related to the other two methods in that the space of inverse-consistent DVF is a subset of the space of diffeomorphic transforms. Also, it used the same iterative optimization framework that uses a gradient descent method. But the methods differ in how they optimal DVF is computed for any target-source pair. While most registration methods use a physical process to obtain the DVF, HDW uses a bayesian MAP method to estimate the optimal DVF for registration. Constraints are enforced using priors which determine how likely any update DVF is and penalizes unlikely deformations.

**1.2.2 Log-Euclidean Strain Vectors (LeSV)**

DBM analysis is on some morphometric metric computed from the DVF obtained as a result of registration. As explained in Chapter ??, we have used the LeSV which allows us to measure both volume and directional changes at each voxel.

**1.2.3 Multivariate Statistics**

Group differences were analyzed by applying per-voxel Hotelling’s two-sample $T^2$ test on the set of Log-Euclidean strain vector maps. A voxel-wise testing procedure results in a huge multiple testing problem. Any significance tests done on the P-value maps obtained will have to be compensated for multiple comparisons.

In order to be able to visualize the significance maps of the tests, their False Discovery Rates
(FDR) were controlled at 0.05 using the Benjamini-Hochberg procedure [22]. In this procedure the tolerable proportion $\alpha$ of false positives (among the total number of significant voxels discovered) is fixed and the critical p-value ($p_{\text{crit}}$) is calculated. The procedure to calculate is as follows:

1. Arrange all $m$ p-values in ascending order such that $P(1) \leq P(2) \leq \ldots \leq P(m)$

2. The largest $P(k)$ that satisfies the inequality $P(k) \leq \frac{k}{m} \alpha$ is determined. This value of $P(k) = P_{\text{crit}}$

The p-value maps are thresholded at $P_{\text{crit}}$ to obtain significance maps in which the total number of false positives are controlled at $\alpha \times 100\%$ on average.

1.3 Methods

1.3.1 Registration

As the first step, all images were skull-stripped and affinely (12 degrees of freedom (DOF)) registered to an arbitrary target from the control group. Affine registration can be done using FSL tool FLIRT (version 5.0) [23]. The initial affine registration removes global differences among the subjects and increases the focus of the non-linear registration methods on local, anatomical changes only.

The affinely registered subject images were registered to the target using six different non-linear registration method yielding 3 sets of registration results per subject. We have described below the parameters to be used for each of the registration methods used. The choice of an appropriate target has an effect on the outcome of TBM studies. When picking a target from the control group it must be verified that the chosen subject is not an outlier in the group and is a fair representative of the subjects in that group. Lepore et al [24] have proposed a method to choose a template for TBM studies. While it is not required to choose the best template, the suitability of the chosen target can be efficiently evaluated through visual inspection.

DD: An implementation of the diffeomorphic demons is available as a part of the from the Insight Journal - 2007 MICCAI Open Science Workshop 1. The user controlled parameters of the implementation were set to: diffeomorphic update rule with symmeterized gradient computation. The default values of the smoothing sigma for both the update and deformation field provided the

\footnote{1http://hdl.handle.net/1926/510}
best registration for the two datasets (evaluated through visual inspection) and hence we chose to use these values for registration.

**DARTEL:** An open-source implementation of DARTEL is available with the SPM package [25]. In order to be used with DARTEL, the fixed and moving images must have identical dimensions, isotropic voxels and rigidly registered. DARTEL can be used in only two ways: to create tissue templates from a population or to warp a population of images to an already existing template. In order to be used in this study, a template was created using the target image alone. The subjects images were then aligned to this template.

**HDW:** The SPM5 package, which includes an implementation of HDW, suggests that HDW is best used for intra-subject registration because the method assumes that there are no intensity variations between the source and target. By modifying the default parameters we were able to optimize the tool for inter-subject registration with satisfactory results. The number of iterations for the bias correction step was increased to 10. The number of iteration for the warping was increased to 12 while the regularization value was decreased to 2 such that smoothness was sacrificed for higher accuracy. A batch operation script was written to warp each of the affinely registered subject images to the target.

A four-level multiresolution registration framework was used for the three methods. The pyramid scheme which downsamples (and smooths) the fixed and moving images by a factor of two at each level. The number of iterations at each level were kept constant for all the three methods. In a typical DBM study, the users have limited control over the parametrization of the DVF for any registration method. We have therefore restricted out study to only those registration parameters that are accessible to users.

The C++ implementation of the LeSV, the statistical and permutation tests were based on the ITK toolkit [26]. The implementation of all the tools used in this chapter will soon be made available for public download from our lab website [http://www.bsl.ece.vt.edu](http://www.bsl.ece.vt.edu).

### 1.4 Results and Discussion

#### 1.4.1 Aging Dataset

Figure 1.1 (a-c) show the result of DBM analysis on the three registration methods. We notice that the DD and DARTEL (b-c) detect similar patterns of significant p-values though not identical
ones. While both methods detect changes in the regions surrounding the thalamus, ventricles and the corpus callosum, they differ in the extent of change and the exact regions. DD studies show that the entire thalamus has shown significant changes and genu of the corpus callosum while results using the DARTEL method yield a large area of significant changes in the ventricles that is absent in DD. The main characteristic exhibited in this region is the diffused nature of changes in DARTEL while DD shows changes with very low p-values concentrated at the site of change. This can be explained by the difference in regularization in both methods. We discussed earlier that regularization constraint is better enforced in DARTEL than in DD i.e the DVF from DARTEL show more smoothness than those from DD. This explains why significant regions shown in Figure 1.1 (c) have regions of low p-value in the center with increasing p-values as we move away from the center. Another effect of the regularization constraint is the significant changes detect outside the brain. In order to satisfy the smoothness constrain, the changes occurring inside the brain could be diffused throughout the DVF. DBM methods declare significant those regions near the edges of the images where the changes are limited by boundary conditions. While these voxels outside the brain are difficult to interpret, it would not be wise to completely ignore all of them. The voxels changes outside the brain region, under careful analysis, could provide limited but additional information on the direction of the changes in the brain. For example, 1.1 (b) & (c) show discrepancies outside the frontal lobe that are directly collinear to the changes detected in the ventricular regions of the brain. This suggests the DVF have been pulled inward near the ventricles and corpus callosum in the aged dataset. Since p-values obtained from DBM studies provide information on where changes occurred, carefully study of the pattern of changes both inside and out the brain allows us to gain information on the nature of those changes. Both methods detect changes in the occipital lobe of the brain between the two groups but the exact location of change is inconsistent. In Figure 1.1 (b) the changes in the occipital lobe are distributed and reaching in the parietal lobe. It also shows some changes in the cerebellum that are not seen in Figure 1.1 (c).

Figure 1.1 (a) shows changes detected when using HDW registration method. The large discrepancy in the result compared to 1.1 (b & c) shows that any changes in the constraints and assumptions of the registration methods yields completely different DBM results. If a study included only results Figure 1.1 (a), it would concluded that there are no major changes between the two groups in the aging data set. Earlier studies and Figures 1.1 (b & c) have shown us that this is not the case. We hypothesize the inverse-consistency constraints inhibits the ability of the algorithm to characterize changes on one subject with reference to the template. While inverse-consistency maybe a good property to include for many applications, it does not fit well into the framework of DBM.
Another interesting feature about the results from HDW is the total lack of significant voxels outside the brain. This is because, the SPM implementation of HDW forces voxels outside the brain to have a prior of 0 which sets the deformation at those voxels to 0. The bayesian framework of this registration method, makes it possible to incorporate such a constraint on the registration method. A good understanding of the constraints used in a registration method is necessary to predict its implications on the DVF and hence the DBM analysis. Persons interpreting the results of DBM studies should consider carefully the properties of the registration methods used before arriving at any conclusions about test results.

