Multi-Fascicle Models of the Brain Microstructure for Population Studies: Acquisition, Estimation, Registration and Statistical Analysis

Maxime Taquet

UNIVERSITE CATHOLIQUE DE LOUVAIN
Ecole Polytechnique de Louvain
ICTEAM Institute

Thesis submitted in partial fulfillment of the requirements for the degree of
Doctor in Applied Sciences

Ph.D. Jury

Benoît Macq	ICTEAM, UCL, Louvain-la-Neuve, Belgium
Simon K. Warfield	Boston Children's Hospital, Harvard Medical School, Boston, USA
Laurent Jacques	ICTEAM, UCL, Louvain-la-Neuve, Belgium
Christian Barillot	INRIA, INSERM, Rennes, France
Jean-Philippe Thiran	EPFL, Lausanne, Switzerland
Etienne Olivier	IoNS, UCL, Louvain-la-Neuve, Belgium
Michel Verleysen	ICTEAM, UCL, Louvain-la-Neuve, Belgium

December 2013

Abstract

The root causes of various neurological disorders are not identified yet. While there is a consensus that alterations in the central nervous system must be involved in these disorders, the exact nature and location of these alterations is often unknown. When those properties are discovered, a novel biomarker may be defined, enabling early and objective diagnoses of the disease.

In this context, brain imaging is often used to characterize brain tissues non-invasively and in vivo. In particular, recent developments in diffusion-weighted imaging have led to the advent of models of the brain microstructure. These models provide an unprecedented insight into the complex organization of cellular structures in the brain. They are therefore of strong interest in population studies to characterize neurological disorders.

Conducting population studies from models of the brain microstructure, however, raises numerous challenges that pertain to the required image acquisition, the estimation, registration and statistical analysis of those models. This thesis addresses these challenges and defines a comprehensive framework that harnesses multi-fascicle models in population studies of the brain microstructure. In particular, this framework includes methods to estimate multi-fascicle models from widely available clinical data, to reliably select an appropriate model of the brain microstructure, to register multi-fascicle models and to perform statistical analyses of microstructural properties. This framework is tested in population studies where it unravels alterations of the brain microstructure associated with autism.

The presented framework opens opportunities for new investigations of the brain microstructure in normal development and in disease and injury. It paves the way to the definition of microstructure-based biomarkers of neurological disorders.

Acknowledgments

First and foremost, I would like to thank Simon Warfield, advisor of my thesis who supervised most of my research. Simon has had a significant ($p<0.05$) impact on my research and my development as a researcher. His thorough insight into the field of medical imaging, his incredible knowledge of the literature as recent as one hour old or as old as me ("I think [some author I had never heard of] wrote something about [a topic I would never have guessed he would be aware of] in the [journal that may not even exist anymore] in 1986"), our long email threads to discuss papers and future research directions, our discussions at three or four packed for up to two hours in his office and our ironic methods designed to circumvent corrections for multiple comparisons[1] ("If the reviewer asks for extra comparisons, we do not need to correct for them since they were not part of our intentions"), have all contributed to make the research of my thesis enjoyable, fascinating, and thrilling. Owing to his great support and stimulating research atmosphere, what was supposed to be a three-month internship at Harvard Medical School ended up being a two-year stay and a fruitful and ongoing collaboration of over three years.

My meeting with Simon would probably never have happened without my supervisor, Benoît Macq, who dragged me into a research career and contributed to my early development as a researcher and has since followed all of my research with great interest. His endless optimism and his remarkable support have played an ever important role in my conducting of a PhD thesis. His mastering of soft skills and his world-spanning collaboration network is an invaluable asset for his "top-cerfs". I want to thank him for all of that.

My gratitude also goes to my two closest collaborators and friends at the Computational Radiology Laboratory: Benoît Scherrer, PhD, who can travel for a few days, throw a party at his place and then write a paper in a week that will have a strong impact on the field; and Jurriaan Peters, MD, a neurologist who could have been an engineer, writes code and enjoys statistics. Benoît introduced me to the amazing field of microstructure imaging and I enjoyed a lot writing every one of our papers (especially the IPMI paper on which we managed to work 24/7 thanks to time differences). Jurriaan introduced me to the fascinating medical part of the research and even medical practice,

[1] I guarantee that we never used those in any of our papers.

giving my research a stronger meaning. He also substantially contributed to my networking by introducing me to some of his outstanding friends and colleagues. Writing papers side-by-side (literally!), preparing presentations that deal with cucumbers, belts and suspenders to introduce concepts of microstructure and connectivity to the lay person and grabbing red-eyes, have all been (in their own ways) enjoyable and instructive.

I am also grateful to those who first introduced me to medical imaging and medical imaging research: Marinko Sarunic, Jonathan Orban, Guillaume Janssens, Laurent Paul and Olivier Cartiaux. Without them and the passion they have for medical imaging, I would probably be doing something very different (and possibly less enjoyable) now. I also want to acknowledge the master students and research interns that I have had the chance to supervise, who brought some fresh prospectives on my research: Augustin Cosse, Ruobin Gong, Jolene Singh, Robin Lefrant and Stéphanie Guérit.

There is a number of people who contributed to my development as a researcher through very interesting and insightful discussions: Yves-Alexandre de Montjoye, PhD candidate at MIT, a friend and colleague, who "never says no to talk about statistics or have a triple espresso", Martin Desseilles, MD and Professor at the University of Namur, a friend and colleague who is always enthusiastic to discuss novel ideas to map the brain or probe behavior, Jordi Quoidbach, PhD and Professor at Pompeu Fabra University, a friend and colleague, who conducts researches in social psychology, knows his statistics, and is always happy to discuss novel intricate experiments to understand human behavior, and Nicolas Boumal who can write equations on a white board for two hours, take snapshots with his phone, and call it fun.

I would also like to thank the members of my PhD committee: Xavier Banse, Sylvain Jaume and Laurent Jacques. Xavier Banse has been very supportive and vey enthusiastic about my research. He is the very first MD who invited me to assist medical interventions, which added a dimension to my research. Sylvain Jaume was incredibly dedicated when welcoming me in Boston, even when looking for a hostel at 2:00AM when the hostel I booked refused to give me a key to my room. Laurent Jacques has been a fantastic mentor in my research and my scientific development in general. Besides our discussions about my thesis, our debates about homeopathy, global warming, research practice, entropy, causality, free will, space, time, Erdos, statistics, and Feynman have all been very instructive.

Finally, I would like to thank all those who supported me outside of the

lab borders where what is supposed to be "life besides research in the lab" sometimes becomes "research besides the life in the lab". I am therefore very grateful to Sophie for her awesome support (she actually did help me write grants!), her dedicated optimism and her capability to make me laugh on top of my lungs in stressful times. I also want to thank my sisters, Charlotte and Lucie, and my sister's husband, Arnaud, who have the capacity to make hard time a lot easier, both my parents for their great support and interest in my research, and my friends who have the patience to wait for a deadline to pass before throwing a party: Xavier, Francois (fma10), Jon, Maëlle, Sosotte, Quentin, LP, Arnaud, Marie, and the Tem's team.

Contents

List of Figures

List of Tables

Acronyms and Symbols

Acronyms

ASD	Autism Spectrum Disorders
CHARMED	Composite Hindered and Restricted Model of Diffusion
CSF	Cerebrospinal Fluid
DIAMOND	Distribution of Anisotropic Microstructural Environments with Diffusion-Weighted Imaging
DTI	Diffusion Tensor Imaging
DWI	Diffusion-Weighted Imaging/Images
EEG	Electroencephalography
FA	Fractional Anisotropy
FBSS	Fascicle-Based Spatial Statistics
FOV	Field of View
GCC	Generalized Correlation Coefficient
HARDI	High Angular Resolution Diffusion Imaging
IDA	Isotropic Diffusion Analysis
LSME	Large-Scale Microstructural Environment
MD	Mean Diffusivity
MFM	Multi-Fascicle Model
MRI	Magnetic Resonance Imaging
NODDI	Neurite Orientation Dispersion and Density Imaging
SNR	Signal to Noise Ratio
TSC	Tuberous Sclerosis Complex
TSC+ASD	Tuberous Sclerosis Complex with a comorbid diagnosis of autism
TSC−ASD	Tuberous Sclerosis Complex without a comorbid diagnosis of autism

Symbols

a, A	Scalar values	*e.g.* S the attenuated signal
b	B-value	
λ	Eigenavlues	*e.g.* λ_1 the largest eingevalue
$\boldsymbol{a}$	Vector values	*e.g.* $\boldsymbol{g}$ the gradient vector
$\boldsymbol{x}$	Spatial location	*e.g.* $\boldsymbol{D}(\boldsymbol{x})$ the matrix value at $\boldsymbol{x}$
$\boldsymbol{A}$	Matrix values	*e.g.* $\boldsymbol{D}$ a diffusion matrix
$\mathrm{Tr}(\boldsymbol{A})$	Trace of matrix $\boldsymbol{A}$	
$\lvert\boldsymbol{A}\rvert$	Determinant of matrix $\boldsymbol{A}$	
$\boldsymbol{I}_3$	3×3 Identity matrix	
$\mathcal{M}$	Multi-fascicle models	
a	Scalar random variable	*e.g.* f_1 the random variable representing a volumetric fraction
$\mathbf{a}$	Random vector	*e.g.* $\mathbf{f}$ the random variable representing all volumetric fractions
$\mathbf{A}$	Random matrix	*e.g.* $\mathbf{D}$ the random variable representing a tensor
$\mathbb{1}_a$	Indicator function returning 1 if a is true and 0 otherwise	*e.g.* $\mathbb{1}_{\boldsymbol{f} \in \mathcal{S}}$

Notice that the spatial location variable $\boldsymbol{x}$ may be omitted for clarity.

Chapter 1

Introduction

The incidence of Autism Spectrum Disorders (ASD), a neurological disease affecting thousands of children across the globe, continues to increase at an alarming rate [112, 16], rising from 4 per 10,000 in the 60's to almost 1 in 100 children nowadays [97]. Whether the cause of this increase is a better ascertainment, changes in diagnostic criteria or increased incidence remains unclear [16]. However, with almost 1% of the child population affected, autism has become a major challenge of public health. Many children with this troubling diagnosis suffer from significant developmental delays, learning disabilities, and communication impairments. ASD places severe emotional stress upon the family as well as a staggering financial burden upon the entire social service network and health care system. With an estimated total annual cost of $35 billion to assist individuals with autism over their lifetimes [39], there is now intense pressure upon the medical research community to develop more accurate methods of diagnosing this devastating illness in its earliest stages, monitoring its progression, and assessing response-to-therapy. Early intervention is critical to achieving better outcomes for these children, especially with respect to quality of life.

In radiology, we generally rely upon biomarkers that point to a particular disease state based on what is observed on magnetic resonance imaging (MRI) —the modality most commonly used to explore the brain. In current radiologic practice, however, no such imaging biomarker for ASD exists. The diagnosis of autism is thus based primarily on subjective measures of signs and symptoms (*e.g.* general affect, social interactions, communication and learning styles). Subjective assessments may be unreliable and dissenting diagnoses are not uncommon. Further, definitive diagnoses of ASD are rarely made in children younger than two year old; and given the importance of early intervention, this poses a particular challenge for health care providers. We are thus strongly motivated to identify an objective, imaging-based biomarker that would enable the early diagnosis of ASD in the youngest children (perhaps even near-term infants), leading to significantly improved patient outcomes.

The inability of conventional MRI to find reliable biomarkers of autism spectrum disorders may stem from its focus on *regional* abnormalities in the brain. Regional abnormalities include altered cortical thickness and the the presence of focal lesions or tubers. Contradictory conclusions in the characterization of the cortical thickness in patients with autism have been found, with some studies reporting increased cortical thickness [48] and another reporting decreased cortical thickness [47] in autism. The location and load of tubers and lesions in patients with tuberous sclerosis complex, a genetic disease with a comorbidity of autism of about 50% [54], was also shown to be unrelated to cognitive impairment or autism [84, 52, 141] (Fig. 1.1). Our motivational study presented in the next

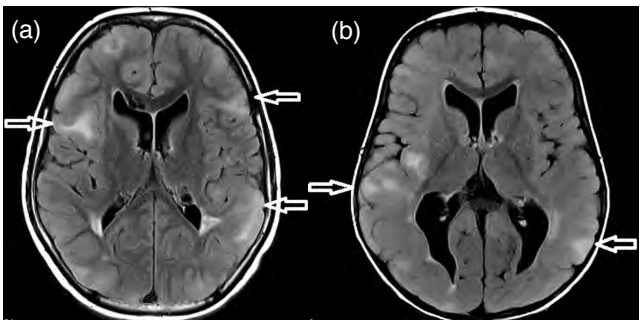


Figure 1.1: Features from conventional MRI fail to predict the patient outcome in tuberous sclerosis. (a) and (b) are both conventional MRI of patients with tuberous sclerosis complex. Both patients have subcortical tubers (arrows) of comparable size and distribution, but the patient in (a) has severe autism, no active seizure disorder and is nonverbal, while patient (b) has mild motor and language delays, no autism and refractory seizures.

section shows that autism may be associated with aberrations in *connectivity* between regions rather than abnormalities in the regions themselves. This study motivates the investigation of the *brain microstructure* that underpins connectivity. The study of the brain microstructure *in vivo* raises mathematical, computational and clinical challenges. This thesis addresses these challenges and proposes a comprehensive framework to investigate microstructural properties of the brain. This framework covers the various steps of studies of the brain microstructure, from the acquisition of images with an MRI scanner to the statistical analysis of these microstructure images.

1.1 Common Patterns of Disconnection in Autism

In the last decade, the *disconnection syndrome* model of autism has been proposed as a neurobiological basis of behavioral features presented in patients with autism [23, 42, 69, 92]. This model posits that autism is caused by the inability for patients to integrate information from distant regions in the brain. This integration is required to bind together various pieces of information into a more global concept, a cognitive ability called *central coherence* [38]. For instance, a child with autism may not understand why a face looks angry because his brain visual center does not communicate properly with his brain emotional center.

Despite the broad acceptance of the disconnection syndrome model, methodological challenges have cast doubts on its validity. On the one hand, recent researches have shown that head motion during functional magnetic resonance imaging (fMRI) makes long-range connections appear weaker and short-range connections appear stronger [87]. This artifactual pattern of connectivity is identical to that observed in fMRI scans of children with autism who typically move more in the scanner [34]. This finding therefore prompts the question of the validity of the disconnection syndrome model. On the other hand, clinical and etiological heterogeneity of autism (*i.e.* differences in the causes of autism and its outcome for the patient) complicates the detection and characterization of neurobiological mechanisms common to all autism spectrum disorders. In particular, autism may be syndromic (*i.e.* caused by a specific genetic disorder and thus identified by a syndrome) or non-syndromic. To advance the understanding of common neurobiological mechanisms in ASD, these mechanisms should be present in subjects with ASD regardless of an underlying neurogenetic abnormality. This is at odds with most researches on autism which focus on non-syndromic ASD.

In a recent study, we investigated whether common patterns of disconnection appear in the brain of children with ASD, regardless of etiology [85, 124]. We included, in a population study, patients with non-syndromic autism and children whose autism is caused by Tuberous Sclerosis Complex (TSC), a genetic disease with a high penetrance of autism (Fig. 1.2A). To avoid artifactual findings related to head motion in fMRI, we studied functional connectivity from electroencephalography (EEG) using 19 electrodes in a 10-20 system of electrode placement (Fig. 1.2B). The common post-processing pipeline of EEG includes manual removal of artifacts by experts to guarantee motion-free recordings.

The population was composed of 43 children with TSC (14 with ASD and 29 without ASD), 16 children with non-syndromic ASD and 46 healthy controls. Groups were identified by two indicator variable: TSC (indicating whether the individual has TSC) and ASD (indicating whether the individual has ASD). The brain was modeled as a network whose nodes are the electrodes and whose edges are the coherence between EEG signals (Fig. 1.2C). Coherence measures the strength of the connection in a particular frequency band. Both conventional and graph-theoretical measures were used to characterize and compare the functional networks [95]. Conventional measures included (i) the mean coherence, (ii) the ratio of long-range (Euclidean distance between electrodes larger than 3 in the grid of Fig. 1.2B) over short-range (Euclidean distance of two) connectivity, and (iii) the ratio of the mean coherence of all corresponding inter-hemispheric electrode pairs over all non-midline intrahemispheric electrode pairs. Graph-theoretical measures included the clustering coefficient, the average path length, the global efficiency and the resilience [95]. The resilience was measured as the fraction of global efficiency remaining when the N_{rem} nodes with highest degree centrality are removed from the network ($1 \leq N_{\text{rem}} \leq 5$).

For each measure, a two-way analysis of covariance (ANCOVA) was used to assess whether significant differences associated with TSD and/or ASD were observed. The ANCOVA is a conventional method to include some sources of stochastic variability in the analysis. The subjects' age was use as a covariate of the analysis so that the generalized linear model reads:

$$y = \bar{y} + \beta_{\text{ASD}}\text{ASD} + \beta_{\text{TSC}}\text{TSC} + \beta_{\text{age}}\text{age},$$

for every measure y given the population average $\bar{y}$. Analyses were conducted in the theta band, the lower and the upper alpha bands.

The results indicate that traits of TSC and ASD are mostly reflected in the lower alpha band (Fig. 1.3). In this band, TSC is associated with a significantly decreased mean connectivity ($p < 0.005$), a significantly decreased clustering coefficient ($p < 0.002$), a significantly increased average path length ($p < 0.008$) and a significantly decreased global efficiency ($p < 0.009$). ASD is associated with a significantly lower long-over-short-range connectivity ratio ($p < 0.0002$) and a significantly increased resilience ($p < 0.05$ for $N_{\text{rem}} = 1, 3, 4, 5$ and $p = 0.06$ for $N_{\text{rem}} = 2$). More importantly, these findings were associated with the condition as a whole (TSC or ASD) and not with a specific subgroup of the population (*e.g.* only children with syndromic ASD). This observation was tested by performing post-hoc t-tests, testing whether significant differences

occur between the subgroups of a condition. No significant subgroup difference was found for the findings detailed above ($p > 0.05$).

These results indicate that autism spectrum disorders, regardless of etiology, share common patterns of disconnection. Since these results were based on an EEG study, they are not affected by head motion (which is detected and removed from EEG signals). This study thus provides evidence that common mechanisms occur in autism spectrum disorders. These mechanisms may therefore lead to the definition of one or several biomarkers. EEG, however, only provides a characterization of the brain at a coarse spatial resolution. This resolution does not allow for accurate localization of brain abnormalities. Increasing the number of electrodes would not indefinitely improve the spatial resolution due to volume conduction of electrical signals and because the skull spatially smoothes the signals. While EEG is thus unlikely to be specific enough to provide reliable biomarkers of autism spectrum disorders, it can guide the search for such biomarkers.

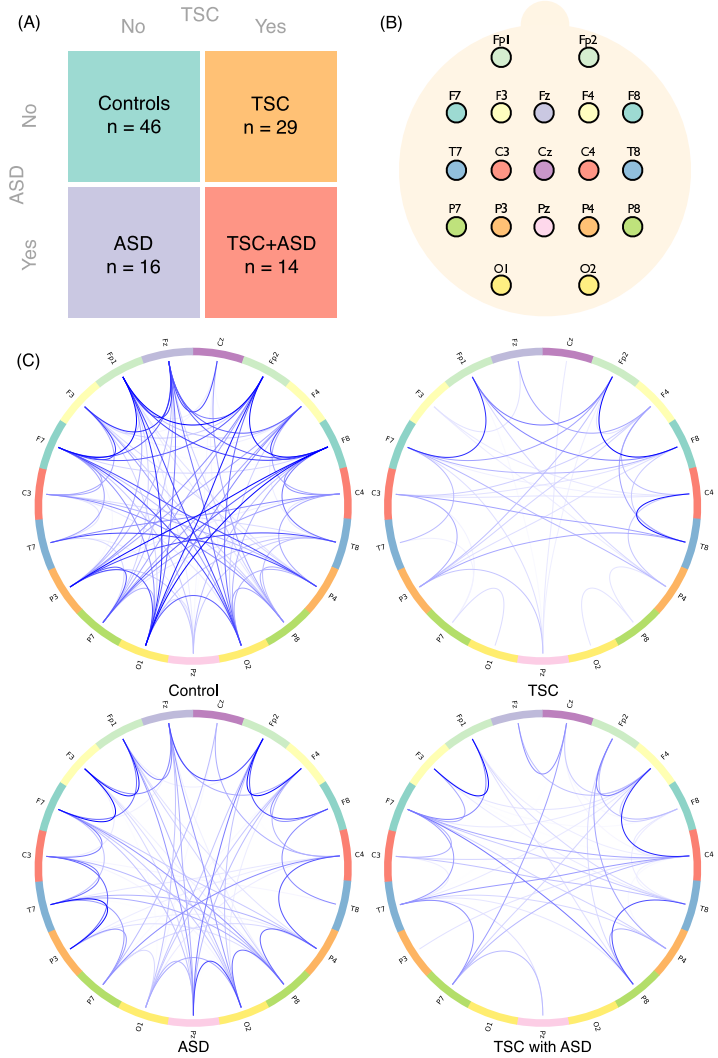


Figure 1.2: Functional networks in autism. (A) In the two-way representation of our population, groups are identified by the subject being diagnosed with or without Tuberous Sclerosis Complex (TSC) and/or with or without Autism Spectrum Disorder (ASD). This structure allows an independent attribution of effects specific to TSC and ASD. (B) Electrode locations from the international 10-20 system of electrode placement are used as nodes in the network. (C) Illustrations of the functional networks of a control subject, a TSC patient, a non-syndromic autistic patient, and a TSC patient diagnosed with autism. Colors on the connection terminations correspond to the colors in (B) and the darkness of the line correlate with the strength of the connection.

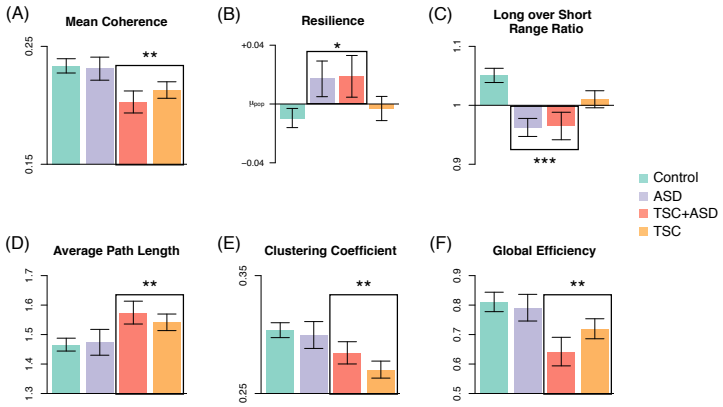


Figure 1.3: Results of the population study of functional networks in autism spectrum disorders (ASD) and tuberous sclerosis complex (TSC). TSC is associated with decreased mean coherence (A), increased average path length (D), decreased clustering coefficient (E) and decreased global efficiency (F). ASD is associated with increased resilience (B) shown for $N_{\mathrm{rem}} = 3$ and decreased long over short range connectivity ratio (C). Remarkably, differences were associated with a condition (TSC or ASD) rather than a specific subgroup.
$\star$: $p < 0.05$, $\star\star$: $p < 0.005$, $\star\star\star$: $p < 0.001$

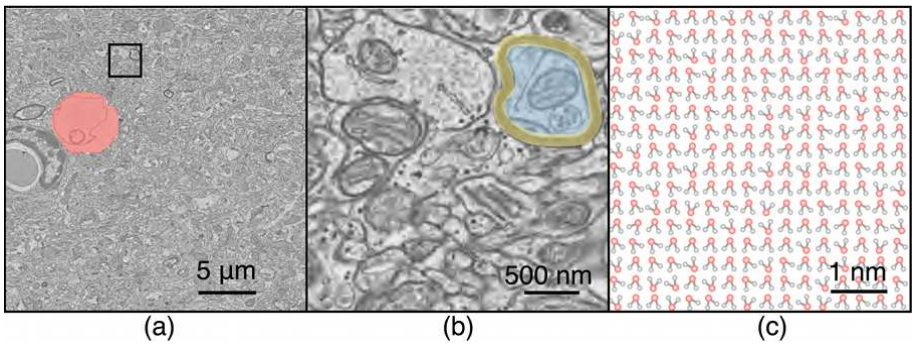


Figure 1.4: Components and organization of the brain microstructure. The brain microstructure is composed of different compartments. Each compartment impose different barriers to the diffusion of water molecules present within them. (a) A piece of the visual cortex of a mouse imaged with an electron microscope showing the heterogeneity of microstructural compartments and a comparatively large blood vessel (in red). (b) A zoomed-in version of (a) showing a single neuron in blue with its myelin sheath in yellow. (c) Each voxel of the electron microscopy image contains thousands of water molecules. Water molecules contain two hydrogen atoms, the protons of which contribute to the magnetic resonance signal.

1.2 *In Vivo* Analysis of the Brain Microstructure

Brain connectivity is supported by the architecture of its neurons and their axons. Axons are projections of the neurons and act as the cables of the brain transmitting the electrical impulse generated by the neuron's body (the *soma*). This transmission is supported by a network of blood vessels that provides the necessary nutrients and oxygen, and by glial cells that maintain homeostasis and provide protections for the neurons. Oligodendrocytes are glial cells that wrap around the axons to form the myelin sheath that increases the propagation speed of electrical impulses along the fibers. The *brain microstructure* is the complex organization of all these compartments: the neurons and their dendrites and axon, the glial cells, the vascular capillary bed and the lattice of extracellular space [13] (Fig. 1.4).

Axons have various sizes, directions and degrees of myelination (Fig. 1.4b). The volume of extracellular space also varies across the brain. Despite this heterogeneity of cellular compartments, the organization of the microstructure is not random and several key properties of the microstructure present detectable

patterns at the millimeter scale. For instance, the corpus callosum is a region of the brain located near the midline, in which densely packed fibers connect the left and right hemispheres of the brain. The microstructure in the corpus callosum therefore contains aligned axons and a small volume of extracellular space. These patterns and properties can be measured with brain imaging at the millimeter scale.

All microstructural environments contain water molecules that move (*diffuse*) according to a Brownian motion. One cubic millimeter of tissue can contain billions of billions ($\sim 10^{19}$) of water molecules (Fig. 1.4c). These water molecules contain two hydrogen atoms whose protons generate the magnetic resonance signal (Fig. 1.4d). Motion of water molecules are detected by diffusion-weighted magnetic resonance imaging, or diffusion-weighted imaging (DWI) for short.

DWI provides indirect measurements of the microstructure by characterizing how water molecules diffuse in the brain. The underlying microstructural properties can be inferred by fitting a model to the DWI data. Various microstructural models have recently been proposed and are reviewed in the next chapter. Regardless of the selected model, numerous challenges arise from their introduction to study the brain microstructure. These challenges relate to the acquisition, estimation and exploitation of microstructural models for population studies.

Models of the brain microstructure are described by several parameters at each voxel. Their optimization therefore requires several scalar diffusion-weighted images to be acquired with various parameters. The duration of DWI is therefore typically much longer than that of conventional MRI and increases with the complexity of the model at hand. Most children with ASD have difficulty in following instructions and remaining still in the MRI scanner. Patient motion creates artifacts in the resulting images that may ultimately jeopardize the use of these images to estimate the parameters of a microstructural model. To minimize patient motion, sedation and anesthesia are often used, substantially increasing the cost of the procedure and the risk to the patient. To offset the persistent limitation of patient motion without resorting to anesthesia, fast imaging of the brain microstructure is thus needed.

Once DWI have been acquired, the microstructural model needs to be estimated. An appropriate model needs to be selected and, for the selected model, the number of compartments needs to be determined. These compartments are described by a number of parameters whose value is to be optimized.

Finally, images of the brain microstructure contain, within each voxel, an

entire model of the local diffusion instead of a single scalar value (as in conventional MRI). Basic tasks commonly performed on MRI (such as interpolating, averaging or spatially aligning images) require methodological advances to be performed on microstructural images. Solving these tasks is a requirement to exploit microstructural images in population studies. Population studies also require microstructural properties to be compared between groups. The definition of a system for statistical microstructure analysis is therefore needed.

1.3 Contributions of this Thesis

As we will present in the next section, microstructural images are commonly represented as multi-fascicle models in which the diffusions of different populations of water molecules have their own parametric representation. In this thesis, we present a framework to conduct population studies from multi-fascicle models of the brain microstructure. The main contributions of this thesis are:

1. A mathematical framework for the registration and analysis of multi-fascicle models. This framework provides definitions of simple operators such as averaging, interpolation and smoothing for multi-fascicle models. It also provides a robust similarity metric between multi-fascicle models to spatially align them. This framework enables the construction of an atlas of the brain microstructure on which multi-fascicle models of subjects can be aligned to assess group differences [125, 123, 127].

2. Formulation of the selection of the appropriate model as the minimization of the generalization error. This formulation enables the reliable identification of an appropriate model and estimation of the number of compartments, by avoiding data overfitting [100].

3. Estimation of multi-fascicle models from single-shell diffusion imaging that is both fast and widely available. We first demonstrate that this estimation is an ill-posed problem and we explore the geometry of the manifold of all equivalent models. We then propose an efficient method to regularize the problem by including prior knowledge from scans of other subjects into a maximum *a posteriori* formulation of the problem [129, 126].

The contributions of this thesis are illustrated in the specific context of detecting relevant group differences in autism spectrum disorders and tuberous sclerosis

complex. The proposed framework, however, can be exploited in various other neurological and psychiatric contexts.

1.4 Thesis Outline

To conduct a population study of the brain microstructure, one typically proceeds as follows:

1. Acquire diffusion-weighted images (DWI) in each subject
2. Estimate a microstructural model from the DWI of each subject
3. Spatially align (*register*) all microstructural models to an atlas
4. Statistically analyze microstructural properties in the atlas reference frame.

These steps require novel assets and capabilities which correspond to the contributions of this thesis (Fig. 1.5). The outline of this thesis starts with the most fundamental assets and capabilities and progressively builds upon new assets to develop the framework.

Chapter 2 reviews the background material needed for the developments of this thesis, including diffusion-weighted imaging, diffusion tensor imaging and its limitations, and microstructural models. Chapter 3 defines the mathematical framework that enables registration and analysis of multi-fascicle models. Chapter 4 formulates the selection of an appropriate microstructural model as the minimization of the generalization error. Chapter 5 defines a method to estimate multi-fascicle models that requires few diffusion-weighted images at a single non-zero b-value. Chapter 6 explores prospective developments in the field of microstructure analysis that go beyond the characterization of multiple fascicles.

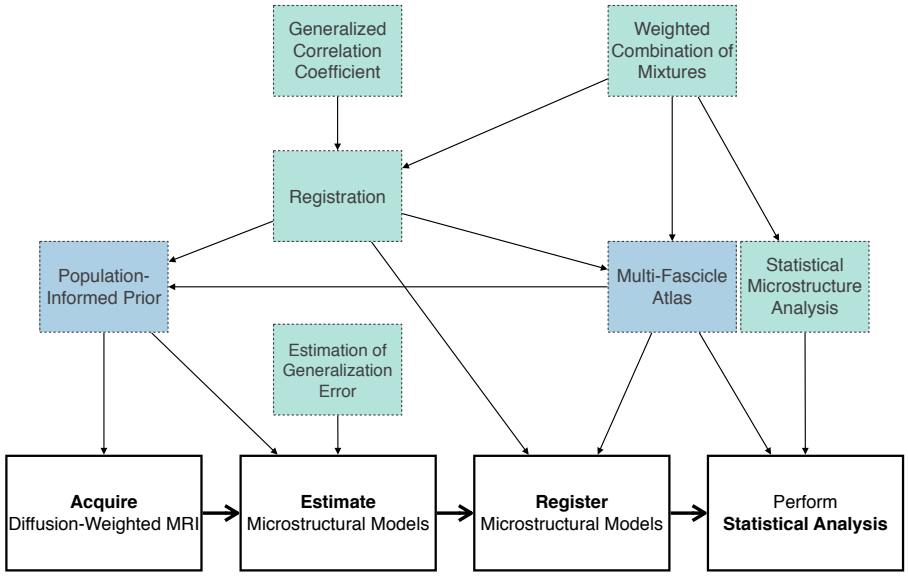


Figure 1.5: Workflow of population studies of the brain microstructure and contributions of this thesis. The steps to conduct population studies of the brain microstructure are: acquisition, estimation, registration and statistical analysis. These steps require some assets (blue) and capabilities (in green) that we contribute in this thesis. Arrow $X \rightarrow Y$ encodes the relation "Y requires X". The thesis outline starts with requirements and progressively describes new assets and capabilities.

Chapter 2

Background

In vivo studies of the brain microstructure are based on acquisition of diffusion weighted images from which a model is estimated. In this chapter, we first explain the basics of diffusion-weighted magnetic resonance imaging. We then describe the most commonly used model of the diffusion signal, called diffusion tensor imaging, and outline its limitations. Finally, we summarize the recent developments made in modeling the diffusion signal.

2.1 Diffusion-Weighted Imaging

Diffusion-weighted imaging (DWI) allows the characterization and quantitative measurement of the diffusion of water molecules in tissues. It enables the distinction between unrestricted diffusion of protons and restricted diffusion of protons, based on the random motion of water molecules in tissue [26].

2.1.1 Physical Basis

In pure water, there is no barrier to the diffusion of water molecules and the diffusion is referred to as *unrestricted*. In contrast, in the brain, local restriction to water diffusion such as caused by the presence of densely organized white matter fiber bundles (also known as *fascicles*), gives rise to *restricted* diffusion. Furthermore, when the diffusion has a directional preference (such as in white matter fascicles where diffusion is less constrained in the direction of the fascicles), it is referred to as *anisotropic* whereas a diffusion with no directional preference is coined *isotropic*. The degree of anisotropy and restriction depends on the microstructure present in the voxel. In highly organized structures (*e.g.* white matter fascicles), diffusion will be highly anisotropic and, in the direction orthogonal to fascicles, highly restricted, as molecules will diffuse preferably along the path of least resistance (*e.g.* along the axon within the myelin sheath). In less coherent structures (*e.g.* in a tuber consisting of poorly organized collection of cells), the diffusion will be almost isotropic and

moderately restricted. In the cerebrospinal fluid (CSF), the diffusion will be unrestricted and isotropic.

In an MRI scanner, protons' spins are initially aligned with the strong magnetic field produced by the magnet. Applying a short magnetic pulse changes this orientation and protons' spins start to precess (much like a spinning humming top deviates from its central axis). This precession generates an electromagnetic signal detectable by an electric coil. The rate of precession depends on the strength of the magnetic field. By applying a magnetic field that varies along a certain direction (adding a so-called *field gradient pulse* to the magnetic field), we can label the spins by different precession rates according to their position along the gradient direction. This variation in precession rates results in an interference between the precessing spins, leading to a signal attenuation. Applying the opposite field gradient pulse would refocus the spins and counterbalance the signal attenuation, only if the protons did not move between the two pulses. However, due to motion, protons' spins are imperfectly refocused and the signal attenuation cannot be completely compensated for. The amount of remaining signal attenuation is related to the amount of motion that occurs in the gradient direction. Measuring the signal attenuation therefore measures the diffusion of protons (or water molecules that contain them). This is the physical basis of diffusion-weighted imaging.

2.1.2 B-value, Echo Time and Acquisition Time

The diffusion-weighted MRI signal in an homogenous medium is given by the Stejskal-Tanner equation [119]:

$$S = S_0 e^{-\frac{T_E}{T_2}} e^{-\gamma^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right) \|\boldsymbol{g}\|^2 D}, \tag{2.1}$$

where S_0 is the signal obtained if no signal attenuation occurred, D is the diffusion coefficient, γ is the gyromagnetic ratio that depends on the tissue, T_2 is the time constant for spin-spin relaxation that also depends on the tissue, $\|\boldsymbol{g}\|$ is the strength of the gradient pulse, T_E is the echo time, δ is the duration of the field gradient pulse and Δ is the separation in time between two pulses. The latter three parameters (T_E, δ and Δ) are chosen for the acquisition and must be such that $T_E > 2\Delta > 2\delta$. The gradient pulse is developed by three coils generating a field in three orthogonal directions. Each coil can generate a field of a magnitude up to $g_{\max}$, so that the maximum magnitude of the three-dimensional gradient pulse $\boldsymbol{g}$ is $\sqrt{3} g_{\max}$ and is obtained when the

three coils produce their maximum field. We let $b_{\text{nominal}} = \gamma^2\delta^2(\Delta - \delta/3)g_{\max}^2$ and $b = b_{\text{nominal}}\|\boldsymbol{g}\|^2/g_{\max}^2$ that we denote the *nominal* and *effective* b-values respectively [102]. We have $0 \leq b \leq 3b_{\text{nominal}}$.

Diffusion-weighted images (DWI) are scalar images that contain, at each voxel, the value of the signal S for a given effective b-value and a given gradient direction $\hat{\boldsymbol{g}}$. In heterogeneous microstructures, multiple DWI are typically acquired with various gradient pulse directions $\hat{\boldsymbol{g}}$ and various effective b-values. The gradient field is generated by three electromagnetic coils in the three cartesian directions. Each coil can create a field of a maximum amplitude $g_{\max}$ that is a scanner constraint. Due to this constraint, increasing the b-value is typically done by increasing δ and/or Δ. Increasing δ and/or Δ, in turn, increases the echo time T_E. The increase in echo time has two adverse effects: it increases the acquisition time of the images and it decreases the signal-to-noise ratio (SNR) of the DWI due to the factor $e^{\frac{-T_E}{T_2}}$ in (2.1). Too low an SNR can be compensated for by averaging multiple DWI acquired with the same parameters. Averaging N DWI results in an SNR increased by a factor $\sqrt{N}$. Acquiring N images takes N times longer so that both adverse effects related to increased T_E result in larger increases in acquisition time.

To illustrate the extent of this effect, let us imagine that we need to triple the nominal b-value. Since the nominal b-value is proportional to $\delta^2(\Delta - \delta/3)$, this can be achieved by multiplying δ and Δ by a factor of $3^{1/3} = 1.44$. Let us assume that the corresponding T_E also increases by a factor 1.44, going from 100 ms to 144 ms. For a typical T_2 in the white matter of 80 ms [139], the SNR resulting from this increase in T_E is only 58% of the initial SNR. This decrease in SNR can be compensated for by acquiring three times as many images ($1/0.58^2$). Since each image takes approximately 1.44 times longer to be acquired (due to increase in T_E), the whole process results in an imaging time that is 4.33 longer than the initial imaging time. In other words, a scan that initially takes 6 minutes would take over 25 minutes if the b-value needs to be multiplied by 3 by increasing δ and Δ. While asking for children —and children with autism in particular— to remain still for 6 minutes is challenging, asking them not to move for 25 minutes is typically futile.

Recently, a novel acquisition scheme, namely cube and sphere (CUSP) imaging, has been proposed to acquire DWI at multiple b-values without increasing the T_E [102]. The essence of CUSP is to increase the effective b-value without increasing the nominal b-value, by increasing the strength of the gradient field, $\|\boldsymbol{g}\|$. The maximum squared gradient magnitude is $3g_{\max}^2$. This

maximum amplitude can only be achieved for gradients located at the eight corners of the cube in the space of gradients $\boldsymbol{g}$: $\boldsymbol{g} = (\pm g_{\max}, \pm g_{\max}, \pm g_{\max})$. All the gradients lying on this cube have an effective b-value between b_{nominal} and $3b_{\text{nominal}}$ and, in each direction $\boldsymbol{g}$, the magnitude of the gradient field is maximized for a given T_E. CUSP combines gradients on the cube and gradient on the enclosed sphere (also called *shell*) described by $\|\boldsymbol{g}\| = g_{\max}$. The latter all have gradient field strength equal to $g_{\max}$ yielding b-values equal to b_{nominal}. By increasing the b-value without increasing the T_E, CUSP enables the acquisition of DWI at multiple b-values in clinically achievable times.

Models of the brain microstructure relate the diffusion signal S to the gradient direction $\hat{\boldsymbol{g}}$ and the effective b-value b. In the sequel, to align our notations with those commonly used in the literature, the term *b-value* refers to the effective b-value and the gradient directions $\hat{\boldsymbol{g}}$ is simply denoted $\boldsymbol{g}$ (with the actual squared norm of the magnetic gradient included in the b-value). The factor $e^{-\frac{T_E}{T_2}}$ is also implied so that S_0 actually refers to $S_0 e^{-\frac{T_E}{T_2}}$. With these notations, the equation (2.1) for isotropic diffusion reads:

$$S = S_0 e^{-bD}. \tag{2.2}$$

Increasing the number of acquired DWI by increasing the number of directions $\boldsymbol{g}$ and the number of b-values enables better characterization of the brain microstructure, by providing more data points to estimate the parameters of the model $S(b, \boldsymbol{g})$. This increase always comes at the expense of longer acquisition times that may not be practical in a clinical setting. In Chapter 5, we will propose a method to accurately estimate microstructural models with few acquisitions at a single b-value.

2.2 Diffusion Tensor Imaging and its Limitations

Diffusion Tensor Imaging (DTI) is the first and most widely used model of the diffusion signal in tissues [20]. It enables the definition of scalar measures that can then be used to detect group differences in population studies. DTI, however, is affected by various weaknesses which limit their use and mislead conclusions from population studies.

