Gray Matter Surface-based Spatial Statistics in Neuroimaging Studies

By

Prasanna Parvathaneni

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

May 10, 2019

Nashville, Tennessee

Approved:

Bennett A. Landman, Ph.D.

Adam W. Anderson, Ph.D.

Neil D. Woodward, Ph.D.

Benoit M. Dawant, Ph.D.

Ilwoo Lyu, Ph.D.

Hakmook Kang, Ph.D.

Yuankai Huo, Ph.D.

ACKNOWLEDGEMENTS

I have a deep respect for the research happening in this field and I am grateful for having an opportunity to be part of this valuable research. Firstly, I express my sincere gratitude to my advisor Dr. Landman for the continuous support of my Ph.D study and related research, for his patience, motivation, immense knowledge and energy for science. His vision and guidance helped me immensely in gaining the research experience in the best possible way. I could not have imagined having a better advisor and mentor for my Ph.D study. Besides my advisor, I express my appreciation to Dr. Woodward for letting me work on the clinical dataset and for providing timely feedback with clarity of thought on do's and don'ts of approaching such analysis. It is a learning experience and I sincerely appreciate all the support that has been offered on this front. I would like to thank Dr. Lyu for his immense support for helping me to gain experience on shape-based analysis and research in general. His patience, constant feedback and easy going work style with enormous technical expertise is a great asset for making progress in my dissertation proposal. I would like to thank Dr. Anderson for being a constant support from the beginning of my research journey. My sincere thanks to Dr. Dawant for offering hands on course that helped me to understand the intricate details on the medical image processing and also for the valuable feedback and support whenever in need. I also would like to thank Dr. Kang for the insightful comments and encouragement. I thank Dr. Huo, who is a friend, colleague and mentor who is very approachable and provides simple solutions to the complex problems at hand.

My sincere thanks also goes to Dr. Nelson and Dr. Lupo from UCSF, who had provided me an opportunity to join their team as research assistant. It was a memorable and enriching experience that helped me to gain clinical perspective on the research applications. I am thankful to the entire team for providing such a friendly environment and opening doors for clinical collaborations.

I thank my fellow lab members starting with Zhoubing Xu and Kurt Schilling for providing a welcoming support into the research world by helping me to understand medical imaging concepts. I also would like to thank Peizhen Sun for all his research efforts and for bringing me up to speed on the first

SPIE publication. I would like to thank Andrew Plassard and Rob Harrigan for providing an insight into MASI lab and helping to keep checks on lab etiquettes. I appreciate all the discussions both technical and personal with my fellow lab members Yuankai Huo, Shikha Changati, Vishwesh Nath, Allison Hainline, Shunxing Bao, Camilo Bermudez, Peijun, Sandra, Roza and Colin. I also would like to thank Ben, Steve, Justin and Karthik for providing needed technical support in tirelessly honoring our requests for spider updates and Xnat queries. Thank you for your efforts in maintaining the systems and helping the lab to focus on the research goals and deadlines. I also would like to thank Linda Koger and Michele Bender for taking care of all the administrative paper work helping us to make smooth transitions in research domain.

Last but not the least, I would like to thank my parents, all of my family members, friends and well wishers who had helped me directly or indirectly in making this journey possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I. INTRODUCTION

1. Big picture

The human brain is the central organ of the nervous system that controls most of the activities of the body by integrating, processing and coordinating the information that is received from the sense organs and making decisions on the instructions to be sent out to the rest of the body. Understanding the development of brain microstructure is essential to understanding neurodevelopment. Neuroimaging is used to visualize the brain and record brain activity providing information related to both the structure and function of the brain. Neuroimaging analysis provides an opportunity to gain valuable insight into the microstructural changes in the brain associated with healthy growth and neurological disorders when conducting longitudinal or cross-sectional studies involving multiple subjects.

New imaging techniques [1, 2] or protocols [3] have been introduced over the past few years that make it possible to capture microstructural complexity of dendrites and axons at millimeter scale in clinical settings. Moreover, recent advances [4-6] in the medical imaging community have made it possible to access large scale datasets acquired at various sites across the world, including different modalities and parameters of interest like tissue properties and blood oxygenation. With the increasing number of publicly available neuroimaging databases, including the Adolescent Brain Cognitive Development (ABCD) [7], the UK BioBank [8] and Connectome studies [9], there is an opportunity to combine large-scale imaging resources to increase the power of statistical analyses to test the common biological hypothesis. Community-wide efforts are underway to address standardization of acquisitions and analyses for imaging biomarkers as described in imaging biomarker road map for cancer studies [10]. These innovations bring exciting opportunities to probe into understanding the brain microstructure in a multidimensional setting with extensive demographic information. The data not only help in standardizing the expected tissue microstructure composition in a given population but also aid in

personalized therapeutic treatments based on the deviation observed from the healthy population.

However, the data interpretation in this area needs to be approached with extreme care as it poses challenges because of group bias and heteroskedastic variance. Three main factors that lead to biased conclusions in research findings, limiting the validity of published results in imaging studies with neuroscience applications are: widespread low statistical power, selective outcome and selective analysis reporting [11]. As part of this dissertation, we propose strategies that focus on improving statistical power in quantifying brain microarchitecture features for conducting group studies. We propose to demonstrate the application of these techniques in three different clinical applications to carry out a multivariate analysis of brain microstructure features in psychosis population, to study the effects of aging on brain microstructure features on a large-scale population study, and improving human brain sulcal curve labeling of aging population in a large-scale MRI dataset. Our goal is to show that in both cases we achieve improved spatial statistics and better predictability.

The rest of the chapter is structured in the following order. First, anatomical background relevant to neuroscience or psychiatry is presented in the context of this dissertation. In the next section, neuroimaging techniques appropriate for the goal of our analysis are briefly covered. Then, data processing steps involved in converting the signal from imaging the brain to relevant metrics in the context of understanding brain microstructure are discussed. These cover both intra-subject, (i.e., dealing with multi-modal data from a single subject) and inter-subject (i.e., analyzing data across multiple subjects) analyses. Then various types of group analysis techniques are presented that include the region of interest (ROI) based, parametric and non-parametric tests that are commonly used in neuroimaging along with the trends and challenges associated with each of them. A brief rationale of why we are leaning to non-parametric tests as part of this study is stated. Then a list of possible sources of bias in the data analysis for population studies is presented that includes template bias and the variability arising from site/scanner differences. The challenges/gaps that we are trying to address as part of this dissertation are presented. These challenges are targeted for improving the statistical power while incorporating the multi-modal analysis in brain microstructure analysis. Then we present the targeted clinical applications

from this study.

2. The anatomical context in neuroscience or psychiatry

The human brain is composed of more than 100 billion neurons [12] that vary in size, shape, and function and they form networks for information processing that are responsible for all of our thoughts, sensations, feelings, and actions [13]. The cortical layers of neurons constitute much of the brain's gray matter (GM), which are interconnected to other brain areas by axons (WM). The mature human brain has a characteristic pattern of folds called the gyri and the grooves in between them are called the sulci.

Much of the brain (approximately 45%) comprises WM, the predominantly myelinated axon bundles that originate in the neurons and carry electrical impulses between adjacent gyri and between different lobes of the brain [14]. The human brain continues to undergo myelination (which implies WM microstructure changes are prevalent in the context of neuroimaging) until at least the third decade of age [15]. The frontal regions of the cerebral cortex, which carry out higher-level executive functions, are the last to become myelinated [16]. GM constitutes both cortical and subcortical structures. GM cerebral cortex is about < 5 mm thick and is key to many executive functions of the brain. Any alterations in this region with neuropathologic conditions could be a potential biomarker to facilitate early diagnosis and assessment of disease severity. Another critical component is cerebrospinal fluid (CSF) that is a colorless fluid that fills and surrounds the brain and spinal cord. It provides a mechanical barrier against shock. An increase in the volume of blood or brain tissue results in a corresponding decrease in the fluid and vice versa. Illustration of these regions is shown in Figure I-1 with a zoomed in the region showing the axonal pathways entering and exiting the GM.

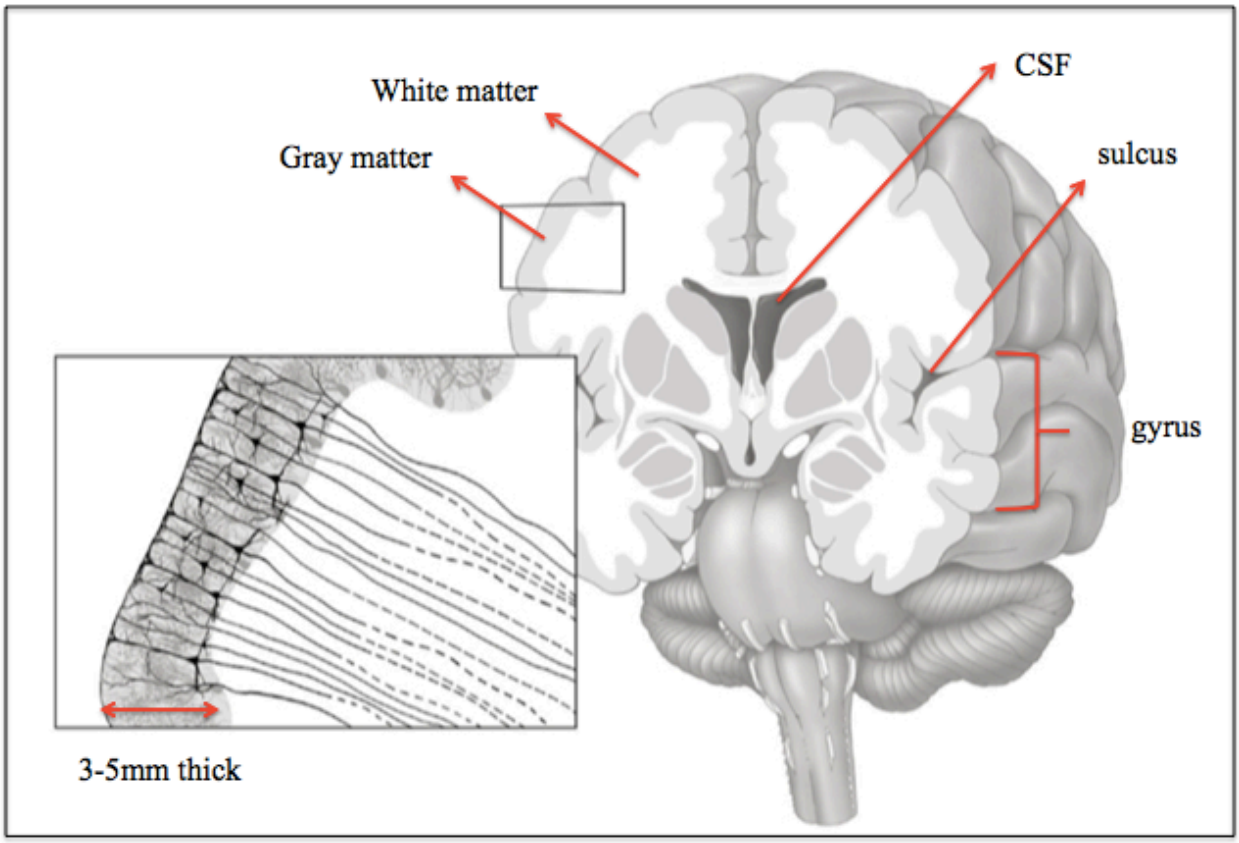

Figure I-1 Schematic illustration of brain anatomy with (a) gray matter which is < 5 mm thick (b) white matter (c) CSF (d) sulcus and (e) gyrus as indicated by the arrows. The boxed image is zoomed in to show a portion of GM and WM region with an illustration of axonal pathways entering and exiting the gray matter. Dotted lines indicate the pathways exiting GM. (By Ms. Emma Vought [17])

3. Types of Neuroimaging

Neuroimaging is the use of imaging techniques to either directly or indirectly image the structure or function of the central nervous system. While there are other types of neuroimaging, like computed tomography (CT), magnetoencephalography (MEG), single-photon emission computed tomography (SPECT), in this section, we focus on magnetic resonance imaging (MRI) as it is relevant in the context of neuroscience or psychiatry applications to study brain microstructure features. Structural imaging deals with the structure of the nervous system and helps in the diagnosis of diseases such as brain tumor and injury. Functional imaging is used to diagnose metabolic diseases at a finer scale in addition to neurological and cognitive psychology research. In recent years, the development of neuroimaging

techniques such as high-resolution magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), diffusion-weighted magnetic resonance imaging (dMRI), positron emission tomography (PET) or single photon emission tomography (SPECT) have promoted the identification of structural and functional characteristics of the developing brain and underlying mental disorders [18-24].

3.1. T1 weighted MRI

T1 weighted MRI (Figure I-2) is one of the essential contrast mechanisms in MRI that provides differences in the T1 relaxation times of different tissues. T1 images are typically acquired at a high resolution around 1mm.

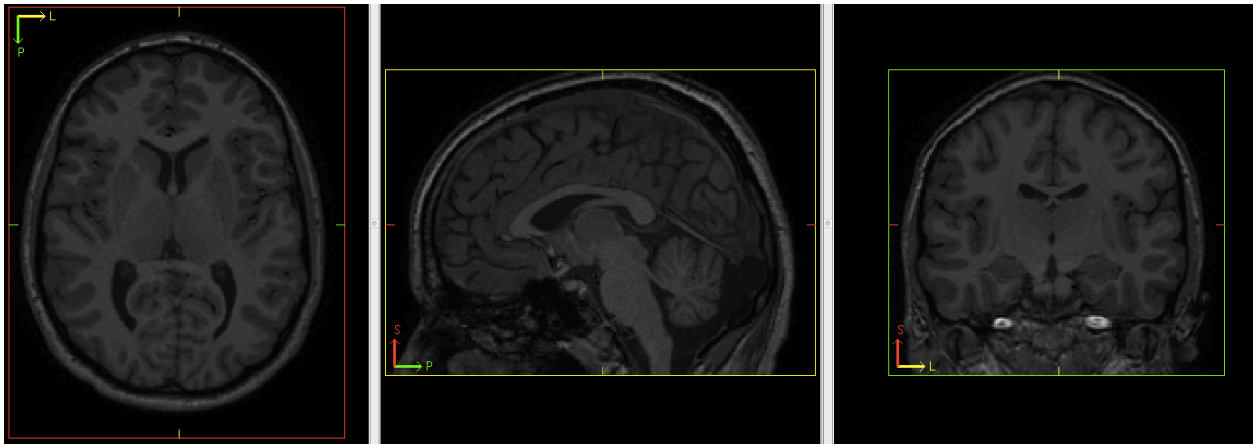

Figure I-2 T1 weighted image showing the anatomical structure of the human brain in axial, sagittal and coronal planes.

T1 weighted images are typically used for segmentation of brain into different tissues like WM, GM and CSF. Image segmentation is a critical step in clinical applications that is used for visualizing specific anatomical regions in the brain or for analyzing brain related changes. There are different segmentation methods like intensity-based approaches [25, 26], atlas-based approaches [27] or surface-based approaches [28] that can further label the images into smaller anatomical regions. In this study, we used multi-atlas segmentation [29] according to the BrainCOLOR protocol [30]: 132 brain regions and 1 background (Figure I-3) to automatically label each voxel of the image.

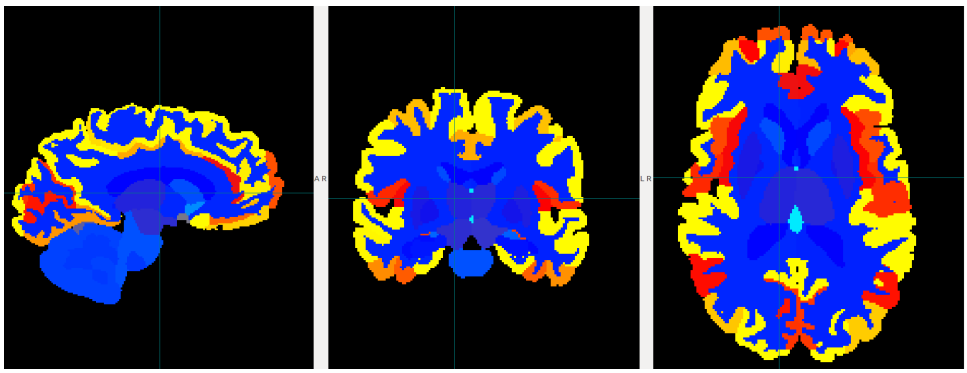

Figure I-3 Segmentation of T1 weighted image according to brainColor protocol with 133 labels.

T1 images are also used for cortical surface reconstruction that is a 2D representation of brain surface with corresponding cortical parcellation (Figure I-4).

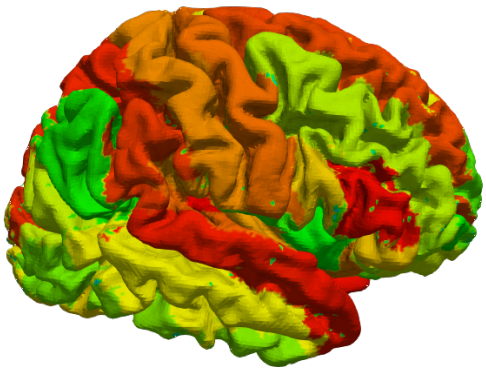

Figure I-4 Cortical parcellation from brainColor protocol

The cerebral cortex is the outermost layer of neural gray matter and is critical for many brain functions including memory, attention, cognition, language, and consciousness[31]. Features that characterize the cortex (including sulcal curves, gyral curves, sulcal depth, curvature, and cortical thickness) are essential in neuroimaging studies involving these regions [32-34].

3.2. Diffusion MRI

Diffusion-weighted magnetic resonance imaging (DW-MRI) is an advanced imaging technique to characterize the diffusion of water in biological tissues non-invasively [35]. Diffusion MRI has been used

to produce images of brain tissue structure and connectivity, in the healthy and diseased brain. This technology has revolutionized the management of acute brain ischemia (stroke), saving the lives of many patients and sparing them significant disabilities as diffusion MRI was able to capture the dying brain cells early on highlighting them as bright spots while no abnormality is seen in standard MRI images [36]. Besides, diffusion MRI is now widely used for the detection of cancers and metastases (breast, prostate, liver) [37, 38]. Another major field of application of diffusion MRI is in studies on the connectivity of the brain [39]. Diffusion tensor imaging (DTI) [40] is a DW-MRI technique that enables the measurement of the diffusion anisotropy of water in tissue for producing neuronal tract images and has become an established method for assessing white matter 'integrity' and connectivity in healthy and unhealthy brain structure (Figure I-5).

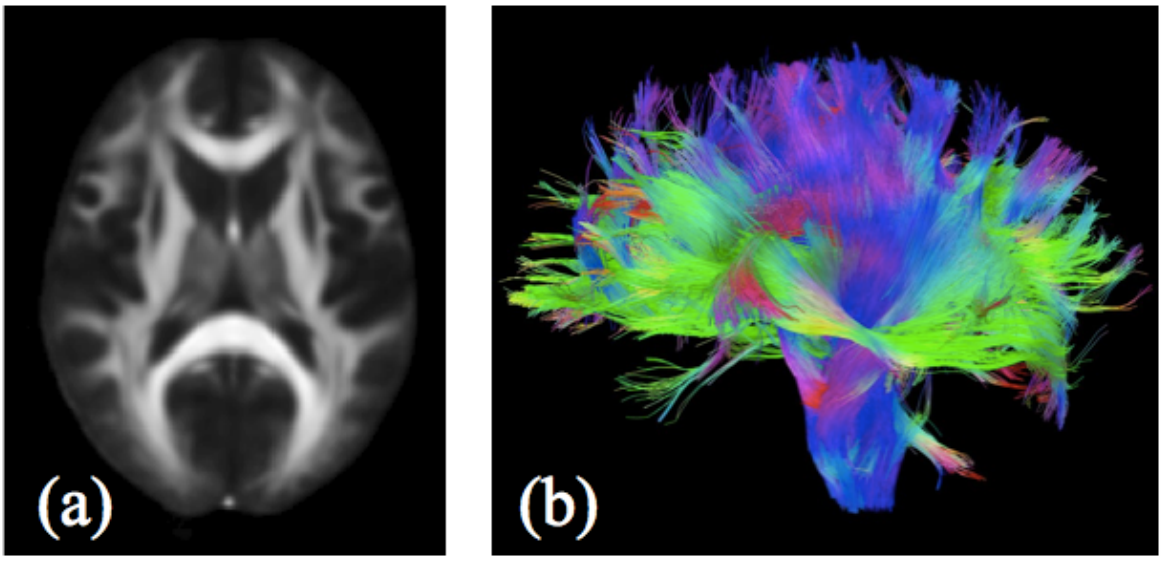

Figure I-5 Illustration of DTI based measures typically used in most studies (a) Fractional anisotropy (b) Tractography representing neural tracts.

However, DTI assumes that diffusion follows a tri-variate Gaussian distribution containing a single, coherently oriented bundle of white matter axons (e.g., the diffusion "ellipsoid") thus leading to non-specificity of the underlying microstructure of neurite-specific measures such as their density and orientation dispersion.

3.3. Advanced diffusion MRI

Recent advances in diffusion MRI modeling have led to the development of several state-of-the-art methods that allow for distinct features of the diffusion to be computed, which in turn reflect properties of fibrous tissue in the brain and other organs [14]. Microstructure imaging aims to address this limitation by fitting more descriptive models of tissue (e.g., axons, glial cells, and extra-cellular space) that relate to specific parameters in different tissue compartments [41-50]. Neurite orientation dispersion density imaging (NODDI) [48] has emerged as a clinically feasible MRI technique for estimating the microstructural complexity by combining a three-compartment tissue model with a two-shell high-angular-resolution diffusion imaging (HARDI) protocol [48]. Briefly, NODDI provides quantification of the volume fraction of the intracellular compartment with restricted diffusion, extra-cellular compartment with hindered diffusion and isotropic compartment along with an orientation dispersion index (ODI) of the neurites.

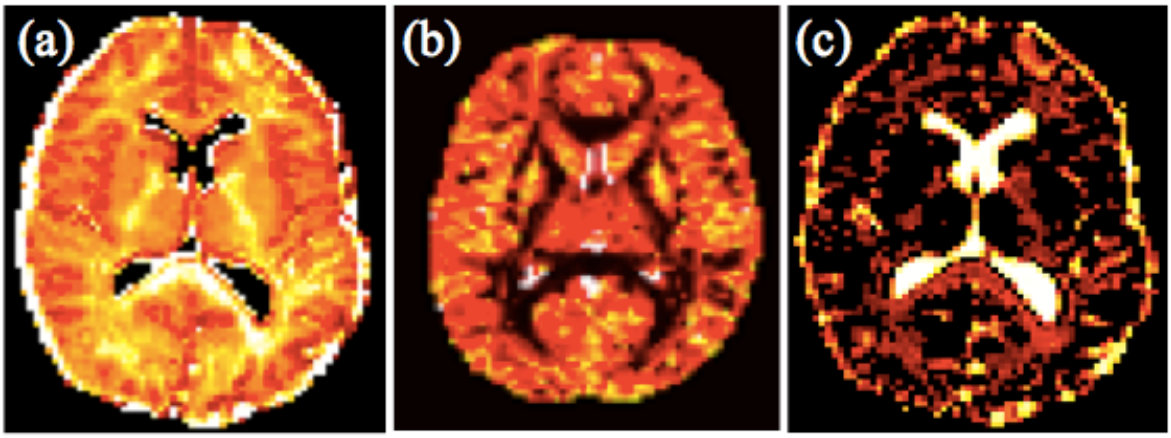

Figure I-6 Illustration of the metrics obtained from NODDI model (a) Intracellular volume fraction (V_{ic}) providing neurite density index (b) Orientation dispersion index (ODI) (c) isotropic volume fraction (V_{iso}) providing volume fraction of CSF compartment.

Intracellular volume fraction (V_{ic}) is modeled to represent the compartment defined by axons and dendrites, which are collectively called neurites. Extra-cellular volume fraction (V_{ec}) refers to the space around the neurites that is occupied by various types of glial cells and cell bodies (somas) in gray matter.

Isotropic volume fraction (V_{iso}) refers to the volume fraction of the cerebrospinal fluid (CSF) compartment in each voxel[48]. Voxel-wise quantification of these metrics over the whole brain (Figure I-6) presents new opportunities for understanding brain development and disorders.

3.4. Functional MRI

Functional MRI is based on the essential observation that when a neural activity increased in a particular area of the brain, MR signal is also increased by a small amount. It is currently the most widely used method for brain mapping and for studying mechanisms of human cognition. The vascular architecture of the brain limits biological spatial resolution of fMRI. Neural correspondence, measured as a correlation between cortical activity and the evoked BOLD response, falls to less than 50% for voxel sizes below 2.8 mm^3 [51]

4. Data processing steps in Neuroimaging

There are various steps involved in the quantification of brain microarchitecture through multivariate image processing (Figure I-7). These steps range from acquiring the brain MRI from a scanner to the final analysis for hypotheses based testing. Quality check procedures are a vital part in carrying out any analysis as each step involves data manipulation in preparing it for the next step.

4.1. Intra-Subject data processing

A preliminary step is the intra-subject processing of the data as illustrated in Figure I-7. Different types of images like T1 weighted, DWI, and fMRI could be acquired from an MRI scanner based on the scan protocol settings. T1w images are used to segment the brain into WM, GM and CSF regions and can further be extended to cortical reconstruction [52] to obtain GM surface. The segmentation can further be divided into smaller regions using multi-atlas segmentation approaches [53]. DWI images are processed to obtain model specific parameters like fractional anisotropy from DTI or ODI from NODDI. The first level analysis is performed on fMRI data to obtain contrast maps. Images from all modalities are co-registered to T1w image. All the images are then transferred to a standard template preparing it for cross-

subject analysis.

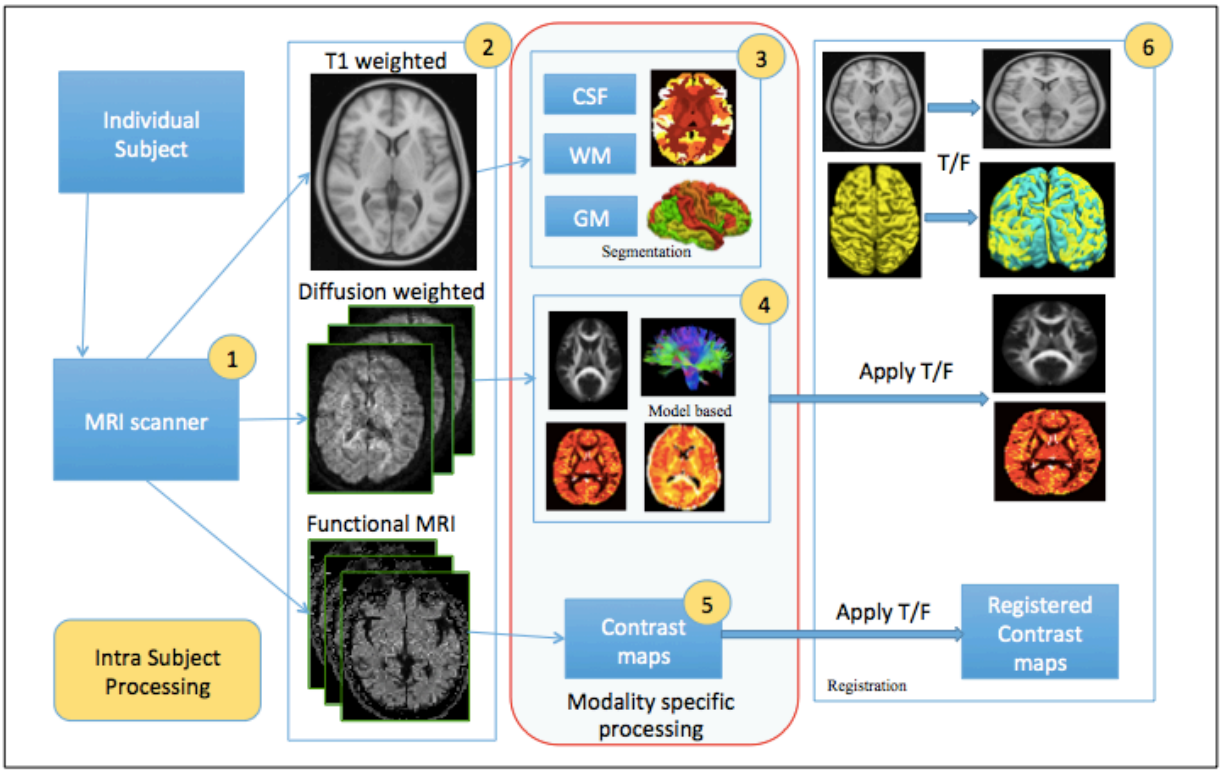

Figure I-7 Intra subject processing workflow: Illustrated are the steps involved in intra-subject processing from image acquisition through the quantification across modalities. 1) Setting up scan protocols to obtain the imaging parameters of interest. 2) Image modalities including T1 weighted tissue contrast, 4-dimensional diffusion-weighted images with the corresponding gradient in the time domain and four-dimensional functional MRI imaging bold signal. 3) Segmentation of T1 weighted image into CSF, WM and GM components and also further segmenting them into multi atlas labels both in volume and cortical surface domain. 4) Obtain model based images from DWI images top row showing DTI based metrics like FA and corresponding tracts in white matter and bottom row showing NODDI outputs obtained with HARDI sequences. 5) First level analysis of functional MRI images to obtain contrast maps. 6) Both volume and surface-based registration to MNI space from T1 weighted images and applying the transform to other modalities to transfer all metrics to MNI space.

4.2. Inter-Subject data processing

The human brain varies in shape and size across different subjects. While intra-subject analysis addresses spatial alignment issues within the same subject, it does not necessarily involve shape or size differences. Inter-subject analysis, on the other hand, addresses the difference in shape and size

information of the brain that could cause potential inconsistencies for carrying out of group analysis. It is common to normalize anatomical images to a standard template to overcome the brain shape variability. Several techniques are proposed for carrying out this step including both intensity-based normalization [54] and cortical surface (shape) based normalization [33, 34, 55-59].

4.2.1. Volume-based Registration (VBR)

Volume-based registration [54] is achieved based on maximum overlapping voxels between the template image and individual subject anatomical image. The ultimate aim is to transform the individual brains into a common space to make them spatially comparable to each other as needed for group analysis. The transformation that is achieved from the anatomical images to normalize to the template space is applied to the low-resolution images from the same subject to align all the modalities to a template space.

4.2.2. Surface-based Registration (SBR)

The cerebral cortex is the outermost layer of neural gray matter and is critical for many brain functions. Cortical surfaces are widely used for analyses in this region as they preserve topology [33, 34, 55-59]. It is common to align all the cortical surfaces to a common space [60, 61] to study group differences in these regions between control and clinical samples. Representative template space-based approaches have been proposed to study local individual differences in cortical morphometric measurement due to their ability to represent data involving cortical patterns and other model-based voxel wise parameters mapped onto a common surface in both normal and clinical populations [62, 63].

5. Group analysis methodologies in Neuroimaging

There are different types of cross-subject analysis performed for group studies as shown in Figure I-8. Data from each modality across all subjects are aligned to target template either an Montreal Neurological Institute, MNI template or population-based template as mentioned in the above section. Below are different types of group analysis carried out at each level.

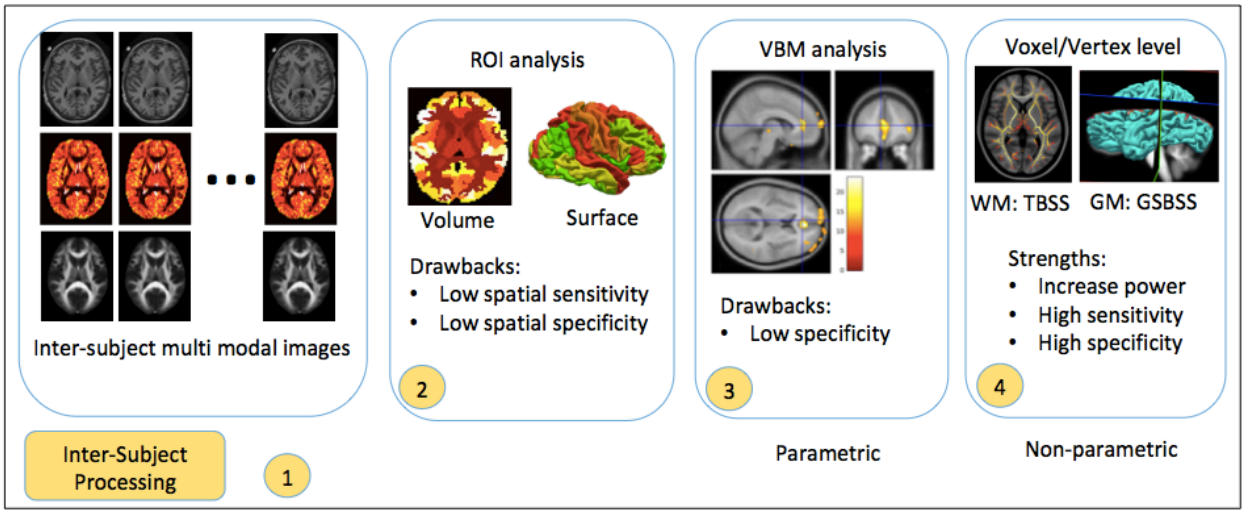

Figure I-8 Inter-subject analysis: Different levels involved in conducting a group analysis across subjects is shown. 1) Multi-modal data from all the subjects need to be aligned to a common template space before conducting group studies to derive metrics of interest. 2) Statistical metrics can be derived from each of those subjects at regions of interest (ROI) level to compare across subjects or between groups. ROI analysis can be done using volume segmentation or cortical surface segmentation. As the results are often averages across multiple voxels/vertices, there is low sensitivity or specificity to report underlying microarchitecture differences. 3) Voxel-based morphometry (VBM) analysis is performed across the entire volume at voxel level by applying smoothing to overcome registration artifacts and partial volume effects. VBM improves the sensitivity of the study but loses specificity due to smoothing. 4) Tract-based spatial statistics for WM or Gray matter based spatial statistics (GSBSS) for gray matter is carried out at voxel/vertex level retaining both sensitivity and specificity by overcoming registration artifacts and partial volume effects. This analysis has the potential to increase statistical power as the voxels/vertices get reduced to highly probable WM/GM units.

5.1. Region of interest (ROI) analysis

For the region of interest (ROI) analysis [61, 64, 65], statistical metrics are derived from each segmentation label from volume segmentation or cortical surface parcellation. Another approach would be to consider metrics along the tracts [66] based on diffusion tensor imaging (DTI) or based on sulcal curves [62].

5.2. Whole brain voxel-based analysis/Parametric tests

Voxel-based morphometry (VBM) analysis [67], considers the whole volume and performs parametric tests across the corresponding voxels in an image. This technique allows investigation of local differences in brain anatomy using a statistical approach using statistical parametric mapping [68]. In

volumetric neuroimaging analyses, spatial smoothing is generally performed to improve image alignment and statistical sensitivity. However, this could lose specificity of the underlying region of interest [69]. As the GM of healthy adult subjects is typically only <5 mm thick, spatial smoothing needs to be carefully performed to retain the sensitivity and specificity of the underlying changes [70, 71].

5.3. Non-parametric analysis

Non-parametric tests [72] the data at hand to perform inference making as few assumptions as possible thus providing a flexible methodology for neuroimaging. These tests include nonparametric equivalent to two-sample t-tests [73], one-sample t-tests [74], paired t-tests [75] with techniques like bootstrap and permutation tests that are computationally intensive. Statistical nonparametric testing [76] is equivalent to SPM using permutation tests instead of random field theory. FSL randomize [77] is another popular tool that is used for carrying out non-parametric permutation-based tests. Threshold-free cluster environment (TFCE) [78] is the most used thresholding option for finding clusters where cluster-like structures are enhanced, but the image remains primarily voxel wise. In order to gain power in non-parametric tests involving many voxels from the whole brain analysis, and also to improve the sensitivity of the analysis, skeleton based approaches are introduced where the data is analyzed in 2D domain. Following analysis fall into that category of performing non-parametric tests using TFCE option in 2D domain.

5.3.1. Skeleton-based analysis

TBSS Tract-based spatial statistics (TBSS) has proven to be a popular technique for performing voxel-wise statistical analysis that aims to improve sensitivity and interpretability of analysis of multi-subject diffusion imaging studies in white matter. Gray matter based spatial statistics method (GBSS) was first proposed in studying the cortical microstructure development of preterm human brain [79]. This method adapted tract-based spatial statistics (TBSS) [80] framework to overcome partial volume contamination in diffusion measures within cortical microstructure where spatially transformed individual cortical measurements are projected onto the mean cortical skeleton by searching in a direction

perpendicular to the cortical skeleton identifying maximally probable cortical voxels. Unlike TBSS where projection was based on high fractional anisotropy (FA), GBSS method seeks voxels with a high probability of gray matter (GM).

5.3.2. Surface-based analysis

Surface-based approaches have been proposed with improved sensitivity in cortical morphometry [63, 69, 81-84] over volumetric neuroimaging in both fMRI and cortical features of interest. In surface based approaches morphometric measures are derived from the geometry of the cortical surfaces. Cortical surfaces are inherently two-dimensional in nature with set of vertices/coordinates that are derived from the MRI during cortical surface reconstruction process.

5.4. Strengths and weaknesses of each of the group analysis methods

Once the ROIs are defined either through the segmentation labels or with the custom region of interest like boundary around a brain tumor region, statistical metrics like mean and standard deviation within this region are captured. For example, in a single subject, for metric M within an ROI of n voxels, statistical mean is obtained as the sum of all the intensity values within the ROI as,

$$ROI_{mean} = \frac{1}{n}\sum_{i=1}^{n} M_i \tag{1.1}$$

The difference in the statistical metrics across all the subjects within a single group or between different groups is compared for hypothesis testing [64] [85]. However, if the region of effect is much smaller than the ROI, then it is possible that the measures taken in this approach can get obscured. As the results are often averages across multiple voxels/vertices [86], it is possible to lose the spatial sensitivity and specificity of underlying microstructure feature changes.

Localized regional abnormalities can be captured in VBM using statistical parametric mapping with Generalized linear model (GLM) [87] that is used to construct statistical maps. Gaussian random field theory [88] is used to resolve multiple comparison issues. In GLM, each outcome variable Y is assumed to be generated of the dependent variable X from a particular exponential family of distributions

(like Poison or normal).

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \epsilon_i \tag{1.2}$$

Random field theory [88] has the characteristic of data under Null hypothesis H_0, which holds that all data activations are merely by chance, and each voxel has a random number. If H_0 is assumed to be true, and the probability of falsely rejecting H_0 is controlled by some fixed threshold α, then, the probability of making one or more Type I errors in a family of tests is given by,

$$FWER = P(\cup \underset{i=1,..m}{\{T_i \geq u_\alpha\}} | H_0) \leq \alpha \tag{1.3}$$

where m is the number of voxels in the image. As this necessitates knowing the distribution of t-statistic T_i under the null, several methods are attempted to approximate this distribution. The most conservative approach for dealing with multiple comparisons is to use the Bonferroni correction [89] defined as,

$$P(T_i \geq u_\alpha | H_0) \leq \frac{\alpha}{m} \tag{1.4}$$

With Boole's inequality [90], we are guaranteed to control FWER at threshold α using Bonferroni correction as,

$$FWER \leq \sum_i P(T_i \geq u | H_0) \leq \sum_i \frac{\alpha}{m} = \alpha \tag{1.5}$$

In random field theory (RFT) the image of voxel wise test statistics are assumed to be a discrete sampling of a continuous smooth random field. Smoothness of the image is defined in terms of resolution elements called resels, which are roughly equal to number of comparisons. In a two dimensional image, with volume of search region as V, and full width half maximum (FWHM) representing estimated smoothness, resels are defined as,

$$R = \frac{V}{FWHM_x + FWHM_y} \tag{1.6}$$

RFT uses information on smoothness of the image and resel counts to compute Euler characteristic, a topological property at different thresholds u to determine the needed threshold α. While each of these correction methods are acceptable in the neuroimaging field, RFT also provides cluster-level correction. As these two models are parametric that make number of assumptions (normal

distribution, derivative of multivariate Gaussian, etc), the results of such approach are influenced by the validity of the assumptions.

Non-parametric tests, on the other hand, do not make assumptions on the distribution of underlying data thus making it a better choice compared to that of parametric tests. These are shown to provide improvements in power and validity, especially in small datasets. Permutation tests can be used for getting approximate of the distribution of the maximum statistics by repeatedly resampling the data under the null hypothesis. The only assumption needed for permutation tests is the exchangeability of the data.

6. Possible sources of error in group analysis for population studies

6.1. Model reproducibility

Different scan protocols could have an impact on the data quality specific to the model of interest. Moreover. Jones et al. [91] argues "Knowing the uncertainties, higher moments, and probability distributions of various DT-MRI parameters could improve our ability to glean more information from DT-MRI data and to design DT-MRI experiments more efficiently, particularly longitudinal or multi-site studies". As more studies consider the NODDI model, it is essential to investigate the empirical reproducibility and performance of this model with respect to scan time.

6.2. Harmonization of diffusion microstructure features

Using diffusion tensor imaging (DTI), ENIGMA's disease working groups have begun to compile evidence across cohorts for differences in a range of DTI measures and discovering factors that consistently affect brain structure and function [92]. However, the community of clinicians, engineers, and physicists is not yet ready to agree on a single best practice approach to advanced DW-MRI. Therefore, it is imperative to understand when (and how) diverse protocols can be analyzed to enable comparison and optimization across protocols based on practical study design and imaging constraints.

An application of advanced DW-MRI methods in clinical research studies has been hindered by a

lack of consensus on best protocols. DW-MRI suffers from between-scanner variation that hinders comparisons of images across imaging sites, scanners and over time, a model that can preserve biological variability and remove the unwanted variation introduced by the site is needed. Using fractional anisotropy (FA) and mean diffusivity (MD) maps, Fortin et al. compared several harmonization approaches and found that ComBat performed best at inter-site variability correction [93]. ComBat has also been applied to harmonization of cortical thickness measures across scanners and sites [94] and multicenter radiomic studies with positron emission tomography [95].

6.3. Bias in template selection

Prior work has addressed the importance of template surface selection from the perspective of pairwise registration [56, 57, 96]. In a template based registration approach, each surface is mapped to a common template surface in coordinate space by regularizing based on feature information. However, surface-based analyses employing a predefined template might yield undesirable results if the selected template surface is substantially different from the population or if it is biased towards a particular set of surfaces [97]. Template-based registration is dependent on the a priori template specification thus constraining the underlying data to be biased to the selected template. Methods have addressed the issue of dissimilarity between template surface and surfaces of population under consideration by organizing the population of cortical surfaces into pairs with high shape similarity to achieve a higher accuracy by only corresponding such similar pairs [96], while others factored in the pattern of folding across the entire cortical surface in considering the inter-subject average [98]. But these approaches are still prone to bias towards the majority representation of the underlying population that could pose a problem in cross subject analysis.

7. Contributions

In this study, we proposed techniques to improve statistical power for population studies to quantify brain microstructure with a focus on gray matter. The main contributions of this dissertation are (1) to show that our gray matter surface based spatial statistics (GSBSS) approach improves the statistical

power for performing group studies in neuroimaging compared to that of traditional registration methods (**Chapter III, IV**), (2) to address source of bias and variance in group studies by investigating empirical reproducibility of the NODDI model (**Chapter II**), correcting for inter-scanner variability effects of diffusion microstructure features (**Chapter V**) and constructing unbiased feature based cortical surface template (**Chapter VI**), (3) to apply deep learning techniques on cortical surfaces for improved sulcal curve labeling on large datasets (**Chapter VII**).

More specifically:

- **Empirical model reproducibility** is performed on NODDI model in **Chapter II** as the optimized NODDI protocol can be carried out in a clinical setting. Moreover, as we delve into understanding brain tissue microstructure including GM. This model was chosen as it disentangles neurite density and orientation dispersion estimates yielding more specific markers.
- The **GSBSS approach** was developed using gray matter surfaces in **Chapter III** to overcome the challenges of alignment issues and partial volume effects in low-resolution images like diffusion/functional MRI. Application of this technique was shown in **diffusion MRI in psychosis population**. This effort will facilitate getting the corresponding metrics of interest from each modality (T1, fMRI, DWI) in subject space and map them onto a target surface template in MNI space by making use of high-resolution T1 images for cortical surface segmentation and registration.
- We enhanced the GSBSS pipeline with a **search methodology** and also showed the **application in functional MRI** in **Chapter VI**. We showed that this approach improves statistical power compared to that of volumetric (VBM) or surface-based registration methods and human connectome project minimum processing pipeline.
- We performed **data harmonization of diffusion microstructure features** in both white matter and gray matter with data across multiple sites and scanners. While data harmonization has been

addressed in different contexts in previous works [93, 99], the novel part of our contribution is to apply this technique to spatial statistics derived from TBSS and GSBSS as shown in **Chapter V**.

- We showed the approach for **constructing statistically unbiased cortical surface templates** using feature-space covariance in **Chapter VI** to address population bias in template selection. Since we are dealing with cortical surfaces for GM spatial statistics, we have chosen to address the issue of bias in choosing a template for population studies.
- We presented how the **cortical surface features are integrated into a deep neural network** to achieve **improved brain sulcal curve labeling** for achieving improved speed, higher accuracy and better consistency in **Chapter VII** in large-scale brain MRI datasets.

Finally, observations of GSBSS reproducibility and lessons learned with different clinical applications are captured in an observation chapter followed by the conclusion.

7.1. Clinical applications

The motivation behind each of the contribution is to apply in clinical studies ultimately. We show three different clinical applications with improvement in statistical power. The first application is in psychosis population to study the differences between healthy vs. psychosis, the second application is to validate aging effects in diffusion microstructure features, and third application to label sulcal curves on a large-scale MRI dataset in an ageing population.

7.1.1. Application I: Hypothesis testing on Psychosis population

Based on post-mortem studies, microstructure changes were seen in the prefrontal cortex, in individuals with psychosis [100, 101]. Recent in vivo studies have demonstrated the ability of NODDI technique in identifying the group differences between healthy and psychosis groups in white matter [102] and gray matter (GM) regions [103, 104]. Using the NODDI model, neurite orientation dispersion index (ODI), a putative marker of dendritic structure and complexity, was calculated and compared between healthy controls and individuals with psychosis. Prior post-mortem studies have shown reduced dendritic length in prefrontal cortex region in schizophrenia and bipolar disorder population. To validate

the results, statistical tests are compared between GSBSS and gray matter based spatial statistics using NODDI (N-GBSS) to study the differences between healthy and psychosis population. Significant results confirming the microstructural changes are presented. GSBSS results show higher sensitivity to group differences between healthy and psychosis population in previously known regions.

7.1.2. Application II: Ageing effects of brain microstructure in white and gray matter

Aging effects on the white matter have been studied on large-scale datasets using UK Biobank with NODDI metrics [105]. With the increasing number of publicly available neuroimaging databases, there is an opportunity to combine large-scale imaging studies to increase the power of statistical analyses to test the common biological hypothesis. However, cross-study, cross-sectional analyses are confounded by inter-scanner variability that can cause both spatially and anatomically dependent signal aberrations. In particular, scanner related differences in the diffusion-weighted magnetic resonance imaging (DW-MRI) signal are substantially different for sub-cortical gray versus the neighboring white matter region or cortical gray matter region. There is a scope for studying brain microarchitecture features in gray matter using the approaches that are mentioned in earlier sections. We applied data harmonization techniques to correct for scanner differences and show that the biological variability with age is retained or improved while correcting for variability across scanners.

7.1.3. Application III: Sulcal curve labeling of the aging population on a large scale MRI

The human cortex is one of the most complex anatomical structures with substantial variation in shape across individuals. Despite its complexity, the cortical sulci are known as relatively stable regions that embed consistent cortical folds [106-108] by which the cerebral cortex can be subdivided into functionally and structurally homogeneous regions. From a morphological view, each sulcus (or sulcal region) can be well represented as a curve by tracing its sulcal fundus. Despite their advantages, there has not yet been a common agreement of a rigorous definition of the cortical sulci as they exhibit phenotypic variability. Such variability hampers consistent labeling of such cortical sulci and is challenging even for neuroanatomists. Moreover, creating a complete set of sulcal delineation is practically time-consuming

and requires qualified experts. In this study, we showed a new application of deep learning on cortical surfaces using a U-Net model. We achieved substantively faster labeling with better accuracy than multi-atlas approaches in an aging population.

7.2. Previous publications

Many contributions of this dissertation have been previously published. Empirical reproducibility of NODDI is evaluated and presented on a single subject with verification using the second subject for two-shell protocol and also for single shell ODI [109]. Gray matter surface based spatial statistics in diffusion microstructure approach with an application in psychosis is presented [110, 111]. Data harmonization of diffusion microstructure features in gray matter, and white matter is presented [112]. Constructing statistically unbiased cortical surface template using feature space covariance is presented [113]. Contributions related to improved GSBSS with search methodology and application to functional MRI (**Chapter IV**), and improved brain sulcal curve labeling using deep neural networks (**Chapter VII**) are currently in submitted status.