![Image of brain with p-values overlayed](image)

Figure 1.1: p-Value map overlayed on structural image - Aging Data

### 1.4.2 Oasis Dataset

Figure 1.2 (a-c) show the results of DBM analysis for the Oasis dataset. All three methods are in agreement about the changes seen in the anterior aspect of the lateral ventricle. Comparing
the results from DD and DARTEL in Figure 1.2 (b-c) respectively, we see that the significant area around the splenium of the corpus callosum (CC) is missing in the result from DARTEL. Also both methods show small islands of significant voxels all over the image with no discernible pattern. It should also be noted that the small islands do no occur on corresponding locations in Figures 1.2 (b & c) and cannot be interpreted as cortical changes. All these islands are caused by regularization of the deformation corresponding to the changes along the CC. For example, the 1.2 b, the large area of significant voxels in the inferior end of the frontal pole corresponds to “edge effect” cause by regularizing changes in the genu. In both figures 1.2 (b & c), based on the pattern of outward radiation of the significant voxels, we conclude that changes centered at the posterior aspect of the lateral ventricle are caused by dilation of the ventricle.

*Figure 1.2: p-Value map overlayed on structural image - Oasis Data*
1.4.3 Comparison of properties across datasets

All three registration methods studied here are shown to have consistent registration performance. But their performance for DBM will vary based on the registration problem and the ability of the method to find an optimal DVF within its constrained search space. This implies that the virtues and vices of any one registration method will not translate directly across applications. To illustrate this, we compare Figures 1.1 (a) and 1.2 (a). The HDW was unable to delineate any of the changes between the two groups in the aging study but for the oasis dataset, HDW was much in agreement with the other methods. The physical process chosen to generate a DVF and the associated constraints will effect its DBM performance differently for different datasets. When one registration method is not sensitive for a particular application, it would be unwise to conclude that it will not be sensitive to changes of interest in another application.

1.5 Conclusion

The objective of this study is not to prove the superiority of any registration algorithm over another. We would like to illustrate that the choice of registration algorithm and associated constraints greatly influence the results of group difference studies. In our opinion, studies that use different registration methods cannot be reliably compared or contrasted. In the absence of a GOLD Standard, there is no means to suggest which algorithm is most suitable for any studies using DBM. All changes seen at the end of DBM study do not always correspond to actual changes in the structure. It is of vital importance that person/persons interpreting any DBM results be completely aware of the mechanics of the registration methods. Prior knowledge about the changes between two groups will help users discard those methods whose mechanics inhibit accurate characterization of the expected changes. If no prior knowledge is available, we suggest that users study results from multiple registration methods before finalizing any conclusions. For these reasons, we encourage very cautious use of off-the-shelf DBM software.
Chapter 2

Combining Morphometric Evidence from Multiple Registration Methods using Dempster-Shafer Theory

2.1 Introduction

We have seen in Chapter 1, that DBM results are influenced by the mathematical model, properties and constraints of the registration method used in the study. Chapter 1 also shows that, a number of registration methods available perform equally in terms of registration accuracy. The mathematical properties of various registration methods greatly influence how anatomical changes are characterized in their deformation transforms. In the absence of a justified physical model for brain deformation there exists no objective procedure for choosing one registration method over another for any application. In such a situation, using information from multiple registration methods guarantees a more complete study of the dataset than using a single method alone.

In this chapter we propose a method to integrate the information from multiple registration methods using the Dempster-Shafer evidence theory. By treating the various registration methods as independent, imprecise and conflicting information sources our method acts as an information fusion tool.

One of the main drawbacks of using standard statistical tests is the choice of decision-making threshold (p-value threshold) for these tests is always arbitrary [27]. While many authors have taken to reporting the actual statistical values, these values are prone to misinterpretation because of their complex relationship to the actual likelihood of change. Therefore this tool represents change as belief maps, reducing the ambiguity and subjectivity in interpreting results from a study and allows for easy incorporation of prior knowledge.
2.2 Background – Dempster-Shafer Theory

Dempster-Shafer theory (DST) [28] is a generalization of Bayesian theory used to combine conflicting pieces of information. In this chapter, we use the DST as a means of information fusion. In DST, Θ, called the frame of discernment, is the mutually exclusive and exhaustive set of all possible outcomes/events. The power set, \( \Omega(\Theta) \), is the set of all possible subsets of \( \Theta \). The probability assignment function for \( \Omega(\Theta) \), \( m : \Omega(\Theta) \to [0, 1] \) uses available evidence to map the power set to a value between 0 and 1. For any subset of the power set \( A \), \( m(A) \) is the support that any particular element of \( \Theta \) belongs to \( A \) but not any other subset of \( A \) [29] and

\[
\sum_{A \subseteq \Omega(\Theta)} m(A) = 1; 
\]  
(2.1)

The support for any event/outcome is expressed as an interval representing the actual probability of that event. The lower and upper limit of the support are known as Belief and Plausibility respectively and are defined as

\[
Bel(A) = \sum_{B \subseteq A} m(B) \quad \text{and} \quad Pls(A) = \sum_{B \cap A \neq \phi} m(B). 
\]  
(2.2)

**Modified Rule of Combination** [30] : For the same \( \Theta \), multiple mass functions \( m_1, m_2, \ldots, m_N \) (computed from multiple sources of information) can be combined by using the following modified combination rule:

\[
m(A) = (m_1 \oplus m_2 \oplus \ldots \oplus m_N)(A) = \frac{\sum_{\cap_{i=1}^n E_i = A} m_1(E_1)m_2(E_2)\ldots m_N(E_n)}{K}, \]

where

\[
K = \sum_{\cap_{i=1}^n E_i = \phi} m_1(E_1)m_2(E_2)\ldots m_N(E_n). \]  
(2.4)

In DBM, the state of a subject with respect to the reference image is quantified using specific metrics calculated from the deformation fields. We use DST to fuse this information from these registration methods and compute the belief functions for the various states given the data sets.

2.3 Methods

To illustrate the use of DST for DBM we have conducted two experiments typical of previous DBM applications. The objective of this experiment is to detect structural differences between two groups of human brains. Local volume change between the two groups is measured using Jacobian
determinant maps. Using DST to combine information from the various registration methods, we calculate the belief of voxel-wise change between the two groups.

The following preprocessing steps were applied. All images were skull-stripped using BET [31] and affinely (12 degrees of freedom (DOF)) registered to an arbitrary target from the control group using the FSL tool FLIRT (version 5.0) [23]. The initial affine registration removes global differences among the subjects and increases the focus of the non-linear registration methods on local anatomical changes only. While there are methods to choose the most appropriate target from the control group [24] or to build an unbiased template for registration, we arbitrarily picked a target from the control group.

The affinely registered subject images were then registered to the target using six different non-linear registration methods yielding six sets of registration results per subject. In this study, we have only included only those registration methods whose implementations are freely available to the general community:

- Four registration methods that are implemented in Insight Segmentation and Registration Toolkit (ITK) [26]: Demons registration [11], Symmetric Demons [12], diffeomorphic demons with symmetric constraint [13] and the curvature registration method [14].
- Two registration methods included in the SPM5 toolkit [25]: High Dimension Warping and DARTEL

We have not included the ITK implementations of level-set registration [32], fast block matching [33], and B-spline registration, nor HAMMER [17] because these methods could be optimized to yield visually comparable registration accuracy for our particular data set.