2.2.1 Tensor Model

DTI models the average diffusion direction and strength at each voxel with a tensor, which can be thought of as an ellipsoid (Fig. 2.1a). This ellipsoid is

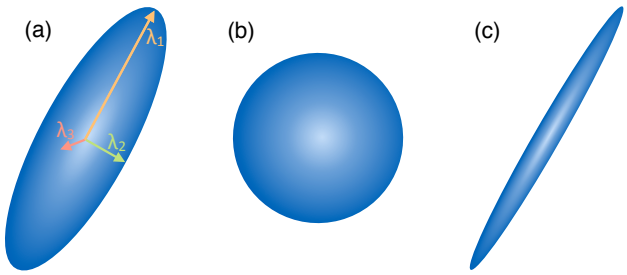


Figure 2.1: Ellipsoid representation of the diffusion tensor model. (a) Diffusion tensor imaging can be represented as an ellipsoid that consists of three axes of diffusion and the corresponding diffusivities (here λ_1, λ_2 and λ_3). The shape of the ellipsoids provides information about the type of diffusion present in the voxel. (b) An isotropic diffusion leads to a spherical tensor. (c) Diffusion that is highly restricted in two directions and favored in one direction will present as an elongated tensor with very small second and third diffusivities.

characterized by a principal direction along which diffusion is the strongest. In the two orthogonal directions, diffusion is more restricted and its magnitude is given by the width of the ellipsoid in those directions. A total of six parameters are required to fully define the ellipsoid: three parameters for the widths and length (these are also called diffusivities), two parameters to define the direction of strongest diffusivity and one parameter to define the rotation of the ellipsoid around its principal axis. The shape of the ellipsoid provides information on the nature of the diffusion occurring in the corresponding voxel. Isotropic diffusion (as that occurring in free water) gives rise to a spherical tensor (Fig. 2.1b), and its diffusivities are equal in all directions. Highly anisotropic diffusion gives rise to long and thin ellipsoids (Fig. 2.1c), indicating that the diffusion is highly favored along a principal direction and highly restricted in the other two directions.

Mathematically, ellipsoids are represented as 3×3 symmetric positive-definite matrices that we denote $\boldsymbol{D}$. Diffusivities correspond to the eigenvalues of this matrix and are denoted $\lambda_1 \geq \lambda_2 \geq \lambda_3$ (Fig. 2.1a). The directions along which the diffusivity is equal to λ_i is the i-th eigenvector of $\boldsymbol{D}$ and is denoted $\hat{\boldsymbol{e}}_i$. For a general gradient direction $\hat{\boldsymbol{g}}$ (a 3-dimensional vector with unit norm), the diffusion coefficient is $\hat{\boldsymbol{g}}^T \boldsymbol{D} \hat{\boldsymbol{g}}$ and we can verify that $\hat{\boldsymbol{e}}_i^T \boldsymbol{D} \hat{\boldsymbol{e}}_i = \lambda_i$. A generalization of (2.2) that accounts for the anisotropy of the diffusion coefficient

is therefore:

$$S(b, \boldsymbol{g}) = S_0 e^{-b\boldsymbol{g}^T \boldsymbol{D}\boldsymbol{g}}. \tag{2.3}$$

Under the assumption that δ is negligible compared to Δ (*narrow pulse approximation*), the diffusion signal can be directly related to the diffusion process by the inverse Fourier transform [28]. Specifically, the probability density of protons to undergo a displacement $\boldsymbol{r}$ in a time Δ is

$$p_\Delta(\boldsymbol{r}) = S_0^{-1}(2\pi)^{-3} \int_{\mathbb{R}^3} |S_\Delta(\boldsymbol{q})| e^{-i\boldsymbol{q}\boldsymbol{r}} d\boldsymbol{q},$$

where $\boldsymbol{q} = \gamma\delta\|g\|\hat{\boldsymbol{g}}$ and $S_\Delta(\boldsymbol{q}) = S(b = \|q\|^2\Delta, \boldsymbol{g} = \hat{\boldsymbol{q}})$. Under this interpretation, the multivariate Gaussian function of (2.3) implies that the diffusion of water molecules follows a Gaussian process whose covariance matrix is $\boldsymbol{D}$ (since the Fourier transform of a multivariate Gaussian is a multivariate Gaussian with a covariance matrix equal to the precision matrix of the initial Gaussian). The ellipsoids represented by $\boldsymbol{D}$ in this interpretation correspond to a surface of equal probability to find a proton that was initially located at the center of the ellipsoid. The narrow pulse approximation is often violated because the separation Δ between pulses tends to be minimized ($\Delta \approx \delta$) to minimize T_E and thereby maximize the SNR [65]. The inverse Fourier transform of the diffusion signal thus only provides intuitions (and not measurements) about the underlying diffusion process.

2.2.2 DTI-Based Measures

DTI-based measures quantify the shape of the ellipsoids in diffusion tensor imaging. Two main measures are commonly used: the mean diffusivity (MD) and the fractional anisotropy (FA). The mean diffusivity is the average diffusion in all three directions. In terms of the tensor formalism, it is defined as:

$$\text{MD} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} = \frac{1}{3}\text{Tr}\boldsymbol{D}.$$

The mean diffusivity is an intrinsic property of tissues. For example, the MD of demyelinated white matter is increased as there is more extracellular water and a weaker biological barrier to diffusion [116]. The fractional anisotropy (FA) reflects the degree of asymmetry between diffusivities in all three directions. If the diffusion is completely isotropic (Fig. 2.1b), then FA is equal to zero. Conversely, if the diffusion is extremely anisotropic (water molecules can only

diffuse in one direction and not in the other two directions), then FA is equal to one. In terms of the tensor formalism, the FA is defined as:

$$\mathrm{FA} = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \mathrm{MD})^2 + (\lambda_2 - \mathrm{MD})^2 + (\lambda_3 - \mathrm{MD})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}.$$

Diffusion within the white matter axons is restricted to the longitudinal axis by cell membranes and by the the myelin sheath which form a biological barrier, resulting in a high FA. When neurons or myelin sheaths are damaged, the FA decreases: there is less preferential directionality of diffusion because the fluid can move freely along various axes. MD and FA can be altered by any pathological process that modifies tissue integrity and leads to an attenuation of structural barriers to water motion.

The FA and MD are computed in each voxel individually, yielding scalar images of the measure (Fig. 2.2b-c). An alternative way of displaying DTI is to color-code the FA by the principal direction of the tensor. A tensor aligned with the left-right axis of the brain gets a red color, one that is oriented along the anterior-posterior axis gets a green color, and a tensor aligned with the superior-inferior axis gets a blue color. Tensors that are in between these axis receive a color that is a weighted sum of these principal colors. The value of the color (in a hue-saturation-value encoding) encodes the FA so that voxels with low FA are dark and values with high FA are bright (Fig. 2.2d).

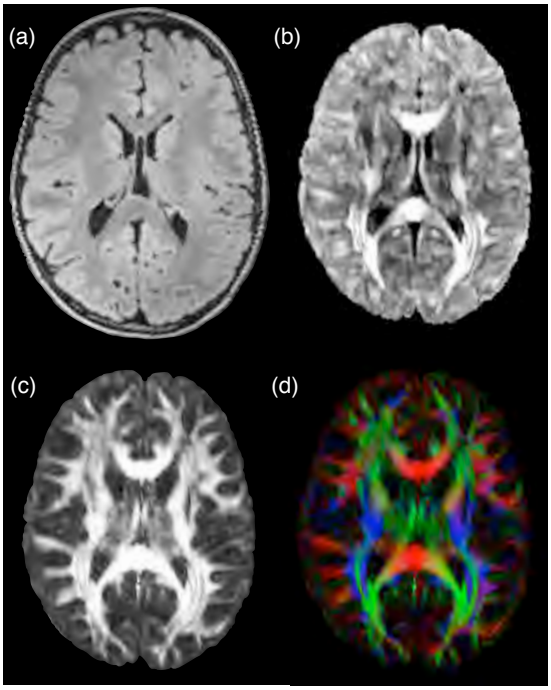


Figure 2.2: Diffusion tensor imaging (DTI) and DTI-based measures. (a) Conventional, structural MRI. (b) Mean diffusivity (MD) image showing that MD is especially large in the corpus callosum and in corticospinal tracts. (c) Fractional anisotropy (FA) image shows where, in the brain, diffusion is more (white) or less (black) anisotropic. FA in the white matter is high due to the presence of highly structured white matter fascicles. By contrast, gray matter presents in axons with various directions, resulting in a lower FA. (d) FA can be colored based on the directions of the fascicle in each voxel: red means the fascicle is oriented left to right, green represents fascicles that are oriented along the anterior-posterior axis and blue represents the superior-inferior axis.

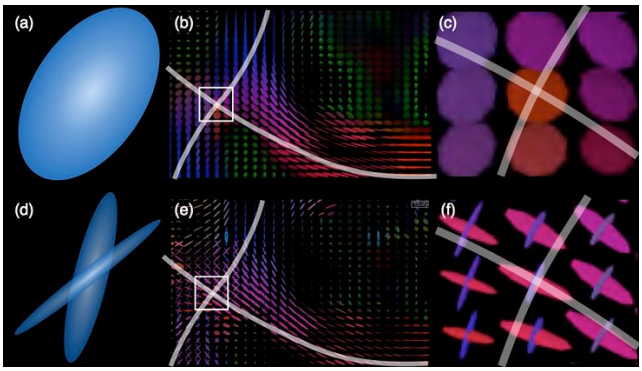


Figure 2.3: Single tensor and multi-fascicle models. (a-c) Unlike assumptions of the diffusion tensor models, fascicles in the voxels may have more than one preferential direction. Diffusion tensor imaging model assumes that the diffusion is either purely unrestricted or purely restricted within a single fascicle. This assumption is violated in regions where fascicles cross, such as (b) the corona radiata. In those regions, tensors are abnormally inflated to capture the signal arising from (c) each fascicle, resulting in a lower fractional anisotropy that may be misleadingly interpreted. By contrast, (d-f) multi-fascicle models represent each fascicle independently and are, therefore, able to characterize regions with crossing fascicles.

2.2.3 Processing Diffusion Tensors

The processing of diffusion tensors (averaging, interpolating, smoothing, regularizing, etc.) is challenging because any operations must yield positive definite matrices. This constraint is not respected by Euclidean metrics, *i.e.* performing element-wise operations on the entries of the matrices. For instance, the element-wise subtraction of two tensors may lead to null or negative eigenvalues.

Furthermore, the determinant of a diffusion tensor, which is related to the volume of the confidence region of finding a water molecule that has diffused, is not preserved by Euclidean processing. The Euclidean average of two tensors may lead to a tensor with a higher determinant than those of both initial tensors, an effect known as the *swelling effect* [8].

For these reasons, tensors are typically processed in the log-Euclidean domain, which is an interesting and efficient alternative to the Euclidean processing [7, 8, 106]. Processing tensors in the log-Euclidean domain simply consists in computing operations on the matrix logarithm of the tensors and subsequently computing the matrix exponential of the result. As an illustration,

the logarithmic version of the average of N tensors $\boldsymbol{D}_i$ is:

$$\mu_{\boldsymbol{D}} = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log \boldsymbol{D}_i\right).$$

The matrix exponential is defined as the Taylor series of the exponential function applied to the matrix. The logarithm, $\log \boldsymbol{D}_i$, of a matrix $\boldsymbol{D}_i$ is defined as the matrix whose exponential is $\boldsymbol{D}_i$. The logarithm $(\log \boldsymbol{D}_i)$ has the same eigenvectors as $\boldsymbol{D}_i$ and eigenvalues that are equal to the scalar logarithm of the eigenvalues of $\boldsymbol{D}_i$. Interestingly, the determinant of the log-Euclidean average of N tensors is exactly equal to the scalar geometric mean of the determinants of the original tensors [8].

The log-Euclidean processing of diffusion tensors does not, however, preserve anisotropy. The FA of the log-Euclidean average of diffusion tensors, for instance, may be lower than the FA of all original tensors. Another processing scheme has recently been proposed to circumvent this problem, based on the processing of spectral-quaternion [31].

2.2.4 Limitations of the DTI Model

The diffusion tensor model relies on the assumption that all the water molecules in one voxel follow the same diffusion process and are equally constrained by the brain tissues. We have already seen in Section 1.2 that *microscopic* observations of the brain microstructure contradict these assumptions (because of the heterogeneity of axonal diameters, orientations, degree of myelination and tortuosity, and the presence of an extra-axonal matrix). The utmost limitations of the DTI model, however, are related to two *macroscopic* effects: the presence of heterogeneous fascicle orientations and the partial volume effect [131].

a) Heterogeneous Fascicle Orientations

The DTI model assumes that, at each voxel, the diffusion is Gaussian with either no preferential direction (isotropic diffusion) or one preferential direction. This assumption is reasonable only when all axons in the voxel are contained into one fascicle with a specific orientation. However, owing to the presence of complex fascicle organization, heterogeneous fascicle orientations can be present in one voxel. In the corona radiata, for instance, corticospinal tracts (tracts connecting the cortex to the spinal cord) cross fascicles of the corpus callosum (the horizontal tracts that connect the left and right hemispheres)

(Fig. 2.3a-c). Another example are the pyramidal projections that give rise to fanning fascicles (i.e., fascicles that follow different directions from an original point at which they are aligned) [131]. Recent studies estimate the prevalence of those heterogeneities to range between 60 and 90% of voxels in the white matter at typical DWI resolution [55]. When fascicles are crossing, interpretation of the DTI-based measures (MD and FA) may be misleading [136]. For instance, in the presence of two crossing fascicles, a single overly wide tensor would be estimated resulting in a decreased FA. This decreased FA is not related to a property of the fascicle and, if interpreted this way, may incorrectly lead to the assumption that the myelin is altered for that fascicle.

b) Partial Volume Effect

Voxels that are at the interface between different tissues (gray and white matter), between adjacent fiber bundles or between a tissue and cerebrospinal fluid (CSF) suffer another problem called *partial voluming*. The diffusion signal arising from protons in the different compartments (CSF, gray or white matter) will be averaged into a single value that is observed in DWI. Because DTI assumes that the diffusion of water molecules is either purely unrestricted or purely restricted within a single fascicle in the voxel, influences of different compartments will conflate, resulting in an inflated tensor with a lower FA. As with heterogeneous fascicle orientations, this decreased FA may be misleadingly interpreted as altered myelin.

Even in voxels that are not at the interface between tissues, some fraction of the water molecules will diffuse in the extra-axonal space (either within other cells or in the extracellular lattice). Their diffusion process is different from the diffusion within axons and results in a different diffusion signal. The partial volume effect therefore also applies within a fascicle, between molecules inside and outside axons. The partial volume effect is essentially a resolution limitation. At an infinite resolution, the diffusion model would pertain to a single water molecule and would be relatively simple compared to the models of the brain microstructure that we describe below.

2.3 Models of the Brain Microstructure

To overcome the limitations of the diffusion tensor model, various models of the brain microstructure have been proposed. We start by expressing a fundamental

assumption that is used in most models of the brain microstructure. We then discuss models that extend the tensor model for a single fascicle. We finally describe multi-fascicle models.

2.3.1 Non-Exchangeability of Water Molecules

Most limitations of the diffusion tensor model amount to the presence of populations of water molecules present in different compartments and having different diffusion patterns. These compartments are typically not completely impermeable and water molecules may transit from one compartment to another at a certain *exchange rate*. However, if this exchange rate is slow (compared to the diffusion time), a *non-exchangeability* assumption can be made. Under this assumption, the fraction of water molecules in each compartment remains constant throughout the diffusion process. The measured diffusion signal is then simply the sum of the signals S_i contributed by each population of water molecules, weighted by the fraction f_i of water molecules in each population:

$$S = \sum_{i=1}^{N} f_i S_i,$$

with $\sum_{i=1}^{N} f_i = 1$. Since the non-attenuated signal S_0 is the same for all compartment (it does not depend on any diffusion process), we also have

$$S = S_0 \sum_{i=1}^{N} f_i A_i,$$

where A_i is the attenuation of the signal in the i-the compartment.

The validity of this non-exchangeability assumption depends on the actual rate of exchange of water molecules between compartments [131]. As we will see in the next two subsections, the scale of these compartments vary and, thereby, so does their exchange rate. Macroscopic compartments consist in fascicles and/or actual fractions of the voxel (in the case of partial voluming with other tissues or CSF). Microscopic compartments consist in individual axons and the surrounding extra-axonal space. The non-exchangeability assumption is commonly made at every scale, mostly owing to the assumption that individual axons are almost impermeable [14].

2.3.2 Models of a Single Fascicle

Recent diffusion models of a single fascicle have been proposed to overcome the limitations of DTI at the microscopic level. These imperfection can be attributed to three main reasons: the presence of an extra-axonal space, the non-Gaussianity of the diffusion process and the heterogeneity of axonal properties.

a) Modeling the Extra-Axonal Space

The brain does not only consist of axons. Outside the axons, in the *extra-axonal space*, the diffusion of water molecules is different from that within the axons. The measured diffusion signal receives contributions from all water molecules. Therefore, the signal arising from water molecules in the extra-axonal space ought to be modeled separately. Under the non-exchangeability assumption, the diffusion model becomes:

$$S = S_0(f_{\text{intra}}A_{\text{intra}} + f_{\text{extra}}A_{\text{extra}}), \tag{2.4}$$

where f_{intra} and f_{extra} are the fractions of water molecules in the intra-axonal and extra-axonal space, respectively (with $f_{\text{intra}} + f_{\text{extra}} = 1$), and A_{intra} and A_{extra} are the corresponding signal attenuations.

Water molecules outside the axons are either in the extracellular matrix or trapped in other cells (mostly glial cells). The diffusion of water molecules in the extracellular space is hindered by the surrounding tissue microstructure including cells, membranes, axons, etc. This diffusion is well modeled by a diffusion tensor. Isotropic [76, 22, 102, 82, 80], cylindrically symmetric [2] or full [14] tensors have been proposed in this context. The diffusion in glial cells depends on the shape of the cell but is typically modeled as isotropic (either as an isotropic tensor [102], as a sphere [118], as dots [5] or as a uniform spherical distribution of infinitely anisotropic tensors, known as astrosticks [79]).

As an example, Pasternak *et al.* [80] use an isotropic tensor to model the extracellular space and an anisotropic tensor to represent the diffusion in the supposedly unique fascicle:

$$S = S_0(f_{\text{intra}}e^{-b\boldsymbol{g}^T\boldsymbol{D}_{\text{intra}}\boldsymbol{g}} + f_{\text{extra}}e^{-bD_{\text{extra}}}),$$

and fix $D_{\text{extra}} = 3 \times 10^{-3}\text{mm}^2/\text{s}$ as the diffusion of free water at body temperature (37°C).

b) Modeling Non-Gaussian Diffusion

The assumption of Gaussian diffusivity of water molecules within axons has been questioned [76, 12, 72, 14]. In most cases, however, the non-Gaussianity of the diffusion signal has been explained by the presence of an extra-axonal space [76] and therefore relates to the models described above. It remains unclear whether any other source of non-Gaussian diffusivity can be detected with clinical DWI acquisitions once a model of the extra-axonal space has been included. In a recent study, we have argued to the contrary: from DWI acquired in a clinical setting (with $b \leq 3000\text{s/mm}^2$), once the unrestricted diffusion in the extra-axonal space is accounted for, the remaining signal is accurately modeled as a Gaussian function and the whole signal is therefore well represented by a mixture of Gaussian functions [99].

In non-clincial settings, where more acquisitions and higher b-values can be achieved, non-Gaussian models of the intra-axonal diffusion have been proposed [14, 5, 145]. These models posit that water molecules in the axons are restricted in a cylinder rendering a free axial diffusion (parallel to the cylinder axis) and a restricted radial diffusion. The radial diffusivity is modeled following the early developments by Van Gelderen *et al.* [134], Neuman [74] and Codd and Callaghan [29]. In the extra-axonal space, these models use one of the models described above so that only A_{intra} in (2.4) is modified.

Due to their increased complexity, the cylindrical models of diffusion are typically made at the expense of other assumptions on the diffusion signal. The first and obvious assumption of cylindrical models is cylindrical diffusivity, *i.e.* they assume that the diffusion is equal in all directions orthogonal to the principal direction of diffusion. While this assumption may be reasonable, constraints and mechanical tensions during morphogenesis may possibly lead to asymmetrical axons as it is observed for other cells [51]. More importantly, assumptions about the radial diffusivity and the axon diameter are also typically made. For instance, the Composite Hindered And Restricted Model of Diffusion (CHARMED) [14] fixes *a priori* both the distribution of axon diameters and the radial diffusivity for the entire image. As for the cylindrical model of Alexander *et al.* [2, 5], it assumes that all axons in one voxel have the same diameter.

c) Modeling the Heterogeneity of Axon Properties

Axons within a single fascicle have different properties (Fig. 1.4). They differ in shape, orientation and degree of myelination. Different models have recently attempted to take this heterogeneity into account. In the Neurite Orientation Dispersion and Density Imaging (NODDI) model [146], the dispersion of axon orientations is modeled by a Watson distribution of pre-defined sticks to model intra-axonal diffusion. The account for the orientation dispersion thus comes at the expense of losing all information about the fascicle-specific properties (such as diffusion anisotropy, mean diffusivity, radial diffusivity , etc.) since intra-axonal diffusivities are fixed *a priori*.

The AxCaliber model [11] extends the CHARMED model [14] to account for the heterogeneity of axon diameters. A Gamma distribution is fitted to these diameters. This extension adds to the complexity of the CHARMED model and AxCaliber has therefore primarily been designed to describe the diffusion in a two-dimensional plane perpendicular to the axons. This limitation does not allow its use in whole brain population studies.

More recently, the DIstribution of Anisotropic MicrO-Structural eNvironments in Diffusion (DIAMOND) model [98] has been introduced to account for the heterogeneity of microstructural environments at different spatial scales. DIAMOND extends the tensor model by replacing the matrix $\boldsymbol{D}$ by a matrix-variate Gamma distribution over matrices $\boldsymbol{D}$ with a mean $\boldsymbol{D}_0$. The mean of the matrix-variate Gamma distribution remains a tensor but the variability around this mean represents the heterogeneity in microstructural properties. DIAMOND therefore proposes a phenomenological, rather than biophysical, model of the heterogeneity in axonal properties. Doing so enables DIAMOND to simultaneously account for all sources of heterogeneity but requires further assumptions when associating a detected alteration of the diffusion signal to a specific feature of the microstructure [77].

2.3.3 Models of Multiple Fascicles

All the models described above relate to a single fascicle. Due to the limited resolution of DWI, most voxels in the white matter and the grey matter contain multiple fascicles with heterogenous directions [55]. Multi-fascicle models are required to represent the diffusion signal arising from these fascicles. In theory, any single-fascicle model can be extended to a multi-fascicle models by virtue

of the non-exchangeability assumption. The diffusion model becomes:

$$S = S_0 \sum_{k=1}^{K} f_k A_k,$$

where f_k and A_k are the fraction and signal attenuation in the k-th fascicle. The different fascicles present in one voxel are independent so that a full single-fascicle model needs to be estimated for each fascicle.

Owing to the increased number of parameters involved by multiplying the number of fascicles, not all single-fascicle models can be practically generalized to multi-fascicle models. A broad and widely used class of multi-fascicle models is represented by multivariate Gaussian mixtures [59, 102, 105, 132, 137]. In this class, the diffusion signal S is modeled as:

$$S(b, \boldsymbol{g}) = S_0 \sum_{i=1}^{N} f_i e^{-b\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}}. \tag{2.5}$$

In this expression, the number N of compartments may be larger than the number K of fascicles if each fascicle is itself represented by several compartments. For instance, one compartment of hindered diffusion can be included for each fascicle to represent the diffusion of extra-axonal water molecules in the vicinity of the axons. A compartment of isotropic diffusion can also be included to represent water molecules outside fascicles, in a partial volume of CSF or in large glial cells. This multi-fascicle model meets the requirement of providing a full independent model for each fascicle present in the voxel (Fig. 2.3d-f). It enables the separate characterization of fascicle properties and thereby the identification of group differences that affect a specific fascicle and not the fascicles that cross it.

Generalization of the CHARMED model to two fascicles has also been proposed [14] but the distribution of axonal radii is fixed and equal for both fascicles in all voxels so that separate characterization of fascicles is not possible. A multi-fascicle DIAMOND model has also been proposed, replacing matrix-variate Gamma distributions by mixtures of Gamma [98]. This generalization provides a full single-fascicle model of the diffusion signal and accounts for the heterogeneity of microstructural environments in each fascicle independently.

In this thesis, we will focus on the Gaussian mixture models, due to their wide use and their relevance to characterizing independent fascicles. Most of the developed techniques can (and have been [27]) adapted to other types of

models. In Chapter 6, we discuss, in more details, extensions to the multi-fascicle DIAMOND model.

2.4 Population Studies of the Brain Microstructure

In this final section of the chapter dedicated to background material, we explain how features of the diffusion model correlate with the brain microstructure and its alterations. We then describe how population studies of the brain microstructure are conducted.

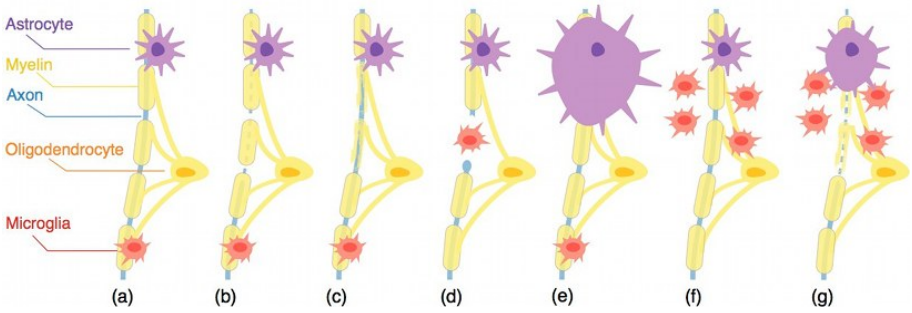


Figure 2.4: Main alterations of the brain microstructure. (a) Healthy myelinated axon in the presence of a normal amount of glial cells, leading to a baseline diffusion model. (b) Axonal injury leading to decreased axial diffusivity. (c) Demyelination resulting in increased radial diffusivity. (d) Axon loss leading to increased fraction of isotropic diffusion and decreased longitudinal diffusivity. (e) Cytotoxic edema leading to increased water molecules within astrocytes. (f) Neuroinflammation leading to increased restricted isotropic diffusion. (g) Combination of axon and myelin injury with neuroinflammatory response leading to a specific alteration of the diffusion mixture model.

2.4.1 Diffusion Signal Correlates of Microstructure Alterations

Analyzing the parameters of the diffusion model of (2.5) provides insight into alterations of the brain microstructure. The brain microstructure is essentially composed of neurons (and their axons and dendrites), glial cells and an extracellular matrix. Three main types of glial cells are present in the brain: (i) *astroctytes* that play different roles including the provision of nutrients to the neurons, the maintenance of extracellular ion balance and the scarring and repair damaged areas, (ii) *oligodendrocytes* that provide a protection and

insulation to axons by creating the myelin sheath, and (iii) *microglia* that play the role of active immune defense by engulfing and digesting cell debris.

These different components are illustrated in Fig. 2.4a in the baseline of a healthy microstructure. In this section, we discuss the most common pathologies of the brain microstructure: axon injury, demyelination, axon loss, edema and neuroinflammation.

a) Axon Injury

Axonal injury (or axonal degeneration) in the absence of myelin degeneration results in water molecules that are no more free to move along the direction of the axon (Fig. 2.4b). This leads to a decrease in the diffusivity λ_1 (corresponding to $\hat{\boldsymbol{e}}_1$ aligned with the axon) of the tensor representing the water molecules within the axon. This diffusivity is known as *axial diffusivity*. The radial diffusivities (λ_2 and λ_3) remain unchanged because water molecules continue to be restricted by the myelin sheath. This correlation between axonal injury and axial diffusivity was proposed and validated in a study on a mouse model of retinal ischemia [115]. This pathology causes an acute inner retinal degeneration whose subsequent demyelination is delayed so that axonal injury can be observed in the absence of demyelination.

b) Demyelination

Demyelination (also called dysmelination, myelin degradation or myelin injury) in the absence of axonal injury results in water molecules in the myelin whose radial diffusivity is less constrained because of the permeability of the damaged myelin (Fig. 2.4c). This leads to increased radial diffusivities (λ_2 and λ_3) of the tensor representing water molecules in the myelin sheath (typically the same tensor as that used to represent the axons). The axial diffusivity λ_1 remains unchanged. This correlation between demyelination and radial diffusivity was proposed and validated in a study of shiverer mice, an animal model of demyelination [116]. Demyelination occurs in various brain diseases including Canavan disease, a genetic degenerative disorder [61]. Demyelination deteriorates the transmission of signals within the brain and may have consequences in sensation, motion and cognition.

c) Axonal Loss

Axonal loss is the complete loss of the whole or part of the axon. It results in an increased fraction of water molecules that are free to diffuse in the extracellular space (Fig. 2.4d). Axonal loss also results in a migration of microglia towards the site of the lesion to engulf cell debris. Axonal loss leads to an increased fraction of isotropic diffusion [137]. Its occurrence is frequent in traumatic brain injury and chronic neurodegenerative diseases, among others [30]. The presence of activated microglial cells around the lesion may lead to a metabolic cascade potentially involving edema and neuroinflammation.

d) Cerebral Edema

Edema are abnormal accumulations of fluid in the extracellular or intracellular space that may result from three separate pathophysiological processes: cytotoxic edema, vasogenic edema and osmotic edema. In cytotoxic edema, water molecules migrate from the extracellular space to accumulate inside cells (neurons or astrocytes) [64]. These water molecules are driven by an influx of cations (including Na^+) due to increased cation permeability of the cell membrane or the failure of active ion-pumps due to energy depletion [133]. The accumulation of water molecules in cytotoxic edema leads to a cell swelling and a corresponding reduction in the amount of water molecules diffusing in the extracellular space (Fig. 2.4e). Cytotoxic edema resulting in cell swelling is more prominent in astrocytes than in neurons [64]. The impact on the diffusion model is an increased volumetric fraction of molecules undergoing restricted anisotropic diffusion in the astrocytes and a corresponding decreased volumetric fraction of water molecules diffusing in the extracellular space. These changes in volumetric fraction may be accompanied by a modification of the parameters related to the diffusion within astrocytes, due to their increased radius following cell swelling. Cytotoxic edema occur, for instance, in ischemic stroke where a lack of oxygen and glucose leads to the failure of ion-pumps [94].

In vasogenic edema, unlike cytotoxic edema, the blood-brain barrier (which is a physiological separation between blood and the extracellular space) is damaged due to injury or autodestructive mediators [133]. This breakdown allows intravascular fluid to penetrate into the extracellular space. Consequently, the volumetric fraction corresponding to the extracellular space in the diffusion model increases. Vasogenic and cytotoxic edema both occur in traumatic brain injury and may be lethal [133]. Vasogenic edema may also result from brain

cancer due to the presence of glioma (cancerous glial cells) whose secretion of specific proteins weakens the blood-brain barrier [50].

Finally, osmotic edema results from osmotic imbalance (that is the difference in concentrations of various solutes or solvents) between the serum and the brain. Osmotic imbalance creates a pressure gradient that results in penetration of water into the brain. As with vasogenic edema, osmotic edema results in an increased fraction of water molecules that diffuse in the extracellular space. Osmotic edema have also been observed in cases of traumatic brain injury [142].

e) Neuroinflammation

Neuroinflammation may refer to the normal migration of microglia to a specific part of the brain to engulf cellular debris or dead cells, in which case it is coined *acute neuroinflammation* [120]. However, abnormal excessive neuroinflammatory response may also occur, in which case the microglial cells exacerbate the neurodestructive process [120]. We refer to this process as *chronic neuroinflammation*. The presence of microglial cells may potentially be detected by a fraction of the signal being unattenuated since the diffusion of water molecules trapped in microglia is highly restricted [5]. When microglial cells are activated (differentiated into macrophages), they release cytotoxic metabolites [17] which, in turn, may lead to cytotoxic edema and cell swelling [64]. The cytotoxic metabolites released by activated microglia include small signaling proteins called cytokines. Interestingly, recent studies have revealed an increased level of cytokines in the plasma of children with autism supporting the hypothesis that autism spectrum disorder may be caused, in part, by an immune dysfunction [10].

f) Detecting and distinguishing microstructural pathologies

The multi-fascicle model of the brain microstructure in (2.5) may distinguish the aforementioned alterations. For instance, an increased radial diffusivity without decreased axial diffusivity points to the presence of demyelination. An increased fraction of isotropic diffusivity points to the presence of a vasogenic or osmotic edema. Axon and myelin alterations can be analyzed separately for each fascicle present in one voxel. This property is of paramount importance in population studies since specific altered fascicles may cross normal unaltered fiber bundles. To fully account for all possible sources of microstructural alterations, a model with sufficient granularity (reflected by the number N of compartments in (2.5))

is required. Typically, one needs to trade-off the granularity of the model with the available data, since higher granularities imply a higher number of parameters to be estimated. A lower model granularity can be compensated for by prior knowledge about the specific pathogenesis based on biopsies or post-mortem histology. For instance, Pastrenak *et al.* suggested the presence of a neuroinflammatory response in schizophrenia onset with a model consisting of a single anisotropic tensor and an isotropic tensor accounting for all water molecules in the extra-axonal space [82].

In etiologically complex diseases and conditions, a combination of microstructure alterations may occur. For instance, multiple sclerosis may be related to an axonal injury and demyelination accompanied with a neuroinflammatory response [137] (Fig. 2.4e). In other diseases, such as schizophrenia and Alzheimer's disease, the neurological basis are still unknown and different hypotheses have been proposed. In Alzheimer's disease, hypotheses of a neuroinflammatory response [138] and hypotheses of a myelin breakdown [19] have been proposed. In schizophrenia, both myelin alterations [60] and neuroinflammation [73] have been proposed.

The neurological basis underpinning functional observations in autism spectrum disorders remains largely unknown. It is yet to be determined whether axons and/or myelin abnormalities are involved or whether ASD results from an autoimmune response leading to neuroinflammation [140, 135], or whether a combination of these causes is involved. Population studies based on multi-fascicle models have the potential to distinguish the different sources of microstructure alterations, thereby providing answers to these critical questions.

2.4.2 Workflow of a Microstructure Population Study

To access the brain microstructure properties in a specific individual, DWI first need to be acquired and a multi-fascicle model needs to be estimated. To compare multi-fascicle models with each other and to detect group differences that point to a specific alteration, multi-fascicle models need to be spatially aligned. This spatial alignment step, also known as *registration*, is critical since misalignment results in an artificial increased variability within the groups and therefore reduces our capacity to detect group differences. By registering all multi-fascicle models to each other, a multi-fascicle atlas can be constructed. This atlas represents the microstructure in a standard average anatomy. It enables the definition of standard fascicles along which microstructural properties

can be defined. Finally, sound statistical analysis need to be performed to characterize group differences. In particular, as explained in the previous section, fascicle properties should be performed on a per-fascicle basis, while comparisons of volumetric properties can be performed on a per-voxel basis.

The four steps of the workflow for population studies of the brain microstructure (acquisition, estimation registration, and statistical analysis) are depicted in Fig. 1.5. In the next chapters, we will address the challenges that arise at each step of this workflow, providing new assets and capabilities that, together, yield a comprehensive and effective framework for population studies of the brain microstructure.

Chapter 3

Registration and Analysis of Multi-Fascicle Models

Multi-fascicle models are of great interest in population studies to characterize and compare properties of the brain microstructure. Central to population studies is the construction of an atlas and the registration of all subjects to it. However, the appropriate definition of registration and atlasing methods for multi-fascicle models have proven challenging. In this chapter, we propose a mathematical framework to register and analyze multi-fascicle models. Specifically, we define novel operators to achieve interpolation, smoothing and averaging of multi-fascicle models. We also define a novel similarity metric to spatially align multi-fascicle models. Our framework enables simultaneous comparisons of different microstructural properties that are confounded in conventional DTI. The framework is validated on multi-fascicle models from 24 healthy subjects and 38 patients with tuberous sclerosis complex, 10 of whom have autism. We demonstrate the use of the multi-fascicle model registration and analysis framework in a population study of autism spectrum disorder.

3.1 Related Work

Conducting population studies based on multi-fascicle models (MFM) requires the alignment of all models to a common coordinate system (the *atlas*). Registering and atlasing multi-tensor images are known to be challenging and many studies attempt to perform population analyses without resorting to them [89, 113, 25, 53]. In [89], a T2-weighted MRI of the subject is registered to a DWI at $b = 0$ s/mm^2 and correspondence between subjects is achieved by segmenting the anatomy based on a T1-weighted atlas. In tract-based spatial statistics (TBSS) [113], single-tensor images are estimated and FA images are used to spatially align subjects. To interpret anisotropies in crossing fiber areas, heuristics based on the mode and FA of the tensor are used. In crossing-fiber TBSS [53], a ball-and-sticks model is estimated but spatial alignment is still

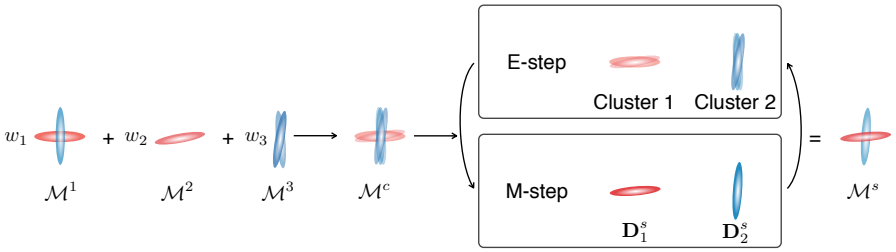


Figure 3.1: Weighted combinations of multi-fascicle models. Computing weighted combinations of multi-fascicle models amounts to computing the complete mixture $\mathcal{M}^c$ and simplifying it in an EM scheme to obtain $\mathcal{M}^s$. The E-step is a clustering problem and the M-step consists in averaging log-tensors in each cluster.

based on single-tensor FA images. None of these approaches attempt to register multi-tensor models directly.

Direct registration of multi-fascicle models is important since the latter provide increased contrast in areas where T2-weighted images and FA images are almost constant (as will be shown in Section 3.7.4). Furthermore, multi-tensor image registration can be made invariant with respect to differences in FA and MD, which is important when those properties need to be compared after alignment. The challenges of registering and analyzing multi-fascicle models stems from difficulties in processing multi-tensors. In particular, interpolating, averaging, smoothing and defining robust similarity metrics for multi-fascicle models cannot be directly extended from the single-tensor case. This is because the j-th tensor in one voxel does not necessarily correspond to the same fascicle as the j-th tensor in another voxel. Furthermore neighboring voxels may contain different number of fascicles.

Interpolation and *spatial smoothing* are required in registration to apply transforms and to prevent aliasing in multi-scale approaches. Building an atlas further requires *averaging* MFM. From a mathematical perspective, interpolating, smoothing and averaging all amount to computing weighted combinations of MFM. In this chapter, we propose a mathematical framework to compute weighted combinations of MFM and a similarity metric to register them. These developments enable registration and analysis of multi-fascicle models which open new opportunities for population studies of microstructural properties. These contributions are based on our recent publications [123, 125, 127].

3.2 Weighted Combinations of Multi-Fascicle Models

Computing weighted combinations of multi-fascicle models is at the basis of interpolation, smoothing and averaging. The linear combination of K mixture models, $\mathcal{M}^k$, with weights w_k, each having N components (with fraction f_j^k and tensor $\boldsymbol{D}_j^k$ leading to a signal S_j^k) results in a mixture with KN components that we call the *complete model*:

$$\mathcal{M}^c = \sum_{k=1}^{K} w_k \mathcal{M}^k = \sum_{k=1}^{K} w_k \sum_{j=1}^{N} f_j^k S_j^k(\boldsymbol{g}) \tag{3.1}$$

$$\equiv \sum_{i=1}^{KN} f_i^c S_i^c(\boldsymbol{g}) = S_0^c \sum_{i=1}^{KN} f_i^c e^{-b\boldsymbol{g}^T \boldsymbol{D}_i^c \boldsymbol{g}}. \tag{3.2}$$

In most practical applications, increasing from N to KN the number of components is not desirable. We therefore estimate a *simplified model*, $\mathcal{M}^s$, with N components which best approximates the complete model:

$$\mathcal{M}^s = \sum_{j=1}^{N} f_j^s S_j^s(\boldsymbol{g}) = S_0^s \sum_{j=1}^{N} f_j^s e^{-b\boldsymbol{g}^T \boldsymbol{D}_j^s \boldsymbol{g}} \tag{3.3}$$

$$= \arg\min_{\mathcal{M}^s} D(\mathcal{M}^c, \mathcal{M}^s), \tag{3.4}$$

where $D(.,.)$ is some discrepancy measure between the complete and simplified models. This problem is known as *mixture model simplification* for which efficient approaches have recently been proposed [18, 35, 125]. In [18], the simplified mixture model is defined as that which minimizes the cumulative differential relative entropy between the complete and simplified models:

$$D\left(\mathcal{M}^c, \mathcal{M}^s\right) \hat{=} \sum_{j=1}^{N} \sum_{i:\pi_i=j} f_i^c D\left(S_i^c(\boldsymbol{g}) \| S_j^s(\boldsymbol{g})\right) \tag{3.5}$$

$$= \sum_{j=1}^{N} \sum_{i:\pi_i=j} f_i^c \int_{\mathbb{R}^3} S_i^c(\boldsymbol{g}) \log \frac{S_i^c(\boldsymbol{g})}{S_j^s(\boldsymbol{g})} d\boldsymbol{g}.$$

The variables π_i cluster the components S_i^c of the complete mixture into N clusters each represented by a single component of the simplified mixture, S_j^s; $\pi_i = j$ means that S_i^c is best represented by S_j^s. Banerjee *et al.* [18] showed that, as long as $S_i^c(\boldsymbol{g})$ and $S_j^s(\boldsymbol{g})$ belong to the exponential family, Equations (3.4) and (3.5) can be optimized in an expectation-maximization scheme for

which both the E-step and the M-step can be solved in closed form (Fig. 3.1). This makes the computation of weighted combinations of multi-fascicle models tractable. Mixtures of distributions from the exponential family is a wide class of mixtures which includes Gaussian mixtures [132], ball-and-sticks models [22], composite hindered and restricted models [14], diffusion directions models [117], Watson and Bingham distributions [146].