CHAPTER II. EMPIRICAL REPRODUCIBILITY, SENSITIVITY, AND OPTIMIZATION OF ACQUISITION PROTOCOL, FOR NEURITE ORIENTATION DISPERSION AND DENSITY IMAGING USING AMICO

1. Introduction

Diffusion weighted magnetic resonance imaging (DW-MRI) is an advanced imaging technique to characterize the diffusion of water in biological tissues non-invasively [1]. Diffusion tensor imaging (DTI) [2] is a DW-MRI technique that enables the measurement of the diffusion anisotropy of water in tissue for producing neuronal tract images and has become an established method for assessing white matter 'integrity' and connectivity in normal and unhealthy brain structure. However, DTI assumes that diffusion follows a tri-variate Gaussian distribution containing a single, coherently oriented bundle of white matter axons (e.g., the diffusion "ellipsoid") thus leading to non-specificity of underlying microstructure of neurite-specific measures such as their density and orientation dispersion. Microstructure imaging aims to address this limitation by fitting more descriptive models of tissue (e.g., axons, glial cells and extra-cellular space) that relate to specific parameters in different tissue compartments [3-12].

Neurite orientation dispersion density imaging (NODDI) [10] has emerged as a clinically feasible MRI technique for estimating the microstructural complexity by combining a three-compartment tissue model with a two-shell high-angular-resolution diffusion imaging (HARDI) protocol[10]. Briefly, NODDI provides quantification of the volume fraction of the intra-cellular compartment with restricted diffusion, extra-cellular compartment with hindered diffusion and isotropic compartment along with an orientation dispersion index (ODI) of the neurites. Intracellular volume fraction (V_{ic}) is modeled to represent the compartment defined by axons and dendrites, which are collectively called neurites. Extra-

cellular volume fraction (V_{ec}) refers to the space around the neurites that is occupied by various types of glial cells and cell bodies (somas) in gray matter. Isotropic volume fraction (V_{iso}) refers to the volume fraction of the cerebrospinal fluid compartment in each voxel[10]. Voxel-wise quantification of these metrics over the whole brain presents new opportunities for understanding brain development and disorders. To accelerate the fitting process, the accelerated microstructure imaging via convex optimization (AMICO)[13] analysis algorithm fits a linearized version of the NODDI model that greatly reduces the calculation time to approximately 1/40–1/80th of the original time. AMICO can achieve slightly more accurate and robust parameter estimates than the original technique and guarantees convergence to a global minimum[13].

NODDI has been applied in clinical applications including normal brain development and ageing [14-16], epilepsy[17, 18], traumatic brain injury[19], stroke[20], brain tumor[21], neonates[4], Alzheimer's disease[22], Parkinson's disease[23], functional connectivity predictors[24], multiple sclerosis [25, 26], age effects and sex differences [27], as it allows quantification of microstructure changes in both grey matter (GM) and white matter (WM). The originally proposed acquisition protocol is a two-shell sequence with a 711 s/mm^2 shell with 30 gradient directions and a 2855 s/mm^2 shell with 60 directions for maximum gradient (G_{max}) of 60 mT/m or b=2000 s/mm^2 shell for G_{max}=40 mT/m with 9 b=0 measurements. Yet, the acquisition protocols in the above referenced studies range from b-values ~300-1000 s/mm^2 (9-126 gradient directions) in the inner shell and ~2000 - 9375 s/mm^2 (30-126 gradient directions) on the outer shell b-values and are implemented on different scanner models and gradient strengths.

As more studies consider the NODDI model, it is essential to understand the empirical reproducibility and performance of this model with respect to scan time. Alexander et al. [28] designed an optimized protocol under an acquisition time constraint of 30 min that was considered for clinical viability in the original work [10]. While the scan-rescan analysis was performed on synthetic data in the original NODDI paper [10] over different acquisition schemes, the empirical data validation was limited to a single scan over four shells and focused on the number of shells and directions needed within the

optimized protocol range. Later work addressed scan-rescan reproducibility using a single subject on 3T across 2 sessions [29], within session reproducibility and between subject variability with magnetic strengths at 1.5T and 3T [30, 31] using the optimized protocol for NODDI. However extensive reproducibility analysis with empirical data across various acquisition protocols has not yet been performed on the NODDI model as has been done in the case of DTI [32-35]. Moreover, as mentioned in the original paper [10], ODI can be computed from single shell data. Similarly, the reproducibility of the estimated single shell ODI has not yet been evaluated systematically in empirical studies.

In this study, we perform the first evaluation of empirical reproducibility of AMICO/NODDI metrics for a two-shell HARDI protocol and characterize the impacts of protocol relative to acquisition time by comparing with a gold standard (extended scan time) dataset. We investigate how the number of directions in each shell and the selection of outer shell b-value (in a well-balanced DW scheme), affects NODDI-derived metrics through direct in vivo analyses of empirical data and provide optimized protocol settings to strike a practical balance between scan time and error. We perform similar analysis using single shell for ODI across all b-values. For reproducibility and sensitivity analysis, we scanned a single subject across multiple sessions including multiple repeats in each session as in [34] and analyze the results both qualitatively and quantitatively using the mean and standard deviation of bias within chosen ROI and report the root mean square error (RMSE) of each protocol with respect to the gold standard data. Recommended protocol settings are proposed based on the overlap of 95% confidence interval of each protocol with minimum RMSE across all the protocols within an ROI. These findings are validated with a dataset from second subject with a different scan protocol. Significant differences seen between the subjects and across protocols were compared using paired Wilcoxon ranksum tests and performance of each protocol with respect to gold standard is reported with Pearson pairwise correlation coefficient.

2. Methods

2.1. Data Acquisition

Diffusion weighted MRI data was acquired on a single healthy subject (S1) in three different

sessions, 2 hours each on a 3T scanner (Achieva, Philips Medical Systems, Best, The Netherlands) with a 32-channel head coil on consecutive days. The study protocol was performed four times in each of two sessions and three times in one session resulting in a total of 11 equivalent sets of scan data. The scan protocol included a B0 map, five diffusion shells with b-values of 3000, 2500, 2000, 1500 and 1000 s/mm^2, each with 96 diffusion-weighted directions. The data were acquired at 2.5mm isotropic resolution with a matrix of 96 x 96 and 38 slices using DW PGSE-EPI sequence (TR=2.65s, TE=101ms) and Gmax=~37.5mT/m with multiband (factor 2). Each scan was preceded by a scanner average of 10 minimally weighted ("b0") reference images. Scan time for each shell was ~4.7 minutes and the total scan time for each session was ~106 minutes/session. Second subject (S2) considered for validation is from Tractography-reproducibility Challenge with Empirical Data (TraCED) (https://my.vanderbilt.edu/ismrmtraced2017/publications/) whose scan protocol included three shells with b=1000 s/mm^2, 2000 s/mm^2, 3000 s/mm^2 with 20, 48 and 64 evenly distributed shells respectively and 15 b0 images (interspersed with 5 per shell one for each repeat). This data is acquired at 2.5mm resolution as well with a matrix of 96X96 and 44 slices (TR=2.92s, TE=99ms). Scan time of each shell with 48 gradient directions is ~2.25 minutes and equivalent scan time for 96 directions would be ~4.5 minutes. Total scan time for each session was ~99 minutes/session. Reported times are for multiband factor of 2 and can vary based on availability of simultaneous multi-slice and hardware capabilities. TR was fixed for all shells and TE held constant across scanners. Additionally, 5 reverse phase-encoded b0 images and 3 diffusion-weighted directions were acquired to aid in distortion correction. Cardiac and respiratory gating was not used. All data were acquired in accordance with the Vanderbilt University Institutional Review Board (IRB) guidelines and with the signed consent of the volunteer.

2.2. Preprocessing

Data from all sessions were corrected for patient movement, eddy current distortion and susceptibility distortion through FSL's topup and eddy functions [36]. The b-vectors were rotated by the rotation component from the transformation obtained from the correction.

Data from both intra- and inter-session scans were brought into the same target space by first registering the mean b0 images in scanner space to T2 in 2.5 mm isotropic MNI-152 space, thus obtaining an in-plane matrix of 78x93 and 75 slices using the FMRIB Software Library's (FSL 5.0) FLIRT [37]. Then, for each session, all b0's were combined to obtain a mean b0 image for individual sessions. The mean b0 of first session was registered to MNI T2 using 6 degrees of freedom (DOF) thus getting b0 into target space (b0_S1_T2_Reg). The mean b0 for the other two sessions were then registered to b0_S1_T2_Reg with 6 DOF. All individual b0's and diffusion-weighted images from the first session were brought to target space by applying the same transform. Accordingly, all b0's and diffusion-weighted images from the other two sessions were moved to target space using their corresponding transforms obtained from registration. All preprocessing was implemented with custom scripts using the FSL software (http://www.fmrib.ox.ac.uk/fsl).

All the b0 images from all sessions were combined together to form a mean b0 ($b0_m$) image in target space by compensating for scanner gain factors. Briefly, weights of each scan, W_s, were calculated by taking the inverse of the median of individual $b0_s$ images divided by the $b0_m$ image intensity to remove the scanner gain variations observed across scans. The MNI weighted mean b0 ($b0_{wm}$) image was generated using equation (2.1)

$$I(b0_{wm}) = \frac{\sum_{i=1}^{n}(w_i * \mathrm{I}(b0_s))}{\sum_{i=1}^{n} w_i} \tag{2.1}$$

Each DWI was then normalized using equation (2.2)

$$DWI_{norm} = I(b0_{wm}) * \frac{DWI_s}{I(b0_s)} \tag{2.2}$$

2.3. Segmentation:

In a multi-atlas segmentation framework, each target image was first affinely registered to the MNI305 atlas [38] and then N4 bias correction [39] using the Advanced Normalization Tools (ANTs) [40] software package was applied to the atlas and target images. Atlas images were non-rigidly registered to the target image with the ANTs package and the symmetric image normalization algorithm (SyN) [41]. Deformation of image and label volumes for the atlas to the target space was performed with

bi-cubic and nearest-neighbor interpolation respectively. The registered label volumes were then combined together with non-local spatial STAPLE [42, 43] and AdaBoost correction [44] was applied. Each voxel in the brain was then assigned to one of the 133 labels (including background) using the BrainCOLOR protocol [45]. T1 image labels were down-sampled using the multi-label interpolation technique in ANTs and the segmented image was brought back to the original target space by computing the ANTs inverse transformation. Then, all cortical gray matter (GM) labels were combined into one GM region of interest (ROI) while all white matter (WM) labels are combined to one WM ROI. Additional ROI images of the splenium of the corpus callosum (SCC), internal capsule (IC) and centrum semiovale (CS) were hand drawn using ITK-SNAP [46] for further analysis of white matter regions (Figure II-2b).

2.4. NODDI/AMICO Processing

The NODDI algorithm provides a unified model for gray and white matter microstructure. It uses a three compartment model representing restricted, hindered and isotropic components in each voxel and adopts the Watson distribution to model the orientation distribution of the sticks[10]. The full normalized signal (A), as described in the original work [10], is

$$A = (1 - V_{iso})(V_{ic}A_{ic} + (1 - V_{ic})A_{ec}) + V_{iso}A_{iso} \tag{2.3}$$

where A_{ic} and V_{ic} are the normalized signal and volume fraction of the intra-cellular compartment; A_{ec} is the normalized signal of the extracellular compartment; and A_{iso} and V_{iso} are the normalized signal and volume fraction of the CSF compartment. Orientation dispersion index is redefined from first proposed in [47] as below

$$OD = \frac{2}{\prod}\arctan\left({}^{1}/_{K}\right) \tag{2.4}$$

where K is the concentration parameter of the Watson distribution. The default diffusivities are fixed to their respective typical values in vivo: $d_{\parallel} = 1.7 \times 10^{-3}$ mm^2 s^{-1} and $d_{iso} = 3.0 \times 10^{-3}$ mm^{2}s^{-1} and the remaining parameters are estimated from the fitting procedure.

Data from each shell was split into multiple datasets of different number of directions, from 16 to

88 in increments of 8 for b=1000 s/mm^2 which were used as the inner shell and from 24 to 96 directions in all other shells that were considered outer shells (i.e., more directions were used in the outer shell than the inner shell). We used "orderpoints" [48] from Camino software [49] to get the desired subset of gradient directions such that a partial scan has evenly spaced gradient directions on a sphere. It takes a set of N gradient directions as inputs and searches for an ordering that minimizes the electrostatic energy of the first P (where $5 < P < N$) directions. It uses simulated-annealing optimization to search for the best configuration. . For each session and repetition, the inner shell data with b=1000s/mm^2 was combined with corresponding data in outer shells for different combinations of gradient directions. Outer shell gradient directions were chosen to be greater than the inner shell gradient directions, thus providing 220 unique combinations across outer shell b-values of 1500, 2000, 2500 and 3000s/mm2. With 11 repetitions across the sessions, this yielded a total of 2420 datasets that were considered for analysis. The NODDI model was applied and processed using the Python version of AMICO [https://github.com/daducci/AMICO/tree/master/python] on these datasets to estimate V_{ic}, V_{iso} and ODI parameters. As a gold standard, the data from all shells with all 96 directions were combined and processed through AMICO with the NODDI model to get the parameter maps V_{ic}_gs, V_{iso}_gs, and ODI_gs. To investigate the single shell ODI analysis, each shell was split into multiple datasets of various numbers of directions from 16 to 96 in increments of 8 for each session and repetition, across all b-values.

2.5. Statistical Analysis

Statistical analysis was carried out based on scan-rescan reproducibility across multiple repeats of each combination and performance of each of the NODDI parameters within an ROI, is assessed based on voxel-wise analysis. The whole brain voxel-wise mean (V_{mean}) and standard deviation (V_{std}) of each parameter was calculated across the 11 repeats for all combinations.

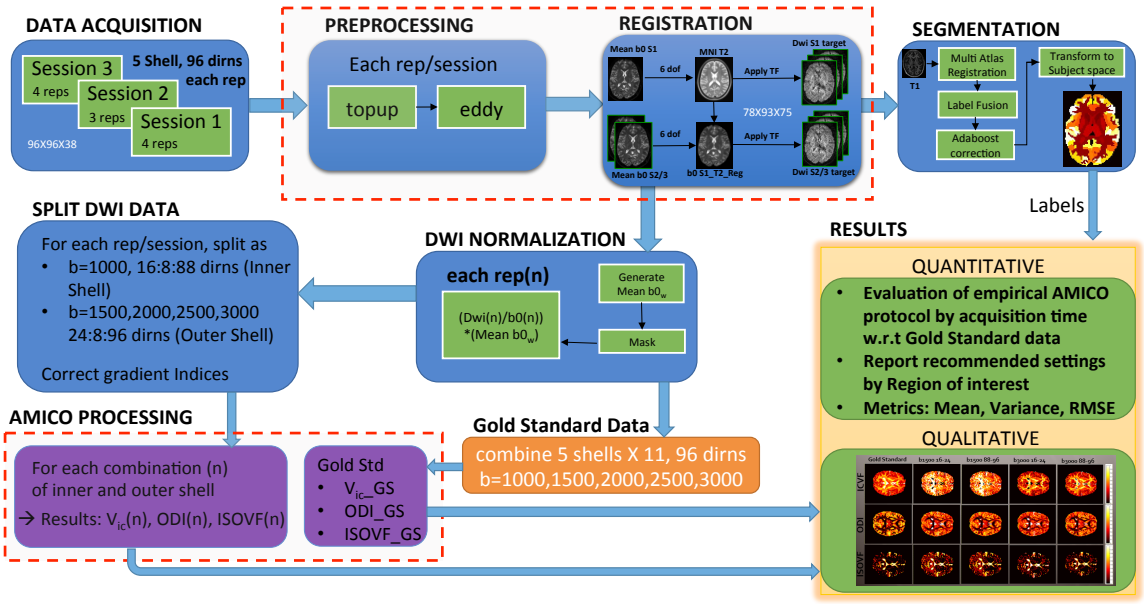

Figure II-1 Data acquisition included 3 sessions of 5 shells with b values of 1000, 1500, 2000, 2500 and 3000s/mm2 each with 96 gradient directions and a total of 11 repeats. Preprocessing included topup and eddy current correction after which all the images are registered to MNI template. Mean b0 is generated from these images and DWI normalization is performed. For each repetition and session, normalized DWI data is then split into n multiple combinations including inner and outer shells based on corrected gradient indices to account for uniform sampling. AMICO processing is performed on each of these combinations thus yielding metrics Vic(n), ODI(n) and ISOVF(n) for each combination. Data from all shells and repeats are combined to obtain full protocol that is considered to be the gold standard dataset. with Vic_GS, ODI_GS and ISOVF_GS as corresponding metrics from AMICO processing. Multi atlas label segmentation is performed on the T1 image and labels are obtained through label fusion method, which is used in further quantitative analysis. Results included both qualitative and quantitative metrics.

The voxel-wise bias (V_{bias}) was calculated by subtracting V_{mean} from the gold standard data for each parameter V_{ic}, V_{iso} and ODI as,

$$V_{ic}\_bias = V_{ic}\_gs - V_{ic}\_Vmean \quad (2.5)$$

$$V_{iso}\_bias = V_{iso}\_gs - V_{iso}\_Vmean \quad (2.6)$$

$$ODI\_bias = ODI\_gs - ODI\_Vmean \quad (2.7)$$

For evaluating the empirical reproducibility and sensitivity of each acquisition protocol, the following metrics were reported for each of the microstructure imaging parameters across all ROIs where n is the number of voxels in an ROI:

$$BiasM_{roi} = \frac{\sum_n (V_{bias})}{n} \quad (2.8)$$

$$BiasSD_{roi} = \sigma_n (V_{bias}) \quad (2.9)$$

$$\text{RMSE}_{\text{roi}} = \sqrt{\frac{\sum_n (V_{bias})^2}{n}} \tag{2.10}$$

Confidence intervals to describe RMSE values were calculated from $\boldsymbol{\chi^2}$ distribution table with (r-1) degrees of freedom where r is the number of repeats as in equation 2.11. Since we are considering the confidence interval of the error metric this approach is considered to be similar to that of calculating 95% confidence interval of the variance. For a 95% confidence interval, we get 2.5% of the area at either end of chi-square distribution. For S1 with 11 repeats we have 10 degrees of freedom and based on the chi-square table, we got $\chi^2_{0.975}$ =3.247 and $\chi^2_{0.025}$ =20.483 that are used in evaluating the confidence intervals based on below equations.

$$\sqrt{\frac{(r-1)*MSE}{\chi^2_{0.025}}} < RMSE < \sqrt{\frac{(r-1)*MSE}{\chi^2_{0.975}}} \tag{2.11}$$

Similarly for S2, we have 10 repeats thus yielding 9 degrees of freedom for which the chi-square table values are $\chi^2_{0.975}$ =2.7 and $\chi^2_{0.025}$ =19.023 that are used in evaluating 95% confidence interval for S2.

Within each ROI, minimum RMSE across all protocols is computed for individual NODDI parameters. We made use of the confidence intervals of the RMSE values in arriving at the recommended settings as it captures the possible range of RMSE values across multiple repeats. If the 95% confidence interval limit of any protocol overlaps with minimum RMSE value it is considered to be in the recommended range.

Performance of different protocols with respect to gold standard is calculated based on pairwise Pearson correlation coefficient and linear correlation coefficient (R) values are reported. Significant group differences across subjects are computed using Wilcoxon rank-sum test and corresponding p values are reported. P value < 0.05 is considered to be significant.

2.6. Second subject processing

A second subject (S2) was scanned with a slightly different scan protocol consisting of three shells with b=1000 s/mm^2, 2000 s/mm^2 and 3000 s/mm^2 and evenly distributed gradient directions of 20,

48, and 64 respectively. The data from these shells were split into multiple datasets similar to that of S1. Since we have fewer gradient directions in S2 for each shell, lower shell with b=1000 s/mm^2 was split to have 16 gradient directions to match with S1 while outer shells were split in the increments of 4 from 20 to 48 gradient directions for b=2000 s/mm^2 and 20 to 64 directions for b=3000 s/mm^2. Combining the inner shell data of b=1000 s/mm^2 with corresponding data in outer shells yielded 20 unique combinations for each session and repetition. With 10 repetitions across the 2 sessions, this yielded a total of 200 datasets that were considered for analysis. The NODDI model was applied and processed on these datasets to get V_{ic}, ODI and V_{iso} parameters. Data from all shells including all gradient directions for each session and repetition were combined and processed with NODDI model and the parameters V_{ic}_gs, V_{iso}_gs, and ODI_gs thus obtained were considered to be gold standard for second dataset S2.

Whole brain voxel-wise mean (V_{mean}) and standard deviation (V_{std}) of each parameter were calculated across the 10 repeats for all combinations. Voxel-wise bias (V_{bias}) with respect to gold standard was calculated using equations (2.5-2.8). For empirical reproducibility and sensitivity analysis of each acquisition protocol in S2, $BiasM_{roi}$, $BiasSD_{roi,}$ $RMSE_{roi}$ values were calculated using equations (2.8-2.10). To compare the reproducibility results across different subjects, a subset of data from S1 that match with unique combinations yielded from S2 data processing was considered for further analysis of results.

SNR was calculated by the difference method as indicated in [50]. First, an average image and a difference image were computed voxel wise with two consecutive b0 images. Then SNR was taken to be the mean of the average image divided by the standard deviation of the difference image within an ROI. This was repeated for all sequential b0 images within a session and a scan and the median of these measurements was reported for the corresponding SNR by ROI in Table 1. It is to be noted that the b0 reported below is based on the scanner average of 10 b0's. S1 and S2 represent the first and second subjects respectively.

3. Results

We examined the performance of microstructural measures obtained from the NODDI model

based on the acquisition protocol and characterized the impacts of acquisition time with respect to the extended scan time gold standard data. Evaluation was performed in terms of change in outer shell b-values and variation of gradient directions in inner and outer shells.

3.1. Qualitative results

A qualitative representation of microstructural parameters from a single session with respect to the gold standard across different acquisition protocols is indicated in Figure II-2a. Gold standard results are shown in column 1 where clear variation between WM and GM can be seen in both V_{ic} and ODI and V_{iso} is higher in cerebrospinal fluid (CSF) spaces. Columns 4 and 5 outlined with blue dotted lines were close to the proposed protocol from the original NODDI paper [10]. Lower shell of all the results is fixed at b=1000 s/mm^2. Number of gradient directions in lower shell and outer shell are indicated besides each b-value for each column. Column 2 has results from b=1500 s/mm^2 in outer shell with 24 gradient directions while inner shell has 16 gradient directions. Similarly, column 3 has results from b=1500 s/mm^2 as outer shell with 96 and 88 gradient directions respectively in outer and inner shells. Last 2 columns have results from b=3000 s/mm^2 as outer shell with the minimum (16 and 24) and maximum number of gradient directions (88 and 96) in inner and outer shells respectively. Residual images of V_{ic} and ODI, shown in rows 2 and 4, were obtained by subtracting the result image from gold standard of the corresponding metric. Most variation was observed in V_{ic} from overestimation at low b-values to underestimation at high b-values in the outer shell. However, qualitatively, there was little difference between minimum and maximum number of gradient directions chosen within a given outer shell b-value. ODI showed better performance (R=0.94) with maximum number of directions when compared to the minimum number of gradient directions(R=0.9), as expected, and also no significant variation was noticed across outer shell b-values (R=0.98).

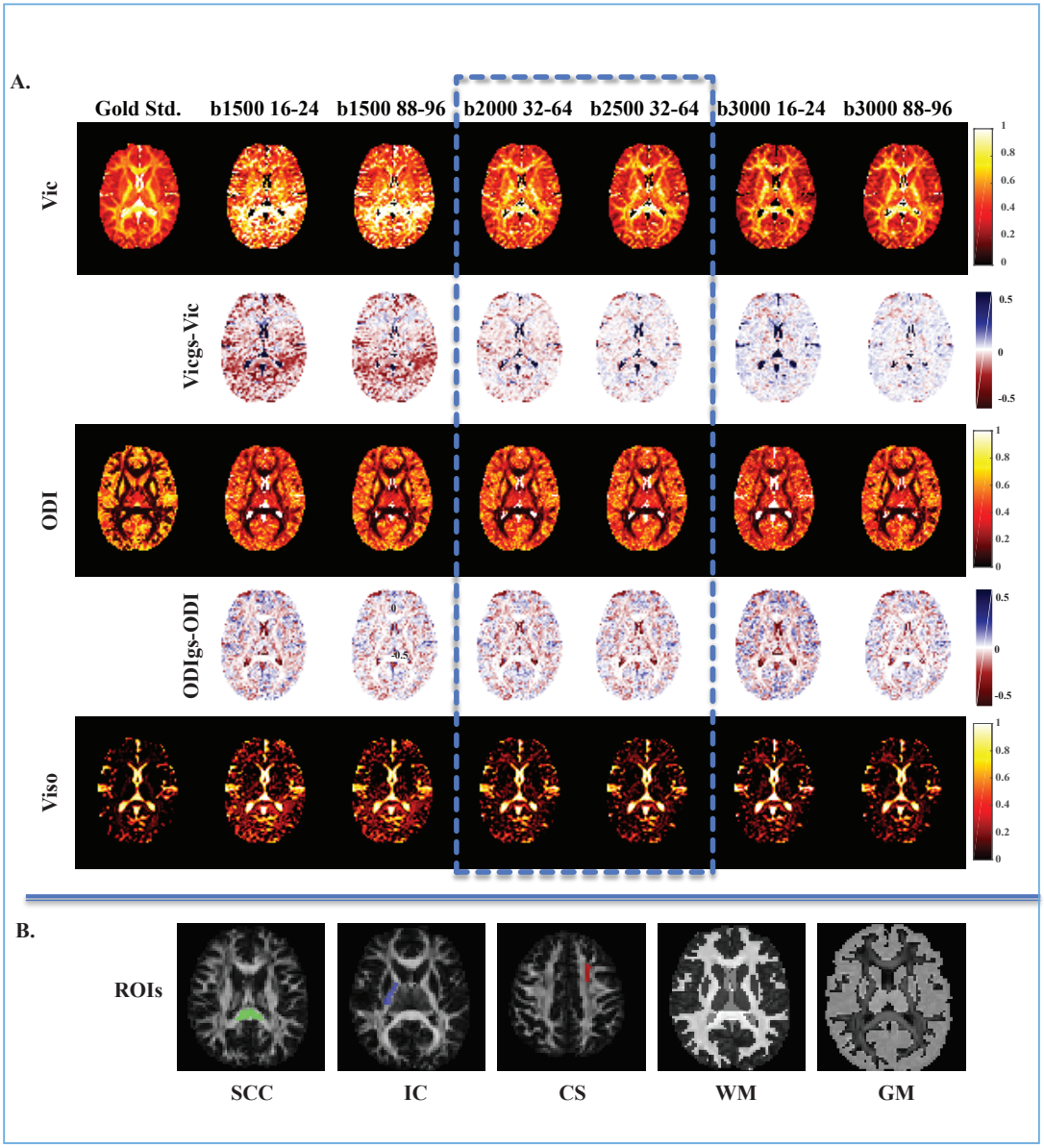

Figure II-2 Qualitative results for V_{ic}, ODI and V_{iso} maps and ROIs. (A) Qualitative results of NODDI maps with respect to gold standard data for different protocols with columns: (1) Gold Standard results with 5 shell data (2) Session results with b=1500 s/mm^2 (16,24) (3) Session results with b=1500 s/mm^2 (88,96) (4) Session results with b=2000 s/mm2 (32,64) (5) Session results with b=2500 s/mm^2 (32,64). (6) Session results with b=3000 s/mm^2 (16,24). (7) Session results with b=3000 s/mm^2 (88,96). All session data has fixed inner shell of b=1000 s/mm^2. Difference between gold standard and Vic and ODI parameters for corresponding protocols are represented in rows 2 and 4 respectively. Protocols with b=2000s/mm^2 and b=2500 s/mm^2 in outer shells similar to the proposed optimized protocol from original NODDI is highlighted in blue dotted line. (B) ROIs used in the study (a) Splenium of corpus callosum (SCC) (b) Internal capsule (IC) (c) Centrum semiovale (d) White matter (WM) (e) Gray matter.

One interesting observation with the minimum number of gradient directions was that the deviation of ODI from the gold standard was higher for b=3000 s/mm^2 (R=0.9) compared to lower outer shell b-values (R=0.91). V_{iso} showed better performance (R=0.95 at b=3000s/mm^2, R=0.88 at b=1500s/mm^2) with increased outer shell b-values, but is still considerably higher than the gold standard V_{iso}_gs. Figure II-2b shows the ROIs that were used for analysis in this study.

Qualitative results of ODI from a single shell with the highest number of gradient directions across b-values are presented in Figure II-3. As can be seen in the second row of Figure II-3, ODI values from single shell data were underestimated at lower b-values in GM regions but overestimated across the brain towards higher b-values. When compared with the gold standard, 2-shell ODI was always overestimated (row 3). When compared with 2-shell results (row 4), single shell ODI with b=2000 s/mm^2 was optimal where any b-value below lead to underestimation and any value higher produced overestimation.

3.2. Quantitative results

Empirical data were quantitatively evaluated by first analyzing the variation of estimated microstructure parameters V_{ic}, ODI and V_{iso} with respect to the gold standard for all combinations of acquisition protocols within WM, GM and CSF regions. A similar analysis was then extended to evaluate these parameters in specific white matter regions, SCC, IC and CS, which were chosen to represent a range of complexity of the underlying white matter fibers in those regions.

Bias statistics with respect to the gold standard dataset are presented in the following section. In each plot, the x-axis indicates acquisition time considering the number of gradient directions in both inner and outer shells.

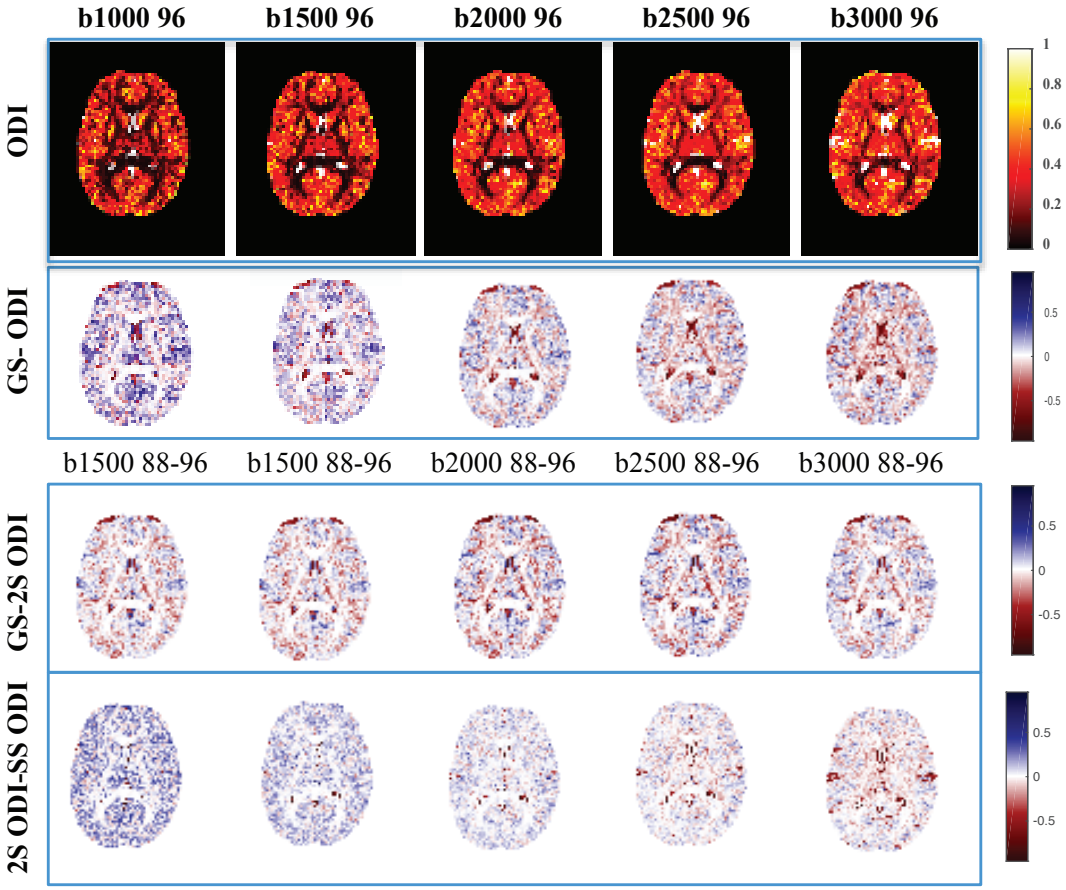

Figure II-3 Qualitative results with single shell ODI. (a) ODI from single shell data with 96 gradient directions for b=1000, 1500, 2000, 2500 and 3000 s/mm^2. (b) Difference ODI images from gold standard for single shell protocols as indicated (a). (c) Difference ODI images from gold standard for 2-shell protocols with b=1500, 2000, 2500 and 3000 s/mm^2 in outer shell with 88 and 96 gradient directions in inner and outer shells. (d) Difference ODI images from 2 shell and single shell protocols.

3.3. Whole Brain Analysis

Figure II-4a-b plots the bias statistics of V_{ic} for each combination within WM and GM regions with respect to the gold standard protocol. Based on the observed results, for WM outer-shell b-values of 2500 or 3000 s/mm^2 were optimal with b=2000 s/mm^2 also falling within the recommended range whereas b=1500 s/mm^2 showed the poorest performance. In case of GM, b=2500 s/mm^2 was optimal while b=2000 s/mm^2 and 3000 s/mm^2 were within the recommended range. Bias mean and standard deviation (provided in supplementary results) reflected similar information showing overestimation of V_{ic}

in both WM and GM with less than the optimal b-value and underestimation in GM with higher b-values. Error metrics were slightly improved with an increase in the number of gradient directions. Within GM, at lower number of gradient directions, b=3000/2000 s/mm^2 showed similar performance.

Figure II-5 a-b plots the bias statistics of ODI for each combination within WM and GM regions with respect to the gold standard protocol. Compared to higher b-values in outer shell, RMSE of ODI is low at lower b-values. However with increase in number of gradient directions, RMSE value is reduced for all b-values in both WM and GM In both cases, the ODI was better estimated with an increase in the number of gradient directions. However, based on the bias mean and standard deviation (details in supplementary results), ODI was always overestimated in WM. Also, lower numbers of gradient directions with highest b-values had the poorest performance, thus b=3000 s/mm^2 with (16,24) directions had the highest error as shown in qualitative images in Figure II-2a.

3.4. WM Regions Analysis

ROIs considered for analysis were chosen to be representative of the level of complexity of white matter regions based on the underlying tissue microstructure. SCC is highly anisotropic with a mean FA of 0.92. IC is another white matter region with a moderate FA of 0.73. CS, a complicated white matter region containing crossing and fanning fibers that has been used in many studies [30] for evaluating the performance of HARDI methods had a mean FA of 0.37. Reported FA values were calculated for the gold standard data.

Figure II-4 c-e plots the RMSE values of V_{ic} for each combination within SCC, IC and CS regions with respect to the gold standard protocol. For SCC, b=3000 s/mm^2 was found to be optimal followed by b=2500 s/mm^2 with lower b-values having the poorest performance stabilizing around 32 gradient directions for inner shell based on RMSE. In this region, b=1500 and 2000 s/mm^2 had same performance across different acquisition times. V_{ic} had lower RMSE for both b=2500 s/mm^2 and 3000 s/mm^2 in IC and highest error for b=1500 s/mm^2. Moreover, with an increase in number of directions, b=2000 to 3000 s/mm^2 showed similar performance.

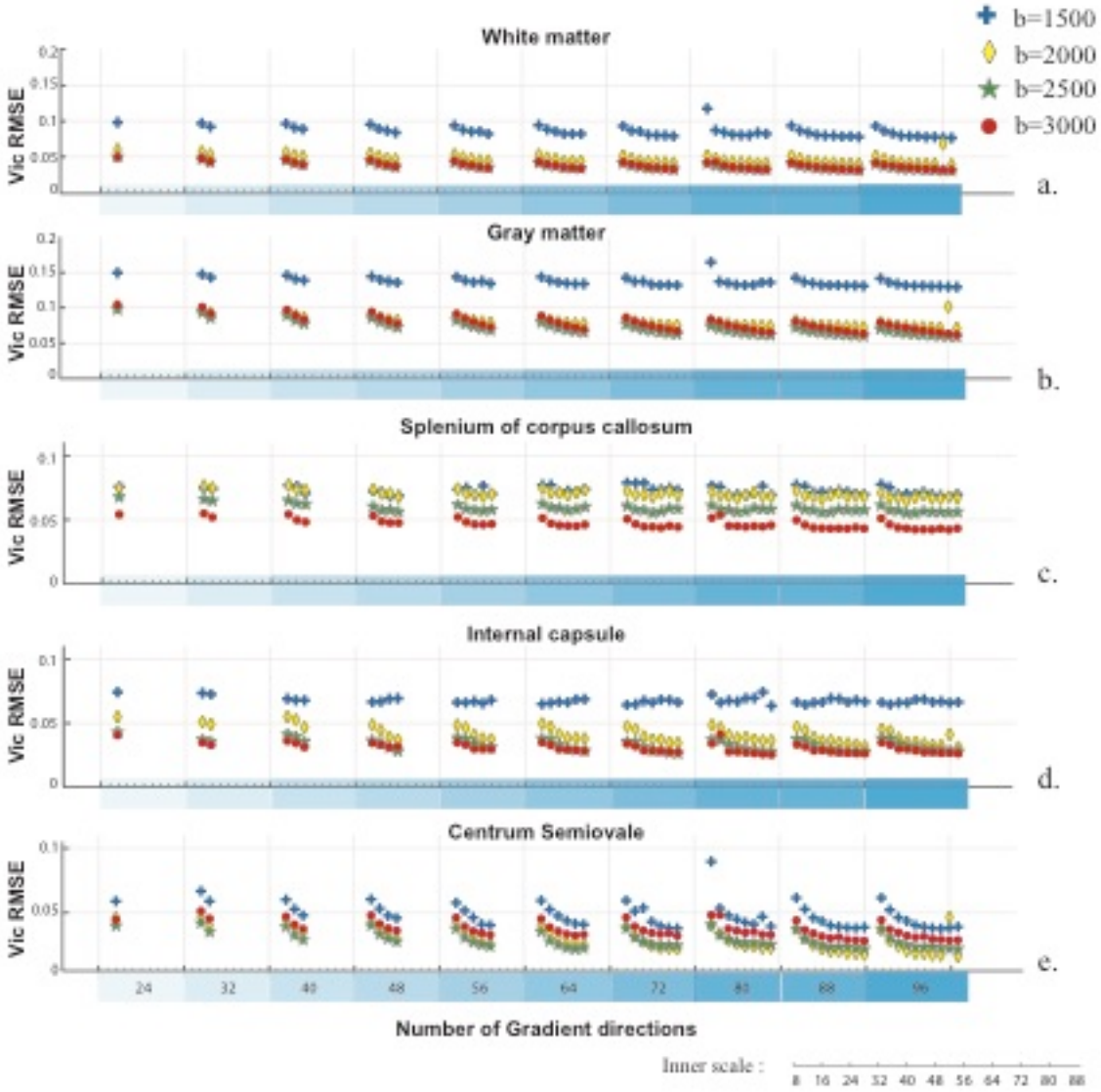

Figure II-4 The x-axis indicates the combination of inner and outer shells represented as acquisition time. Labels on x-axis within shaded region indicate the outer shell gradient directions while the inner scale within each shade indicates the inner shell gradient directions. Legend indicates the outer-shell b-values (a) RMSE of V_{ic} in WM (b) RMSE of V_{ic} in GM. (c) RMSE of V_{ic} in splenium of corpus callosum (d) RMSE of V_{ic} in internal capsule (e) RMSE of V_{ic} in centrum semioval

With low number of gradient directions, V_{ic} in CS was optimal for b=2500 s/mm^2 whereas b=2000 s/mm^2 displayed lower error with increase in both inner and outer shell gradient directions. Poorest performance in this region was seen with b=1500 s/mm^2 followed by b=3000 s/mm^2. However, based on the bias (provided in supplementary results), V_{ic} was underestimated for all protocols in SCC.

When compared to the gold standard, V_{ic} in IC was optimal for b=3000 s/mm^2, overestimated for both b=2500 and 2000 s/mm^2 and underestimated for b=1500 s/mm^2. For CS, V_{ic} was overestimated overall and b=1500 s/mm^2 had the highest standard deviation. Inner shell with 16 gradient directions showed the poorest performance across all b-values.

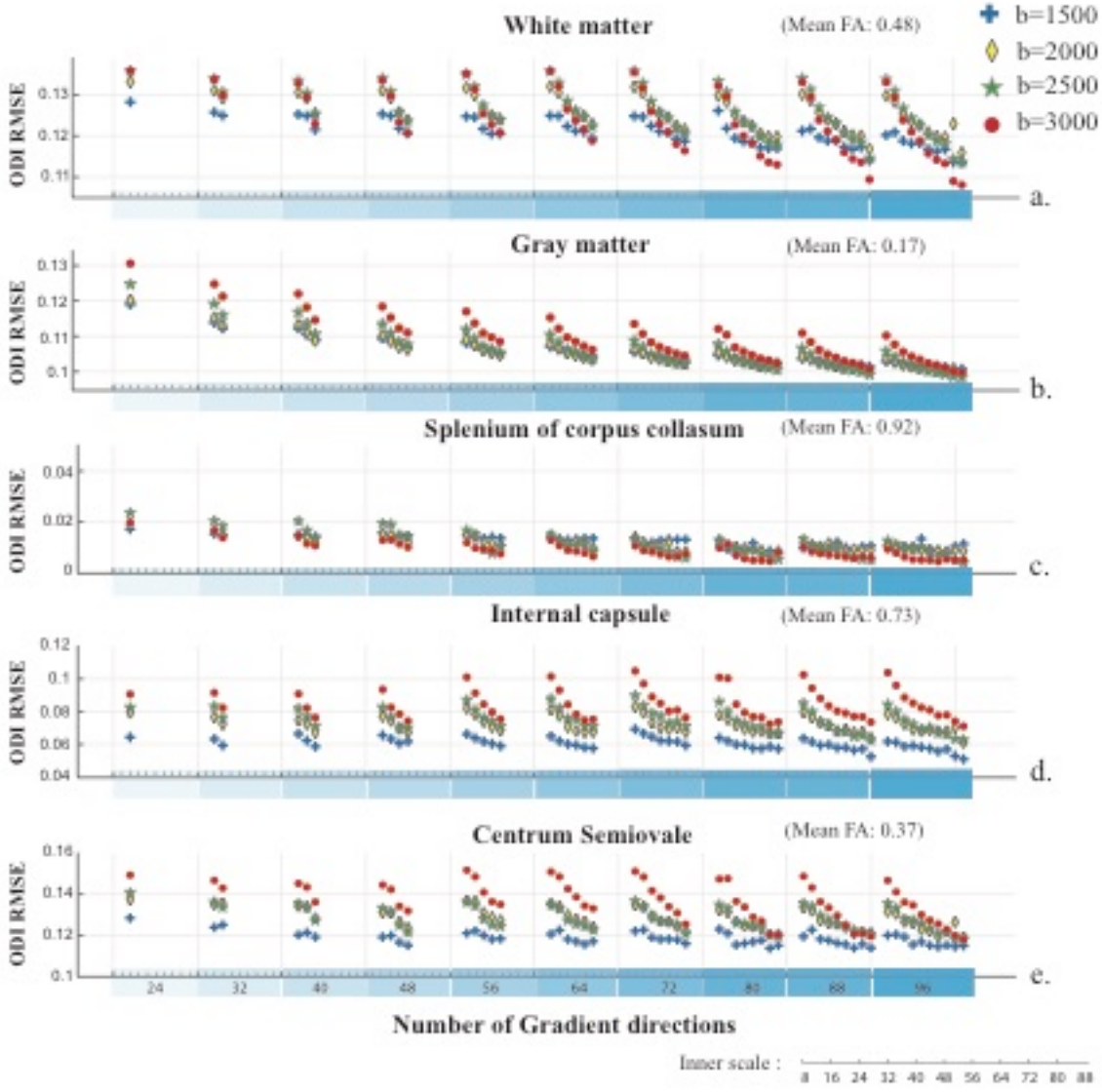

Figure II-5 RMSE of ODI across white matter regions of interest for multiple b-values and acquisition times. X-axis indicates the combination of inner and outer shells represented as acquisition time. Labels on x-axis within shaded region indicate the outer shell gradient directions while the inner scale within each shade indicates the inner shell gradient directions. Legend indicates the outer-shell b-values (a) RMSE of ODI in WM (b) RMSE of ODI in GM. (c) RMSE of ODI in splenium of corpus callosum (d) RMSE of ODI in internal capsule (e) RMSE of ODI in centrum semiovale. Mean FA values for each of the ROIs are indicated at the top of the figure.

Figure II-5c-e plots the bias statistics of ODI for each combination of acquisition protocol within SCC, IC and CS regions with respect to the gold standard protocol. In case of ODI within SCC, b=3000 s/mm^2 with higher number of gradient directions had better performance; however, b=1500 s/mm^2 performed consistently across all gradient directions which was also reflected with lower bias mean values but with high standard deviation. ODI was overestimated at lower number of gradient directions and slightly improved with increase in acquisition time/gradient directions. Within IC, ODI had lower RMSE with b=1500 s/mm^2 and poor performance with b=3000 s/mm^2 and improved with increased acquisition times. Overall ODI was overestimated and standard deviation increased with higher b-values. ODI in CS also showed similar performance with b=1500 s/mm^2 being optimal and b = 3000 s/mm^2 having highest error. Moreover, it was more sensitive to both outer and inner gradient directions.

3.5. Single-Shell ODI vs. Two-Shell ODI quantitative result

Figure II-6a-b plots the RMSE values of single and 2-shell ODI estimates across WM and GM. As seen in the qualitative results, lower b-values showed low RMSE compared to 2-shell, while high b-values in single shell were comparable with higher numbers of gradient directions. Figure II-6c-e plots the RMSE values of single and 2-shell ODI in WM regions of SCC, IC and CS. Overall ODI in SCC had lowest er-ror compared to the other ROIs, IC and CS. Yet, within SCC b=3000 s/mm^2 had the poorest performance. In IC, b=1000 s/mm^2 had lower error and increased with b-value and number of gradient directions after 54 seem to have similar performance. There is high variation in ODI values with higher b-values based on the bias mean and standard deviation for both single shell and 2 shell data.

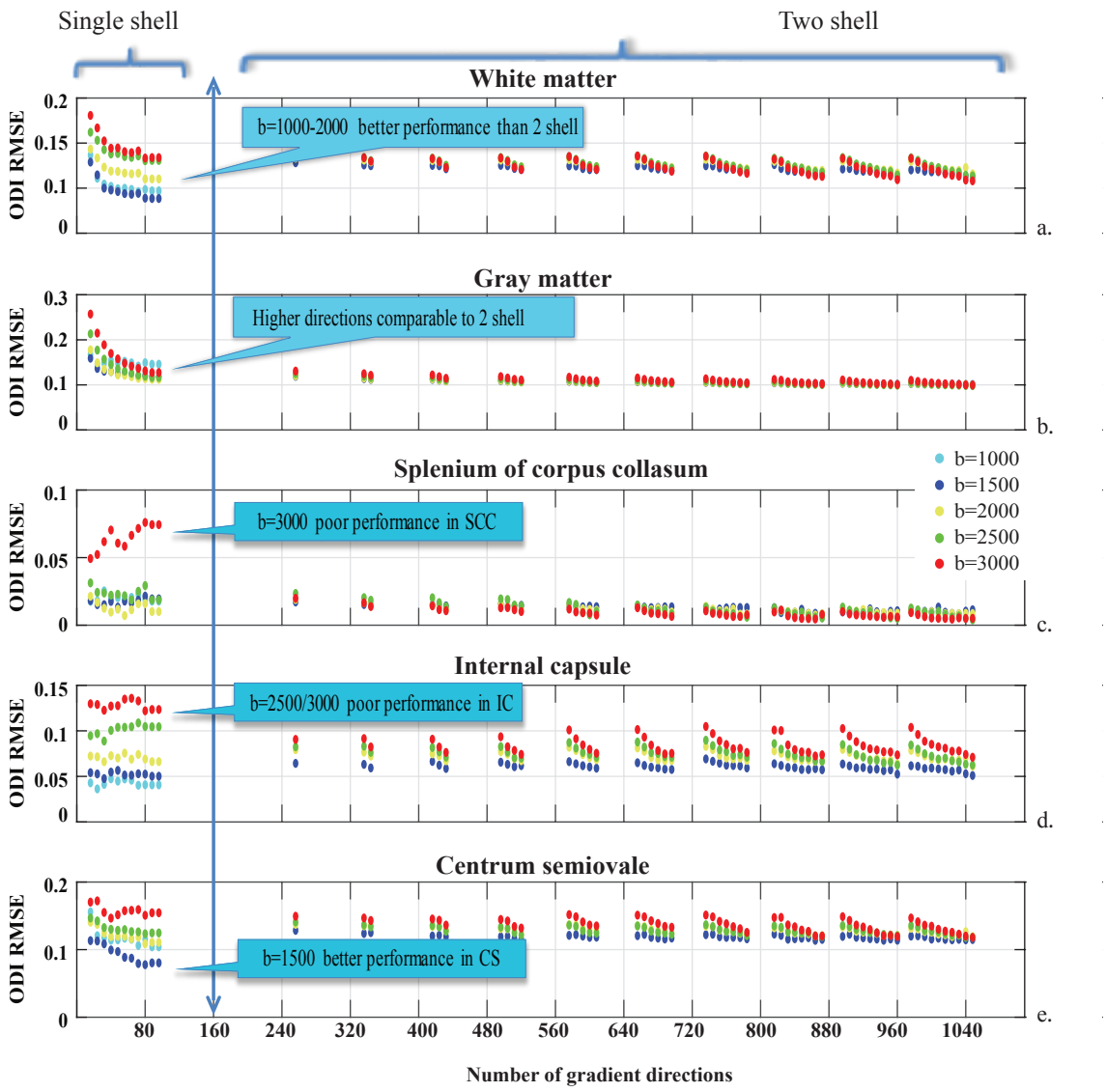

Figure II-6 Single and 2-shell RMSE of ODI. (a) White matter (b) Gray matter (c) Splenium of corpus collosum (d) Internal capsule (e) Centrum semiovale. Breakout boxes indicate qualitative results. Labels on x-axis shaded region indicate the combination of inner and outer shell gradient directions. For visual representation of results from both single and 2-shell scenario, these are scaled as inner shell gradient directions +10*outer shell gradient directions. For example, the combination of 88 directions in inner shell and 96 in outer shell is represented with 88+10 ⍰ 96=1048 on x-axis.