The deformation fields were smoothed and then used to compute voxel-wise Jacobian determinant values (using the appropriate ITK [26] filter) yielding Jacobian determinant maps. This resulted in 6 Jacobian determinant values at each voxel for each subject. In order to ensure that Jacobian determinant values were always positive, we computed the determinant from the positive definite Jacobian matrix \((J^TJ)^{\frac{1}{2}}\) at each voxel.

### 2.3.1 Combining Morphometric Information using DST

At each voxel, we assume that we receive information (Jacobian Determinant value) from \(N\) independent sources (registration methods). The morphometric evidence of each group is combined
independently. Assume there are $M$ subjects in one of the groups. Our frame of discernment is defined as $\Theta = \{\text{No Change (N)}, \text{Expansion (E)}, \text{Contraction (C)}\}$. Since the options are mutually exclusive, $\Omega(\Theta)$ is the same as $\Theta$. If $J_{ij}^k$ is the Jacobian determinant at the $i^{th}$ voxel of the $j^{th}$ subject from $k^{th}$ registration method, we know that geometrically if $J_{ij}^k = 1$ then the $i^{th}$ voxel has not changed with respect to the target. If $J_{ij}^k < 1$ then there is local contraction and $J_{ij}^k > 1$ indicates local expansion. Any small variations in $J_{ij}^k$ from 1 are most likely caused by noise than by actual changes between images. To compensate for effect if noise we incorporate a noise limiting factor $\varepsilon$ in the construction of the frequency matrix

$$H_i = \begin{bmatrix} x_{\text{no}}^1 & x_{\text{no}}^2 & \ldots & x_{\text{no}}^N \\ x_{\text{exp}}^1 & x_{\text{exp}}^2 & \ldots & x_{\text{exp}}^N \\ x_{\text{con}}^1 & x_{\text{con}}^2 & \ldots & x_{\text{con}}^N \end{bmatrix}$$

where the $\text{no}$ indicates the no change category, $\text{exp}$ indicates the expansion category and $\text{con}$ indicates the contraction category. Each entry is given by

$$x_{\text{no}}^k = \frac{|1 - \varepsilon < J_{ij}^k \geq 1 + \varepsilon|}{M} \quad x_{\text{exp}}^k = \frac{|J_{ij}^k > 1 + \varepsilon|}{M} \quad x_{\text{con}}^k = \frac{|J_{ij}^k < 1 - \varepsilon|}{M},$$

where $||$ indicates cardinality. For example, $x_{\text{no}}^k$ is the frequency across the group where the determinant from the $k^{th}$ registration method was one (noise limited by $\varepsilon$ which we set to 0.0001), indicating no change. The membership vector consolidates evidence supporting the state of the $i^{th}$ voxel given $M$ images of one group. The membership vector is then

$$\hat{m}_i = [\hat{m}_{\text{no}}^i \quad \hat{m}_{\text{exp}}^i \quad \hat{m}_{\text{con}}^i]$$

where $\hat{m}_{\text{no}}^i = x_{\text{no}}^1 \times x_{\text{no}}^2 \times \ldots \times x_{\text{no}}^N$, $\hat{m}_{\text{exp}}^i = x_{\text{exp}}^1 \times x_{\text{exp}}^2 \times \ldots \times x_{\text{exp}}^N$, $\hat{m}_{\text{con}}^i = x_{\text{con}}^1 \times x_{\text{con}}^2 \times \ldots \times x_{\text{con}}^N$

We then apply the modified combination rule to create the mass function vector at this voxel.

$$m_i = \left[ \begin{array}{c} \frac{\hat{m}_{\text{no}}^i}{\hat{m}_{\text{no}}^i + \hat{m}_{\text{exp}}^i + \hat{m}_{\text{con}}^i} \\ \frac{\hat{m}_{\text{exp}}^i}{\hat{m}_{\text{no}}^i + \hat{m}_{\text{exp}}^i + \hat{m}_{\text{con}}^i} \\ \frac{\hat{m}_{\text{con}}^i}{\hat{m}_{\text{no}}^i + \hat{m}_{\text{exp}}^i + \hat{m}_{\text{con}}^i} \end{array} \right]$$

For this particular $\Theta$, belief and plausibility of an event/outcome are equal to the mass function. The mass function image is computed for each of the two groups to be compared.

The objective of our study was to detect differences between a group of control subjects and a group of treated subjects. With three distinct events in the frame of discernment, we can build a 3x3 confusion matrix for the various outcomes from comparing the two groups. Let $m_i^C = \begin{bmatrix} m_{\text{no}}^C & m_{\text{exp}}^C & m_{\text{con}}^C \end{bmatrix}$ be the mass function vector at the $i^{th}$ voxel for the control group and $m_i^T = \begin{bmatrix} m_{\text{no}}^T & m_{\text{exp}}^T & m_{\text{con}}^T \end{bmatrix}$ be the corresponding vector for the treated group. The confusion matrix is defined as
The diagonal elements of this matrix correspond to no change in state between the control and treated groups. The other elements of the confusion matrix gives belief and plausibility of change from one state to another between the control and treated group for all combinations.

We validate our method using results from two datasets:

- Structural difference caused by healthy aging in humans
- Structural difference caused by demential

### 2.4 Results and Discussion

Figure 2.1 shows the results of group difference testing using combined morphometric evidence. Each image corresponds to belief in change from one state to another between the control and treated groups. The belief maps have been thresholded at 0.1 and overlaid on an axial slice to help visualize the changes. When evidence from all the registration methods are in agreement, the belief in the results of that region increases.

Figure 2.1(a,b,c,e) shows that there is little evidence for no change in either group. Figure 2.1(d) shows that there is a small amount of evidence for expansion in the control group and contraction in the treated group. The ventricular region of Figure 2.1(f) indicates regions where the belief in contraction in the control group and expansion in the treated group is close to 1, consistent with other findings in the aging literature that show ventricular enlargement.

In order to illustrate the need to combine information from various registration methods, Figures 2.2, 2.3, 2.4 shows individual (per registration method) DST-DBM results from three of the methods included. It is evident that the methods do not present identical results, and in the absence of ground truth, no conclusions can be made about the reliability of the individual methods.

We repeated the experiment with another dataset. The objective of this experiment was to identify regional volume changes caused by dementia. Figure 2.5 shows the results of DBM using DST
Figure 2.1: DBM results using combined morphometric evidence – belief maps overlaid on a structural image. None = No Change. Exp = Expansion. Cont = Contraction.

for the Oasis dataset. Figures 2.7, 2.6 and 2.8 show inconsistent results for the same dataset. We would like to point out the large amount of conflict among the method with respect to the expansion-expansion state in the confusion matrix (image (e) in Figures 2.6-2.8). The use of DST allows us to use information from all the available sources and arrive at a decision based on the total belief for that state. We notice that the no change category usually has 0 belief because it is unlikely that the Jacobian determinant values at any voxel will stay at 1 after registration. This is also a proof of significant intra-group variability.

Another advantage of this method, over conventional DBM analysis, is that information regarding similarity between the two groups is also available. For example, in Figure 2.5 e and i, we notice that for much of the image there is no change of state between the two groups. This allows us to infer that the two groups are largely similar except for certain areas where differences occur. At this point, we would like to caution the reader about sensitivity of this method to the thresholds chosen to differentiate between two states. Another caveat of using this method is that, the metric chosen for this method should allow the system to be categorized into distinct states.
In Chapter 1, we established that studies of anatomical group variances that use DBM with only one registration are incomplete and are subject to biased results based on the mathematics of the chosen registration method. We present in this chapter, an alternative approach to DBM using Dempster-Shafer theory to fuse information from different registration methods. Using information from various multiple methods minimizes the appearance of certain “significant changes” that are in actuality artifacts of the registration mathematics.