In the case of multi-tensor models, the E-step consists in optimizing for the clustering variables π_i assuming S_j^s are known, based on the Burg divergence $B(.,.)$ between covariances $\mathbf{\Sigma} = \boldsymbol{D}^{-1}$:

$$\begin{aligned} \pi_i &= \arg\min_j B\left(\mathbf{\Sigma}_i^c, \mathbf{\Sigma}_j^s\right) \\ &= \arg\min_j \left[\mathrm{Tr}\left(\mathbf{\Sigma}_i^c \mathbf{\Sigma}_j^{s\,-1}\right) - \log\left|\mathbf{\Sigma}_i^c \mathbf{\Sigma}_j^{s\,-1}\right|\right]. \end{aligned} \tag{3.6}$$

The M-step then consists in optimizing the parameters of the simplified mixture (that is $\boldsymbol{D}_j^s$ and f_j^s) to minimize (3.5) providing that we know π_i. Davis and Dhillon [35] proved that this step amounts to computing the weighted average of covariance matrices and fractions in each cluster:

$$\mathbf{\Sigma}_j^s = \frac{\sum_{i:\pi_i=j} f_i^c \mathbf{\Sigma}_i^c}{\sum_{i:\pi_i=j} f_i^c} \qquad \text{and} \qquad f_j^s = \sum_{i:\pi_i=j} f_i^c. \tag{3.7}$$

Alternating the E-step (3.6) and M-step (3.7) until convergence provides the parameters (f_j^s and $\boldsymbol{D}_j^s$) of the weighted combination of mixtures. Initialization is required to control the local minimum to which EM will converge. We initialize the clustering variables π_i by spectral clustering using the cosine similarity matrix between the primary eigenvector $\boldsymbol{e}_i$ of each tensor $\boldsymbol{D}_i^c$ [110]. We found this initialization to be efficient in our experiments. However, since the algorithm typically converges in a few steps, one may consider running it multiple times with various initializations and selecting the result that yields the lowest cumulative differential relative entropy.

One may be concerned about the swelling effect due to averaging covariance matrices in (3.7). This motivates the definition of a log-Euclidean version of the mixture model simplification described above, as it has been defined for single-tensor interpolation [8]. This is achieved by replacing the covariance matrices by their matrix logarithm before performing the EM. The update of the covariance matrices now reads:

$$\log \mathbf{\Sigma}_j^s = \frac{\sum_{i:\pi_i=j} f_i^c \log \mathbf{\Sigma}_i^c}{\sum_{i:\pi_i=j} f_i^c}.$$

Since $\log \boldsymbol{\Sigma} = \log \boldsymbol{D}^{-1} = -\log \boldsymbol{D}$, the logarithmic version of the weighted combination of multi-fascicle models is equivalent to its single-tensor counterpart in voxels with only one tensor. This is not the case in the Euclidean version since covariance matrices, rather than tensors, are averaged. A pseudocode of the method is presented in Algorithm 1.

Importantly, due to the construction of a complete model, the framework described above does not depend on the label i assigned to tensors in the multi-tensor model and accounts for cases where the number of tensors differs between voxels.

Algorithm 1 Weighted Combinations in one voxel

Input: K multi-fascicle models $\mathcal{M}^k$ with weights w_k and the number N of fascicles in the output.
Output: A multi-fascicle model: $\sum_{j=1}^{N} f_j^s e^{-b\boldsymbol{g}^T \boldsymbol{D}_j^s \boldsymbol{g}}$
for k in 1 to K **do** ▷ Construct the complete model $\mathcal{M}^c$
    for j in 1 to N **do**
        $i \leftarrow (k-1)N + j$
        $f_i^c \leftarrow w_k f_j^k$
        $\boldsymbol{D}_i^c \leftarrow \boldsymbol{D}_j^k$
    end for
end for
$\boldsymbol{\pi} \leftarrow$ Initialization$(\{D_i^c, f_i^c\}_{j \leq NK})$ ▷ Initialize clustering
while $\boldsymbol{\pi}$ has not converged **do**
    for j in 1 to N **do** ▷ M-Step
        $\log \boldsymbol{D}_j^s \leftarrow \left(\sum_{i:\pi_i=j} f_i^c \log \boldsymbol{D}_i^c\right) \Big/ \left(\sum_{i:\pi_i=j} f_i^c\right)$
        $f_j^s \leftarrow \sum_{i:\pi_i=j} f_i^c$
    end for
    for i in 1 to KN **do** ▷ E-Step
        for j in 1 to N **do**
            $B_i(j) \leftarrow \mathrm{Tr}\left(\boldsymbol{D}_i^{c\,-1} \boldsymbol{D}_j^s\right) - \log \left|\boldsymbol{D}_i^{c\,-1} \boldsymbol{D}_j^s\right|$
        end for
        $\pi_i \leftarrow \arg\min_l B_i(l)$
    end for
end while

3.3 A Similarity Metric for Multi-Fascicle Models

To register multi-fascicle models, a similarity metric between multi-tensor images needs to be defined. Since registration is used for population studies, the similarity metric must be invariant to inter-subject variability. In particular, since mean diffusivity and fractional anisotropy are typically used as potential biomarkers for diseases, the similarity metric must be invariant to changes in FA and MD. This observation has lead Zhang *et al.* to define a single-tensor similarity metric based on deviatoric tensors, making it invariant to changes in MD [144], though not robust to other differences in diffusivity profiles. In this section, we generalize the correlation coefficient widely used in scalar images when intensities differ between subjects and we show that this similarity metric is invariant under changes in FA and MD.

The correlation coefficient as a similarity metric for block matching is defined as the scalar product between the normalized blocks. For voxels with values in $\mathbb{R}$, the blocks R and S defined over a domain Ω with $|\Omega|$ voxels are elements of $\mathbb{R} \times ... \times \mathbb{R} = \mathbb{R}^{|\Omega|}$ and the correlation coefficient reads:

$$\rho(R,S) = \left\langle \frac{R-\mu_R}{\|R-\mu_R\|}, \frac{S-\mu_S}{\|S-\mu_S\|} \right\rangle, \tag{3.8}$$

where μ is the mean of the image values in the block and $\langle .,. \rangle$ is the canonical scalar product in $\mathbb{R}^{|\Omega|}$. It is invariant if R is replaced by $aR+b$.

For vector images with values in $\mathbb{R}^n$, blocks are elements of $\mathbb{R}^n \times ... \times \mathbb{R}^n = (\mathbb{R}^n)^{|\Omega|}$. The correlation coefficient can be generalized to vector images by redefining the means μ_R and μ_S as the projection of the block onto a block $T \in (\mathbb{R}^n)^{|\Omega|}$ that has a constant value at each voxel, *i.e.* $T(\boldsymbol{x}) = t_0$ for all $\boldsymbol{x} \in \Omega$ [96]:

$$R - \mu_R = R - \frac{\langle R,T \rangle}{\|T\|^2} T. \tag{3.9}$$

The factor $\langle R,T \rangle / \|T\|^2$ is a scalar that we call the *scalar mean* and is equal to μ_R for scalar images. Equation (3.8) is therefore a particular case of (3.9) for $n=1$ and $t_0=1$. This generalized correlation coefficient can be used in any vector space endowed with an inner product. It is invariant if R is replaced by $aR+bT$ where a and b are scalars and T is the chosen constant block.

Let us first generalize the correlation coefficient to single-tensor diffusion images which will prove useful for the generalization to multi-tensor images. Single-tensor blocks are elements of $(\mathcal{S}_3^+)^{|\Omega|}$, where S_3^+ is the space of 3×3

symmetric positive definite matrices. It is typically more convenient to work in the log-tensor space in which blocks are elements of $(\mathcal{S}_3)^{|\Omega|}$. This space is endowed with the Frobenius inner product and the correlation coefficient of (3.9) can be readily applied. Choosing $t_0 = \boldsymbol{I}_3$ (the identity matrix), the correlation coefficient is invariant under linear transformations of the log-tensor eigenvalues: $\log \lambda_i \rightarrow \log \lambda_i' = a \log \lambda_i + \log b$ due to the invariance $\log \boldsymbol{D} \rightarrow \log \boldsymbol{D}' = a \log \boldsymbol{D} + \log b \boldsymbol{I}_3$. It is instructive to observe what the definition of the scalar mean becomes in this space with the Frobenius inner product. We have:

$$\begin{aligned}
\mu^{\mathcal{S}_3} &= \frac{\langle \log \boldsymbol{D}, \boldsymbol{T} \rangle}{\|\boldsymbol{T}\|^2} && (3.10)\\
&= \frac{\sum_{\boldsymbol{x}\in\Omega} \langle \log \boldsymbol{D}(\boldsymbol{x}), \boldsymbol{I}_3 \rangle_{\mathrm{F}}}{\sum_{\boldsymbol{x}\in\Omega} \langle \boldsymbol{I}_3, \boldsymbol{I}_3 \rangle_{\mathrm{F}}} && (3.11)\\
&= \log \left(\prod_{\boldsymbol{x}\in\Omega} \lambda_1(\boldsymbol{x}) \lambda_2(\boldsymbol{x}) \lambda_3(\boldsymbol{x}) \right)^{\frac{1}{3|\Omega|}} && (3.12)\\
&\hat{=} \log \left(\lambda_\Omega^G \right). && (3.13)
\end{aligned}$$

The generalized scalar mean for blocks of single-tensors is therefore the logarithm of the geometric mean λ_Ω^G of diffusivities over the domain Ω.

Defining a scalar product in the space $(\mathcal{S}_\mathcal{M})^{|\Omega|}$ of blocks of multi-tensors seems impractical if not impossible. We further generalize the correlation coefficient (3.9) by substituting the inner product $\langle ., . \rangle$, by a more general scalar mapping: $m(.,.) : \mathcal{S}^{|\Omega|} \times \mathcal{S}^{|\Omega|} \rightarrow \mathbb{R}$ for any space $\mathcal{S}^{|\Omega|}$. The generalized correlation coefficient becomes:

$$\rho(R, S) = m\left(\frac{R - m(R,T)T}{n_m(R - m(R,T)T)}, \frac{S - m(S,T)T}{n_m(S - m(S,T)T)} \right),$$

where $n_m(X)^2 = m(X, X)$ is a generalization of the norm, and T is assumed normalized ($n_m(T) = 1$).

This expression does not guarantee the invariance of the generalized correlation coefficient (GCC) with respect to linear changes of the blocks: $R \rightarrow R' = aR + bT$. Furthermore, in order to remain interpretable, the GCC must be symmetric, equal to one in case of perfect match and lower than one in any

other case:

$$\begin{aligned} \rho(aR + bT, S) &= \rho(R, S) && (3.14)\\ \rho(R, S) &= \rho(S, R) && (3.15)\\ \rho(R, R) &= 1 && (3.16)\\ |\rho(R, S)| &\leq 1. && (3.17) \end{aligned}$$

These constraints on ρ impose constraints on the scalar mapping m. One can show that constraints (3.14-3.17) are satisfied if the following constraints are respected by m (the proof is provided in Appendix A):

$$\begin{aligned} m(R + bT, T) &= m(R, T) + b\, m(T, T) && (3.18)\\ m(aR, S) &= a\, m(R, S) && (3.19)\\ m(R, S) &= m(S, R) && (3.20)\\ |m(R, S)| &\leq n_m(R) n_m(S). && (3.21) \end{aligned}$$

The latter generalizes the Cauchy-Schwartz inequality. Being a scalar product is a sufficient but unnecessary condition to respect these constraints. Therefore, constraints (3.18-3.21) as well as the choice of a constant block T and suitable basic operations (to define the multiplication by a scalar and the addition of the constant block T), stand together as a model to generate correlation coefficients in potentially any space, even when an inner product cannot be defined.

In the case of multi-fascicle models, we further want the similarity metric between two multi-tensor blocks $\mathcal{M}^1(\boldsymbol{x}) = \sum_{i=1}^{N} f_i^1(\boldsymbol{x}) e^{-b\boldsymbol{g}^T \boldsymbol{D}_i^1(\boldsymbol{x})\boldsymbol{g}}$ and $\mathcal{M}^2(\boldsymbol{x}) = \sum_{i=1}^{N} f_i^2(\boldsymbol{x}) e^{-b\boldsymbol{g}^T \boldsymbol{D}_i^2(\boldsymbol{x})\boldsymbol{g}}$ to be equal to the single-tensor similarity metric if the blocks contain only one tensor in each voxel. This can be achieved if the scalar mapping is equal to the Frobenius inner product when all but one fractions are equal to zero. We therefore add a fifth constraint on the scalar mapping:

$$\begin{aligned} &\text{If } f_j^1 = 1, f_{i\neq j}^1 = 0, f_k^2 = 1, f_{i\neq k}^2 = 0,\\ &\Rightarrow m(\mathcal{M}^1, \mathcal{M}^2) = \sum_{\boldsymbol{x}\in\Omega} \left\langle \log \boldsymbol{D}_j^1(\boldsymbol{x}), \log \boldsymbol{D}_k^2(\boldsymbol{x}) \right\rangle_{\mathrm{F}}. \end{aligned} \quad (3.22)$$

We define the multiplication of multi-tensors by a scalar a as the multiplication of all log-tensors by a and the addition of the constant block T as the addition of t_0 to all log-tensors. These definitions naturally generalize the single-tensor case. A generalized scalar mapping m comes by computing pairwise scalar

products between corresponding tensors. This requires to pair tensors between the two blocks at each voxel. We introduce the following notation:

$$d(p, \boldsymbol{x}) = \sum_{i=1}^{N} f_i^1(\boldsymbol{x}) f_{p(i)}^2(\boldsymbol{x}) \left\langle \log \boldsymbol{D}_i^1(\boldsymbol{x}), \log \boldsymbol{D}_{p(i)}^2(\boldsymbol{x}) \right\rangle,$$

where p is the pairing function which associates one and only one tensor of $\mathcal{M}^1$ to one and only one tensor of $\mathcal{M}^2$. For $N-$fascicle models, there are $N!$ such pairings. We define the scalar mapping for multi-tensor images as:

$$m(\mathcal{M}^1, \mathcal{M}^2) = \sum_{\boldsymbol{x} \in \Omega} d\Big(\arg\max_{p} \big| d(p, \boldsymbol{x}) \big|, \boldsymbol{x} \Big). \tag{3.23}$$

In practice, the values of $d(p, \boldsymbol{x})$ for all $N!$ pairings p are computed and we select the one with the highest absolute value. This scalar mapping satisfies (3.18-3.22). The absolute value is required by Condition (3.22) for cases where the Frobenius inner product between the tensors is negative. To better interpret this generalized scalar mapping, it is instructive to assess how it generalizes the concept of scalar means and norm to multi-fascicle models. The generalized scalar mean is given by:

$$\begin{aligned}
\mu^{\mathcal{S}_{\mathcal{M}}} &= \frac{m(\mathcal{M}, T)}{m(T, T)} \\
&= \frac{\sum_{\boldsymbol{x} \in \Omega} \sum_{i=1}^{N} f_i(\boldsymbol{x}) \frac{1}{N} \left\langle \log \boldsymbol{D}_i(\boldsymbol{x}), \boldsymbol{I}_3 \right\rangle_{\mathrm{F}}}{\sum_{\boldsymbol{x} \in \Omega} \sum_{i=1}^{N} \frac{1}{N^2} \left\langle \boldsymbol{I}_3, \boldsymbol{I}_3 \right\rangle_{\mathrm{F}}} \\
&= \log \left(\prod_{\boldsymbol{x} \in \Omega} \prod_{i=1}^{N} \left(\lambda_1^{(i)}(\boldsymbol{x}) \lambda_2^{(i)}(\boldsymbol{x}) \lambda_3^{(i)}(\boldsymbol{x}) \right)^{\frac{f_i(\boldsymbol{x})}{3|\Omega|}} \right).
\end{aligned} \tag{3.24}$$

Remarkably, the generalized scalar mean for multi-fascicle model is the geometric mean of the diffusivities within the block for which all fascicles contribute in a ratio that is equal to their volumetric fraction f_i in their voxel. As for the generalized norm of multi-fascicle models, it is given by:

$$n_m^2(\mathcal{M}) = m(\mathcal{M}, \mathcal{M}) = \sum_{\boldsymbol{x} \in \Omega} \sum_{i=1}^{N} f_i(\boldsymbol{x})^2 \| \log \boldsymbol{D}_i(\boldsymbol{x}) \|_{\mathrm{F}}^2, \tag{3.25}$$

that is the sum of the Frobenius norms of each log-tensor, weighted by the squared fractions. To demonstrate the latter expression, we need to show that the absolute value of d in (3.23) is maximized if the pairing p pairs a fascicle

(in $\mathcal{M}_1 = \mathcal{M}$) with itself (in $\mathcal{M}_2 = \mathcal{M}$). The proof is straightforward using the Cauchy-Schwartz inequality on elements $X = (f_1 \log \boldsymbol{D}_1, ..., f_N \log \boldsymbol{D}_N)$ and $Y_p = (f_{p(1)} \log \boldsymbol{D}_{p(1)}, ..., f_{p(N)} \log \boldsymbol{D}_{p(N)})$. Both the generalized scalar mean and the generalized norm therefore have direct interpretations in terms of multi-fascicle models.

More importantly, because the generalized correlation coefficient is invariant under linear transformations ($\rho(aR+bT, S) = \rho(R, S)$), the proposed generalized scalar mapping for multi-fascicle models leads to a GCC that is invariant under linear transformations of the eigenvalues of each log-tensor. Indeed, for a block R of multi-log-tensors whose k-th eigenvalue of the j-th log-tensor is $\log \lambda_k^{(j)}$, $aR + bT$ contains log-tensors with eigenvalues equal to $a \log \lambda_k^{(j)} + b$. In other words, ρ for multi-fascicle models is invariant under the following transformations ($k = 1, 2, 3$):

$$\log \lambda_k^{(j)} \rightarrow a \log \lambda_k^{(j)} + \log b, \quad \forall a \in \mathbb{R} \text{ and } \forall b \in \mathbb{R}^+,$$

or, equivalently,

$$\lambda_k^{(j)} \rightarrow b \left(\lambda_k^{(j)}\right)^a, \quad \forall a \in \mathbb{R} \text{ and } \forall b \in \mathbb{R}^+, \tag{3.26}$$

for any fascicle j. In particular, this invariance property encompasses differences in mean diffusivity (MD) for unchanged FA if $a \neq 1$ and $b = 1$. Similarly, changes in FA with unchanged MD can be obtained by varying a and b in a specific manner. Indeed, MD is preserved under changes of the eigenvalues following Equation (3.26), if $b(\lambda_1^a + \lambda_2^a + \lambda_3^a) = \lambda_1 + \lambda_2 + \lambda_3$. For any given set of eigenvalues and any given a, there exists a b that satisfies this relation. One can therefore fix a to match the desired FA and subsequently fix b to respect this MD-preserving relation (since b does not affect the FA). Finally, by varying both a and b in an unconstrained manner, various changes in MD and FA can be accounted for by the invariance property of the GCC.

This similarity metric therefore allows registration of subjects with locally different diffusivity profiles. Importantly, because of the presence of the fractions f^1 and f^2 in the scalar mapping, the GCC accounts for cases where the number of tensors is different in different voxels (the corresponding fraction will simply be set to zero).

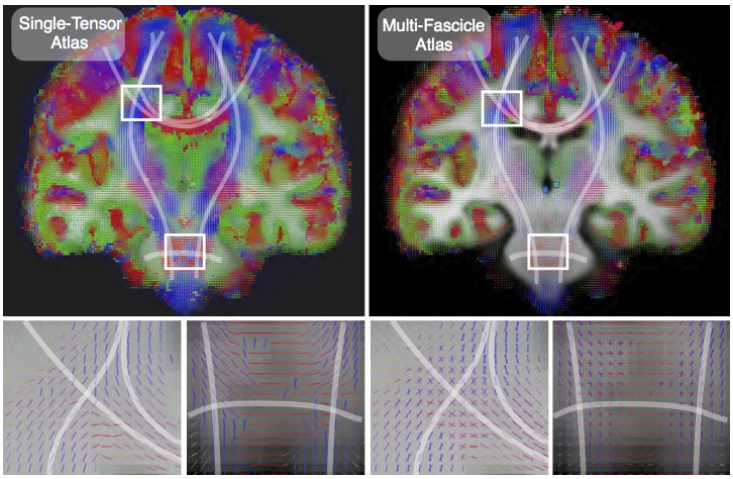


Figure 3.2: Atlas of the brain microstructure Single-tensor and multi-fascicle atlases overlaid on the T1-weighted MRI atlas. The multi-fascicle atlas presents tensors with higher fractional anisotropies than the single-tensor atlas. This is due to the account of both the free water diffusion in the isotropic compartment and the multiple fascicle present in the voxel. The highlighted regions represent the corona radiata where projections of the corpus callosum cross corticospinal tracts, and a region where the pyramidal tracts (vertical lines) and the medial cerebellar peduncle (horizontal lines) cross.

3.4 Registration and Atlasing

With the framework developed in the previous section, we can now perform population studies by constructing a multi-fascicle atlas and registering all subjects to it.

Multi-fascicle models are estimated in the coordinate system of a T1-weighted MRI of the same subject (see Section 3.7.1). The registration between multi-fascicle models is initialized by affine registration of the T1-weighted MRI using the Baladin method [78], yielding a transformation T^0.

The weighted combinations of multi-fascicle models and the GCC are introduced in a robust multi-scale block matching registration algorithm developed in [32]. A dense deformation field is estimated through the following steps:

- For each pyramid level $p = 1, ..., P$
    - For each iteration $i = 1, ..., N$

- Estimate sparse pairings C between R and $F \circ T^{i-1}$ by block-matching
- Interpolate a dense correction field δT^i from C using a Gaussian kernel and weighted by the confidence in the matches as in [41].
- Reject a fixed amount of outliers from C based on their dissimilarity with the estimated δT^i
- Estimate an outlier-free correction $\delta \tilde{T}^i$
- Compose the correction $\delta \tilde{T}^i$ with the current estimate of the transform $T^i = T^{i-1} \circ \delta \tilde{T}^i$
- Apply elastic regularization to the field T^i

In our implementation, $P = 4$, $N = 10$, block sizes are $5 \times 5 \times 5$, and the outlier removal rate is 20%. The weighted combinations of multi-fascicle models are used to interpolate multi-tensor images when applying the deformation or constructing the multi-scale representation of the image. When warping tensor images (and hence multi-tensor images), tensors need to be reoriented. This reorientation is performed using the finite-strain rationale [6].

Registration is then used iteratively to build an atlas based on the method developed in [45]. This method essentially alternates between aligning and averaging images. To average multi-tensor images, we use the weighted combination of multi-fascicle models described above. Ten iterations are used to build the final atlas. The resulting atlas for single-tensor and multi-tensor images are depicted in Fig. 3.2.

Based on the registration and atlasing methods, we can employ our novel operators including interpolation and averaging of multi-fascicle models to perform different statistical analyses of the brain microstructural properties.

3.5 Fascicle-Based Spatial Statistics

With all subjects aligned to the multi-fascicle atlas, we can compare properties of the aligned tracts through *fascicle-based spatial statistics* (FBSS) (Fig. 3.3). Tractography is performed once on the atlas using the multi-fascicle tractography method described in [83, 63, 24] and adapted to include the multi-fascicle interpolation. For each registered subject, the tensor most aligned with the tract is selected and its property of interest (FA, MD, etc.) is computed. This provides, for each subject, a vector of length n (the number of points on the tract), representing the microstructural property along the fascicle.

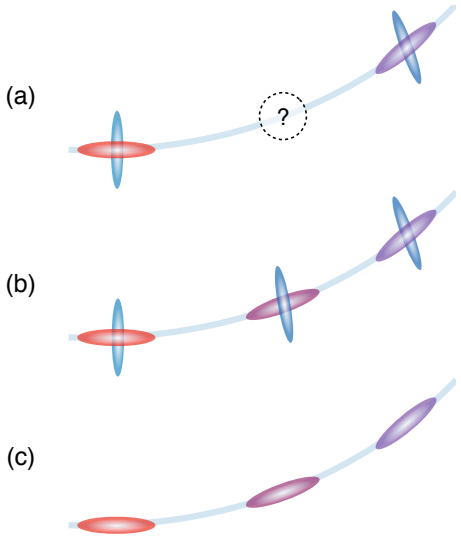


Figure 3.3: Fascicle-based spatial statistics (FBSS) proceeds in three steps. (a) Fascicles (blue line) are drawn on the atlas with a sub-voxel resolution. The point in the middle is at a non-grid location. (b) Multi-fascicle models are interpolated at non-grid locations. (c) At each location along the fascicle, the tensor most aligned to the fascicle is selected to compute the property of interest (FA, MD, etc.).

Point-by-point t-tests are carried out along the tract to compare its properties between the two groups. This yields a vector $\boldsymbol{t}$ of n t-scores. Since the smoothness of the tract property depends on the individual, the tract and the resolution of the tractography, we use a non-parametric correction for multiple comparisons based on cluster-based statistics [75]. This method assumes that differences along the tract occur in clusters of adjacent points and proceeds as follows:

1. Define a threshold t_0 on the t-statistics.
2. Define a binary vector $\boldsymbol{b}$ of supra-threshold t-statistics:

$$\boldsymbol{b} = (\boldsymbol{t} > t_0).$$

3. Detect the connected components $C = \{c_i\}$ in $\boldsymbol{b}$.
4. For each connected component c_i, compute its size s_i (number of points) and its mass m_i (sum of t-scores).

5. Randomly permute N_p times the subjects in the groups (*i.e.* randomly reassign subjects to either groups) and perform Steps 1-4 for each permutation. For each permutation k, record the maximum size s_p^k and the maximum mass m_p^k among the detected clusters.

6. The recorded s_p^k and m_p^k describe the null distributions of the size and mass of the clusters. Corrected for multiple comparisons, the p-values of each connected component c_i testing for its likelihood to be due to chance alone are:

$$p_i^s = \frac{1}{N_p+1}\left(1+\sum_{k=1}^{N_p}\delta(s_p^k > s_i)\right)$$
$$\text{and} \quad p_i^m = \frac{1}{N_p+1}\left(1+\sum_{k=1}^{N_p}\delta(m_p^k > m_i)\right)$$

for the size statistics and the mass statistics, respectively.

With FBSS, local fascicle segments where the two groups significantly differ can be discovered. Several $t-$thresholds t_0 are typically used to assess the robustness of the findings. Higher t_0 yield smaller clusters of stronger differences.

3.6 Isotropic Diffusion Analysis

Large isotropic fraction f_{iso} indicates an excessive extracellular volume [80] which is in turn a surrogate for the presence of edema or neuroinflammation [82]. Isotropic diffusion analysis (IDA), *i.e.* the statistical analysis of the isotropic fraction, is thus of strong interest for population studies of disease involving these pathologies. The isotropic fraction is non-Gaussian since it ranges between 0 and 1. We apply the *logit* transform to f_{iso} prior to computing t-tests. This transform brings the distribution of f_{iso} closer to normality. Specifically, we transform f_{iso}-maps into l_{iso}-maps where:

$$l_{\text{iso}} = \text{logit}(f_{\text{iso}}) = \log\left(\frac{f_{\text{iso}}}{1-f_{\text{iso}}}\right).$$

To prevent l_{iso} to take on infinite values when $f_{\text{iso}} = 0$ or 1, we bound the latter within $[10^{-6}, 1-10^{-6}]$. We then carry cluster-based statistics on the l_{iso}-maps with the cluster size and cluster masses as quantities of interest, as described in [75].

3.7 Validation Experiments

In this section, we validate the presented framework for multi-fascicle models. We systematically compare our results with those obtained when multi-tensor images are seen as a stack of single-tensor images on which multi-channel approaches to registration and averaging can be applied.

3.7.1 In vivo data

In vivo DWI were acquired on a Siemens 3T Trio scanner with a 32 channel head coil using the CUSP-45 gradient sequence [102]. This sequence includes 30 diffusion-encoding gradients on a shell at $b = 1000\text{s/mm}^2$ and 15 extra gradients in the enclosing cube of constant TE with b-values up to 3000s/mm^2. Eddy current distortion was minimized using a twice-refocused spin echo sequence [90]. Other acquisition parameters were set to FOV= 220mm, matrix= 128×128, number of slices=68, resolution = $1.7 \times 1.7 \times 2\text{mm}^3$. Data acquisition was conducted using a protocol approved by the Institutional Review Board (IRB). The DW images were aligned to the $1{\times}1{\times}1\text{mm}^3$ T1-weighted MRI with rigid registration (using the mean $b = 0$ image as a moving image) and the gradients were reoriented appropriately. This compensates for patient head motion and for residual geometric distortions due to magnetic field inhomogeneity and eddy current.

A multi-fascicle model with three tensors including an isotropic compartment were estimated as in [102]. Images were acquired for 24 healthy controls and 38 patients with tuberous sclerosis complex (TSC): 10 diagnosed with autism (TSC+ASD), 17 diagnosed without (TSC-ASD) and 11 too young for diagnosis (see details in Appendix B).

3.7.2 Relabeling Invariance Study

Multi-fascicle models assign arbitrary labels i to tensors in Equation (2.5). The framework must therefore be invariant under relabeling of tensors. In this experiment, we randomly relabeled tensors 10 times for each of the 24 healthy controls and performed registration between the result and the original image. Using the proposed framework, the deformation fields obtained were exactly the identity. By contrast, using the multi-channel registration, a significantly non-zero deformation field resulted from the registration of relabeled multi-fascicle models (Fig. 3.4). This result demonstrates the failure of multi-channel

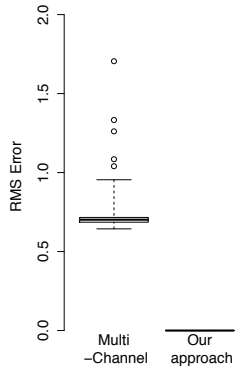


Figure 3.4: Results of the relabeling invariance study Average RMS error of the deformation fields obtained by registering a multi-fascicle model with itself after randomly relabeling tensors. Our framework is invariant under relabeling leading to an error that is exactly zero. By contrast, multi-channel registration yields non-zero deformation fields.

registration for multi-fascicle models. In what follows, tensors are labeled based on their FA ($\boldsymbol{D}_1$ has the highest FA and $\boldsymbol{D}_N$ the lowest) to allow a fair comparison between the two approaches.

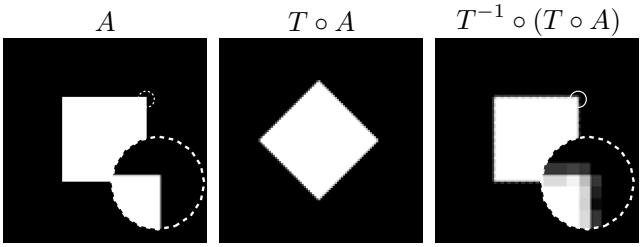


Figure 3.5: Toy example of the $T^{-1} \circ T$ experiment to estimate the interpolation error. An original image A is transformed with T and the result is transformed backward with T^{-1}. The end result, $T^{-1} \circ (T \circ A)$, is equal to the original image corrupted by interpolation errors and can therefore be used to estimate the level of interpolation errors.

3.7.3 $T^{-1} \circ T$ Study: Assessment of the Interpolation Error

For any transformation T, the composition $T^{-1} \circ T$ is equal to the identity. Therefore, for any image A, $T^{-1} \circ T \circ A = A$. However, if we first compute

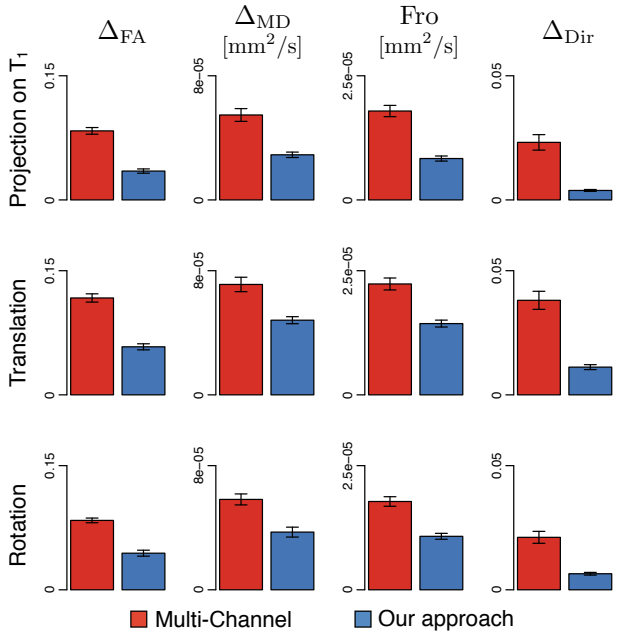


Figure 3.6: The interpolation error is significantly smaller when the proposed approach is used. Compared to the multi-channel alternative, the proposed approach to compute linear combinations of multi-fascicle models is significantly better for all four metrics and under all three transformations.

$(T \circ A)$ and then apply T^{-1} to the result, we do not obtain exactly the original image A due to interpolation error. Comparing the result $\tilde{A} = T^{-1} \circ (T \circ A)$ to the original A thus provides estimates of the interpolation error, independently from the similarity metric (Fig. 3.5).

We investigated the residual error of $T^{-1} \circ T \circ A$ to compare the interpolation error of the proposed approach for linear combinations and the multi-channel alternative. The experiment was conducted with three different affine transformations T that were applied to the multi-fascicle models of the 24 healthy controls: (1) a transformation that maps the DWI to the T1-weighted image, (2) a translation by half a voxel in all directions, (3) a rotation of 45 degrees around the vertical axis.

The result $\tilde{A}$ is compared to the original multi-fascicle model A in terms of

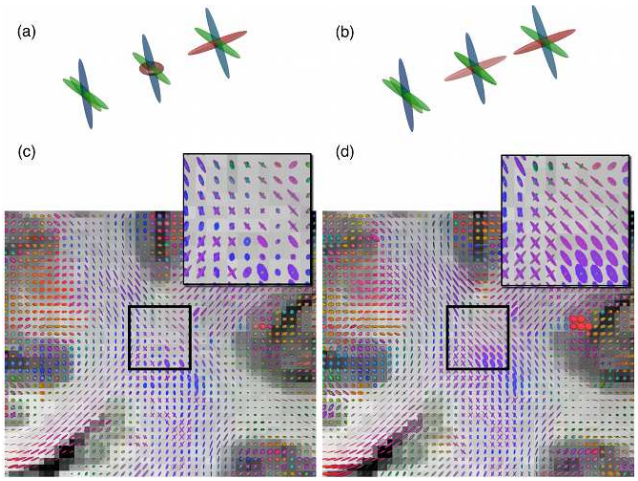


Figure 3.7: Examples of interpolation results obtained with the multi-channel approach (left) and the proposed approach (right). Performing interpolation for each tensor independently confounds fascicles resulting in an inflated result. Weighted combination of multi-fascicle models introduced in our mathematical framework clusters similar fascicles to avoid the inflation effect. (a-b) Results obtained on synthetic data by interpolating the multi-fascicle models at the extremities. (c-d) Results obtained on *in vivo* data by applying a linear transform to a multi-fascicle model.

the following four similarity metric computed at each voxel:

$$\Delta^2_{\mathrm{FA}} = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} (\mathrm{FA}(\boldsymbol{D}_i) - \mathrm{FA}(\tilde{\boldsymbol{D}}_{\boldsymbol{i}}))^2 \tag{3.27}$$

$$\Delta^2_{\mathrm{MD}} = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} (\mathrm{MD}(\boldsymbol{D}_i) - \mathrm{MD}(\tilde{\boldsymbol{D}}_{\boldsymbol{i}}))^2 \tag{3.28}$$

$$\mathrm{Fro}^2 = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \|\boldsymbol{D}_i - \tilde{\boldsymbol{D}}_{\boldsymbol{i}}\|^2_{\mathrm{F}} \tag{3.29}$$

$$\Delta_{\mathrm{Dir}} = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \left(1 - |\boldsymbol{e}_{\boldsymbol{1,i}}.\tilde{\boldsymbol{e}}_{\boldsymbol{1,i}}|\right), \tag{3.30}$$

where $\boldsymbol{e}_{1,i}$ is the principal eigenvector of tensor $\boldsymbol{D}_i$ with unit norm. The last equation assesses how aligned the resulting tensors are to the original tensors.

The results, summarized in Fig. 3.6, demonstrate that the use of the proposed

method decreases the interpolation error as compared to the multi-channel alternative, for all similarity metrics and for all three transformations. On average, this decrease varies in magnitude from 38% for Δ_{MD} to 73% for Δ_{Dir}. One-tailed paired t-tests indicate that the decreases are significant in all cases ($p < 10^{-12}$). An example of interpolation obtained with both methods is depicted in Fig. 3.7 and presents a portion of the corona radiata where fascicles cross. In this region, the multi-channel approach confounds the fascicles and fails to interpolate the multi-fascicle model.

3.7.4 Scan-Rescan Study: Evaluation of the Similarity Metric

In this section, we independently assess the accuracy of the similarity metric. We exploited two sets of 45 DWI acquired on the same subject during the same scanning session. The subject was required not to move and remained still throughout the acquisition. The two sets of DWI are thus intrinsically aligned. A multi-fascicle model as well as a single-tensor DTI were estimated from each set. The two multi-fascicle models differ due to acquisition noise, artifacts, and estimation errors. This scan-rescan experiment therefore provides a unique opportunity to estimate the accuracy of the proposed similarity metric in a realistic scenario.

A total of 495 landmarks were defined on a regular grid within the first image (Fig. 3.8(a)). Landmarks were spaced 7 voxels apart in all directions. Blocks of size $5 \times 5 \times 5$ were defined around each landmark and correspondence were sought in a neighborhood of size $21 \times 21 \times 21$ in the second image. Since there is no transformation between the two images, the true correspondence $\boldsymbol{x}_{\mathrm{true}}$ is located at the center of the neighborhood. The accuracy of the best match $\boldsymbol{x}_{\mathrm{max}}$ (that maximizes the similarity $C(\boldsymbol{x})$) and the saliency of the true match are:

$$\begin{aligned} \text{Accuracy} &= \|\boldsymbol{x}_{\mathrm{max}} - \boldsymbol{x}_{\mathrm{true}}\| \\ \text{Saliency} &= \frac{C(\boldsymbol{x}_{\mathrm{true}}) - \bar{C}}{\sigma_c}, \end{aligned} \tag{3.31}$$

where $\bar{C}$ and σ_c are the mean and standard deviation of the similarity metric within the neighborhood.