3.6. Second healthy subject analysis results

RMSE values for S1 and S2 were used in plotting the performance of each protocol setting for individual subjects based on the unique combinations of b-values and combined gradient directions from inner and outer shells. To represent the results on the same scale, each RMSE value is divided by the

minimum RMSE value ($RMSE_{min}$) within the region of interest for each subject. Overall RMSE value in S2 is higher compared to S1, which could be due to the lower SNR of data acquisition in S2 (Table 1). Minimum RMSE value associated with each subject for corresponding ROIs are provided in supplementary material. For both the subjects, ratios of RMSE value with respect to $RMSE_{min}$ for V_{ic} were plotted in Figure II-7 in different ROIs. RMSE ratio for ODI in different ROIs for different combinations of b-values and combined gradient directions were plotted in Figure II-8. As b=3000 s/mm^2 shell had 64 gradient directions, we have 12 total unique combinations with total gradient directions (36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80) whereas there are 8 unique combinations (36, 40, 44, 48, 52, 56, 60, 64) for b=2000 s/mm^2 with 48 gradient directions when combined with inner shell of 16 gradient directions.

X-axis in Figure II-7 and II-8 indicates combined (inner + outer shell) gradient directions with fixed inner shell of 16 directions. Y-axis is the ratio of RMSE values divided by the $RMSE_{min}$ for each subject by ROI. Black line indicates the minimum ratio of 1 for $RMSE/RMSE_{min}$. Blue and green indicate b=2000 s/mm^2 and b=3000 s/mm^2 for S2, orange and yellow indicate b=2000 s/mm^2 and b=3000 s/mm^2 for S1. Overall trend between the two subjects based on varying gradient directions in outer shell was matched in most of the ROIs for both V_{ic} and ODI ($p<0.001$) in spite of scan differences and SNR differences. Recommended protocol settings based on 95% confidence interval overlap of RMSE value with RMSE min for each ROI and NODDI parameter are listed in Table 4. The trends seen in Figure II-7 matches with the recommended settings where there is higher RMSE ratio of V_{ic} in SCC for b=2000 s/mm^2 for S1 and higher RMSE ratio of ODI in SCC for b=3000 s/mm^2.

4. Discussion

The findings confirmed [10] the sensitivity of V_{ic} to the choice of outer shell b-value in contrast to ODI, which was more sensitive to the number of gradient directions. They suggested that the choice of b-value and number of gradient directions in conventional NODDI acquisition could further be optimized depending on the specific tissue type and region of interest or targeted pathologies.

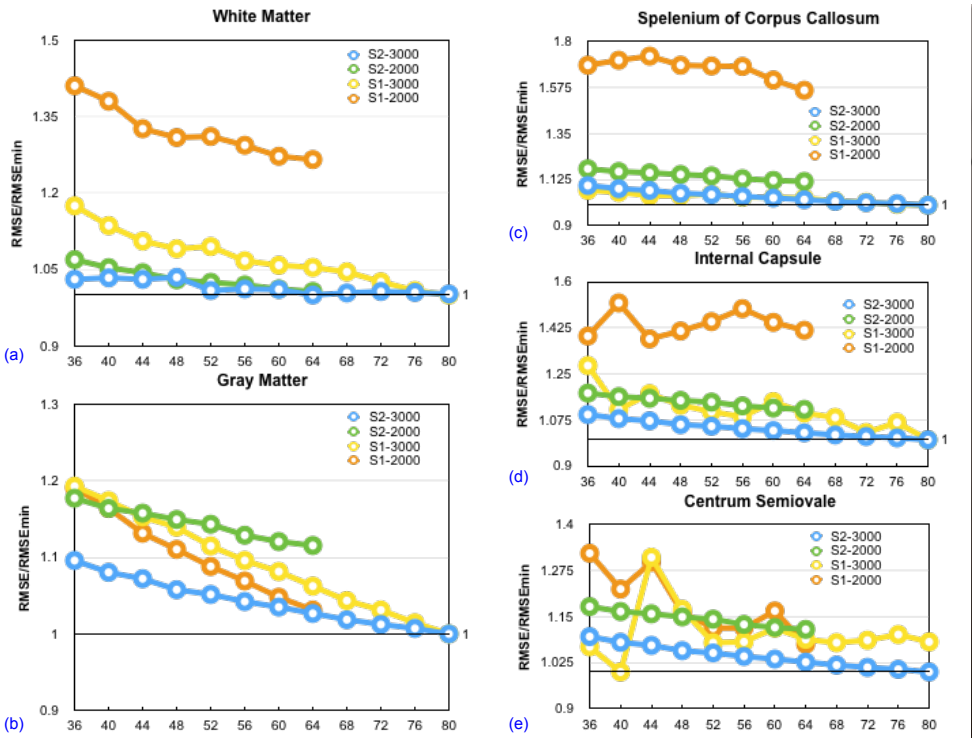

Figure II-7 RMSE values in Vic for the two subjects S1 and S2 across ROIs for multiple b-values and acquisition times. X-axis indicates combined (inner+outer shell) gradient directions with fixed inner shell of 16 directions. Y-axis is the ratio of RMSE values divided by the $RMSE_{min}$ for each subject by ROI. Black line indicates the minimum ratio.

However, it is to be noted that caution must be taken in interpreting NODDI indices across studies with different acquisition schemes. Unlike V_{ic}, single shell ODI results were comparable to 2 shell results and a recent study [51] also confirmed that the single shell ODI analysis was able to reproduce the group differences from multi-shell analysis (Figure II-3).

V_{ic} in white matter and gray matter was in agreement with the results from the optimized protocol for NODDI [10]. However, within WM ROIs, V_{ic} was less sensitive to gradient directions in SCC and IC compared to CS. This could be because of the complexity of the underlying microstructure in CS. Thus, additional care should be taken when considering the protocol selection in such regions (Figure II-4). Interestingly, V_{ic} at b=1500 s/mm^2 was less sensitive to gradient directions compared to higher b-values in the IC region.

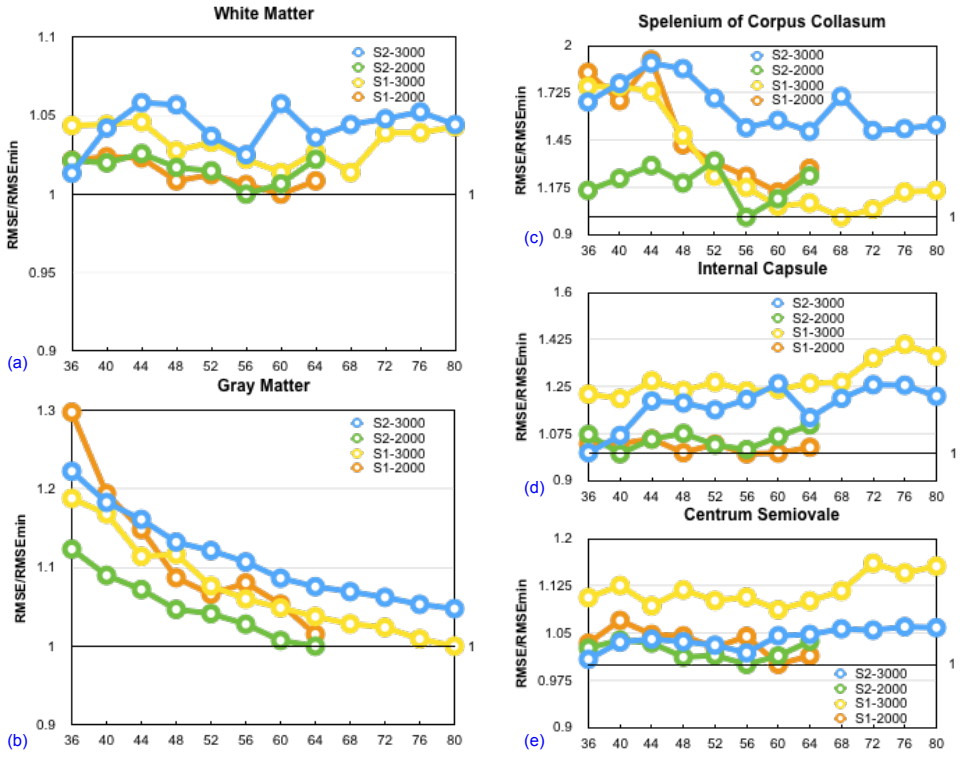

Figure II-8 RMSE values in ODI for two subjects S1 and S2 across ROIs for multiple b-values and acquisition times. X-axis indicates combined (inner+outer shell) gradient directions with fixed inner shell of 16 directions. Y-axis is the ratio of RMSE values divided by the $RMSE_{min}$ for each subject by ROI. Black line indicates the minimum ratio and dotted red line 5% range from the minimum value.

ODI in white matter was more sensitive to the number of directions compared to gray matter (Figure II-5). ODI represents spatial configuration of the neurite structures thus by increasing the number of directions it could capture the variation in the underlying complex structure. This is apparent with the variation seen within WM ROIs with least variation in SCC to highest variation in CS by changing the number of directions. ODI was highly dependent on the underlying complexity of the WM ROI. SCC results were similar to WM, as expected, which can be attributed to its high anisotropy. In IC and CS, ODI was optimal with b=1500 s/mm^2 and the error showed an increasing trend with increases in b-values. In IC, ODI was mainly dependent on inner shell directions, consistently improving with increases in number of inner shell directions. However in CS, it showed dependency on both inner and outer shell directions with 16 and 24 directions in inner shell having the poorest performance. Thus for considering

the analysis in CS, the inner shell acquisition should at least have 32 directions for better performance.

Single shell ODI results comparison with 2-shell results yielded some interesting findings as shown in Figure II-6. In WM, lower b-values showed lower RMSE compared to both single shell higher b-values and 2-shell results. Based on the WM ROIs, this performance was reflected in IC and CS regions. Moreover, in SCC 2-shell protocol always showed better performance compared to all single shell protocols. The implication of ODI variation within WM regions needs to be carefully reviewed and additional care should be taken when interpreting ODI results from these regions based on the protocol settings. In gray matter, ODI results were largely consistent across b-values with improved performance with increase in the number of gradient directions. However, 2-shell data always had better performance even when compared with the highest number of gradient directions in a single shell in gray matter.

4.1. Optimized NODDI Protocols

Recommended protocol settings based on acquisition time and performance by error statistics are presented in Tables 2, 3 and 4. Based on Table 2, when considering both V_{ic} across all ROIs, b=2500 s/mm^2 with highest number of gradient directions (>104) was optimal whereas for ODI, b=2500/3000 s/mm^2 was optimal with total gradient directions (>128). Considering the scan time equivalent (STE) of ~4.5 minutes for 96 directions, the corresponding minimum scan time for each scenario is indicated in parentheses in the last column. Considering the performance of different NODDI acquisition parameters in all ROIs, overall recommended settings for acquisition is listed to have b-value of 2500 s/mm^2 with at least 128 total gradient directions.

A subset of original scan protocol is considered in further analysis for comparing with a second subject S2 from TRACED dataset. The data and analysis was carried out separately for two subjects under consideration. Even though we are comparing the overall trend between the subjects, and added the second subject for validity/reproducibility check of the results, these results are not combined into a quantitative comparison. Also the idea is to confirm similar trend of performance in NODDI across different data acquisitions.

Table II-1: Recommended protocol settings by acquisition time are reported based on RMSE values with respect to gold standard. $Bias_{min}$ column indicates the acquisition parameters (outer shell b-value, inner shell gradient directions, outer shell gradient directions) that had lowest bias mean. $RMSE_{min}$ column indicates the acquisition parameters (outer shell b-value, inner shell gradient directions, outer shell gradient directions) that had lowest RMSE. Last two columns were based on 95% confidence interval overlap of RMSE value with RMSE min for each ROI and NODDI parameter. Rows highlighted in gray under each parameter indicate recommended settings based on all the ROIs. Overall recommended scan protocol based on all parameters is highlighted in green in the last row. Acquisition time in the last column is sum of gradient directions in both shells (inner shell gradient directions + outer shell gradient directions). Minimum scan time based on scan time equivalent of ~4.5 minutes for 96 directions is indicated in the parenthesis.

Parameter	ROI	$Bias_{min}$	$RMSE_{min}$	Recommended	Acquisition time by
				b-value	total gradient directions
ICVF	WM	1500,16,80	2500,88,96	[2000;2500;3000]	>=56 (~2.5 mins)
ICVF	GM	1500,16,80	2500,88,96	[2000;2500;3000]	>=64 (~3 mins)
ICVF	CSF	2000,80,96	2000,80,96	[1500;2000;2500;3000]	>=48 (~2 mins)
ICVF	SCC	3000,16,96	3000,80,96	[2500;3000]	>=40 (~ 2mins)
ICVF	IC	2000,16,24	3000,72,80	[2000;2500;3000]	>=48 (~2 mins)
ICVF	CS	1500,16,80	2000,88,96	[2000;2500]	>=104 (~5 mins)
ICVF - REC				2500	>=104 (~5 mins)
ODI	WM	2500,16,[64;72]	3000,88,96	[1500;2000;2500;3000]	>=40(~2 mins)
ODI	GM	2000,80,96	2500,88,96	[1500;2000;2500;3000]	>=40(~2mins)
ODI	CSF	3000,16,24	2000,80,96	[1500;2000;2500;3000]	>=72 (~3 mins)
ODI	SCC	3000,16,24	2500,88,96	[2500;3000]	>=128 (~6 mins)
ODI	IC	3000,16,72	1500,88,96	[1500;2000;2500;3000]	>=40(~2 mins)
ODI	CS	3000,16,56	1500,64,80	[1500;2000;2500;3000]	>=40(~2 mins)
ODI - REC				2500/3000	>=128 (~6 mins)
ISOVF	WM	1500,16,24	3000,88,96	[1500;2000;2500;3000]	>=40(~2 mins)
ISOVF	GM	1500,16,80	3000,88,96	[2500;3000]	>=40(~2 mins)
ISOVF	CSF	2500,16,24	2000,80,96	[1500;2000;2500;3000]	>=88(~4 mins)
ISOVF	SCC	3000,16,24	3000,40,48	[1500;2000;2500;3000]	>=40(~2 mins)
ISOVF	IC	3000,24,80	2500,80,88	[2000;2500;3000]	>=56 (~2.5 mins)
ISOVF	CS	3000,16,32	2000,80,96	[1500;2000;2500]	>=96(~4.5 mins)
ISOVF - REC				2500	>=96 (~4.5 mins)
Overall REC				**2500**	**>=128 (~6 mins)**

Recommended settings are in agreement between the two subjects based on most of the ROIs as we reach higher number of gradient directions except for ODI in IC and CS where the trend seem to be

favoring lower number of gradient directions. RMSE of V_{ic} in b=2000 s/mm^2 for S1 is higher compared to that of b=3000 s/mm^2. Thus, even though it matched the overall trend when compared to S2, it was not within the recommended settings range. There is a general decrease in RMSE of V_{ic} with increasing gradient directions. However, when looking at the crossing fiber regions of IC and CS, S1 showed uncertain pattern with varying gradient directions specially in b=2000 s/mm^2 thus favoring higher b-values and higher gradient directions.

Table II-2 Recommended protocol settings by acquisition time for single shell data are reported based on RMSE values with respect to gold standard. $Bias_{min}$ column indicates the acquisition parameters (b-value, gradient directions) that had lowest bias mean. $RMSE_{min}$ column indicates the acquisition parameters (b-value, gradient directions) that had lowest RMSE. Last two columns were based on 95% confidence interval overlap of RMSE value with RMSE min for each ROI in ODI. Overall recommended scan protocol based on all ROIs is highlighted in green in the last row. Acquisition time in the last column is sum of gradient directions in both shells (inner shell gradient directions + outer shell gradient directions). Minimum scan time based on scan time equivalent of ~4.5 minutes for 96 directions is indicated in the parenthesis.

Parameter	ROI	$Bias_{min}$	$RMSE_{min}$	Recommended b-value	Acquisition time by total gradient directions
ODI	WM	3000,24	1500, 96	1000,1500,2000	>=24 (~1.5 mins)
ODI	GM	3000,16	2000,96	1000,1500,2000, 2500,3000	>=24 (~1.5 mins)
ODI	SCC	3000,96	1000,16	1000,1500,2000	>=16 (~1 min)
ODI	IC	3000,24	1000,96	1000,1500	>=16 (~1 min)
ODI	CS	3000,24	1500,80	1500	>= 40 (2 mins)
Overall REC				**1500**	**>=40 (~ 2mins)**

ODI is lower in white matter regions (highly aligned axons) compared to crossing fiber regions within WM or that of gray matter (highly dispersed neuritis) where there is higher dispersion [10]. Within ODI, even though there is not a general trend in WM region with varying gradient directions, most of the scan protocol combinations fell within the 95% confidence interval overlap with respect to minimum RMSE value. Similar uncertainty is seen in highly oriented SCC region but most of the scan protocol combinations fall outside the recommended setting range making it an unreliable metric in this region.

Table II-3 Recommended protocol settings by acquisition time are reported based on RMSE for each protocol for both the subjects S1 and S2 with respect to gold standard. These settings were based on 95% confidence interval overlap of RMSE value with RMSE min for each ROI and NODDI parameter. Rows highlighted in gray under each parameter indicate recommended settings based on all the ROIs. Overall recommended scan protocol based on all parameters is highlighted in green in the last row. Acquisition time is sum of gradient directions in both shells (inner shell gradient directions + outer shell gradient directions). Minimum scan time based on scan time equivalent of ~4.5 minutes for 96 directions is indicated in the parenthesis.

Parameter	ROI	Recommended B-value – S1	Acquisition time by Gradient directions – S1	Recommended B-value – S2	Acquisition time by Gradient directions S2
ICVF	WM	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ICVF	GM	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ICVF	CSF	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ICVF	SCC	3000	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ICVF	IC	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ICVF	CS	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ICVF - REC		3000	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ODI	WM	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ODI	GM	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ODI	CSF	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ODI	SCC	[2000;3000]	>=48 (~2 mins)	2000	>=36 (~2 mins)
ODI	IC	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ODI	CS	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ODI - REC		[2000;3000]	>=48 (~2 mins)	2000	>=36 (~2 mins)
ISOVF	WM	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ISOVF	GM	3000	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ISOVF	CSF	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ISOVF	SCC	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ISOVF	IC	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ISOVF	CS	[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
ISOVF - REC		[2000;3000]	>=36 (~2 mins)	[2000;3000]	>=36 (~2 mins)
Overall REC		**3000**	**>=48 (~2 mins)**	**2000**	**>=36 (~2 mins)**

In GM region, higher gradient directions in both b=2000 s/mm^2 and b=3000 s/mm^2 yield low RMSE values. Moreover, with high b-value, more number of gradient directions were needed to yield lower error. In crossing fiber regions of IC and CS, higher gradient directions were leading to higher error, thus making this metric most optimal with lower gradient directions in these regions.

One of the limitations of this study is the use of same subject for repeating measurements instead of acquiring data on multiple subjects that could have lead to higher confidence and impact. But because of the duration of scan time and availability of willing subjects to participate in such lengthy protocol we have limited this to a single subject study. However we added second subject to validate the trend observed in NODDI protocol performance. As shown in the results irrespective of the scan protocol differences the two subjects are in agreement with the recommended settings range for NODDI acquisition.

5. Conclusion

The performance of the microstructure based NODDI model's estimated parameters was evaluated for a range of acquisition protocol parameters. These results are based on a sub-sampling scheme, which assumes that a subset of a reduced number of gradient directions is largely equivalent to the same number of directions directly achieved on an MR-scanner. It was observed that the performance of these parameters was sensitive to the acquisition protocol based on the regions of interest. We provide guidelines for the selection of optimal protocols based on the type of study and acquisition time. Further work could aim at extending this to inter-session, intra-session and inter-site variability using the rich dataset that was acquired in the current project. Additional error metrics could also be considered in evaluating robust performance across these protocols. All work was performed based on the default settings offered in NODDI and AMICO, and could be extended to evaluate adjustment of the regularization parameters in optimization by acquisition time. Another limitation of this work is the number of ROIs that were included in the study, future studies could aim at more representative regions that were used in the literature based on the white matter tracts such as the superior longitudinal fasciculus (SLF), etc.

CHAPTER III. GRAY MATTER SURFACE BASED SPATIAL STATISTICS (GS-BSS) IN DIFFUSION MICROSTRUCTURE

1. Introduction

Gray matter based spatial statistics method (GBSS) was first proposed in studying the cortical microstructure development of preterm human brain [1]. This method adapted tract-based spatial statistics (TBSS) [2] framework to overcome partial volume contamination in diffusion measures within cortical microstructure where spatially transformed individual cortical measurements are projected onto the mean cortical skeleton by searching in a direction perpendicular to the cortical skeleton identifying maximally probable cortical voxels. Unlike TBSS where projection was based on high fractional anisotropy (FA), GBSS method seeks voxels with a high probability of gray matter (GM). This approach was customized as NODDI-GBSS (will be called as N-GBSS in rest of the chapter for ease of reference), in a recent study [3] to take advantage of indices provided in NODDI in segmenting the GM instead of using T1 weighted images. GM concentration is then used along with FA to construct T1 like the image in diffusion space that is then used to normalize to standard space (Montreal Neurological Institute, MNI). Thinning of the voxels based on the TBSS approach is advantageous in improving the sensitivity of diffusion measures in GM; however it could lead to missing some of the significant regions of interest within GM. Also, registration approach followed in N-GBSS has the possibility of compromising on the sensitivity of voxel-based correspondence when compared to using better registration and segmentation methods. Caveats associated with using TBSS based approach have been presented in a recent literature [4] some of which could hold for GBSS approaches as well.

In this chapter, we present GM surface based spatial statistics (GS-BSS) approach where GM central surface is derived from structural scan using Multi-Atlas Cortical Reconstruction Using Implicit

Surface Evolution (MaCRUISE) [5] method. MaCRUISE has been shown to have superior performance in achieving surface accuracy by combining multi-atlas segmentation with cortical surface reconstruction. Structural images are segmented and normalized to MNI template space with diffeomorphic anatomical registration using exponentiated lie algebra (DARTEL) method [6]. Also, instead of using TBSS based approach, smooth correspondence between cortical surfaces is obtained with a diffeomorphic spectral matching algorithm [7] proposed for cortical surfaces. This method retains the speed advantage of spectral matching methods while assuring smooth correspondence between surfaces by performing spectral analysis of the joint graph Laplacian after an initial surface match. To validate our results, we have applied this method for identifying underlying microstructural changes in gray matter regions between healthy and psychosis groups.

Based on post-mortem studies, microstructure changes are exhibited in the prefrontal cortex, in individuals with psychosis [8, 9]. Recent in vivo studies have demonstrated the ability of NODDI technique in identifying the group differences between healthy and psychosis groups in white matter [10] and gray matter (GM) regions [3, 11]. Using the NODDI model, neurite orientation dispersion index (ODI), a putative marker of dendritic structure and complexity, was calculated and compared between healthy controls and individuals with psychosis.

When compared to N-GBSS, GM surface-based approach has yielded 1) double the number of most probable gray matter voxels within cortical regions that can be used for further analysis, 2) better sensitivity in VBM analysis with registration based on structural image, and 3) significant regions that are reported to have altered microstructure in psychosis group from post mortem studies.

2. Theory and implementation

Flowchart of the approach followed in our method is described in Figure III-1 where registration and segmentation steps are performed on T1 weighted image.

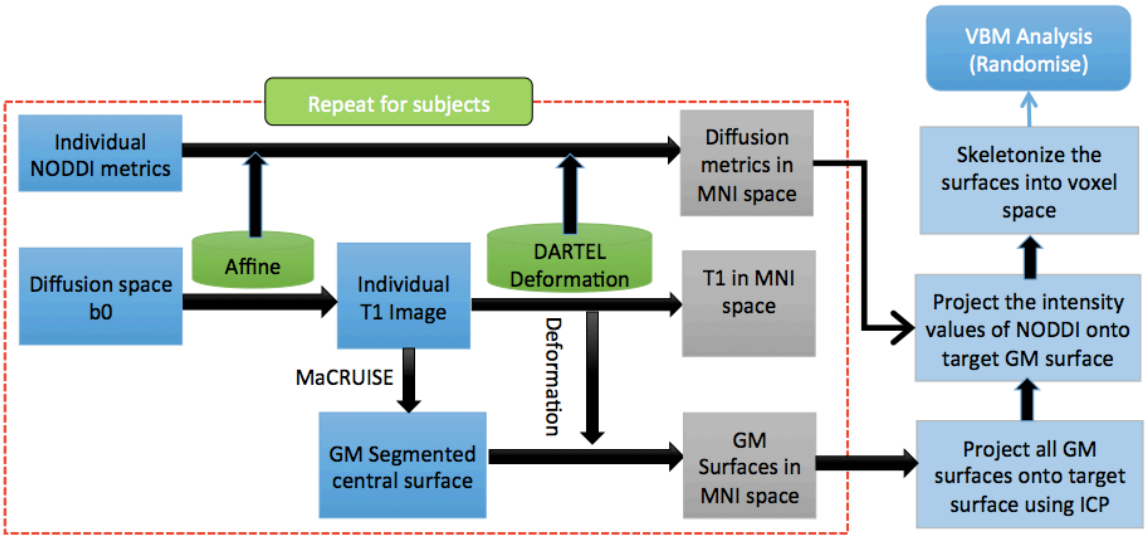

Figure III-1 Data processing flowchart.

2.1. Registration

Each subject's structural scan was segmented into gray matter, white matter, and cerebral-spinal fluid (CSF) tissue classes using the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm/) for SPM12 (http://www.fil.ion.ucl.ac.uk/spm). After an initial affine registration of the DARTEL [6] templates to the corresponding tissue probability maps in MNI space (http://www.mni.mcgill.ca/), non-linear warping of the segmented images were then performed to match the corresponding MNI space DARTEL templates (GM, WM). Forward and inverse deformation fields are saved that can be used in further processing.

2.2. Segmentation

Individual T1 images are segmented and gray matter surfaces are derived using MaCRUISE as described in [12] where inner, central and outer cortical surfaces are reconstructed by using the topology-preserving geometric deformable surface model. Gray matter central surface is then normalized to MNI space by applying inverse deformation field to the vertices.

2.3. Surface matching approach

As the cortical surfaces derived between subjects do not have a shape correspondence, non-rigid surface registration is required after normalizing to MNI space. Following procedure is followed to account for misalignment of vertices and get voxel-based correspondence needed for group analysis.

Mahalanobis distance (MD) is calculated from each point in a source surface to target surface which is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D. Mean of this metric as shown in Eq (2) is taken to be the distance measure between surfaces for comparison. The distance metric, d from surface Y to surface X is computed by,

$$d = \sum \frac{(Y - \mu x)}{Cx} * (Y - \mu x)' , \quad (2)$$

where μx and Cx are the mean and covariance of X. Average distance of the individual surface from all other surfaces is calculated by taking the mean of distance measures for that subject. Finally target surface T_s is chosen based on the minimum distance measure.

2.4. Surface projection

Let $S_s = \{(x_1, y_i), (x_2, y_2), .., (x_k, y_k)\}$ be the vertices on each surface S. Delaunay triangulation $DT(S)$ based on the convex hull of points in S [13] is generated which ensures that the circumcircle associated with each triangle contains no other point in its interior. N-D nearest point search is performed to obtain initial match K_1, between source surface S and the target surface T_s using $DT(S)$. Similarly, corresponding indices K_2, from T_s to S are also captured. K_1 and K_2 are used as a starting point for execution of diffeomorphic spectral matching [7]. Weighted adjacency matrices are calculated for both surfaces as W_1 and W_2 based on vertices and edge information from each surface. Single correspondence Gc graph is then obtained based on the volumetric entity formed by 2 surfaces and with associated links obtained through K_1 and K_2. The spectral matching algorithm yields the final transformation files T_{12} and T_{21} that can be used to register T_s to S and vice versa.

For each surface, S_i let C_i be the correlation map T_{12} obtained in the above step, then the intensity values I_k are projected onto the T_s. Individual measurements from GM surface can then be projected onto

the target surface. Surface coordinates are converted to actual volume co-ordinates by applying the inverse transform of the matrix from image volume in MNI space to the vertices.

2.5. NODDI processing

The NODDI model was applied and processed using Accelerated Microstructure Imaging via Convex Optimization (AMICO) [https://github.com/daducci/AMICO_matlab] to estimate intracellular volume fraction (V_{ic}), isotropic volume fraction (V_{iso}) and orientation dispersion index (ODI) which is defined as

$$OD = \frac{2}{\prod}\arctan\left(1/K\right) \quad (1)$$

where K is the concentration parameter of the Watson distribution. ODI is the metric of interest in further analysis for this study. NODDI metrics are first co-registered to intra-subject T1 through b0 image using normalized mutual information as a similarity metric. These are then transformed to MNI space by applying forward deformation field obtained from the DARTEL registration step with nearest neighbor interpolation.

2.6. N-GBSS processing

GM fraction based on ODI and FA and a corresponding psuedoT1 are generated in diffusion space that is used for the registration to MNI space as proposed in N-GBSS [3]. A threshold of 0.55 is used for GM mask.

2.7. Experimental setup/Statistical Analysis

Data Acquisition and Preprocessing: T1 and diffusion-weighted MRI were acquired on a 3T scanner (Achieva, Philips Medical Systems, Best, The Netherlands) with a 32-channel head coil. The data were acquired at 2.5mm isotropic resolution with FOV of 96 x 96 using DW PGSE-EPI sequence (TR=2.65s, TE=101ms) and Gmax = ~37.5 mT/m. The diffusion scan protocol included a B0 map, two diffusion shells with b-values of 1000, s/mm^2 (24 directions), and a HARDI shell with b-value of 2000 s/mm^2 (60 directions). This raw data is then corrected for patient movement, eddy current distortion and susceptibility distortion through FSL's topup and eddy functions [14, 15].

The proposed method is tested on the schizophrenic population in identifying the significant regions of decreased ODI between healthy and psychosis groups. In this study, we have considered a dataset with 114 subjects of whom 47 are healthy, and 67 are classified as a psychosis group. To compare the sensitivity retained after registration in identifying significant regions between the groups, VBM analysis is performed in SPM. Each individual ODI maps are merged into a 4D volume to prepare for voxel-based analysis within gray matter surface skeleton regions. The design matrix is made with two groups having 47 healthy subjects and 67 psychosis subjects with age as a covariate. Permutation-based statistics are performed on skeletonized ODI for each method using FSL randomize (10000 permutations) and results are presented within a different region of interests (ROIs) in Table 1. P-value < 0.05 corrected for multiple comparisons is considered to be significant.

3. Results

GBSS based skeleton and GS-BSS based skeleton are overlaid on mean ODI as shown in Figure III-2. As can be noticed from the qualitative image, the coverage of GBSS based skeleton is limited especially in the cortical folding regions.

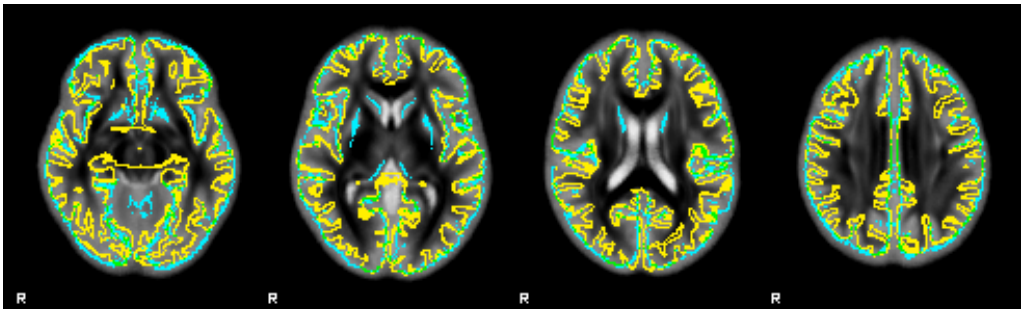

Figure III-2 Mean ODI image with an overlay of GBSS based skeleton (blue) and GS-BSS based skeleton (yellow).

Mean voxel count after skeletonization is around ~56K from GBSS whereas it is around ~101K for GS-BSS based approach. However, GBSS does capture regions like putamen and caudate which are not overlapped with surface-based skeleton from GS-BSS. Significant results from group comparison

between healthy and psychosis patients are shown in Figure III-3. Based on the ROI analysis, ODI for the medial PFC sub-region is reduced in psychosis patients. To validate the group differences specific to these regions, permutation tests are performed by passing the mask of these ROIs to both GBSS and GS-BSS skeleton images. Significant values are presented in Table 1. GBSS did show significance in anterior cingulate gyrus (ACG) but with less number of voxels in the cluster compared to that of GS-BSS. Also, medial frontal cortex (mFC) was identified to be significant with $p<0.001$ using GS-BSS approach whereas none of the voxels showed significance from GBSS approach. Figure III-4a shows the overlay of mFC region with skeletons from both the methods where it is noticeable that there is little overlap of GBSS skeleton on this region that could be causing this issue.

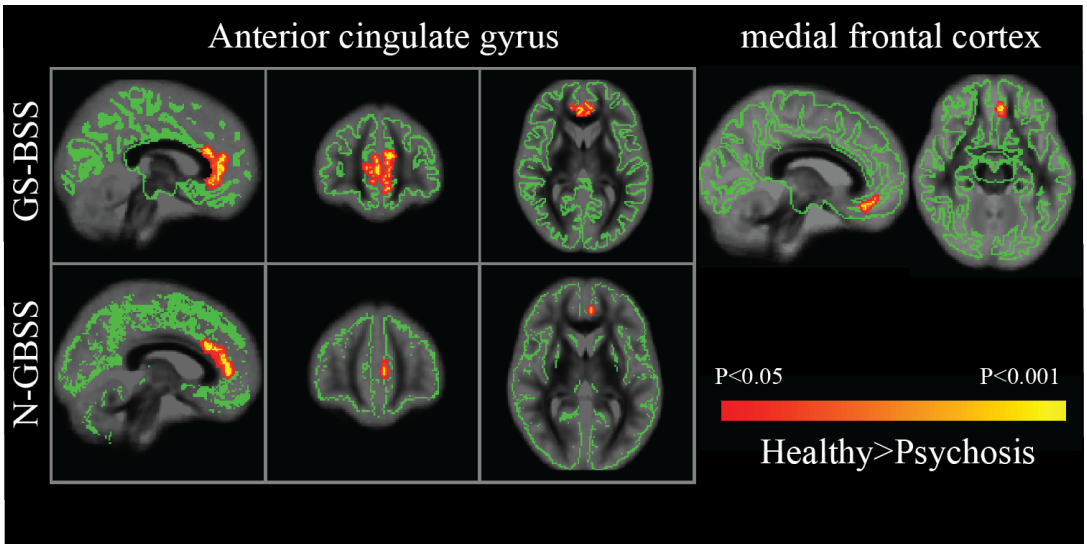

Figure III-3 Significant regions with FWE corrected p<0.05 where ODI in healthy>psychosis.

Table III-1 Significant regions of between-group analysis with healthy>psychosis using threshold-free cluster enhancement with FSL randomize cluster analysis

ROI	GS-BSS			N-GBSS		
	Voxel#	p-value	X, Y, Z	Voxel#	p-value	X, Y, Z
ACG	539	<0.001	58,111,51	103	0.002	63,115,51
mFC	29	0.006	64,109,38	-		

4. Discussion

In this chapter, we have presented a technique to carry out voxel wise spatial statistics using gray matter surfaces. We have then presented the application of this approach to NODDI based metrics in identifying the group differences between healthy and psychosis subjects. We have shown the difference in performance between N-GBSS and GS-BSS based approaches.

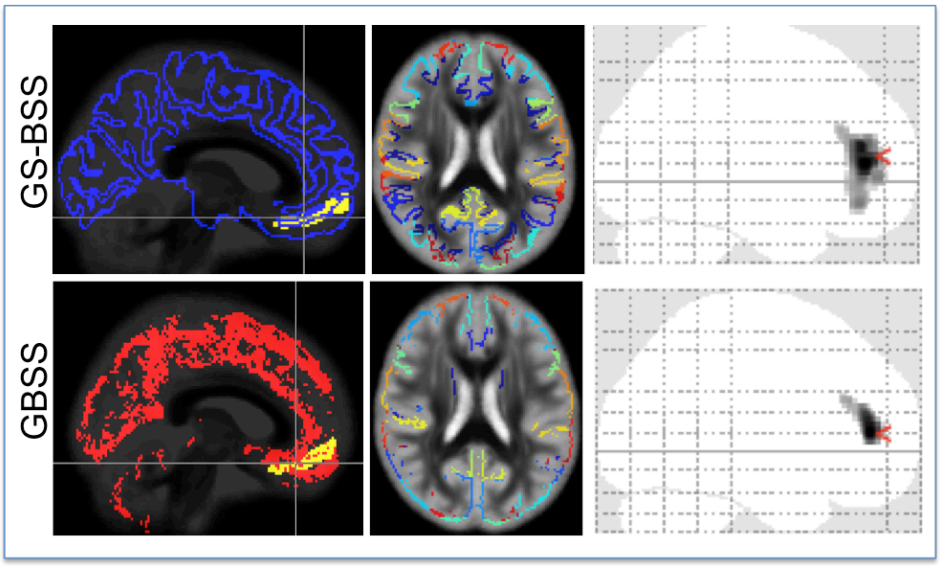

Figure III-4 Results indicating a comparison between GS-BSS vs. N-GBSS methods a) mFC region (yellow) overlaid on mean ODI and GS-BSS (blue) and GBSS skeleton (red) b) Segmented gray matter surfaces c) VBM results for whole brain analysis after registration with a threshold at 250 voxels.

While N-GBSS method is yielding highly probable gray matter voxels that are sensitive to underlying microstructural changes, it still has some limitations. These issues, with possible improvement from GS-BSS based approach, are shown in Figure III-4. N-GBSS was not able to capture the mFC that has shown to be significant in GS-BSS approach with high probability ($p<0.01$). Secondly, GS-BSS yields more number of highly probable gray matter voxels on the overall skeleton (Figure III-4b) that could be beneficial in further analysis. VBM based ODI results are shown in Figure III-4c (uncorrected

$p<0.001$ thresholding at 250 voxels) After applying the correction for multiple comparisons, none of the regions are shown to be significant in N-GBSS whereas significant clusters (FWE_{corr} $p<0.005$ at cluster level with 1797 voxels) in prefrontal cortex are seen in the overall ODI based on the registration approach followed in GS-BSS. Lastly, though the results are validated by applying on NODDI derived indices, GS-BSS approach is by no means limited to this particular model and can be applied to any parameter of interest for analyzing the results in the gray matter.

In future work, we would like to improve the accuracy of the surfaces for capturing additional regions like putamen and also incorporate detailed validation to account for any compromise in variance with increased voxel count in gray matter regions.

5. Conclusion

We have presented a generalized approach of carrying out voxel wise spatial statistics using gray matter surfaces from structural images. It can be applied to any model parameter of interest in understanding the underlying diffusion microstructural changes in gray matter. We have shown that by utilizing the established methods of registration and segmentation, the sensitivity can be retained for performing voxel-wise group analysis compared to N-GBSS method. Moreover, we have shown that using the gray matter surface as opposed TBSS based skeleton provides more number of highly probable gray matter voxels that can be helpful in further analysis.

CHAPTER IV. IMPROVED GRAY MATTER SURFACE BASED SPATIAL STATISTICS IN NEUROIMAGING STUDIES

1. Introduction

Gray matter (GM) in the cerebral cortex is key to many sensory, cognitive, and motor functions of the brain. Detecting cortical alterations with neuropathologic conditions could provide potential biomarkers to facilitate early diagnosis and assessment of disease severity. In recent years, the development of neuroimaging techniques, such as high-resolution magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), diffusion-weighted magnetic resonance imaging (DW-MRI), positron emission tomography (PET) or single photon emission computed tomography (SPECT), have promoted the identification of structural and functional characteristics of the developing brain and underlying mental disorders [1-7]. An increasing number of studies have shown structural and functional gray matter changes in clinical applications - e.g., amyotrophic lateral sclerosis [8], schizophrenia and bipolar disorder [9, 10], age-related effects [11], attention deficit hyperactivity disorder [12], and Alzheimer's disease [13]. While T1 images can be acquired at high resolution (e.g., 1 mm isotropic), clinical imaging in other modalities (such as DW-MRI and fMRI) is constrained by imaging and physiological factors leading to a lower resolution (2-3 mm isotropic). As the cortex is about 1.6 – 4.5 mm thick [14-16] within the gray matter tissue region between white and pial surfaces, significant challenges arise with cross-subject analysis involving registration artifacts and partial volume effects [17]. The individual cortical anatomy may not be sufficiently aligned after non-rigid volumetric registration since it is quite challenging to incorporate spatial coherence in the volumetric images (see Figure IV-1a). In particular, volumetric smoothing potentially introduces partial volume effects since the cortical structure is thinner, as seen in Figure IV-1b. This issue was successfully addressed in WM using tract-

based spatial statistics (TBSS) [18], which has proven to be a popular technique for performing voxel-wise statistical analysis with improved sensitivity and interpretability of analysis of multi-subject diffusion imaging studies in white matter (WM) [19-23].

Gray matter based spatial statistics (GBSS) adapted the TBSS framework for GM using neurite orientation dispersion and density imaging (NODDI) [11] to perform voxel-wise statistical analysis on GM microstructure in diffusion studies. GBSS employs skeletonized cortical ribbon to capture diffusion metrics along its trajectories. However, this approach could yield low sensitivity to the cross-sectional differences around the cortical sulci since GM skeletonization is extracted only along highly overlapping regions. To overcome this issue, we proposed an alternative approach known as gray matter surface based spatial statistics (GS-BSS) [24] that employs a cortical surface to increase the number of highly probable GM vertices that closely follow the cortex (Figure IV-1b).

In volumetric neuroimaging analyses, spatial smoothing is generally performed to improve image alignment and statistical sensitivity, at the cost of specificity of the underlying region of interest [25]. As the GM of healthy adult subjects is typically < 5 mm thick, spatial smoothing needs to be carefully performed to retain the sensitivity and specificity of the underlying changes [26, 27]. Even if optimal spatial smoothing exists, the individual cortical anatomy may not be sufficiently aligned after non-rigid volumetric registration. Consequently, surface-based approaches have been proposed with improved sensitivity in cortical morphometry [25, 28-33] over volumetric neuroimaging in both fMRI and cortical features of interest. There is a wide agreement that the surface-based morphometric (SBM) analyses [34-36] have theoretical and empirical advantages over traditional voxel-based morphometry (VBM) approaches for addressing the problem of inference in group studies. However, substantial inter-subject variation in the shapes of local features (e.g., mean curvature) still hampers accurate cortical surface registration.

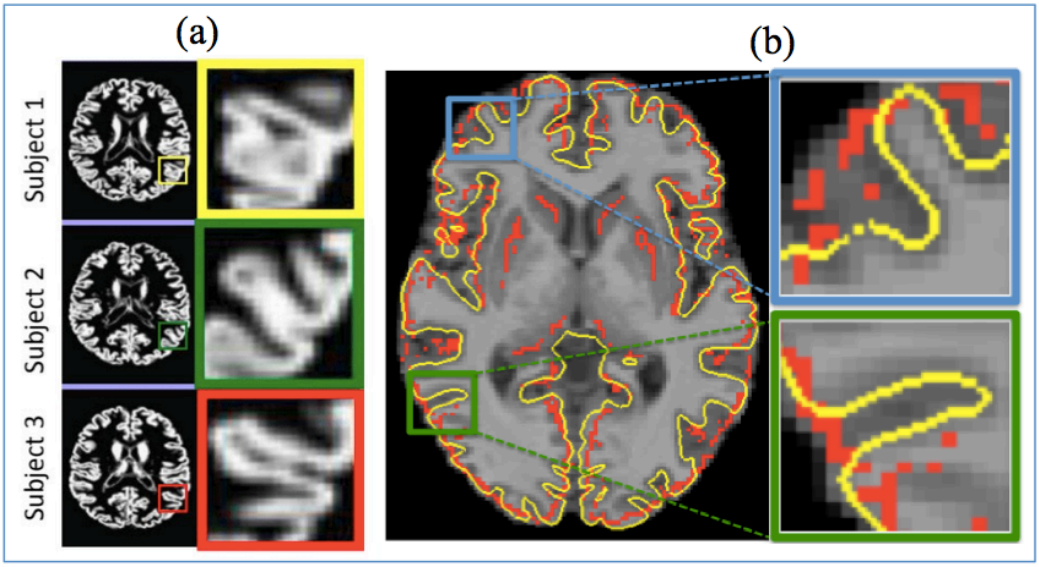

Figure IV-1 (a) Non-rigid image registration of GM probability maps of three subjects. Each color box highlights the corresponding region of interest. The right column shows detailed differences in cortical folding patterns across the subjects. (b) Skeletonized GM (red) and cortical central surface (yellow) are overlaid on T1 image. GM central surface closely follows the cortical structure compared to that of skeletonized GM approach. Two examples are highlighted in blue and green boxes where GM cortical surface closely follows the cortical structure compared to the volumetric based GM skeletonization approach. Darker regions on T1 indicate GM and lighter regions represent WM.

A majority of studies focus on volume- or surface-based analysis on a particular modality [9, 37]. Few studies [32, 38, 39, 40] have incorporated multi-modalities into a single integrated pipeline of surface-based analyses. The desire to better understand structural-functional relationships drives the need for robust analysis frameworks. The Human Connectome Project (HCP) minimal preprocessing pipeline [38] is one such approach that integrates multimodal data for cross-subject analysis. It is built upon the FreeSurfer software tool (https://surfer.nmr.mgh.harvard.edu) for surface generation and alignment to standard space in addition to defining Connectivity Informatics Technology Initiative (CIFTI) format and gray matter coordinate system that incorporates cortical and subcortical information. In a recent study, multimodal surface matching (MSM) [39] registration is incorporated into a pipeline that uses multimodal registration features containing myelin maps (Glasser and Van Essen, 2011), resting-state networks (RSNs) and visuotopic features to drive alignment to a group template. In the HCP approach [38], the data acquisition protocol is customized and often requires newly developed preprocessing methods,

unlike conventional data acquisition schemes.

In this chapter, we propose N-GSBSS for carrying out localized statistical testing of neuroimaging data across multiple modalities in GM. Unlike the skeletonization approach in GBSS, cortical surfaces reconstructed from high-resolution T1 images are employed to facilitate cross-subject analysis. This method provides a bridge between volume and surface registration approaches to achieve cross-subject correspondence of low-resolution image data. This method is an extension of our previous work, GS-BSS [24]. While conceptually similar, improvements are made in registration methodology that allows mapping of the metrics of interest in subject space. The key idea in this method is to incorporate normal search from the cortical surface to get metrics from highly probable GM voxels using the orientation dispersion index (ODI) from the NODDI model. ODI is higher in GM compared to that of WM [41], thus searching for higher ODI could help to locate underlying highly probable GM. Toward this end, we show an application to the statistical analysis of fMRI data. To test the sensitivity of the approach, a simulation study is performed by varying region of interest (ROI) size and percentage change of intensity values within the ROI. It is presented as a full end-to-end pipeline to perform such spatial statistics in group analysis. The source code for N-GSBSS is made available at https://github.com/MASILab/N-GSBSS/. The computational time of N-GSBSS is 68 times faster than that of traditional SBM or 86 times faster than the HCP minimum-processing pipeline.

2. Methods

2.1. Background

GS-BSS method was proposed to perform voxel-based statistical analysis of diffusion microstructure features acquired at low resolution on GM surfaces using high-resolution T1 images. Structural images are segmented and normalized to MNI template space using diffeomorphic anatomical registration using exponentiated lie algebra (DARTEL) method [42]. Diffusion metrics of interest are co-registered to structural T1 and transformed to MNI template space using forward deformation. GM surfaces are deformed to MNI template space using inverse transformation obtained from the registration

step. Correspondence between cortical surfaces is obtained with diffeomorphic spectral matching DSM [44] and the mapping is applied to the deformed diffusion microstructure data in MNI template space to project onto the target surface for group analysis. GS-BSS is shown to yield better performance compared to that of VBM or the skeletonization approach of GBSS, which is based on alignment invariant skeleton projection. However, some methodological limitations could impact the sensitivity of such an analysis. Firstly, the possibility of having any misalignment between diffusion microstructure and structural images after co-registration could impact the sensitivity of the analysis to be performed on highly probable GM region. Secondly, the diffusion metrics of interest are projected onto the GM cortical surface in MNI template space that could allow the prospect of including distortions caused in the data from the volume registration step. Finally, while the GM surfaces are used for achieving cortical correspondence, all the data is mapped back into voxel-space before performing statistical analysis.