Another advantage of this method over traditional DBM is its representation of results as belief maps of where change has occurred in all possible combinations between the control and treated group. This method provides a more intuitive visualization of results compared to p-value change maps. The representation of results as a confusion matrix allows the user to understand not only where changes occur between the two groups but also what kind of changes have occurred. Another added advantage is the ability to visualize regions of similarity between the two groups without the need for any additional computation.

Figure 2.2: DBM results using diffeomorphic demons – belief maps overlaid on a structural image. None = No Change. Exp = Expansion. Cont = Contraction.

2.5 Conclusion
Our results show that this DS approach gives results that are superior to using single registration methods and traditional DBM. The DBM-DS framework allows us to minimize bias introduced by arbitrary choice of a particular registration method and presents only those results which exhibit “truth by consensus”.

*Figure 2.3*: DBM results using curvature registration method – belief maps overlaid on a structural image. None = No Change. Exp = Expansion. Cont = Contraction.
Figure 2.4: DBM results using HDW – belief maps overlaid on a structural image. None = No Change. Exp = Expansion. Cont = Contraction.
Figure 2.5: DBM results using combined morphometric evidence for Oasis dataset – belief maps overlaid on a structural image. None = No Change. Exp = Expansion. Cont = Contraction.
Figure 2.6: DBM results using DD for Oasis dataset – belief maps overlaid on a structural image. None = No Change. Exp = Expansion. Cont = Contraction.
Figure 2.7: DBM results using curvature registration for Oasis dataset – belief maps overlaid on a structural image. None = No Change. Exp = Expansion. Cont = Contraction.
Figure 2.8: DBM results using HDW for Oasis dataset – belief maps overlaid on a structural image. None = No Change. Exp = Expansion. Cont = Contraction.
Chapter 3

Distribution Independent Statistical Testing for Deformation Based Morphometry

3.1 Introduction

In Chapter 1, we studied the effect of the registration algorithms on DBM analysis. The function of the registration method acts as a tool to quantitatively represent variations between each subject and the template image. Once quantified, it is necessary to sift population wide changes from individual subject variations. Using data from the registration methods, statistical hypothesis perform tests to check where (in the image) statistically significant differences occur between the two groups. Researchers use a number of statistical tests for this purpose such as the t-test [34], ANOVA [35], ANCOVA [35] [36], MANCOVA [37] and the Hotelling’s $T^2$ test [38] [39].

The parametric statistical tests used above, which are variations of the student’s t-test, are not a reasonable choice for DBM because of the following:

Scope of Tests: When comparing two groups of data using the tests mentioned, the null hypothesis is that the mean of both data sets is the same. It is necessary to consider if this test provides us enough information to characterize the changes in the data. Using only one parameter to characterize the differences between DBM metrics is insufficient and provides incomplete testing. This leads to our next reason regarding various other assumptions made by the test.

Violating assumptions of the tests: The family of student t-tests assume that data from the two groups being tested are:
1. normal or Gaussian
2. homoscedastic (having the same variance)

Published DBM literature do not explicit mention normality tests. We believe they follow the conjecture: “normal unless proven otherwise”. The sample size of DBM data fall under the small sample category \( n \leq 30 \) making normality testing unreliable. At this sample size, data would most likely pass standard normality tests but the results would, at best, be inconclusive. In these cases, failing the normality test is very serious. Since DBM is a voxel-wise test, all voxels will have to pass the normality test for the family of \( t \)-test to be valid. The same difficulty applies to testing for homoscedastic. It can be said that it is very unlikely for data from the two groups being tested will be homoscedastic. Violating the fundamental assumptions of the hypothesis tests presents a whole new set of challenges. Further, the study of the effects of erroneous distributional assumption on DBM results is hampered by the unavailability of ground truth. The consequence of violating the assumptions also depends on the extent of violation. In a voxel-wise, this effect could vary spatially this making an estimate of the effect complicated. The effect of violation of the fundamental assumptions of the statistical tests can vary based on the data set. All these factors together make the assumption of normality an unreasonable one for DBM.

**Reasonability if normality assumption:** Given the data used in the test, it is necessary to examine if normality is a reasonable assumption irrespective of the results of the normality tests. For example, the Jacobian determinant, a popular metric in DBM, is constrained to be non-negative. Even if the data passes normality tests, it is not reasonable to expect the data to be normally distributed.

The alternative to parametric hypothesis tests are non-parametric, distribution independent statistical tests that make fewer assumptions about the distribution of the data. Non-parametric tests have wider applicability and yield more robust statistics because of the few assumptions. Non-parametric statistics are simpler to use in that they require less to be known about the data being tested. This can be in the form of prior data or prior testing on the data to validate assumptions. This increase in flexibility makes non-parametric tests more conservative in detecting differences when compared to parametric. For DBM, conservative results are preferred over unreliable results. The voxel-wise testing framework of DBM, benefits from the distribution-free parametric results. Results of DBM influence critical process like drug or treatment choices for various conditions. Under such circumstances, it would unwise to use parametric methods where the fundamental as-
sumption are bound to be violated. The absence of ground truth means we cannot assess the effect of these violations or quantify the error in the results. Under such circumstances, using a statistical test with fewer assumptions provides a more flexible framework for DBM.

In this chapter we show, using two different data sets, that DBM data cannot be reliably declared as coming from normal distributions. Using the Hotelling’s $T^2$ test, we perform DBM on two theoretically equivalent registration methods for the non-Gaussian data. When the registration parameter is constant, the DBM results are affected only by the nature of the statistical test used. We show that violating the fundamental assumptions of the statistical tests causes theoretically similar registration methods to show largely conflicting DBM results. To remedy this, we propose the use of Cramer von Mises (CvM) two sample test to detect inter group variability in DBM. The CvM test is a popular distribution independent test of inter-group variability for univariate samples. The CvM test compares two sets of data by measuring the integral of the squared distance between the probability distributions of the two data sets. This allows us to test for difference between two distributions as opposed to testing for one parameter in the distribution. The final result of the CvM test based DBM shows minimal conflict between the registration methods used. The obvious conservative nature of the CvM test over the Hotelling’s $T^2$ test is also evident in the results.

3.2 Background

3.2.1 Distance from Identity transform

In DBM, the jacobian determinant of the deformation transforms are used to study local changes in volume between the control and treated group. The obvious drawback of this metric is that it is not sensitive to volume preserving shape changes. An example of this has been illustrated in Lepore et al. [39]. To remedy this, we use the deformation strain matrix $((J^T J)^{0.5})$ of a voxel where $J$ is the jacobian matrix of the deformation vector at a voxel. Since strain vectors do not form a linear vector space, we transform them to the ‘Log Euclidean space’ proposed by Arsigny et al. [40]. The transformed Log euclidean strain vectors (LeSv) form a linear vector space thus allowing euclidean operations to performed on them. We compute the distance between the strain vector at a voxel and the identity strain vector in the log euclidean space to measure the magnitude of deformation.
3.2.2 Test of Normality

Normality tests allow us to determine if the data being studied is well-modelled by a normal distribution. Normality tests are goodness-of-fit tests which determine the normality of a given data set by measuring its departure from expected values. The D’Agostino’s omnibus $K^2$ statistic [41] is a goodness-of-fit statistic that uses transformation of the skewness and kurtosis of the hypothesized distribution of the dataset under test. Skewness is a measure of the symmetry of a distribution; kurtosis measure the degree of “peakedness” of the distribution. Since a normal distribution is symmetric it has a skewness of 0. Every normal distribution is mesokurtic with a kurtosis value of 0. The D’Agostino’s omnibus $K^2$ statistic, under the null hypothesis that the samples being tested are normal, is distributed as the $\chi^2$ distribution with two degrees of freedom.