Results for these two indices were compared amongst four different metric (1) the correlation coefficient applied to FA images (C_{FA}), (2) the correlation coefficient generalized to single-tensor DTI (C_{DTI}), (3) the multi-channel correlation

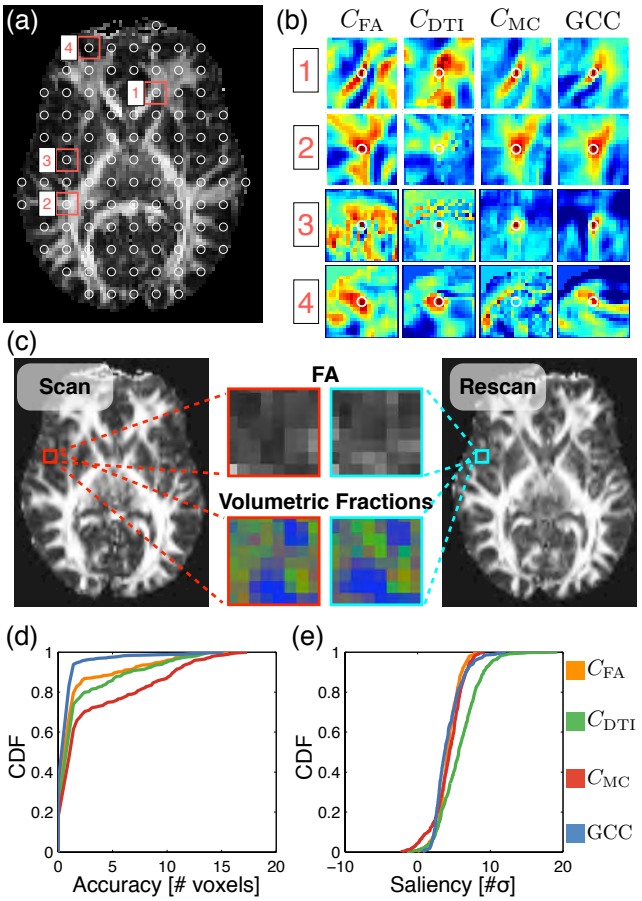


Figure 3.8: Evaluation of the generalized correlation coefficient (GCC) as a similarity metric for multi-fascicle models. The GCC for multi-fascicle models outperforms other metrics in terms of registration accuracy, as assessed by a scan-rescan experiment. (a) 495 regularly spaced landmarks are used for the experiment. (b) Similarity maps in four neighborhoods (circles indicate true matches) showing that GCC is the most specific metric. (c) In regions with no contrast in FA, the GCC is able to find correct matches due to robust patterns observed in multi-fascicle models. (d) The accuracy of the GCC is significantly better than all other metrics as seen by the cumulative distribution function (CDF). (e) No significant difference in saliency between the metrics are observed, except for a significantly higher saliency with C_{DTI}.

coefficient applied to multi-tensors (C_{MC}) and the GCC for multi-fascicle models (GCC). Fig. 3.8(b) depicts the similarity maps for four different neighborhoods with each metric.

Neighborhood 1 in Fig. 3.8(b) illustrates the case of a specific white matter landmark, for which all four metrics perform equally well. Neighborhood 2 illustrates a case of a specific white matter landmark located at the intersection of crossing fascicles. In this case, both metrics based on multi-fascicle models find the correct match. Matching based on FA has more spurious maxima and matching based on DTI is offset because the single-tensor is a poor model of the diffusion signal in this region. Finally, Neighborhoods 3 and 4 show landmarks located at the boundary between the white and grey matter. In this area the microstructure is more complex. Multi-fascicle models are required in these regions to find a correct match. The multi-channel metric fails to detect the correct correspondence if tensors are not properly paired (see Neighborhood 4).

Fig. 3.8(c) depicts the volumetric fraction (shown as a color image) in areas where the FA displays no contrast. The fractions show a clear pattern of alternation between isotropic diffusion (blue) and single (green) or multi-fascicle (brown) orientation. These patterns are repeated in both the scan and the rescan and therefore enable accurate matching.

Quantitatively, the GCC for multi-fascicle models significantly outperforms all other metrics in terms of accuracy (one-tailed paired t-test: $p < 10^{-8}$), as depicted in Fig. 3.8(d) and summarized in Table 3.1. The average gain in accuracy C_{FA} is 45%. Importantly, the probability for the accuracy to be lower or equal to 1, that is the fraction of landmarks for which the best match was found in the direct neighborhood of the true match, is $P(\text{Accuracy} \leq 1) > 80\%$ for the GCC while it is lower than 70% for all other metrics. These results suggest that the remaining registration error would likely be eliminated by regularization and outlier removal in the registration algorithm. No significant

Table 3.1: Summary statistics of the target registration error in the scan-rescan study

Metric	**Mean**	**St. dev.**	**P(Accuracy $\leq$ 1)**
C_{FA}	1.79	2.79	67.3%
C_{DTI}	2.18	3.05	59.4%
C_{MC}	3.34	4.18	47.5%
GCC	**0.98**	**1.66**	**83.4%**

difference was observed in terms of saliency except for a significantly larger saliency for C_{DTI} which may partially counterbalance its poorer accuracy in registration.

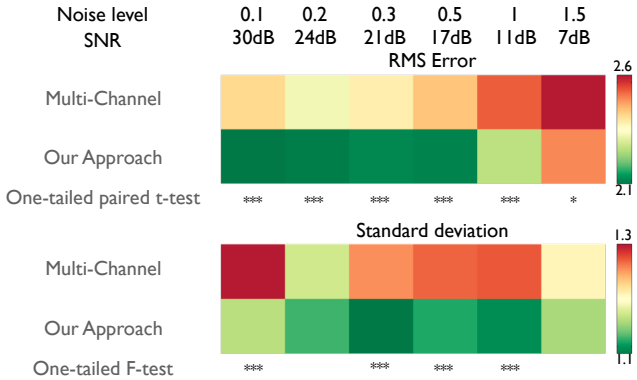


Figure 3.9: Our mathematical framework leads to higher registration accuracies than the multi-channel alternative. The RMS errors (top) and its variance (bottom) are significantly lower. Results are shown for 1,440 registrations performed at various SNR for synthetic deformation fields ($\star: p < 0.05$, $\star\star: p < 0.005$, $\star\star\star: p < 0.001$).

3.7.5 Synthetic Fields Study

In this experiment, we compare the registration accuracies when synthetic deformation fields are applied to multi-fascicle models. Ten random log-Euclidean polyaffine deformation fields are generated by drawing parameters from a Gaussian with zero mean and 0.05 standard deviation for 27 regularly spaced affine components (this results in a field with a mean magnitude of 6.9 voxels and a maximum magnitude of 42 voxels) [122]. Each of the ten deformations are applied to the 24 multi-fascicle models of the control subjects. Symmetric matrices of Gaussian noise with zero mean and standard deviation at six different levels (0.1, 0.2, 0.3; and 0.5, 1.0, 1.5) were then added to the log of all tensors in both the original and the transformed image, corresponding to SNR of (30dB, 24dB, 21dB, 17dB, 11dB, 7dB). The original and the transformed images were then registered and the resulting deformation field was compared to the initial synthetic field in term of its root mean squared (RMS) error. All 1,440 registrations (24 subjects $\times$ 10 deformation fields $\times$ 6 noise levels) were

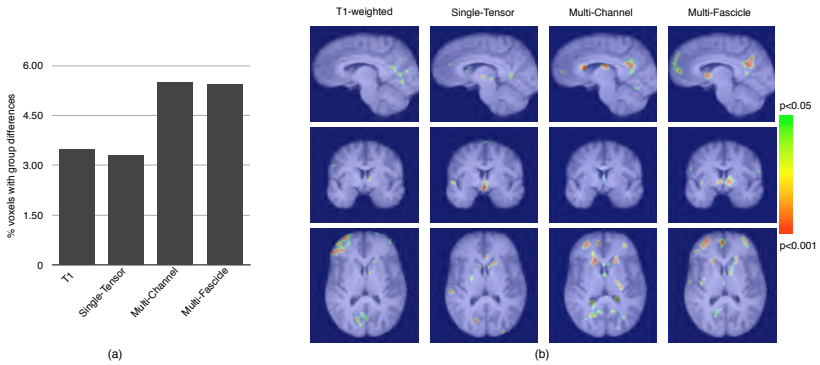


Figure 3.10: Morphometry results show areas with a significant volume deficit within the grey and white matter of TSC patients. (a) Multi-fascicle registration and multi-channel registration reveal more differences than single-tensor and the T1-weighted registrations. (b) The differences observed with multi-fascicle registration are more consistent with the known anatomy than those observed with multi-channel registration, as seen for example in the left and right internal capsules.

performed with both the proposed framework and the multi-channel alternative.

On average, the root mean squared (RMS) error of the deformation field is 17% higher when the multi-channel registration is used instead of the proposed framework. A one-tailed paired t-test on the RMS for the transformation at each SNR shows that the difference in RMS between the approaches is significant: $p < 10^{-5}$ for all SNR between 11dB and 30dB and $p = 0.02$ for SNR=7dB (Fig. 3.9-*Top*). The variance of the RMS is also decreased by 45% on average. A one-tailed F-test reveals that this decrease is significant for all SNR ($p < 0.001$) except for SNR=24dB and 7dB (Fig. 3.9-*Bottom*). This experiment indicates that even when tensors are labeled based of their FA, the proposed framework outperforms the multi-channel alternative.

3.7.6 Morphometric Contrast Study

The deformation field obtained by registering a subject to an atlas provides a measurement of the local morphometric difference between the subject and a standard anatomy. The determinant of the Jacobian $|J|$ of the deformation fields at every voxel provides information about the amount of local volume differences ($|J| < 1$ indicates a decreased volume and $|J| > 1$ indicates an increased volume).

Widespread volume deficits in the white and grey matter of patients with TSC have been previously reported [91]. However, the amount of differences detected depends on the accuracy of the registration because the statistical power of the test depends on the registration accuracy [43]. Because of the increased level of microstructure they represent, we expect multi-fascicle models to reveal more morphometric differences than single-tensor models and scalar T1-weighted MRI. To assess the statistical power of all modalities (T1, single-tensor DTI and multi-fascicle models), we used the common voxel-based morphometry method [9]: register all subjects to the atlas, compute the log-jacobian, smooth it by a kernel of 8mm FWHM and correct for family-wise error rate at $p = 0.05$.

Results in Fig. 3.10(a) show that multi-fascicle models reveal more differences than single-tensors and T1-weighted MRI, as expected. This is likely due to an increased statistical power resulting from a higher registration accuracy when the structure of the white matter is better represented. The number of significant voxels does not differ between the proposed approach and the multi-channel alternative. However, the spatial distribution of the volume deficit findings (Fig. 3.10(b)) better follows the anatomy than the multi-channel alternative as seen, for example, in the left and right internal capsules.

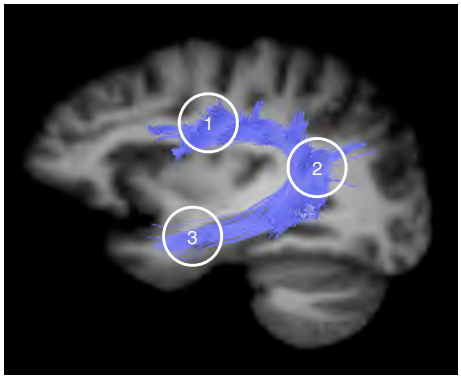


Figure 3.11: The dorsal language circuit is a set of white matter pathways involved in language. The median tract was manually selected from those tracts to perform fascicle-based spatial statistics (FBSS). Region 1 is the white matter in the vicinity of Broca's territory. Region 2 is the white matter in the vicinity of Geschwind's territory. Region 3 is the white matter in the vicinity of Wernicke's territory.

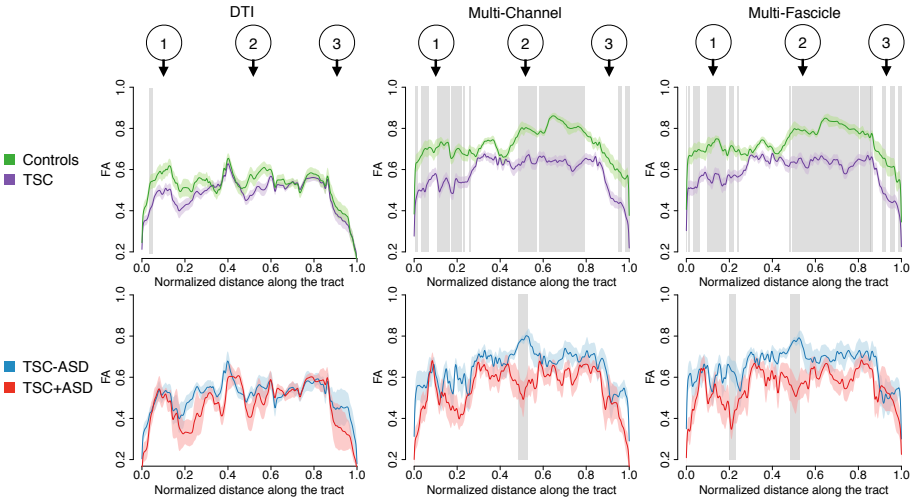


Figure 3.12: FBSS of multi-fascicle models reveal local differences in the dorsal language circuit that single tensor DTI fails to detect. Curves show the mean FA along the median tract of the dorsal language circuit in each group. Shaded area along the curves represent two standard errors. Grey rectangles indicate that the FA in that cluster is significantly different between the two groups. The top row studies differences between patients with tuberous sclerosis complex (TSC) and healthy controls. The bottom row further investigates differences between TSC patients with (TSC+ASD) and without autism (TSC-ASD). Landmarks 1, 2 and 3 correspond to those in Fig. 3.11.

3.8 Applications to Population Studies

The designed framework enables population studies of various properties of the brain microstructure. In this section, we propose two such population studies related to fascicle properties in the dorsal language circuit and the isotropic diffusion in the white matter.

3.8.1 FBSS of the Dorsal Language Circuit

FBSS can detect local abnormalities in white matter pathways, which helps defining foci of neurological disorders. The dorsal language circuit is a set of white matter fascicles involved in language (Fig. 3.11). In this section, we investigate whether local decreases in FA along the dorsal language circuit can be discovered by FBSS.

Tractography of the dorsal language circuit was performed using the automatic seeding method of [121, 63]. A representative tract that captures the geometry of the bundle was manually selected. One-tailed fascicle-based spatial statistics was first performed between the 38 patients with TSC and the 24 healthy controls to test whether TSC patients have lower FA along the tract than healthy controls (Fig. 3.12-top row). Results found with both the multi-fascicle approach and the multi-channel approach consistently show that differences between TSC patients and healthy controls are widespread over the tract. This is consistent with recent models of tuberous sclerosis complex presented as a widespread decreased white matter microstructural integrity [83] and a global loss of connectivity [85]. Analysis based on single-tensor images did not reveal significant differences between the groups (except for a small cluster near the dorsal end of the tract). This is probably due to DTI being unable to distinguish the signal arising from each fascicle (one of them generating the group difference) and from free diffusion.

One-tailed fascicle-based spatial statistics was also performed between TSC+ASD and TSC-ASD patients to further understand the impact of autism on the properties of fascicles in the language system (Fig. 3.12-bottom row). A cluster of significantly lower FA was found in the middle of the tract, *i.e.* in the white matter close to the Geschwind's territory, a region that has previously been associated with the interpretation of facial emotions [88]. Furthermore, using the proposed framework for multi-fascicle registration and analysis, a second cluster of significantly lower FA was found in the white matter close to Broca's area, a cortical region associated with speech production whose activity was shown to be impaired in patients with autism spectrum disorder [33]. Again, no local difference was observed based on single-tensor images.

Findings of lower FA in TSC+ASD compared to TSC-ASD were previously reported in the literature [63, 83]. However, for the first time, our framework enables the detection of local differences, improving our knowledge of alterations in the brain microstructure related to autism spectrum disorder.

3.8.2 Isotropic Diffusion Analysis in Autism

Isotropic diffusion analysis allows whole-brain inspection of differences in isotropic fraction f_{iso} whose excess relates to the presence of neuroinflammation and edema among others. To investigate *in vivo* whether autism spectrum disorder may result from a neuroinflammatory response (as suggested by post-

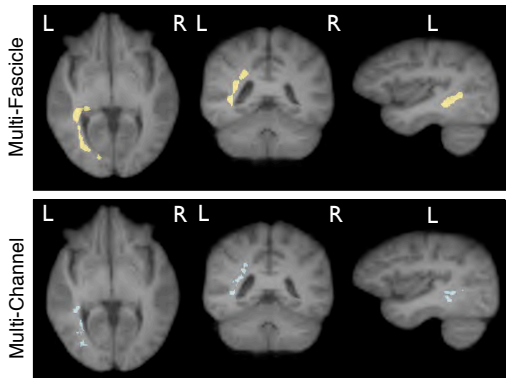


Figure 3.13: Multi-fascicle models reveal clusters of increased isotropic fraction in autism, potentially indicating the presence of neuroinflammation. Clusters found with our framework (top) are larger and more coherent than those obtained with the multi-channel alternative (results shown for t_0=3).

mortem studies [135]), we performed isotropic diffusion analysis to compare the TSC+ASD and TSC-ASD groups. Cluster-based statistics was performed at four different thresholds: $t_0 = 2$, 2.5, 3, and 3.5 to assess the robustness of the findings with respect to the threshold used.

Consistently for all thresholds, clusters of significantly higher f_{iso} were detected in patients with autism (Fig. 3.13). Both the size and mass of these clusters show significant departure from the null distribution ($p < 0.05$, Table 3.2). The multi-channel approach also found significant clusters but these were smaller in size and more sensitive to the choice of threshold (no significant cluster was found for $t_0 = 3.5$). The location of the significant cluster detected with both methods coincide and corresponds to part of the visual system (Fig. 3.13).

These findings are consistent with recent studies of autism in children which have demonstrated that appropriate maturation of visual system is crucial for social cognition development [58]. Furthermore, while autism is believed to potentially result from a neuroinflammatory process, *in vivo* evidence of such neurological mechanism are missing. These results illustrate how the proposed techniques for the analysis of multi-fascicle models can provide new insights into the brain microstructure. The validation of the neuroinflammatory process in autism would however require further studies including more subjects and other imaging modalities (such as PET imaging and T2 mapping).

Table 3.2: Size and mass statistics of the significant clusters from isotropic diffusion analysis.

Threshold	Cluster	Cluster Size				Cluster Mass			
		Multi-Channel		Multi-Fascicle		Multi-Channel		Multi-Fascicle	
		Value	p-value	Value	p-value	Value	p-value	Value	p-value
2	Cluster 1	22408	0.046	**30927**	0.05	58457	0.047	**82224**	0.047
2.5	Cluster 1	5126	0.039	14034	0.028	15766	0.043	43581	0.028
	Cluster 2	7448	0.025	-	-	22875	0.026	-	-
	Total	12574		**14034**		38641		**43581**	
3	Cluster 1	1837	0.035	6270	0.017	6503	0.038	22219	0.017
	Cluster 2	2004	0.029	-	-	6983	0.034	-	-
	Total	3841		**6270**		13486		**22219**	
3.5	Cluster 1	-	-	**1864**	0.02	-	-	**7667**	0.021

3.9 Summary

Diffusion tensor imaging confounds the diffusion signal arising from different compartments and may therefore not be reliable for population studies of the brain microstructure. In particular, studies based on DTI cannot separate differences in properties of the fascicles due to demyelination or axonal injury, from differences in extracellular volume fraction due to neuroinflammation, edema or partial voluming with CSF. By representing the signal arising from different compartments with distinct parameterizations, multi-fascicle models are able to explain the origin of the observed differences. This property makes multi-fascicle models of great interest for population studies of the brain microstructure.

The cornerstone of image-based population studies is the construction of an atlas and the registration of all subjects to it. In this chapter, we introduced a framework for registration and atlasing of multi-fascicle models. A mixture model simplification method was introduced to compute weighted combinations of multi-fascicle models, as used for interpolation, smoothing and averaging. As a similarity metric, a generalized correlation coefficient was developed to be invariant under linear transformations of the eigenvalues of each fascicle in the log-domain, making it robust to inter-subject variability.

Once all subjects are aligned to the atlas, population studies can be carried out to investigate microstructural properties in brain diseases. We introduce a system of two statistical analyses of the brain microstructure: fascicle-based spatial statistics (FBSS) and isotropic diffusion analysis (IDA). The former allows discoveries of local differences in the microstructural properties of the fascicle in a specific pathway. The latter allows detection of differences in extracellular volume fraction which may relate to neuroinflammation and edema. Together, these analyses allow for comprehensive investigation of the brain microstructure. We illustrated its use in a population study of autism spectrum disorder related to tuberous sclerosis complex and showed that the use of multi-fascicle models in this context increases the sensitivity of the statistical tests.

Chapter 4

Selection of the Appropriate Model

Chapter 2 presented various models that overcome the limitations of the diffusion tensor model. The development of these models is an active field of research and we urgently need a rationale to select between these models. This selection requires to account for the data at hand and the most appropriate model may depend on the amount and kind of available data.

The selection can be performed at the level of model category, where one wants to select between different classes of models such as multi-tensor models [102], CHARMED [14], NODDI [146], DIAMOND [98], and combinations of different restricted, hindered and unrestricted environments [79]. The selection may also operate at the level of model granularity, where the number of fascicles or compartments needs to be determined.

The identification of the appropriate model category and granularity has proven challenging. In this context, different model selection approaches have been proposed. Most approaches attempt to maximize the quality of fit while penalizing complex models to avoid overfitting. However, the choice of a penalization strategy and the trade-off between penalization and quality of fit are rather arbitrary and produce highly variable results. In this chapter, we propose to select the appropriate model at each voxel by assessing the generalization error. This criterion naturally prevents overfitting by comparing how the models predict new data not used to estimate the model. Since the generalization error cannot be directly computed, we propose to estimate it by the 632 bootstrap technique which has low bias and low variance. Results on synthetic phantoms and *in vivo* data show that our approach performs better than existing techniques, and is robust to the choice of a decision threshold.

4.1 Related Work

Assessing the appropriate diffusion model at each voxel is a model selection problem. To date, most approaches select between diffusion models of different

complexity by minimizing the fitting error. Because both the model estimation and assessment is achieved on the same dataset, this strategy favors complex models and may overfit the data. For this reason, the criterion usually integrates a component penalizing complex models.

The most common criterion for the selection of diffusion models is the F-test. Alexander *et al.* [4] compared the spherical harmonic expansion of the average diffusion coefficient (ADC) truncated at different orders by means of a series of ANOVA F-Tests. In this strategy, complex models are penalized by their necessity to significantly decrease the fitting error when compared to simpler models. Kreher *et al.* [59] used F-tests to select the appropriate number of fascicles by observing the variance of the ADC. Scherrer and Warfield [102] used a similar F-test strategy applied to the signal residuals rather than the ADC.

Besides the F-test, other approaches based on the quality of fit have been proposed. Behrens *et al.* [21] used a Bayesian Automatic Relevance Determination (ARD) approach which starts with the most complex model and gradually prune the unnecessary variables. However, this was shown inefficient in tractography and required to manually force the number of fascicles [68]. A recent Bayesian approach set a sparse prior on the volumetric fraction f_i in a maximum a posteriori estimation of the model to automatically select the number of fascicles [147]. Neither of these approaches can be generalized to the selection amongst a broad range of non nested models of different expressions (*e.g.* to select between DIAMOND and a multi-fascicle model) since their formulation implies setting priors on parameters of the same nature (*e.g.* the volumetric fractions). The Bayesian Information Criterion (BIC), a weighted sum of the fitting error and a penalizing term, has been suggested as well but was shown to yield suboptimal results, even on synthetic data [105].

A more reliable and general paradigm to avoid overfitting when selecting between models is to compare how each model performs for *new data* not included in the model estimation. This relates to the *generalization error*. Typically, a model not complex enough to represent a dataset will have a large generalization error, and so will too complex a model which overfits the data. Minimization of the generalization error has never been used to determine the number of fascicles or, more generally, the complexity of the diffusion model at each voxel. Leave-one-out cross-validation follows this paradigm. However, it does not lead to a consistent estimate of the model [109] and its results are highly variable. Other cross-validation methods, such as K-fold cross-validation, reduce this variance at the expense of a higher bias. By contrast, the 632 bootstrap

method proposed by Efron [36] reduces this variability while remaining almost unbiased.

In this chapter, we use, for the first time, the 632 bootstrap (B632) method to determine the number of fascicles present in each voxel from diffusion-weighted images. We first introduce how the generalization error can be estimated (Section 4.2) and we then illustrate how it can be used to select the number of fascicles (Section 4.3).

4.2 Estimation of the Generalization Error

In this section, we present the generalization error minimization framework. We start by explaining the fundamental difference between generalization error and fitting error. We then introduce different methods to estimate the generalization error, and explain why the 632 bootstrap should be used in this context. We subsequently provide an expression to estimate the standard error of the estimation of the generalization error.

4.2.1 Generalization Error and Fitting Error

Let $\boldsymbol{z} = \{z_1, ..., z_n\}$ with $z_i = (x_i, y_i)$ be the set of n observed data points, in which x_i are inputs to the model (*e.g.* the b-values b and gradient directions $\boldsymbol{g}$ in diffusion images) and y_i are outputs (*e.g.* the signal attenuation in diffusion images). These data are used to build a generative model $r_{\boldsymbol{z}}(x)$ that attempts to predict the output y from an input x. This prediction rule is based on a finite set $\boldsymbol{z}$ of observations and the hope is that the learnt model $r_{\boldsymbol{z}}(x)$ generalizes well to new data. Ideally, the optimal model would thus minimize the *generalization error*, that is the error made on a new hypothetical data point $\mathsf{z}_0 = (\mathsf{x}_0, \mathsf{y}_0)$. The generalization error conditional on the observed data is :

$$E_g|\boldsymbol{z} = E_{\mathsf{z}_0 \sim F}\left[|y_0 - r_{\boldsymbol{z}}(x_0)|^2 \big| \boldsymbol{z}\right], \tag{4.1}$$

where $E[.]$ is the statistical expectation and $\mathsf{z}_0 \sim F$ indicates that the expectation is taken over the new data point that follows the (unknown) distribution F. To account for the variability of the observed data points, the unconditional generalization error can be defined as the expectation of (4.1) over all possible values of z :

$$E_{g,n} = E_{\mathsf{z}_i \overset{\text{iid}}{\sim} F}\Big\{E_g|\boldsymbol{z}\Big\} = E_{\mathsf{z}_i \overset{\text{iid}}{\sim} F}\Big\{E_{\mathsf{z}_0 \sim F}\left[|y_0 - r_{\boldsymbol{z}}(x_0)|^2 \big| \boldsymbol{z}\right]\Big\}, \tag{4.2}$$

where the index n indicates that n samples were used to optimize the model $r_{\boldsymbol{z}}$. The generalization errors (4.1) and (4.2) cannot be directly computed because the distribution F is unknown. One simple solution would be to estimate $F(z)$ by the empirical distribution $\hat{F}(z) = \frac{1}{n}$ for all $z \in \boldsymbol{z}$. For the conditional generalization error (4.1), this yields the following estimate:

$$\begin{aligned} \hat{E}_g^{\text{fit}} &= E_{\mathrm{z}_0 \sim \hat{F}} \left[|y_0 - r_{\boldsymbol{z}}(x_0)|^2 \middle| \boldsymbol{z} \right] \\ &= \frac{1}{n} \sum_{i=1}^{n} |y_i - r_{\boldsymbol{z}}(x_i)|^2, \end{aligned} \tag{4.3}$$

that is the common *fitting error*. This estimate is a biased estimate of E_g since the data $\boldsymbol{z}$ are used both to optimize the parameters of the model $r_{\boldsymbol{z}}$ and to estimate its error. In particular, in many modeling problem (including multi-fascicle modeling), it is always possible to find a model that yields $\hat{E}_g^{\text{fit}} = 0$ provided that it is complex enough. In the following sections, we will therefore explore other estimates of E_g and comment on their bias and variance.

4.2.2 Cross-Validation Estimates

To circumvent the overfitting problem of $\hat{E}_g^{\text{fit}}$, one could estimate the model by omitting one data point in the training sample $\boldsymbol{z}$ and evaluating the model prediction for this data point. This is the idea behind the leave-one-out cross-validation (LOOCV) method. Let $\boldsymbol{z}_{|i}$ be the training samples without z_i. The resulting estimate of the generalization error reads:

$$\hat{E}_g^{\text{CV}} = \frac{1}{n} \sum_{i=1}^{n} |y_i - r_{\boldsymbol{z}_{|i}}(x_i)|^2. \tag{4.4}$$

This is an unbiased estimator of the generalization error $E_{g,n-1}$. For large n, the bias of $\hat{E}_g^{\text{CV}}$ as an estimator of $E_{g,n}$ is positive but low. Its variance, however, is large, leading to high root mean squared errors, despite the low bias [36].

The variance of the LOOCV can be decreased by keeping more than one element out of the dataset at each iteration of model training and testing. This approach, called K-fold cross-validation, results in unbiased estimates of $E_{g,n-\frac{n}{K}}$ which, however, present an increased bias for the estimation of $E_{g,n}$.

4.2.3 632 Bootstrap

The bootstrap smoothing method can be used to lower the variance of the cross-validation estimate [36]. This technique estimates E_g in (4.2) by providing two different estimates for the distribution F. The point z_0 is drawn from a distribution F approximated by the empirical distribution $\hat{F}$. The distribution F of $\boldsymbol{z}$ is then estimated from bootstrap samples $\boldsymbol{z}^*_{|i}$ of the empirical distribution $\hat{F}_{|i}$ which excludes the sample z_i drawn for the testing point z_0. Formally, an expression of the estimator comes by inverting the order of the expectations in (4.2) and by subsequently replacing the distributions F by their estimates:

$$\hat{E}_g^{\mathrm{BS}} = \frac{1}{n}\sum_{i=1}^{n} E_{\hat{F}_{|i}}\left[|y_i - r_{\boldsymbol{z}^*_{|i}}(x_i)|^2\right].$$

An equivalent expression of $\hat{E}_g^{\mathrm{BS}}$ that is closer to its implementation is:

$$\hat{E}_g^{\mathrm{BS}} = \sum_{i=1}^{n}\left[\sum_{b=1}^{B}\delta(N_i^b)\left|y_i - r_{\boldsymbol{z}^*_{|i}}(x_i)\right|^2 \Big/ \sum_{b=1}^{B}\delta(N_i^b)\right], \tag{4.5}$$

where N_i^b is the number of times sample i is used in the training set of the b^{th} bootstrap replicate and $\delta(x)$ is the Dirac function. The factor $\delta(N_i^b)$ guarantees that sample i can be used as a testing sample in bootstrap replicate b.

Much like cross-validation, the bootstrap estimate is biased because it relies on fewer point than the number n of available samples. LOOCV uses $(n-1)$ points and its bias is therefore limited. By contrast, $\hat{E}_g^{\mathrm{BS}}$ uses, on average, $[1-(1-\frac{1}{n})^n]n$ points which is approximately equal to $0.632n$ for large n. This makes the bias of $\hat{E}_g^{\mathrm{BS}}$ more critical. Efron [36] proposed to counterbalance the positive bias of $\hat{E}_g^{\mathrm{BS}}$ by the negative bias of $\hat{E}_g^{\mathrm{fit}}$, introducing the 632 bootstrap estimator:

$$\hat{E}_g^{632} = 0.368\,\hat{E}_g^{\mathrm{fit}} + 0.632\,\hat{E}_g^{\mathrm{BS}}. \tag{4.6}$$

The coefficients are heuristically defined so that the testing samples used to estimate $\hat{E}_g^{632}$ are at the same average distance from the training sample as would be a random point drawn directly from F [36]. This estimator has outperformed others in many applications, mostly when the signal-to-noise ratio is low [70].

4.2.4 Standard Error of the Difference Estimator

If we had access to the generalization error for each model, its direct comparison would enable to select the model with the lowest generalization error. The 632

bootstrap expression is, however, a noisy estimate of the generalization error. A lower 632 bootstrap estimate may be obtained for one model by chance alone. This effect is especially concerning when models are nested (simpler models are particular cases of more complex ones) since more complex models can be arbitrarily close to simpler ones (for instance, a 3-fascicle model with one volumetric fraction close to zero is very close to a 2-fascicle model). To reliably select between the two models, we need to assess whether any improvement in the generalization error is statistically significant.

Assessing the statistical significance of the difference in generalization error estimates between a model A and a model B, $\hat{\Delta}^{632}_{AB} = \hat{E}^{632}_{g,A} - \hat{E}^{632}_{g,B}$, requires the standard error of this difference to be estimated. After rearranging the terms of $\hat{\Delta}^{632}_{AB}$, taking advantage of the linearity of the expectation, we have:

$$\begin{aligned}
\hat{\Delta}^{632}_{AB} &= \frac{0.368}{n}\sum_{i=1}^{n} E_{\mathbf{z}_i \overset{\text{iid}}{\sim} \hat{F}}\left[\left|y_i - r^A_{\boldsymbol{z}}(x_i)\right|^2 - \left|y_i - r^B_{\boldsymbol{z}}(x_i)\right|^2\right] \\
&+ \frac{0.632}{n}\sum_{i=1}^{n} E_{\mathbf{z}_i \overset{\text{iid}}{\sim} \hat{F}_{|i}}\left[\left|y_i - r^A_{\boldsymbol{z}^*_{|i}}(x_i)\right|^2 - \left|y_i - r^B_{\boldsymbol{z}^*_{|i}}(x_i)\right|^2\right] \\
&\triangleq \frac{0.368}{n}\sum_{i=1}^{n} \hat{\Delta}^{\text{fit}}_{AB,i} + \frac{0.632}{n}\sum_{i=1}^{n} \hat{\Delta}^{\text{BS}}_{AB,i} \\
&\triangleq 0.368\,\hat{\Delta}^{\text{fit}}_{AB} + 0.632\,\hat{\Delta}^{\text{BS}}_{AB} \qquad (4.7)
\end{aligned}$$

One could estimate the standard error of $\hat{\Delta}^{\text{BS}}_{AB}$ as $[\sum_i (\hat{\Delta}^{\text{BS}}_{AB,i} - \hat{\Delta}^{\text{BS}}_{AB})^2/n^2]^{1/2}$. This would assume that the $\hat{\Delta}^{\text{BS}}_{AB,i}$ are independent, which is not the case. A better estimate can be obtained by the *delta-method-after-bootstrap* approach [36]. This method is nonparametric and allows the computation of the standard error for any statistics that (1) is smooth in the observed data $\boldsymbol{z}$, (2) is invariant under permutations of the points z_i and, (3) only depends on the empirical distribution $\hat{F}$. With this method, one can show that the standard error of $\hat{\Delta}^{\text{BS}}_{AB}$ can be estimated by:

$$\hat{SE}^{\text{BS}} = \left[\sum_{i=1}^{n} \hat{D}_i^2\right]^{1/2} \text{ with } \quad \hat{D}_i = \left(2 + \frac{1}{n-1}\right)\frac{\hat{\Delta}^{\text{BS}}_{AB,i} - \hat{\Delta}^{\text{BS}}_{AB}}{n} + \frac{\sum_{b=1}^{B}(N_i^b - \bar{N}_i)\bar{q}^b}{\sum_{b=1}^{B}\delta(N_i^b)}$$

$$\text{and } \quad \bar{q}^b = \sum_{i=1}^{n} \delta(N_i^b)\left[\left|y_i - r^A_{\boldsymbol{z}^*_{|i}}(x_i)\right|^2 - \left|y_i - r^B_{\boldsymbol{z}^*_{|i}}(x_i)\right|^2\right],$$

where $\bar{N}_i$ is the average N_i^b over all B bootstrap replicates. The same approach cannot be used for the standard error of $\hat{\Delta}^{\text{fit}}_{AB}$ because it is a non-smooth

function of the samples $\boldsymbol{z}$. Efron [36] proposes to estimate the standard error of $\hat{\Delta}_{AB}^{632}$ as:

$$\hat{SE}^{632} \approx \frac{\hat{\Delta}_{AB}^{632}}{\hat{\Delta}_{AB}^{\mathrm{BS}}} \hat{SE}^{\mathrm{BS}}. \tag{4.8}$$

To infer whether a model A is better than a model B, the estimate of the difference between their generalization errors (4.7) can be compared to the estimate of the standard error of this difference (4.8). This is the cornerstone of the selection of the number of fascicles problem.

4.3 Application: Estimation of the Number of Fascicles

In this section, we illustrate how the 632 bootstrap estimation of the generalization error can be used to identify the number of fascicles present in each voxel. We validate the approach on synthetic and *in vivo* data.

4.3.1 Selection of the Number of Fascicles

In this section, we provide a method to select the number N of fascicles in the multi-fascicle model by assessing the generalization error:

$$S(b, \boldsymbol{g}) = \sum_{i=1}^{N} f_i S_i(b, \boldsymbol{g}).$$

The main idea to identify the number of fascicles at each voxel is to progressively increase the complexity of the model as long as a substantial decrease in the generalization error can be achieved and to stop when the decrease is no more significant or when the generalization error starts to increase. More specifically, the steps for the selection of the number of fascicles are:

1. For each pair of consecutive models (model with $m-1$ fascicles and m fascicles), compute the difference of generalization error estimates using (4.7). Let $\Delta_m = \hat{\Delta}_{m-1,m}^{632}$ be this difference. To use expression (4.5) for this estimate, the bootstrap replicates must be identical for all models.

2. Compute the standard error s_m of the estimate Δ_m using (4.8).

3. Select the model with m_{opt} fascicles such that:

$$m_{\mathrm{opt}} = \inf \left\{ m \middle| \Delta_m - \theta^{\mathrm{B632}} s_m \geq 0, \Delta_{m+1} - \theta^{\mathrm{B632}} s_{m+1} < 0 \right\}, \tag{4.9}$$

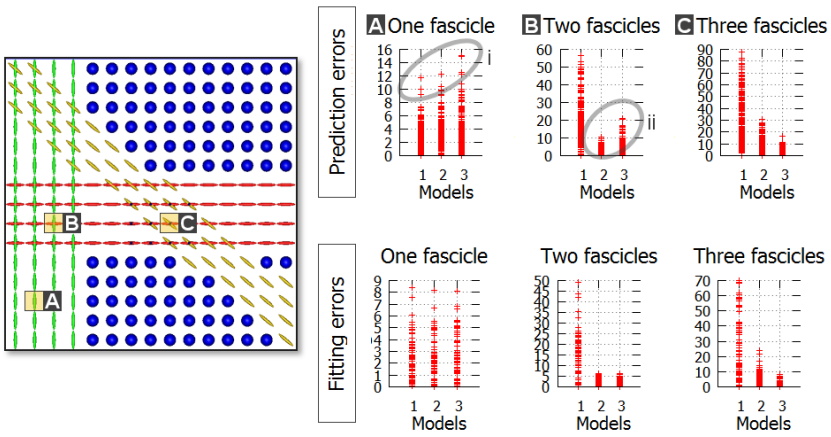


Figure 4.1: Results of the synthetic phantom experiment show that estimating the generalization error allows reliable selection of the number of fascicles. (Left) Synthetic phantom used in our experiment, containing isotropic areas (blue balls), 1-fascicle areas [A], 2-fascicle areas [B] and three fascicle areas [C]. (Right) The repartition of the prediction errors, as used in the 632 bootstrap estimator, discriminates the three models, since more complex models may have higher prediction errors. On the contrary, the fitting error, as used in the F-test model selection, always decreases when the model complexity increases, due to overfitting.

where θ^{B632} is the number of standard errors above which the difference should be to be deemed significant. For this expression to hold, we set $s_0 = \Delta_0 = \Delta_M = 0$ and $s_M = 1$, where M is the maximum number of fascicles that we authorize for our models.

We could compare every model to every other one and extend the selection rule (4.9) to express the need for a model to be significantly better than all the simpler ones and not significantly worse than the more complex ones. In our experiment, we did not observe a difference between the two rules. Rule (4.9) has the advantage that it can be applied as the model complexity is progressively increased, avoiding the need to optimize further complex models at voxels where a simple model has already been selected.

4.3.2 Experimental Setup

Our proposed approach for the selection of the number of fascicles could be used with any model of the diffusion signal. In this work, we considered a

multi-fascicle model in which each fascicle is represented by a tensor and the diffusion in the extra-axonal space is represented by an isotropic tensor. This amounts to considering the following model for the formation of the diffusion signal S for a b-value b and a gradient direction $\boldsymbol{g}$:

$$S = S_0 \left(f_0 e^{-bD_{\text{iso}}} + \sum_{i=1}^{m} f_i e^{-b\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}} \right),$$

where $D_{\text{iso}} = 3.0 \times 10^{-3}\text{mm}^2/s$ is the diffusion of free water in the brain at 37^oC and m is the number of fascicles of the model. Such model requires a DWI acquisition that images multiple non-zero b-values [102, 126]. We employed the CUSP gradient encoding scheme [102] composed of five $b = 0$ s/mm^2, 30 DWI at $b = 1000$ s/mm^2 and 30 DWI with gradients on the enclosing cube with b-values between $b = 1000$ s/mm^2 and $b = 3000$ s/mm^2. The parameters of each model were estimated using a maximum a posteriori approach (see Chapter 5). We focused on model complexity ranging from $m = 0$ (isotropic diffusion only) to $m = 3$ fascicles. We investigated the performance of our B632 model selection approach with both synthetic phantoms and in vivo data. We compared it to the F-test on the signal residuals [102], for which the null hypothesis is that the fitting error of models with $m - 1$ and m fascicles are equivalent by assessing the F-score:

$$F_{m-1,m} = \frac{n - 1 - |\mathcal{M}_m|}{|\mathcal{M}_m| - |\mathcal{M}_{m-1}|} \frac{\text{SSE}_{m-1} - \text{SSE}_m}{\text{SSE}_{m-1}} > \theta^{\text{F-test}}, \tag{4.10}$$

where n is the number of data, $\theta^{\text{F-test}}$ is the F-score threshold above which the null hypothesis is rejected, and $|\mathcal{M}_m|$ and SSE_m are respectively the number of parameters and the sum of squared errors (fitting error) for a model with m fascicles. Various synthetic phantoms of size 15×15 were generated. The tensor profile $\boldsymbol{D}_i$ representing an individual fascicle was chosen to match typical in vivo data (trace of $2.1 \times 10^{-3}\text{mm}^2/s$ and FA of 0.8). We considered regions with 0, 1, 2 and 3 fascicles (Fig. 4.1). The simulated DWI were corrupted by various Rician-noise levels. In vivo imaging was achieved on a healthy volunteer using a Siemens 3T Trio scanner with a 32-channel head coil and the following parameters : FOV=220mm, 68 slices, matrix=128×128, resolution=$1.72 \times 1.7 \times 2$ mm^3.