In this chapter, the goal is to improve spatial statistics in GM by projecting all the metrics of interest from each modality onto a single target cortical surface and carry out vertex based statistical analysis. Current work addressed the limitations of GS-BSS and provided an improvement in the following areas,

- To overcome possible alignment issues from the co-registration step and improve intra-subject correspondence, the cortical search is proposed that can further improve the sensitivity of the method.
- To minimize distortions and keep the data as close to the raw images that are acquired as possible, metrics of interest are mapped onto the cortical surface in subject space, unlike the GS-BSS method where the metrics of interest are mapped from the volume image in MNI space onto the deformed cortical surface in MNI template space.
- To perform spatial statistics on vertices, unlike the voxel-based spatial statistics that is performed in GS-BSS.
- To show the applicability of the method in additional modalities like fMRI.

2.2. Subjects and neuroimaging data acquisition

Neuroimaging data were collected on 30 healthy subjects (average age = 31.94 (male, n=18) / 35.83 (female, n=12)) who participated in an on-going study of brain connectivity in neuropsychiatric disorders. The Vanderbilt University Institutional Review Board approved the study, and all participants provided written informed consent before enrolling in the study. Neuroimaging data were acquired on a 3T scanner (Achieva, Philips Medical Systems, Best, The Netherlands) equipped with a 32-channel head coil located at the Vanderbilt University Institute of Imaging Sciences. The following neuroimaging data were acquired on each subject: 1) a T1-weighted MPRAGE anatomical scan (1mm isotropic resolution, TE=2ms, TR=8.95 ms and TI=643 ms), 2) up to 6 functional EPI scans (3 mm resolution during which subjects completed an event related spatial working memory task (described below), and 3) a diffusion-weighted imaging scan protocol (2.5 mm isotropic resolution, FOV=96 x 96, TR=2.65s, TE=101ms, Gmax = 37.5 mT/m) that included two diffusion shells with b-values of 1000 s/mm2 (24 directions) and 2000 s/mm2 (60 directions). Cardiac and respiratory gating was not used.

2.3. Preprocessing

2.3.1. T1 anatomical data processing

Each structural scan was segmented into GM, WM, and cerebrospinal fluid (CSF) tissue classes using the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm/) from SPM12 (http://www.fil.ion.ucl.ac.uk/spm). Additionally, each voxel of the images was automatically labeled using multi-atlas segmentation [45] according to the BrainCOLOR protocol [46] into 132 brain regions and one background. The white, central and pial cortical surfaces were reconstructed by MaCRUISE [47] using the topology-preserving geometric deformable surface model. The central surfaces were used in further surface-based processing including registration and mapping volume data onto the surfaces.

2.3.2. Diffusion data processing

Diffusion-weighted images (DWI) were first affinely registered to b0 with 12 degrees of freedom

using FMRIB Software Library's (FSL 5.0) FLIRT [48] for eddy current correction. The registration matrix of each DWI was used to measure patient movement, and the gradient table was rotated accordingly. For diffusion data processing, the data from 2 shells were combined into a single DWI file and corresponding b-values and b-vectors were concatenated accordingly. A scheme file was generated using the fsl2scheme command from Camino (http://camino.cs.ucl.ac.uk). A brain mask was created using the FSL brain extraction tool [49].

For NODDI processing, the DWI file, scheme file, and mask (generated as described above) were passed to the AMICO package (https://github.com/daducci/AMICO/tree/master/matlab), which is a fast implementation of NODDI [42] with a linear approximation. Intracellular volume fraction, isotropic volume fraction (V_{iso}), and orientation dispersion index (ODI) obtained from the above method were co-registered to the individual structural T1-weighted scan using SPM12. These ODI and V_{iso} maps from multiple subjects were used in further analysis and validation of N-GSBSS.

2.3.3. Working memory fMRI processing

During the functional EPI scans, subjects completed a slow event-related spatial working memory task. Briefly, on each trial, three filled circles were presented sequentially, one at a time, during a 3-second encoding phase. The encoding phase was followed by a 16 second delay period during which a fixation dot was shown. Following the delay period, a probe (open circle) was presented for 1 second and subjects had to indicate with a button press whether or not the probe matched one of the previously encoded locations. Each trial was followed by a 14 second inter-trial interval. Subjects complete 30 working memory trials and 18 control trials. The working memory and control trials were identical, except for the fact that subjects were asked not to memorize the locations during the cue period of the control trials and pressed both the yes and no button during the probe period. Different colored circles, red and grey, were used to alert subjects to working memory and control trials, respectively. Preprocessing and generation of first-level, subject-specific statistical parametric maps were performed in SPM12 [43]. Preprocessing included slice timing and motion correction, and co-registration of subject's

functional EPI scans to their anatomical T1-weighted scans. Subject-specific, voxel-wise maps showing relative difference in the BOLD response between working memory and un-modeled baseline for the cue, maintenance, and probe conditions were generated by modeling each subject's time series data. Note that the contrast maps for the cue, maintenance, and probe conditions were kept in the individual subject-specific space co-registered to T1 before being entered into the N-GSBSS pipeline described below.

2.4. N-GSBSS pipeline

The steps involved in carrying out the spatial statistics starting from the preprocessed multi-modal data to transferring all the metrics of interest onto a single target surface are illustrated in this section. The data from the co-registered volume images are projected onto the GM central surface using enclosing voxel approach. If there are alignment issues after co-registration, it would introduce partial volume effects or outliers by fetching data from the voxels that may not belong to highly probable GM. In order to overcome this limitation, the cortical search is proposed using ODI measure as it has been shown to be higher in GM compared to that of WM [41].

2.4.1. Cortical search using NODDI maps

Diffusion microstructure indices from NODDI including ODI and V_{iso} are used in the cortical search. First ODI is masked with V_{iso} to exclude any voxels with the isotropic volume fraction of greater than 0.5 indicating CSF regions. The surface normal is calculated at each vertex on the central surface. As the T1 was acquired at 1 mm resolution and the cortical thickness is < 5 mm thick, we search the maximum ODI at each vertex along positive and negative normal directions (2 mm at maximum range with an interval of 1 mm). We create a search map by collecting these enclosing voxels that the normal directions point out. The metrics of interest in other modalities are finally transferred onto the central surface via the search map. Figure IV-2a illustrates this approach and corresponding histogram of masked ODI is shown in Figure IV-2b before and after the search.

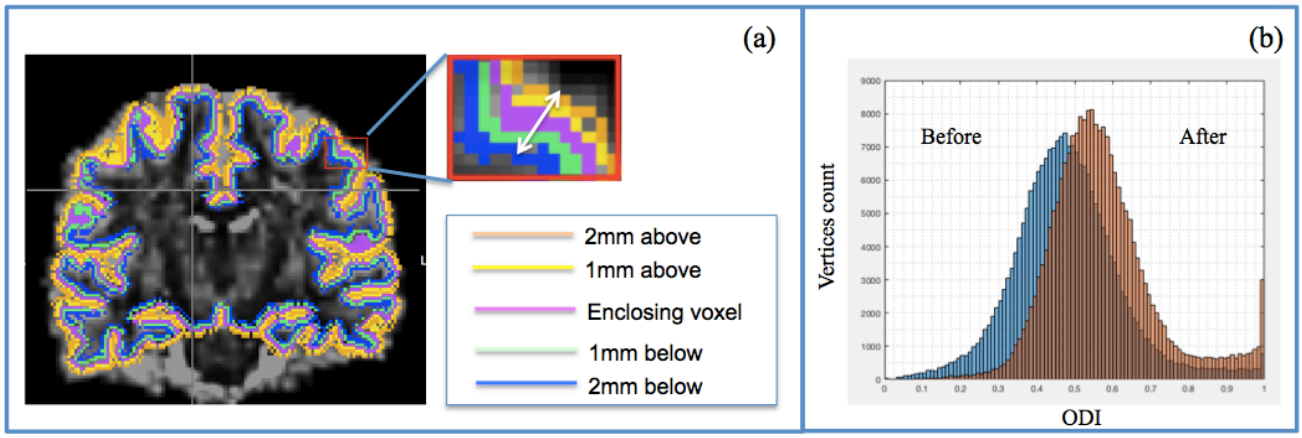

Figure IV-2 (a) ODI overlaid with cortical surface mapping based on enclosing voxels, 1mm above, 2mm above, 1mm below and 2mm below of central surface obtained using normal search. At each vertex, maximum ODI value is selected from these five values along the vertex normal (white arrow in zoomed in box) and corresponding map is used for projecting the diffusion metrics on to the cortical surface. (b) Histogram of ODI projected on to the cortical surface on a single subject before and after ODI search.

2.4.2. Cortical correspondence on the target surface

Cortical surfaces are highly variable, so roughly similar surfaces would be useful for surface registration. As preprocessing volume registration can provide reasonably well-aligned surfaces, structural T1 is non-linearly registered with MNI template using ANTs SyN registration method [52]. Corresponding inverse deformation is applied to the surface as the first step. The vertex coordinates of the surface are converted to RAS format before applying "antsApplyTransformsToPoints" from ANTs toolbox. The deformed coordinates are converted back into original format thus transforming the surface from subject space to MNI space (#2 from Figure IV-3). However, as shown in Figure IV-1a, the cortical anatomy is not yet well aligned across the subjects after volume deformation. Then, we refine/update the correspondence using surface registration step [44] in the same way as [24], which is expected to establish better correspondence. It provides mapping information of the cortical surface from each subject onto the target surface (#3 from Figure IV-3) on which spatial statistics can be performed.

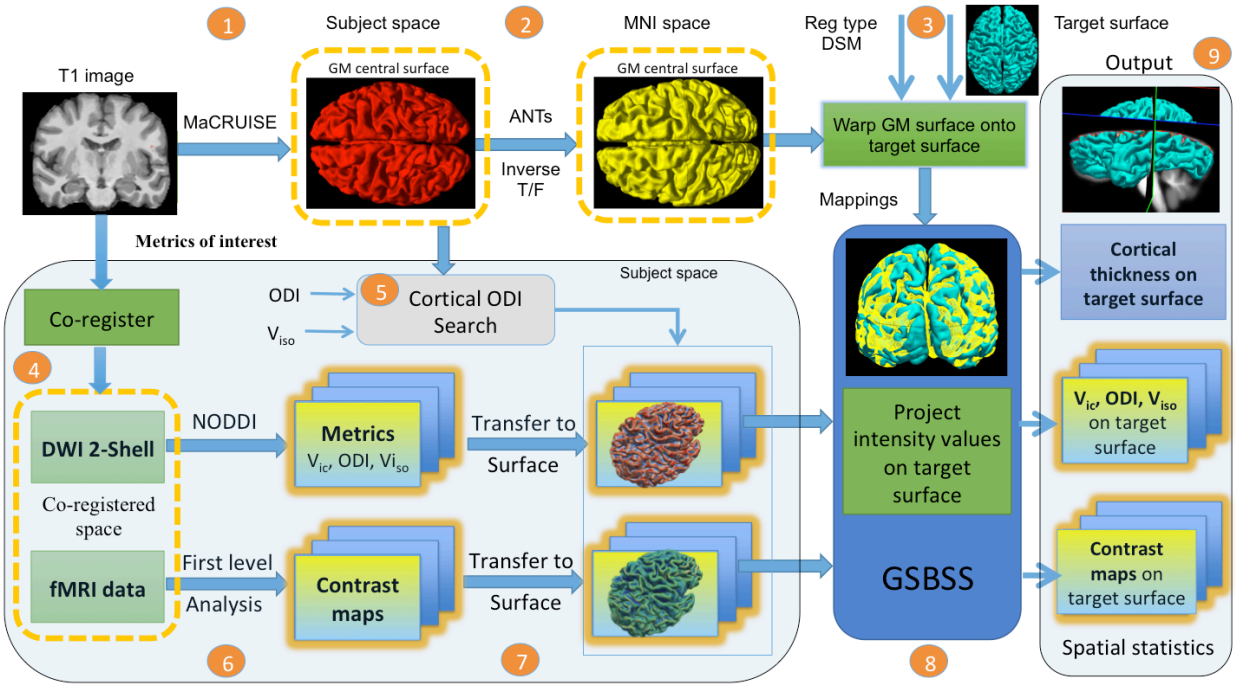

Figure IV-3 Flowchart of the N-GSBSS data processing for each subject. (1) The central surface is reconstructed via MaCRUISE (red) (2) and transformed to the MNI space (yellow) using ANTs volume registration. (3) These volumes are diffeomorphically registered to a single target surface. (4) Metrics of interest in other modalities are co-registered to corresponding anatomical T1-weighted image. (5) Cortical ODI search is performed using ODI and V_{iso} from NODDI metrics to search for higher ODI excluding V_{iso} within a given range (6) Data are processed for each modality (NODDI for diffusion microstructure and first level analysis for working memory tasks) to derive metrics of interest for cross-sectional analysis. (7) Metrics of interest are mapped onto the individual surface. (8) The mappings from shape correspondence are used to project the intensity values of metrics of interest to the target surface (blue). (9) Vertex-wise spatial statistics on all projected data are performed on the target surface.

2.4.3. Project metrics of interest on the target surface

As cortical anatomical properties such as cortical thickness were derived from the surface, they were already assigned to each vertex. These properties were then projected onto the target surface via the established shape correspondence from step 3. Images from different modalities are co-registered to T1 anatomical images before proceeding with further analysis as shown in step 4. Cortical ODI search is performed by taking in ODI and V_{iso} measures from the NODDI model to get the corresponding map of highly probable GM vertices for co-registered images (step 5 in Figure IV-3). Step 6 illustrates the first level analysis carried out on each modality to derive metrics of interest. In the volume images, the metrics

of interest were mapped onto the individual GM surface (step 7 in Figure IV-3) from the voxel that encloses the corresponding vertex coordinate obtained from the cortical ODI search (step 5 in Figure IV-3). Both DW-MRI based NODDI metrics and fMRI based working memory contrast maps were thus projected via the vertex coordinates and the mapped properties were then transferred onto a common target surface (Step 8 in Figure IV-3). Spatial statistics across the subjects is performed on the target surface by applying a 2mm smoothing kernel for cross-subject analysis. We adapted the Gaussian kernel smoothing proposed by [45], where each vertex was weighted based on data from the neighboring vertices and scaled by the vertex area.

2.4.4. Summary of highlighting enhancements

A novel ODI search along the surface normal for maximum ODI value is used to probe for highly probable GM regions in the co-registered image. Additionally, enhancements that are made to the earlier method [24] are the transfer of metrics of interest on to the GM cortical surface in the individual subject space instead of MNI space, to reduce the error that could occur with volume and surface deformation to the MNI template. While [24] showed the application to diffusion microarchitecture features, this work extends the applications to fMRI data, thus enabling multimodal analysis across structural and functional changes. Group analysis is performed at vertex level on the target surface.

The evaluation of the approach is carried out in the following ways. 1) We compare qualitative mean ODI, a diffusion microstructure feature, for N-GSBSS with and without cortical ODI based search in comparison with HCP minimum processing pipeline. 2) We perform non-parametric permutation testing on contrast maps obtained from first level analysis of working memory tasks in fMRI. 3) We perform a simulation study in structural MRI to evaluate the sensitivity and specificity of the approach.

2.5. Spatial statistics

Once all the properties from different modalities were projected on the target surface, GM based vertex-wise spatial statistics were calculated using the Permutation Analysis of Linear Models (PALM) [51] package from the FSL software library (FMRIB; http://www.fmrib.ox.ac.uk/fsl/) which performs

inference through permutation. Significant results are reported after controlling for family-wise error (FWE) with $p<0.05$ through threshold-free cluster environment (TFCE).

2.6. Baseline methods

2.6.1. Volume-based registration (VBR) processing:

Volume images of metrics of interest from other modalities were registered to MNI template by applying the non-rigid transformation obtained from anatomical T1-weighted images. A GM mask was calculated based on 0.5 thresholds on the GM probability map in each subject and 70 percent overlap across all the subjects to filter the number of voxels to retain highly probable GM voxels. A Gaussian kernel smoothing of 2 mm was applied before performing spatial statistics. Nonparametric permutation-based testing was performed on smoothed volume data within a brain mask using FSL PALM [51] (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM). Statistical results were projected onto the target surface based on the enclosing voxel approach for visualization and comparison with surface-based results.

2.6.2. Surface-based registration (SBR) processing:

In order to compare the proposed approach, we used the FreeSurfer surface registration method [30] for cortical shape correspondence. Metrics of interest from volume data in subject space were projected onto the central surface using the enclosing voxel approach. These metrics were transferred to the target surface via the shape correspondence and then smoothed on the target surface for cross-sectional analysis. In order to make a fair comparison with N-GSBSS results with optimal multiple comparison correction, metrics of interest from two hemispheres were considered as a single data set before carrying out the permutation-based statistical tests.

2.6.3. HCP minimum processing pipeline

As the tools (https://github.com/Washington-University/HCPpipelines) provided for the HCP pipeline [38] are specific to the data acquired for the Human Connectome Project, the ciftify (https://github.com/edickie/ciftify) tool has been developed by HCP community to adapt the HCP

pipeline to non-HCP datasets. For preprocessing, surface reconstruction is carried out using recon-all from FreeSurfer 6.0 (https://surfer.nmr.mgh.harvard.edu) by supplying a raw structural T1 image as we do not have T2 weighted images as required by HCP pipeline for recon-all. Since the acquired diffusion data do not have phase-encoding reversed pairs for distortion correction as suggested for the HCP pipeline, distortion corrected images are obtained from diffusion data processing mentioned in the above section. Then the diffusion to the structural script provided in the HCP pipeline is adapted to process epi_reg_dof from the HCP pipeline to register DW-MRI images to the structural T1 image.

WM segmentation obtained from FreeSurfer processing is supplied as an input argument. For fMRI processing, since we do not have field maps as required by the HCP pipeline, preprocessed first-level images that are co-registered to the T1 structural image using SPM12 are used as a starting point. Conversion tools provided in ciftify toolbox are used to put preprocessed T1, dMRI data and fMRI data into an HCP-like folder structure. Volume to the surface mapping of diffusion measures is performed using a ribbon mapping method that uses white and pial surfaces to get the voxels within the GM ribbon, as described in [38]. As we do not have T2 weighted images in our custom dataset, myelin maps are not calculated and thus myelin-style volume to surface mapping is not considered for diffusion analysis. Also, the results are based on the gray ordinates at 2 mm resolution with a total of ~64000 cortical vertices for both hemispheres as proposed by the default settings.

Processing time comparison between N-GSBSS and SBR using FreeSurfer are reported in Table 1. We used a single thread (Intel Xeon CPU E5-2630 v4 @ 2.20GHz and 32 GB of RAM) on an Ubuntu 16.04 LTS Linux Workstation.

Table IV-4: Processing time comparison for SBR, HCP and N-GSBSS based approaches. In SBR, a spherical mapping was conducted for each hemisphere followed by spherical registration. HCP minimum processing pipeline involves the recon-all step and ciftify step to convert to gray ordinates. Details of the time taken for each step are provided in the processing details column.

Pipeline	Processing steps details	Total time
SBR	Per hemisphere: FSRUNTIME@ mris_sphere 1.48 hours, 1 thread FSRUNTIME@ mris_register 0.80 hours, 1 thread	~273.6 mins
HCP	ReconAll (mris_sphere and mris_register) : 4.71 hours hrs, 1 thread Ciftify : 1hr 5 mins, 1 thread	~345 mins
N-GSBSS	ANTs volume registration: ~2.12 mins, 1 thread DSM surface registration: ~1.49 mins, 1 thread	~4 mins

2.7. Simulation study setup

The spherical masks with a radius of 3, 4, and 5 mm were created in template space and transferred back to subject space via the inverse transformation from ANTs SyN [52] registration for each subject. This range was chosen since the cortex is around 3-5 mm thick and because capturing the ROIs with different radii could reflect the differences in accounting for partial volume effects in the GM and WM border regions. The location was chosen to contain cortical folding that is variable across multiple subjects to account for partial volume effects when performing cross subject studies.

The GM probability maps for the 30 subjects were randomly divided into two groups, G1 and G2, with 15 subjects in each group. The GM probability data in G2 were then modified in the subject space to simulate percentage change of intensity values in intervals of 10% in the corresponding mask regions. A total of 27 combinations (3 masks and 9 different scalings) were considered for evaluation.

With 0% change, the images in G2 were the same as original images. Thus, we considered the difference between the groups as a baseline. We excluded 100% change of the region of interest in G2, which is completely reduced to zero. With 50% change, the intensity values were half of the original values in ROIs from G2 images.

GM probability data from each of the 27 combinations in G2 were then processed through N-

GSBSS to place all the data on the target surface for cross-sectional analysis. GM probability data were also evaluated for VBR, SBR and HCP pipeline for comparison with the same parameter/experimental settings, including 2 mm Gaussian kernel smoothing. Non-parametric permutation tests were then performed between G1 and G2 for all combinations using FSL's PALM [51] package with 5000 iterations.

To assess the sensitivity of the approaches, we examined the ratio of maximum t-statistic ("t-stat ratio"), which was defined as the amount of scaling with respect to the baseline. To have a single metric with comparable result across all the methods, we reported the ratio with respect to the baseline. A baseline is where we performed second level analysis for group differences across the two groups where no changes are applied to original GM probability maps.

3. Results

In this section, we present the results of all the N-GSBSS analysis as follows: 1) Qualitative results of mean ODI with and without search in comparison with the HCP processing pipeline 2) Application in fMRI to identify active regions in task-based working memory. 3) GM simulation results in structural MRI based on different ROI size and intensity differences.

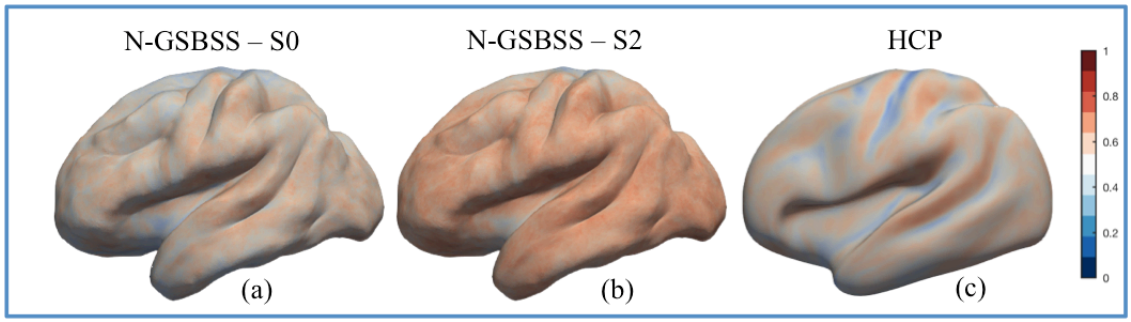

Figure IV-4 Mean ODI across 30 healthy subjects using (a) N-GSBSS – S0 with no search (a) N-GSBSS - S2 including ODI search of 2mm (c) HCP minimal processing pipeline. The HCP results are based on the "gray ordinates" with 64 thousand vertices (the suggested tessellation) on both left and right hemispheres for HCP while the target surface template used in N-GSBSS has about 261 thousand vertices.

Mean ODI values across 30 subjects are shown on the target surface (Figure IV-4) for N-GSBSS

without search, with cortical ODI search and the HCP pipeline. With cortical ODI search, partial volume effects are addressed reflecting higher ODI across the cortex compared to that of other two approaches.

3.1. Working memory fMRI results

As an application of N-GSBSS in fMRI, working memory data was processed for 30 healthy subjects in cue, probe and delay tasks. We compared significant regions revealed by VBR, SBR, the HCP pipeline and N-GSBSS methods as shown in Figure IV-5. For all these tasks, the overall activation pattern is comparable across different methods. As expected, the significant vertices in VBR are fewer and more scattered than the cortical surface-based approaches of SBR, HCP, and N-GSBSS.

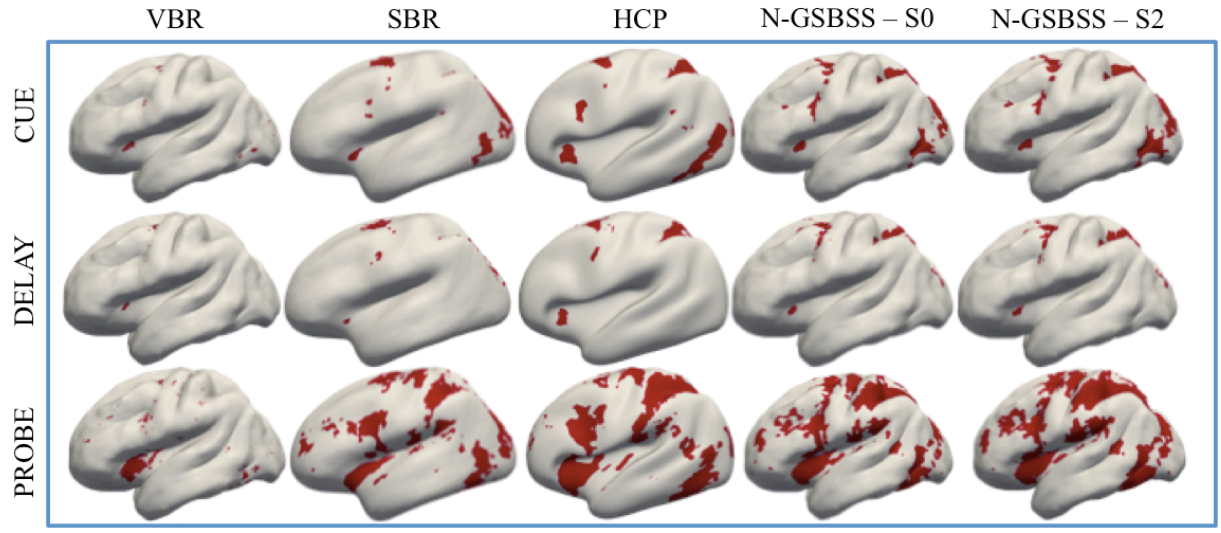

Figure IV-5 Working memory fMRI data were processed for 30 healthy controls and results are reported for (a) correct cue, (b) correct delay (c) correct probe tasks with 2 mm smoothing for VBR, SBR, HCP, N-GSBSS -S0 with no search and N-GSBSS-S2 with 2mm search methods. Significant p-values after FWE correction based on nonparametric randomize one sample t-test with 10000 iterations are reported. P_{fwe} <0.05 are highlighted in red.

Quantitative representation of the number of significant vertices with p<0.05 for all the three tasks are shown in Figure IV-6. It is noteworthy that N-GSBSS has a higher number of significant vertices in all the tasks than VBR, SBR, and HCP pipeline results. The HCP pipeline results are comparable to that of N-GSBSS more

than VBR or SBR approaches. Applying cortical ODI search further improved the activation percentage in N-GSBSS.

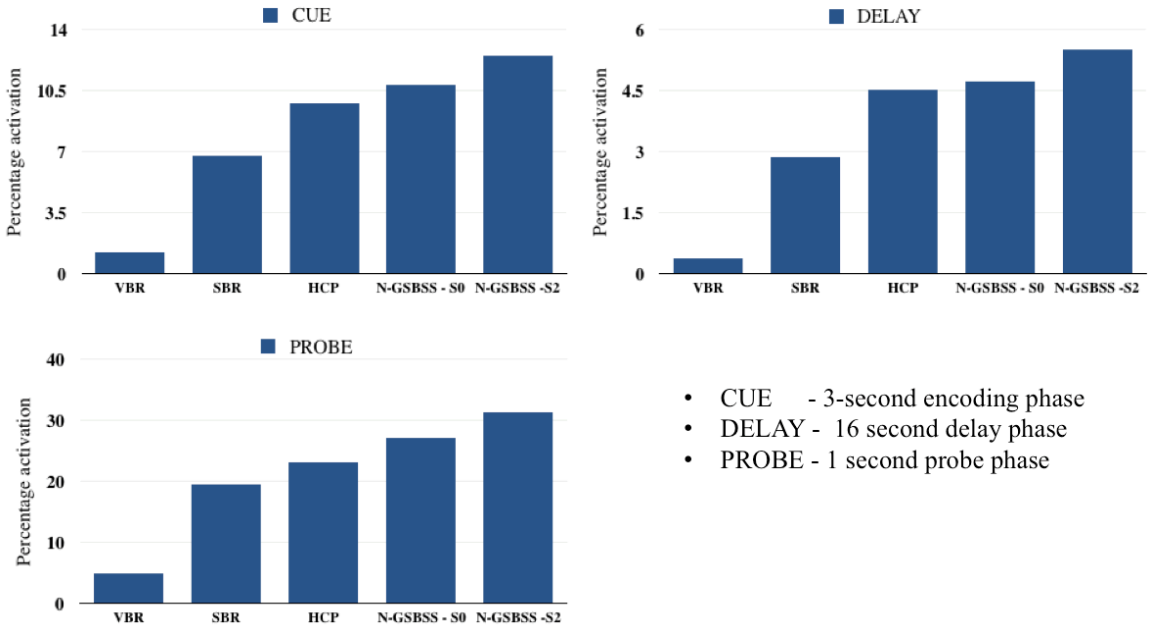

Figure IV-6 Percentage activation of working memory fMRI data were processed for 30 healthy controls and results are reported for (a) correct cue, (b) correct delay (c) correct probe tasks with 2 mm smoothing for VBR, SBR, HCP, N-GSBSS –S0 with no search and N-GSBSS –S2 with 2mm search methods. The number of significant vertices, with p-values < 0.05 after FWE correction based on nonparametric randomize one sample t-test with 10000 iterations, is divided by a total number of vertices and the percentage is reported.

3.2. A simulation study in structural MRI with changes in regions of interest

Here, we evaluate N-GSBSS with respect to VBR, SBR and HCP pipeline techniques in identifying sensitivity and specificity of changes in GM voxels located in spherical ROIs of 3, 4, and 5 mm radius located in a region of the frontal cortex. Figure IV-7 illustrates spheres with a radius of 5 and 3 mm.

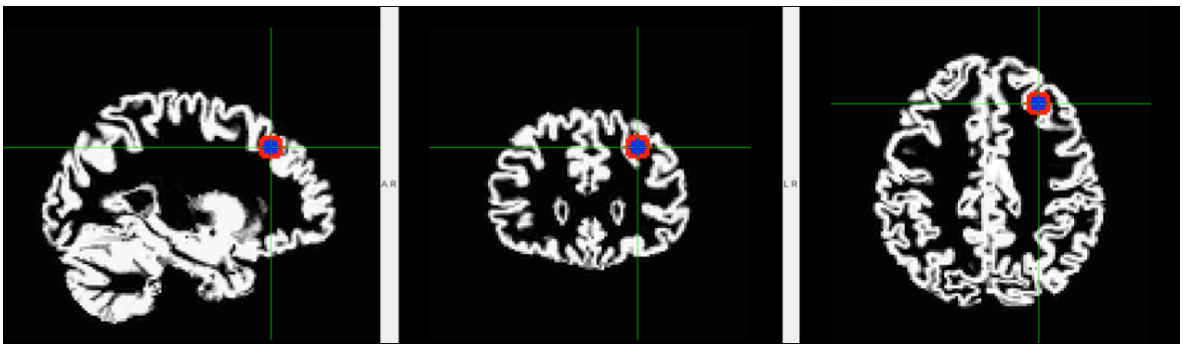

Figure IV-7 The gray matter probability map shows the simulated effect as an overlay mask of 5 mm (red) and 3 mm (dark blue) spheres.

Quantitative results in Figure IV-8 show the t-statistics ratio for varying ROI sizes of 3mm, 4mm, and 5mm, and percentage change in the GM probability values from 10% to 90% in the intervals of 10%.

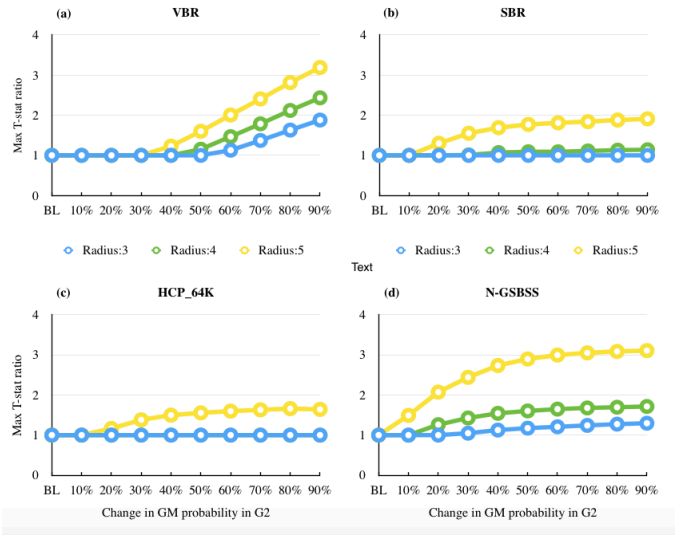

Figure IV-8 Quantitative results for statistical group differences over the change in lesion size from 3 to 5mm and percentage change from 10 percent to 90 percent. (a) Results from VBR analysis. (b) Results from Free surfer registration analysis. (c) Results from HCP minimum processing pipeline with default gray ordinates. (d) Results from GSBSS based analysis. Y-axis indicates maximum t-statistic ratio with respect to the baseline. The x-axis indicates the percentage change of GM probability in G2 with respect to original GM probability images in G2.

A t-stat ratio is the maximum t-statistic for each scenario with respect to the baseline to reflect how much it was scaled with induced changes in the region of interest. A baseline is chosen to be the differences between the two groups in the current experiment. For VBR, to capture the intensity

difference between groups, the probability change must be at least 40% with 5 mm spherical ROI, 50% for 4 mm, and 60% for 3 mm.

GSBSS results are much more sensitive starting at 10% with 5 mm ROI, 20% with 4 mm and 30% for 3 mm spherical ROI. N-GSBSS also showed higher maximum t-statistics than SBR. With higher intensity differences starting at 70%, VBR results have higher t-statistic ratio than that of N-GSBSS. In all other cases, N-GSBSS has a higher maximum t-statistic ratio and better sensitivity.

4. Discussion

Herein, we describe an approach for carrying out multi-modal spatial statistics in low-resolution images by taking advantage of high-resolution T1 weighted images that are acquired as part of the scan protocol. This approach favorably compares with traditional volume-based analyses and with respect to the FreeSurfer surface registration approach along with the HCP minimum processing pipeline. Our approach offers an advantage over VBM by achieving improved cortical alignment in agreement with other surface-based registration techniques [25, 28-33]. Moreover, in comparison with FreeSurfer, SBR, and HCP pipelines, the N-GSBSS approach showed an improvement in sensitivity. It suggests that the initial alignment obtained by non-rigid deformation from the T1 image makes surface registration much easier. Consequently, this improves the statistical power compared to existing approaches.

The key aspect of this work is the addition of NODDI based search, which ensures that metrics from low-resolution images are retrieved from highly probable GM. It is achieved by use of the ODI measure from NODDI, which is known to be higher in GM compared to that of WM [42]. Thus by searching for maximum ODI, alignment issues after co-registration or PVE effects from underlying voxels are addressed. As we are interested in low resolution with dMRI acquired at 2.5 mm resolution and fMRI at 3 mm resolution, we assume that after co-registration to T1, the underlying data is roughly aligned at the voxel level. Thus we utilize the search map obtained from diffusion modality to apply to fMRI for getting the data based on enclosing voxel approach. The reported fMRI t-statistics suggest an improvement in sensitivity with N-GSBSS.

The simulation study is set-up to perform sensitivity or specificity check for N-GSBSS to the underlying changes in tissue microstructure. As we are interested in performing analysis in psychiatric applications including schizophrenia [53, 54] that are known to have changes in the prefrontal region, the ROI is chosen from this region. The GM probability map is chosen as the parameter of interest and the intensity changes are simulated within an ROI region. Compared to the baseline methods, N-GSBSS showed superior sensitivity to the underlying changes in both intensity and the size of the ROI as shown in Figure IV-8. While the volume-based analysis was not able to detect any significant differences between groups for at least up to 50% change in the GM probability values, N-GSBSS was able to capture differences starting from 10% change with ROI size of 5 mm, 20% for 4 mm and 30% for 3 mm. The low performance of VBM could be potentially due to partial volume effects prevalent in the volume-based approach even after applying the GM mask to limit the analysis to highly probable GM regions.

In the simulation study, SBR analysis showed a similar pattern as N-GSBSS. However, the sensitivity of this approach is not as high as N-GSBSS. Differences between the methods are likely due to different registration approaches since both of them used the same surface to obtain corresponding GM probability values from the volume image. The HCP pipeline results are similar to those of SBR, which is expected since the HCP pipeline uses FreeSurfer registration.

The subtle differences seen between HCP and SBR results could be due to the difference in preprocessing steps including surface reconstruction performed in each of these methods. The data of interest in the current analysis is not acquired as per the HCP pipeline guidelines. So the preprocessing is carried out using third party tool called "ciftify" as recommended by the HCP community. This tool adapted the HCP minimum-processing pipeline to make it work for non-HCP datasets. Once the legacy data is brought into the gray ordinates space, the second level analysis is performed as applicable to the HCP data processing. For a fair comparison, we used the HCP pipeline with default parameters to the extent possible. For example, the analysis results for HCP are based on the "gray ordinates" with 64 thousand vertices (the suggested tessellation) on both left and right hemispheres. It differed from the target central surface used in SBR and N-GSBSS analysis that has about 261 thousand vertices for both

hemispheres. These could have contributed to the lower sensitivity of this pipeline in this simulation study with GM probability percentage change greater than 40% with an ROI size of radius 5mm.

Overall significant regions captured by N-GSBSS agree with those of VBR, SBR, and HCP pipelines across different modalities while achieving high spatial specificity. It is highly likely that the volumetric transformation already deformed cortical surfaces into similar shapes (geometry) before the surface registration, which results in better shape correspondence by greatly reducing the local anatomical ambiguity in the surface registration. Particularly, the proposed approach showed significantly improved performance in our study. However extensive validation in different studies might be also interesting in future work. Finally, N-GSBSS possesses high flexibility that allows any registration method as well as multiple modalities. We expect that such a feature can be generally extended to various modalities in general neuroimaging studies.

The source code for N-GSBSS is posted in a Docker image: (https://github.com/MASILab/N-GSBSS/).

CHAPTER V. HARMONIZATION OF WHITE AND GRAY MATTER FEATURES IN DIFFUSION MICROARCHITECTURE FOR CROSS-SECTIONAL STUDIES

1. Introduction

Understanding the development of white matter microstructure is essential in neurodevelopmental studies. With the increasing number of publicly available neuroimaging databases, including the Adolescent Brain Cognitive Development (ABCD), the UK BioBank and Connectome studies, there is an opportunity to combine large-scale imaging resources to increase the power of statistical analyses to test the common biological hypothesis. Community-wide efforts are underway to address standardization of acquisitions and analyses for imaging biomarkers as described in imaging biomarker road map for cancer studies [10]. Using diffusion tensor imaging (DTI), ENIGMA's disease working groups have begun to compile evidence across cohorts for differences in a range of DTI measures and discovering factors that consistently affect brain structure and function [88]. However, the community of clinicians, engineers, and physicists is not yet ready to agree on a single best practice approach to advanced DW-MRI. Therefore, it is imperative to understand when (and how) different protocols can be analyzed to enable comparison and optimization across protocols based on practical study design and imaging constraints.

Application of advanced DW-MRI methods in clinical research studies has been hindered by a lack of consensus on best protocols. DW-MRI suffers from between-scanner variation that hinders comparisons of images across imaging sites, scanners and over time. A model that can preserve biological variability and remove the unwanted variation introduced by the site is needed. Harmonization of data across scan protocols and site differences is an important preliminary step to conduct group analysis involving cross-sectional data acquired across different regions or on different scan protocols (Figure V-1). Using fractional anisotropy (FA) and mean diffusivity (MD) maps, Fortin et al. compared several

harmonization approaches and found that ComBat performed best at inter-site variability correction [89]. ComBat has also been applied to harmonization of cortical thickness measures across scanners and sites [90] and multicenter radiomic studies with positron emission tomography [91]. With the advent of advanced diffusion MRI models, e.g., the neurite orientation dispersion density imaging (NODDI), along with white matter, it is also possible to analyze microstructural changes within gray matter (GM).

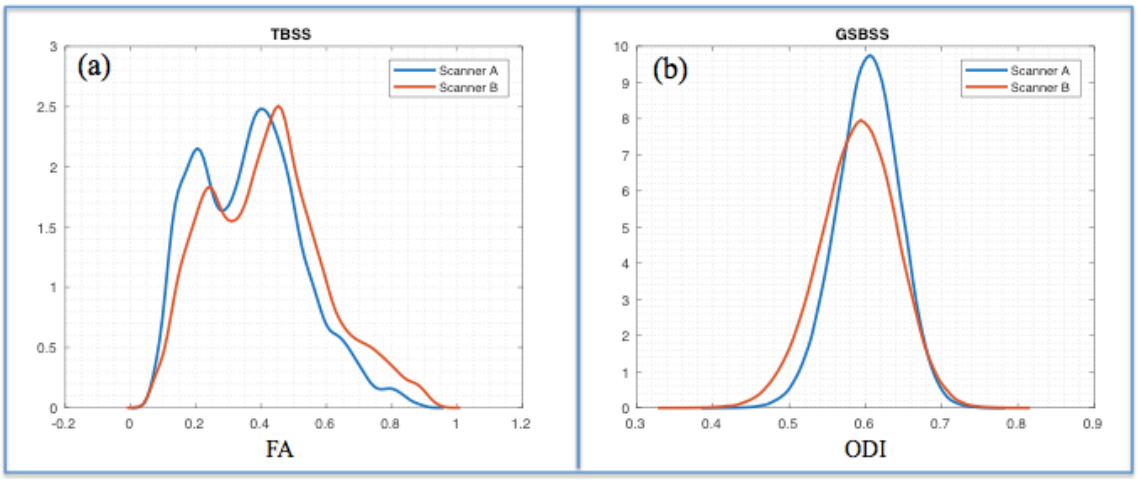

Figure V-1 Scanner differences are illustrated in a) Study 2: FA in white matter using TBSS between scanners with 1.5T and 3T field strengths b) Study 3: ODI in gray matter using GSBSS across two distinct 3T scanners.

In this study, we propose three different applications of ComBat for diffusion microstructure. First, we propose to extend the applicability of ComBat in white matter using the tract-based spatial statistics (TBSS) [76]. Second, we evaluate the performance of ComBat for correction of FA maps in white matter across two scanners with different gradient strengths of 1.5T and 3T using TBSS. Finally, we propose to use ComBat [89] in GM for correcting the scanner variability of orientation dispersion index (ODI) using GSBSS.

2. Methods

2.1. Data

This manuscript examines three independent datasets (noted as Study 1, Study 2 and Study 3). Study 1 included young adults (average age= 26.02 (male, n=194), 26 (female, n=223)) ranging from 23 to 34 years old with data acquired from two different 3T scanners. Study 1 has diffusion-weighted imaging scan protocol with b-value 1000 s/mm^2 with 33 directions. Study 2 included 119 subjects from elderly population (average age= 78.29 (male, n=45), 71.38 (female, n=66)) acquired on two different scanners with gradient strengths of 1.5T and 3T with b-value 700 s/mm^2 with 30 and 32 gradient directions, respectively. Study 3 included 30 healthy subjects (average age= 31.94 (male, n=18), 35.83 (female, n=12)) with data acquired from two different 3T scanners. For each subject in Study 3, we have: 1) a T1-weighted anatomical scan (MPRAGE, 1mm isotropic resolution, TE=2ms, TR=8.95 ms and TI=643 ms); 2) a diffusion-weighted imaging scan (2.5 mm isotropic resolution, FOV=96 x 96, TR=2.65s, TE=101ms, Gmax = 37.5 mT/m) that included two diffusion shells with b-values of 1000 s/mm^2 (24 directions) and 2000 s/mm^2 (60 directions). Diffusion-weighted images (DWI) were affinely registered to b0 using FMRIB Software Library's (FSL 5.0) FLIRT for eddy current correction. The data was then processed using DTIFIT from FSL to obtain diffusion tensor that yields FA maps and other tensor-based metrics. Overall processing flow used in this approach is presented in Figure V-2.

2.1.1. TBSS and GSBSS processing:

Individual FA maps were registered to the FMRIB58 template [76]. For group analysis, the mean FA-image was created following registration and thinned to represent the mean FA skeleton. Individual FA maps were projected onto this common skeleton.

NODDI processing was carried out using the AMICO package which is a fast implementation of NODDI [45]. Orientation dispersion index (ODI) obtained from the above method was co-registered to the individual structural T1-weighted scan using SPM12 (http://www.fil.ion.ucl.ac.uk/spm). Cortical surfaces were obtained from the MaCRUISE pipeline [172] and initially deformed via non-rigid transformation obtained by template-based volume registration using ANTs SyN [137] registration. Then,

surface registration (DSM) [148] was performed to establish cortical shape correspondence across cortical surfaces as mentioned in GSBSS [106]. ODI was mapped onto the individual gray matter (GM) surface and evaluated for cross-subject analysis.

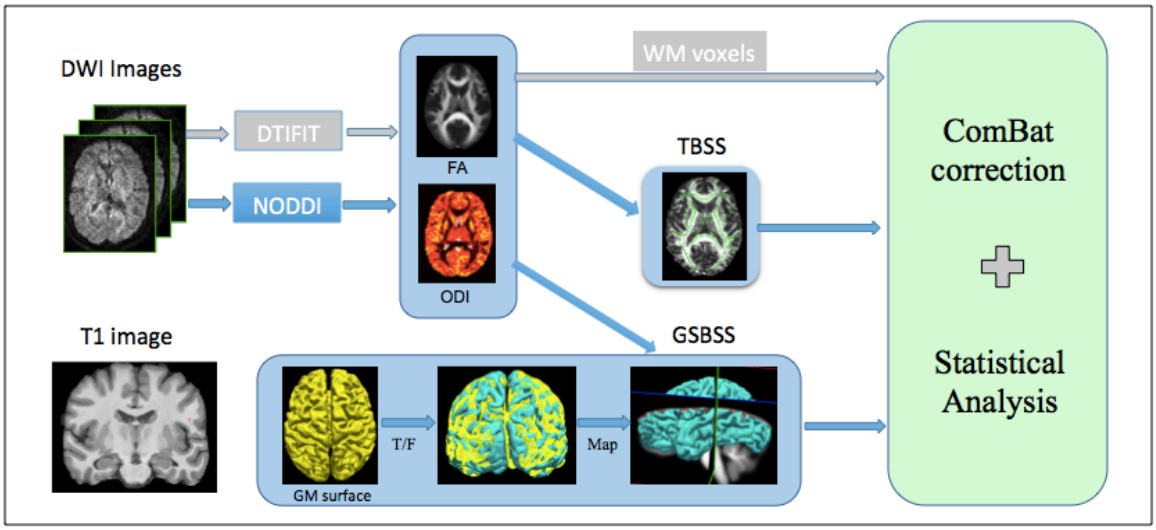

Figure V-2 DWI/HARDI images are processed with DTIFIT to extract FA map and NODDI model to extract ODI. FA maps are then processed using TBSS to extract skeletonized WM voxels. T1 image is used for generating GM surfaces that are used for GSBSS. ODI maps are then processed using GSBSS to compute mapped values on the target surface. These measures are passed into ComBat tool for variability correction for statistical analysis. Gray arrows indicate the previous application of ComBat to DTI data.

2.1.2. Variability correction and statistical analysis:

FA skeletonized voxels from TBSS and ODI indices along the target surface were then corrected using ComBat as reformulated in the context of DTI images. Additive (γ) and multiplicative (δ) effects of the site were included in the formulation to account for scanner/site variability as shown in below equation from [89]. In our study, the outcome would be FA in white matter or ODI in GM with scanner effect. For each site i, scan session j and at voxel v, corrected measure is given by,

$$y_{ijv}^{ComBat} = \frac{y_{ijv} - \alpha_v - X_{ij}\beta_v - \gamma_{iv}}{\delta_{iv}} + \alpha_v + X_{ij}\beta_v \quad (5.1)$$

A design matrix was formed with unpaired 2 sample t-tests between the groups (Site A and B) adjusting

for age. Permutation-based statistics were performed on skeletonized FA for each method using FSL randomise [173] (5,000 permutations). Significant results ($p<0.05$) comprising a threshold-free cluster environment (TFCE) [74] clustering were presented after accounting for multiple comparisons by controlling for family-wise error (FWE) rate. GM based vertex-wise spatial statistics were calculated using the Permutation Analysis of Linear Models (PALM) [173] using 5000 permutations followed by a generalized Pareto distribution to the tail of the approximation distribution [174]. The significance is reported for $p<0.05$ (FDR corrected).

3. Results and Conclusion

For study 1 with two scanners having same field strength, Figure V-3 shows significant differences between scanner A and scanner B (~20450 voxels) before correction with an overlay of negative effects of age (~37700 voxels). As noted, the number of significant voxels between scanners is reduced to zero voxels while retaining the age effects (~37500 voxels) after ComBat correction.

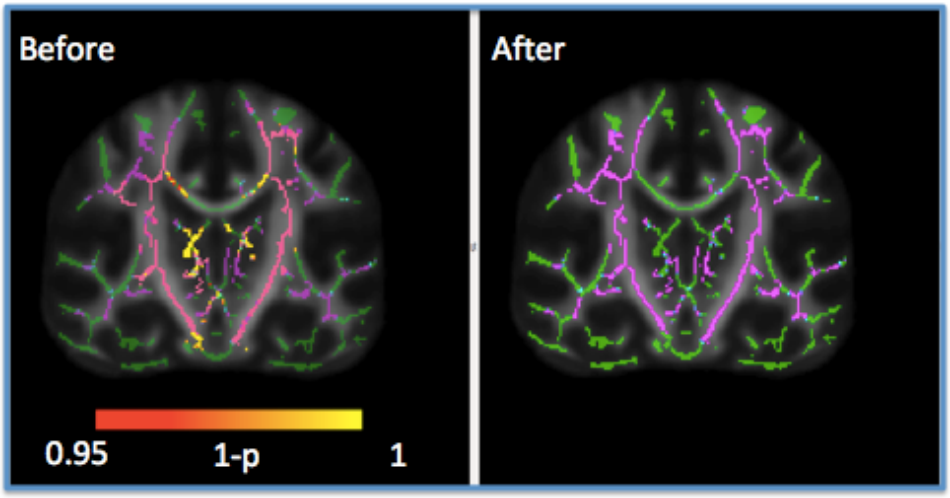

Figure V-3 Study 1: Overlay of the significance of scanner effects (red-yellow) with age (magenta) with FWE. The left image shows before correction and the right image shows after ComBat correction. Note that ComBat eliminates the scanner effect, but the age effect is essentially unchanged.