Doornik et al [42] proposed a simple to implement, omnibus, univariate D’Agostino’s omnibus $K^2$ statistic. The test uses a transformation of the skewness and kurtosis so as make the $\chi^2$ pertinent for small samples.

\[ \sqrt{\beta_1} = \frac{\mu_3}{\mu_4^{3/2}} \]  
\[ \sqrt{\beta_2} = \frac{\mu_4}{\mu_4^{3/2}} \]  
\[ K^2 = \left( Z\left(\sqrt{\beta_1}\right) \right)^2 + (Z(b_2))^2 \approx \chi^2(2) \]

where $\sqrt{\beta_1}$ is the skewness which is a function of the third moment about the mean ($\mu_3$). $\sqrt{\beta_2}$ is the kurtosis which is a function of the fourth moment about the mean ($\mu_4$). $Z(\sqrt{\beta_1})$ and $Z(b_2)$ are the transformations of the skewness and kurtosis respectively. The transformation for the skewness is given in Equation 3.8 and transformation for the kurtosis is given in Equation 3.15

\[ \beta = \frac{(3n^2 - 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)} \]  
\[ \omega^2 = -1 + 2(\beta - 1)^{0.5} \]  
\[ \delta = \frac{1}{\log(\sqrt{\omega^2})^{0.5}} \]  
\[ y = \sqrt{b_1} \omega^2 - 1 \frac{(n+1)(n+3)}{2} \frac{1}{6(n-2)} \]  
\[ Z_1 = \delta logy + (y^2 + 1)^{0.5} \]
\[ \delta = (n+3)(n+1)(n^2 + 15n - 4) \]  
\[ a = \frac{(n-2)(n+5)(n+7)(n^2 + 27n - 70)}{6\delta} \]  
\[ c = \frac{(n-7)(n+5)(n+7)(n^2 + 2n - 5)}{6\delta} \]  
\[ k = \frac{(n+5)(n+7)(n^3 + 37n^2 + 11n - 313)}{12\delta} \]  
\[ \alpha = a + b_1c \]  
\[ \chi = (b_2 - 1 - b_1)2k \]  
\[ z_2 = [\left(\frac{\chi}{2\alpha}\right)^{0.5} - 1 + \frac{1}{9\alpha}]^{(9\alpha)^{0.5}} \]

This particular transformation has shown good power and size characteristics for sample sizes \( n \) as low as 10. The drawback of this method is that when a dataset fails the test, the omnibus statistic cannot specify the type of non-normality is not specified.

### 3.2.3 Cramer von Mises Test

The two sample Cramer von Mises tests measure the L-distance between two empirical, continuous distributions \( \mathcal{X} \) and \( \mathcal{Y} \). The distribution free property and the large power of this method makes it a popular choice for testing equality of two empirical distributions. The two sample generalization of the CvM goodness of fit test was proposed by Anderson [43]. Consider \( n \) samples from the empirical distribution \( \mathcal{X} = x_1, x_2, ..., x_n \) and \( m \) samples from the empirical distribution \( \mathcal{Y} = y_1, y_2, ..., y_m \). Based on the samples, the method tests the null hypothesis that \( \mathcal{X} = \mathcal{Y} \). The Cvm test statistics is given by:

\[ T^2 = \frac{mn}{m+n} \left\{ \sum_{i=1}^{n} (\mathcal{X}_n(x_i) - \mathcal{Y}_m(x_i))^2 + \sum_{j=1}^{m} (\mathcal{X}_n(y_j) - \mathcal{Y}_m(y_j))^2 \right\} \]

The null hypothesis is rejected for large values of the \( T^2 \) statistic. The distribution free property and the large power of this method makes it a popular choice for testing equality of two empirical distributions. The two sample generalization of the CvM goodness of fit test was proposed by Anderson [43]. The advantage of using this test was it robust performance for small to medium sample sizes [42] which are typical in DBM applications.
Power Comparison for CvM test and Hotelling’s $T^2$ test

The power of a statistical test is the probability that the test will reject the null hypothesis when the alternative hypothesis is true. This test the probability of a Type II error not occurring. Power of the statistical test is important for DBM because it measures the ability of the test to accurately detect differences between the two groups. Below, we compare the power of the CvM and the hotelling’s test for both gaussian and non-gaussian data.

In this test, we used R to compute the power of the CvM test and the hotelling’s $T^2$ test. The tests were performed as follows:

1. For each sample, we generated 20 samples for two distributions $X$ and $Y$.
2. Samples were generated as follows: $X$ and $Y$ are both normal distributions. $X = \mathcal{N}(0, 1)$ and $Y = \mathcal{N}(\mu_i, \sigma_1 = 1)$. The mean of $Y$ is varies from 0 to 4. The variance is held a constant.
3. Using the samples generated, the test statistics are computed for both the CvM and the hotelling’s $T^2$ test.
4. Steps 1-3 are repeated 10,000 times. The number of times when the null hypothesis gets rejected at a significance level of 0.05 is divided by 10 000 and plotted.

The experiment described above was repeated by varying step 2 as follows:

- As mentioned above, $X = \mathcal{N}(0, 1)$ and $Y = \mathcal{N}(\mu_Y, \sigma_Y = 1)$. The mean of $Y$ is varies from 0 to 4. The $\sigma_Y$ is held a constant.
- $X = \mathcal{N}(0, 1)$ and $Y = \mathcal{N}(\mu_Y, \sigma_Y = 1)$. The $\mu_Y$ is held constant at 0. The $\sigma_Y$ is varied from 1 to 5.
- $X$ and $Y$ are both lorenzian distributions. The scale of both distributions is held a constant at 1. The location parameter of of $X$ is fixed at 0.05 while that of $Y$ is varied from 0.05 to 3.
- $X$ and $Y$ are both gamma distributions. The scale of both distributions is held a constant at 1. The location parameter of of $X$ is fixed at 1 while that of $Y$ is varied from 0.05 to 4.
a) X and Y are normal distributions - $\mu_Y : \{0, 4\}$, $\sigma_Y = 1$

b) X and Y are normal distributions - $\sigma_Y : \{1, 5\}$, $\mu_Y = 0$

c) X and Y are lorentz distributions. Constant scale and varying location for Y.

d) X and Y are gamma distributions. Constant scale and varying shape for Y.

*Figure 3.1:* Power comparison for Hotelling’s $T^2$ (blue line) and CvM (purple line) tests
Results and Discussion

In Figure 3.1, we see the power comparison plots for Hotelling’s $T^2$ test (represented in blue line) and CvM test (purple line). In Figure 3.1 (a), both distributions X and Y follow fulfill the assumptions of Hotelling’s $T^2$ test and hence the test shows good power characteristics. It can be seen that the power of the CvM test is also comparable to the power of the hotelling’s $T^2$ test for this case. In Figure 3.1 (b), we notice that the power of the hotelling’s $T^2$ test drops drastically when the homoskedasticity condition is violated. But the power characteristics of the CvM test remains relatively higher compared to Figure 3.1 (a). In Figure 3.1 (c), also shows that the performance of Hotelling’s test is very poor when the data distributions varies significantly from a normal distribution. In Figure 3.1 (d), the gamma distribution begins to resemble the normal distribution as the scale parameters vary, hence the improvement in the performance of the hotelling’s test. We conclude that the power of the hotelling’s $T^2$ test drops significantly for any variations from its fundamental assumption. The cramer von mises is better suited for data where the distribution information is unclear or unknown.