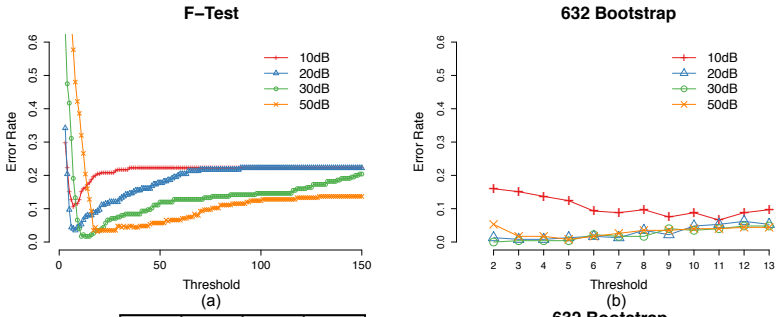
4.3.3 Synthetic Phantom Experiments

Synthetic phantoms offer a ground truth against which results of the model selection can be compared. We investigated the performance of the B632 and

F-test approaches under four different SNR : 10dB, 20dB, 30dB and 50dB. Both the B632 and the F-test approaches require determination of a threshold (see (4.9) and (4.10)). We investigated the influence of $\theta^{\mathrm{F-test}}$ by evaluating the F-test model selection with $3 < \theta^{\mathrm{F-test}} < 150$. Similarly, we evaluated the influence of θ^{B632} by computing the B632 model selection with $0 < \theta^{\mathrm{B632}} < 13$. Note that $\theta^{\mathrm{F-test}}$ is a threshold on the F-score while θ^{B632} is the number of standard errors above which a model m is considered better than a model $m-1$ (see (4.9)). The maximum number of bootstrap replicates for B632 was set to 150. We counted the number of errors between the ground truth and the automatic model selection results and reported the error rate.

The overall minimum error (obtained when the threshold can be chosen independently for different SNR) is consistently higher with the F-test (Fig. 4.2a) than with B632 (Fig. 4.2b). The table in Fig. 4.2c summarizes those errors. In practice, the SNR is unknown and so the choice of threshold cannot depend on it. The overall minimum error rate are therefore lower bounds for what can actually be achieved in practice. The increase in error rate compared to this lower bound, due to the choice of a single threshold, is more dramatical with the F-test than with B632 (Fig. 4.2c, bottom rows). In particular, at 10dB, the error rate almost doubles compared to its lower bound with the F-test, while it increases only by a few percents with B632.

Finally, the evolution of the error rate with the number of bootstrap replicates assesses the stability of the estimate and allows the definition of a minimum number of bootstrap replicates required to achieve good performances. Fig. 4.2d shows that the error rate becomes stable after approximately 50 replicates.



		10dB	20dB	30dB	50dB
overall minimum	F-Test	24 (10%)	8 (3%)	4 (2%)	7 (3%)
	632 Bootstrap	**15 (7%)**	**2 (0.9%)**	**0 (0%)**	**2 (0.9%)**
single threshold	F-Test	44 (19%)	20 (9%)	6 (3%)	8 (4%)
	632 Bootstrap	**20 (9%)**	**3 (1%)**	**4 (2%)**	**6 (3%)**

(c)

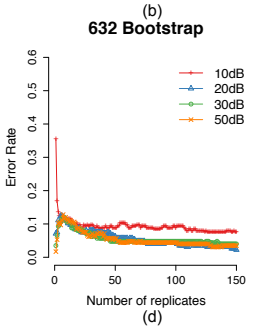


Figure 4.2: Evaluation on simulated data for various noise levels show that selection based on the generalization error is more reliable and robust than selection based on F-tests. (a) For the F-test, the dependence of the error rate with the threshold $\theta^{\mathrm{F-test}}$ shows that no single threshold can be used to achieve good performance at all noise levels. (b) The error rate is less sensitive to the threshold θ^{B632} set on the generalization error estimate. (c) Number of errors out of the 225 voxels and error rate for optimal thresholds chosen independently for each SNR (top rows) and jointly for all SNR (bottom rows). B632 leads to fewer error than the F-test, in both scenarios. The difference is more striking when a single threshold is used for all SNR. (d) The 632 bootstrap estimate reaches a close to optimal value after about 50 replicates.

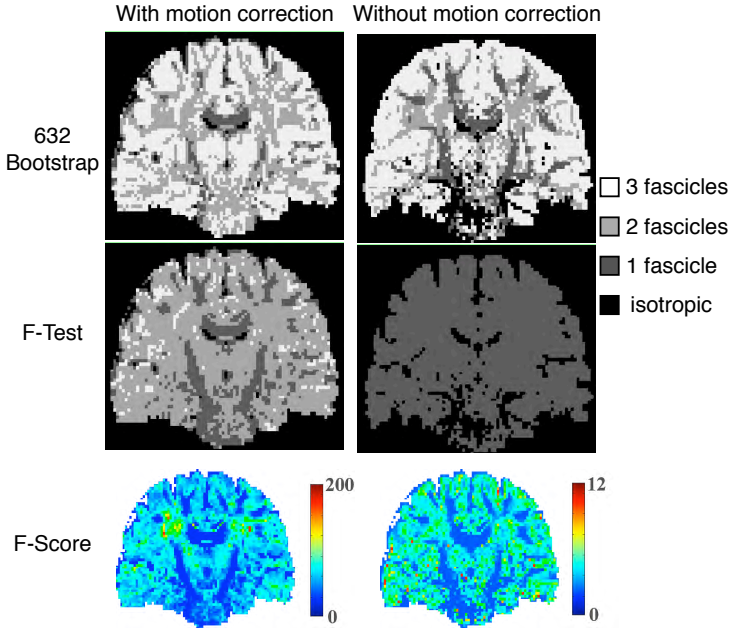


Figure 4.3: Evaluation of the model section approaches on *in vivo* data. Maps of the number of fascicles as detected with B632 and F-test for DWI with and without motion correction. Motion correction mostly introduced interpolation since no significant motion was present in this scan. B632 yields similar results in both cases, while the F-test selection fails to detect areas of more than one fascicle. This is due to the reduced perceived signal to noise ratio after motion correction, which affects the result of the F-test, as shown by the map of the F-score for the comparison between one and two-fascicle models.

4.3.4 *In Vivo* Data: Robustness to Pre-Processing

The results on the synthetic phantom suggest that the F-test model selection is not robust to changes in the characteristics of the image. In particular, if a threshold is optimized for some SNR, it will yield suboptimal results at another SNR. The acquisition of DWI is usually followed by several steps of pre-processing before the diffusion model is estimated. Some of these steps aim at improving the SNR. We may wonder whether the model selection is robust to these pre-processing step.

As an illustration, we applied the model selection methods on an in vivo acquisition before and after automatic motion correction based on coregistration of all the DWI. The acquisition was not corrupted by any significant subject motion, and therefore this step mostly introduces a smoothing due to the interpolation when coregistering the images. Results in Fig. 4.3 show that the F-test model selection is strongly affected by this preprocessing step. With a threshold of $\theta^{\mathrm{F-test}} = 15$, the map of the number of fascicles after motion correction does not resemble that before motion correction. In the latter, only one-tensor models and isotropic diffusion models were selected, while two- and three-tensor models are detected at many locations after motion correction. To observe two- or three-tensor models in the second map, one would need to decrease $\theta^{\mathrm{F-test}}$ since the SNR is lower, which is consistent with the synthetic results of Fig. 4.2(a). By contrast, for a constant $\theta^{\mathrm{B632}} = 8$, the maps of the number of fascicles detected with B632 are similar in both images and follow the traits of the anatomy.

Interestingly, the B632 approach favors a three-tensor models throughout the grey matter. Such a model is consistent with the known cortical anatomy where vertical columns cross horizontally spreading axons. However, such structures are probably too small and their geometry varies at too high a spatial frequency to be detected by clinical diffusion MRI. The systematic detection of three tensors in the cortex may, instead, be an artifact of the set of models being considered. The model with zero tensor has an isotropic compartment with a fixed diffusivity (that of unrestricted water at 37°C). As such, it has no free parameter. Diffusion of water molecules in the cortex may be approximately isotropic but is not unrestricted because axons and neurons hinder their motion. An appropriate model would probably be, in this case, an isotropic compartment with a variable diffusivity. This model would have a single free parameter.

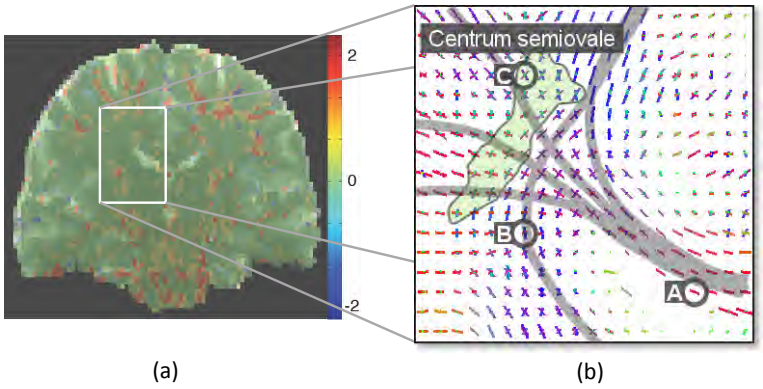


Figure 4.4: Validation with cross-testing indicates that selection based on the generalization error is, on average, significantly better than selection based on the F-test. (a) Comparison of the F-test and B632 model selection approaches using cross-testing. The map shows the difference between the testing error when F-test or B632 are used. It shows that the testing error is significantly larger when using the F-test. (b) Illustration that B632 and MFM estimation enables reliable detection of the number of fascicles that matches the known anatomy. [A] Body of the corpus callosum, [B] Crossing of the corpus callosum and the cortico-spinal tracts and [C] Centrum semiovale which contains three fascicle orientations.

4.3.5 *In Vivo* Data: Cross-Testing Validation

Experiments based on in vivo data cannot rely on any ground truth to assess the error rates of the methods. To objectively compare the performance of the F-test and the B632 model selection approaches, we performed a cross-testing analysis. This procedure consists in repeatedly splitting the dataset into an *estimation set* and a *testing set*. In our experiments, we considered 70% of the data for estimation and the remaining 30% for testing. Both the model selection and estimation of the MFM parameters were carried out with the estimation set while the testing set was used to assess the performance of the two approaches. The threshold parameters were set to respectively $\theta^{\mathrm{F-test}} = 15$ and $\theta^{\mathrm{B632}} = 8$. The performance of the approaches was assessed by computing the mean-square prediction error on the testing set. We repeated the estimation-testing process 30 times and computed the average testing error. Fig 4.4 shows that the testing error with B632 is lower than with F-test. More precisely, a paired t-test on the differences between the testing errors at each voxel shows that B632 is significantly better than F-test ($p < 10^{-12}$) with a mean improvement of 0.56.

4.4 Summary and Discussion

The estimation of the generalization error allows a reliable selection of the optimal model. Results on both synthetic and in vivo data show the improved performance over model selection based on the F-test.

Validating the models by means of an external dataset, as done in Section 4.3.5, seems the most objective validation method. Arguably, the generalization error is therefore what all model selection approaches attempt to minimize. However, unlike the fitting error, the generalization error cannot be computed and model selection criteria can be viewed as bypasses to this conceptual limitation. In this interpretation, the fitting error is, itself, an estimate of the generalization error. Complexity penalization can then be seen as heuristic methods to correct for the high bias of this estimate. The 632 bootstrap estimate, on the other hand, directly estimates the generalization error and is designed to have low bias and variance.

Computing the fitting error is a lot faster than computing the 632 bootstrap which requires several estimations of the model (one for each of the B bootstrap replicates). For this reason, model selection and estimation using B632 is about B times slower. Different approximations of the bootstrap estimates can be used to decrease the computational time of the bootstrap estimates [111]. A stopping criterion on the number of bootstrap replicates based on the current estimate of the generalization error can also be used.

Chapter 5

Estimation of Multi-Fascicle Models

When MRI need to be acquired in patients, acquisition time is a factor of paramount importance, and even more so for children who have difficulties to remain still in the scanner. Diffusion imaging, as we discussed in Chapter 2, requires more time than conventional MRI because multiple DWI with different gradient orientations need to be acquired. Reducing the number of acquisitions to its bare minimum while having sufficient DWI to reliably estimate a multi-fascicle model is therefore a challenge of prime importance.

Existing methods fail to estimate the parameters of multi-fascicle models from conventional diffusion sequences with the desired accuracy. In this chapter, we first demonstrate that, with conventional acquisitions at a single b-value, there is a manifold of models that all produce the same diffusion signal and thus cannot be identified. To regularize this problem, we propose to learn a prior over the model parameters from data acquired at several b-values in an external population of subjects. We show that this population-informed prior enables, for the first time, accurate estimation of multi-fascicle models from single non-zero b-value data typically acquired in clinical practice. The approach is validated on synthetic and *in vivo* data of healthy subjects and patients with autism. We apply it in population studies of the white matter microstructure in autism spectrum disorders and show that the introduction of a population-informed prior leads to reliable detection of group differences. This approach enables novel investigations from large existing DWI datasets both in normal development and in disease and injury.

5.1 Related Work

Estimating the parameters of a multi-fascicle model is known to be unstable [108]. It was originally thought that multi-tensor models can be estimated from DWI acquired at a single non-zero b-value [3]. However, it was later shown that

this is an ill-posed problem [59, 101, 102, 105], justifying why the estimation of the model parameters appears unstable. Kreher *et al.* [59] estimate a model with two anisotropic tensors and one isotropic tensor without accounting for the ill-posedness problem, so that only pseudo tensors (whose eigenvalues are undefined) are estimated. Furthermore, to increase the stability of their numerical estimation, they assume that the two lowest eigenvalues of each tensor are equal. These assumptions result in a model with 9 degrees of freedom instead of the initial 15. This model does not allow measurements of diffusion properties (such as the fractional anisotropy and the mean diffusivity) for each fascicle, limiting its use in population studies of the brain microstructure.

Another approach to circumvent the ill-posedness problem is to fix *a priori* either the volumetric fraction of each fascicle [66, 143] or the properties of the fascicles [21, 105, 132]. In [132], the tensor eigenvalues are completely fixed *a priori* so that only the directions need to be estimated. In [21, 105], a ball-and-stick model is estimated. This model assumes that the fascicles are all identical and that there is no diffusion in directions orthogonal to the fascicle orientation ($\lambda_1 > 0$ and $\lambda_2 = \lambda_3 = 0$). In [66] and [143], the volumetric fractions are arbitrarily fixed to an equal value. Because the ill-posedness problem affects both the volumetric fractions and the tensor eigenvalues, unreasonable assumptions on any of those parameters necessarily impact the estimation of the other parameters.

Assumptions about the volumetric fractions are arbitrary since the location of the boundary between fiber bundles, or between the white and the grey matter, are not known in advance and depend on the position of the subject in the scanner, and on the image resolution. As for assumptions about the fascicle properties, there is a long-standing body of evidence showing that fascicles differ throughout the brain [62, 1] and across subjects due to normal development [40] and disease [49, 148]. These changes in axon diameter, axon density and degree of myelination have direct influences on diffusion properties of each fascicle.

Some researchers have proposed to rely on spatial priors to regularize the estimation problem. Spatial priors assume that the multi-tensors in adjacent voxels are not independent. Pasternak *et al.* proposed a piece-wise smoothness constraint to regularize the estimation of a single-tensor model with a free-water compartment [81, 80]. In [102], an anisotropic spatial prior is used to regularize a two-tensor field. As we will see in Section 5.2.2, these spatial priors do not solve the ill-posedness problem when estimating a general N-fascicle model from single b-value data. More importantly, we will show in Section 5.5 that relying

solely on spatial priors may mislead the conclusions drawn from population studies.

To regularize the estimation problem while keeping all the degrees of freedom of the model, Scherrer and Warfield proposed to acquire data at several b-values [102]. To achieve high signal-to-noise ratio with multiple non-zero b-values, they introduced a novel acquisition sequence that combines gradients on a shell at the nominal b-value and gradients on the enclosing cube of constant TE (reaching effective b-values up to three times the nominal b-value). Pasternak *et al.* showed that even for a simpler model including only one anisotropic tensor, inclusion of multiple non-zero b-values improves the accuracy of the estimation [80].

However, many large diffusion imaging datasets have been acquired with a single-shell HARDI acquisition sequence. For instance, the recent Pediatric Imaging, Neurocognition, and Genetics Study (PING) includes diffusion images from 1,400 children at $b = 0$ s/mm^2 and $b = 1000$ s/mm^2 [37]. Such large datasets hold unprecedented promise to better understanding the relation between the brain microstructure and the neuropsychological development in children. The inability to estimate multi-fascicle models from single b-value data, however, jeopardizes the potential of these data to reveal novel insight into the brain microstructure. There is therefore a strong need for a method to estimate multi-fascicle models from data acquired at a single non-zero b-value. A method to estimate a general N-tensor model can be applied to many generative models since most of them include, as part of the model, a multi-tensor of some kind.

In this chapter, we propose a method to estimate a general N-fascicle model from single-shell data. The ill-posedness is removed from the estimation by leveraging prior knowledge from data acquired at multiple b-values in an external population of subjects. Section 5.2 explores the geometry of the ill-posed problem and explain why spatial priors alone do not regularize it. Section 5.3 describes how data acquired in an external set of subjects can be used to build a population-informed prior that regularizes the estimation problem. Section 5.4 validates the proposed approach in synthetic and *in vivo* data. Section 5.5 illustrates the use of the population-informed prior to conduct population studies of the brain microstructure. Finally, Section 5.6 summarizes the contributions and results of this section.

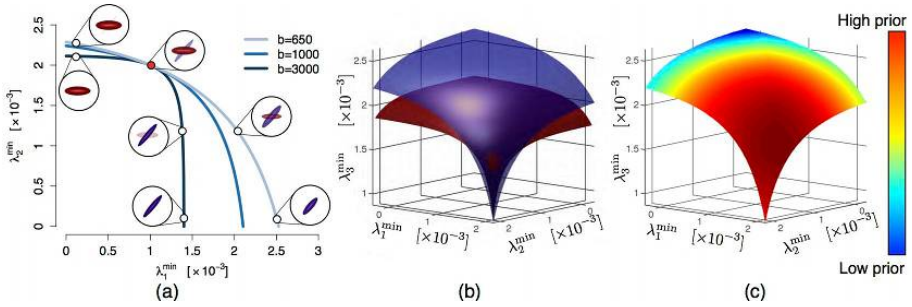


Figure 5.1: Geometry of the manifolds of indistinguishable multi-fascicle models. (a) At a given b-value, an infinite number of models with very different properties produce the same diffusion signal and form a manifold. At the true underlying model (red dot), the manifolds at different b-values intersect tangentially. (b) For an N-fascicle model (here N=3), the manifolds are $(N-1)$-dimensional hypersurfaces which share the same tangent hyperplane at their intersection making the estimation problem very sensitive to noise. (c) The population-informed prior (here restricted to the manifold) associates different probabilities to otherwise equivalent models.

5.2 Geometry of the Ill-Posed Estimation Problem

In this section, we analyze the problem of estimating N-fascicle models from a geometric point of view. We then demonstrate why spatial priors cannot be used to remove the ill-posedness from the estimation.

5.2.1 Manifolds of Equivalent Models at a Given B-Value

A multi-fascicle model is represented as a mixture of single fascicle models. In the multi-tensor formalism, we can represent each fascicle by a tensor [132]. In this setting, the generative model for the formation of the diffusion signal S for a b-value b and a gradient direction $\boldsymbol{g}$ (with $\|\boldsymbol{g}\| = 1$) is:

$$S = S_0 \sum_{i=1}^{N} f_i e^{-b\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}}, \tag{5.1}$$

where $\boldsymbol{D}_i$ and f_i are the tensor and the volumetric fraction of the i-th fascicle. Since, for any $\gamma_i > 0$, $\gamma_i e^{-\log \gamma_i} = 1$, all multi-fascicle models with fractions $\gamma_i f_i$ and tensors $\boldsymbol{D}_i + \frac{\log \gamma_i}{b}\boldsymbol{I}$ produce the same signal:

$$S = S_0 \sum_{i=1}^{N} \gamma_i f_i e^{-b\boldsymbol{g}^T \left(\boldsymbol{D}_i + \frac{\log \gamma_i}{b}\boldsymbol{I}\right)\boldsymbol{g}}, \quad \text{with the constraint } \sum_{i=1}^{N} \gamma_i f_i = 1. \tag{5.2}$$

The tensors remain positive definite as long as $\gamma_i > e^{-b\lambda_i^{\min}}$, where $\lambda_i^{\min}$ is the lowest eigenvalue of $\boldsymbol{D}_i$. These models are all equally compatible with the observed signal S and selection of a particular model based solely on the measurements is purely arbitrary. Fixing the value of the γ's amounts to fixing the lowest eigenvalues ($\lambda_i^{\min}$) of each tensor ($\boldsymbol{D}_i$). Each model satisfying (5.2) is thus uniquely identified by its vector $(\lambda_1^{\min}, \ldots, \lambda_N^{\min})$. The set of all models respecting equation (5.2) is a manifold of dimension $(N-1)$ defined by the implicit equations (we let $\lambda_i \triangleq \lambda_i^{\min}$ for clarity):

$$\begin{cases} \lambda_i = \lambda_i^{\text{true}} + \frac{1}{b}\log(\gamma_i) & \text{for } i = 1, \ldots, N \\ \\ \sum_{i=1}^{N} \gamma_i f_i = 1, \end{cases} \tag{5.3}$$

where $(\lambda_1^{\text{true}}, \ldots, \lambda_N^{\text{true}})$ is the true unknown model obtained if we let $\gamma_1 = ... = \gamma_N = 1$ (Fig. 5.1(a)). Since these equations depend on b, so will the manifold. Acquiring diffusion images at different b-values amounts to defining different such manifolds that intersect at the true underlying model $\lambda_i = \lambda_i^{\text{true}}$. Let us analyze how those manifolds intersect at that point. The explicit equation of the hypersurface $\lambda_N(\lambda_1, \ldots, \lambda_{N-1})$ obtained by eliminating the γ's between equations (5.3) is:

$$\lambda_N(\lambda_1, \ldots, \lambda_{N-1}) = \lambda_N^{\text{true}} + \frac{1}{b}\log\left(\frac{1 - \sum_{i=1}^{N-1} f_i e^{b(\lambda_i - \lambda_i^{\text{true}})}}{f_N}\right). \tag{5.4}$$

The normal vector to the hypersurface is $\boldsymbol{\eta} = \left(\frac{\partial\lambda_N}{\partial\lambda_1}, \ldots, \frac{\partial\lambda_N}{\partial\lambda_{N-1}}, -1\right)$. Its k-th component is:

$$\eta_k = \frac{\partial\lambda_N}{\partial\lambda_k} = \frac{-f_k e^{b(\lambda_k - \lambda_k^{\text{true}})}}{1 - \sum_{i=1}^{N-1} f_i e^{b(\lambda_i - \lambda_i^{\text{true}})}},\ k = 1, ..., N-1.$$

At the true underlying model, the normal vector is:

$$\eta_k\Big|_{\lambda_i = \lambda_i^{\text{true}}, \forall i} = \frac{\partial\lambda_N}{\partial\lambda_k}\Big|_{\lambda_i = \lambda_i^{\text{true}}, \forall i} = \frac{-f_k}{f_N}. \tag{5.5}$$

Remarkably, this normal vector (hence the tangent hyperplane) does not depend on b at the point of interest. In other words, at the first-order approximation, the manifolds at all b-values coincide locally, explaining the high sensitivity to noise encountered when optimizing the parameters of a multi-fascicle model (Fig. 5.1(b)). This is akin to the sensitivity of the intersection point of two

almost-parallel lines in the plane: small perturbations of their slope can have a dramatic effect, whereas for nearly orthogonal lines, the effect would be far smaller.

At the second-order approximation, the manifold is characterized by the Hessian matrix of $\lambda_N(\lambda_1, \ldots, \lambda_{N-1})$:

$$\left(\boldsymbol{H}\right)_{ij} = \frac{\partial^2 \lambda_N}{\partial_i \partial_j},$$

and a typical measure of the curvature in direction $\boldsymbol{s}$ is $\kappa_b(\boldsymbol{s}) = \boldsymbol{s}^T \boldsymbol{H} \boldsymbol{s}$. At the true underlying model, the Hessian has the following expression:

$$\boldsymbol{H}\Big|_{\lambda_i = \lambda_i^{\text{true}}} = \frac{-b}{f_N^2} \left(\tilde{\boldsymbol{f}} \tilde{\boldsymbol{f}}^T + f_N \ \text{diag}(\tilde{\boldsymbol{f}}) \right),$$

where $\tilde{\boldsymbol{f}} = [f_1, \ldots, f_{N-1}]^T$. The difference between the Hessian matrices at the intersection point for two different b-values, b and $b' > b$, is positive definite since, for all $\boldsymbol{s} \neq 0$, we have

$$\kappa_b(\boldsymbol{s}) - \kappa_{b'}(\boldsymbol{s}) = \boldsymbol{s}^T \left(\boldsymbol{H}(b) - \boldsymbol{H}(b')\right) \boldsymbol{s} = \frac{b' - b}{f_N^2} \left((\tilde{\boldsymbol{f}}^T \boldsymbol{s})^2 + f_N \boldsymbol{s}^T \text{diag}(\tilde{\boldsymbol{f}}) \boldsymbol{s} \right) > 0. \tag{5.6}$$

Therefore, there exists no direction $\boldsymbol{s}$ along which the two manifolds have the same curvature. Consequently, the true model is locally the only intersection of all manifolds. Given the difference (5.6), it appears that a wider range of b-values leads to a larger difference between their manifolds, which should in turn improve the accuracy of the estimation (ignoring the potential impact of b on noise). Indeed, (5.6) indicates that manifolds with large b-values are strictly more curved than manifolds with small b-values, in all directions. Continuing with a different analogy, this is akin to having a small ball stuck inside a larger ball, such that they intersect at a single point. Moving the small ball gently displaces the intersection point somewhat. On the other hand, if the inner ball is nearly as large as the outer one (almost the same b-value), then small perturbations (due to noise) can lead to large displacements of the contact point.

When an isotropic compartment $f_{\text{iso}} e^{-bD_{\text{iso}}}$ is added to the model, one can show that the above development remains valid with an unchanged N if D_{iso} is known (typically set to the diffusion of water at 37°C, that is $3 \times 10^{-3} mm^2/s$) and considering an $(N+1)$-fascicle model if D_{iso} needs also be optimized.

In summary, estimating multi-fascicle models from single b-value data is an ill-posed problem that has an infinite number of equivalent solutions. However, due to the particular geometry of the manifold of these solutions, acquiring data at two b-values is sufficient to resolve the particular equivalence relation of (5.3).

5.2.2 Ill-Posedness with Spatial Priors

In this section, we demonstrate that the use of the most common spatial priors, as described in [102, 81, 80], does not alleviate the ill-posedness problem of the estimation. In their most general expression, these spatial priors impose a penalty on multi-fascicle models based on some discrepancy between the multi-tensor at one voxel $\mathcal{M}(\boldsymbol{x})$ and the multi-tensors in the neighborhood $\Omega_{\boldsymbol{x}}$, $\{\mathcal{M}(\boldsymbol{x}')|\boldsymbol{x}' \in \Omega_{\boldsymbol{x}}\}$:

$$P(\{\mathcal{M}(\boldsymbol{x}')|\boldsymbol{x}' \in \Omega_{\boldsymbol{x}}\}) = f(\{d(\mathcal{M}(\boldsymbol{x}'), \mathcal{M}(\boldsymbol{x}))|\boldsymbol{x}' \in \Omega_{\boldsymbol{x}}\}), \tag{5.7}$$

where $d(\mathcal{M}(\boldsymbol{x}'), \mathcal{M}(\boldsymbol{x}))$ is some discrepancy function between the multi-tensors $\mathcal{M}(\boldsymbol{x})$ and $\mathcal{M}(\boldsymbol{x}')$. The expressions for the spatial prior may vary (through the choice of f, d and/or $\Omega_{\boldsymbol{x}}$) but they share the property of reaching a maximum if the local neighborhood is constant (*i.e.* if $\mathcal{M}(\boldsymbol{x}') = \mathcal{M}(\boldsymbol{x}),\ \forall \boldsymbol{x}' \in \Omega_{\boldsymbol{x}}$), since, typically, $d(\mathcal{M}, \mathcal{M}) = 0$, for any $\mathcal{M}$. For instance, the prior used in [102] for a multi-fascicle model that has N tensors $\boldsymbol{D}_i(\boldsymbol{x})$ at location $\boldsymbol{x}$ is:

$$P_{\text{spatial}}(\boldsymbol{D}) \propto exp\left\{-\frac{1}{2\sigma_s^2}\sum_{\boldsymbol{x}}\sum_{i=1}^{N}\|\nabla \log \boldsymbol{D}_i(\boldsymbol{x})\|_{\text{F}}^2\right\}, \tag{5.8}$$

where ∇ is the spatial gradient with respect to $\boldsymbol{x}$ approximated by finite differences [57], with $\boldsymbol{D}\hat{=}(\boldsymbol{D}_1, ..., \boldsymbol{D}_N)$ and σ_s weights the confidence put in the prior. The gradient ∇ is computed along the fascicle by selecting for each $\boldsymbol{D}_i$ at $\boldsymbol{x}$, the two tensors in $\Omega_{\boldsymbol{x}}$ most similar to $\boldsymbol{D}_i(\boldsymbol{x})$.

In a constant field of multi-tensor models, all gradients are equal to zero and the prior is maximum. Therefore, two constant multi-tensor fields must have equal spatial priors. Given a constant multi-fascicle model, one can generate an infinite number of constant multi-fascicle models from (5.3) that the data alone cannot distinguish. Since the spatial prior is unable to distinguish them either, the ill-posedness problem persists.

Spatial priors that do not follow the expression (5.7) may be proposed. For instance, one may propagate properties of tensors from single-tensor to multi-tensor areas [104]. Since the estimation in single-tensor areas is not ill-posed,

such a prior would, in theory, avoid the ill-posedness problem. However, this regularization comes at the expense of strong assumptions about the homogeneity of diffusion properties along the fascicles and would lead to estimation errors that vary in space (voxels that are further away from single-tensor areas probably have larger estimation errors).

5.3 Model Estimation with a Population-Informed Prior

While all models of Equation (5.3) are equally compatible with the observed DWI at a given b-value, they are not all equally likely from a biological point of view. This knowledge can be learnt from available observations of multi-fascicle models at multiple non-zero b-values in other subjects. Since these subjects are imaged at multiple non-zero b-values, their multi-fascicle models are not affected by the ill-posedness problem and they are therefore reliable observations of the anatomy. In this section, we explain how this prior knowledge can be encoded and integrated in the estimation.

5.3.1 Posterior Predictive Distribution of the Parameters

For a fascicle i at a specific voxel, we denote by $\mathcal{F}_i = \{f_i^0, \dots, f_i^m\}$ the observations of its i-th fraction and by $\mathcal{D}_i = \{\boldsymbol{D}_i^0, \dots, \boldsymbol{D}_i^{m_i}\}$ the observations of its i-th tensor in m other subjects. Notice that the number of observations for the fraction is always equal to the number m of subjects while the number of observed tensors is $m_i \leq m$ since, if a fraction is zero, the corresponding tensor is not observed. We let $\boldsymbol{\mathcal{F}} = (\mathcal{F}_1, ..., \mathcal{F}_N)$ and $\boldsymbol{\mathcal{D}} = (\mathcal{D}_1, ..., \mathcal{D}_N)$ for the observations of all fractions and all tensors in one voxel. These observations can be incorporated into the estimation of a multi-fascicle model in a new subject, as a prior over the parameters $(\mathsf{f}_i, \mathbf{D}_i)$ (Fig. 5.1(c)). In this section, we will derive an expression for the prior probability $P_{\mathbf{f},\mathbf{D}}(\boldsymbol{f}, \boldsymbol{D}|\boldsymbol{\mathcal{F}}, \boldsymbol{\mathcal{D}})$ for the fractions and tensors in one voxel given the observations at the same voxel in other subjects.

We assume that the fascicle properties are independent of partial voluming and that the properties of one fascicle are independent of those of another. The prior can therefore be expressed as:

$$P_{\mathbf{f},\mathbf{D}}(\boldsymbol{f}, \boldsymbol{D}; \boldsymbol{\theta}) = P_{\mathsf{f}}(\boldsymbol{f}; \boldsymbol{\theta}_f) \prod_{i=1}^{N} P_{\mathbf{D}_i}(\boldsymbol{D}_i; \boldsymbol{\theta}_i), \tag{5.9}$$

where $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_i$ are parameters of the distributions. The fractions are not independent since they sum to 1. However, we assume that any fraction f_i is

independent of the relative proportions of others $\mathsf{f}_j/(1-\mathsf{f}_i)$. This *neutral vector* assumption naturally leads to the Dirichlet distribution:

$$P_{\mathsf{f}}(\boldsymbol{f};\boldsymbol{\alpha}) = \frac{\mathbb{1}_{\boldsymbol{f}\in\mathcal{S}}}{B(\boldsymbol{\alpha})}\prod_{i=1}^{N} f_i^{\alpha_i-1}, \text{ where } \mathcal{S} = \left\{ \boldsymbol{x}\in\mathbb{R}^N : x_i>0, \sum_{i=1}^{N} x_i = 1 \right\}, \tag{5.10}$$

where $\mathbb{1}_{\boldsymbol{f}\in\mathcal{S}}$ equals one if only if $\boldsymbol{f}\in\mathcal{S}$ and zero otherwise. Calculus of diffusion tensor data is typically performed in the log-Euclidean domain since it resembles the Euclidean space and avoids nonsensical negative eigenvalues [8]. The prior knowledge about $\mathbf{D}_i$ can be described as a multivariate Gaussian distribution over their logarithm [107, 8]:

$$\mathbf{L}_i = \log \mathbf{D}_i \sim \mathcal{N}\left(\mathbf{M}_i, \boldsymbol{\Sigma}_i\right). \tag{5.11}$$

In this notation, $\mathbf{L}_i$ is vectorized as $\big(\mathrm{diag}(\mathbf{L}_i), \sqrt{2}\ \mathrm{offdiag}(\mathbf{L}_i)\big) \in \mathbb{R}^6$. In general, $\boldsymbol{\Sigma}_i$ is a 6×6 symmetric matrix and has 21 free parameters, which may overfit the usually small training dataset. For DTI, Schwartzman *et al.* suggested in [107] to constrain $\boldsymbol{\Sigma}_i$ to be orthogonally invariant, imposing the following structure that depends only on σ_i and τ_i:

$$\boldsymbol{\Sigma}_i = \sigma_i^2 \begin{pmatrix} \boldsymbol{I}_3 + \frac{\tau_i}{1-3\tau_i}\boldsymbol{I}_3 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_3 \end{pmatrix} \triangleq \boldsymbol{B}(\sigma_i,\tau_i).$$

Orthogonal invariance implies that a deviation $\boldsymbol{\Delta}_i$ from the mean $\mathbf{M}_i$ has the same likelihood to occur as any of its rotated versions $\boldsymbol{Q}\boldsymbol{\Delta}_i\boldsymbol{Q}^T$ for any orthogonal matrix $\boldsymbol{Q}$. In other words, the observation of a tensor $\mathbf{M}_i+\boldsymbol{\Delta}_i$ is equally likely as the observation of $\mathbf{M}_i+\boldsymbol{Q}\boldsymbol{\Delta}_i\boldsymbol{Q}^T$, which is a notion of isotropy. This structure yields a closed-form solution for the maximum likelihood (ML) estimator [107]:

$$\hat{\boldsymbol{M}}_i = \bar{\boldsymbol{L}}_i = \frac{1}{m_i}\sum_{k=1}^{m_i}\boldsymbol{L}_i^k \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_i = \boldsymbol{B}(\hat{\sigma}_i,\hat{\tau}_i), \text{with} \tag{5.12}$$

$$\hat{\tau}_i = \frac{-\sum_{i=1}^{m_i}\|\boldsymbol{L}_i^k-\bar{\boldsymbol{L}}_i\|_2^2}{5\sum_{i=1}^{m_i}\left[\mathrm{Tr}(\boldsymbol{L}_i^k-\bar{\boldsymbol{L}}_i)\right]^2} \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{6m_i}\sum_{i=1}^{m_i}\|\boldsymbol{L}_i^k-\bar{\boldsymbol{L}}_i\|_{\hat{\tau}_i}^2, \tag{5.13}$$

where $\|A\|_t^2 = \langle A,A\rangle_t$ and $\langle \boldsymbol{A},\boldsymbol{B}\rangle_t = \mathrm{Tr}(\boldsymbol{A}\boldsymbol{B}) - t\,\mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{B})$. The ML estimator may be unreliable for compartments with only a few observations. This uncertainty is accounted for by replacing point estimates of $\boldsymbol{\theta}$ by posterior

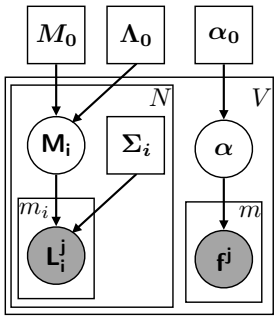


Figure 5.2: Graphical model of the posterior predictive distribution model that serves as a prior for new observations. Circles indicate random variables, shaded circles indicate observed variables, and boxes indicate parameters. The graphical model is repeated for all V voxels.

distributions and integrating over all possible $\boldsymbol{\theta}$. This yields the *posterior predictive distribution* which contains all the knowledge about new observations that we learn from previous observations. Its derivation requires the definition of hyperpriors over $\boldsymbol{\theta}$ and is closed-form if we select conjugate hyperpriors. $\mathbf{M}_i \sim \mathcal{N}(\boldsymbol{M}_0, \boldsymbol{\Lambda}_0)$ is a conjugate hyperprior for the tensor part of (5.9) assuming a deterministic $\boldsymbol{\Sigma}_i = \hat{\boldsymbol{\Sigma}}_i$. We set $\boldsymbol{\Lambda}_0$=$\boldsymbol{B}(1,0)$ and $\boldsymbol{M}_0$=$\log \boldsymbol{D}_{\text{iso}}$ to keep it weakly informative (this hyperprior merely encodes the order of magnitude of diffusivity at $37°C$). The posterior predictive distribution over the tensors becomes:

$$\mathbf{L}_i|\mathcal{D}_i \sim \mathcal{N}(\boldsymbol{M}_i^{m_i}, \boldsymbol{\Lambda}_i^{m_i}), \tag{5.14}$$

$$\text{with} \quad \boldsymbol{\Lambda}_i^{m_i} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}} + \left(\boldsymbol{\Lambda}_0^{-1} + m_i \hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}^{-1}\right)^{-1} \triangleq \boldsymbol{B}(\tilde{\sigma}_i, \tilde{\tau}_i), \tag{5.15}$$

$$\text{and} \quad \boldsymbol{M}_i^{m_i} = \left(\boldsymbol{\Lambda}_0^{-1} + m_i \hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}^{-1}\right)^{-1} \left(\boldsymbol{\Lambda}_0^{-1} \boldsymbol{M}_0 + m_i \hat{\boldsymbol{\Sigma}}_{\boldsymbol{i}}^{-1} \bar{L}_i\right). \tag{5.16}$$

For the parameters α_i in (5.10), a conjugate hyperprior is the Dirichlet distribution. We set all its parameters to 1, making it uniform over the simplex $\mathcal{S}$. The resulting posterior predictive distribution is a Dirichlet with parameters $1 + \sum_{k=1}^{m} f_i^k$. In this expression, f_i^k are considered frequency counts since they are samples of f_i rather than samples from a multinomial parameterized by f_i. The complete posterior predictive distribution is represented as a graphical

model in Fig. 5.2 and its expression is (with $C_{\mathcal{F},\mathcal{D}}$ constant):

$$P_{\mathbf{f},\mathbf{D}}(\boldsymbol{f},\boldsymbol{D}|\mathcal{F},\mathcal{D}) = C_{\mathcal{F},\mathcal{D}}\mathbb{1}_{\boldsymbol{f}\in\mathcal{S}}\prod_{i=1}^{N} f_i^{\sum_{k=1}^{m} f_i^k}\prod_{i=1}^{N}\exp\left\{-\frac{\|\boldsymbol{L}_i-\boldsymbol{M}_i^{m_i}\|^2_{\tilde{\tau}_i}}{2\tilde{\sigma}_i^2}\right\}. \tag{5.17}$$

This prior pertains to a single voxel. The population-informed prior for the entire multi-fascicle model is obtained by multiplying this prior for all location $\boldsymbol{x}$. Dependence between adjacent voxels can be accounted for by a separate spatial prior. This population-informed prior encodes our *a priori* knowledge of the brain microstructure at every location. In the Section 5.3.3, we will see how this prior is incorporated in a maximum *a posteriori* estimate of the model parameters.