For the scanners of different gradient strengths from Study 2, significant results of scanner effects

before correction (~97100 voxels) were reduced to zero while improving the number of significant voxels with negative effects of age from ~70800 before correction to ~72200 voxels as shown in Figure V-4.

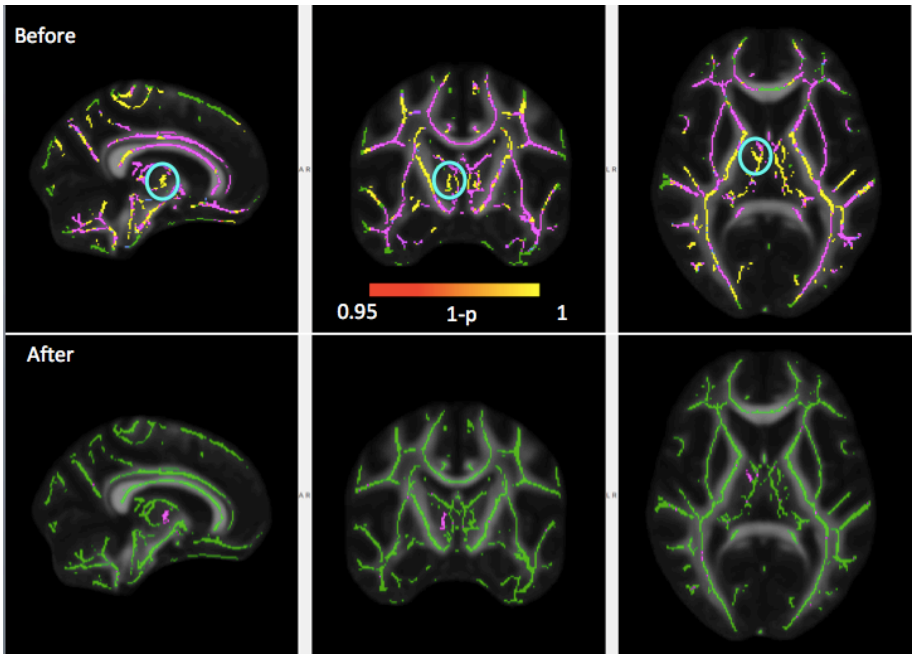

Figure V-4 Study 2: Overlay of the significance of scanner effects (red-yellow) with age (magenta) with FWE. The top row shows before correction and lower row shows the *additional* significant effects of age with ComBat correction. Note that ComBat eliminates the scanner effect, while the age effect is increased.

Finally, scanner differences are shown for ODI in the gray matter before and after correction in Figure V-5. Since the population in this study has a narrow distribution of age range, age-related effects analysis is not included for this study.

Figure V-6 illustrates the effect size as the difference in mean values of the parameter of interest. Figure V-6a shows the effect size of FA values in WM before and after correction from Study 1. Figure V-6(b) shows the effect size of FA values in WM from scanner A and B with field strengths 1.5T and 3T from study 2. Figure V-6c shows the ODI effect size shown as the difference in mean ODI values between the two scanners in gray matter from Study 3 with two different 3T scanners.

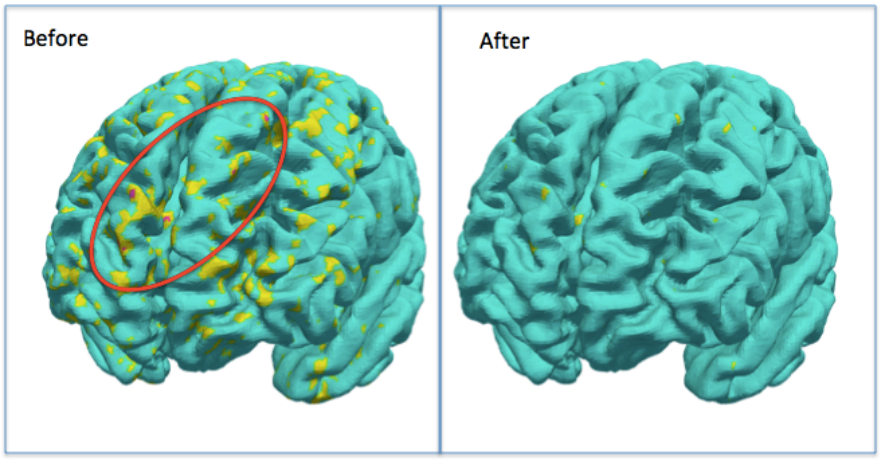

Figure V-5 Study 3: Significance of scanner effects is shown in magenta before correction (~179 vertices with FDR) overlaid on the cluster of uncorrected p-values (yellow). After ComBat correction for ODI, there is no significant scanner effect.

In all the cases, we see the reduction in the mean difference between the two groups after applying ComBat. Mean FA and ODI values between Scanner A and B after applying variability correction are shown in Figure V-7.

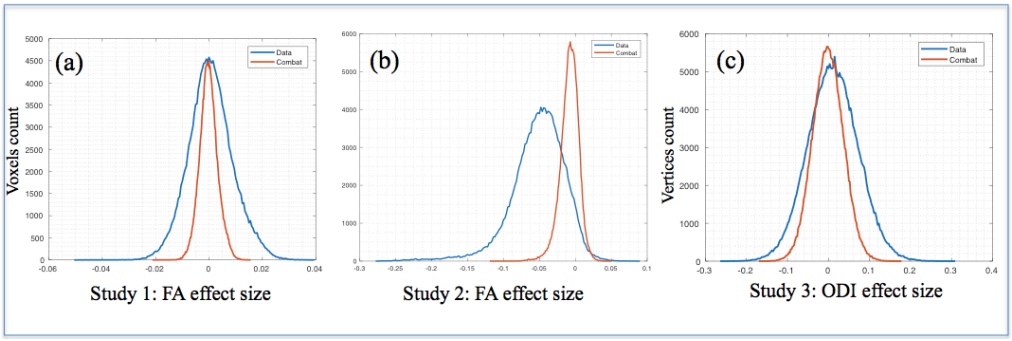

Figure V-6 Effect size values for before and after correction are shown for (a) Study 1: before correction, there was a ~0.02 FA effect between the two and 3T scanners (blue), while with ComBat, both the voxel-wise variance and size (red) were reduced by ~2-fold. (b) Study 2: before correction, there was a ~0.05 FA effect between the 1.5T and 3T scanners (blue), while with ComBat, both the voxel-wise variance and size (red) were reduced by ~5-fold. (c) Study 3: before correction, there was a ~0.2 ODI effect between the two 3T scanners (blue), while with ComBat, vertex wise ODI effect is dropped to ~0.1 with the corresponding reduction in variance.

Harmonization of WM and GM indices from DWMRI models using ComBat are shown in 1) Study 1 by applying ComBat to TBSS for correcting FA values in DTI for two different scanners with a

field strength of 3T 2) Study 2 characterizing at field strength differences showing the positive effects of ComBat correction while improving the age effect. 3) Study 3 showing the applicability of ComBat in GM with advanced DWMRI models like NODDI. In the first two studies, the scanner effect is corrected while retaining the biological variability with age. Detailed analysis in this domain could provide an opportunity to combine large-scale imaging studies to increase the power of statistical analyses to test the common biological hypothesis of understanding the development of brain microarchitecture.

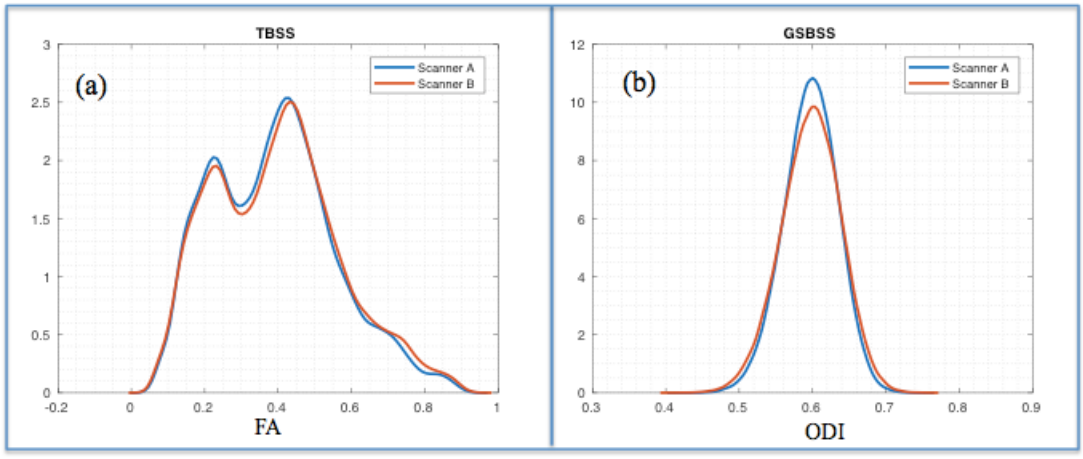

Figure V-7 Mean values between Scanner A and B are shown after applying variability correction in a) FA in white matter using TBSS between 1.5T and 3T gradient strengths b) ODI in gray matter using GSBSS within 3T scanners.

While we have shown the possibility of application of ComBat technique in three different scenarios, a detailed analysis is needed to interpret the results in a meaningful way. In the first study where we have shown the applicability of the technique to TBSS, further analysis showing the comparison with the initially proposed method as the baseline is needed if this approach indeed improves the statistical power. We are considering carrying out a detailed analysis of these approaches in future studies.

CHAPTER VI. CONSTRUCTING STATISTICALLY UNBIASED CORTICAL SURFACE TEMPLATES USING FEATURE-SPACE COVARIANCE

1. Introduction

The cerebral cortex is the outermost layer of neural gray matter critical for many brain functions including memory, attention, cognition, language and consciousness [28]. Features that characterize the cortex (including sulcal curves, gyral curves, sulcal depth, curvature and cortical thickness) are important in neuroimaging studies involving these regions [29-31]. Cortical surfaces are widely used for such analysis as they preserve topology [30, 31, 53-57]. In order to study group differences in these regions between control and clinical samples, it is common to align all the cortical surfaces to a common space [94, 175-177]. In this context, representative template space-based approaches have been proposed to studying local individual differences in cortical morphometric measurement due to their ability to represent data involving cortical patterns and other model-based voxel wise parameters mapped onto a common surface in both normal and clinical populations [52, 178]. Prior work has addressed the importance of template surface selection from the perspective of pairwise registration [54, 55, 92]. In a template based registration approach, each surface is mapped to a common template surface in coordinate space by regularizing based on feature information. However, surface-based analyses employing a predefined template might yield undesirable results if the selected template surface is substantially different from the population or if it is biased towards a particular set of surfaces [93]. Template-based registration is dependent on the a priori template specification thus constraining the underlying data to be biased to the selected template. Methods have addressed the issue of dissimilarity between template surface and surfaces of population under consideration by organizing the population of cortical surfaces into pairs with high shape similarity to achieve a higher accuracy by only corresponding such similar pairs [92], while others factored in the pattern of folding across the entire cortical surface in considering

the inter-subject average [94].

However, these approaches are still prone to bias towards the majority representation of the underlying population that could pose a problem in cross-subject analysis. We propose to address the above limitations using an a priori covariance matrix approach that de-weights subjects with similar features instead of treating all observations as independent instances. This approach uses inverse covariance weighting under the assumption that there is one latent feature of interest from the representative group. An optimal weighted mean is then reconstructed on the basis of that assumption. Features considered in this approach are mean curvature and sulcal depth. Mean curvature captures the mean amount of change with respect to surface normal [56]. Sulcal depth measures the closest distance between a cortical surface and its cerebral hull [177]. In this chapter we present an approach for constructing an unbiased mean of cortical surfaces in feature space that is representative of the underlying population while not being biased to multiple representations of the same feature from multiple surfaces by using a priori based covariance information. To simplify the analysis, we use the correlation matrix between scans as an approximation of the true covariance.

This framework is flexible and scalable for selecting the target template space, involving cross-sectional subject analysis, or performing template-based registration. The proposed technique can be factored into group-wise registration [29] to include de-weighting based on population information.

2. Methods

2.1. Data Acquisition

We considered the Kirby dataset [179] that was acquired with scan-rescan imaging sessions on 21 control volunteers. The acquisition protocol includes T1 MPRAGE employing a gradient echo read out with a short TE value (TR/TE/TI=6.7/3.1/842ms) with 240 X 204 X 256 mm FOV and 1 X 1 X 1.2 mm^3 resolution acquired in sagittal plane. No fat saturation was employed and the total scan time was 5 min and 56 s. We also included a second dataset with 10 control subjects and 10 individuals with schizophrenia for analysis. The scan protocol for this project included T1 MPRAGE (256 X 256 mm

FOV, 1 X 1 X 1 mm, TE=2ms, TR=8.95 ms and TI=643 ms) acquired on a 3T scanner (Achieva, Philips Medical Systems, Best, The Netherlands) with a 32-channel head coil.

2.2. Preprocessing

T1 images are bias-corrected using N4 bias correction [136] to account for spatial inhomogeneity. Individual T1 images are then segmented using multi-atlas segmentation [26] that segments the images into 133 BrainColor labels with 132 brain regions and a background [180]. After segmentation, the GM surfaces are derived using multi-atlas segmentation to surface method proposed as Multi-atlas Cortical Reconstruction Using Implicit Surface Evolution (MaCRUISE) [181] where inner, central and outer cortical surfaces are reconstructed by using the topology-preserving geometric deformable surface model. These central surfaces are used in further cortical surface-based analysis.

2.3. Feature-based template space selection approach

Cortical surfaces are initially aligned to MNI space (http://www.mni.mcgill.ca/) using affine transformation acquired from T1 image using trilinear interpolation**.** For each T1, gray matter (GM) central surface is reconstructed via an MaCRUISE pipeline and then mapped onto a unit sphere of a standard icosahedron subdivision with 163,842 vertices [176]. Features are generated on central surface after applying 3 smoothing iterations to reduce local noise influences. We compute weighted and un-weighted mean from these features. The weights are represented as the approximated covariance matrix as shown in Figure VI-1.

Again our goal is to build an unbiased mean of surface in feature space. So the idea is to de-weight multiple representations of similar data while capturing maximum variance in the population. From this perspective, an a priori covariance matrix is built based on the population information (e.g., demographics, patient status, cortical shapes). For example, in reproducibility analysis scan and rescan entries are provided with the same correspondence in the off-diagonal elements. Similarly for psychosis population, subjects belonging to same group are assigned to have correspondence based on setting the off-diagonal elements to have the same weighting as the diagonal elements for that group in building the

approximated covariance matrix.

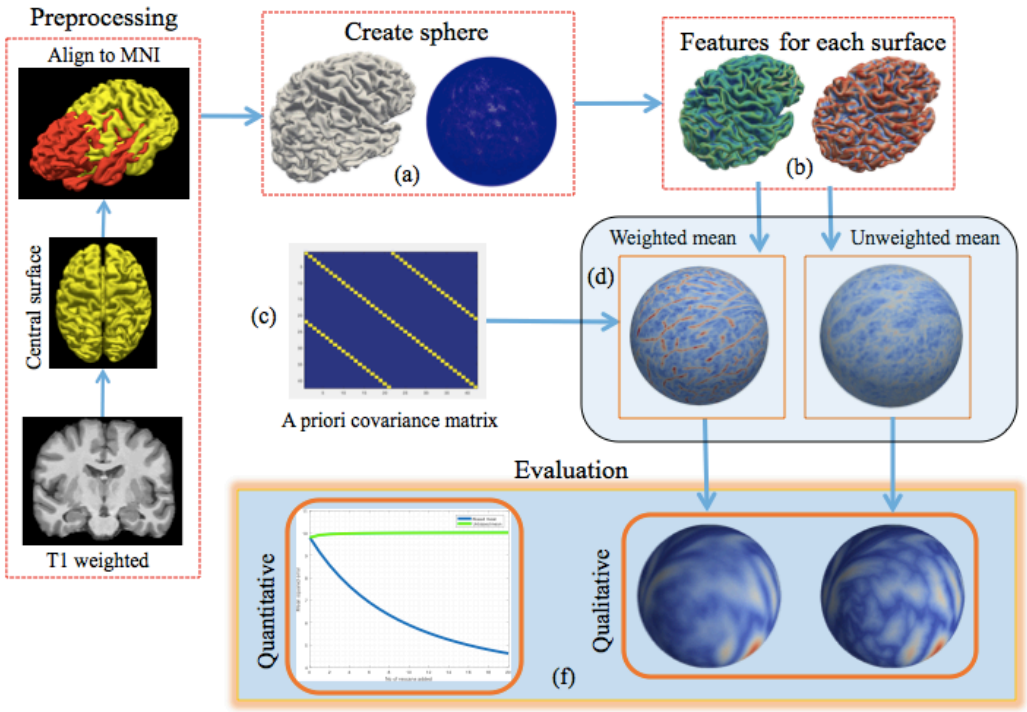

Figure VI-1 After preprocessing: a) A spherical representation is generated based on central surface. b) Features are computed and resampled along with central surface into 163,842 vertices via icosahedron subdivision. c) The covariance matrix is constructed. d) The weighted mean of features is computed based on weights from the covariance matrix. e) The un-weighted mean is computed f) Qualitative and quantitative analysis are performed based on weighted and un-weighted mean information.

By taking inverse covariance weighting approach, elements belonging to same group are de-weighted to make it a single representation for underlying population.

We compute the mean feature based template using the weights obtained in the previous step. This weighted mean of the surfaces in feature space is then compared with to un-weighted mean based on a vertex wise relative distance metric with respect to baseline for evaluation.

The idea behind the bias compensation is illustrated in a toy example in Figure VI-2. Here we have six points in 2-D space where three of them belong to one group. If we do not consider this information then the un-weighted average (red *) is biased towards the single group containing three points.

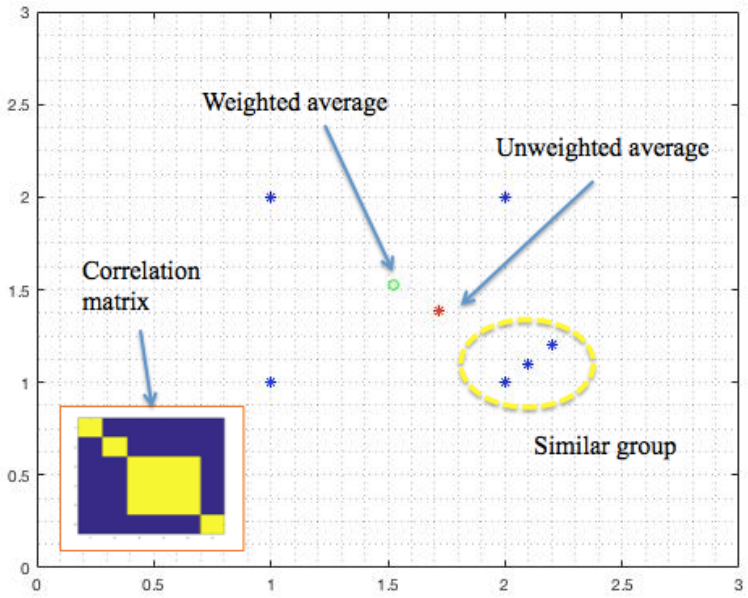

Figure VI-2 This figure shows a comparison between un-weighted average versus a weighted average for a toy example on the 2D plane. The equal weighting is given to off-diagonal elements belonging to a similar group (yellow dashed oval) as illustrated in the correlation matrix presented in the lower left-hand corner.

By factoring in the information about underlying data in a similarity matrix $\Sigma(X)$ as indicated below, we can compute a weighted average, by down-weighting three points in the same group to be a single representation thus yielding unbiased mean (green o).

$$\Sigma(X) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

By taking the pseudo inverse also called the Moore-Penrose inverse of above a similarity matrix $\Sigma(X)\ and$ taking the sum of the elements we can compute weights associated with each point which are then used in computing vertex wise weighted average as below,

$$X_m = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{6.1}$$

A similar approach can be adapted when computing the average in feature space for template space selection in pairwise registration or group-wise registration. Thus, one can address explicit bias towards multiple representations of similar data in the underlying population.

2.4. Quantitative Analysis

The results are validated on a reproducibility dataset using scan-rescan protocols. In order to quantitatively evaluate the proposed unbiased template, we used cortical surface features (mean curvature and sulcal depth) that are commonly used in surface registration approaches. We used two different distance metrics for this quantitative evaluation: 1) Mean square error (MSE) (L2 norm) of the weighted and un-weighted average with respect to baseline data as described for each scenario as described in eq (2) below where X_i is the baseline data and Y_i is the corresponding average at each vertex i. 2) The relative distance of weighted average WMD_i and relative distance of un-weighted average as MD_i in i^{th} iteration from the group averages HC_{mean} and SZ_{mean} with an equal number of subjects in each group. Vertex-wise differences of the feature measurements are captured based on the relative distance measure with respect to baseline for evaluation.

$$MSE = \sum_{i=1}^{n}(Y_i - X_i)^2 \tag{6.2}$$

$$MD_i = \frac{abs(M_i - HC_{mean}) - abs(M_i - SZ_{mean})}{abs(HC_{mean} - SZ_{mean})} \tag{6.3}$$

$$WMD_i = \frac{abs(WM_i - HC_{mean}) - abs(WM_i - SZ_{mean})}{abs(HC_{mean} - SZ_{mean})} \tag{6.4}$$

3. Results

In the Kirby dataset, rescan data is given the same weighting as corresponding scan data in the off-diagonal elements. Inverse covariance weighting from this approximated matrix is taken which is used to compute the weighted average. In order to test the reproducibility of the approach, we have taken 21 subjects with scan data as a baseline and computed both weighted and un-weighted mean.

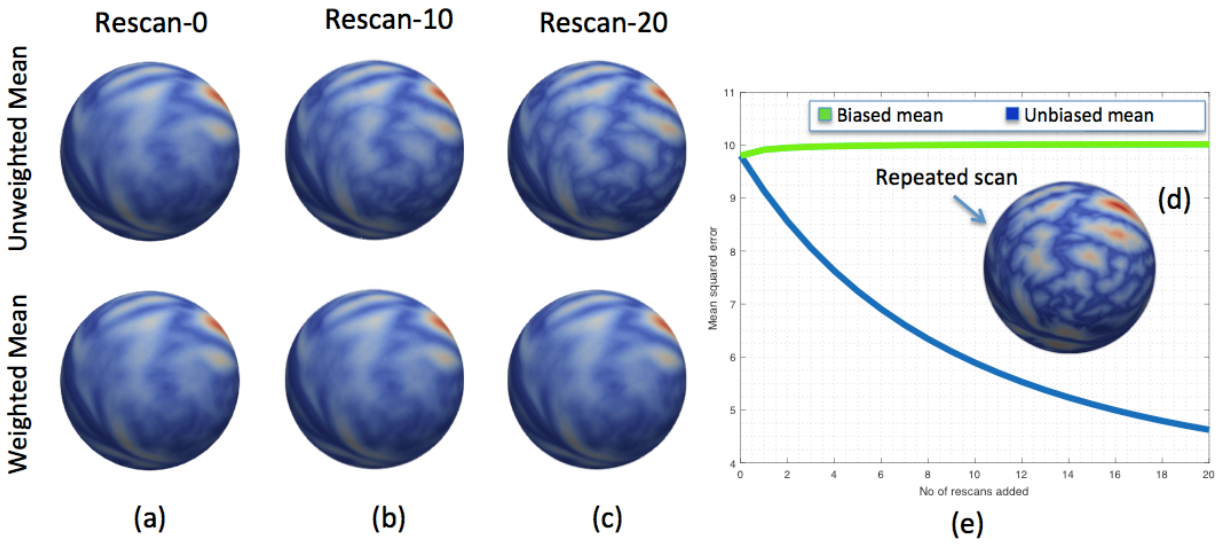

Figure VI-3 Iterative rescan data example of sulcal depth feature from Kirby dataset. Row 1 is un-weighted mean and row 2 is weighted mean from (a) to (c) with each scenario containing (a) 21 subjects with no rescan data, (b) 21 subjects with 10 repeats of rescan from one of the subjects, and (c) 21 subjects with 20 repeats of rescan from one of the subjects. The inlay (d) shows the sulcal depth of subject whose rescan is added iteratively. Plot (e) presents the mean squared distance to rescan subject from the un-weighted mean (blue) and weighted mean (green).

As seen in Figure VI-3a the means of the sulcal depth feature obtained with both these approaches are the same when no rescan data is added and 21 subjects are considered to be independent of each other. Then, we added rescan data of one subject giving it equal weighting as its scan data in the approximated covariance matrix. We repeated the addition iteratively for 20 times and captured the weighted and un-weighted means at each iteration. Figure VI-3b and Figure VI-3c present the qualitative results at 10 and 20 iterations. Figure VI-3d shows the sulcal depth feature information on the sphere for the rescan subject that was used in these iterations. With each iteration of adding rescan data, un-weighted mean comes closer to the vertex-wise feature information of the corresponding subject making it biased towards that subject. However, the weighted mean remains unchanged irrespective of the number of items as it de-weights additional duplicate scans based on approximated covariance information. Quantitative values of mean squared error distance of each of the mean to the rescanned subject feature information are presented in Figure VI-3e.

In the second application, we used two groups with a control and schizophrenia population. The

diagnosis is considered as the prior information in this dataset.

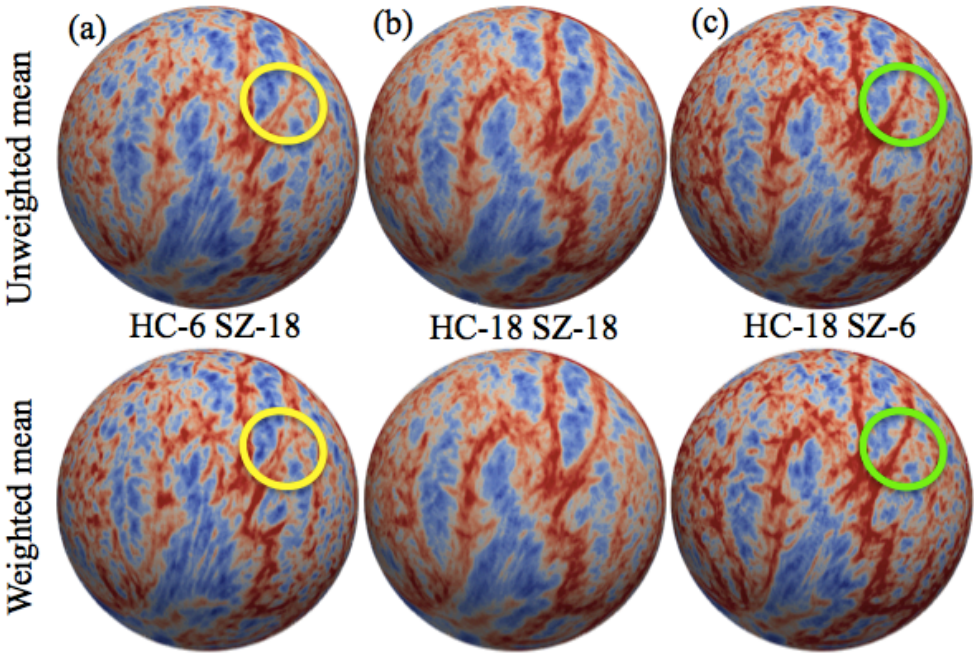

Figure VI-4 Mean curvature qualitative plot of un-weighted mean and weighted mean with a different number of subjects in control and schizophrenia populations. Feature data in each scenario included (a) 6 controls and 18 schizophrenic patients, (b) 18 controls and 18 schizophrenic patients, and (c) 18 controls and 6 schizophrenic patients. Both unweighted and weighted mean are similar with equal number of subjects in each group. However, the un-weighted mean had higher variance across the sampling strategies. The ovals emphasize qualitative areas of difference.

For this analysis mean curvature is employed as an evaluation metric and approximated covariance matrix is built based on the diagnosis information. Weights are calculated based on the inverse covariance weighting and corresponding covariance based mean is computed.

Weighted and unweighted means of mean curvature feature for a different number of subjects in each group are shown in the qualitative plot (Figure VI-4). For the same number of subjects in each group, it can be seen that weighted and unweighted means are equal. The weighted mean shows less variance compared to the unweighted mean, while increasing/decreasing the number of subjects in each group.

To compare the effect of the varying number of subjects in each group and evaluate bias

information, 10 subjects from each group are considered. Mean curvature feature based mean for 10 control subjects HC_{mean} and similarly mean for 10 schizophrenic subjects SZ_{mean} are calculated. As we increase the number of subjects in each group, the un-weighted mean becomes biased towards the group with a higher number of subjects compared to the other as shown in Figure VI-5. The plot shows the difference in vertex-wise absolute distance of un-weighted and weighted mean with respect to HC_{mean} and SZ_{mean}. By adding more control subjects, the un-weighted mean becomes closer to HC_{mean} as opposed to SZ_{mean} and vice versa. For the equal number of subjects from each group (HC=5 and SZ=5) the relative distance to both the means from corresponding group averages, HC_{mean} and SZ_{mean} are equal as highlighted in the red box. However, as the number of subjects in one group increases compared to other un-weighted mean is biased towards that group which is also reflected with a lower mean square error in Figure VI-5d and Figure VI-5e. Green lines in the plot indicate L2 norm distance of weighted mean from HC_{mean} and SZ_{mean}. Blue lines show L2 norm of the un-weighted mean from the mean of each group.

4. Discussion

In both the applications, the effect of bias towards the underlying dataset is shown when considering the un-weighted mean, while the weighted mean is stable when capturing the details of the representative features.

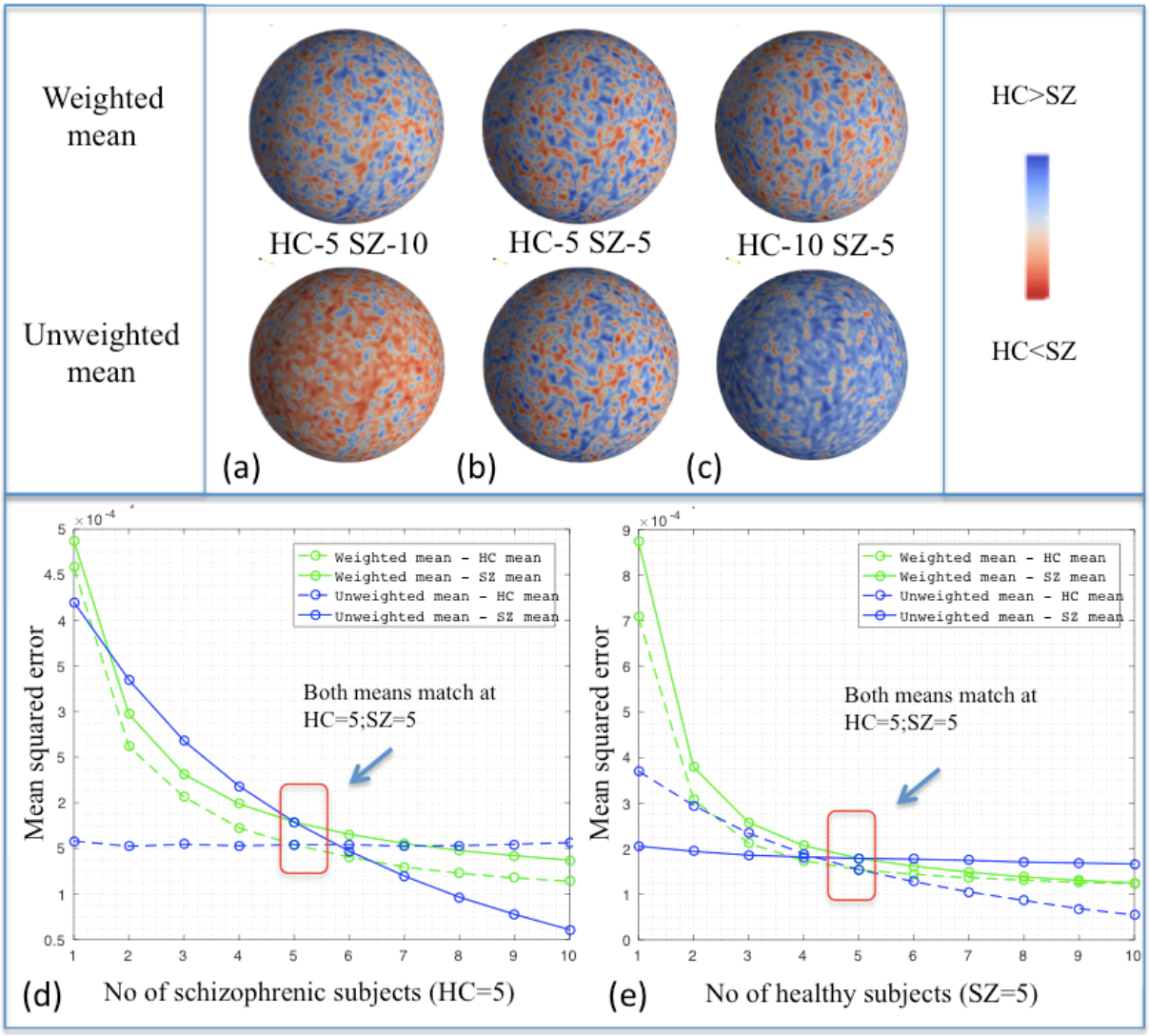

Figure VI-5 Mean curvature quantitative plot with a relative absolute distance of un-weighted mean and the weighted mean between control and schizophrenic means. Data are normalized between -1 and 1 between patients with schizophrenia and controls. Feature data in each scenario from qualitative plot included (a) 5 controls and 10 patients with schizophrenia (b) 5 controls and 5 patients with schizophrenia, and (c) 10 controls and 5 patients with schizophrenia. The color bar on the side indicates how close the relative distance is with respect to control mean (blue) and schizophrenia mean (red). The top row is from weighted mean while the lower row is from the un-weighted mean. Mean square error of mean curvature values with respect to control and schizophrenic means with a varying number of subjects in each group is shown below. In (d), the number of controls was fixed at 5 and the number of patients with schizophrenia varied from 1 to 10. In (e), the number of patients with schizophrenia was fixed at 5 and the number of control subjects varied from 1 to 10. When the number of subjects in each group is equal, then both the un-weighted and weighted means are equal as highlighted in the red box.

While we have presented the effect of incorporating covariance in computing the feature based unbiased average for template-based pairwise registration, this approach is also adaptable to group-wise registration methods [54] where no prior template is needed. In these methods, group-wise cortical correspondence is achieved by making use of various cortical features while preserving the topology. As the result still has the possibility of having a bias towards the representation of the majority of the population, incorporating covariance information at the stage of feature averaging could aid in reducing such bias.

5. Conclusion

We have presented feature based unbiased average template surface approach using an a priori covariance matrix. The proposed approach is compared with a typical un-weighted mean by applying to two different applications; one with scan/rescan data and another with clinical data with two groups. In both the cases, the weighted average is shown to be more stable and less biased when measured in terms of relative distance from the group mean or mean squared error. Incorporating covariance based approach at template selection level or when considering the mean of features in group registration methods could potentially minimize the bias. Much work remains to effectively estimate appropriate covariance structures either from study designs or in a data-driven manner.

CHAPTER VII. IMPROVING BRAIN SULCAL CURVE LABELING IN LARGE SCALE CROSS-SECTIONAL MRI USING DEEP NEURAL NETWORKS

1. Introduction

The human cortex is one of the most complex anatomical structures with substantial variation in shape across individuals. Despite its complexity, the cortical sulci are known as relatively stable regions that embed consistent cortical folds [102-104] by which the cerebral cortex can be subdivided into functionally and structurally homogeneous regions. From a morphological view, each individual sulcus (or sulcal region) can be well represented as a curve by tracing its sulcal fundus. By taking advantage of such a representation, sulcal curves have played key roles as a distinguishing indicator in cortical surface registration [57], brain development and degeneration [102, 182] and morphological variability [183, 184]. Yet, a concrete representation of cortical sulci is not commonly agreed upon due to an unclear anatomic boundary between sulci and gyri [53, 185, 186]. With recent advance of 3D cortical surface reconstruction techniques; [4, 16-18] cortical geometric features greatly support sulcal curve extraction [15,19-21].

Despite the success in extracting sulcal curves, labeling of these curves is still an open and challenging problem because of high complexity and variability in cortical folding patterns [187]. For example, the inferior temporal sulcus is highly variable with several discontinuous pieces [188]. Such variability hampers consistent labeling of cortical sulci, which is challenging even for neuro-anatomists. Sulcal labeling could be sensitive to variations in labeling protocols, and even in the same protocol, delineation of sulci can vary across experts (e.g., inconsistent endpoint delineation) [188-190]. Although one could employ cortical parcellation for sulcal labeling, it could still have challenges, as sulci do not

always obey cortical parcellation boundaries. Accurate labeling is essential to derive meaningful inferences from brain-related changes in health and disease

Manual sulcal labeling is a tedious task and needs a very high level of expertise in achieving high accuracy that agrees with sulcal nomenclature [188]. This necessitates well-developed neuroanatomical conventions in the existence of high individual variability [188-190]. Joshi et al [191] proposed a multi-atlas labeling method that uses predefined curves. In these approaches, sulcal curves from an atlas brain surface are projected onto the target subject surface and evolve along sulcal fundi via a level-set approach. Statistical shape models [192] define a shape variation prior as projections of landmarks onto tangent planes to the sphere. A watershed approach [193] extracts regions around sulcal fundi that embed some meaningful geometric characteristics like geodesic depth. However, their method is not fully automated, as sulcal regions are manually selected from the extracted sulcal regions. Another semi-automated approach [194] guides the user through a cortical surface delineation protocol implemented as a tool with an interface incorporated into BrainSuite software [28] reducing the inter-rater variability. All these manual/semi-automatic methods need manual intervention in addition to very long processing time.

In the last decades, fully automated techniques have been proposed to overcome the need for manual intervention. For example, Shi et al. [30] proposed a probabilistic graphical model of sulci, from which major sulcal curves are jointly labeled by solving a maximum a posteriori (MAP) estimation. Tu et al. [195] used a discriminative model using a boosting tree to extract major cortical sulci without employing any user-defined rules. On the other hand, Joshi et al. [196] proposed an atlas registration method that transfers the sulcal curves from an atlas surface to the subject surface and then refines the sulcal curves' locations to closely follow the sulcal fundi using geodesic curvature flow. However, this approach uses only a single subject atlas. Lyu et al. [53] proposed a multi-atlas sulcal curve labeling using spectral point matching. This method does not require surface registration. Mangin et al. [197] noted that "the future of sulcus recognition is in pattern matching methods informed by a very large dataset of manually labeled sulci". In this context, readily accessible large-scale neuroimaging dataset is highly

likely to capture sulcal variability with the improved accuracy of labeling the sulcal curves on unseen subjects. As the gold standard is not available, learning variability in the cortical folding patterns from big data can better infer new models that encode the complex folding patterns. However, handling such huge data with existing methods could be practically implausible or requires a well-defined prior model that can fully incorporate sulcal variability [188, 198, 199].

Recently, deep neural networks (DNN) have become popular in the medical imaging field [200, 201] involving large datasets due to the accuracy, speed and flexibility offered by these models. In neuroimaging, deep learning has shown its success in a wide range of applications including anatomical brain segmentation [202], brain tumor segmentation [203], biological psychiatry [204], deep unsupervised learning in traumatic brain injury [205], epileptic discharge detection for EEG-fMRI [206], segmentation of deep brain regions in MRI and ultrasound [207], inter-scanner harmonization in diffusion MRI [208], understanding sensory cortex [209], et cetera. Convolution neural nets (CNN) are a specific type of DNN that uses convolution and pooling layers. These are widely used for image recognition tasks as they reflect the translation-invariant nature of most images. U-Net [210] is a deep CNN model that is adapted from fully convolutional networks. It can work with very few training images and provide more precise segmentations. CNNs are successfully applied for segmentation and other applications in volume images. However, relatively little work has applied these approaches to 2-manifolds of cortical surfaces. For example, the application of CNNs is extended onto non-structured data with geometric deep learning [211, 212]. Seong et al. [213] implemented a graph CNN that samples the data over a surface and reshapes the data to make it compatible with conventional CNN toolbox. Cucurull et al. [214] used graph based methods for performing parcellation of two regions (44 and 45) in Broca's area using structural and functional features on the cortical surface patches.

In this study, our goal is to label sulcal curves by using a U-Net model [210]. The novelty of our approach is a new application of deep learning on cortical surfaces using a U-Net model. We aim to achieve substantively faster labeling with better accuracy than multi-atlas approaches [53, 191]. Briefly,

we use a large dataset of healthy controls to train a neural network for performing sulcal curve extraction and labeling. We train samples with prelabeled sulcal curves and geometric feature maps. These features are mapped onto a 2-D polar plane to fully utilize a 2-D U-Net model. Specifically, we use the following geometric features: sulcal curve distance map, mean curvature, surface parcels, and spectral features (eigenfunctions of 2-manifold). We enhance the accuracy of the model using an independent dataset of manual labels. In the experiments, we show that our approach outperforms the standard curve labeling method [53] that is considered as reference.

2. Methods

2.1. Data Acquisition

We analyzed structural data from two cohorts of participants. We first constructed a sulcal curve atlas dataset using 21 subjects from the publicly available KIRBY21 database (11 males; 10 females; age range=22-61) [49]. Structural images in the atlas dataset were acquired with (3T Philips MPRAGE sequence with a 1x1x1.2 mm^3 resolution and an FOV of 240x204x256 mm^3). Our primary analyses were carried out using data from 784 participants in the Baltimore Longitudinal Study of Aging (BLSA) (349 males; 435 females; mean age=72 years; range = 25-99) [215]. Participants in this study were scanned 1-7 times over a period of 8.5 years on a Philips 3T scanner using a 3D "magnetization prepared rapid gradient echo" (MPRAGE) sequence. Each image had 170x256x256 voxels with 1.2x1x1 mm^3 resolution. The local Institutional Review Boards approved the study, and all participants provided written informed consent at each visit. From the BLSA dataset, we created 3 samples of participants. The training set consisted of 1-7 longitudinal scans from 759 individuals for a total of 1374 separate MRI sessions. The validation dataset was constructed from single sessions from 28 individuals that are not included as part of the training set. Single session data from 22 participants not included in the training or validation datasets were used to create a test dataset.

2.1.1. Sulcal curve atlas

For sulcal curve atlases from the KIRBY21 dataset, eight primary sulcal curves were manually labeled on both hemispheres by an expert according to a published sulcal curve labeling protocol [199]: central sulcus (CS), superior temporal sulcus (STS), superior frontal sulcus (SFS), inferior frontal sulcus (IFS), occipitotemporal sulcus (OTS), cingulate sulcus (CingS), calcarine sulci (CalcS), and olfactory sulcus (OLF) (see Figure VII-5 for examples).

2.2. Preprocessing

Structural data were processed for cortical surface reconstruction using the MaCRUISE pipeline [49]. Sulcal curves were extracted using the TRACE method [187] . Sulcal curves were labeled using a curve labeling protocol [53].

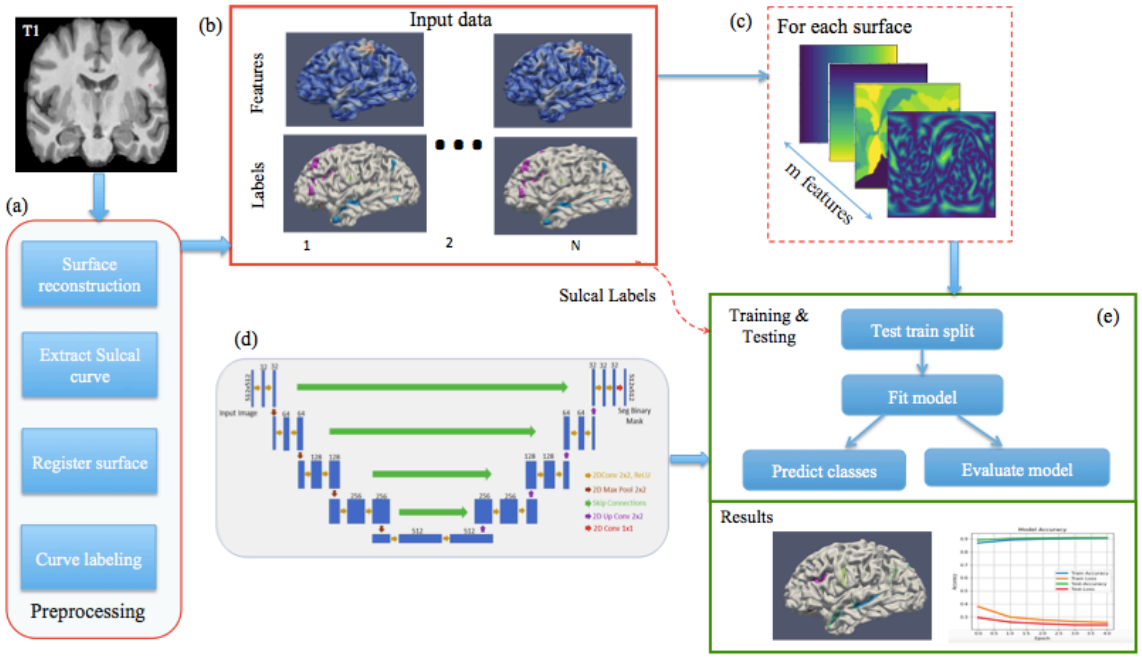

Figure VII-1. An overview of the proposed method: (a) Preprocessing steps include surface reconstruction, sulcal curve extraction, surface registration, and curve-labeling. (b) Cortical features are generated including distance maps, spectral features, mean curvature, parcellation, and sulcal labels. (c) Features defined on spherical polar coordinates are mapped onto a uniformly spaced grid. (d) The features are then passed into 2-D U-net model and fitted with batch size of 10 for maximum of 20 epochs. (e) Labels are predicted on the test set for evaluation..

All datasets were quality checked for each label and any faulty surfaces or extremely inconsistent curve labeling results are excluded. Note that we used these labels for training the deep learning model using various geometric features including distance map, surface parcels, mean curvature and spectral features. Cortical features were finally projected onto a plane to feed CNN for training with the U-Net model. An overview of the proposed workflow is shown in Figure VII-1.

2.2.1. Sulcal curve extraction

The TRACE method [187] was applied on each hemisphere to extract a set of sulcal curves along the sulcal fundic regions. Briefly, candidate sulcal points were selected to form a topological graph of the points, and Dijkstra trajectories over the graph to delineate optimal sulcal curves. This includes primary and secondary sulcal curves as a set of points (a subset of vertices) without curve labels, which needs to be pruned and labeled through the proposed pipeline. In this chapter, we used default parameter settings as suggested in [187]. The code is available at https://github.com/ilwoolyu/CurveExtraction.

2.2.2. Surface registration

Although surface registration is not a mandatory step for the reference method [53] it could improve the quality of sulcal curve-labeling after surface registration. We first mapped each hemisphere onto the unit sphere while minimizing area distortion [78] and then established cortical surface correspondence [216] (https://github.com/ilwoolyu/HSD). A custom template was obtained by averaging in a group-wise fashion [216] of co-registered 21 subjects from the publicly available dataset [179] of the Kirby Research Center for Functional Brain Imaging in Baltimore (http://mri.kennedykrieger.org/databases.html).

2.2.3. Sulcal curve labeling for model training

We used the standard curve labeling method [53] to assign labels to primary sulcal curves by pruning minor sulcal curves and branches. The reference method employs multi-atlas to determine a label for each individual sulcus. It establishes a point-wise curve correspondence with each atlas and finds the best match across the established correspondence. However, we found in this work that a majority vote shows better performance since the best match can sometimes work poorly if only partial perfect match (few points) with a particular atlas yields the highest score among the other atlases. Therefore, final labels were assigned if at least half of atlases agree.

2.3. Deep Neural Network for Sulcal Curve Labeling

We used cortical geometric features to capture both cortical folding patterns and individual variability. First, the sulcal curve distance map (geodesic distances between sulcal curves) was generated to represent cortical folding patterns. As complementary features to cortical folds, we used mean curvature and surface parcellation labels [217]. In addition, spectral features were computed to capture intrinsic geometric characteristics in the embedding space being spanned by the eigenvectors associated with the first five smallest eigenvalues. These features were fed into the neural network for sulcal label prediction.