3.3 Methods

We validated our method using two data sets:

- Structural changes caused by healthy aging:
- Structural changes caused by Dementia:

The images from the control and treated group were affinely registered to a manually chosen target image using FLIRT [44]. This removes the global differences between the images and the target. While no physical model for brain deformation has been established, it has been shown that non-linear deformation methods with diffeomorphic constraints are the most suitable for statistical analysis of inter-group variations [45] [46]. To this end, we have used only diffeomorphic or inverse consistent registration methods in this paper. We used the DARTEL method proposed in [47], the diffeomorphic demons (DD) [48] and the inverse consistent high dimensional warping method (HDW) [16]. Hernandez et al. [49] have shown that DD and DARTEL are theoretical close and practically equivalent for registration purposes. The property of inverse-consistency encompasses the diffeomorphic property. Hence, in theory, HDW transforms to be theoretically similar to the diffeomorphic transforms. A more detailed analysis of the similarities of the registration methods can be found in Chapter 2.
The Log euclidean strain vector maps are computed from the smoothed deformation fields from each of the three methods. The distance of the log euclidean strain vectors from the identity strain vectors were computed.

We performed a voxel wise normality test on the distance data from each group. We used the D’Agostino’s $k^2$ normality test [42]. Test were performed on each group separately. At each voxel, the distance values were concatenated into a vector. Normality test was performed on the vector at each voxel; the output of the test was a binary image indicating which voxels had failed the normality test. The number of voxels that failed the test were then obtained from the output image. Once the test was completed on both groups, we compared this with the result of the Hotelling’s $T^2$ two sample test to obtain the number of voxels which passed the significance threshold but failed the normality test.

We performed voxel-wise Hotelling’s $T^2$ two sample test on the data from each of the registration methods. The resulting p-value maps were corrected for multiple testing by controlling the FDR at 0.05. To compute the Cramér von Mises statistic and the corresponding p-values we used the implementation from [50]. The resulting p-value maps were corrected for multiple testing by controlling the FDR at 0.05. The C++ implementation of the LeSV, the normality and hypothesis tests were based on the ITK toolkit [26]. The implementation of all the tools used in this chapter will soon be made available for public download from our lab website [http://www.bsl.ece.vt.edu](http://www.bsl.ece.vt.edu).

### 3.4 Results and Discussion

#### 3.4.1 Normality Tests

The result of the normality tests showed that not all voxels that passed the $T^2$ test were normally distributed. Tables 3.1 and 3.2 shows the number of voxels that failed the normality tests in each of the data sets. We observed that the locations of the non-Gaussian voxels did not form any anatomically significant patter and therefore found numeric data more meaningful than images. Each table lists the number of voxels where the metrics failed the normality test. The table shows the number of voxels that failed the normality test for three cases: for the control group only, for the treated group only and voxels which were declared significant by the Hotelling’s $T^2$ two sample test but failed the normality test. The number of non-Gaussian voxels are different for the different methods for reasons mentioned on Chapter 2. The distance metric is computed from the DVF therefore the distances from the various methods vary for the same reasons are the DVF. In
both Tables 3.1 and 3.2, we notice that there are a finite number of voxels that have been declared significant in the Hotelling’s $T^2$ two sample test. This validity of the significance result at these voxels is now suspect because one or both of the input data has violated the assumptions of the test. At this point, we would like to mention that many readers could call into question the small sample size used for the tests. But it is to overcome this drawback that we have used a test particularly designed for small samples and hence the results of the test hold. At first glance, it seems that the number of non-Gaussian voxels in the HDW case seem to be comparable to the other two methods, but it must be noted that HDW forces the DVF of the background voxels to be zero. Hence the proportion of the voxels being tested in HDW is far lower than for the other two methods. Another feature we observed in the normality results is that the number of non-Gaussian voxels are greater for the Oasis data than the aging dataset (after normalization for image size). As discussed in Chapter 2, the properties of the DVF and its derived metrics vary with the data set being studied. *The normality tests have given us evidence to believe that hypothesis testing with a normality assumption is not a reasonable option for DBM.*

<table>
<thead>
<tr>
<th>Registration Method</th>
<th>Control Group (%)</th>
<th>Treated Group (%)</th>
<th>Significant in Hotelling’s $T^2$ test (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>79441 (1.9 × 10^{-2})</td>
<td>14917 (3.5 × 10^{-3})</td>
<td>2640 (0.0125)</td>
</tr>
<tr>
<td>DARTEL</td>
<td>65359 (1.5 × 10^{-2})</td>
<td>11331 (2.7 × 10^{-3})</td>
<td>5400 (5.4 × 10^{-3})</td>
</tr>
<tr>
<td>HDW</td>
<td>47160 (1.1 × 10^{-2})</td>
<td>443 (1.06 × 10^{-4})</td>
<td>41840 (0.4135)</td>
</tr>
</tbody>
</table>

*Table 3.1: Number of Voxels that failed the normality test for the Aging data set*

<table>
<thead>
<tr>
<th>Registration Method</th>
<th>Control Group (%)</th>
<th>Treated Group (%)</th>
<th>Significant in Hotelling’s $T^2$ test (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>107983 (1.6 × 10^{-2})</td>
<td>74409 (1.4 × 10^{-3})</td>
<td>64040 (0.01)</td>
</tr>
<tr>
<td>DARTEL</td>
<td>186626 (2.9 × 10^{-2})</td>
<td>146268 (2.3 × 10^{-2})</td>
<td>191310 (0.03)</td>
</tr>
<tr>
<td>HDW</td>
<td>80118 (1.2 × 10^{-2})</td>
<td>55102 (8.6 × 10^{-3})</td>
<td>78450 (0.012)</td>
</tr>
</tbody>
</table>

*Table 3.2: Number of Voxels that failed the normality test for the OASIS data set*

Figure 3.2 shows the p-value maps of the Hotelling’s $T^2$ test overlaid on the target image of the Aging data set for the three registration methods. Of particular interest is the comparison of results between DD and DARTEL which we expected to be identical because of their theoretic similarity. The inconsistency between the results of DD and DARTEL illustrate effect of erroneous normality
assumptions on DBM results. There is large conflict between the results of the HDW and the diffeomorphic methods in spite of the theoretic similarity between the methods. The large significant section in the thalamus, seen in both DD and DARTEL, is missing in HDW results.

In Figure 3.3, we show the results of the CvM test on the Aging data set. The FDR controlled p-value map is overlaid on the target image. Again, we particularly notice the similarity between the results of DD and DARTEL. The result of the HDW method seems largely inconsistent with the diffeomorphic registration results. In theory, a transform that is inverse-consistent is also diffeomorphic. But algorithms that use the two constraints independently produce vastly different transforms to characterize the same changes. The HDW algorithm shows that while there are a few significant changes between the two data sets, the evidence against the null hypothesis is small. The changes near the edges of the image maybe caused by the property of the HDW algorithm which forces the deformation outside the brain to be 0. This effect is seen in all the results corresponding to HDW (Figures 3.2-3.5 (c)). While the total number of voxels being declared as significant is reduced, we notice the reduction in the number of small, islands of significant voxels seen in Figure 3.2.

In Figure 3.4, we observe the use Hotelling’s $T^2$ to study the structural differences between a subjects with and without dementia. We observe that there are no anatomically specific pattern to the regions marked as significant except at the genu of the corpus callosum. As with the aging data, the results of the DD and DARTEL methods are not identical as expected. The results of the HDW method is vastly different from both the diffeomorphic methods. The significant region in the genu of the corpus callosum is not visible here. The result of the CvM test is shown in Figure 3.5. Here all the three methods are in agreement that there is no significant differences between the two groups being studied. In essence, the result of the CvM test is that there is discernible difference between the two groups based on their difference from the identity transform. Here we emphasize that the result of the test does not imply that dementia causes no structural changes. It shows that for the given group of subjects, the distance from the identity transform show no difference in distributions. Changing the metric used or increasing the number of samples in each group will yield different results.