5.3.2 Building and Evaluating the Population-Informed Prior

The population-informed prior is built from data acquired in completely different subjects at several b-values. These data are used to compute multi-fascicle models for each subject. These multi-fascicle models are then registered to a multi-fascicle atlas, using the method described in [127]. Following alignment, tensors from all subjects at each voxel are clustered in N compartments by minimizing the cumulative relative differential entropy as described in [125]. The number N of compartments is set to the maximum number of tensors observed among subjects at that location. Each cluster represents the sets $\mathcal{F}_i$ and $\mathcal{D}_i$ of available observations for the i-th fascicle and the prior is built following the procedure described in Section 5.3.1.

To evaluate the prior at a given voxel $\boldsymbol{x}$ for a given candidate model $(\boldsymbol{f}(\boldsymbol{x}),\boldsymbol{D}(\boldsymbol{x}))$, we need the prior to be aligned to the subject whose multi-fascicle model is being estimated. We also need to identify which fascicle of the prior corresponds to what fascicle of the model being estimated. Alignment of the prior is performed by first estimating a multi-fascicle model without the population-informed prior and by performing multi-fascicle registration [127] between this first estimate and the multi-fascicle atlas. Associations between the fascicles of the model and the fascicles of the prior is achieved by computing the prior for all possible associations of fascicles and recording the maximum value obtained.

5.3.3 Maximum A Posteriori Estimation

The population-informed prior is incorporated in the estimation of the model as a prior over the parameters. The maximum *a posteriori* estimate of the parameters becomes:

$$\hat{\boldsymbol{f}}, \hat{\boldsymbol{D}} = \underset{\boldsymbol{f},\boldsymbol{D}}{\arg\max} \Big\{ P_{\mathbf{f},\mathbf{D}}(\boldsymbol{f}, \boldsymbol{D}|\mathcal{F}, \mathcal{D}).P_{\text{spatial}}(\boldsymbol{f}, \boldsymbol{D}).P_{\text{likelihood}}(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{D}) \Big\}.$$

We use the spatial prior $P_{\text{spatial}}(\boldsymbol{f}, \boldsymbol{D})$ defined by Scherrer and Warfield [102] and recalled in Equation (5.8). The likelihood density is over measurements $\boldsymbol{y} = (y_1, ..., y_K)$, that is the set of K DWI. The expression of the likelihood density depends on the noise model. Since the noise impacts DWI acquired at a single b-value that is typically low (*e.g.* b=1000 s/mm^2), the Gaussian approximation of the noise is valid [44]. We denote its variance by $\boldsymbol{\sigma}^2_{\text{noise}}$, leading to:

$$P_{\text{likelihood}}(\boldsymbol{y}|\boldsymbol{f}, \boldsymbol{D}) \propto \prod_{\boldsymbol{x}} \prod_{k=1}^{K} \exp\left(-\frac{1}{2\sigma^2_{\text{noise}}} \Big(S_k(\boldsymbol{f}(\boldsymbol{x}), \boldsymbol{D}(\boldsymbol{x})) - y_k(\boldsymbol{x}) \Big)^2 \right),$$

where S_k is the modeled signal obtained with Equation (5.1) given the b-value (b_k) and the gradient orientation ($\boldsymbol{g}_k$) of the measured y_k.

The maximum *a posteriori* estimate at each voxel amounts to maximizing the following for $\boldsymbol{f}$ and $\boldsymbol{D}$:

$$\begin{aligned} L(\boldsymbol{f}, \boldsymbol{D}) = \sum_{\boldsymbol{x}} \quad \Bigg\{ & \quad \log(\mathbb{1}_{\boldsymbol{f} \in \mathcal{S}}) + \sum_{i=1}^{N} \sum_{j=1}^{m} f_i^j(\boldsymbol{x}) f_i(\boldsymbol{x}) - \sum_{i=1}^{N} \frac{||\log \boldsymbol{D}_i(\boldsymbol{x}) - \boldsymbol{M}_i^{m_i}(\boldsymbol{x})||^2_{\tilde{\tau}_i}}{2\tilde{\sigma}_i^2} \\ & - \quad \frac{1}{2\sigma_s^2} \sum_{i=1}^{N} \|\nabla \log \boldsymbol{D}_i(\boldsymbol{x})\|^2_{\text{F}} \\ & - \quad \frac{1}{2\sigma^2_{\text{noise}}} \sum_{k=1}^{K} \left(S_0 \left[\sum_{i=1}^{N} f_i(\boldsymbol{x}) e^{-b_k \boldsymbol{g}_k^T \boldsymbol{D}_i(\boldsymbol{x}) \boldsymbol{g}_k} \right] - y_k(\boldsymbol{x}) \right)^2 \Bigg\}. \end{aligned} \tag{5.18}$$

The first line is the population-informed prior, the second line is the spatial prior and the third line is the likelihood. Notice that the population-informed prior is learnt in closed form from data and therefore does not introduce additional parameters. The estimation of the optimal values for parameters σ_s^2 and σ^2_{noise} is investigated in the next section.

As in [102], (5.18) is maximized using the BOBYQA algorithm [86] that allows the introduction of constraints. The first term of (5.18), $\log(\mathbb{1}_{\boldsymbol{f} \in \mathcal{S}})$, is

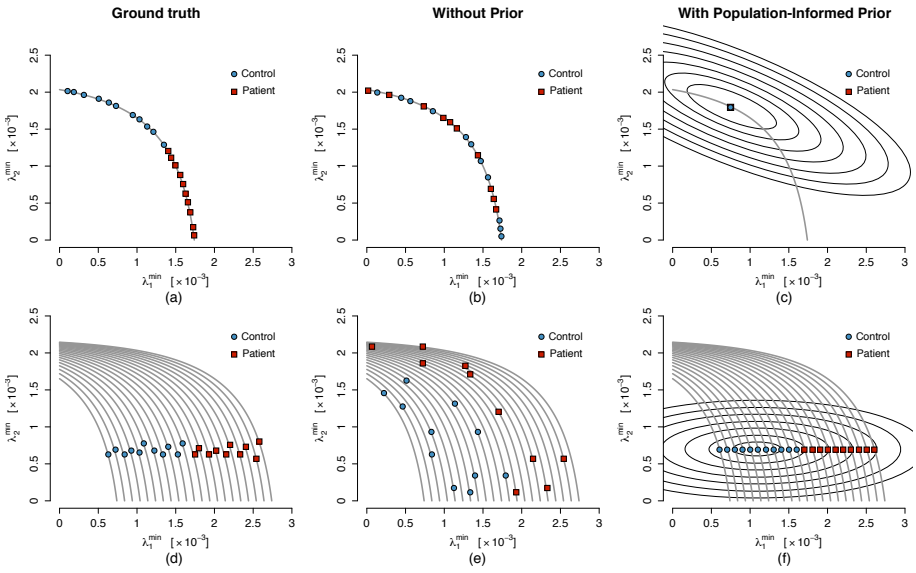


Figure 5.3: Effect of the ill-posedness problem and the use of a prior on the detectability of group differences. (a-c) If the group differences are aligned with a manifold of equivalent models, then the use of a prior will unavoidably collapse all models to a single point, preventing the detection of group differences. (d-f) If differences are not aligned with the manifold, however, the use of a population-informed prior will simply select the most likely model from the manifold of equivalent models. To ease the understanding of which model gets selected by the prior, isosurfaces of the population-informed prior are illustrated as black lines in (c) and (f).

thus set as a constraint on fractions $\boldsymbol{f}$ and removed from the objective function. The selection of the number N of fascicles is an important aspect of the model estimation. In this chapter, we use an F-test with a threshold $t = 25$ as in [102] due to its widespread use [59, 4, 102] to isolate the impact of the population-informed prior when comparing the results with those of earlier methods. Recent developments have, however, enabled more reliable selection of the number of fascicles [100, 103] and can be integrated with the proposed estimation.

5.3.4 Detectability of Group Differences

The introduction of an informed prior in the estimation implies that estimated models will be closer to the mean of the population than they would be if

the estimation problem was well-posed and if no prior was used. This is the *shrinkage towards the mean* effect. In this section, we investigate, theoretically, the impact of this effect on the detectability group differences.

Let us first imagine a population study in which the multi-tensor models of all subjects (in a particular voxel) lie on a single manifold described by (5.3) with the control subjects characterized by $\lambda_1^{\min} < t$ (for some threshold t) and the patients characterized by $\lambda_1^{\min} > t$ (Fig. 5.3(a)). If the ground truth was available, this distribution of $\lambda_1^{\min}$ would lead to a significant group difference. However, because all models lie on the same manifold, estimations from single b-value data with a population-informed prior would collapse all models to a single point (assuming that the measurement noise is negligible) and the group difference would not be detectable (Fig. 5.3(c)). With data at a single b-value, the absence of a prior would result in arbitrary estimations along the manifold which, too, would prevent the detection of group differences (Fig. 5.3(b)).

The situation described above is very unlikely, for models from each subject would need to follow exactly equation (5.3). This equation imposes a strict relationship between the eigenvalues of all tensors as well as their volumetric fractions. In particular, subjects with a higher $\lambda_1^{\min}$ would need to have a lower $\lambda_2^{\min}$ and a higher volumetric fraction f_1. More commonly, differences between groups are reflected into some properties of one of the tensors with all other properties (other tensors and volumetric fractions) kept relatively constant between groups (Fig. 5.3(d-f)). In this case, the models of each subject would lie on different manifolds (Fig. 5.3(d)). Assuming zero noise in the data, the use of a prior would then align the estimation to the most likely model along each manifold and the group differences would remain detectable.

Situations where group differences are not aligned with the manifolds of equivalent models are infinitely more likely to occur than situations where differences are confounded with the manifold. Indeed, although the space of first eigenvalues $(\lambda_1^{\min}, ..., \lambda_N^{\min})$ is sufficient to consider the equation of the manifolds, N-tensor models live in a $(7N-1)$-dimensional space spanned by $6N$ parameters for tensors and $N-1$ volumetric fractions. The manifold of equivalent models is therefore an $(N-1)$-dimensional manifold embedded in a space of $(7N-1)$ dimensions. The space orthogonal to the manifold is therefore $6N$-dimensional. For a given multi-tensor model $\mathcal{M}$, the set of models lying orthogonally to the manifold of models equivalent to $\mathcal{M}$ therefore contains infinitely more elements than the set of models lying on that manifold.

5.4 Validation Experiments

This section validates the proposed approach and compares its results with those obtained when no prior or only a spatial prior is used.

5.4.1 Comparison Metrics

In this section, different estimations of the multi-fascicle models will be compared. A multi-fascicle model $(\boldsymbol{f}, \boldsymbol{D})$ is compared to another $(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{D}})$, according to the following six metrics computed at every voxel:

$$\Delta^2_{\mathrm{FA}} = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \left(\mathrm{FA}(\boldsymbol{D}_i) - \mathrm{FA}(\tilde{\boldsymbol{D}}_i) \right)^2, \tag{5.19}$$

$$\Delta^2_{\mathrm{MD}} = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \left(\mathrm{MD}(\boldsymbol{D}_i) - \mathrm{MD}(\tilde{\boldsymbol{D}}_i) \right)^2, \tag{5.20}$$

$$\mathrm{Fro}^2 = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \|\boldsymbol{D}_i - \tilde{\boldsymbol{D}}_i\|^2_F, \tag{5.21}$$

$$\Delta_{\mathrm{Dir}} = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \left(1 - |\boldsymbol{e}_{1,i}.\tilde{\boldsymbol{e}}_{1,i}|\right), \tag{5.22}$$

$$\Delta^2_{\mathrm{F}} = \sum_{i=1}^{N} (f_i - \tilde{f}_i)^2, \tag{5.23}$$

$$\Delta^2_{\mathrm{iso}} = (f_{\mathrm{iso}} - \tilde{f}_{\mathrm{iso}})^2, \tag{5.24}$$

where $\boldsymbol{e}_{1,i}$ is the principal eigenvector of tensor $\boldsymbol{D}_i$ with unit norm.

5.4.2 Synthetic Phantom and *In Vivo* Data

Validation of the methods is based on experiments performed with a synthetic phantom and *in vivo* data in both healthy subjects and children with autism. The synthetic phantom, as previously used in [100] and [126], consists of a $16 \times 16 \times 16$ multi-fascicle model containing an isotropic compartment and 1, 2 or 3 tensors with various properties (Fig. 5.5). S_0 is set to 400. This phantom enables the generation of synthetic DWI at different noise levels with different gradient directions and b-values.

In vivo DWI were acquired on a Siemens 3T Trio scanner with a 32 channel head coil using the CUSP-45 gradient sequence [102]. This sequence includes 30

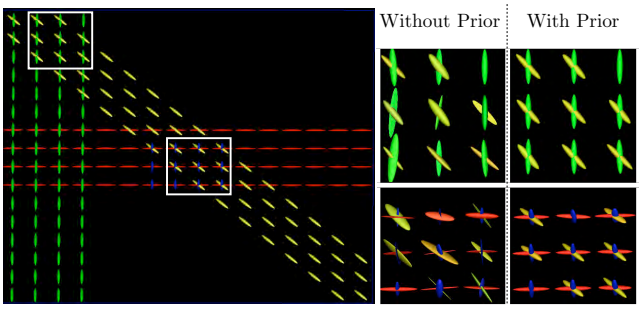


Figure 5.4: Visualization of the increased accuracy brought by the population-informed prior on synthetic data. (Left) Incorporating the prior in the estimation significantly improves the accuracy of the estimated model under the three simulated scenarios and for all five comparison metrics (distributions are shown for 20 datasets simulated for each set of parameters). (Right) The better accuracy mostly affects the diffusion properties of tensors and not their directions, as predicted by equation (5.2).

diffusion-encoding gradients on a shell at $b = 1000$ s/mm^2 and 15 extra gradients in the enclosing cube of constant TE with b-values up to 3000 s/mm^2. Eddy current distortion was minimized using a twice-refocused spin echo sequence [90]. Other acquisition parameters were set to FOV= 220mm, matrix= 128×128, number of slices=68, resolution = $1.7 \times 1.7 \times 2$mm^3. Data acquisition was conducted using a protocol approved by the Institutional Review Board (IRB). Images were acquired for 31 healthy controls and 10 children diagnosed with a syndromic form of autism spectrum disorders (ASD) related to Tuberous Sclerosis Complex (see details in Appendix B). No significant difference in age and gender were observed between the groups. A total of 13 healthy controls were used to construct the population-informed prior and the remaining 18 healthy controls as well as the children with ASD were used to test the performance of the estimation method.

5.4.3 Synthetic Phantom Experiment

In this experiment, we evaluated the impact of noise, group differences, and registration error on the models estimated with and without the population-informed prior. We constructed the prior from 20 sets of 95 DWI (5 at b =0 s/mm^2, 30 at b =1000 s/mm^2, 30 at b =2000 s/mm^2, and 30 at b =3000 s/mm^2) simulated by the synthetic phantom under a Rician noise with squared scale

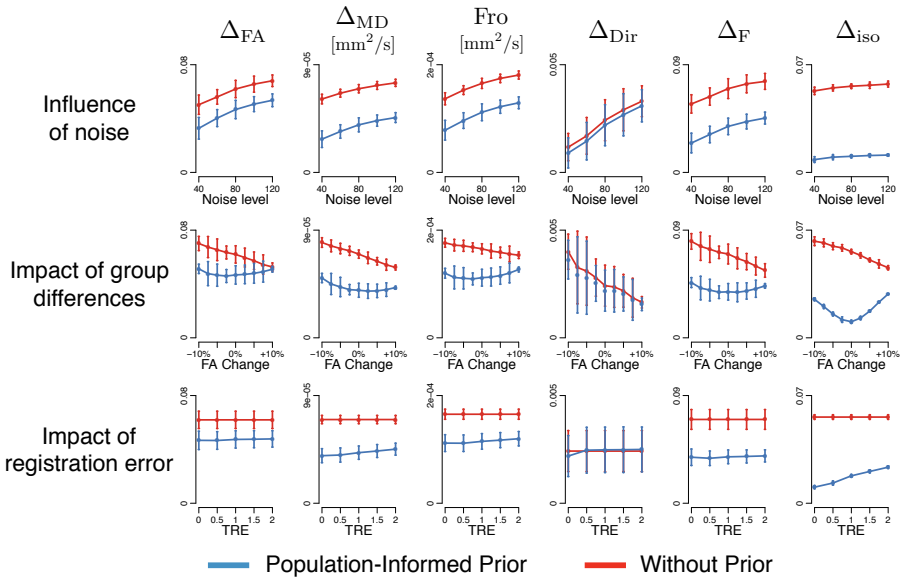


Figure 5.5: Evolution of the six comparison metrics under the influence of noise, group differences and registration error for the estimation of the synthetic phantom. Significantly better results were obtained by utilizing the population-informed prior. The directions are not affected by the absence of a population-informed prior, as predicted by Equation (5.2).

$\sigma^2 = 80$. This prior was then used to estimate multi-fascicle models from 30 DWI simulated at a single non-zero b-value (b =1000 s/mm^2) and 5 DWI at b =0 s/mm^2. The estimation was performed in three different scenarios: (i) under the influence of different levels of Rician noise, (ii) under the simulation of group differences and (iii) under the simulation of registration errors. Each estimation was repeated 20 times, each time with 35 DWI simulated to test the performance of the estimation. In these experiments, no spatial prior was used (*i.e.* $\sigma_s \to \infty$ in Eq. (5.18)). The parameter σ^2_{noise} in Experiments (ii) and (iii) was set to 80. Summary statistics for the comparison metrics are reported in Table 5.1.

a) Influence of noise

The influence of noise level was assessed by simulating Rician noise of squared scale varying from 40 to 120 by steps of 20. As depicted in Fig. 5.5, the

Table 5.1: Average improvement achieved by utilizing the population-informed prior in the estimation of the synthetic phantom (in percents of the value obtained without the prior).

Experiment \ Metric	Δ_{FA}	Δ_{MD}	Fro	Δ_{Dir}	Δ_{F}	Δ_{iso}
Influence of noise	26.2%	45.0%	34.1%	12.2%	46.9%	81.7%
Group differences	20.4%	38.7%	29.3%	8.3%	39.6%	64.0%
Registration error	23.6%	38.7%	29.9%	-2.4%	44.6%	66.5%

population-informed prior significantly improves the accuracy of the estimation for the six comparison metrics (one-tail paired t-test: $p < 10^{-10}$). The difference is less pronounced for Δ_{Dir}, since directions are not affected by the ill-posedness problem, as predicted by Equation (5.2).

b) Impact of group differences

Estimating multi-fascicle models from patients' data implies that the prior (built in healthy controls) may be offset in terms of fascicle properties. To simulate the impact of these group differences, we offset the FA of the tensors in the phantom by -10% to +10% with steps of 2.5%. The datasets of DWI used for testing were simulated from these offset phantoms while the population-informed prior was left unchanged. The results, depicted in Fig. 5.5, show that the population-informed prior significantly improves the accuracy of the estimation for the six metrics (one-tail paired t-test: $p < 10^{-5}$ for Δ_{Dir} and $p < 10^{-10}$ for the other metrics). As for the previous experiment, the improvement in Δ_{Dir} is less pronounced than improvements in other metrics.

c) Impact of registration errors

The prior needs to be registered to the subject whose multi-fascicle model is to be estimated. To simulate the impact of registration errors on the estimated multi-fascicle model, we applied random deformations to the prior. The three components of the deformation at each voxel were sampled from an uniform distribution between 0 and $a = 0, 0.5, 1, 1.5$, and 2 voxels. The results, depicted

in Fig. 5.5, show that the population-informed prior significantly improves the accuracy of the estimation for the five metrics other than Δ_{Dir} (one-tail paired t-test: $p < 10^{-5}$). For Δ_{Dir}, a small decrease in accuracy is observed because estimates without prior are not affected by registration errors. The angular magnitude of this decrease, however, is 0.048°, which is negligible in practice.

These experiments demonstrate that the population-informed prior significantly improves the estimation of multi-fascicle models, even when registration errors and group differences occur. In the next section, we will see that the same is true for *in vivo* data.

5.4.4 *In vivo* Data Experiment

For *in vivo* data, there is no absolute ground truth. However, the CUSP acquisition sequence allows us to estimate the multi-fascicle model from the *full set* of DWI (5 DWI at $b = 0$ s/mm^2, 30 DWI on a shell at $b = 1000$ s/mm^2 and 15 DWI on the enclosing cube with b-values up to 3000 s/mm^2) and from the *restricted set* of DWI at a single non-zero b-value (5 DWI at $b = 0$ s/mm^2, 30 DWI on a shell at $b = 1000$ s/mm^2). The multi-fascicle models estimated with the *full sets* of DWI are not affected by the ill-posedness problem. We consider them as ground truths for the experiments and we compare estimations from the *restricted sets* to it. In this section, we optimize the parameters for the estimation and we compare our approach with the results obtained with (i) no prior, (ii) a spatial prior only and (iii) a fixed response function as proposed in the literature [132]. A summary of the performance of all methods is depicted at the end of this section, in Fig. 5.9.

a) Optimization of the weight ratio

Two parameters need to be fixed for the estimation of the maximum *a posteriori* in equation (5.18): the weight associated with the likelihood, $1/2\sigma^2_{\mathrm{noise}}$, and the weight associated with the spatial prior, $1/2\sigma^2_{\mathrm{s}}$. The value of $\sigma^2_{\mathrm{noise}}$ could be estimated from the data directly assuming ergodicity. However, the noise in diffusion images is only approximately ergodic [67] and we therefore prefer to estimate the optimal weights based on comparisons to the ground truth. When no population-informed prior is used, only the ratio of these weights, $\sigma^2_{\mathrm{noise}}/\sigma^2_{\mathrm{s}}$, needs to be fixed. The ground truths (based on the full sets of DWI) were estimated using a ratio $\sigma^2_{\mathrm{noise}}/\sigma^2_{\mathrm{s}} = 0.4$ as we observed that it yields smooth multi-fascicle models while preserving important structures of the white matter.

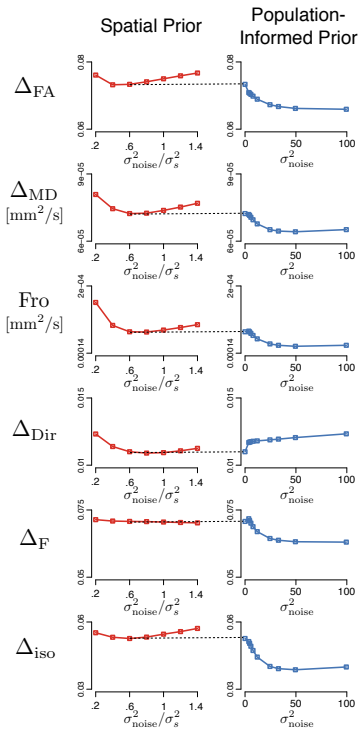


Figure 5.6: Additional increase in accuracy when the population-informed prior is introduced after optimizing the weight of the spatial prior. (Left) Evolution of the six comparison metrics when only a spatial prior is used, for different weight ratios $\sigma^2_{\text{noise}}/\sigma^2_s$. (Right) Evolution of the six comparison metrics when both a spatial prior and the population-informed prior are used, for different values of the noise level σ^2_{noise} with$\sigma^2_{\text{noise}}/\sigma^2_s = 0.6$. The use of a population-informed prior significantly improves the estimation accuracy for all five non-directional metrics. The evolution is smooth and monotonic with respect to σ^2_{noise} so that moderate errors in the estimation of σ^2_{noise} do not have a large impact on the result.

We first estimated multi-fascicle models from single b-value data using a spatial priors only, removing the first term of Equation (5.18). Because of the ill-posedness problem, the optimal weight ratio ($\sigma^2_{\text{noise}}/\sigma^2_s$) may differ from that used to build the ground truth. We therefore estimated multi-fascicle models with weight ratios of $\sigma^2_{\text{noise}}/\sigma^2_{\text{s}} = 0.2$, 0.4, 0.6, 0.8, 1.0, 1.2 and 1.4. The evolutions of the six similarity metrics with the weight ratio are depicted in Fig. 5.6. These results show that the best overall results are obtained for a

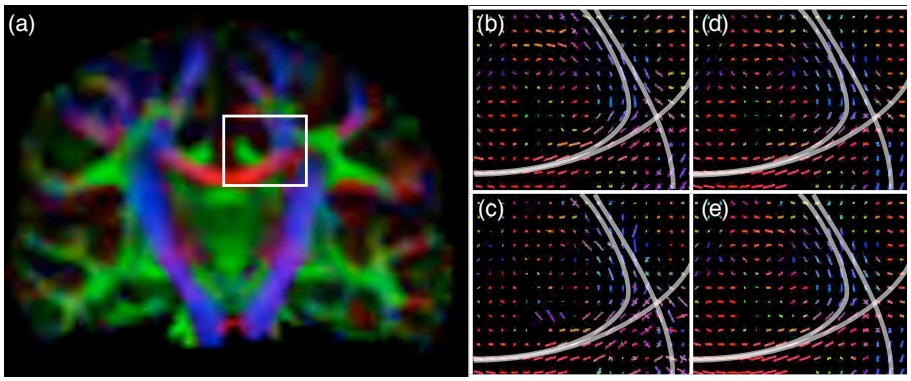


Figure 5.7: In terms of direction, the spatial prior and the population-informed prior perform equally well. (a) Coronal slice of a subject's color FA highlighting the centrum semiovale. (b) Directions estimated with the DWI at multiple b-values. Directions estimated with DWI at a single b-value (c) without prior, (d) with a spatial prior only, (e) with both spatial and population-informed priors.

weight ratio of 0.6, which is optimal for Δ_{MD}, and Δ_{iso} and close to optimal for Δ_{Dir}, and Fro.

b) Comparison between population-informed and a spatial priors

As explained in Section 5.2.2, typical spatial priors alone do not solve the ill-posedness problem in the estimation of multi-fascicle models from single b-value data. In this experiment, we explored whether the population-informed prior increases the accuracy of the estimation. We kept $\sigma^2_{\mathrm{noise}}/\sigma^2_{\mathrm{s}}$ to its optimal value of 0.6 and we estimated multi-fascicle models for $\sigma^2_{\mathrm{noise}} = 0$, 3.7, 4.5, 5.75, 7.83, 12, 24.5, 32.83, 49.5 and 99.5. These values stem from our implementation which consists in fixing the trade-off between the population-informed prior (assigning it a weight w) and the likelihood (assigning it a weight $(1-w)$), with linearly increasing w (0, 0.88, 0.9, 0.92, 0.94, 0.96, 0.98, 0.985, 0.99, 0.995). When $\sigma^2_{\mathrm{noise}} = 0$, the population-informed prior is mute and the estimation corresponds to the estimation with the spatial prior only (dashed lines in Fig. 5.6).

Results in Fig. 5.6 demonstrate that the population-informed prior improves the accuracy of the estimation for all metrics except for Δ_{Dir}. Optimal performances occur for $\sigma^2_{\mathrm{noise}} = 49.5$ (corresponding to a trade-off $w = 0.99$). Importantly, the performances improve monotonically and smoothly with σ_{noise} so that slightly suboptimal estimates of the noise level will not dramatically

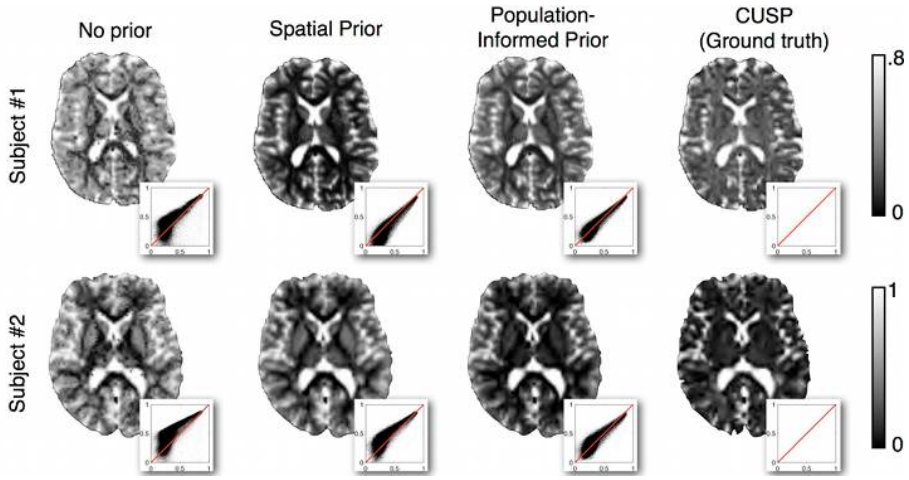


Figure 5.8: The maps of isotropic fraction show that population-informer prior results in maps with a lesser bias than those obtained with the spatial prior alone. The spatial prior leads to maps that are visually plausible but affected by a global bias (negative for Subject 1 and positive for Subject 2). This bias is due to the arbitrary choice of a model and the propagation of this choice through neighboring voxels. By contrast, the population-informed prior results in maps of the isotropic fraction that are closer to the ground truth. Scatter plots next to every image depicts the joint distribution of isotropic fraction in the ground truth (x-axis) and in the estimation (y-axis) highlighting the bias observed in estimations with a spatial prior.

affect the results. For $\sigma^2_{\text{noise}} = 49.5$, the improvement in the group of healthy controls ranges from 7.8% for Fro to 28.8% for Δ_{iso}. One-tail paired t-tests indicate that these improvements are significant ($p < 0.005$ for all five metrics). Importantly, similar results were obtained for the estimation of multi-fascicle models in patients with autism (Fig. 5.9). Improvements were observed for all five non-directional metrics with improvements ranging from 3.2% for Δ_{FA} to 18.2% for Δ_{iso}. One-tail paired t-tests indicate that these improvements are significant ($p < 0.005$ for all metrics except $p = 0.05$ for Δ_{F}).

To better appraise the improvement brought by the population-informed prior, Fig. 5.10 depicts the spatial distribution of the improvement in estimation error (as measured by Δ_{F}). The improvement is widespread and is more important in areas with crossing fascicles than in areas with a single fascicle (in those areas, estimating the model with single b-value data is not ill-posed). Importantly, even in areas with low contrast in diffusion-weighted images, the

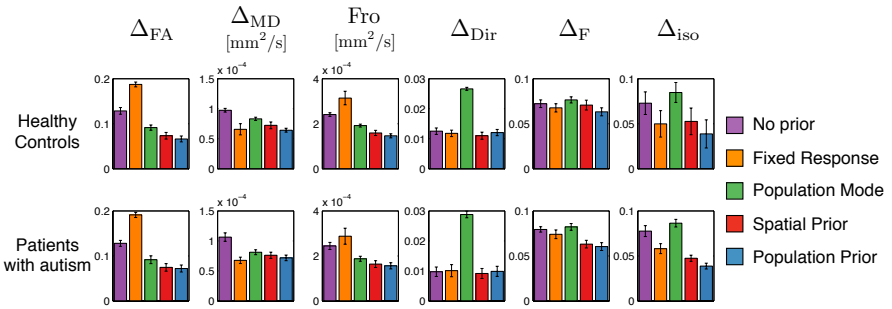


Figure 5.9: Accuracies of the estimation of a multi-fascicle model from single-shell HARDI data with the five methods under comparison. (1) estimation without prior, (2) estimation with a fixed response (fixed eigenvalues), (3) estimation by setting all parameters to the mode of the population, (4) estimation with a spatial prior only and (5) estimation with a population-informed prior. Results with the population-informed prior have overall better performances than all other methods.

estimation remains better with a population-informed prior, probably owing to the reliability of the proposed registration in those areas [128].

The slightly worse Δ_{Dir} observed with the population-informed prior corresponds to an angular difference of 0.4°, which is negligible in most practical applications. The equivalent accuracies in Δ_{Dir} is explained by Equation (5.2) which implies that directions are not affected by the ill-posedness problem. Figure 5.7(d-e) illustrate the difference in directions between estimations with and without the population-informed prior.

Consistently for both synthetic data and *in vivo* data in healthy controls and patients with autism, the isotropic fraction is the metric most beneficial of the introduction of the population-informed prior. The reason for this dramatic improvement can be understood from Equation (5.3). Equation (5.3) implies that models with large γ_i have the i-th fascicle close to an isotropic tensor (due to the offset of all eigenvalues by $\log(\gamma_i)/b$) and a volumetric fraction $(\gamma_i f_i)$. Due to noise, those fascicles may be associated with an isotropic compartment, thereby increasing the isotropic fraction. Conversely, a multi-fascicle model that has a large isotropic fraction may be misinterpreted as one with a small isotropic fraction and a tensor with a low fractional anisotropy. The choice between the two models is arbitrary. The spatial prior only imposes that a consistent choice

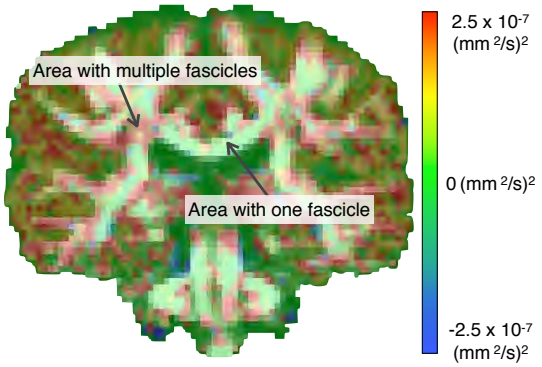


Figure 5.10: Spatial distribution of the improvement in accuracy brought by the population-informed prior. Difference in Δ_F^2 between the estimation without the prior and with the population-informed prior. Values larger than zero indicate voxels where the population-informed prior improves the estimation. As indicated by the arrows, the improvement is more important in areas with crossing fascicles. In areas without crossing fascicles, the estimation without prior is not ill-posed and results in models with a similar accuracy as those estimated with a population-informed prior.

be made between adjacent voxels, and will lead to globally biased isotropic fractions if the wrong model is picked. This phenomenon is observed in Fig. 5.8, where, in the first row, a negative bias is observed throughout the map of isotropic fraction, and, in the second row, a positive bias is observed. These global biases may mislead population studies as will be demonstrated in the next section.

c) Comparison with a fixed response function

One way to remove the ill-posedness problem is to fix *a priori* the tensor eigenvalues and to focus on optimizing the directions [132]. We assessed this strategy by estimating, for each subject individually, a typical response function, *i.e.* a tensor that represents the typical fascicle within the individual's brain. This response was estimated in a similar fashion as [130]. First, we selected the 300 voxels with the highest FA. We assumed that a single fascicle was present in these voxels. A tensor was then estimated in each of these voxels and we computed its principal eigenvector. This eigenvector was used to re-orient the fascicle with the z-axis. The rotation was then applied to all DWI voxels and

the corresponding gradients were reoriented accordingly. Finally, the typical response tensor was estimated from all 300 voxels of all the reoriented DWI. This response function was plugged into the maximum *a posteriori* estimate as a hard constraint on the tensor eigenvalues. Results in Fig. 5.9 demonstrate that such a fixed response strategy fails to accurately estimate a multi-fascicle model. In particular, the estimated FA is on average worse than that estimated without any prior.

d) Assessment of the shrinkage towards the population mean

Bayesian parameter estimation leads to models that are closer to the mean of the population than they would be without a prior. One could be concerned that this *shrinkage*, if too severe, would jeopardize the usefulness of the models for population studies. To assess the extent of the shrinkage created by the population-informed prior, we compared the models estimated with our method with the mode of the population-informed prior. If the shrinkage is large, the models would be very close to this mode.

Results in Fig. 5.9 show that the estimation accuracies of the population mode are significantly worse than those obtained with the proposed population-informed prior. The departure of our results from those obtained with the population mode shows that the Bayesian inference operates in a regime that trades off information from both the prior and the likelihood. Remarkably, our estimation performs equally good than the spatial prior estimate in terms of Δ_{Dir} despite the poor results obtained by the population mode.

These results suggest that our maximum *a posteriori* expression correctly integrate information from the likelihood, the spatial prior and the population-informed prior. In particular, if the population-informed prior indicates a large dispersion in fascicle orientations in one voxel, the corresponding variance $\tilde{\sigma}_i$ will be large and more weight will be put on the spatial prior and the likelihood.

5.5 Application to Population Studies

Multi-fascicle models enable population studies of various properties of the brain microstructure. In this section, we propose two population studies of the brain microstructure comparing patients with autism spectrum disorders to neurotypical controls. The increased accuracy brought by the population-informed prior translates –as we will see– into a more reliable inference of group

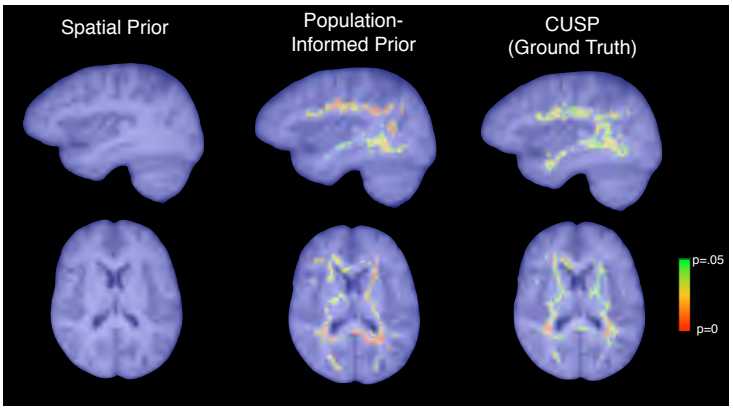


Figure 5.11: Maps of the p-values of isotropic diffusion analysis, thresholded at $p < 0.05$ after correction for family-wise error rate. The population-informed prior leads to inference of group differences that are remarkably close to those obtained with the data at several b-values (CUSP). A spatial prior alone fails to detect any difference due to biases in estimations of the isotropic fraction.

differences. All results presented in this section were based on $\sigma^2_{\text{noise}}/\sigma^2_s = 0.6$ and $\sigma^2_{\text{noise}} = 49.5$.

5.5.1 IDA in the white matter

Isotropic diffusion analysis [126] allows whole-brain inspection of differences in isotropic fraction f_{iso}. Excessive f_{iso} relates to the presence of neuroinflammation and edema among others [137, 82]. We performed isotropic diffusion analysis to compare the group of patients with ASD to the group of neurotypical controls. The statistics of interest was the t-statistics transformed with threshold-free cluster enhancement [114]. This statistics is more sensitive to group differences occurring in clusters of neighboring voxels, without relying on the (typically arbitrary) choice of a threshold. The p-values were corrected for family-wise error rate using exact tests with 5000 permutations.

Widespread regions of significantly higher isotropic fraction were observed in patients with autism, as compared to controls (see the maps of corrected p-values thresholded at $p < 0.05$ depicted in Fig. 5.11). These regions correspond to the arcuate fasciculi, the corpus callosum and cortico-spinal tracts. These widespread differences are consistent with finding of widespread white matter abnormalities in patients with tuberous sclerosis complex [91] as well as

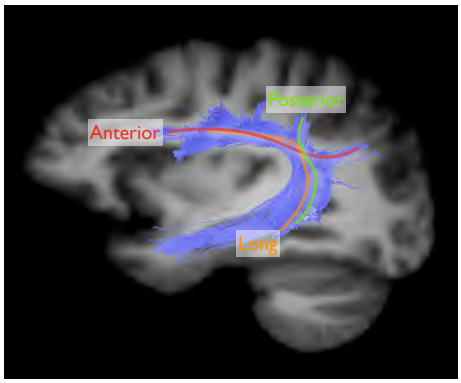


Figure 5.12: The dorsal language circuit is composed of three main fascicles. These fascicles are thought to connect Broca's area in the frontal lobe, Wernicke's area in the temporal lobe, and Geschwind's territory in the parietal lobe.

recent findings of impaired language pathways [63] and loss of corpus callosum integrity [83] in children with syndromic autism. However, those recent results were based on single-tensor DTI analysis which conflates differences in fascicle properties (such as the FA) and differences in isotropic fraction. DTI studies therefore cannot separate axon/myelin injury from increased cellularity associated with neuroinflammation [137]. Our findings of increased isotropic fraction in children with autism suggest that previous findings of impaired white matter in children with syndromic autism may be in part due to increased cellularity or edema that may point to a neuroinflammatory process in those regions, as suggested by post-mortem studies of autism [135].

Remarkably, the differences found with the population-informed prior match the differences found with the ground truth models. By contrast, no significant differences were detected with the spatial prior, probably owing to the arbitrariness of the choice of a model from the manifold of equivalent models as explained in the previous section and depicted in Fig. 5.8. Changes in volumetric fractions across the manifold of equivalent models are compensated by changes in fascicle properties. Group difference in volumetric fraction may therefore be incorrectly interpreted as differences in fascicle properties. In the next experiment, we will see that this misleading effect occurs in our population when spatial priors are being used alone.