2.3.1. Sulcal curve distance map

To represent cortical folding patterns, we computed a geodesic distance map $u: \Re^3 \to \Re$ on the cortical surface $\mathbf{\Omega}$ between the sulcal curves. We set all the extracted sulcal curves as a source $C \subseteq \mathbf{\Omega}$. By letting $c \subseteq C$, the minimum travel time $u(\mathbf{x}): \Re^3 \longrightarrow \Re$ from the source to any point $\mathbf{x} \in \mathbf{\Omega}$ can be obtained from the following propagation equation with some speed function $\exists F \, \epsilon \, \Re^+$,

$$u(c) = 0 \quad , \tag{7.1}$$

$$\| \nabla u(\mathbf{x}) \| \, F\left(\mathbf{x}, \frac{\nabla u(\mathbf{x})}{\|\nabla u(\mathbf{x})\|}\right) = 1 \quad . \tag{7.2}$$

This wavefront propagation formulation is a static Hamilton-Jacobi partial differential equation. We set a constant speed function in every direction by letting F=1 as follows:

$$\| \nabla u(\mathbf{x}) \| = 1 \quad . \tag{7.3}$$

By solving equation (7.3), we have the minimum travel time with unit speed. This is equivalent to the geodesic distance between sulcal curves, which generates the geodesic distance map u of the sulcal curves. Typically, gyral regions have high values whereas sulcal regions have a zero distance. We later feed this map to the network for training.

Given a threshold $\xi \, \epsilon \, \Re^{+}$, sulcal regions can be segmented by a binary mask $M: \Re \longrightarrow [0,1]$ as follows

$$M(u;\xi) = \; f(\mathbf{x}) = \begin{cases} 1, \text{ if } u \leqq \xi \\ 0, \text{ otherwise} \end{cases} \quad . \tag{7.4}$$

This threshold widens the extracted sulcal curves to prevent them to being too narrow to provide sufficient geometric information within sulcal regions. We empirically set ξ =10 to sufficiently cover regions along a single sulcus similar to the previous studies [187, 218].

2.3.2. Spectral features of cortical surface

Spectral features were generated for each cortical surface to feed the neural network. We first build a graph G = {V, E}, where V is a set of vertices and E is a set of edges. We then setup a |V | by |V | weighted adjacency matrix W that stores node affinities. A diagonal node degree matrix D encodes the sum of all the point affinities (vertex degree) at point *i*.

$$d_i = \sum_{j=1}^{|V|} W_{ij}. \tag{7.5}$$

A diagonal node-weighting matrix *G* is given by the exponential of negative mean curvature. The node weighting at point *i* is defined as,

$$g_i = e^{-h_i}, \tag{7.6}$$

where h_i is mean curvature at point i. Finally a general Laplacian operator is formulated on the connectivity adjacency matrix as the following |V | by |V | matrix:

$$L = G^{-1}(D - W). \tag{7.7}$$

Since L is symmetric, positive, and semi definite, its eigenvalues Λ and their associated eigenvectors U hold the following form:

$$L = U\Lambda U^{-1}\ . \tag{7.8}$$

We then use [148] to correct sign ambiguity in eigenvectors. Here we refer to spectra of a fixed subject being arbitrarily chosen from the dataset. The spectral features were given by the eigenvectors associated with the first five nonzero smallest eigenvalues, which were ultimately fed as five additional input channels into the neural network.

2.3.3. Planar mapping

In general, the cortical surface has a genus-zero form that is not yet fully compatible with a neural network optimized for a uniform grid representation. To address this issue, we represent a cortical surface with polar coordinates (θ, ϕ) at (x, y, z) as follows.

$$\theta = \arctan\left(\frac{\mathbf{y}}{\mathbf{x}}\right), \tag{7.9}$$

$$\phi = \arctan\left(\frac{\mathbf{z}}{\sqrt{\mathbf{x^2+y^2}}}\right), \tag{7.10}$$

where θ is the azimuth angle and ϕ is the elevation angle. However, a polar coordinate system inherently has non-uniform angular representations, which yields substantial length distortion around the poles. To reduce such distortion, we rotate each sphere such that the poles are located around insular and ventricle regions so as to minimize length distortion of sulcal curves around the poles [219]. The inspiration for choosing a pole is drawn from Auzias et al.'s approach [220]. Given a set of sulcal curves $C \subseteq \Omega$ for each of N subjects, we approximate global pole location by maximizing the following

objective function:

$$\hat{z} = \text{argmax}_{\|z\|=1} \sum_{i=1}^{N} \sum_{c \in C_i} \arccos^2(z \cdot c), \tag{7.11}$$

with underlying assumptions that the cortical anatomy on the spheres is roughly aligned and the initial value for **z** is chosen such that it is roughly located away from most of the sulcal curves (i.e., insular and ventricle regions). Spherical reparameterization is done by updating the pole as obtained from equation (2.11). In this way, we reduce the planar projection distortion on the primary sulci of interest.

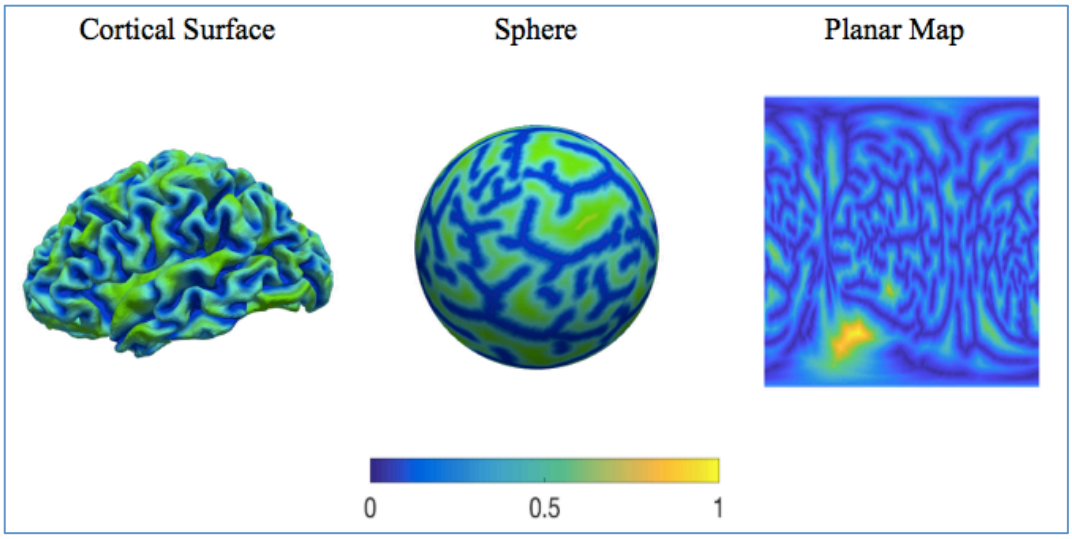

Figure VII-2 Qualitative representation showing normalized distance map feature on cortical surface, sphere and planar map.

After the planar projection, we resampled the plane with a 512 x 512 resolution that is optimized for convolutional layers of the neural network. The cortical features and label maps were then used for training. Figure VII-2 illustrates the distance map normalized between 0 and 1 on a cortical surface, a sphere and after projection to a planar map.

2.3.4. Training

The network architecture is inspired and motivated from the original U-Net design that has been

known to perform well for image segmentation tasks [221]. A 5-level architecture has been used for each sulcal label with curve overlap as the loss function. A batch size of 10 was used to meet with GPU memory constraints while also achieving better performance. Hence, the network weights are saved using check pointing and batch images are fed into the training model using a fit generator.

The network consists of 2D convolution layers followed by batch normalization and a rectified linear unit (ReLU) activation, which yielded better performance during preliminary evaluations. Max pooling is used after each convolution layer. In the last layer, 1×1 convolution produces probability of sulcal curve label from a sigmoid activation function. We can threshold this probability map to derive desirable label information. In this study the threshold was empirically set to 0.7. Each network was trained until convergence for a maximum of 20 epochs. Finally, as the network was trained on planar maps, the predicted labels were mapped back onto the unit sphere:

$$x = \cos\theta \cos\emptyset, \tag{7.12}$$

$$y = \cos\theta \sin\emptyset, \tag{7.13}$$

$$z = \sin\theta. \tag{7.14}$$

2.3.5. Post Processing

Predicted labels are generated based on the various features supplied to the training model and do not consider cortical shape to delineate labels to only primary sulcal curves. This can yield to issues such as partial labeling on the primary curves or labeling extra minor branches due to the limited ability in cortical shape representation in the U-Net.

As our goal is to label primary sulcal curves, the extra minor branches need to be pruned, and missing label information needs to be filled before further analysis (see Figure VII-3). Since a sulcal fundus is represented by a single curve, it is reasonable for final sulcal label decision to include or exclude whole branches as major sulci are generally delineated along the entire sulcal fundi. We propose

curve length-based pruning to trim the minor branches. Briefly, we measure both the length (i.e., geodesic distance on Ω) of the extracted sulcal curves in C by TRACE and that of the predicted sulcal curves by the proposed method. For each predicted curve, we compute a length ratio r proportional to its corresponding extracted curve in C. If the extracted and predicted curves are perfectly matched, we have $r = 1$, whereas $r < 1$ if the extracted curves are partially labeled. Thus, this quantity is used to prune the predicted curves ($r < 0.2$) or fill partially missing labels along the extracted curves ($r > 0.8$). Although we empirically choose such thresholds in this study, they could be learned from datasets for better decision-making.

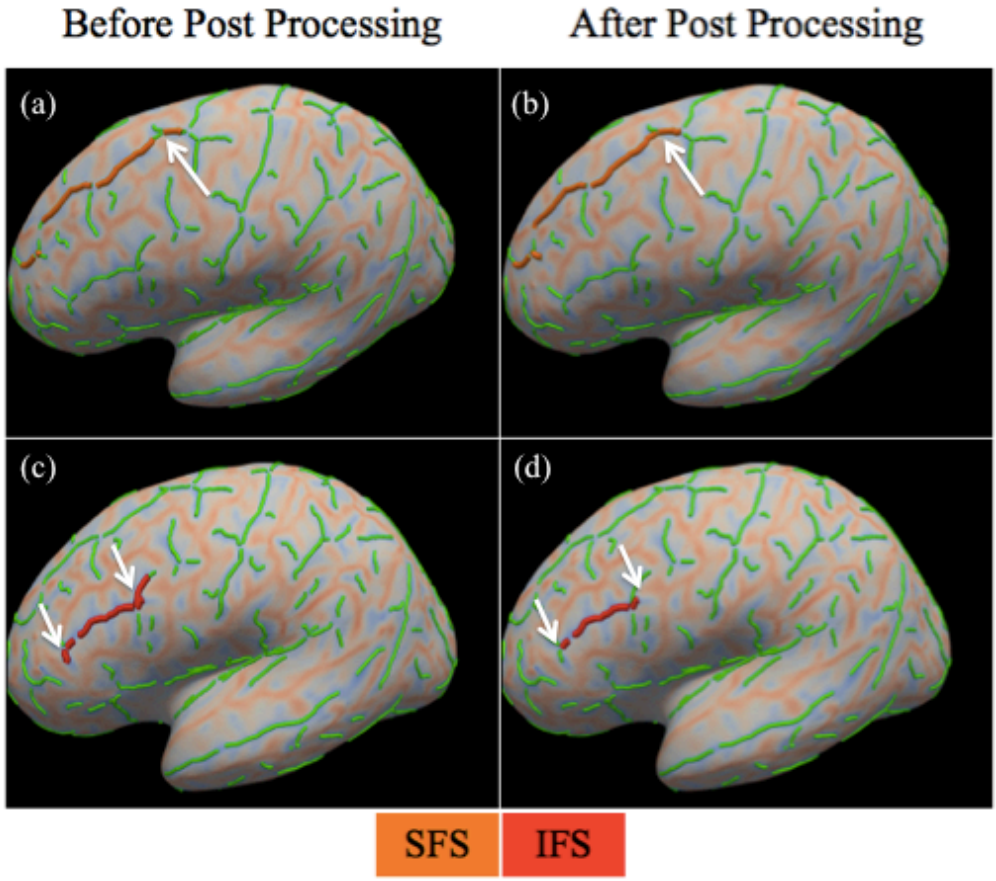

Figure VII-3. Example of pruning and filling of sulcal curves in post processing. Inflated surfaces are used for better visualization. Mean curvature is shown as the background on the inflated surface. Sulcal curves in C by TRACE are shown in green lines. Arrows indicate the areas of change before and after post processing. (a) SFS curve with missing branch before post processing. (b) SFS curve after filling. (c) IFS curve with extra branches before post processing. (d) IFS curve after pruning extra branches.

2.3.6. Evaluation

We used the sulcal curve editor (https://github.com/ilwoolyu/SulcalCurveEditor) to manually label major sulcal curves on 22 subjects via the sulcal delineation protocol [199]. These labels were used for evaluation to validate the proposed DNN predicted labels. These results were compared with a standard curve labeling method [53]. Since true curve correspondence was unavailable between major curves, the closest distance measures at sulcal points may not be able to capture the missing or extra curves. In particular, we have computed the sulcal Dice coefficient (SDC) as described below for evaluation. This measure summarizes labeling accuracy as well as false positives in a single metric. From equation (7.4), we have two masks M_A and M_B on a cortical surface Ω, which cover two corresponding major curves A and B, respectively, as follows.

$$M_A(\mathbf{x}) = M(u_A(\mathbf{x}); \xi) \quad , \tag{7.15}$$

$$M_B(\mathbf{x}) = M(u_B(\mathbf{x}); \xi), \tag{7.16}$$

where u_A and u_B are geodesic distance from A and B. SDC can be computed by taking surface integral of M_A and M_B.

$$SDC = \frac{\iint M_A(\mathbf{x}) \cdot M_B(\mathbf{x})\, \partial\Omega}{\iint M_A(\mathbf{x})\, \partial\Omega + \iint M_B(\mathbf{x})\, \partial\Omega}. \tag{7.17}$$

It is challenging to derive analytical solutions to the above surface integrals on an arbitrary cortical surface. In triangular mesh, we can instead approximate the solution by counting the numbers of vertices within the masks, where the numerator is the intersection between the vertices belonging to both M_A and M_B.

In evaluation, we computed SDC for the proposed and reference methods with respect to manual labeling. For each label, we summarized mean SDC values across all the subjects. Two sample t-tests are performed between the reference and proposed methods. Finally, we used false discovery rate (FDR)

[222] on resulting SDC for multi-comparison correction.

3. Results

We found that the proposed method significantly improved sulcal curve labeling in 9 out of 16 sulcal curves across the left and right hemispheres compared to the reference method [6], after multi-comparison correction via false discovery rate [223]. The average SDC is improved by 12.5 percent for the left hemisphere and 20.6 percent for the right hemisphere. Figure VII-4 shows mean SDC values for each of the eight curves on both hemispheres. There is marginal, but not statistically significant, improvement of SDC across both hemispheres in SFS (12.8%), STS (11.3%), and IFS (15.3%) using the proposed method. The highest improvement of SDC is seen in CS with over 28.8% in left hemisphere and 45.3% in right hemisphere with SDC above 0.93 in both hemispheres ($p<0.05$).

Figure VII-5 shows qualitative comparisons of the reference and proposed methods with manual labeling for a single subject on the left hemisphere. The predicted CS from the proposed method is consistent with manual labeling while the reference method only captures about a half of the total curve. Some false positives on minor branches are exhibited in the reference method for STS and SFS that are not present in the proposed method. The proposed method consistently label IFS agreed with manual labeling whereas the reference method includes extra branches (false positives) or is missing a portion of the corresponding curve (false negatives). CingS is consistent with manual labeling for both methods. There is a false positive branch apparent towards the frontal region in both methods (Figure VII-5), while the standard curve labeling method also has a false negative in the parietal region in which the curve label is not captured. OTS and OLF are consistent across all the methods except for extra minor branch on OLF.

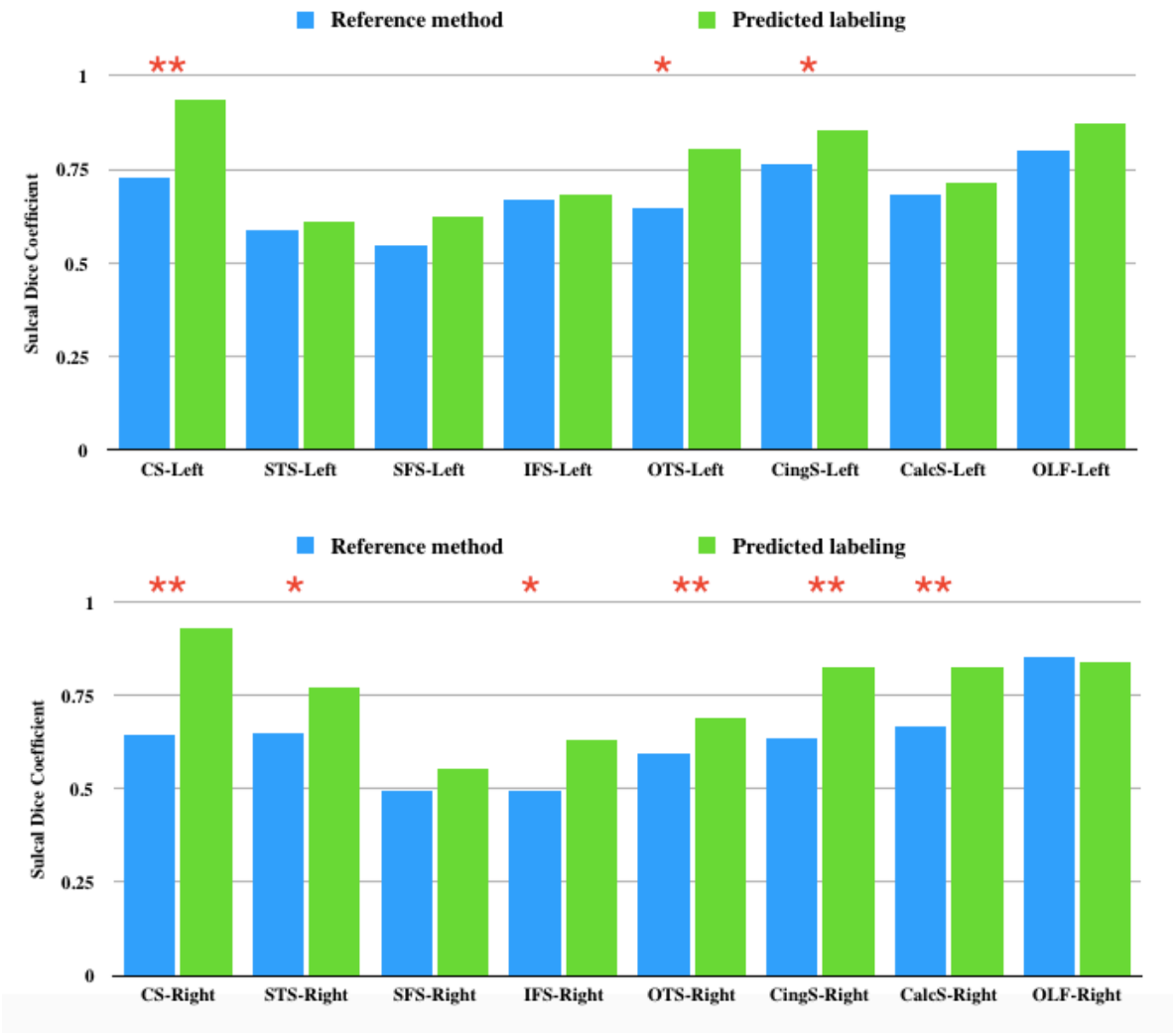

Figure VII-4. Sulcal Dice coefficients are shown across eight curves in both left and right hemispheres. * p<0.05. ** p < 0.005. Reference method shown in blue bar and Predicted labeling from DNN model is shown in green.

Although sulcal labels by the proposed method reasonably well agree with manual labeling, the proposed method sometimes misses a portion of the corresponding curve or shows extra branches of the curve, as shown in Figure VII-6. CS and OLF are well matched with manual labels (Figure VII-6a). While OTS is also well matched to the manual labels in this example, there are apparent false positives or false negatives in the remaining five curves (Figure VII-6b-g).

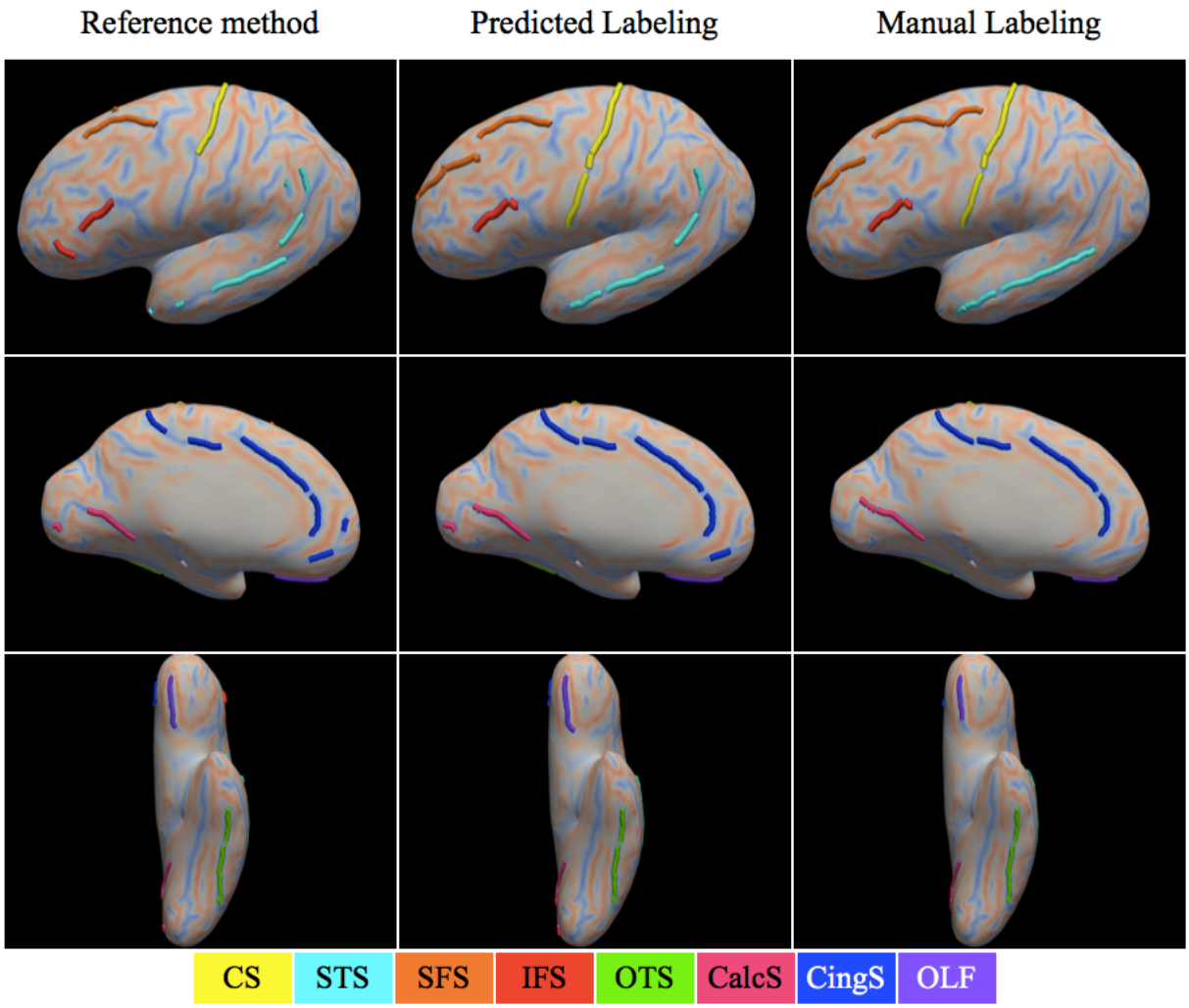

Figure VII-5. Qualitative results of eight curves on the left hemisphere for a single subject. Inflated surfaces are used for better visualization. Results are shown in 3 different views for reference, predicted labeling and manual labeling methods for comparison. The color code at the bottom indicates the color associated with each curve. Mean curvature is shown as the background on the inflated surface.

Figure VII-7 shows an overlay of all eight curves using manual labeling from 22 subjects. They are overlaid on the resampled template surface from 21 subjects in the Kirby21 dataset [49] after surface registration [48]. Heterogeneity is observed in most of the sulcal curves. In particular, variability in IFS appears to be higher in the right hemisphere compared to that of left hemisphere. Similarly, higher variability is seen across the frontal region of the SFS curve and the entire region of STS curves across the subjects.

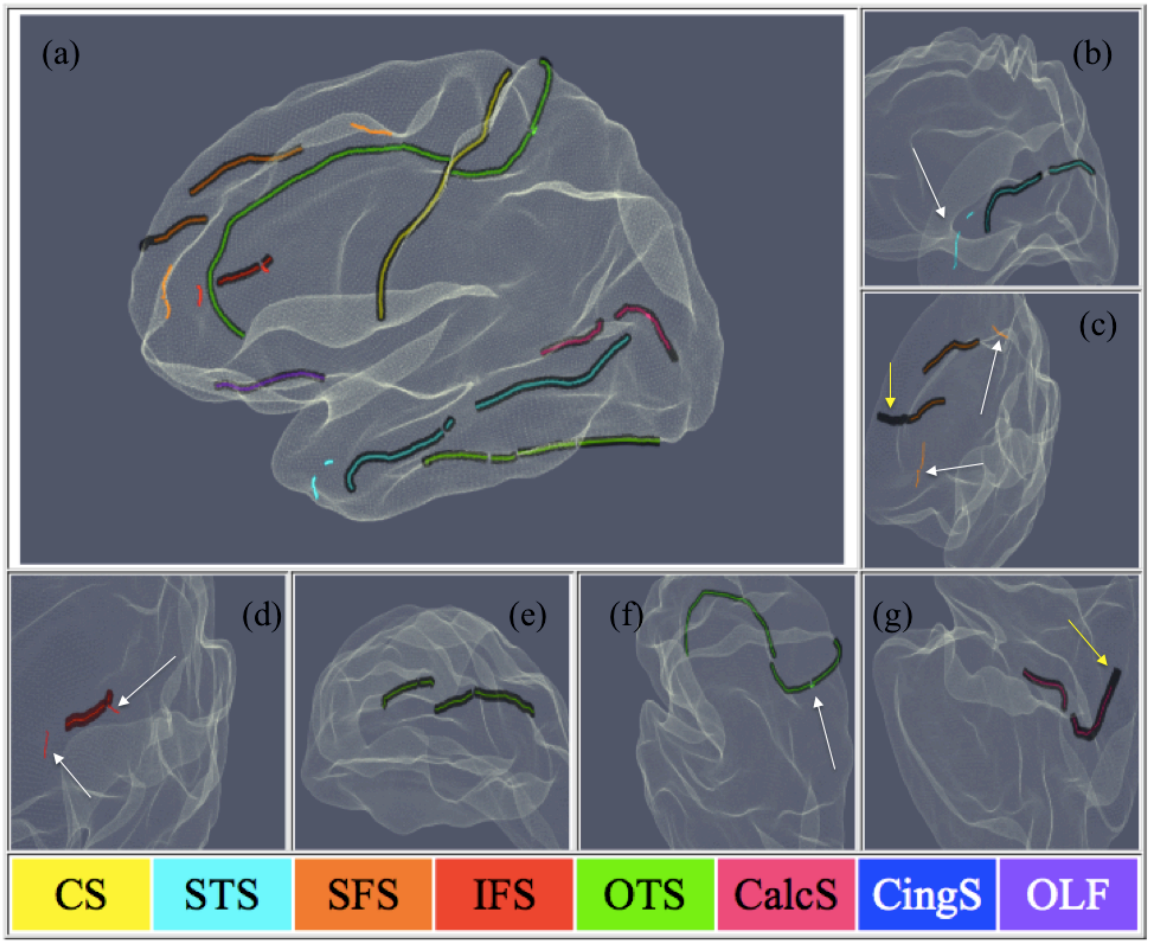

Figure VII-6: Illustrative example of inconsistencies seen in predicted labeling. Yellow arrows indicate false negatives and white arrows indicate false positives of predicted labeling compared to that of manual labeling. (a) Overlay of eight manual curves (black tube representation) and predicted curves (line representation) on a single subject. (b) STS with false positives. (c) SFS with a false negative and two false positive branches. (d) IFS with false positives. (e) OTS with no false positives or false negatives. (f) CingS with false positive. (g) CalcS with false negative. All the curves in (b-g) are rotated in the best view for the corresponding curves.

Figure VII-8 highlights the disparities in left and right hemisphere between the proposed method and manual labeling. Shorter curves are seen in the frontal region of SFS for labeling from the proposed method compared to that of manual labeling in the left hemisphere. In the right hemisphere, the difference in curve lengths of the two methods is not as pronounced as seen in the left hemisphere [224, 225]; however, there is higher curve variability in the frontal region. In contrast, STS displays false positives

with longer curves in labeling from the proposed method compared to that of manual labeling in the left hemisphere. Moreover, there is also higher curve variability in the left hemisphere compared to that of the right hemisphere. This is in agreement with the higher SDC seen in right hemisphere for STS compared to that of the left hemisphere (Figure VII-4). Similarly, CingS has higher variability in the predicted curves in the left hemisphere compared to that of the right hemisphere.

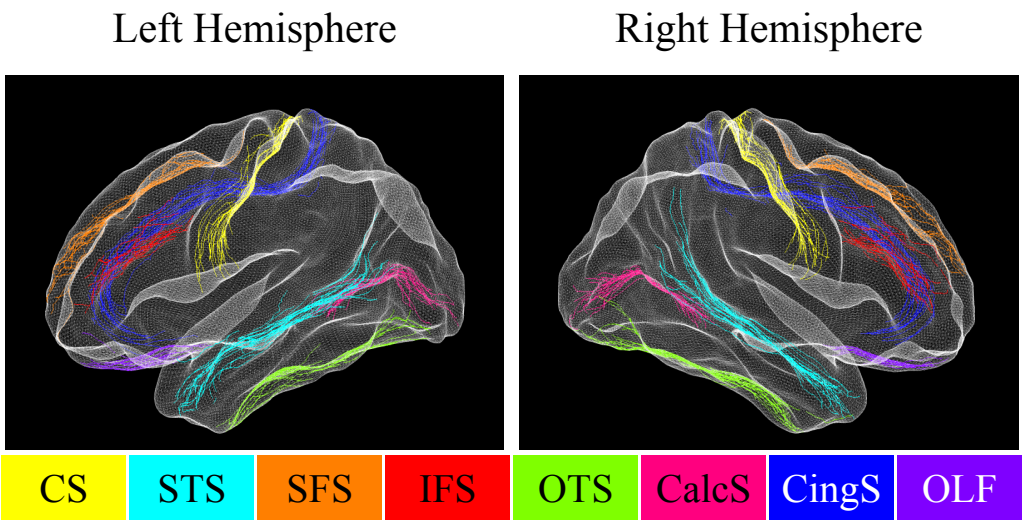

Figure VII-7. Eight major sulcal curves manually labeled across 22 subjects in left and right hemispheres.

1. DISCUSSION

Deep learning has shown high efficiency and scalability on large datasets [200, 226]. In supervised learning approaches, initial labeling is essential in training for segmentation or classification tasks in medical imaging [221, 227]. This is time-consuming for manual sulcal labeling or semi-automated approaches. In the proposed method, we collected large datasets and employed automated processing pipelines [53, 187] for curve extraction and labeling to produce training sets with reasonably plausible initial labels. We further performed liberal manual QA of all the curves to ensure filtering out extreme outliers in the data. This step helps prevent the model in learning from outlier examples while

retaining the variability needed for training. As initial labels were generated automatically, the liberal QA was not as time-consuming as the manual or semi-automated approaches. However, it could further be improved by incorporating labeling uncertainty as proposed in [228, 229] that can guide in filtering out the data used for training. This will be promising in future work.

A higher SDC is achieved in all the sulcal curves in both hemispheres except for OLF in the right hemisphere (Figure VII-4). This might be because the reference method already has good performance on OLF in the right hemisphere (SDC=0.85). The false positives or negatives shown in Figure VII-6 could have been partially caused because of the heterogeneity across the population among these major sulcal curves, as illustrated in Figure VII-7. Such variability has been acknowledged and measured in many previous studies [189, 230, 231]. Although the focus in our study is not to measure such variability, a qualitative analysis would be helpful in understanding the major curves. For example, different variability in IFS between hemispheres may explain lower SDC in the right hemisphere than that of the left hemisphere (Figure VII-3).

Low performance in certain sulcal curves could also be attributed to labeling inconsistency obtained in the proposed method as highlighted in Figure VII-8. Possibly, the inconsistency across the predicted labels in certain anatomical regions has risen from the training data acquired by the reference method even after liberal manual QA. The high variability in sulcal curves with several discontinuous pieces [188] could be another contributing factor. Despite the existence of such high variability, the proposed method achieved better SDC values compared to that of reference method. The contributing factors for improved performance could be the use of multiple geometric features and parcellation labels for training the large dataset used in the proposed method. We have included multiple (1-7) longitudinal sessions for individual subjects in training data to create a large dataset needed for deep learning to perform well allowing it to generalize to new examples without over fitting. Also having the regularization step though layers like drop out is another factor that is known to help generalizing to new examples. Even though multiple sessions per subject are used in training, we have only included one

session per subject in testing for fair comparison. In this work, we have shown the possibility of using a deep learning approach that integrates both sulcal curve extraction and labeling to achieve improved curve-labeling results.

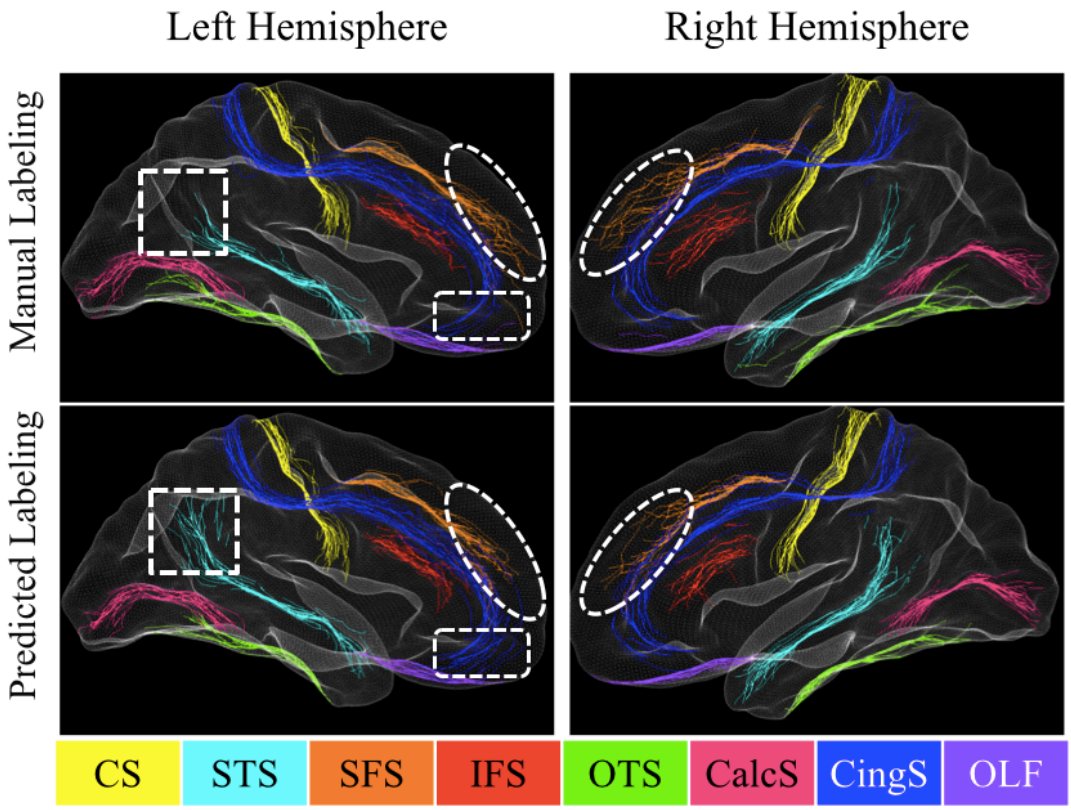

Figure VII-8. Illustration highlighting Sulcal curves variability differences between manual and proposed predicted labeling in left and right hemispheres. Dotted squares show differences in STS curves in the left hemisphere. Dotted ovals highlight the differences in SFS curves in both hemispheres. Dotted rectangles indicate differences in CingS curves towards the frontal region.

Although the proposed method outperformed the reference method in most curves, there is still a room for improvement. First, in order to use existing architecture for U-Net models designed for uniform grid data, we have mapped cortical surface features onto a plane, which causes mapping distortions. We alleviated distortions to some extent by choosing pole locations and placing the labels away from the poles, yet they are not fully addressed but attenuated. Second, while results from [53] are based on a

young healthy population, the majority of our sample is middle to older age adults. This difference in training and test population demographics could have led to failures in labeling certain curves due to high variation in sulcal shapes between the two age groups. Although we removed outlier cortical surfaces and labeling results from training data, there is still variability and inconsistency within the data used for training this model.

In future work, we will investigate recent spherical CNN methods [232-234] or graph CNN methods [235] in training these datasets to overcome the distortion issues caused with planar mapping of such data. Our approach is not limited to a curve-labeling problem, but it could be extended to tackle other problems like cortical parcellation.

4. Conclusion

In this method, we proposed a DNN predicted labeling method using a U-Net model for labeling eight major sulcal curves on each hemisphere. The proposed method has shown improvement in most of the curves in both left and right hemispheres. We observed significant improvement in CS, OTS, and CingS in the left hemisphere and six curves in the right hemisphere including CS, STS, IFS, OTS, CingS, and CalcS ($p<0.05$). The overall SDC is improved by 16.6 percent for both hemispheres. CS showed the highest improvement with over 28.8 and 45.3 percent in the left and right hemispheres, respectively, compared to those of the reference method. This provides an opportunity to explore the application of deep learning techniques to methods involving cortical surfaces.

The trained model and code are available on the Github project: (https://github.com/MASILab/slabelDNN).

CHAPTER VIII. QUALITY ASSESSMENT OF GSBSS AND CAPTURING OBSERVATIONS ON DIFFERENT CLINICAL DATASETS

1. Introduction

Understanding changes in the gray matter through imaging parameters is an essential aspect of many clinical studies,[152, 236] including mindfulness meditation [237-239] and aging effects [236]. However, partial volume effects and alignment issues need to be corrected before considering for further analysis [152, 240]. GSBSS has been shown to overcome partial volume effects to improve the sensitivity of detecting underlying changes in imaging parameters [241]. Method reproducibility is essential for advancing the science, as it is the primary basis for verifying a scientific result. Therefore, it is important to understand the individual components and specifications involved in the GSBSS process in preparation for transferring the data onto a target surface for cross-subject analysis. Preliminary analysis showed improved statistical power for DW-MRI and functional MRI datasets using GSBSS on a single study (Chapters III and IV). The accuracy of the method needs to be carefully reviewed before applying to various clinical datasets for hypothesis testing.

In this chapter, we first describe the preprocessing steps for GSBSS (Section#2) and then develop and describe quality-checking measures with examples that can be used for evaluation of GSBSS results before considering for group analysis (Section#3). Since structural T1 is used for registering subject space metrics to a target surface, segmentation labels obtained on the T1 image are used for evaluation of reproducibility and robustness of the approach. In addition, reproducibility of metrics of interest from other modalities is reported in terms of accuracy and coefficient of variation (CV) on a scan rescan dataset (Section#4). Robustness of the method is evaluated using Dice similarity coefficient (DSC) on multiple clinical datasets that are acquired on different scanners and sites in diverse population

(Section#5). Finally, interpretability of the results is presented on clinical applications with expected outcomes (Section#6).

2. Preprocessing

In this section, the preprocessing steps involved in obtaining the input data for GSBSS processing are described. Individual T1 images are segmented using multi-atlas segmentation into 132 brain regions and 1 background. Gray matter surfaces are derived using MaCRUISE [49] and the central surfaces are used for further processing in GSBSS. Diffusion-weighted images (DWI) are processed using a protocol described in [242]. For NODDI processing [45], the DWI file, scheme file, and mask are passed to the AMICO package (https://github.com/daducci/AMICO/tree/master/matlab), which is a fast implementation of NODDI with a linear approximation. Intracellular volume fraction V_{ic}, isotropic volume fraction (V_{iso}), and orientation dispersion index (ODI) are obtained from AMICO method.

3. Quality assurance

In this section, quality assurance components involved in GSBSS processing are described. When applying GSBSS to new datasets each step needs to be reviewed. Quality can be assessed based on both qualitative and quantitative measures as described in Figure VIII-1. We assess quality at four levels, first to ensure the quality of input data from structural T1 and also for metrics of interest from other modalities, then to ensure the quality of intermediate volume registration and corresponding surface deformation, and finally to ensure the quality of output data through both qualitative and quantitative measures. These processes are described below.

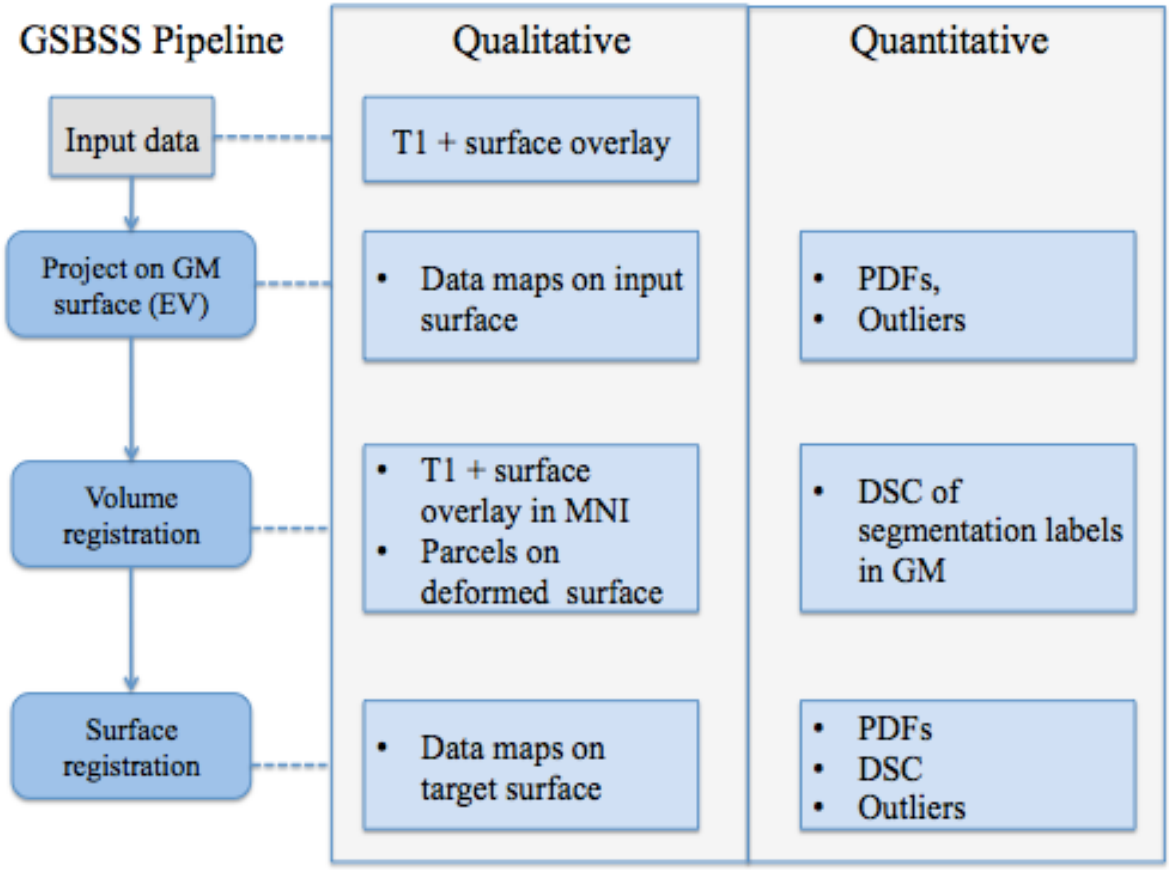

Figure VIII-1. Flowchart indicating the steps in GSBSS processing and corresponding qualitative and quantitative QA measures. The first column shows the steps involved in GSBSS and the dotted line indicates corresponding QA measures produced at that step.

3.1. Quality of input data

First, the overlay of the central surface on structural T1 images ensures that cortical surface well represents the underlying cortical structure. Then, data maps projected on the input cortical surface provide a qualitative assessment of input data in gray matter. These are obtained by projecting volume data onto the cortical surface based on the enclosing voxel approach. Our report includes metrics of interest with data from other modalities like diffusion MRI and functional MRI along with the data on structural T1 like segmentation labels and gray matter fraction images. Images from other modalities need to be first co-registered to structural T1 to take advantage of the mapping from native space onto the target surface. This is an important preprocessing step and any failure happening at this step could

seriously impact further analysis. To address any issue arising from this step, complementary measures for quality checking are incorporated that can take advantage of a priori information of the expected image intensity values in the GM region. For example, the orientation dispersion index (ODI) is higher in gray matter and lower in white matter. So lower values of ODI on the cortical surface could indicate partial volume effects from white matter regions.

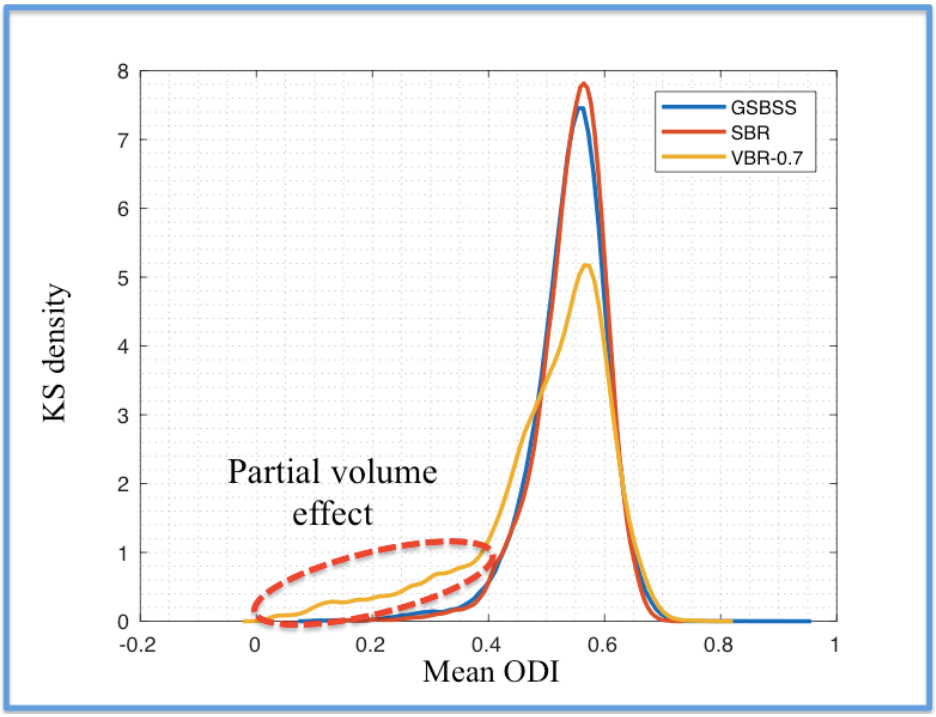

Figure VIII-2. ODI intensity distribution is indicated using kernel-smoothing function estimate for volume based metric with the application of 0.7 thresholds for GM masks along with SBR and GSBSS based ODI.

The probability density estimate of mean ODI measure over the target surface is shown in Figure VIII-2 for GSBSS. Surface-based registration (SBR) and volume based registration (VBR) methods provide an expected range of ODI on GM surface. It can be seen that GSBSS and SBR measures have higher ODI and do not have partial volume effects in comparison with that of the volume based registration measure.

Using this criterion, the percentage of outliers is captured with a customizable threshold T to check if data on the surfaces is falling outside the expected range. In equation (8.1), N indicated the

number of vertices of the surface Ω, and p is the metric of interest on Ω,

$$Outliers = \frac{1}{N}\sum_{i=1}^{N}(p_i < T)\times 100 \quad (8.1)$$

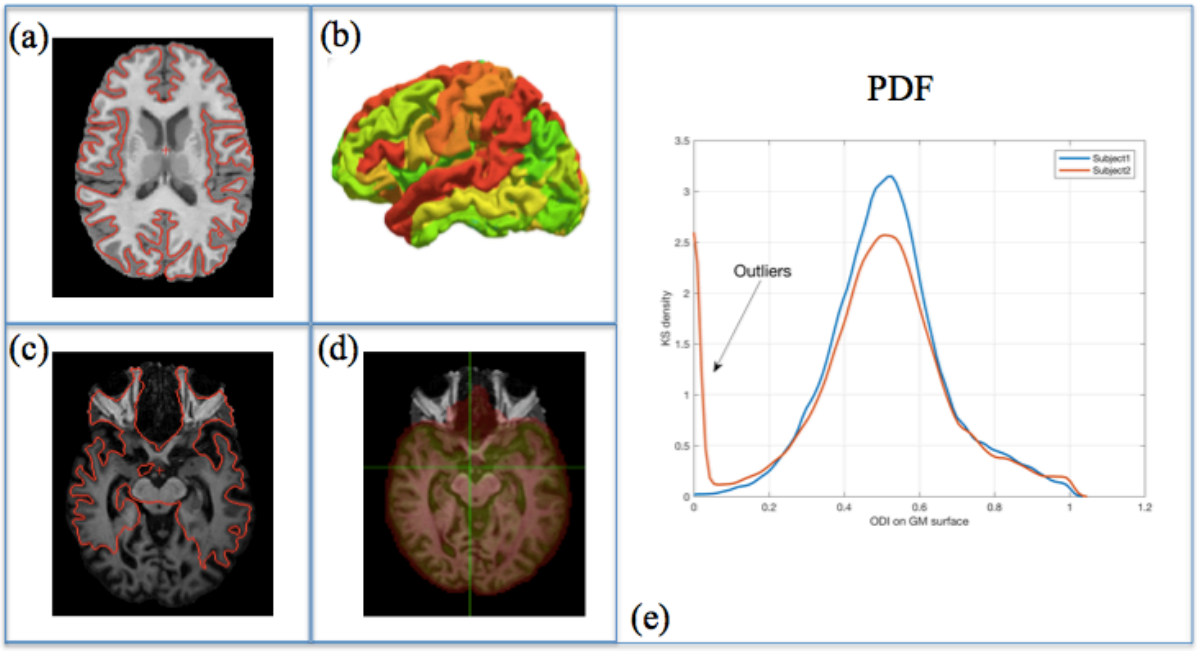

Figure VIII-3. Example of quality checking measures in two subjects. (a) Overlay of the cortical surface on structural T1 for Subject#1. (b) Segmentation label map on input cortical surface of Subject#1. (c) Overlay of the cortical surface on structural T1 for Subject#2. (d) Overlay of co-registered ODI on structural T1. (e) A PDF map of ODI for Subject#1 and Subject#2. Outliers are identified in Subject#2 due to acquisition quality issues in structural T1 and problems due to segmentation.