### 3.5 Conclusion

In this chapter we showed that the use of parametric hypothesis tests for DBM analysis is not an optimal choice because:
- the comparison of the means of the two distributions alone does not provide sufficient information about the variability in the two groups.

- the assumptions of normality and homoscedasticity are violated for DBM data.

- for most DBM metrics, the assumption of normality is an unreasonable one.

We have introduced distribution independent testing of inter-group variability in DBM using the Cramer von mises test. The CvM test has produced a more conservative result which is also consistent across registration methods. This agreement in results allows us to conclude that much of the unreliability in DBM is caused by the inappropriate use of parametric statistical tests.
Figure 3.3: Results of CvM test – Aging Data set
Figure 3.4: Results of Hotelling’s $T^2$ test – Oasis Data set
Figure 3.5: Results of CVM test – Oasis Data set
Chapter 4

Distribution Independent Multivariate Statistical Testing for DBM

4.1 Introduction

In Chapter 3, we discussed why parametric hypothesis are not appropriate for DBM tests. We also proposed the use of univariate, non-parametric CvM for DBM when using scalar morphometric metrics. Researchers are now shifting towards the use of multivariate analysis of DBM. For the same reasons discussed in 3, we believe multivariate parametric tests should be replaced with non-parametric tests.

Non-parametric multivariate analysis is not as widely used as their univariate cousins. In the past, the computational complexity of these methods hindered their wide-spread use. But with the current availability of very inexpensive computational power, the use of non-parametric tests for multivariate analysis is now possible. In this chapter we propose the use of the Jupp multivariate two sample test (JMTS) for DBM analysis. The JMTS test is a non-parametric measure of the distance between two empirical distributions from a hyperplane. The tests is distribution-free in that the statistic does not follow a particular distribution under the null hypothesis. If conversion to p-values is desired it can be done through permutation tests.

The computational complexity of this method is overcome by the use of multicore programming. Since in DBM, the result at each voxel is considered to be independent of another, voxel-wise testing can be implemented in a multi-threaded fashion thereby reducing computational time by many scale factors.
4.2 Background

Univariate non-parametric hypothesis tests are are constructed on the principle of ordering of pairs of samples $o(x_i, y_j)$ which can be realized as $\text{sign}(x_i - y_j)$. The Jupp's multivariate two sample (JMTS) test generalize the idea of ordering of unidimensional samples to orientation of simplexes formed by the multi-dimensional samples. Consider two empirical distributions $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$ where each sample is a $p$-dimensional vector. Using JMTS test, we would like to test the null hypothesis that $X = Y$. The test determines equality of the distributions by determining if the samples of $Y$ lie on the same side of a hyperplane as the samples of $X$. This is measured by computing the orientation of simplexes formed the samples of $Y$ with the samples of $X$.

**Computing orientation of Simplexes:** Consider the $p \times (p + 1)$ matrix formed by the $p$ dimensional samples $(z_1, z_2, \ldots, z_p, z_{p+1})$. The simplex formed by the vertices $(z_1, z_2, \ldots, z_p, z_{p+1})$ has a signed volume $\frac{\text{det}(M)}{p!}$. $M = (u_1, \ldots, u_p)$is a $p \times p$ square matrix where the columns of $M$ are defined as $u_i = z_{i+1} - z_1$. The orientation of the simplex, $o(z_1, z_2, \ldots, z_p, z_{p+1}) = \text{sign}(\text{det}(M))$.

**Computing the JMTS statistic:** Consider the samples $y_{j_1}$ and $y_{j_2}$. The extent to which these two samples lie on the same side of a hyperplane as the the samples $x_{i_1}, \ldots, x_{i_p}$ is measured by

$$s(x_{i_1}, \ldots, x_{i_p}; y_{j_1}, y_{j_2}) = o(x_{i_1}, \ldots, x_{i_p}, y_{j_1}) o(x_{i_1}, \ldots, x_{i_p}, y_{j_2})$$  \hspace{1cm} (4.1)

If the samples are on the same side of a hyperplane then $s(x_{i_1}, \ldots, x_{i_p}; y_{j_1}, y_{j_2}) = -1$ and if they are separated by the hyperplane then $s(x_{i_1}, \ldots, x_{i_p}; y_{j_1}, y_{j_2}) = 1$. This is generalized to a two-sample statistic by averaging the orientation of simplexes formed by all possible combinations of the samples from both distributions. The statistic is given by:

$$S(x_{i_1}, \ldots, x_{i_n}; y_{j_1}, \ldots, y_{j_m}) = \frac{\sum \sum s(x_{i_1}, \ldots, x_{i_p}; y_{j_1}, y_{j_2})}{\binom{n}{p} \binom{m}{2}}$$  \hspace{1cm} (4.2)

where $1 \leq i_1 < i_2 \leq n$ and $1 \leq j_1 < j_2 \leq m$. The symmeterized statistic, $T(x_{i_1}, \ldots, x_{i_n}; y_{j_1}, \ldots, y_{j_m})$, which is used to test the null hypothesis in the JMTS test is given by:

$$T = \frac{n(m - 1)S(x_{i_1}, \ldots, x_{i_n}; y_{j_1}, \ldots, y_{j_m}) + m(n - 1)S(y_{j_1}, \ldots, y_{j_m}); x_{i_1}, \ldots, x_{i_n}}{2mn - n - m}$$  \hspace{1cm} (4.3)

The author claims that for the scalar case, it can be shown through algebraic manipulation that the statistic $T$ is a function of the Cramer-von-mises statistic (as discussed in the previous chapter).
Since $T$ is a symmeterized statistic, $-1 \leq T \leq 1$. When $T = 1$, there is a hyperplane separating $X$ and $Y$ and larger values of $T$ indicates that the two distributions are well-separated. The author suggests that the null hypothesis of the test be rejected for large values of $T$. Since the $T$ statistic, under the null hypothesis, does not follow a particular distribution, calculation of p-values and choice of rejection regions are done through permutation tests.

### 4.2.1 An illustrative example for the JMTS test

**Computing the $T$ Statistic:** In order to better illustrate the procedure to compute of the $T$-statistic, we use it in a toy example. In Figure 4.1, let $(x_1,x_2)$ and $(y_1,y_2,y_3)$ represent samples from two empirical distributions. Since the data is 2-D, they form 3 simplexes which are triangles. The orientation of each simplex is indicated in the figure. The position of these simplexes with respect to $(x_1,x_2)$ determine the orientation values for each simplex. Once the orientation of the simplexes are available we calculate $T$ as shown in Equation 4.4. The orientation values for all $\alpha(x_1,x_2,y_j)$ are shown in Figure 4.1.
Figure 4.1: Simplexes formed with 2-D data

\[
S_{xy} = \frac{s(x_1, x_2; y_1, y_2) + s(x_1, x_2; y_1, y_3) + s(x_1, x_2; y_2, y_3)}{\binom{2}{2} \binom{3}{2}}
\]

\[
S_{xy} = o(x_1, x_2; y_1) o(x_1, x_2; y_2) + o(x_1, x_2; y_1) o(x_1, x_2; y_3) + o(x_1, x_2; y_2) o(x_1, x_2; y_3)
\]

\[
S_{xy} = \frac{1(-1) + 1(-1) + (-1)(-1)}{\binom{2}{2} \binom{3}{2}}
= \frac{-1}{3}
\]

\[
S_{yx} = \frac{s(y_1, y_2; x_1, x_2) + s(y_1, y_3; x_1, x_2) + s(y_2, y_3; x_1, x_2)}{\binom{3}{2} \binom{2}{2}}
\]