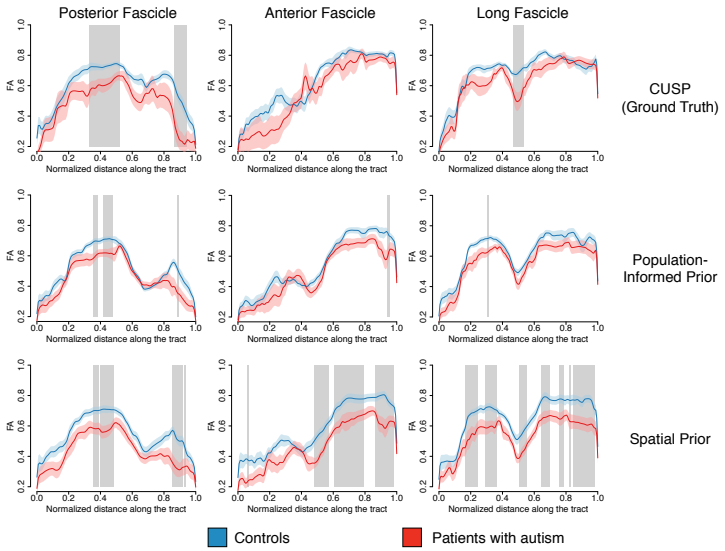


Figure 5.13: Results of fascicle-based spatial statistics in the dorsal language circuit. The mean FA along fascicles are shown as dark lines and their confidence intervals are colored shadowed curves. Segments of significant differences are identified by grey rectangles. Estimations with a population-informed prior detect most differences observed with the ground true and does not lead to excessive false discovery rates. By contrast, estimations with a spatial prior generate false positives at a high rate, covering the anterior fascicle and the long fascicle by over 35%.

5.5.2 FBSS in the dorsal language circuit

Fascicle-based spatial statistics (FBSS) enables to analyze white matter properties along individual fascicles in the presence of crossing fascicles. As explained in Chapter 3, FBSS proceeds in four steps. First, all multi-fascicle models are registered to a multi-fascicle atlas. Second, tractography is performed on the atlas and specific fascicles are selected. Third, properties of the fascicles are extracted from each subject by selecting at every point the tensor most aligned with the fascicle. Fourth, statistical analysis is performed on the fascicle properties by computing t-tests at every point of the fascicle and by performing permutation tests on the threshold-free cluster enhanced t-statistics, similar to [113], to control for multiple comparisons.

We performed FBSS for the FA of the dorsal language circuit (Fig. 5.12). This set of pathways involved in language is organized around three main

fascicles [63]: the posterior, anterior and long fascicles. These fascicles are thought to connect Broca's area in the frontal lobe, Wernicke's area in the temporal lobe, and Geschwind's territory in the parietal lobe. FBSS was performed on these three fascicles individually.

Along the posterior fascicle, two clusters of significantly lower FA were observed ($p < 0.05$ after correction, Fig. 5.13). Similar findings were obtained with estimations using the population-informed prior and the spatial prior alone. Along the anterior fascicle, no significant difference was observed (Fig. 5.13). Estimations with the population-informed prior only found a small false positive cluster covering 1.4% of the anterior fascicle. With the spatial prior, three large false positive clusters were observed, covering 39.4% of the fascicle. This high prevalence of false discoveries with the spatial prior can be explained by its inability to detect the significant differences in isotropic fractions. These differences are incorrectly reflected on the FA of the fascicle. A similar effect is observed for the long fascicle. A cluster of significantly lower FA was observed using the ground truth. This cluster was not detected by the population-informed prior and only partially detected (30% overlap) by the spatial prior which also incorrectly detects other clusters covering 38.9% of the fascicle.

5.6 Summary

Multi-fascicle models cannot be estimated from conventional single-shell HARDI data alone because a manifold of different models are equally compatible with the measurable diffusion signals, making them indistinguishable. Estimation with a spatial prior arbitrarily selects a model from the manifold of equivalent models and thereby conflates differences in fascicle properties and differences in volumetric fractions. This conflation misleads conclusions of population studies in a similar way as single tensor DTI. With single b-value data, spatial priors alone therefore fail to harness the novel insight provided by multi-fascicle models. By contrast, we showed that the introduction of a population-informed prior generates more accurate multi-fascicle models by removing the ill-posedness problem from the estimation. These models, in turn, lead to a more reliable inference in population studies, distinguishing differences in isotropic fractions from differences in fascicle properties. The population-informed prior therefore enables novel investigations of properties extracted from multi-fascicle models with single-shell diffusion data. This method thus opens new opportunities for population studies with the large number of available clinical diffusion images.

Chapter 6

Beyond Multiple Fascicles: Distribution of Microstructural Environments

The multi-fascicle model defined in (2.5) accounts for the presence of multiple fascicles and one or several extra-axonal spaces. However, it makes the assumption that, within a particular compartment, the diffusion of all water molecules can be described by the same tensor (isotropic or anisotropic). In this chapter, we challenge this assumption.

We describe our recently developed DIAMOND model that accounts for the heterogeneity of the signal contributions from different groups of water molecules (called *spin packets* in reference to the spin of the protons that generates the diffusion signal) within a single compartment [98]. We then show, using the method described in Chapter 4, that the DIAMOND has a lower generalization error than the multi-fascicle model in specific parts of the white matter. We subsequently provide a geometric insight into the space of DIAMOND models that generate the diffusion signals. This insight sheds light on the required DWI acquisition and provides a method to assess the uncertainty in the model parameters.

6.1 Distribution of Anisotropic Microstructural Environments in Diffusion Imaging (DIAMOND)

In this section we explain how, mathematically, the DIAMOND model represents the heterogeneity occurring at various scales in the diffusion of water molecules. While the DIAMOND model is a framework (rather than a specific model) to represent this heterogeneity, we also propose a specific form of the model in which the major sources of diffusivity are explicitly identified and labelled and in which variations of such diffusivities are considered.

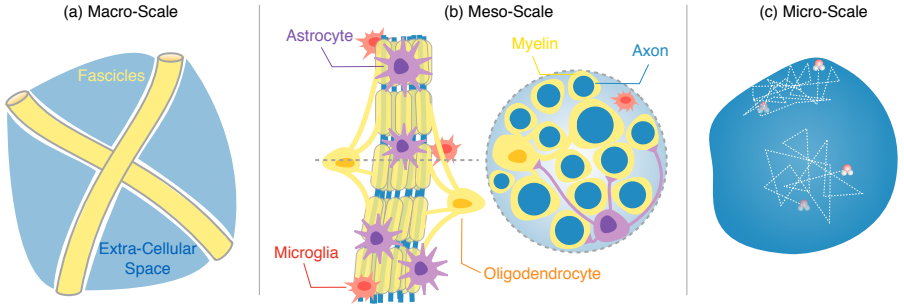


Figure 6.1: Heterogeneity of the spin packets at different scales. The different spin packets generate heterogeneous diffusion signal. This heterogeneity occurs at three spatial scales. (a) At the macro-scale, heterogeneity is due to the presence of multiple fascicles, an extra-cellular space and partial volumes of CSF and/or grey matter. (b) At the meso-scale, heterogeneity is due to the various shapes and sizes of axons, their degree of myelination and the presence of other cells including astrocytes, microglia and oligodendrocytes. (c) At the micro-scale, heterogeneity is due to the distance from water molecules to the cell membranes, as well as intra-cellular obstacles to diffusion.

6.1.1 Heterogeneity in the Diffusion of Water Molecules

Heterogeneity of the diffusion of water molecules arises at different scales. The multi-fascicle models has been defined to capture the heterogeneity at the *macro-scale* (Fig. 6.1a). At this scale, heterogeneity in the diffusion of spin packets arises from the presence, within one voxel, of multiple compartments (crossing fascicles, an extra-axonal space or a partial volume of CSF or gray matter). These are large scale microstructural environments (LSME).

At the so-called *meso-scale*, each LSME presents an heterogeneous microstructure due to the presence of glial cells and axons with various radii, orientations and degree of myelination (Fig. 6.1b). The diffusion signal of water molecules interacting with an homogeneous portion of this microstructure is well modeled by a mono-exponential decay leading to a tensor representation. However, the signal arising from the whole LSME may deviate from a mono-exponential decay due to the heterogeneity in the interaction of spin packets. At the *micro-scale*, other biophysical mechanisms such as heterogenities of cellular structure and the proximity of spin packets to the cell membranes may further impact the diffusion signal (Fig. 6.1c).

The DIAMOND model proposes to account for the various sources of het-

erogeneity in a phenomenological manner, rather than a biophysical way. A phenomenological approach has the advantage of simultaneously capturing all sources of diffusion heterogeneity without adding many extra parameters. Biophysical models of the heterogeneity would require separate models for each source of heterogeneity, resulting in a consequential increase in the number of parameters. However, unlike biophysical models, interpretation of phenomenological parameters in terms of microstructural features may require additional assumptions because of the one-to-many correspondence between the model parameters and the sources of heterogeneity.

6.1.2 DIAMOND: a Framework to Model Heterogeneity

To model the heterogeneity of the diffusion of spin packets, we consider that the signal attenuation due to the diffusion of one spin packet is represented by a single tensor response $e^{-b\boldsymbol{g}^T\boldsymbol{D}\boldsymbol{g}}$ but that different spin packets may be characterized by different tensors $\boldsymbol{D}$. The fraction of water molecules being characterized by a same tensor $\boldsymbol{D}$ is a matrix-variate distribution $p(\boldsymbol{D})d\boldsymbol{D}$ and the diffusion signal in one voxel is obtained by integrating the contributions from each spin packet:

$$S(b,\boldsymbol{g}) = S_0 \int_{\mathcal{S}_3^+} p(\boldsymbol{D}) e^{-b\boldsymbol{g}^T\boldsymbol{D}\boldsymbol{g}} d\boldsymbol{D}, \tag{6.1}$$

where $\mathcal{S}_3^+$ is the space of 3×3 symmetric positive-definite matrices. The single tensor model of (2.3) with tensor $\boldsymbol{D}_0$ is a particular case of (6.2) for which $p(\boldsymbol{D})$ is an indicator function $\mathbb{1}_{(\boldsymbol{D}=\boldsymbol{D}_0)}$ and the multi-fascicle model of (2.5) is a particular case for which $p(\boldsymbol{D})$ is a sum of indicator functions $\sum_{i=1}^{N} \mathbb{1}_{(\boldsymbol{D}=\boldsymbol{D}_i)}$.

Generally, the diffusion signal in each LSME can be characterized by a peak shaped-distribution over matrices $\boldsymbol{D}$. The diffusion signal arising in one voxel from all LSME together is thus characterized by a mixture of the peak-shaped distributions:

$$S(b,\boldsymbol{g}) = S_0 \sum_{i=1}^{N} f_i \int_{\mathcal{S}_3^+} p_{\boldsymbol{\theta}_i}(\boldsymbol{D}) e^{-b\boldsymbol{g}^T\boldsymbol{D}\boldsymbol{g}} d\boldsymbol{D}. \tag{6.2}$$

In this equation the mixture weights f_i play the same role as the volumetric fraction in the multi-fascicle model of (2.5). The distributions $p_{\boldsymbol{\theta}_i}(\boldsymbol{D})$ are parameterized by a set $\boldsymbol{\theta}_i$ of parameters whose value may differ between LSME.

A natural peak-shaped distribution for $n \times n$ symmetric positive definite

matrices is the matrix-variate Gamma distribution [46, 71, 106]:

$$p_{\kappa,\boldsymbol{\Sigma}}(\boldsymbol{D}) = \frac{|\boldsymbol{D}|^{\kappa-(n+1)/2}}{|\boldsymbol{\Sigma}|^{\kappa}\Gamma_n(\kappa)} e^{-\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{D})},$$

where $\kappa > (n-1)/2$ is the concentration parameter, $\boldsymbol{\Sigma} \in \mathcal{S}_n^+$ is the scale parameter, $|.|$ is the matrix determinant and Γ_n is the multivariate gamma function.

The expectation of the matrix-variate Gamma distribution is $\boldsymbol{D}_m = \kappa\boldsymbol{\Sigma}$ and the concentration parameter κ is larger if matrices $\boldsymbol{D}$ are more concentrated around the mean $\boldsymbol{D}_m$. In the DIAMOND model [98], the heterogeneity within each LSME is characterized by a matrix-variate Gamma distribution $p_{\kappa_i,\boldsymbol{\Sigma}_i}(\boldsymbol{D})$:

$$S(b,\boldsymbol{g}) = S_0 \sum_{i=1}^{N} f_i \int_{\mathcal{S}_3^+} p_{\kappa_i,\boldsymbol{\Sigma}_i}(\boldsymbol{D}) e^{-b\boldsymbol{g}^T\boldsymbol{D}\boldsymbol{g}} d\boldsymbol{D}.$$

Importantly, the integrals in this expression are Laplace transforms of $p_{\kappa_i,\boldsymbol{\Sigma}_i}(\boldsymbol{D})$ which have a known analytical expression [56] and the signal model S reads (with $\boldsymbol{D}_i \hat{=} \kappa_i\boldsymbol{\Sigma}_i$):

$$S(b,\boldsymbol{g}) = S_0 \sum_{i=1}^{N} f_i \left(1 + \frac{b\boldsymbol{g}^T\boldsymbol{D}_i\boldsymbol{g}}{\kappa_i}\right)^{-\kappa_i}. \tag{6.3}$$

The matrix $\boldsymbol{D}_i$ in this equation represents the average mono-exeponential decay of all spin packets in the i-th LSME. The heterogeneity in the spin packets local microenvironment is accounted for by the concentration parameter κ_i. In particular, for $\kappa_i \to \infty$, the DIAMOND model of (6.3) is equal to the multi-fascicle model of (2.5).

One of the LSME in (6.3) typically represents the extra-axonal space with an isotropic tensor $\boldsymbol{D}_0 = d_{\mathrm{iso}} I_3$. In this LSME, $\boldsymbol{D}_0$ is the average diffusion of water molecules in the extra-axonal space. Using a finite concentration parameter κ_{iso} enables the representation of a spectrum of isotropic diffusivities (Fig. 6.2). While the mean diffusivity in this compartment is given by $\boldsymbol{D}_0 = d_{\mathrm{iso}} I_3$, the mode equals $\boldsymbol{D}_{\mathrm{mode}} = d_{\mathrm{iso}}(1 - \frac{2}{\kappa_{\mathrm{iso}}}) I_3$ because the matrix-variate Gamma distribution is not symmetric around its mode (Fig. 6.2). This diffusivity spectrum may reflect isotropic diffusion in the extra-cellular space as well as isotropic diffusion in glial cells (Fig. 6.1b).

To account for the heterogeneity of the diffusion process governing different spin packets, the DIAMOND model requires one extra parameter per compartment (κ_i). To assess the added value of this parameter in representing

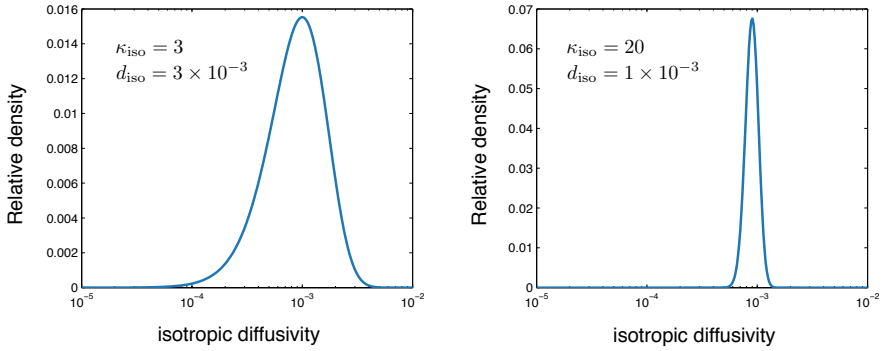


Figure 6.2: Spectrum of isotropic diffusivities with the DIAMOND model. The isotropic large scale microstructural environment encompasses a spectrum of diffusivities that may represent diffusion within different glial cells and in the extra-cellular space. The plotted curves are relative densities of the isotropic diffusivities for two different values of κ. The mode of the distribution is given by $d_{\text{iso}}\left(1 - \frac{2}{\kappa_{\text{iso}}}\right)$.

the diffusion signal of each LSME, we will compare, in the next section, the generalization error of the DIAMOND model with that of the multi-fascicle model of (2.5).

6.1.3 A Specific DIAMOND Model

In models of the microstructure of the brain, it is possible to identify a number of compartments, each of which may be suitably described as a continuous mixture of diffusion tensors through a matrix-variate gamma distribution. For example, isotropic free water and isotropic restricted water diffusion can each be represented by a matrix-variate gamma distribution with an isotropic mode. The diffusivity of restricted intra-axonal molecules in a fascicle, and the surrounding hindered extra-axonal molecules can each be represented by a matrix-variate gamma distribution, and the similarity of orientation between the restricted and hindered diffusivity can be accounted for by, for example, using the same eigenvectors in the mode of the distributions. In this model, a finite mixture of matrix-variate gamma distributions of diffusion tensors provides a generative model that captures isotropic free and restricted diffusivity, and describes the diffusivity of each fascicle that is present with two matrix-variate gamma distributions.

In general, whenever the tissue microstructure compartments give rise to

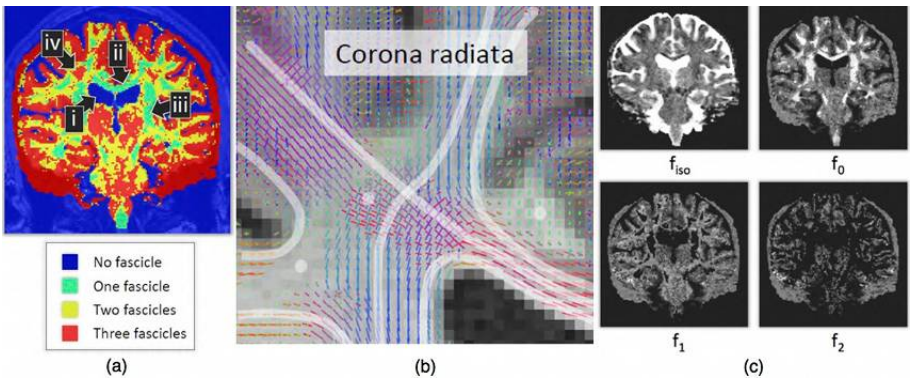


Figure 6.3: Number of fascicles detected by minimization of the generalization error using DIAMOND model (a) Map showing the number of fascicles detected. Consistently with the known anatomy, diffusion in the ventricles (i), a single fascicle is present in the corpus callosum (ii) and in the cortico-spinal tracts (iii) and up to three fascicles are found in the corona radiata (iv). (b) A zoomed-in display of fascicles orientation in the corona radiata. (c) Volumetric fractions of each compartment. (Figure adapted from [98])

heterogeneity that is described by a matrix-variate gamma distribution that is broader than we desire, it is possible to consider a mixture of matrix-variate gamma distributions to better capture the heterogeneity as a sum of more homogeneous compartments. It is then important to consider the nature of the imaging strategy that enables the identification of the model parameters, the practicality of such an imaging scheme for clinical or research application, and alternative mechanisms for identifying suitable model parameters such as using parameter values from the literature or other experiments.

6.2 Generalization Error of the DIAMOND Model

The DIAMOND model has one extra parameter by compartment (κ_i) compared to the multi-fascicle model of (2.5) to capture the heterogeneity in the diffusion of spin packets. To assess whether this extra parameter captures novel observable diffusion phenomena not captured by the multi-fascicle models, we used the model selection approach described in Chapter 4.

We acquired 395 DWI with a multi-shell HARDI sequences (gradient directions located on multiple spheres each with a specific b-value): 5 DWI at $b =$

0 s/mm^2 and 15 shells of 26 gradient directions each with $b \in [200, 3000]$ s/mm^2 by increments of 200 s/mm^2. We repeated the estimation-testing process 100 times using, at each iteration, 70% of the data for estimation and 30% for testing. Within the estimation set, the number of fascicles was estimated with the 632 bootstrap with 50 bootstrap replicates (Fig. 6.3). Using the test set at each iteration, we computed the mean squared prediction error at each voxel. The above procedure was performed with both the DIAMOND model and the multi-fascicle model of (2.5). Mean squared prediction errors of both models were then compared at each voxel.

The resulting map of difference between the mean squared prediction errors is depicted in Fig. 6.4. On average, the DIAMOND model is better than the multi-fascicle model of (2.5) with a mean improvement corresponding to 8% of the mean error. To test whether the improvement is statistically significant, one-tailed paired t-tests were used. This test requires independent measurements, which is not the case of voxels that share spatial dependence (smoothness). To increase independence between samples, the t-test was conducted by selecting only one voxel out of five along each dimension of the image, resulting in a p-value $p < 0.002$ for the test. This test indicates that DIAMOND is, on average, significantly better than the multi-fascicle model. The generalization error is mostly decreased along corpus callosum and the cortico-spinal tract (arrows in Fig. 6.4).

6.3 Acquisition and Estimation of DIAMOND Models

In Chapter 5, we showed that there is a manifold of different models that all generate the same measurable diffusion signals at a given b-value and therefore cannot be distinguished. The same is certainly true for DIAMOND models since they include multi-fascicle models as a particular case (for $\kappa_i \to \infty$, $\forall i$). The particular geometry of the manifolds at the intersection point which, for multi-fascicle models, guarantees the local unicity of the intersection of two manifolds at two different b-values, may be different for DIAMOND models. In this section, we elucidate the geometry of the manifold of equivalent DIAMOND models and demonstrate that two b-values are no more sufficient to uniquely identify the model parameters.

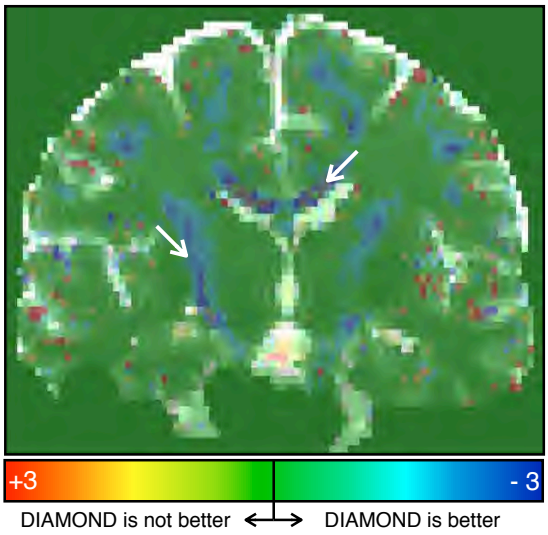


Figure 6.4: The DIAMOND model has, on average, a lower generalization error than the multi-fascicle model. The map shows the difference in generalization errors measured by cross-testing in 395 DWI. The difference is more notable in the corpus callosum (topmost arrow) and in the cortico-spinal tracts (lowermost arrow). (Figure adapted from [98])

6.3.1 Manifold of Equivalent DIAMOND Models

As for the multi-fascicle model, let us start by rewriting the signal generation model of (6.3) by multiplying each term by $\frac{\gamma_i}{\gamma_i}$:

$$\begin{aligned}
\frac{S(b,\boldsymbol{g})}{S_0} &= \sum_{i=1}^{N} \frac{\gamma_i}{\gamma_i} f_i \left(1 + \frac{b}{\kappa_i}\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}\right)^{-\kappa_i} \\
&= \sum_{i=1}^{N} \gamma_i f_i \left(\gamma_i^{\frac{1}{\kappa_i}} + \gamma_i^{\frac{1}{\kappa_i}} \frac{b}{\kappa_i}\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}\right)^{-\kappa_i} \\
&= \sum_{i=1}^{N} \gamma_i f_i \left(1 + \frac{b}{\kappa_i}\boldsymbol{g}^T \left\{\gamma_i^{\frac{1}{\kappa_i}} \boldsymbol{D}_i + \frac{\kappa_i}{b}\left[\gamma_i^{\frac{1}{\kappa_i}} - 1\right]\boldsymbol{I}\right\} \boldsymbol{g}\right)^{-\kappa_i} . \quad (6.4)
\end{aligned}$$

The DIAMOND model with parameters $\{f_i, \kappa_i, \boldsymbol{D}_i\}_{(1\leq i\leq N)}$ therefore generates the same diffusion signal at a given b-value as all the DIAMOND models with

parameters $\{f'_i, \kappa'_i, \boldsymbol{D}'_i\}_{(1\leq i\leq N)}$ such that:

$$f'_i = \gamma_i f_i \tag{6.5}$$

$$\kappa'_i = \kappa_i \tag{6.6}$$

$$\boldsymbol{D}'_i = \gamma_i^{\frac{1}{\kappa_i}} \boldsymbol{D}_i + \frac{\kappa_i}{b}\left[\gamma_i^{\frac{1}{\kappa_i}} - 1\right]\boldsymbol{I}, \tag{6.7}$$

for all $\gamma_i > 0$, insofar as the expression of $\boldsymbol{D}'_i$ remains positive definite and as $\sum_{i=1}^{N} \gamma_i f_i = 1$. Estimating the parameters of a DIAMOND model from DWI acquired at a single b-value is therefore an ill-posed problem. Since $\boldsymbol{D}'_i$ is a linear combination of $\boldsymbol{D}_i$ and the identity $\boldsymbol{I}$, it has the same eigenvectors as $\boldsymbol{D}_i$. Only the estimation of the eigenvalues and the volumetric fractions is thus ill-posed. All three eigenvalues of $\boldsymbol{D}'_i$ are related to the eigenvalues of $\boldsymbol{D}_i$ by the same linear transformation. Therefore, all models of the set of equivalent models described by (6.5-6.7) are uniquely identified by the vector $(\lambda_1^{\min}, ..., \lambda_N^{\min})$ where $\lambda_i^{\min}$ is the lowest eigenvalue of the mean tensor $\boldsymbol{D}_i$. In terms of these eigenvalues (that we denote $\lambda_i \triangleq \lambda_i^{\min}$ for clarity), we have the following implicit equation for the manifold of equivalent DIAMOND models:

$$\begin{cases} \lambda_i = \gamma_i^{\frac{1}{\kappa_i}} \lambda_i^{\text{true}} + \frac{\kappa_i}{b}\left[\gamma_i^{\frac{1}{\kappa_i}} - 1\right] \\ \\ \sum_{i=1}^{N} \gamma_i f_i = 1. \end{cases} \tag{6.8}$$

As expected, for $\kappa_i \to \infty$, the implicit equations (6.8) are equivalent to the implicit equations (5.3) for multi-fascicle models, since

$$\lim_{\kappa\to\infty} \kappa\left(\gamma^{\frac{1}{\kappa}} - 1\right) = \log\gamma.$$

By eliminating the γ's between equations (6.8), we obtain the following compact expression for the manifold of equivalent DIAMOND models:

$$\sum_{i=1}^{N} f_i \left(\frac{\lambda_i + \frac{\kappa_i}{b}}{\lambda_i^{\text{true}} + \frac{\kappa_i}{b}}\right)^{\kappa_i} = 1. \tag{6.9}$$

As in the multi-fascicle case, identifying a DIAMOND model requires acquisitions at multiple b-values so as to have different manifolds intersecting at the true underlying model for which $\lambda_i = \lambda_i^{\text{true}}$, $\forall i$. The robustness and uniqueness of the intersection can be analyzed by exploring the geometry of the manifolds at different b-values.

The expression (6.9) has the form $F(\lambda_1, ..., \lambda_N) = 1$. A normal vector to the manifold of equivalent DIAMOND models is given by $\boldsymbol{\eta} = \boldsymbol{\nabla} F = \left(\frac{\partial F}{\partial \lambda_1}, \ldots, \frac{\partial F}{\partial \lambda_N}\right)$. Its k-th component computed at the point of the true underlying model ($\lambda_i = \lambda_i^{\text{true}}, \forall i$) is:

$$\eta_k \Big|_{\lambda_i = \lambda_i^{\text{true}}, \forall i} = \frac{\partial F}{\partial \lambda_k} \Big|_{\lambda_i = \lambda_i^{\text{true}}, \forall i} = \frac{f_k \kappa_k b}{b \lambda_k^{\text{true}} + \kappa_k}. \tag{6.10}$$

In contrast with multi-fascicle models, this normal vector depends on the b-value (unless $\kappa_k \to \infty$). Therefore, manifolds of equivalent DIAMOND models for different b-values do not, in general, intersect tangentially. The intersection between two different manifolds will be more robust to noise if the angle between the normal vectors is wider. This is akin to the intersection of two straight lines. Small changes in their slopes will displace the intersection point by a distance that depends on the difference between their slopes. If their slopes strongly differ, then changing them slightly will not strongly impact the location of the intersection (Fig. 6.5a) whereas if their slopes are close to each other, changing them slightly will strongly impact the location of the intersection (Fig. 6.5b).

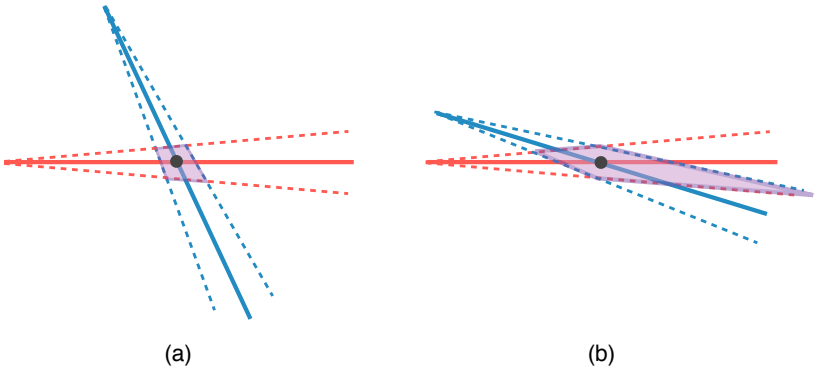


Figure 6.5: The difference between the slopes of manifolds impacts the robustness of their intersection. (a) If the slopes differ strongly, a slight change of the slopes (depicted are dashed lines corresponding to changes of $\pm 5°$) do not strongly impact the location of the intersection. The purple area shows all the possible locations of the intersection is slopes change by up to $5°$. (b) If slopes are close to each other, a slight difference may strongly impact the location of the intersection, as shown by the wider purple area.

Equation (6.10) enables to assess, through simulations, how the choice of b-values impact the angle between the manifolds of equivalent DIAMOND models.

As an illustration, we explored how the angle between the normal vectors to two manifolds evolves as a function of the second b-value when the first b-value is set to $b = 1000$ s/mm^2. We generated 1000 random DIAMOND models by uniformly drawing eigenvalues (λ_i) between 0 and 0.01mm^2/s, concentrations parameters (κ_i) between 1 and 10, volumetric fractions between 0 and 1 and tensor azimuth and zenith angles between 0° and 360°. The second b-value was set between 100 s/mm^{2}and 10,000 s/mm^{2}by steps of 50 s/mm^2. The procedure was repeated for a two-fascicle DIAMOND model and a three fascicle DIAMOND model.

Angular separations between the manifolds at the intersection corresponding to the true model are depicted in Fig. 6.6. This angle increases monotonically as the second b-value is set further apart from the initial one ($b = 1000$ s/mm^2). The rate of change of this angle is higher when b-values are lower than the first one, indicating that acquisitions at smaller b-values may play an important role in estimating the parameters of a DIAMOND model.

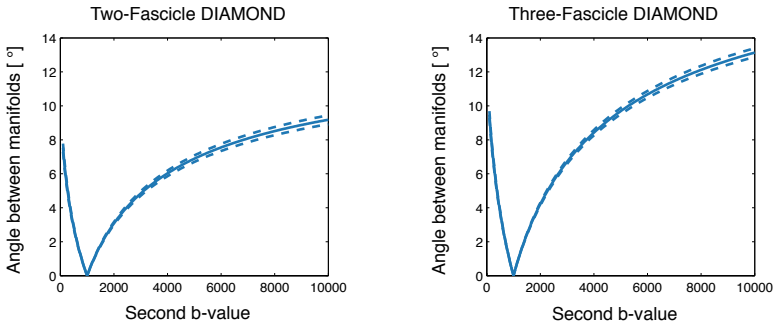


Figure 6.6: More robust intersections are obtained when the second b-value is further apart from the first one. The angle between manifolds, however, remains small. The curves show the relation between the second b-value (when the first one is 1000 s/mm^2) and the angle between the normal vector to the manifolds at the intersection point. Even for large ($b = 10,000$ s/mm^2) and small ($b = 100$ s/mm^2) b-values, the angular separation remains, on average, small ($< 14°$).

For multi-fascicle models, we saw in Chapter 5 that the tangency of manifolds at the intersection point and the positive definiteness of the difference between the Hessian matrices imply that the true model $\lambda_i = \lambda_i^{\text{true}}$ is locally the only intersection between two manifolds. For DIAMOND models, general claims about the dimension of the intersection between two manifolds are less

straightforward. This is because solutions to the equation (6.9) cannot be found other than numerically, to the best of our knowledge. In the next section, we will investigate the locus of intersection points between two manifolds as well as the geometry of manifolds around these points when considering a two-fascicle DIAMOND model. This investigation, as it turns out, will provide important insights into the estimation of DIAMOND models and the required number of b-values.

6.3.2 Investigation of the Two-Fascicle DIAMOND Model

Equation (6.9) for $N = 2$ can be rewritten explicitly as a function $\lambda_2(\lambda_1; b)$ relating the lowest eigenvalue of the first compartment (λ_1) to the lowest eigenvalue of the second compartment (λ_2). Relabeling $f_1 = f$ and $f_2 = 1 - f$, we have:

$$\lambda_2(\lambda_1; b) = \left(\lambda_2^{\text{true}} + \frac{\kappa_2}{b}\right)\left[1 - f\left(\frac{b\lambda_1 + \kappa_1}{b\lambda_1^{\text{true}} + \kappa_1}\right)^{\kappa_1}\right]^{1/\kappa_2}(1-f)^{-1/\kappa_2} - \frac{\kappa_2}{b}. \tag{6.11}$$

This expression holds because $f\left(\frac{b\lambda_1+\kappa_1}{b\lambda_1^{\text{true}}+\kappa_1}\right)^{\kappa_1} = f\gamma_1 < 1$ due to the condition on the volumetric fractions $f_i' = f_i\gamma_i$.

Equation (6.11) describes a curve (the manifold) in the plane (λ_1, λ_2) restricted to its upper-right quadrant. We can show that, for any b-value and any underlying model, $\lambda_2(\lambda_1; b)$ is a decreasing concave function of λ_1 for a given b (Appendix C). This provides an overall characterization of the manifold of equivalent two-fascicle DIAMOND models. Furthermore, when DWI at two distinct b-values are acquired, the set of equivalent two-fascicle DIAMOND models consist, at most, of two points that are the only intersections of the curves $\lambda_2(\lambda_1; b)$ and $\lambda_2(\lambda_1; b')$ for $b \neq b'$ (Proposition 8 in Appendix C).

These findings imply that DWI need to be acquired with at least three distinct non-zero b-values in order for the estimation of a DIAMOND model to be well posed. This is a necessary condition on the number of b-values to be used in the acquisition. For two-fascicle DIAMOND models, this is also an upper bound if and only if the second intersection of the two manifolds depends on their b-values (b and b'). Obtaining an analytic expression for the second intersection is challenging. In our simulations, however, the intersection of three manifolds was always a single point, suggesting that acquisitions at three b-values are theoretically necessary and sufficient to distinguish two-fascicle DIAMOND models.

Although distinct at all points other than the true underlying model, the manifolds for three different b-values may remain extremely close (Fig. 6.7). Albeit well-posed with more than two b-values, the problem of estimating the parameters of a DIAMOND model may thus remain ill-conditioned. The manifolds in Fig. 6.7 were generated with $\lambda_1^{\text{true}} = 0.77 \times 10^{-3}$ mm^2/s, $\lambda_2^{\text{true}} = 1.3 \times 10^{-3}$ mm^2/s, $\kappa_1 = 3.35$, $\kappa_2 = 2.33$, and $f = 0.58$. At the intersection between the manifolds corresponding to $b = 4500$ s/mm^2 and $b = 500$ s/mm^2 (left zoom-in on Fig. 6.7), the third manifold (with $b = 1500$ s/mm^2) is offset by only 0.5% on the λ_2-axis. Reliably estimating the parameters of such a DIAMOND model would thus require λ_2 to be estimated with an error lower than 0.5%. Failing to achieve this accuracy may cause the optimization to converge to the wrong intersection of the manifolds. In the case of the model of Fig. 6.7, choosing the wrong intersection would result in an uncertainty on λ_1 of 52% and an uncertainty on λ_2 of 44%. As for the uncertainty on the fraction f, it can be computed from equation (6.8).

$$f' = \gamma_1 f = \left(\frac{\lambda_1 + \frac{\kappa_1}{b}}{\lambda_1^{\text{true}} + \frac{\kappa_1}{b}} \right)^{\kappa_1} f,$$

which, for the model of Fig. 6.7, corresponds to an uncertainty of 29%.

The uncertainty on the model parameters depends on the b-values used for the acquisition, the true underlying DIAMOND model and the noise on the DWI. Assessing this uncertainty is critical to analyzing the resulting models and to compare its value in population studies. In the following section, we present a simple method to assess the parameter uncertainty based on the equation of the manifolds of equivalent models.

6.3.3 Manifold-Based Estimation of Parameter Uncertainty

The parameters of the models of the brain microstructure may ultimately be used as biomarkers of neurological disorders. Assessing the uncertainty of the parameter estimates is important to evaluate the reliability of inferences based upon these parameters. Non-parametric methods such as bootstrap methods are commonly used in this context. However, the knowledge of the geometry of the manifold of equivalent models provides complementary insights into the uncertainty of parameter estimates. For instance, this knowledge informs us that, at a single b-value, the uncertainty on the parameter estimate describes a manifold even with noise-free measurements. In this section, we develop a simple method to harness the knowledge of the geometry of the space of solutions

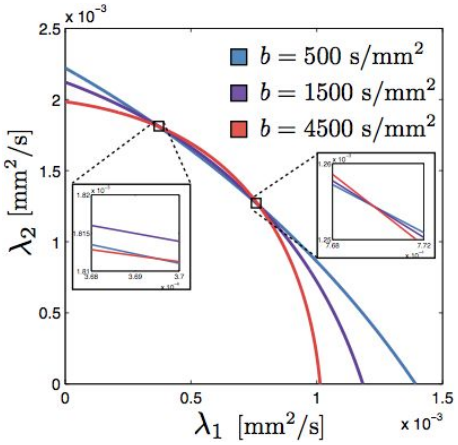


Figure 6.7: Manifolds of equivalent two-fascicle DIAMOND models. The manifolds are decreasing concave curves whose curvature increases (in absolute value) with the b-value b. Three such manifolds obtained with three different b-values intersect at a single point. However, the manifolds may remain close over a wider range of parameters. To single out the true underlying model (avoiding convergence to the second zoomed-in region), a high accuracy in the estimation is required. If this accuracy is not achieved, large uncertainty on the parameter remains, pertaining to the geometry of the manifolds.

in estimations of the model uncertainty. This method is complementary to non-parametric estimates and can be used alongside these methods to provide rich characterizations of parameter uncertainty.

The uncertainty on the model parameters can be fully described by the posterior distribution over the models $\mathcal{M}$, that is $p(\mathcal{M}|\boldsymbol{y})$ where $\boldsymbol{y}$ are the acquired DWI. Insofar as the DWI are independent given a model $\mathcal{M}$, we have:

$$p\Big(\mathcal{M}\Big|\boldsymbol{y}\Big) \propto p\Big(\mathcal{M}\Big) \prod_b p\Big(\boldsymbol{y}_b\Big|\mathcal{M}\Big), \tag{6.12}$$

where $p(\mathcal{M})$ is a prior over the model parameters and $\boldsymbol{y}_b$ is the subset of DWI measurements acquired at a b-value b. We know that $p(\boldsymbol{y}_b|\mathcal{M})$ is equal for all models $\mathcal{M}$ along the manifold of equivalent models at a b-value b. In the noise-free scenario, $p(\boldsymbol{y}_b|\mathcal{M})$ would therefore be uniform along the manifold. Under the influence of noise, the expression of $p(\boldsymbol{y}_b|\mathcal{M})$ depends on the noise

model. Under the Gaussian assumption, for instance, we have:

$$p(\boldsymbol{y}_b|\mathcal{M}) \propto \exp\left(-\frac{1}{2\sigma^2_{\text{noise}}}\sum_{k=1}^{K_b}\left(S(\mathcal{M},\boldsymbol{g}_k,b)-y_{b,k}\right)\right)^2,$$

where $S(\mathcal{M},\boldsymbol{g}_k,b)$ is the signal modeled by (6.3). If $\mathcal{M}'$ lies on the same manifold of equivalent models as $\mathcal{M}$ for a b-value b, we have $S(\mathcal{M}',\boldsymbol{g}_k,b) = S(\mathcal{M},\boldsymbol{g}_k,b)$. Therefore, the distribution $p(\boldsymbol{y}_b|\mathcal{M})$ has level sets that follow the manifold of equation (6.9). The distribution $p(\boldsymbol{y}_b|\mathcal{M})$ can thus be seen as a stack of manifolds each assigned with some likelihood value. This distribution therefore only needs to be estimated in a direction orthogonal to the manifold. An example of such "stack" is depicted in Fig. 6.8(a-b) for two different noise levels.