Figure VIII-3 illustrates two examples of the good and bad quality of input data. Outlier % measures from equation (1) captures this scenario (Figure VIII-3e) indicating > 10% outliers in Subject#2 with ODI threshold set to 0.1.

3.2. Quality assurance at intermediate steps.

The T1 image is non-linearly registered to MNI template to obtain initial deformation for aligning T1 images to template space. The inverse deformation obtained in this step is then applied to the vertices of the cortical surface in native space. This step will provide the deformed surface that is in alignment with the volume image registered to MNI template. Figure VIII-3a illustrates a good example with an

overlay of the central surface on the anatomical image after deformation to the MNI template with cortical surface following the underlying gray matter structure. Quality can be quantified using segmentation labels that are transferred to MNI space after applying volume transformation and projecting the labels onto the deformed cortical surface using enclosing voxel approach. Dice similarity coefficient (DSC) is used as a quantification measure by comparing the overlap of the labels on the deformed surface with that of the reference labels from the MNI template projected onto the deformed surface. Lower mean DSC across all the segmentation labels indicates an issue in the registration step.

$$DSC = \frac{2|X \cap Y|}{|X|+|Y|} \tag{8.2}$$

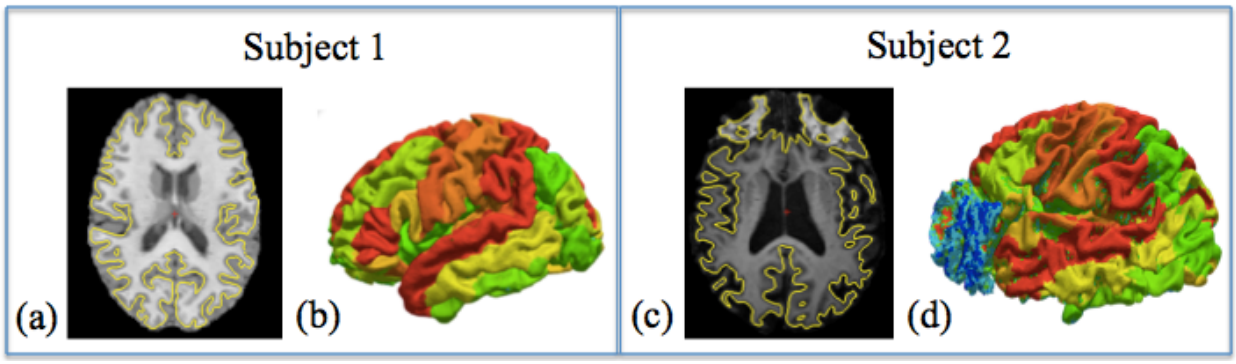

Figure VIII-4. Example of quality checking measures in two subjects in the intermediate step after volume deformation. (a) Overlay of deformed surface on structural T1 in MNI for Subject#1. (b) Segmentation label map on deformed cortical surface of Subject#1. (c) Overlay of deformed surface on structural T1 in MNI for Subject#2. (d) Segmentation label map on deformed cortical surface of Subject#2. (Blue region indicates white matter region that is showing up due to cortical surface reconstruction issue)

3.3. Quality assurance of output data after surface registration

In order to achieve cortical correspondence across multiple subjects, surfaces in MNI space are registered to a target template surface. This step provides a mapping that can be used to transfer the properties from each of the vertices on the subject surface to the target template surface. To evaluate the accuracy of the processing results quantitatively, DSC is computed. Lower DSC indicates lower accuracy. For example, DSC of the two subjects is illustrated in Figure VIII-5 where Subject#1 has higher DSC,

indicating good processing result and the Subject#2 with lower DSC indicating registration failure.

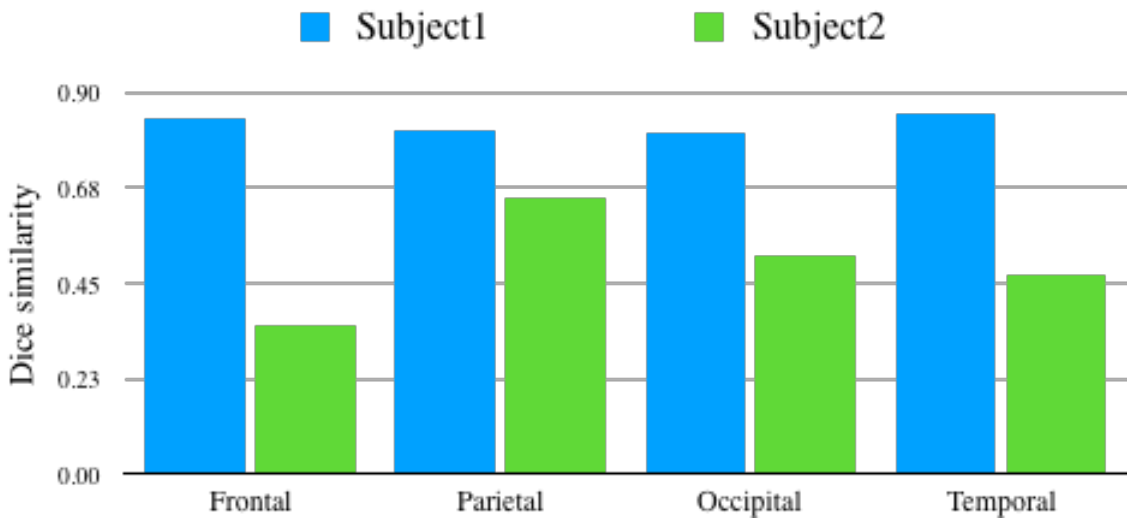

Figure VIII-5. Dice similarity coefficients for two subjects in frontal, parietal, occipital and temporal lobes. Subject#2 has lower DSC across all the four lobes.

Qualitative results of two additional subjects with lower DSC are illustrated in Figure VIII-6.

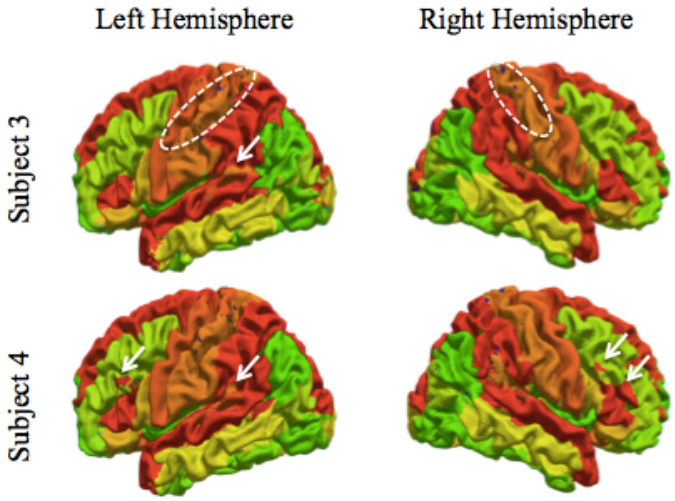

Figure VIII-6. Two example subjects of outliers with lower Dice similarity coefficients. Arrows indicate the regions of mislabels at different locations on the left and right hemispheres. The dotted vowel regions indicate the CSF vertices showing up along the parietal region.

The mislabeling in these subjects could have been caused due to registration issues, unlike the issue that is highlighted in Subject#2 which has a segmentation problem.

In addition to DSC values, the quantitative measures of PDF and outliers % as in equation (1) described in section 3.1 for input data can also be obtained for output data. As part of first level quality checks, the percentage of outliers is calculated as indicated in equation (1) for both inputs and output parameters of GSBSS. Before considering the data for group analysis, these measures could be evaluated to determine the filter criteria.

The top outliers from the underlying dataset can be further reviewed to determine the cause of failure. These measures could then be used for filtering out the major outliers from the group analysis data. Figure VIII-7 illustrates an example of the percentage of outliers in input data and output data from multiple subjects. Outlier1 is from the Subject#2 example that was shown in earlier sections. The difference in scaling could be because of the different number of vertices in the subject surface versus target surface in addition to the mapping effect when transferring the metrics using surface registration. The scatter plot indicates that the outliers from input data are propagated to GSBSS outputs.

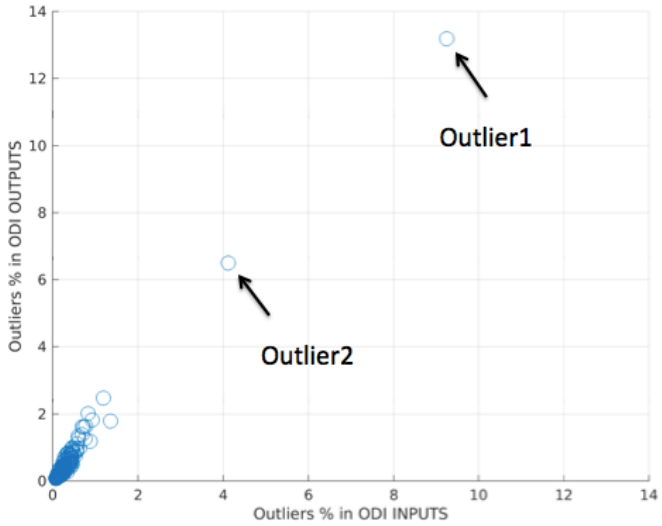

Figure VIII-7. Scatter plot of Outliers % for both inputs and outputs associated with ODI. There are 2 major outliers in the dataset as indicated by the arrows.

3.4. Identifying reason for failure from quantitative measures

Consolidating all the above strategies, we demonstrate three examples that can help identify the underlying issues based on quality checking measures (Figure VIII-8).

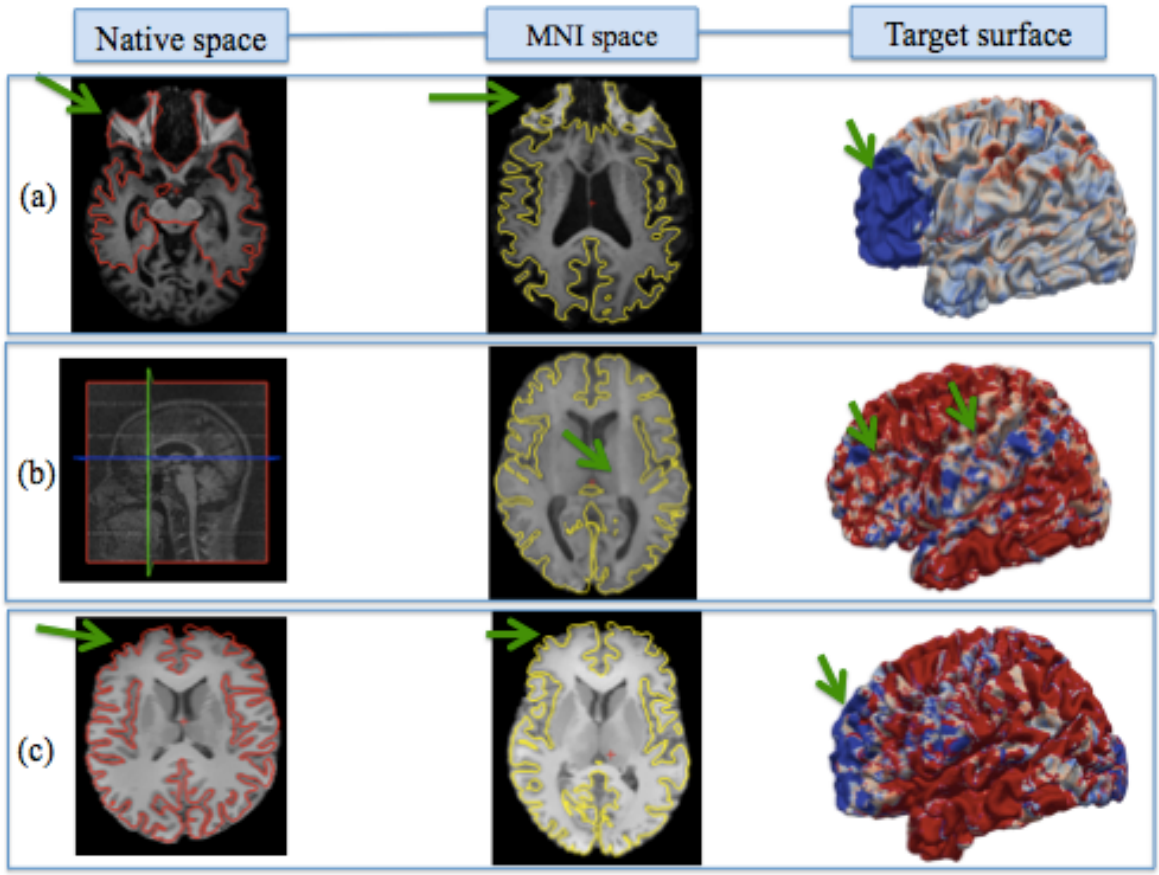

Figure VIII-8 Three examples are illustrating the issues caught from quality checking measures that helped to identify underlying issues. (a) Skull stripping issue identified based on outliers in ODI measure and lower DSC. (b) Image quality issue of structural T1 scan identified based on outliers in gray matter fraction measure. (c) Surface reconstruction issue was leading to missing regions of volume image identified based on outliers in gray matter fraction. Arrows highlight the problem areas in each of these examples. (Blue regions indicate regions of lower GM fraction).

Each of these scenarios is different, and the proposed cumulative approach of assessing quality through multiple measures presented in this section can help identify the underlying problem.

- In the first example (Figure VIII-8a), the subject showed a high percentage of outliers based on intensity threshold criteria on ODI and low DSC on the target surface. This issue is caused

because of the skull stripping issue that led to failure in subsequent steps of segmentation, surface reconstruction and co-registration (Subject#2).

- In the second example (Figure VIII-8b), a higher percentage of outliers are seen on gray matter fraction measure. This issue is caused due to the image quality issues in structural T1.
- The third example (Figure VIII-8c) illustrates an issue with surface reconstruction that caused an increased number of outliers on gray matter fraction in the frontal region.

4. Reproducibility study

In this section, two experiments are conducted (1) to determine the effect of skull stripping of structural T1 on GSBSS reproducibility using DSC measure and (2) to evaluate scan reproducibility and method reproducibility using DSC and reliability measures.

4.1. Data

Six subjects with multiple scan sessions with a total of 42 sessions are considered from the data of scan-rescan study acquired on a 3T (Achieva, Philips Medical Systems, Best, The Netherlands) equipped with a 32-channel head coil located at the Vanderbilt University Institute of Imaging Sciences. Each subject has two repeats of structural T1 (1 mm resolution, FOV=256 x 256, TE=2ms, TR=8.95 ms and TI=643 ms) for each session and a diffusion-weighted imaging scan protocol (2.5 mm isotropic resolution, FOV=96 x 96, TR=2.65s, TE=101ms, Gmax = 37.5 mT/m) with two shells of b=1000 s/mm^2 (32 directions each) and one shell of b=2000 s/mm^2 (60 directions). The Vanderbilt University Institutional Review Board approved the study and all participants provided written informed consent prior to enrolling in the study. Co-registration is performed using "epi_reg_12dof" from FSL toolbox.

4.1.1. Quantitative evaluation with DSC

Any challenges that arise in volume registration step [243] are also applicable to that of GSBSS method. For example, registering the raw images is known to be a much more challenging problem than that of the registration of the images without the skull [244]. The effect of skull-stripping in the overall

accuracy of GSBSS is evaluated using DSC for quantitative validation. The DSC of 42 sessions in 6 subjects was evaluated for two scenarios with,

(a) Raw structural T1 of the subject and a reference template with skull, and

(b) Bias-corrected skull stripped subject T1 and a reference template without the skull.

From Figure VIII-9, DSC of the bias-corrected T1 as input has better performance results in frontal, parietal, occipital and temporal lobes. In all cases, it is noticed that occipital lobe has lower DSC.

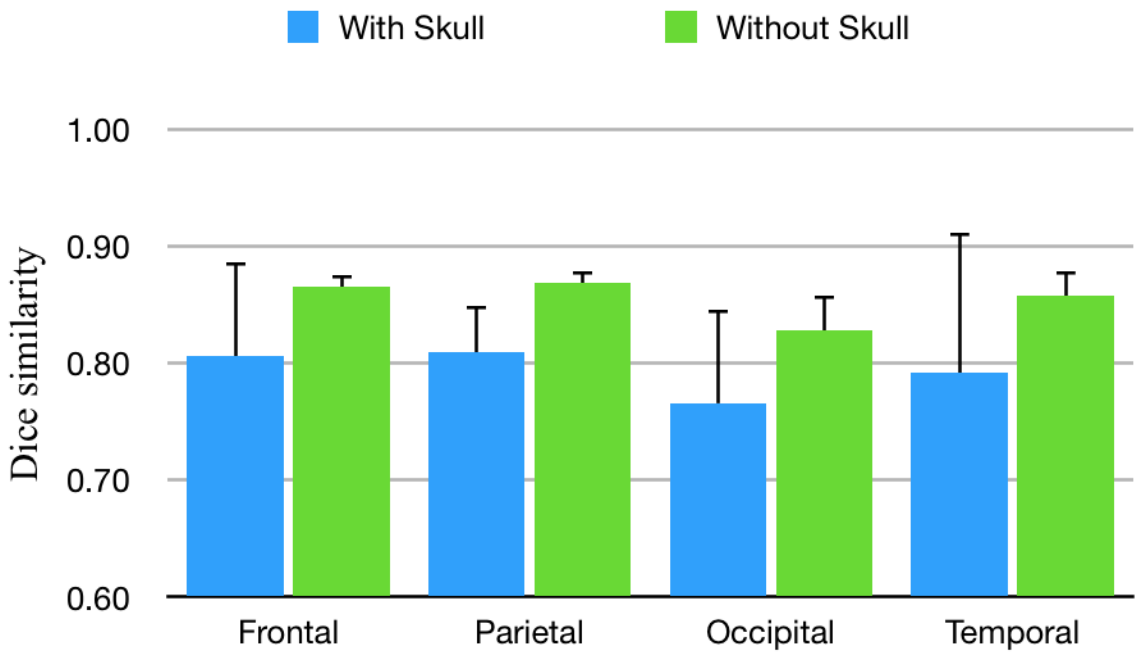

Figure VIII-9 Mean dice similarity coefficient across 42 sessions in 6 subjects is shown for (a) raw structural T1 with reference template with skull and (b) bias-corrected skull stripped structural T1 with skull stripped template.

4.2. Method reproducibility with DSC

The GSBSS pipeline is run twice as run1 (r_1) and run2 (r_2) on the same set of inputs for eight sessions in four subjects. Mean DSC across all the sessions with two repeated GSBSS processing runs is shown for 49 cortical surface labels in both left and right hemispheres in Figure VIII-10.

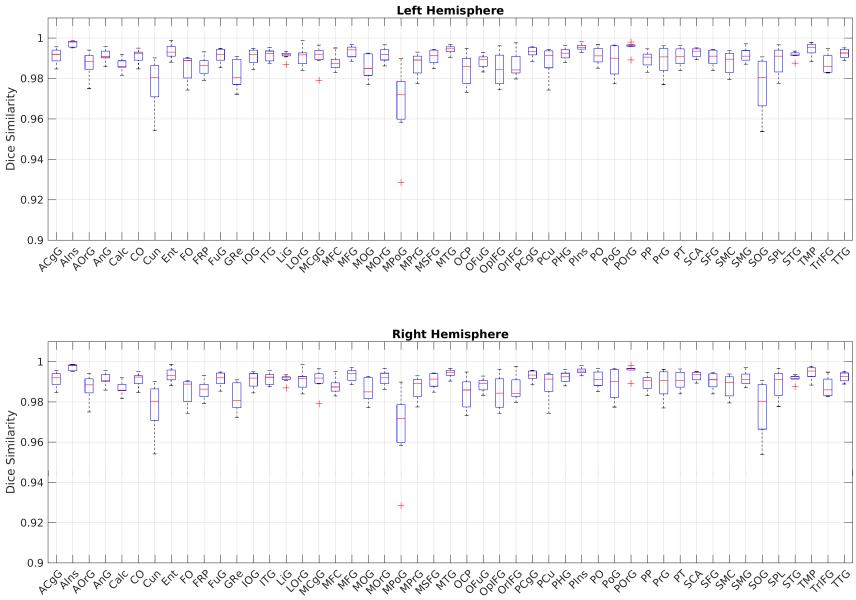

Figure VIII-10 Mean dice similarity coefficient across 49 labels in two hemispheres for repeated processing of GSBSS in 8 sessions.

Overall mean DSC of 0.99 is achieved in both hemispheres. Though the DSC is high in most of the cortical ROIs, there are few ROIs (e.g., MPoG) that exhibited lower DSC in both hemispheres indicating there could be some instability arising with the pipeline in these regions.

4.3. Reliability measures

Scan reproducibility for NODDI metrics and cortical thickness from structural T1 with GSBSS are reported using coefficient of variation (CV) and accuracy measures. Mean and standard deviation between run1 (r_1) and run2 (r_2) is calculated for each of the sessions and reproducibility is evaluated with CV and accuracy measures. Mean and standard deviation values across all the sessions for all subjects are calculated and CV is reported for V_{ic}, ODI, V_{iso}, and cortical thickness measures in GSBSS. In the following equations, n is the number of subjects and m is the number of sessions for each subject. In CV_{scan} measure, for each session s, average of (r_1) and (r_2) is taken as session measure $S_m = \{s_1, s_2, s_m\}$. Percentage of CV is defined as standard deviation over mean as defined in equation (8.2). Similarly, CV for scan reproducibility (CV_{scan}) and GSBSS reproducibility (CV_{gsbss}) are calculated as shown in

equations (8.4) and (8.5).

$$CV = \frac{\sigma}{\mu} \times 100 \tag{8.3}$$

$$CV_{scan} = \sum_{j=1}^{n} \frac{\sigma(S_m)_j}{\mu(S_m)_j} \times 100 \tag{8.4}$$

$$CV_{gsbss} = \sum_{j=1}^{n} \sum_{i=1}^{m} \frac{\sigma(r_1, r_2)_{ij}}{\mu(r_1, r_2)_{ij}} \times 100 \tag{8.5}$$

Accuracy is calculated as shown in equation (8.6) where μ is the mean of the measure from repeated runs of GSBSS processing and ε is the difference in the measurement from the repeated processing at each vertex.

$$accuracy = \frac{\mu}{\mu + abs(\varepsilon)} \tag{8.6}$$

Reproducibility measures of mean accuracy, mean coefficient of variation (CV) for repeated GSBSS processing and for repeated scans are reported in Table 1 for cortical thickness, V_{ic}, ODI, and V_{iso}.

Table VIII-5: Reproducibility measures of accuracy, CV in GSBSS reproducibility and CV in scan reproducibility, are reported for cortical thickness, V_{ic}, ODI, V_{iso}

Reliability measures	Accuracy	CV %	CV – Scan %
Thickness	0.96	2.8	10.7
V_{ic}	0.96	2.9	11.3
ODI	0.98	1.83	7.4
V_{iso}	0.8	23	74

Higher accuracy and lower CV are seen for GSBSS processing reproducibility in V_{ic}, ODI, and cortical thickness measures. However, V_{iso} is unstable with high CV and relatively lower accuracy measure. One reason for higher CV could be because V_{iso} is either too low or has a value closer to 1 on most of the vertices on the cortical surface. So any vertex having a difference would capture higher variance.

Vertex level reproducibility measures are shown in Figure VIII-11 for accuracy and CV of GSBSS

processing reproducibility (Figure VIII-11a-b) and scan reproducibility (Figure VIII-11c). Cortical thickness has lower accuracy and higher CV along the sulcal patterns. V_{ic} displayed lower accuracy and higher CV in the parietal region. ODI has higher accuracy and lower variance compared to that of other measures. V_{iso} is unstable in temporal and frontal regions. For scan reproducibility, CV is relatively higher in occipital region and sulcal patterns are not as obvious as seen in CV of GSBSS reproducibility. V_{ic} has a higher CV in parietal region and ODI has lower CV compared to other measures similar to that of CV from GSBSS reproducibility.

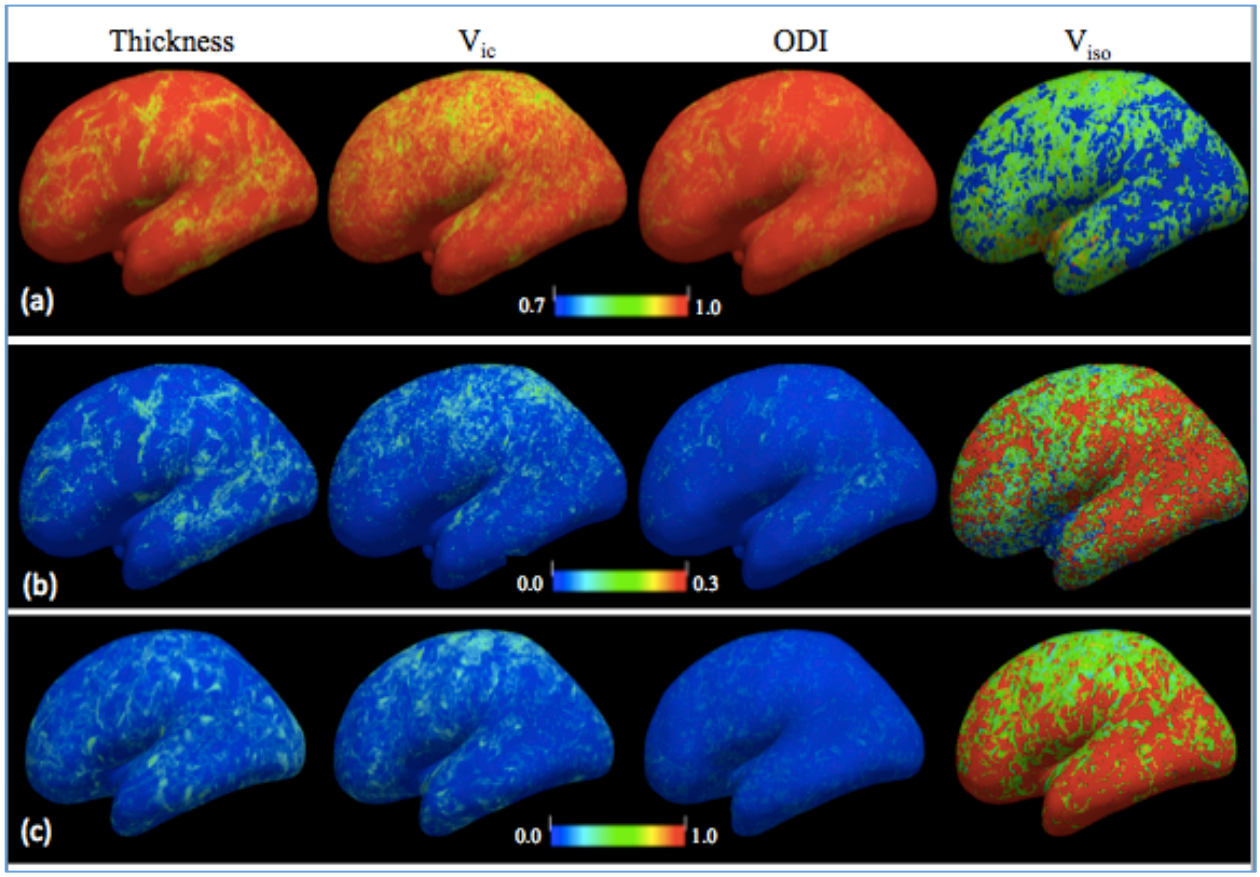

Figure VIII-11 Reproducibility measures at vertex level are shown for cortical thickness, V_{ic}, ODI, and V_{iso}. (a) Accuracy for GSBSS processing reproducibility. (b) The coefficient of variation for GSBSS reproducibility. (c) The coefficient of variation for scan reproducibility using GSBSS.

5. Robustness

For evaluating the robustness of the method across multiple clinical datasets, two experiments are conducted, first quantitative validation is performed using mean DSC, and, then, qualitative validation is performed by computing mean ODI and comparing with mean ODI from HCP minimum processing pipeline.

5.1. Data

In this study, four clinical datasets that are acquired on different scanners, sites and different populations are considered as summarized in Table 2.

Table VIII-6: Data acquisition parameters of four clinical datasets used in the robustness evaluation.

Dataset	Clinical population	Scanner	T1 acquisition parameters	DW-MRI acquisition parameters	Number of sessions
Psychosis	Healthy and psychosis	Phillips 3T	256x256 FOV 1x1x1 mm	b=1000/2000 s/mm^2 24/60 directions 96x96 FOV, 2.5 mm resolution	185
BLSA	Healthy, MCI and AD Ages: 50 – 98 years	Phillips 3T	256x256 FOV 1x1x1.2 mm	b=700/700 s/mm^2 32/32 directions 0.81×0.81×2.2 mm 260x260 FOV	1664
MBSR-1	Healthy controls and MBSR participants Ages; 23 – 52 years	Siemens 1.5T	256x256 FOV 1x1x1.33 mm	NA	66
MBSR-2	Healthy controls and MBSR participants Ages: 19-50		256x256 FOV 1x1x1 mm	NA	152

Study 1 includes psychosis population who participated in an on-going study of brain connectivity in neuropsychiatric disorders with one session per subject. Study 2 data is from the Baltimore Longitudinal Study of Aging (BLSA). Study 3 and 4 has data from individuals enrolled in

mindfulness-based stress reduction (MBSR) courses along with healthy control subjects. The corresponding local Institutional Review Boards approve each of the studies, and all participants provided written informed consent at each visit. De-identified access is obtained to these datasets.

5.2. Experiment 1: Quantitative validation with DSC measure

Mean DSC across all the subjects within each of the four datasets acquired on different scanners are shown in boxplots in Figure VIII-12. These are reported for frontal, parietal, occipital and temporal lobes separately. In all the datasets mean DSC is ~0.85 for frontal lobe except for MBSR data on the 3T scanner that have shown slightly higher DSC.

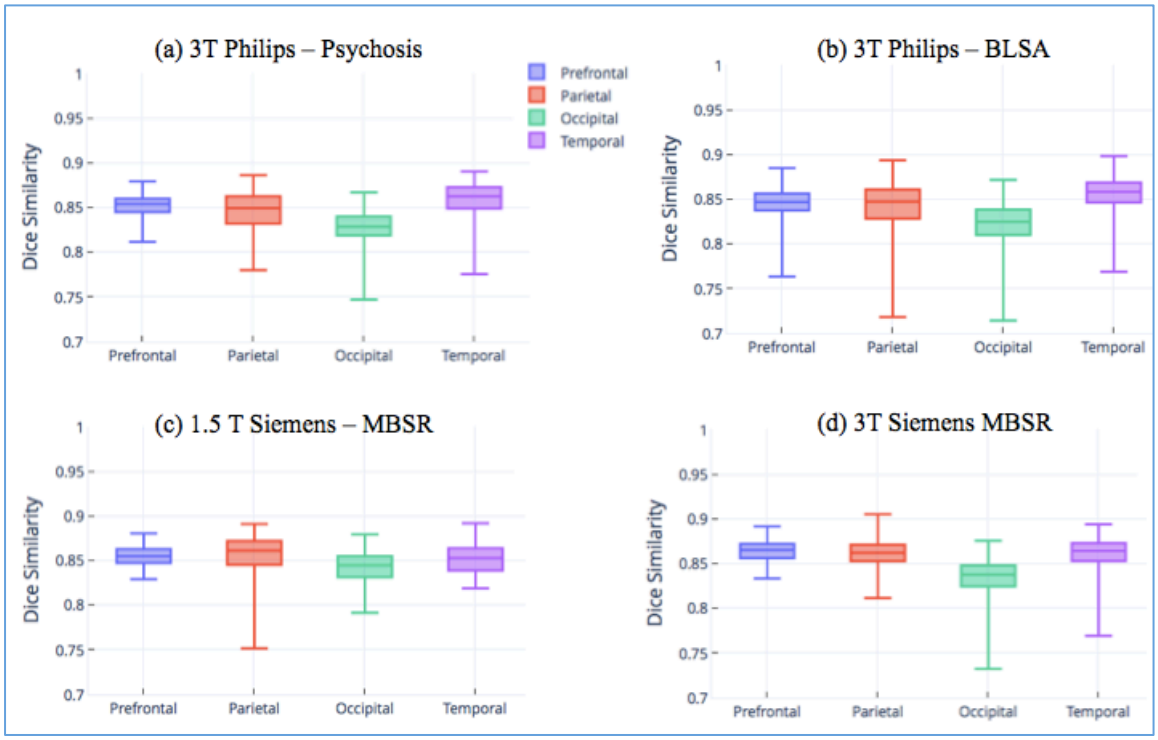

Figure VIII-12 Mean dice similarity coefficient in frontal, parietal, occipital and temporal lobes is shown on different datasets as (a) 185 sessions in psychosis population on 3T Philips scanner (b) 1658 sessions in aging population on 3T Philips scanner. (c) 66 sessions in MBSR data acquired on 1.5T Siemens scanner (d) 152 sessions in MBSR data acquired on 3T Siemens scanner.

While the mean DSC remained similar, higher variation is seen in parietal lobe for BLSA data and MBSR data acquired on 1.5T Siemens scanner. The occipital lobe has lower DSC compared to the

other three lobes.

5.3. Experiment 2: Qualitative validation of mean ODI

In this experiment mean ODI from GSBSS processing with and without search is computed for two clinical datasets of interest and compared with that of mean ODI from HCP minimum processing pipeline.

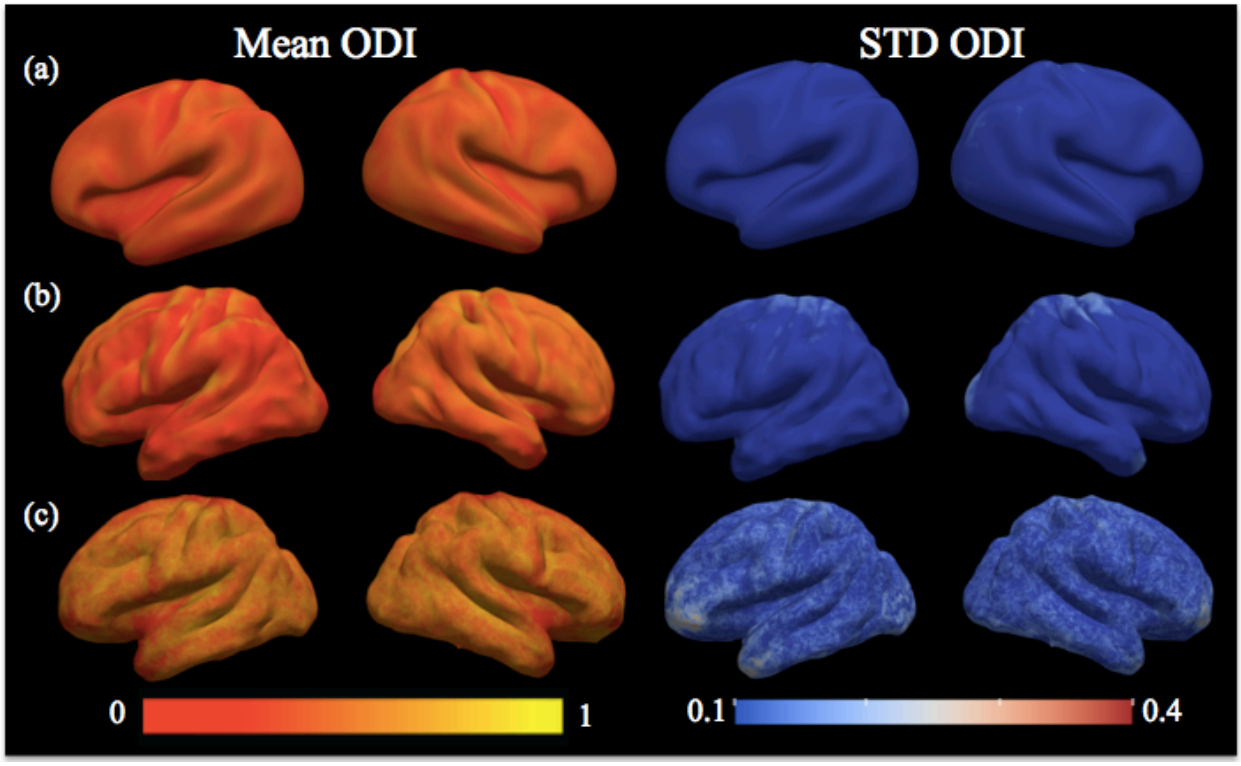

Figure VIII-13 Mean and standard deviation of ODI values are shown from (a) HCP minimum processing pipeline on 163 healthy subjects in clinical dataset 2 of the aging population. (b) GSBSS processing on 163 healthy subjects in clinical dataset 2 of the aging population with no search. (b) GSBSS processing on 28 healthy subjects with the 2mm search.

For this experiment, gender-matched random healthy subjects are chosen from BLSA data and Psychosis dataset. Qualitative results of mean ODI values for 163 healthy subjects from BLSA data and for 28 healthy subjects from psychosis dataset are shown on a template surface. In BLSA data, as the data is acquired on the single shell, no search is performed. Results from both HCP minimum processing pipeline and GSBSS processing results are shown for this dataset (Figure VIII-13a-b). The patterns of

mean ODI are comparable between these methods with higher values along the gyral regions. Overall, GSBSS has higher ODI compared to that of HCP results. However, the standard deviation of ODI values from GSBSS processing is higher in the parietal and occipital regions compared to that of HCP processing. In the psychosis dataset, the mean and standard deviation of ODI from GSBSS processing with 2 mm search are shown in Figure VIII-13c. Here, overall mean ODI is noted to be higher than that of the mean ODI result from BLSA data with no search. And the mean ODI pattern on the cortical structure is not as prominent as that seen in either HCP processing result or that of GSBSS processing result from BLSA dataset. Also, higher standard deviation is seen across the entire surface.

6. Clinical validations

To validate the applicability of the GSBSS processing results in clinical datasets with the expected outcome from the literature, three experiments are conducted and observations are captured.

(a) ODI is negatively associated with FA [245]. In order to validate this expected result, the Pearson correlation between ODI and FA is reported on clinical datasets 1.

(b) Cortical thickness is negatively associated with ODI [245]. In order to validate this expected result, the Pearson correlation between ODI and cortical thickness is reported on clinical datasets 1.

(c) Mean diffusivity is shown to be higher in older age (>70 years) compared to that of the relatively younger (50-59 years) age group [236]. To test this hypothesis, the difference in mean diffusivity between two age groups (a) 50- 59 years with gender-matched 28 subjects and (b) 70-79 years with gender-matched 29 subjects is reported on BLSA data.

With 28 healthy subjects, quantitative validation showing the correlation between ODI and cortical thickness are shown in Figure VIII-14a indicating a negative correlation with ($R = -0.22$). The negative correlation of ODI with FA is shown in Figure VIII-14b with ($R = -0.75$).

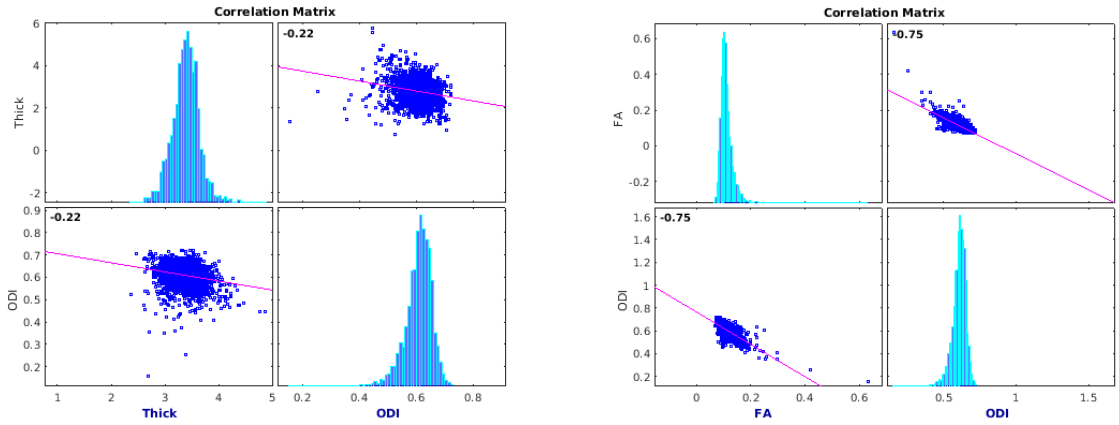

Figure VIII-14 Pearson correlation coefficient between (a) ODI and cortical thickness. (b) ODI and FA.

Finally, the difference in MD values between two groups of ages 50-59 years and 70-79 years is captured to compare with that of the observations reported in previous studies [236]. MD is significantly higher in all the four lobes in the 70-79 years age group compared to that of 50-59 years age group ($p<0.001$).

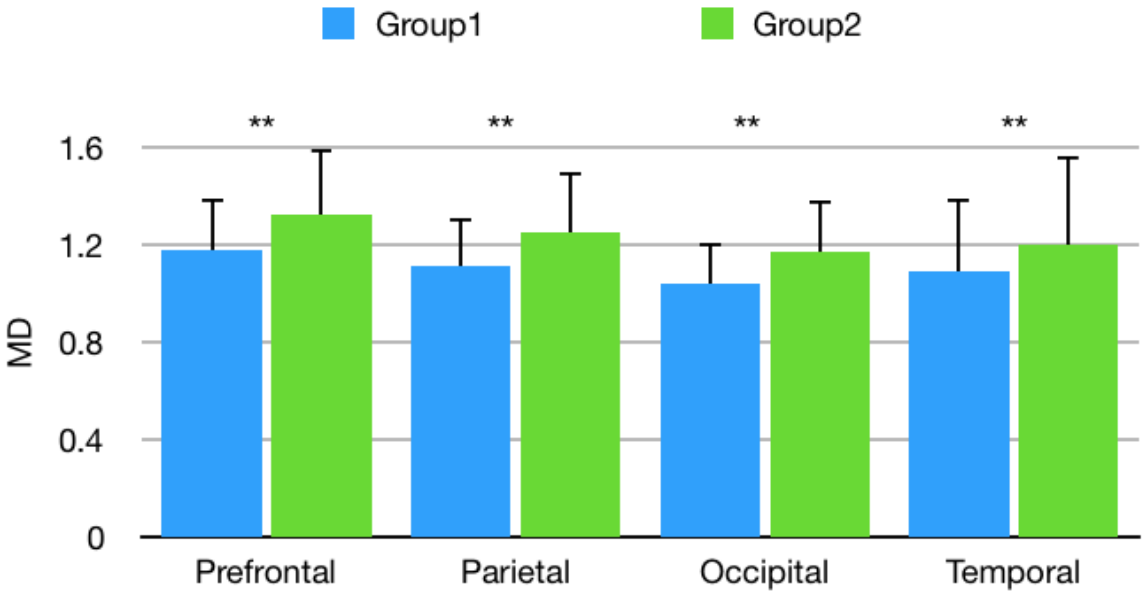

Figure VIII-15 MD in gray matter in four lobes for two groups of 50-59 age range (blue bar) and 70-79 age range (green bar). P<0.001 for a two-sample t-test is indicated by **.

7. Discussion and conclusion

In this chapter, quality checking aspects involved in GSBSS methodology are presented in detail with the goal of encouraging the reproducibility of the method. From various reliability and reproducible measures that are reported, the occipital lobe appears to have lower accuracy compared to that of other regions in all the experiments. CSF vertices are showing up in some of the regions as highlighted in the dotted oval region in Figure VIII-6. While some of these could have occurred from the volume image mapping, further evaluation is needed to check if any of these are appearing as a result of the specific surface registration method that is adapted for this pipeline. Based on the DSC reported across different datasets, GSBSS seems to be consistent across all the four lobes including large datasets (Figure VIII-12). However, based on the robustness check there is higher variance in ODI values for search criteria (VIII-13) that needs further attention in evaluating the accuracy of the measures. The mean ODI values in HCP minimum processing pipeline appear to be less than that of the GSBSS approach with or without search (Figure VIII-13). Lower values could be due to the partial volume effects arising from thinner cortex regions [245].

When compared to that of Fukutumi et al. [245] paper, there is a weaker correlation between ODI and thickness in clinical dataset#1 ($R = -0.22$ vs $R = -0.46$) while ODI vs FA has a stronger negative correlation ($R = -0.75$ vs. $R = -0.53$). The search strategy used in the GSBSS method could be one reason for this deviation in addition to the number of differences between the two datasets in comparison like demographics, data acquisition, and processing.

The experiment conducted for MD differences based on age group is chosen to be consistent with that of the Lauren et al. method [236]. Mean diffusivity values are slightly higher in the clinical dataset#2 compared to that of the [236] method. Significantly increased MD is seen across all the four lobes unlike the two lobes reported in [236] which could indicate improved sensitivity with GSBSS approach.

In this chapter, quality-checking aspects of GSBSS are presented, followed by the evaluation of the method for reproducibility, accuracy, and robustness. Following are the lessons learned capturing the

areas of improvement that need to be considered for further improvement of accuracy and robustness of the method.

a) Registration accuracy in both volume and surface registration is important for the accuracy of GSBSS.

b) There are two steps in GBSSS processing that depend on header information of the structural T1 and corresponding cortical surface reconstructed from structural T1. The first step is the application of inverse deformation from volume registration to the vertices on the surface. The second step is for retrieving the metrics of interest from volume image using enclosing voxel approach to project on the surface. Any discrepancy in the header fields would yield incorrect results and caution need to be exercised in this area when applying to newer datasets.

c) The GSBSS processing and analysis results are presented for a whole brain surface. It may have some drawbacks in causing labeling inconsistencies between the two hemishpheres. However, having a hemi-separation step prior to GSBSS processing could further aid in the flexibility of incorporating surface specific approaches like retesselation and surface registration that can be useful for second level processing.

d) While spectral matching is used for surface registration method, as it is flexible to incorporate whole brain surface without the need for reparameterization as needed for spherical based registration methods, additional surface registration approaches could be incorporated and would need to be validated.

e) The preliminary normal search is proposed based on higher ODI. While this seems to improve sensitivity for the changes occurring in pure gray matter, these results may have to be carefully reviewed if a regional variation is essential for the study of interest. Also, there seems to be higher variance in the ODI when using search. Additional validations need to be performed to improve the accuracy of the measure.

f) V_{iso} metric appears to be having lower accuracy in the preliminary analysis. Further evaluation is needed to establish the validity of this metric in different datasets.

g) While raw data is kept to the extent possible for the metrics of interest in volume images, further evaluation needs to be performed if the median filter or any other regularization technique could aid in reducing the noise in volume images in native space that could reduce some of the variance seen in the search based approach.