\[
S_{yx} = \frac{o(y_1, y_2; x_1) o(y_1, y_2; x_2) + o(y_1, y_3; x_1) o(y_1, y_3; x_2) + o(y_2, y_3; x_1) o(y_2, y_3; x_2)}{\binom{3}{2} \binom{2}{2}}
\]

\[
S_{yx} = \frac{-1}{3}
\]

\[
T = \frac{2(2)(-\frac{1}{3}) + 3(1)(-\frac{1}{3})}{12 - 2 - 3}
= -0.4761
\]
Using T in a two sample test: The authors of the test only show a small example to illustrate the use of the JMTS test. We found the need to do a more detailed study on the use of the T-statistic as a two sample test to discriminate between two distributions. We performed a study using samples from 6 known distributions each with sample size smaller than 20 but greater than 15. The sample size was chosen to mimic the typical number of samples available for DBM studies. Figure 4.2 show visual representation of the data used for this study which encompasses all ranges of separation between distributions. For visualization purposes, we restricted this simulation to 2-D distributions only. In this experiment, we used the following distributions:

\begin{align*}
F_1: & \quad F_1 \sim \mathcal{N}(\mu_1, \Sigma_1) \text{ where } \mu_1 = [1 \ 2] \text{ and } \Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \\
F_2: & \quad F_2 \sim \mathcal{N}(\mu_2, \Sigma_1) \text{ where } \mu_2 = [10 \ 12] \text{ and } \Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \\
F_3: & \quad F_3 \sim \mathcal{N}(\mu_3, \Sigma_1) \text{ where } \mu_3 = [8 \ 7] \text{ and } \Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix} \\
F_4: & \quad F_4 \sim \mathcal{T}(\Gamma, \nu) \text{ where } \nu = 5 \text{ and } \Gamma = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix} \\
F_5: & \quad F_5 \sim \mathcal{N}(\mu_4, \Sigma_4) \text{ where } \mu_4 = [0 \ 0] \text{ and } \Sigma_4 = \begin{bmatrix} 0.25 & 0.3 \\ 0.3 & 1 \end{bmatrix} \\
F_6: & \quad F_6 \sim \mathcal{N}(\mu_4, \Sigma_1) 
\end{align*}

Table 4.1 shows the T values calculated in the simulation study. Each entry in the table, corresponds to the T value obtained for the distributions at the corresponding row and column header. For also calculated the T statistics for comparing the distribution to itself. We notice for most distributions, the T statistic is able to delineate between two disparate distributions. It can also been seen that the T statistic could not distinguish between F4 and F5. Considering the overlap between these two distributions (seen in Figure 4.2) this does not reflect very badly on the power of the statistic itself. It is also observed that the value of the T statistic is very similar for any T(x, F4) and T(x, F5). Since the statistic is based on the concept of distance from a hyperplane, distributions that are closely spaced should have similar distance from other distributions. Another property we noted is that the sensitivity of the T statistic will decrease as the number of sample
and/or the number of dimensions increase. An increase in either of the two values causes the denominator to increase in Equations 4.2 and 4.3, consequently reducing the value of $T$. This test is particularly suited for small sample sizes.

### 4.3 Methods

In order to illustrate the use of the JMTS test for DBM, we have conducted an experiment similar those in Chapter 3. The objective of this experiment is to detect structural differences between young and healthy adult human brains. Here we have used the multidimensional morphometric measure LeSV (Section 1.2.2).

The image from the two groups were first affinely registered to the target image. We then performed
<table>
<thead>
<tr>
<th>x/y →</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>-0.103379</td>
<td>0.242</td>
<td>0.17703</td>
<td>0.114191</td>
<td>0.147119</td>
<td>0.174511</td>
</tr>
<tr>
<td>F2</td>
<td>0.242</td>
<td>-0.168361</td>
<td>0.03471</td>
<td>0.276817</td>
<td>0.291631</td>
<td>0.38019</td>
</tr>
<tr>
<td>F3</td>
<td>0.17703</td>
<td>0.03471</td>
<td>-0.11332</td>
<td>0.276817</td>
<td>0.291198</td>
<td>0.37978</td>
</tr>
<tr>
<td>F4</td>
<td>0.114191</td>
<td>0.276817</td>
<td>0.276817</td>
<td>-0.00800</td>
<td>0.0030</td>
<td>-0.038711</td>
</tr>
<tr>
<td>F5</td>
<td>0.147119</td>
<td>0.29161</td>
<td>0.291198</td>
<td>0.0030</td>
<td>-0.0122189</td>
<td>0.004487</td>
</tr>
<tr>
<td>F6</td>
<td>0.174511</td>
<td>0.38019</td>
<td>0.37978</td>
<td>-0.038711</td>
<td>0.004487</td>
<td>-0.00183824</td>
</tr>
</tbody>
</table>

*Table 4.1: T* values calculated to discriminate between two distributions – Simulation Study

non-linear registration using three different registration methods: DARTEL [47], the diffeomorphic demons (DD) [48] and the inverse consistent high dimensional warping method (HDW) [16]. For justification of the choice of these methods and the Section 3.3. Once the LeSV maps were computed from the DVF obtained from the registration methods, we performed statistical analysis on these LeSV maps using:

**Hotelling $T^2$ Test:** Parametric statistical analysis was performed on the LeSV maps as described in Section 1.3.

**JMTS Test:** We implemented a voxel-wise JMTS test for DBM. At each voxel we computer the $T$ statistic and the map of $T$ values was used to analyze structural differences between the two groups of data. We implemented a multicore version of the test using C++ and Cilk++ [52] [53]. Cilk++ provides a set of simple extensions and utilities that allow extension of C++ applications to multicore applications. Here computation for each voxel is distributed to one core of a multicore machine. The computation of the $T$ statistic maps is time consuming; on an 8 core machine, it took 3 days per test. We were unable to perform conversion to p-values because of the computational time involved. This also prevented us from performing any correction for multiple testing. The values reported are uncorrected $T$ statistic values. From our simulation study, we learnt that the zero crossing of the $T$ statistic is a meaningful threshold for significant differences. Groups that are sufficiently separated have a $T > 0$. Therefore, we thresholded the statistic map at 0 and all $T > 0$ were included for analysis of differences.
4.4 Results and Discussion

Figures 4.3 (a-c) shows the result of the JMTS test for the aging data set. According to the Figures 4.3 (a-b), there are no specific patterns to the differences between the two distributions. The changes are distributed all over the image. For all the three figures, the range of $T$, were the same. The small values of $T$ indicate that there exists some differences between the two groups, the data are not well separated. It must also be noted the statistics will only be able to detect differences between distributions that are linearly separable. The differences between the young and the old group could not be sufficiently discriminated using linear separability alone.

![Figure 4.3: Results of JMTS T test – Aging Data set](image)
4.5 Conclusion

In this chapter we introduced the use of a multivariate, non-parametric test for DBM. The test is based on the separation of two empirical distributions by a hyperplane. The test is distribution free and the statistic, under the null hypothesis, does not follow any particular distribution. The test was used to detect anatomical differences between two groups of young and old people. The test is very computational cost of performing the test prevented the computation of p-values from the statistic maps. For the same reason, p-value correction for multiple testing could not be performed. Parametric testing of morphometric data has a number of drawbacks. While the test presented in this chapter may not be the most optimal alternative to the parametric tests, it is a step in the direction towards increasing the reliability of DBM analysis.
Bibliography


tensor based morphometry study of longitudinal gray matter contraction in ftd. *NeuroImage*, In Press, Corrected Proof. 2