APPENDIX A: PUBLICATIONS

Journal Articles

1. **Prasanna Parvathaneni**, Vishwesh Nath, Justin A. Blaber, Kurt G. Schilling, Allison E. Hainline, Ed Mojahed, Adam W. Anderson, and Bennett A. Landman. "Empirical reproducibility, sensitivity, and optimization of acquisition protocol, for Neurite Orientation Dispersion and Density Imaging using AMICO." Magnetic resonance imaging (2018). https://doi.org/10.1016/j.mri.2018.03.004
2. **Prasanna Parvathaneni**; Ilwoo Lyu; Yuankai Huo; Baxter P. Rogers; Kurt G. Schilling; Vishwesh Nath; Justin A Blaber; Allison E Hainline; Adam W Anderson; Neil D. Woodward; Bennett A Landman. "Improved gray matter surface based spatial statistics in neuroimaging studies." In PLOS One. (Submitted)
3. **Prasanna Parvathaneni**; Vishwesh Nath; Maureen McHugo; Yuankai Huo; Susan Resnick; Neil D. Woodward; Bennett A Landman; Ilwoo Lyu. "Improving Human Cortical Sulcal Curve Labeling in Large Scale Cross-Sectional MRI using Deep Neural Networks." - In Journal of Neuroscience Methods. (Submitted)
4. Vishwesh Nath, Kurt G. Schilling, **Prasanna Parvathaneni**, Justin Blaber, Allison E. Hainline, Zhaohua Ding, Adam Anderson, and Bennett A. Landman. "Empirical estimation of intravoxel structure with persistent angular structure and Q-ball models of diffusion weighted MRI." Journal of Medical Imaging 5, no. 1 (2018). https://doi.org/10.1117/1.JMI.5.1.014005
5. Allison E Hainline, Vishwesh Nath, **Prasanna Parvathaneni**, Justin A. Blaber, Kurt G. Schilling, Adam W. Anderson, Hakmook Kang, and Bennett A. Landman. "Empirical single sample quantification of bias and variance in Q‐ball imaging." Magnetic resonance in medicine (2018). https://doi.org/10.1002/mrm.27115
6. Kurt G Schilling, Vishwesh Nath, Justin A. Blaber, **Prasanna Parvathaneni**, Adam W. Anderson, and Bennett A. Landman. "Empirical consideration of the effects of acquisition parameters and analysis model on clinically feasible Q-ball imaging." Magnetic resonance imaging (2017). https://doi.org/10.1016/j.mri.2017.04.007
7. Yuankai Huo*, Justin Blaber*, Stephen M. Damon, Brian D. Boyd, Shunxing Bao, **Prasanna Parvathaneni**, Camilo Bermudez Noguera, Shikha Chaganti, Vishwesh Nath, Jasmine M. Greer, Ilwoo Lyu, William R. French, Allen T. Newton, Baxter P. Rogers, and Bennett A. Landman. "Towards Portable Large-scale Image Processing with High-Performance Computing". Journal of Digital Imaging (2018). https://doi.org/10.1007/s10278-018-0080-0
8. Schilling, Kurt G., Vishwesh Nath, Colin Hansen, **Prasanna Parvathaneni**, Justin Blaber, Yurui Gao, Peter Neher et al. "Limits to anatomical accuracy of diffusion tractography using modern approaches." NeuroImage 185 (2019). https://doi.org/10.1016/j.neuroimage.2018.10.029
9. Huo, Yuankai, Zhoubing Xu, Shunxing Bao, Camilo Bermudez, Hyeonsoo Moon, **Prasanna Parvathaneni**, Tamara K. Moyo et al. "Splenomegaly Segmentation on Multi-modal MRI using Deep Convolutional Networks." IEEE transactions on medical imaging (2018). https://doi.org/10.1109/TMI.2018.2881110
10. Hansen, Colin B., Vishwesh Nath, Allison E. Hainline, Kurt G. Schilling, **Prasanna Parvathaneni**, Roza G. Bayrak, Justin A. Blaber et al. "Characterization and correlation of signal drift in diffusion weighted MRI." Magnetic resonance imaging 57 (2019). https://doi.org/10.1016/j.mri.2018.11.009

11. Nath, Vishwesh, Kurt Schilling, **Prasanna Parvathaneni**, Allison E. Hainline, Yuankai Huo, Justin A. Blaber, Matt Rowe et al. "Tractography Reproducibility Challenge with Empirical Data (TraCED): The 2017 ISMRM Diffusion Study Group Challenge." BioRxiv (2018): 484543. https://doi.org/10.1101/484543

Highly selective Conference Publications

1. **Parvathaneni, Prasanna**, Baxter P. Rogers, Yuankai Huo, Kurt G. Schilling, Allison E. Hainline, Adam W. Anderson, Neil D. Woodward, and Bennett A. Landman. "Gray Matter Surface based Spatial Statistics (GS-BSS) in diffusion microstructure." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 638-646. Springer, Cham, 2017. https://doi.org/10.1007/978-3-319-66182-7_73
2. Shunxing Bao, **Prasanna Parvathaneni**, Yuankai Huo, Yogesh Barve, Andrew J. Plassard, Yuang Yao, Hongyang Sun, Ilwoo Lyu, David H. Zald, Bennett A. Landman and Aniruddha Gokhale. "Technology Enablers for Cloud-based Multi-level Analysis Applications in Medical Image Processing." IEEE BigData 2018
3. Yuankai Huo, Zhoubing Xu, Katherine Aboud, **Prasanna Parvathaneni**, Shunxing Bao, Camilo Bermudez, Susan M. Resnick, Laurie E. Cutting, and Bennett A. Landman. "Spatially Localized Atlas Network Tiles Enables 3D Whole Brain Segmentation from Limited Data." In International Conference on Medical Image Computing and Computer-Assisted Intervention. vol 11072. pp. 698-705 Springer, Cham, 2018.

Conference Publications

1. **Prasanna Parvathaneni**, Ilwoo Lyu, Yuankai Huo, Justin Blaber, Allison E. Hainline, Hakmook Kang, Neil D. Woodward, and Bennett A. Landman. "Constructing statistically unbiased cortical surface templates using feature-space covariance." In Medical Imaging 2018: Image Processing, vol. 10574, p. 1057406. International Society for Optics and Photonics, 2018.
2. **Prasanna Parvathaneni**, Shunxing Bao, Allison Hainline , Yuankai Huo , Kurt G. Schilling , Hakmook Kang , Owen Williams , Neil D. Woodward , Susan M. Resnick , David H. Zald , Ilwoo Lyu , Bennett A. Landman "Harmonization of white and gray matter features in diffusion microarchitecture for cross sectional studies." In International Conference on Clinical and Medical Image Analysis 2018 (ICCMIA'18).
3. Sun, Peizhen, **Prasanna Parvathaneni**, Kurt G. Schilling, Yurui Gao, Vaibhav Janve, Adam Anderson, and Bennett A. Landman. "Integrating histology and MRI in the first digital brain of common squirrel monkey, Saimiri sciureus." In Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 9417, p. 94171T. International Society for Optics and Photonics, 2015.
4. Gao, Yurui, **Prasanna Parvathaneni**, Kurt G. Schilling, Feng Wang, Iwona Stepniewska, Zhoubing Xu, Ann S. Choe et al. "A 3D high resolution ex vivo white matter atlas of the common squirrel monkey (Saimiri sciureus) based on diffusion tensor imaging." In Medical Imaging 2016: Image Processing, vol. 9784, p. 97843K. International Society for Optics and Photonics, 2016.
5. Huo, Yuankai, Shunxing Bao, **Prasanna Parvathaneni**, and Bennett A. Landman. "Improved stability of whole brain surface parcellation with multi-atlas segmentation." In Medical Imaging 2018: Image Processing, vol. 10574, p. 1057438. International Society for Optics and Photonics, 2018.
6. Nath, Vishwesh, Kurt G. Schilling, Allison E. Hainline, **Prasanna Parvathaneni**, Justin A. Blaber, Ilwoo Lyu, Adam W. Anderson et al. "SHARD: spherical harmonic-based robust

outlier detection for HARDI methods." In Medical Imaging 2018: Image Processing, vol. 10574, p. 105740X. International Society for Optics and Photonics, 2018.

7. Bao, Shunxing, Yuankai Huo, **Prasanna Parvathaneni**, Andrew J. Plassard, Camilo Bermudez, Yuang Yao, Ilwoo Lyu, Aniruddha Gokhale, and Bennett A. Landman. "A data colocation grid framework for big data medical image processing: backend design." In Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, vol. 10579, p. 105790A. International Society for Optics and Photonics, 2018.
8. Hainline, Allison E., Vishwesh Nath, **Prasanna Parvathaneni**, Justin Blaber, Baxter Rogers, Allen Newton, Jeffrey Luci, Heidi Edmonson, Hakmook Kang, and Bennett A. Landman. "Evaluation of inter-site bias and variance in diffusion-weighted MRI." In Medical Imaging 2018: Image Processing, vol. 10574, p. 1057413. International Society for Optics and Photonics, 2018.
9. Vishwesh Nath, **Prasanna Parvathaneni**, Colin B. Hansen, Allison E. Hainline, Camilo Bermudez, Samuel Remedios, Justin A. Blaber, Kurt G. Schilling, Ilwoo Lyu, Vaibhav Janve, Yurui Gao, Iwona Stepniewska, Baxter P. Rogers, Allen T. Newton, L. Taylor Davis, Jeff Luci, Adam W. Anderson and Bennett A. Landman "Inter-Scanner Harmonization of High Angular Resolution DW-MRI using Null Space Deep Learning." (Accepted at Computation Diffusion MRI Workshop at MICCAI 2018).

Conference Abstracts

1. **Prasanna Parvathaneni**, Baxter P. Rogers, Yuankai Huo, Kurt G. Schilling, Allison E. Hainline, Adam W. Anderson, Neil D. Woodward, Bennett A. Landman. "Gray Matter Surface based Spatial Statistics in Neuroimaging Studies". Frontiers in Biomedical Imaging Science VI. May 2017.
2. Neil D. Woodward, **Prasanna Parvathaneni**, Baxter Rogers, Stephen Damon, Bennett Landman. "Neurite orientation dispersion and density imaging (NODDI) of the prefrontal cortex in psychosis". Society of Biological Psychiatry (SoBP) 2017.
3. Neil D. Woodward, **Prasanna Parvathaneni**, Baxter Rogers, Stephen Damon, Bennett Landman "Dendritic organization within the PFC measured in vivo in psychosis using neurite orientation dispersion and density imaging (NODDI)". International Congress on Schizophrenia Research (IOCR) 2017.
4. Sheffield, Julia, **Prasanna Parvathaneni**, Baxter Rogers, Bennett Landman, and Neil Woodward. "T222. Functional Brain Activation and Grey Matter Integrity in Psychosis: A Combined Functional Magnetic Resonance and Neurite Orientation Distribution and Density Imaging Study." Biological Psychiatry 83, no. 9 (2018): S214-S215.
5. Kurt Schilling, Vishwesh Nath, Justin Blaber, **Prasanna Parvathaneni**, Adam Anderson, Bennett Landman. *"Empirical reproducibility of clinically feasible ensemble average propagator imaging".* ISMRM April 2017.
6. **Prasanna Parvathaneni**, Joanna Phillips, Juan R. Cabrera, Anny Shai, Soonmee Cha, Susan Chang, Sarah Nelson, and Janine Lupo "Application of an advanced diffusion-weighted MRI technique to characterize glioma microstructure and relationship to histopathology" Society of Neuro-Oncology (2016): vi134-vi134.
7. **Prasanna Parvathaneni**, Qiuting Wen, Joanna J Phillips, Tracy Luks, Soonmee Cha, Susan M Chang, Sarah J Nelson, Janine M Lupo "Comparison of tumor microstructure derived NODDI and DTI metrics to histopathology in different grades of brain tumor." ISMRM 2016.

REFERENCES

1. Johansen-Berg, H. and T.E. Behrens, *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. 2013: Academic Press.
2. Tuch, D.S., et al., *High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity.* Magnetic resonance in medicine, 2002. **48**(4): p. 577-582.
3. Zhang, H., et al., *NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain.* Neuroimage, 2012. **61**(4): p. 1000-1016.
4. Biswal, B.B., et al., *Toward discovery science of human brain function.* Proceedings of the National Academy of Sciences, 2010. **107**(10): p. 4734-4739.
5. Jahanshad, N., et al., *Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA–DTI working group.* Neuroimage, 2013. **81**: p. 455-469.
6. Van Essen, D.C., et al., *The WU-Minn human connectome project: an overview.* Neuroimage, 2013. **80**: p. 62-79.
7. Casey, B.J., et al., *The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites.* Dev Cogn Neurosci, 2018. **32**: p. 43-54.
8. Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.* PLoS Med, 2015. **12**(3): p. e1001779.
9. Van Essen, D.C., et al., *The Human Connectome Project: a data acquisition perspective.* Neuroimage, 2012. **62**(4): p. 2222-31.
10. O'connor, J.P., et al., *Imaging biomarker roadmap for cancer studies.* Nature reviews Clinical oncology, 2017. **14**(3): p. 169.
11. Dojat, M., D.N. Kennedy, and W. Niessen, *MaPPiNG: Management and Processing of images for Population imagiNG.* Frontiers in ICT, 2017. **4**: p. 18.
12. Pakkenberg, B. and H.J.G. Gundersen, *Neocortical neuron number in humans: effect of sex and age.* Journal of comparative neurology, 1997. **384**(2): p. 312-320.
13. Stiles, J. and T.L. Jernigan, *The basics of brain development.* Neuropsychology review, 2010. **20**(4): p. 327-348.
14. Jones, D.K., *Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI.* Imaging in Medicine, 2010. **2**(3): p. 341.
15. !!! INVALID CITATION !!! [Yakovlev, 1967 #135; Fields, 2005 #136].
16. Fields, R.D., *Myelination: an overlooked mechanism of synaptic plasticity?* Neuroscientist, 2005. **11**(6): p. 528-31.
17. Fields, R.D., *Change in the brain's white matter.* Science, 2010. **330**(6005): p. 768-769.
18. Bonilha, L., et al., *Gray matter axonal connectivity maps.* Frontiers in psychiatry, 2015. **6**: p. 35.
19. Woodward, N.D., *Thalamocortical Functional Connectivity, Cognitive Impairment, and Cognitive Remediation in Schizophrenia.* Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2017. **2**(4): p. 307-309.
20. Gennatas, E.D., et al., *Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood.* Journal of Neuroscience, 2017. **37**(20): p. 5065-5073.
21. Gold, A.L., et al., *Cortical Thickness and Subcortical Gray Matter Volume in Pediatric Anxiety Disorders.* Neuropsychopharmacology, 2017.
22. Gong, N.J., et al., *Differential microstructural and morphological abnormalities in mild cognitive impairment and Alzheimer's disease: Evidence from cortical and deep gray matter.* Human Brain Mapping, 2017. **38**(5): p. 2495-2508.
23. Korponay, C., et al., *Neurobiological correlates of impulsivity in healthy adults: lower prefrontal*

gray matter volume and spontaneous eye-blink rate but greater resting-state functional connectivity in basal ganglia-thalamo-cortical circuitry. NeuroImage, 2017.
24. Tóth, E., et al., *Gray Matter Atrophy Is Primarily Related to Demyelination of Lesions in Multiple Sclerosis: A Diffusion Tensor Imaging MRI Study.* Frontiers in neuroanatomy, 2017. **11**.
25. Calhoun, V.D., et al., *Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data.* Hum Brain Mapp, 2006. **27**(1): p. 47-62.
26. Asman, A.J. and B.A. Landman, *Non-local statistical label fusion for multi-atlas segmentation.* Med Image Anal, 2013. **17**(2): p. 194-208.
27. Klein, A., et al. *Open labels: online feedback for a public resource of manually labeled brain images*. in *16th Annual Meeting for the Organization of Human Brain Mapping*. 2010.
28. Kandel, E.R., et al., *Principles of neural science*. Vol. 4. 2000: McGraw-hill New York.
29. Lyu, I., et al., *Group-wise cortical correspondence via sulcal curve-constrained entropy minimization.* Inf Process Med Imaging, 2013. **23**: p. 364-75.
30. Shi, Y., et al., *Joint sulcal detection on cortical surfaces with graphical models and boosted priors.* IEEE Trans Med Imaging, 2009. **28**(3): p. 361-73.
31. Yun, H.J., et al., *Automated sulcal depth measurement on cortical surface reflecting geometrical properties of sulci.* PLoS One, 2013. **8**(2): p. e55977.
32. Le Bihan, D., et al., *Diffusion tensor imaging: concepts and applications.* J Magn Reson Imaging, 2001. **13**(4): p. 534-46.
33. Le Bihan, D., *Diffusion MRI: what water tells us about the brain.* EMBO molecular medicine, 2014: p. e201404055.
34. Charles-Edwards, E.M. and N.M. deSouza, *Diffusion-weighted magnetic resonance imaging and its application to cancer.* Cancer Imaging, 2006. **6**: p. 135-43.
35. Matsuoka, Y., et al., *Combination of diffusion-weighted magnetic resonance imaging and extended prostate biopsy predicts lobes without significant cancer: application in patient selection for hemiablative focal therapy.* Eur Urol, 2014. **65**(1): p. 186-92.
36. Lebel, C. and C. Beaulieu, *Longitudinal development of human brain wiring continues from childhood into adulthood.* Journal of Neuroscience, 2011. **31**(30): p. 10937-10947.
37. Basser, P.J., J. Mattiello, and D. LeBihan, *MR diffusion tensor spectroscopy and imaging.* Biophysical journal, 1994. **66**(1): p. 259.
38. Fieremans, E., J.H. Jensen, and J.A. Helpern, *White matter characterization with diffusional kurtosis imaging.* Neuroimage, 2011. **58**(1): p. 177-88.
39. Jelescu, I.O., et al., *One diffusion acquisition and different white matter models: how does microstructure change in human early development based on WMTI and NODDI?* Neuroimage, 2015. **107**: p. 242-56.
40. Jespersen, S.N., et al., *Neurite density from magnetic resonance diffusion measurements at ultrahigh field: comparison with light microscopy and electron microscopy.* Neuroimage, 2010. **49**(1): p. 205-16.
41. Jespersen, S.N., et al., *Modeling dendrite density from magnetic resonance diffusion measurements.* Neuroimage, 2007. **34**(4): p. 1473-86.
42. Jespersen, S.N., et al., *Determination of axonal and dendritic orientation distributions within the developing cerebral cortex by diffusion tensor imaging.* IEEE Trans Med Imaging, 2012. **31**(1): p. 16-32.
43. Panagiotaki, E., et al., *Two-compartment models of the diffusion MR signal in brain white matter.* Med Image Comput Comput Assist Interv, 2009. **12**(Pt 1): p. 329-36.
44. Panagiotaki, E., et al., *Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison.* Neuroimage, 2012. **59**(3): p. 2241-54.
45. Zhang, H., et al., *NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain.* Neuroimage, 2012. **61**(4): p. 1000-16.
46. Alexander, D.C., et al., *Orientationally invariant indices of axon diameter and density from*

diffusion MRI. Neuroimage, 2010. **52**(4): p. 1374-89.
47. Barazany, D., P.J. Basser, and Y. Assaf, *In vivo measurement of axon diameter distribution in the corpus callosum of rat brain.* Brain, 2009. **132**(Pt 5): p. 1210-20.
48. Uğurbil, K., L. Toth, and D.-S. Kim, *How accurate is magnetic resonance imaging of brain function?* Trends in neurosciences, 2003. **26**(2): p. 108-114.
49. Huo, Y., et al., *Consistent cortical reconstruction and multi-atlas brain segmentation.* Neuroimage, 2016. **138**: p. 197-210.
50. Wang, H., et al., *Multi-atlas segmentation with joint label fusion.* IEEE transactions on pattern analysis and machine intelligence, 2013. **35**(3): p. 611-623.
51. Avants, B.B., N. Tustison, and G. Song, *Advanced normalization tools (ANTS).* Insight j, 2009. **2**: p. 1-35.
52. Fischl, B., M.I. Sereno, and A.M. Dale, *Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system.* Neuroimage, 1999. **9**(2): p. 195-207.
53. Lyu, I., et al., *Spectral-based automatic labeling and refining of human cortical sulcal curves using expert-provided examples.* Neuroimage, 2010. **52**(1): p. 142-157.
54. Thompson, P.M., et al., *Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia.* Neuroimage, 2004. **23 Suppl 1**: p. S2-18.
55. Mangin, J.F., et al., *A framework to study the cortical folding patterns.* Neuroimage, 2004. **23 Suppl 1**: p. S129-38.
56. Luders, E., et al., *A curvature-based approach to estimate local gyrification on the cortical surface.* Neuroimage, 2006. **29**(4): p. 1224-30.
57. Kim, J.S., et al., *Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification.* Neuroimage, 2005. **27**(1): p. 210-221.
58. Brett, M., et al. *Region of interest analysis using an SPM toolbox*. in *8th international conference on functional mapping of the human brain*. 2002. Sendai.
59. Nieto-Castanon, A., et al., *Region of interest based analysis of functional imaging data.* Neuroimage, 2003. **19**(4): p. 1303-16.
60. Qiu, A., et al., *Region-of-interest-based analysis with application of cortical thickness variation of left planum temporale in schizophrenia and psychotic bipolar disorder.* Hum Brain Mapp, 2008. **29**(8): p. 973-85.
61. Pagani, E., et al., *A method for obtaining tract-specific diffusion tensor MRI measurements in the presence of disease: application to patients with clinically isolated syndromes suggestive of multiple sclerosis.* Neuroimage, 2005. **26**(1): p. 258-65.
62. Levitt, J.G., et al., *Cortical sulcal maps in autism.* Cereb Cortex, 2003. **13**(7): p. 728-35.
63. Ashburner, J. and K.J. Friston, *Voxel-based morphometry—the methods.* Neuroimage, 2000. **11**(6): p. 805-821.
64. Honea, R., et al., *Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies.* Am J Psychiatry, 2005. **162**(12): p. 2233-45.
65. Jo, H.J., et al., *Spatial accuracy of fMRI activation influenced by volume- and surface-based spatial smoothing techniques.* Neuroimage, 2007. **34**(2): p. 550-64.
66. Bookstein, F.L., *"Voxel-based morphometry" should not be used with imperfectly registered images.* Neuroimage, 2001. **14**(6): p. 1454-62.
67. Crum, W.R., et al., *Zen and the art of medical image registration: correspondence, homology, and quality.* Neuroimage, 2003. **20**(3): p. 1425-37.
68. Anderson, M.J., *A new method for non‐parametric multivariate analysis of variance.* Austral ecology, 2001. **26**(1): p. 32-46.
69. Zimmerman, D.W., *Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances.* The Journal of Experimental Education, 1987. **55**(3): p. 171-174.
70. Forrester, J.C. and H.K. Ury, *The Signed-Rank (Wilcoxon) test in the rapid analysis of biological data.* Lancet, 1969. **1**(7588): p. 239-41.

71. Rosner, B., R.J. Glynn, and M.L. Lee, *The Wilcoxon signed rank test for paired comparisons of clustered data.* Biometrics, 2006. **62**(1): p. 185-92.
72. Yang, X., et al., *Biological parametric mapping with robust and non-parametric statistics.* Neuroimage, 2011. **57**(2): p. 423-30.
73. Winkler, A.M., et al., *Permutation inference for the general linear model.* Neuroimage, 2014. **92**: p. 381-97.
74. Smith, S.M. and T.E. Nichols, *Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference.* Neuroimage, 2009. **44**(1): p. 83-98.
75. Ball, G., et al., *Development of cortical microstructure in the preterm human brain.* Proc Natl Acad Sci U S A, 2013. **110**(23): p. 9541-6.
76. Smith, S.M., et al., *Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data.* Neuroimage, 2006. **31**(4): p. 1487-505.
77. Oosterhof, N.N., et al., *A comparison of volume-based and surface-based multi-voxel pattern analysis.* Neuroimage, 2011. **56**(2): p. 593-600.
78. Dale, A.M., B. Fischl, and M.I. Sereno, *Cortical surface-based analysis. I. Segmentation and surface reconstruction.* Neuroimage, 1999. **9**(2): p. 179-94.
79. Wandell, B.A., S. Chial, and B.T. Backus, *Visualization and measurement of the cortical surface.* J Cogn Neurosci, 2000. **12**(5): p. 739-52.
80. Van Essen, D.C., et al., *An integrated software suite for surface-based analyses of cerebral cortex.* J Am Med Inform Assoc, 2001. **8**(5): p. 443-59.
81. Lizarraga, G., et al., *A Neuroimaging Web Services Interface as a Cyber Physical System for Medical Imaging and Data Management in Brain Research: Design Study.* JMIR medical informatics, 2018. **6**(2).
82. Poldrack, R.A. and J.A. Mumford, *Independence in ROI analysis: where is the voodoo?* Soc Cogn Affect Neurosci, 2009. **4**(2): p. 208-13.
83. Nelder, J.A. and R.W. Wedderburn, *Generalized linear models.* Journal of the Royal Statistical Society: Series A (General), 1972. **135**(3): p. 370-384.
84. Brett, M., W. Penny, and S. Kiebel, *Introduction to random field theory.* Human brain function, 2003. **2**.
85. Weisstein, E.W., *Bonferroni correction.* 2004.
86. Prékopa, A., *Boole-Bonferroni inequalities and linear programming.* Operations Research, 1988. **36**(1): p. 145-162.
87. Jones, D.K., T.R. Knösche, and R. Turner, *White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI.* Neuroimage, 2013. **73**: p. 239-254.
88. Thompson, P.M., et al., *ENIGMA and the individual: predicting factors that affect the brain in 35 countries worldwide.* Neuroimage, 2017. **145**: p. 389-408.
89. Fortin, J.-P., et al., *Harmonization of multi-site diffusion tensor imaging data.* Neuroimage, 2017. **161**: p. 149-170.
90. Fortin, J., et al., *Harmonization of cortical thickness measurements across scanners and sites.* NeuroImage, 2017. **167**: p. 104-120.
91. Orlhac, F., et al., *A post-reconstruction harmonization method for multicenter radiomic studies in PET.* Journal of Nuclear Medicine, 2018: p. jnumed. 117.199935.
92. Dalal, P., et al., *Multiple cortical surface correspondence using pairwise shape similarity.* Med Image Comput Comput Assist Interv, 2010. **13**(Pt 1): p. 349-56.
93. Lyu, I., et al., *Robust estimation of group-wise cortical correspondence with an application to macaque and human neuroimaging studies.* Front Neurosci, 2015. **9**: p. 210.
94. Fischl, B., et al., *High-resolution intersubject averaging and a coordinate system for the cortical surface.* Hum Brain Mapp, 1999. **8**(4): p. 272-84.
95. Fortin, J.-P., et al., *Harmonization of cortical thickness measurements across scanners and sites.* NeuroImage, 2018. **167**: p. 104-120.

96. Garey, L.J., et al., *Reduced dendritic spine density on cerebral cortical pyramidal neurons in schizophrenia.* J Neurol Neurosurg Psychiatry, 1998. **65**(4): p. 446-53.
97. Konopaske, G.T., et al., *Prefrontal cortical dendritic spine pathology in schizophrenia and bipolar disorder.* JAMA Psychiatry, 2014. **71**(12): p. 1323-31.
98. Rae, C., et al., *Deficits in neurite density underlie white matter structure abnormalities in first-episode psychosis.* Biological Psychiatry, 2017.
99. Nazeri, A., et al., *Gray Matter Neuritic Microstructure Deficits in Schizophrenia and Bipolar Disorder.* Biol Psychiatry, 2016.
100. Nazeri, A., et al., *Functional consequences of neurite orientation dispersion and density in humans across the adult lifespan.* J Neurosci, 2015. **35**(4): p. 1753-62.
101. Cox, S.R., et al., *Ageing and brain white matter structure in 3,513 UK Biobank participants.* Nature communications, 2016. **7**: p. 13629.
102. Gogtay, N., et al., *Dynamic mapping of human cortical development during childhood through early adulthood.* Proceedings of the National academy of Sciences of the United States of America, 2004. **101**(21): p. 8174-8179.
103. Kochunov, P., et al., *Age‐related morphology trends of cortical sulci.* Human brain mapping, 2005. **26**(3): p. 210-220.
104. Northoff, G., et al., *Cortical sulcal enlargement in catatonic schizophrenia: a planimetric CT study.* Psychiatry Research: Neuroimaging, 1999. **91**(1): p. 45-54.
105. Parvathaneni, P., et al., *Empirical reproducibility, sensitivity, and optimization of acquisition protocol, for Neurite Orientation Dispersion and Density Imaging using AMICO.* Magn Reson Imaging, 2018. **50**: p. 96-109.
106. Parvathaneni, P., et al. *Gray Matter Surface based Spatial Statistics (GS-BSS) in Diffusion Microstructure*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2017. Springer.
107. Woodward, N.D., et al. *Neurite orientation dispersion and density imaging (NODDI) of the prefrontal cortex in psychosis*. in *Biological Psychiatry*. 2017. ELSEVIER SCIENCE INC 360 PARK AVE SOUTH, NEW YORK, NY 10010-1710 USA.
108. Parvathaneni, P., et al., *Harmonization of White and Gray Matter Features in Diffusion Microarchitecture for Cross-Sectional Studies*, in *Computer Aided Intervention and Diagnostics in Clinical and Medical Images*. 2019, Springer. p. 21-29.
109. Parvathaneni, P., et al., *Constructing Statistically Unbiased Cortical Surface Templates Using Feature-Space Covariance.* Proc SPIE Int Soc Opt Eng, 2018. **10574**.
110. Daducci, A., et al., *Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data.* NeuroImage, 2015. **105**: p. 32-44.
111. Nazeri, A., et al., *Functional consequences of neurite orientation dispersion and density in humans across the adult lifespan.* The Journal of Neuroscience, 2015. **35**(4): p. 1753-1762.
112. Chang, Y.S., et al., *White matter changes of neurite density and fiber orientation dispersion during human brain maturation.* PloS one, 2015. **10**(6): p. e0123656.
113. Billiet, T., et al., *Age-related microstructural differences quantified using myelin water imaging and advanced diffusion MRI.* Neurobiology of aging, 2015. **36**(6): p. 2107-2121.
114. Winston, G.P., et al., *Advanced diffusion imaging sequences could aid assessing patients with focal cortical dysplasia and epilepsy.* Epilepsy research, 2014. **108**(2): p. 336-339.
115. Lemkaddem, A., et al., *Connectivity and tissue microstructural alterations in right and left temporal lobe epilepsy revealed by diffusion spectrum imaging.* NeuroImage: Clinical, 2014. **5**: p. 349-358.
116. Flashman, L.A., et al., *Hybrid Diffusion Imaging to Detect Acute White Matter Injury after Mild TBI.* 2016.
117. Adluru, G., et al., *Assessment of white matter microstructure in stroke patients using NODDI.* Conf Proc IEEE Eng Med Biol Soc, 2014. **2014**: p. 742-5.
118. Wen, Q., et al., *Clinically feasible NODDI characterization of glioma using multiband EPI at 7 T.* Neuroimage Clin, 2015. **9**: p. 291-9.

119. Colgan, N., et al., *Application of neurite orientation dispersion and density imaging (NODDI) to a tau pathology model of Alzheimer's disease.* Neuroimage, 2016. **125**: p. 739-744.
120. Kamagata, K., T. Hatano, and S. Aoki, *What is NODDI and what is its role in Parkinson's assessment?* Expert Rev Neurother, 2016. **16**(3): p. 241-3.
121. Deligianni, F., et al., *NODDI and Tensor-Based Microstructural Indices as Predictors of Functional Connectivity.* PLoS One, 2016. **11**(4): p. e0153404.
122. Schneider, T., et al., *Application of multi-shell NODDI in multiple sclerosis.* 2014.
123. Mallik, S., et al., *Imaging outcomes for trials of remyelination in multiple sclerosis.* Journal of Neurology, Neurosurgery & Psychiatry, 2014: p. jnnp-2014-307650.
124. Kodiweera, C., et al., *Age effects and sex differences in human brain white matter of young to middle-aged adults: A DTI, NODDI, and q-space study.* Neuroimage, 2016. **128**: p. 180-192.
125. Alexander, D.C., *A general framework for experiment design in diffusion MRI and its application in measuring direct tissue‐microstructure features.* Magnetic Resonance in Medicine, 2008. **60**(2): p. 439-448.
126. Tariq, M., et al. *Scan-rescan reproducibility of neurite microstructure estimates using NODDI*. in *Medical Image Understanding and Analysis*. 2012.
127. Chung, A.W., K.K. Seunarine, and C.A. Clark, *NODDI reproducibility and variability with magnetic field strength: A comparison between 1.5 T and 3 T.* Human Brain Mapping, 2016.
128. Chung, A.W., *Investigating the effects of high b-value and field strength on diffusion Magnetic Resonance Imaging*. 2014, UCL (University College London).
129. Landman, B.A., et al., *Multi-parametric neuroimaging reproducibility: a 3-T resource study.* Neuroimage, 2011. **54**(4): p. 2854-2866.
130. Vollmar, C., et al., *Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0 T scanners.* Neuroimage, 2010. **51**(4): p. 1384-1394.
131. Landman, B.A., et al., *Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T.* Neuroimage, 2007. **36**(4): p. 1123-1138.
132. Hui, E.S., et al., *B-value dependence of DTI quantitation and sensitivity in detecting neural tissue changes.* Neuroimage, 2010. **49**(3): p. 2366-2374.
133. Andersson, J.L., S. Skare, and J. Ashburner, *How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging.* Neuroimage, 2003. **20**(2): p. 870-888.
134. Jenkinson, M., et al., *Improved optimization for the robust and accurate linear registration and motion correction of brain images.* Neuroimage, 2002. **17**(2): p. 825-41.
135. Holmes, C.J., et al., *Enhancement of MR images using registration for signal averaging.* Journal of computer assisted tomography, 1998. **22**(2): p. 324-333.
136. Tustison, N.J., et al., *N4ITK: improved N3 bias correction.* IEEE Trans Med Imaging, 2010. **29**(6): p. 1310-20.
137. Avants, B.B., et al., *Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain.* Med Image Anal, 2008. **12**(1): p. 26-41.
138. Asman, A.J. and B.A. Landman, *Formulating spatially varying performance in the statistical fusion framework.* IEEE Trans Med Imaging, 2012. **31**(6): p. 1326-36.
139. Freund, Y. and R.E. Schapire. *A desicion-theoretic generalization of on-line learning and an application to boosting*. in *European conference on computational learning theory*. 1995. Springer.
140. Yushkevich, P.A., et al., *User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability.* Neuroimage, 2006. **31**(3): p. 1116-28.
141. Zhang, H., et al., *Axon diameter mapping in the presence of orientation dispersion with diffusion MRI.* Neuroimage, 2011. **56**(3): p. 1301-1315.
142. Cook, P.A., et al., *Optimal acquisition orders of diffusion-weighted MRI measurements.* J Magn Reson Imaging, 2007. **25**(5): p. 1051-8.
143. Cook, P., et al. *Camino: open-source diffusion-MRI reconstruction and processing*. in *14th scientific meeting of the international society for magnetic resonance in medicine*. 2006. Seattle

WA, USA.
144. Reeder, S.B., et al., *Practical approaches to the evaluation of signal-to-noise ratio performance with parallel imaging: application with cardiac imaging and a 32-channel cardiac coil.* Magn Reson Med, 2005. **54**(3): p. 748-54.
145. Timmers, I., et al., *Assessing Microstructural Substrates of White Matter Abnormalities: A Comparative Study Using DTI and NODDI.* PloS one, 2016. **11**(12): p. e0167884.
146. Bach, M., et al., *Methodological considerations on tract-based spatial statistics (TBSS).* Neuroimage, 2014. **100**: p. 358-69.
147. Ashburner, J., *A fast diffeomorphic image registration algorithm.* Neuroimage, 2007. **38**(1): p. 95-113.
148. Lombaert, H., J. Sporring, and K. Siddiqi, *Diffeomorphic spectral matching of cortical surfaces.* Inf Process Med Imaging, 2013. **23**: p. 376-89.
149. Barber, C.B., D.P. Dobkin, and H. Huhdanpaa, *The quickhull algorithm for convex hulls.* ACM Transactions on Mathematical Software (TOMS), 1996. **22**(4): p. 469-483.
150. Peterson, C.T., et al., *Effects of Shambhavi Mahamudra Kriya, a Multicomponent Breath-Based Yogic Practice (Pranayama), on Perceived Stress and General Well-Being.* Journal of evidence-based complementary & alternative medicine, 2017: p. 2156587217730934.
151. Shen, D., et al., *Voxel-Wise Meta-Analysis of Gray Matter Changes in Amyotrophic Lateral Sclerosis.* Front Aging Neurosci, 2016. **8**: p. 64.
152. Nazeri, A., et al., *Gray Matter Neuritic Microstructure Deficits in Schizophrenia and Bipolar Disorder.* Biol Psychiatry, 2017. **82**(10): p. 726-736.
153. Ekman, C.J., et al., *A History of Psychosis in Bipolar Disorder is Associated With Gray Matter Volume Reduction.* Schizophrenia bulletin, 2016. **43**(1): p. 99-107.
154. Carmona, S., et al., *Global and regional gray matter reductions in ADHD: a voxel-based morphometric study.* Neurosci Lett, 2005. **389**(2): p. 88-93.
155. Thompson, P.M., et al., *Dynamics of gray matter loss in Alzheimer's disease.* J Neurosci, 2003. **23**(3): p. 994-1005.
156. Von Economo, C., *The cytoarchitectonics of the human cerebral cortex*. 1929: H. Milford Oxford University Press.
157. Von Economo, C. and G. Koskinas, *The cytoarchitectonics of the adult human cortex.* Vienna and Berlin: Julius Springer Verlag, 1925.
158. Glasser, M.F., et al., *The Human Connectome Project's neuroimaging approach.* Nat Neurosci, 2016. **19**(9): p. 1175-87.
159. Jones, D.K. and M. Cercignani, *Twenty-five pitfalls in the analysis of diffusion MRI data.* NMR Biomed, 2010. **23**(7): p. 803-20.
160. Smith, S.M., et al., *Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data.* Neuroimage, 2006. **31**(4): p. 1487-1505.
161. Bodini, B., et al., *Exploring the relationship between white matter and gray matter damage in early primary progressive multiple sclerosis: an in vivo study with TBSS and VBM.* Human brain mapping, 2009. **30**(9): p. 2852-2861.
162. Westlye, L.T., et al., *Life-span changes of the human brain white matter: diffusion tensor imaging (DTI) and volumetry.* Cerebral cortex, 2009. **20**(9): p. 2055-2068.
163. Olson, E.A., et al., *Disruption of white matter structural integrity and connectivity in posttraumatic stress disorder: A TBSS and tractography study.* Depression and anxiety, 2017. **34**(5): p. 437-445.
164. Bells, S., et al., *Changes in white matter microstructure impact cognition by disrupting the ability of neural assemblies to synchronize.* Journal of Neuroscience, 2017. **37**(34): p. 8227-8238.
165. Goodrich‐Hunsaker, N.J., et al., *Age‐and sex‐related effects in children with mild traumatic brain injury on diffusion magnetic resonance imaging properties: A comparison of voxelwise and tractography methods.* Journal of neuroscience research, 2018. **96**(4): p. 626-641.
166. Yeo, B.T., et al., *Spherical demons: fast diffeomorphic landmark-free surface registration.* IEEE transactions on medical imaging, 2010. **29**(3): p. 650-668.

167. Van Essen, D.C., *A population-average, landmark-and surface-based (PALS) atlas of human cerebral cortex.* Neuroimage, 2005. **28**(3): p. 635-662.
168. Fischl, B., et al., *High-resolution intersubject averaging and a coordinate system for the cortical surface.* Human brain mapping, 1999. **8**(4): p. 272-284.
169. Van Essen, D. and H. Drury, *Structural and functional analyses of human cerebral cortex using a surface-based atlas.* Journal of Neuroscience, 1997. **17**(18): p. 7079-7102.
170. Ashburner, J. and K.J. Friston, *Voxel-based morphometry--the methods.* Neuroimage, 2000. **11**(6 Pt 1): p. 805-21.
171. Glasser, M.F., et al., *The minimal preprocessing pipelines for the Human Connectome Project.* Neuroimage, 2013. **80**: p. 105-24.
172. Huo, Y., et al. *Combining multi-atlas segmentation with brain surface estimation*. in *Proceedings of SPIE--the International Society for Optical Engineering*. 2016. NIH Public Access.
173. Winkler, A.M., et al., *Permutation inference for the general linear model.* Neuroimage, 2014. **92**: p. 381-397.
174. Winkler, A.M., et al., *Faster permutation inference in brain imaging.* NeuroImage, 2016. **141**: p. 502-516.
175. Goebel, R., F. Esposito, and E. Formisano, *Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis.* Hum Brain Mapp, 2006. **27**(5): p. 392-401.
176. Reuter, M., et al., *Within-subject template estimation for unbiased longitudinal image analysis.* Neuroimage, 2012. **61**(4): p. 1402-18.
177. Van Essen, D.C., *Surface-based approaches to spatial localization and registration in primate cerebral cortex.* Neuroimage, 2004. **23 Suppl 1**: p. S97-107.
178. Saad, Z.S. and R.C. Reynolds, *Suma.* Neuroimage, 2012. **62**(2): p. 768-73.
179. Landman, B.A., et al., *Multi-parametric neuroimaging reproducibility: a 3-T resource study.* Neuroimage, 2011. **54**(4): p. 2854-66.
180. Klein, A., et al., *Open labels: online feedback for a public resource of manually labeled brain images. 16th Annu.* Meet. Organ. Hum. Brain Mapp, 2010.
181. Huo, Y., et al., *Combining Multi-atlas Segmentation with Brain Surface Estimation.* Proc SPIE Int Soc Opt Eng, 2016. **9784**.
182. Sowell, E.R., et al., *Mapping cortical change across the human life span.* Nat Neurosci, 2003. **6**(3): p. 309-15.
183. Fillard, P., et al., *Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines.* Neuroimage, 2007. **34**(2): p. 639-650.
184. Amunts, K., et al., *Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps.* Anatomy and embryology, 2005. **210**(5-6): p. 343-352.
185. Rademacher, J., et al., *Human cerebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging.* J Cogn Neurosci, 1992. **4**(4): p. 352-74.
186. Caviness, V.S., Jr., et al., *MRI-Based Topographic Parcellation of Human Neocortex: An Anatomically Specified Method with Estimate of Reliability.* J Cogn Neurosci, 1996. **8**(6): p. 566-87.
187. Lyu, I., et al., *TRACE: A Topological Graph Representation for Automatic Sulcal Curve Extraction.* IEEE Transactions on Medical Imaging, 2017.
188. Ono, M., S. Kubik, and C.D. Abernathey, *Atlas of the cerebral sulci*. 1990: Tps.
189. Thompson, P.M., et al., *Three-dimensional statistical analysis of sulcal variability in the human brain.* J Neurosci, 1996. **16**(13): p. 4261-74.
190. Lohmann, G., D.Y. von Cramon, and H. Steinmetz, *Sulcal variability of twins.* Cerebral Cortex, 1999. **9**(7): p. 754-763.
191. Joshi, A.A., et al. *Geodesic curvature flow on surfaces for automatic sulcal delineation*. in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. 2012. IEEE.

192. Tao, X., J.L. Prince, and C. Davatzikos, *Using a statistical shape model to extract sulcal curves on the outer cortex of the human brain.* IEEE Transactions on Medical Imaging, 2002. **21**(5): p. 513-524.
193. Rettmann, M.E., et al., *Automated sulcal segmentation using watersheds on the cortical surface.* NeuroImage, 2002. **15**(2): p. 329-344.
194. Shattuck, D.W., et al., *Semi-automated method for delineation of landmarks on models of the cerebral cortex.* J Neurosci Methods, 2009. **178**(2): p. 385-92.
195. Tu, Z., et al., *Automated extraction of the cortical sulci based on a supervised learning approach.* IEEE Transactions on Medical Imaging, 2007. **26**(4): p. 541-552.
196. Joshi, A.A., D.W. Shattuck, and R.M. Leahy, *A Method for Automated Cortical Surface Registration and Labeling.* Biomed Image Regist Proc, 2012. **7359**: p. 180-189.
197. Mangin, J.-F., et al., *Sulcus identification and labeling.* Brain mapping: an encyclopedic reference. Academic Press: Elsevier, New York, 2015: p. 365-371.
198. Shattuck, D.W. and R.M. Leahy, *BrainSuite: an automated cortical surface identification tool.* Med Image Anal, 2002. **6**(2): p. 129-42.
199. Pantazis, D., et al., *Comparison of landmark-based and automatic methods for cortical surface registration.* Neuroimage, 2010. **49**(3): p. 2479-93.
200. Greenspan, H., B. Van Ginneken, and R.M. Summers, *Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique.* IEEE Transactions on Medical Imaging, 2016. **35**(5): p. 1153-1159.
201. Litjens, G., et al., *A survey on deep learning in medical image analysis.* Med Image Anal, 2017. **42**: p. 60-88.
202. de Brebisson, A. and G. Montana. *Deep neural networks for anatomical brain segmentation*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015.
203. Havaei, M., et al., *Brain tumor segmentation with Deep Neural Networks.* Med Image Anal, 2017. **35**: p. 18-31.
204. Mechelli, A., *202. Deep Learning Technology: Concepts and Applications in Biological Psychiatry.* Biological Psychiatry, 2018. **83**(9): p. S81-S82.
205. Minaee, S., et al., *A Deep Unsupervised Learning Approach Toward MTBI Identification Using Diffusion MRI.* arXiv preprint arXiv:1802.02925, 2018.
206. Hao, Y., et al., *DeepIED: An epileptic discharge detector for EEG-fMRI based on deep learning.* NeuroImage: Clinical, 2018. **17**: p. 962-975.
207. Milletari, F., et al., *Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound.* Computer Vision and Image Understanding, 2017. **164**: p. 92-102.
208. Nath, V., et al., *Inter-Scanner Harmonization of High Angular Resolution DW-MRI using Null Space Deep Learning.* arXiv preprint arXiv:1810.04260, 2018.
209. Yamins, D.L. and J.J. DiCarlo, *Using goal-driven deep learning models to understand sensory cortex.* Nat Neurosci, 2016. **19**(3): p. 356-65.
210. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
211. Bronstein, M.M., et al., *Geometric deep learning: going beyond euclidean data.* IEEE Signal Processing Magazine, 2017. **34**(4): p. 18-42.
212. Boscaini, D., et al. *Learning shape correspondence with anisotropic convolutional neural networks*. in *Advances in Neural Information Processing Systems*. 2016.
213. Seong, S.-B., C. Pae, and H.-J. Park, *Geometric convolutional neural network for analyzing surface-based neuroimaging data.* Frontiers in Neuroinformatics, 2018. **12**: p. 42.
214. Cucurull, G., et al., *Convolutional neural networks for mesh-based parcellation of the cerebral cortex.* 2018.
215. Shock, N.W., *Normal human aging: The Baltimore longitudinal study of aging.* 1984.

216. Lyu, I., M.A. Styner, and B.A. Landman. *Hierarchical Spherical Deformation for Shape Correspondence*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018. Springer.
217. Huo, Y., et al. *Improved stability of whole brain surface parcellation with multi-atlas segmentation*. in *Medical Imaging 2018: Image Processing*. 2018. International Society for Optics and Photonics.
218. Lebed, E., et al., *Novel surface-smoothing based local gyrification index.* IEEE Trans Med Imaging, 2013. **32**(4): p. 660-9.
219. Joshi, A.A., D.W. Shattuck, and R.M. Leahy. *A method for automated cortical surface registration and labeling*. in *International Workshop on Biomedical Image Registration*. 2012. Springer.
220. Auzias, G., et al., *Model-driven harmonic parameterization of the cortical surface: HIP-HOP.* IEEE Trans Med Imaging, 2013. **32**(5): p. 873-87.
221. Huo, Y., et al., *Spatially Localized Atlas Network Tiles Enables 3D Whole Brain Segmentation from Limited Data.* arXiv preprint arXiv:1806.00546, 2018.
222. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the royal statistical society. Series B (Methodological), 1995: p. 289-300.
223. Storey, J.D., *False discovery rate*, in *International encyclopedia of statistical science*. 2011, Springer. p. 504-508.
224. Sowell, E.R., et al., *Mapping sulcal pattern asymmetry and local cortical surface gray matter distribution in vivo: maturation in perisylvian cortices.* Cereb Cortex, 2002. **12**(1): p. 17-26.
225. Amunts, K., et al., *Interhemispheric asymmetry of the human motor cortex related to handedness and gender.* Neuropsychologia, 2000. **38**(3): p. 304-12.
226. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436-44.
227. Chen, H., et al., *VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images.* Neuroimage, 2018. **170**: p. 446-455.
228. Gros, C., et al., *Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks.* NeuroImage, 2019. **184**: p. 901-915.
229. Wang, G., et al., *Test-time augmentation with uncertainty estimation for deep learning-based medical image segmentation.* 2018.
230. Zilles, K., et al., *Quantitative analysis of sulci in the human cerebral cortex: development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture.* Hum Brain Mapp, 1997. **5**(4): p. 218-21.
231. Juch, H., et al., *Anatomical variability of the lateral frontal lobe surface: implication for intersubject variability in language neuroimaging.* Neuroimage, 2005. **24**(2): p. 504-14.
232. Esteves, C., et al. *Learning so (3) equivariant representations with spherical cnns*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
233. Coors, B., A.P. Condurache, and A. Geiger. *Spherenet: Learning spherical representations for detection and classification in omnidirectional images*. in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.
234. Cohen, T.S., et al., *Spherical CNNs.* arXiv preprint arXiv:1801.10130, 2018.
235. Edwards, M. and X. Xie, *Graph based convolutional neural network.* arXiv preprint arXiv:1609.08965, 2016.
236. Salminen, L.E., et al., *Regional age differences in gray matter diffusivity among healthy older adults.* Brain imaging and behavior, 2016. **10**(1): p. 203-211.
237. Kurth, F., et al., *Brain Gray Matter Changes Associated with Mindfulness Meditation in Older Adults: An Exploratory Pilot Study using Voxel-based Morphometry.* Neuro, 2014. **1**(1): p. 23-26.
238. Holzel, B.K., et al., *Mindfulness practice leads to increases in regional brain gray matter density.* Psychiatry Res, 2011. **191**(1): p. 36-43.
239. Fox, K.C., et al., *Is meditation associated with altered brain structure? A systematic review and meta-analysis of morphometric neuroimaging in meditation practitioners.* Neurosci Biobehav Rev, 2014. **43**: p. 48-73.

240. Salminen, L.E., et al., *Reducing Csf Partial Volume Effects to Enhance Diffusion Tensor Imaging Metrics of Brain Microstructure.* Technol Innov, 2016. **18**(1): p. 5-20.
241. Parvathaneni, P., et al., *Gray Matter Surface based Spatial Statistics (GS-BSS) in Diffusion Microstructure.* Med Image Comput Comput Assist Interv, 2017. **10433**: p. 638-646.
242. Lauzon, C.B., et al., *Simultaneous analysis and quality assurance for diffusion tensor imaging.* PLoS One, 2013. **8**(4): p. e61737.
243. Klein, A., et al., *Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.* Neuroimage, 2009. **46**(3): p. 786-802.
244. Klein, A., et al., *Evaluation of volume-based and surface-based brain image registration methods.* Neuroimage, 2010. **51**(1): p. 214-20.
245. Fukutomi, H., et al., *Neurite imaging reveals microstructural variations in human cerebral cortical gray matter.* Neuroimage, 2018. **182**: p. 488-499.