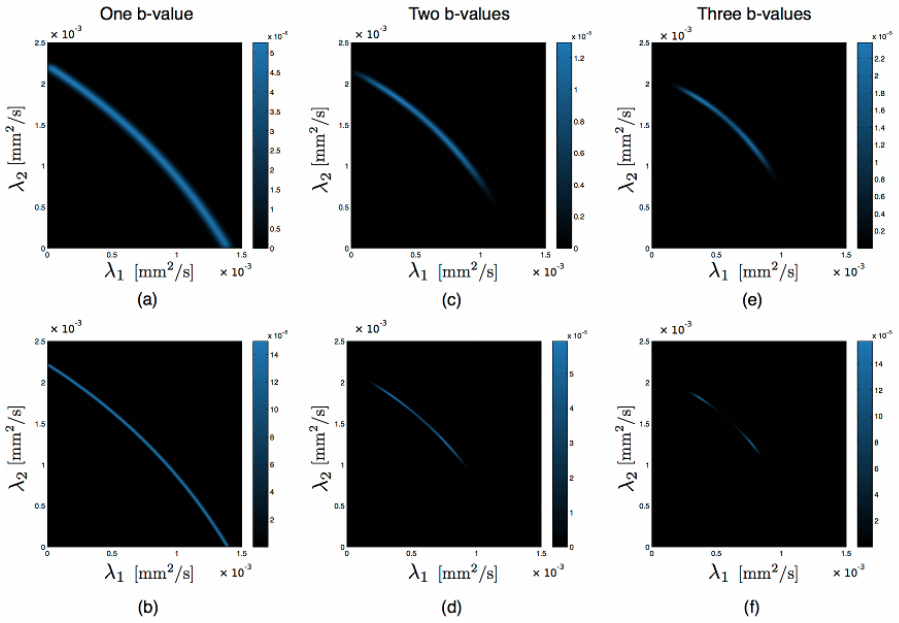


Figure 6.8: Example of posterior distributions over the model parameters when one, two or three b-values are used in the acquisition (a-b) At a single b-value, the posterior distribution has level sets that follow the same geometry as the manifolds, forming a stack of manifolds with different likelihood. (c-f) When multiple b-values are used in the acquisition, the distributions corresponding to each b-value are multiplied, decreasing the uncertainty on the parameter estimates. The resulting posterior distribution may be multimodal, as seen in (d) and (f).

To estimate the manifold-based uncertainty of the parameter estimates, one can proceed as follows:

1. Estimate a model $\mathcal{M}$ using all the acquired DWI $\boldsymbol{y}$ as in [98].
2. For each b-value b, estimate the likelihood $p(\boldsymbol{y}_b|\mathcal{M})$ (of DWI acquired with a b-value b) in the direction orthogonal to the manifold.
3. Compute the likelihood $p(\boldsymbol{y}_b|\mathcal{M})$ in all directions by propagating the values obtained in the previous step along the corresponding manifolds.
4. Multiply the likelihoods obtained in Step 3 for all b-values as in (6.12).
5. Multiply the result by the prior over the models $p(\mathcal{M})$.

Figure 6.8(c-f) depicts the result of this process for a two-fascicle DIAMOND model using DWI at two (c-d) and three (e-f) b-values for two different noise levels (higher in the top row and lower in the bottom row). Interestingly, the geometry of the manifolds may lead to a multimodal posterior distribution (especially visible in Fig. 6.8(f)). This multimodality implies that iterative optimization algorithms (such as the BOBYQA algorithm [86] used to estimate the DIAMOND parameters [98]) may converge to the wrong mode of the posterior distribution. To remove or weaken the confounding modes and single out the true underlying model, an informative prior could be used. In this context, a population-informed prior (as used in Chapter 5) would play an important role even if the estimation problem is ill-conditioned though not ill-posed. The data used to build this population-informed prior would need to be of high quality in order to dismiss the confounding modes of the posterior.

If the optimization algorithm is always initialized close to a single mode of the posterior distribution, it may converge to that mode even with different bootstrap replicates of the DWI data. Therefore, assessing the parameter uncertainty from bootstrap methods only may ignore the presence of other modes in the posterior distribution and mistakingly lead to the conclusion that uncertainty on the DIAMOND parameters is small. By contrast, the geometry of the manifold of equivalent models informs us of the presence of other modes if any. Importantly, this information remains available even if the optimization converges to the wrong mode, thereby warning that another model possibly explains the data.

The estimation of the parameter uncertainty can be performed for each voxel individually, once all DIAMOND models have been estimated. There are voxels

for which the manifolds at two different b-values only have one intersection, leading to monomodal posterior distributions and reduced parameter uncertainty. The uncertainty as computed with the proposed methods only impacts the three eigenvalues and the volumetric fraction of each compartment of the DIAMOND model. Because the orientation and concentration parameters κ_i are not affected by the ill-posedness problem described in (6.8), these parameters have a zero manifold-based uncertainty. Combining manifold-based uncertainty estimation with other uncertainty estimations such as the bootstrap is therefore important to assess the uncertainty in all the parameters.

6.4 Summary and Discussion

The multi-fascicle model used in the previous chapters assumes that the diffusion signals arising from each fascicle and from the extra-axonal space are well represented by a multi-variate Gaussian function. This assumption ignores the potential heterogeneity in each large scale microstructural environment. This heterogeneity may be due to meso-scale effects (such as differences in axonal diameter) and micro-scale effects (such as the distance from water molecules to the cell membrane). The DIAMOND model has been introduced to account for the different sources of heterogeneity. This model compares favorably to the multi-fascicle model in terms of generalization error estimated from 395 DWI.

Similarly to multi-fascicle models, there is an infinite number of DIAMOND models that all produce the same diffusion signal at a single b-value. These models describe a manifold of indistinguishable models. However, unlike the multi-fascicle models, the manifolds at different b-values do not intersect tangentially and may have several intersections. The estimation of the DIAMOND parameters thus remains ill-posed at two b-values. When more than two b-values are acquired, the estimation remains ill-conditioned and the posterior distribution over the model parameters can be multimodal. This multimodality implies that the optimization algorithm may systematically converge to the wrong model if it is initialized away from the mode corresponding to the true underlying model. The uncertainty on the model parameters may therefore be high even when DWI are acquired at multiple b-values.

The parameters of the multi-fascicle model used in the previous chapters may have a lower uncertainty than the parameters of the DIAMOND model. However, the lower uncertainty comes at the expense of an infinite error made on the concentration parameter κ_i. Future investigations should indicate whether

the concentration parameters play a key role in the definition of biomarkers and, if they do not, whether the other parameters can be reliably estimated when setting $\kappa_i \to \infty$.

One way to disentangle the different modes of the posterior distribution of DIAMOND models would be to include an informative prior over the model parameters. This informative prior could be based on biological assumptions (such as minimum and maximum values on the diffusivities) or based on an external population of subjects scanned with multiple b-values and with a high signal-to-noise ratio (by repeating and averaging the acquisitions of DWI, for instance).

The DIAMOND model holds promise to provide better characterization of the brain microstructure. Its phenomenological account of the heterogeneity in microstructural environments may prove useful in the definition of early biomarkers of neurological conditions such as autism spectrum disorders. Our capacity to assess and characterize the uncertainty in estimates of the DIAMOND parameters will be critical to draw reliable inference of abnormalities in this context.

Conclusions

A framework to study the brain microstructure

The brain function is supported by a complex organization of cellular components called the brain microstructure. Alterations of the microstructure is thought to play an important role in the pathogenesis of various neurological disorders. The brain microstructure can be analyzed non-invasively by microstructure imaging that relies on multi-fascicle models estimated from diffusion-weighted images. In this thesis, we proposed that microstructure alterations can be characterized by conducting population studies from clinically available diffusion-weighted images. Population studies based on multi-fascicle models required new capabilities to fulfill unmet needs.

In Chapter 3, we proposed a mathematical framework that enables the spatial alignment of multi-fascicle models and the construction of an atlas of the brain microstructure. These capabilities boil down to the definition of a method to compute weighted combinations of multi-fascicle models and a similarity metric to detect common multi-tensor landmarks. We proposed a scheme based on the simplification of Gaussian mixtures to efficiently compute weighted combinations of multi-fascicle models. As for the similarity metric, we proposed a generalized correlation coefficient that is invariant under linear transformations of diffusivities, thereby enabling the alignment of pathological microstructure images to an atlas based on healthy subjects.

Chapter 3 also proposed a system to perform statistical analysis of features of the brain microstructure. This system is composed of Fascicle-Based Spatial Statistics (FBSS) and Isotropic Diffusion Analysis (IDA). FBSS compares white matter properties on a per-fascicle basis. This property is fundamental to conducting population studies since altered fascicles may cross unaltered fascicles in a single voxel. Together, FBSS and IDA enable rich characterization of microstructural alterations reflected in multi-fascicle models.

In Chapter 4, we proposed a rationale to select between different models of the brain microstructure and to identify the appropriate granularity of the model for a given dataset. This selection is based on the minimization of the generalization error estimated by the 632 bootstrap. In contrast with the F-test commonly used in this context, minimization of the generalization error is shown

to be more robust to image pre-processing and less dependent on the choice of a specific threshold.

Most available clinical dataset have been acquired with a single b-value because of time constraints when imaging patients. In Chapter 5, we proposed a method to estimate multi-fascicle models from these datasets. This problem is known to be ill-posed. To regularize it, we introduced a population-informed prior in the maximum *a posteriori* estimation of the model parameters. This prior encodes the prior knowledge about the brain microstructure that can be learnt from other subjects, scanned in a research setting with multiple b-values. We showed that inference of group differences from the resulting multi-fascicle models are remarkably close to those obtained when images at multiple b-values are available.

Altogether, the model selection of Chapter 4, the estimation of Chapter 5, the registration and statistical analysis of Chapter 3 provide a comprehensive framework to conduct population studies of the brain microstructure from clinically available imaging datasets.

Future Directions

In future work, we will apply the developed framework to the widely available clinical datasets. These studies have the potential to answer pressing questions about normal development as well as advance research of biomarkers of autism spectrum disorders. If ASD is, as suggested by biopsies and post-mortem studies, related to an autoimmune response associated with an increased concentration of microglial cells and the consequential secretion of cytokines leading to cytotoxic response, population studies of the brain microstructure shall reveal systematic and robust differences in corresponding features of the diffusion model. If an autoimmune response is not part of the pathogenesis of (all) autism spectrum disorders, population studies of the brain microstructure shall invalidate this hypothesis and may point to another specific alteration of the brain microstructure.

These studies may require the adaptation of the multi-fascicle models to embrace more complex features of the brain microstructure, such as the heterogeneity of microstructural environments. Another recently active field of research is the characterization of the gray matter microstructure that has long been ignored due to the absence of a principal fascicle direction along which water molecules diffuse. Novel diffusion models with increased complexity may

better characterize the cortical microstructure and reveal patterns of alterations that point to certain diseases. In these cases, the methods developed in this thesis to estimate, register and analyze microstructure images will have to be adapted to the new selected model (may it be DIAMOND or a biophysical model of the diffusion signal).

Current imaging techniques may not be able to reliably estimate the parameters of these models of increased complexity from clinical data. However, fast parallel imaging and increasing resolution of MRI scanner will probably accelerate the development of such complex models and bring them to the realm of clinical practice. Furthermore, the brain microstructure can be imaged with modalities other than diffusion-weighted magnetic resonance imaging. PET imaging, for instance, enables to map the activity of microglial cells, which may then be used to detect neuroinflammatory response. Multimodal microstructure imaging will therefore likely play an ever important role in the identification of pathological pathways of neurological disorders.

Developments of novel methods to harness the multimodal models of brain microstructure in population studies will be required. These developments will aim at answering the very questions asked in this thesis: what acquisitions are required to estimate the selected models, how can we register and analyze them, and what insight in the microstructure do they provide? In this context, we believe that the developments of this thesis will serve as a fertile ground and provide important guidelines for future innovations.

Appendix A

Sufficient Conditions on the Generalized Scalar Mapping

In this appendix, we demonstrate how properties (3.14-3.17) are satisfied if conditions (3.18-3.21) are satisfied. We first recall these properties and conditions. The desired properties on ρ are:

$$\begin{aligned} \rho(aR+bT,S) &= \rho(R,S) && \text{(A.1)}\\ \rho(R,S) &= \rho(S,R) && \text{(A.2)}\\ \rho(R,R) &= 1 && \text{(A.3)}\\ |\rho(R,S)| &\leq 1. && \text{(A.4)} \end{aligned}$$

and the conditions on m that we proposed to be sufficient for those properties to hold are:

$$\begin{aligned} m(R+bT,T) &= m(R,T)+b\,m(T,T) && \text{(A.5)}\\ m(aR,S) &= a\,m(R,S) && \text{(A.6)}\\ m(R,S) &= m(S,R) && \text{(A.7)}\\ |m(R,S)| &\leq n_m(R)n_m(S). && \text{(A.8)} \end{aligned}$$

These properties hold for any generalized correlation ρ of the form:

$$\rho(R,S) = m\left(\frac{R-m(R,T)T}{n_m(R-m(R,T)T)}, \frac{S-m(S,T)T}{n_m(S-m(S,T)T)}\right). \qquad \text{(A.9)}$$

Proposition 1. *Property* (A.1) *is satisfied if conditions* (A.5)*,* (A.6) *and* (A.7) *are satisfied.*

Proof We have

$$\rho(aR+bT,S) = m\left(\frac{aR+bT-m(aR+bT,T)T}{n_m(aR+bT-m(aR+bT,T)T)}, \frac{S-m(S,T)T}{n_m(S-m(S,T)T)}\right)$$

From (A.5) and (A.6), and since $n_m(T) = 1$, we have

$$m(aR + bT, T) = am(R, T) + bm(T, T) = am(R, T) + b,$$

and therefore,

$$\rho(aR + bT, S) = m\left(\frac{aR - am(R,T)T}{n_m(aR - am(R,T)T)}, \frac{S - m(S,T)T}{n_m(S - m(S,T)T)}\right).$$

From (A.6) and (A.7), we have

$$n_m(aR)=\sqrt{m(aR, aR)}=\sqrt{am(R, aR)}=\sqrt{am(aR, R)}=\sqrt{a^2 m(R, R)}=an_m(R).$$

Therefore,

$$\begin{aligned}\rho(aR + bT, S) &= m\left(\frac{a[R - m(R,T)T]}{an_m(R - m(R,T)T)}, \frac{S - m(S,T)T}{n_m(S - m(S,T)T)}\right)\\ &= m\left(\frac{R - m(R,T)T}{n_m(R - m(R,T)T)}, \frac{S - m(S,T)T}{n_m(S - m(S,T)T)}\right)\\ &= \rho(R, S)\end{aligned}$$

■

Proposition 2. *Property* (A.2) *is satisfied if condition* (A.7) *is satisfied.*

Proof This follows directly from the definition (A.9) of $\rho(R, S)$ as a scalar mapping m between two normalized multi-fascicle models. ■

Proposition 3. *Property* (A.3) *is satisfied if conditions* (A.6) *and* (A.7) *are satisfied.*

Proof From (A.6) and (A.7), we have $n_m(aR) = an_m(aR)$ as described in the proof of Proposition 1. Therefore,

$$\begin{aligned}\rho(R, R) &= n_m^2\left(\frac{R - m(R,T)T}{n_m(R - m(R,T)T)}\right)\\ &= \frac{1}{n_m^2(R - m(R,T)T)} n_m^2(R - m(R,T)T)\\ &= 1\end{aligned}$$

■

Proposition 4. *Property* (A.4) *is satisfied if conditions* (A.6), (A.7) *and* (A.8) *are satisfied.*

Proof From (A.6) and (A.7), we have

$$m(aR, bS) = a\, m(R, bS) = a\, m(bS, R) = ab\, m(R, S)$$

Therefore,

$$\begin{aligned}
& |\rho(R,S)| \leq 1 \\
& \left| m\Big(\frac{R - m(R,T)T}{n_m(R - m(R,T)T)}, \frac{S - m(S,T)T}{n_m(S - m(S,T)T)} \Big) \right| \leq 1 \\
\Leftrightarrow \quad & \frac{\Big| m\Big(R - m(R,T)T, S - m(S,T)T\Big) \Big|}{n_m\Big(R - m(R,T)T\Big) n_m\Big(S - m(S,T)T\Big)} \leq 1 \\
\Leftrightarrow \quad & \Big| m\Big(R - m(R,T)T, S - m(S,T)T\Big) \Big| \leq n_m\Big(R - m(R,T)T\Big) n_m\Big(S - m(S,T)T\Big)
\end{aligned}$$

The latter inequality is a particular case of Property (A.8). ■

Appendix B

Details on the Studied Population

The population studies presented in Chapter 3 and 5 involve healthy controls, subjects with TSC and subjects with both TSC and autism. In each case, the groups were age and gender-matched. The age range in both cases was 0.5 - 25 years.

The controls were either recruited specifically as healthy controls or were patients seen at the Boston Children's Hospital who received a clinical MRI for a reason other than TSC or developmental disability. A pediatric neuroradiologist reviewed each MRI; all controls had normal MRI results and normal neurologic examination results. Controls did not undergo neuropsychological evaluation as part of these studies. Recruitment of subjects and data acquisition were conducted using a protocol approved by the institutional review board of Boston Children's Hospital. All patients fulfilled the clinical criteria for definite TSC, as defined by the Tuberous Sclerosis Consensus Conference [93]. All patients with TSC were neurologically examined, and clinical data were obtained during office visits and from review of medical records. ASD diagnoses were based on clinical assessment by a board-certified pediatric neurologist, using the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision [15].

Appendix C

Manifold of Equivalent Two-Fascicle DIAMOND Models

Equation (6.11) describes the manifold of indistinguishable DIAMOND models for a given b-value b:

$$\lambda_2(\lambda_1;b) = \left(\lambda_2^{\text{true}} + \frac{\kappa_2}{b}\right)\left[1 - f\left(\frac{b\lambda_1+\kappa_1}{b\lambda_1^{\text{true}}+\kappa_1}\right)^{\kappa_1}\right]^{1/\kappa_2}(1-f)^{-1/\kappa_2} - \frac{\kappa_2}{b}.$$

In this appendix, we will demonstrate that this curve is a decreasing concave function of λ_1 and $\lambda_2(\lambda_1;b)$ and $\lambda_2(\lambda_1;b')$ have at most two intersections that need to be considered for any $b \neq b'$. We start with two propositions related to the first and second derivative of $\lambda_2(\lambda_1;b)$.

Proposition 5. *$\lambda_2(\lambda_1;b)$ is a decreasing function of λ_1 for a given b.*

Proof If we let

$$u(\lambda_1;b) \hat{=} \frac{b\lambda_1+\kappa_1}{b\lambda_1^{\text{true}}+\kappa_1},$$

then, we have the following expression for the partial derivative of λ_2:

$$\frac{\partial\lambda_2}{\partial\lambda_1}(\lambda_1;b) = -\frac{\kappa_1 f b}{\kappa_2\left(b\lambda_1^{\text{true}}+\kappa_1\right)}(1-f)^{-1/\kappa_2}\left(\lambda_2^{\text{true}}+\frac{\kappa_2}{b}\right)\left(1-fu^{\kappa_1}\right)^{\frac{1}{\kappa_2}-1}u^{\kappa_1-1}.$$

We know that $\kappa_1 > 1$ and $\kappa_2 > 1$ by definition of the matrix-variate Gamma distribution, that $0 < f < 1$ because f is a volumetric fraction, and that $fu^{\kappa_1} = \gamma_1 f_1 < 1$. The partial derivative of λ_2 is therefore negative and λ_2 is a decreasing function of λ_1. ■

Proposition 6. *$\lambda_2(\lambda_1;b)$ is a concave function of λ_1 for a given b.*

Proof The second derivative of λ_2 with respect to λ_1 is:

$$\frac{\partial^2\lambda_2}{\partial\lambda_1^2}(\lambda_1;b) = -K\left(1-fu^{\kappa_1}\right)^{\frac{1}{\kappa_2}-2}u^{\kappa_1-2}\left[\left(1-\frac{\kappa_1}{\kappa_2}\right)fu^{\kappa_1}+\kappa_1-1\right], \quad \text{(C.1)}$$

with

$$K \hat{=} \frac{\kappa_1 f b}{\kappa_2 \left(b\lambda_1^{\text{true}} + \kappa_1\right)^2} \left(\lambda_2^{\text{true}} + \frac{\kappa_2}{b}\right) (1-f)^{-1/\kappa_2}$$

Because each of its factors is positive, K is positive. We know that $fu^{\kappa_1} < 1$ and therefore $(1 - fu^{\kappa_1}) > 0$. Furthermore,

$$v(\lambda_1; b) \hat{=} \left[\left(1 - \frac{\kappa_1}{\kappa_2}\right) fu^{\kappa_1} + \kappa_1 - 1\right] > 0$$

because it is a linear function of $fu^{\kappa_1} \in]0, 1[$ and:

$$\begin{aligned} v\Big|_{fu^{\kappa_1}=0} &= \kappa_1 - 1 > 0 \\ v\Big|_{fu^{\kappa_1}=1} &= \kappa_1 - \frac{\kappa_1}{\kappa_2} > 0. \end{aligned}$$

Therefore, the derivative (C.1) is negative and $\lambda_2(\lambda_1; b)$ is concave. ■

Now, two distinct decreasing concave functions may have an infinite number of intersections. To assess the number of intersections between two curves $\lambda_2(\lambda_1; b)$ and $\lambda_2(\lambda_1; b')$ that need to be considered as potential solutions of the DIAMOND estimation problem, we need to investigate the evolution of the second derivative (C.1) with respect to b. This is the gist of the following proposition.

Proposition 7. *For $\lambda_1 > \lambda_1^{true}$, $\frac{\partial^2 \lambda_2}{\partial \lambda_1^2}(\lambda_1; b)$ is a decreasing function of b.*

Proof The partial derivative of $\frac{\partial^2 \lambda_2}{\partial \lambda_1^2}(\lambda_1; b)$ with respect to b is:

$$\begin{aligned} \frac{\partial}{\partial b}\frac{\partial^2 \lambda_2}{\partial \lambda_1^2}(\lambda_1; b) = w(\lambda_1; b) \Bigg[& (\kappa_2 + b\lambda_2^{\text{true}})\kappa_1 + b\lambda_2^{\text{true}}(\kappa_1 + b\lambda_1^{\text{true}}) \\ & + \kappa_1 b f v^{-1} \left(\frac{\kappa_1}{\kappa_2} - 1\right) (\kappa_2 + b\lambda_2^{\text{true}}) u^{\kappa_1 - 1} \frac{\kappa_1(\lambda_1 - \lambda_1^{\text{true}})}{\kappa_1 + b\lambda_1^{\text{true}}} \\ & - \kappa_1 b f (\kappa_2 + b\lambda_2^{\text{true}}) u^{\kappa_1 - 1} \left(\frac{1}{\kappa_2} - 2\right) (1 - fu^{\kappa_1})^{-1} \frac{\kappa_1(\lambda_1 - \lambda_1^{\text{true}})}{\kappa_1 + b\lambda_1^{\text{true}}} \\ & + b(\kappa_2 + b\lambda_2^{\text{true}}) u^{-1} (\kappa_1 - 2) \frac{\kappa_1(\lambda_1 - \lambda_1^{\text{true}})}{\kappa_1 + b\lambda_1^{\text{true}}} \Bigg], \end{aligned} \tag{C.2}$$

where $w(\lambda_1; b)$ is a negative function defined by:

$$w(\lambda_1; b) \hat{=} -\frac{\kappa_1 f (1 - fu^{\kappa_1})^{\frac{1}{\kappa_2} - 2} u^{\kappa_1 - 2} v}{(1-f)^{1/\kappa_2} (\kappa_1 + b\lambda_1^{\text{true}})^2 \kappa_2}.$$

The last three terms of (C.2) can be grouped leading to:

$$\frac{\partial}{\partial b}\frac{\partial^2\lambda_2}{\partial\lambda_1^2}(\lambda_1;b) = w(\lambda_1;b)\Bigg[(\kappa_2+b\lambda_2^{\text{true}})\kappa_1+b\lambda_2^{\text{true}}(\kappa_1+b\lambda_1^{\text{true}})$$
$$+\,b(\kappa_2+b\lambda_2^{\text{true}})\frac{\kappa_1(\lambda_1-\lambda_1^{\text{true}})}{\kappa_1+b\lambda_1^{\text{true}}}u^{-1}v^{-1}\left(1-fu^{\kappa_1}\right)^{-1}$$
$$\times\left\{\kappa_1 f\left(\frac{\kappa_1}{\kappa_2}-1\right)u^{\kappa_1}\left(1-fu^{\kappa_1}\right)-\kappa_1 fu^{\kappa_1}\left(\frac{1}{\kappa_2}-2\right)v\right.$$
$$\left.+\,(\kappa_1-2)v+(\kappa_1-2)\left(1-fu^{\kappa_1}\right)\right\}\Bigg]. \tag{C.3}$$
$$\tag{C.4}$$

We can show that the expression inside the curly brackets is equal to:

$$\left\{(1-fu^{\kappa_1})(\kappa_1-1)\kappa_1+2(1-fu^{\kappa_1})(\kappa_1-1)v+\kappa_1 f\left(2-\frac{1}{\kappa_2}\right)u^{\kappa_1}v\right\}$$

and is therefore positive since $u>0$, $v>0$, $(1-fu^{\kappa_1})>0$, $\kappa_1>1$ and $\kappa_2>1$. If $\lambda_1>\lambda_1^{\text{true}}$, then all the terms in (C.3) are positive. Since all these positive terms are multiplied by the negative $w(\lambda_1;b)$, we have:

$$\frac{\partial}{\partial b}\frac{\partial^2\lambda_2}{\partial\lambda_1^2}(\lambda_1;b)<0,\quad\forall\,\lambda_1>\lambda_1^{\text{true}}.$$

and therefore $\frac{\partial^2\lambda_2}{\partial\lambda_1^2}$ is a decreasing function of b for $\lambda_1>\lambda_1^{\text{true}}$. ■

Proposition 7 implies that the curvature of the curves $\lambda_2(\lambda_1;b)$ in the plane (λ_1,λ_2) are decreasing when b increases for all $\lambda_1>\lambda_1^{\text{true}}$. Since the curvatures are negative (Proposition 6), their absolute value increases with the b-value. Proposition 7 leads to the following important results.

Proposition 8. *When DWI at two distinct b-values are acquired, the set of equivalent two-fascicle DIAMOND models consist, at most, of two points.*

Proof Let $\lambda_2(\lambda_1;b)$ and $\lambda_2(\lambda_1;b')$ be the two manifolds of equivalent DIAMOND models obtained by acquiring data at two distinct b-values b,b' (we let $b>b'$). These two curves intersect when their difference ($\Delta(\lambda_1)\hat{=}\lambda_2(\lambda_1;b')-\lambda_2(\lambda_1;b)$) equals zero. By construction, $\Delta(\lambda_1^{\text{true}})=0$ because both manifolds

include the true underlying model. For $\lambda_1 > \lambda_1^{\text{true}}$, $\Delta(\lambda_1)$ is a convex function of λ_1 since:

$$\frac{\partial^2 \Delta(\lambda_1)}{\partial \lambda_1^2} = \frac{\partial^2 \lambda_2(\lambda_1; b')}{\partial \lambda_1^2} - \frac{\partial^2 \lambda_2(\lambda_1; b)}{\partial \lambda_1^2} > 0, \quad \forall \lambda_1 > \lambda_1^{\text{true}}, \tag{C.5}$$

due to Proposition 7. Therefore, $\Delta(\lambda_1)$ has at most one other zero for $\lambda_1 > \lambda_1^{\text{true}}$. Now, let us assume that $\Delta(\lambda_1)$ has additional zeros for $\lambda_1 < \lambda_1^{\text{true}}$. If it has a total of two distinct zeros, then the proposition is proven. If it has more zeros, let us label them $\rho_1 < \rho_2 < ... < \rho_m$ with $m \geq 3$. Any ρ_k for $k < m - 1$ cannot be the true underlying model, because if would violate (C.5). Therefore, the true underlying model may only be $(\rho_{m-1}, \lambda_2(\rho_{m-1}; b))$ or $(\rho_m, \lambda_2(\rho_m; b))$.

Now, let some DIAMOND model be characterized by $(\rho_1, \lambda_2(\rho_1; b)$. Because it is on the same manifold as $(\lambda_1^{\text{true}}, \lambda_2^{\text{true}})$, it generates the same diffusion signal as any of the model on that manifold. Therefore, the manifold of DIAMOND models equivalent to $(\rho_1, \lambda_2(\rho_1; b))$ is the same as the manifold of models equivalent to $(\lambda_1^{\text{true}}, \lambda_2^{\text{true}})$ at a b-value b. Since $(\rho_1, \lambda_2(\rho_1; b)) = (\rho_1, \lambda_2(\rho_1; b'))$, the same is true for a b-value b'. From Proposition 7 with $\lambda_1^{\text{true}'} = \rho_1$, there is therefore a maximum of one intersection with $\lambda_1 > \rho_1$. Since there is no intersection with $\lambda_1 < \rho_1$ by definition of ρ_1, the two manifolds described by $\lambda_2(\lambda_1; b)$ and $\lambda_2(\lambda_1; b')$ have at most two intersections, which proves the proposition. ■

Two manifolds spanned from the same true model (with $\lambda_1^{\text{true}} = 0.77 \times 10^{-3}$ mm^2/s and $\lambda_2^{\text{true}} = 1.3 \times 10^{-3}$ mm^2/s) with two intersections are depicted in Fig. C.1. They illustrate all the propositions demonstrated in this appendix.

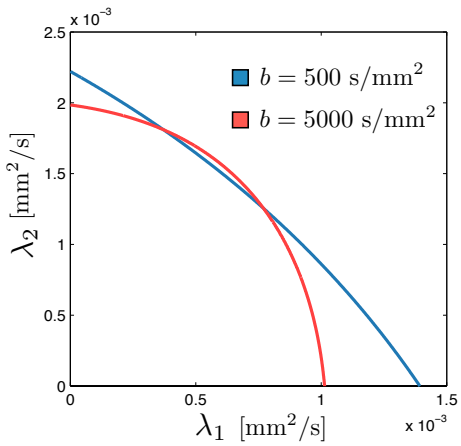


Figure C.1: Manifolds of equivalent two-fascicle DIAMOND models. The manifolds are decreasing concave curves whose curvature increases (in absolute value) with the b-value b. The manifolds obtained with two different b-values intersect at two points, at most.

Related Publications

Journal Papers

1. Taquet, Maxime, Benoit Scherrer, Nicolas Boumal, Benoit Macq, Simon K. Warfield. "Estimation of a Multi-Fascicle Model from Diffusion MRI at a Single B-Value with a Population-Informed Prior.", *under review* (2013)
2. Taquet, Maxime, Benoit Scherrer, Olivier Commowick, Jurriaan M. Peters, Mustafa Sahin, Benoit Macq and Simon K. Warfield. "A Mathematical Framework for the Registration and Analysis of Multi-Fascicle Models for Population Studies of the Brain Microstructure." IEEE Transactions on Medical Imaging, in press, (2013)
3. Peters, Jurriaan M.*, Maxime Taquet*, Anna K. Prohl, Benoit Scherrer, Agnies M. van Eeghen, Sanjay P. Prabhu, Mustafa Sahin, and Simon K. Warfield. "Diffusion tensor imaging and related techniques in tuberous sclerosis complex: review and future directions." Future Neurology 8, no. 5 (2013): 583-597.
4. Taquet, Maxime, Jordi Quoidbach, Yves-Alexandre de Montjoye, and Martin Desseilles. "Mapping Collective Emotions to Make Sense of Collective Behavior." Behavioral and Brain Sciences, in press (2013)
5. Taquet, Maxime, and Jurriaan M. Peters. "The brain functional network of children with autism: redundancy and disconnection." Medecine sciences: M/S 29, no. 6-7 (2013): 567-569.
6. Peters, Jurriaan M.*, Maxime Taquet*, Clemente Vega, Shafali S. Jeste, Ivan Sanchez Fernandez, Jacqueline Tan, Charles A. Nelson, Mustafa Sahin, and Simon K. Warfield. "Brain functional networks in syndromic and non-syndromic autism: a graph theoretical study of EEG connectivity." BMC medicine 11, no. 1 (2013): 54.

Full-Length Peer-Reviewed Conference Papers

1. Taquet, Maxime, Benoit Scherrer, Nicolas Boumal, Benoit Macq, and Simon K. Warfield. "Estimation of a multi-fascicle model from single b-value data with a population-informed prior." In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013, pp. 695-702. Springer Berlin Heidelberg, 2013.
2. Scherrer, Benoit, Armin Schwartzman, Maxime Taquet, Sanjay P. Prabhu, Mustafa Sahin, Alireza Akhondi-Asl, and Simon K. Warfield. "Characterizing the DIstribution of Anisotropic MicrO-structural eNvironments with Diffusion-Weighted Imaging (DIAMOND)." In Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013, pp. 518-526. Springer Berlin Heidelberg, 2013.

* Equal contributions

3. Scherrer, Benoit*, Maxime Taquet*, and Simon K. Warfield. "Reliable selection of the number of fascicles in diffusion images by estimation of the generalization error." In Information Processing in Medical Imaging, pp. 742-753. Springer Berlin Heidelberg, 2013.

4. Taquet, Maxime, Benoit Scherrer, Olivier Commowick, Jurriaan Peters, Mustafa Sahin, Benoit Macq, and Simon K. Warfield. "Registration and Analysis of White Matter Group Differences with a Multi-Fiber Model." In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012, pp. 313-320. Springer Berlin Heidelberg, 2012.

5. Taquet, Maxime, Benoit Scherrer, Christopher Benjamin, Sanjay Prabhu, Benoit Macq, and Simon K. Warfield. "Interpolating multi-fiber models by gaussian mixture simplification." In Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, pp. 928-931. IEEE, 2012.

6. Taquet, Maxime, Benoit Macq, and Simon K. Warfield. "A generalized correlation coefficient: Application to dti and multi-fiber dti." In Mathematical Methods in Biomedical Image Analysis (MMBIA), 2012 IEEE Workshop on, pp. 9-14. IEEE, 2012.

7. Taquet, Maxime, Benoit Macq, and Simon K. Warfield. "Spatially adaptive log-euclidean polyaffine registration based on sparse matches." In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2011, pp. 590-597. Springer Berlin Heidelberg, 2011.

8. Taquet, Maxime, Laurent Jacques, Benoit Macq, and Sylvain Jaume. "Compact rotation invariant image descriptors by spectral trimming." In Image Processing (ICIP), 2011 18th IEEE International Conference on, pp. 2033-2036. IEEE, 2011.

9. Taquet, Maxime. "Feature-based error processing for robust surface registration in computer assisted orthopedic surgery." In MELECON 2010-2010 15th IEEE Mediterranean Electrotechnical Conference, pp. 492-497. IEEE, 2010.

Peer-Reviewed Conference Abstracts

1. Taquet, Maxime, Robin Gong, and Simon K. Warfield. "Contradictory Conclusions in Analysis of Brain Functional Networks: the Role of Image Registration." Proc. Organization for Human Brain Mapping (OHBM), Seattle, USA, 2013

2. Taquet, Maxime, Benoit Scherrer, Benoit Macq, and Simon K. Warfield. "Multi-Fascicle Model Reconstruction from Acquisitions at a Single b-value with a Population-Informed Prior." Proc. the 21th International Symposium on Magnetic Resonance in Medicine (ISMRM), Salt Lake City, USA, 2013

3. Scherrer, Benoit, Maxime Taquet, Onur Afacan, Simon K. Warfield. "Does the signal arising from a single fascicle significantly deviates from a monoexponential decay with a clinical scanner?." Proc. the 21th International Symposium on Magnetic Resonance in Medicine (ISMRM), Salt Lake City, USA, 2013

4. Peters, Jurriaan M., Maxime Taquet, Shafali S. Jeste, Ivan Sanchez-Fernandez, Jacqueline Tan, Charles A. Nelson, III, Mustafa Sahin, and Simon K. Warfield. "Identification of Distinct Network Topology and Resilience Features in Tuberous Sclerosis and

Autism.", the American Academy of Neurology 65th Annual Meeting, San Diego, CA, 2013

5. Peters, Jurriaan M., Maxime Taquet, Ivan Sanchez-Fernandez, Jacqueline Tan, Mustafa Sahin, and Simon K Warfield. "Resting State Cortical Connectivity in Tuberous Sclerosis Complex." In proceedings of 41st Annual Meeting of the Child Neurology Society, 2012

6. Maxime Taquet, Laurent Paul, Bruno Dehez, and Benoit Macq. "Outliers detection and removal for accurate surface-based registration." In proceedings of the 10th International Conference on Computer Assisted Orthopedic Surgery (CAOS), 2010

Bibliography

[1] F. Aboitiz, A. B. Scheibel, R. S. Fisher, and E. Zaidel. Fiber composition of the human corpus callosum. *Brain research*, 598(1):143–153, 1992. 102

[2] D. C. Alexander. A general framework for experiment design in diffusion mri and its application in measuring direct tissue-microstructure features. *Magnetic Resonance in Medicine*, 60(2):439–448, 2008. 45, 46

[3] D. C. Alexander and G. J. Barker. Optimal imaging parameters for fiber-orientation estimation in diffusion mri. *Neuroimage*, 27(2):357–367, 2005. 101

[4] D. C. Alexander, G. J. Barker, and S. R. Arridge. Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data. *Magn Reson Med*, 48(2):331–340, 2002. 86, 113

[5] D. C. Alexander, P. L. Hubbard, M. G. Hall, E. A. Moore, M. Ptito, G. J. Parker, and T. B. Dyrby. Orientationally invariant indices of axon diameter and density from diffusion mri. *Neuroimage*, 52(4):1374–1389, 2010. 45, 46, 52

[6] D. C. Alexander, C. Pierpaoli, P. J. Basser, and J. C. Gee. Spatial transformations of diffusion tensor magnetic resonance images. *Medical Imaging, IEEE Transactions on*, 20(11):1131–1139, 2001. 66

[7] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the Log-Euclidean framework. *MICCAI 2005*, pages 115–122, 2005. 41

[8] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006. 41, 42, 58, 109

[9] J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, and K. Friston. Identifying global anatomical differences: deformation-based morphometry. *Human Brain Mapping*, 6(5-6):348–357, 1998. 78

[10] P. Ashwood, P. Krakowiak, I. Hertz-Picciotto, R. Hansen, I. Pessah, and J. Van de Water. Elevated plasma cytokines in autism spectrum disorders provide evidence of immune dysfunction and are associated with impaired behavioral outcome. *Brain, behavior, and immunity*, 25(1):40–45, 2011. 52

[11] Y. Assaf, T. Blumenfeld-Katzir, Y. Yovel, and P. J. Basser. Axcaliber: a method for measuring axon diameter distribution from diffusion mri. *Magnetic Resonance in Medicine*, 59(6):1347–1354, 2008. 47

[12] Y. Assaf and Y. Cohen. Non-mono-exponential attenuation of water and N-acetyl aspartate signals due to diffusion in brain tissue. *Journal of Magnetic Resonance*, 131(1):69–85, 1998. 46

[13] Y. Assaf and Y. Cohen. Chapter 7 - inferring microstructural information of white matter from diffusion MRI. In H. Johansen-Berg and T. E. Behrens, editors, *Diffusion MRI*, pages 127 – 146. Academic Press, San Diego, 2009. 28

[14] Y. Assaf, R. Freidlin, G. Rohde, and P. Basser. New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine*, 52(5):965–978, 2004. 44, 45, 46, 47, 48, 58, 85

[15] A. P. Association. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR®*. American Psychiatric Pub, 2000. 157

[16] G. Baird, E. Simonoff, A. Pickles, S. Chandler, T. Loucas, D. Meldrum, and T. Charman. Prevalence of disorders of the autism spectrum in a population cohort of children in south thames: the special needs and autism project (snap). *The Lancet*, 368(9531):210 – 215, 2006. 21

[17] R. B. Banati, J. Gehrmann, P. Schubert, and G. W. Kreutzberg. Cytotoxicity of microglia. *Glia*, 7(1):111–118, 1993. 52

[18] A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh. Clustering with Bregman divergences. *The Journal of Machine Learning Research*, 6:1705–1749, 2005. 57

[19] G. Bartzokis. Alzheimer's disease as homeostatic responses to age-related myelin breakdown. *Neurobiology of aging*, 32(8):1341–1371, 2011. 53

[20] P. J. Basser, J. Mattiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, 1994. 36

[21] T. Behrens, H. Berg, S. Jbabdi, M. Rushworth, and M. Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144–155, 2007. 86, 102

[22] T. Behrens, M. Woolrich, M. Jenkinson, H. Johansen-Berg, R. Nunes, S. Clare, P. Matthews, J. Brady, and S. Smith. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50(5):1077–1088, 2003. 45, 58

[23] M. K. Belmonte, G. Allen, A. Beckel-Mitchener, L. M. Boulanger, R. A. Carper, and S. J. Webb. Autism and abnormal development of brain connectivity. *The Journal of Neuroscience*, 24(42):9228–9231, 2004. 23

[24] C. F. Benjamin, J. M. Singh, S. P. Prabhu, and S. K. Warfield. Optimization of tractography of the optic radiations. *Human brain mapping*, 2012. 66

[25] O. Bergmann, G. Kindlmann, S. Peled, and C.-F. Westin. Two-tensor fiber tractography. In *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pages 796–799. IEEE, 2007. 55

[26] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, et al. Mr pulse sequences: What every radiologist wants to know but is afraid to ask1. *Radiographics*, 26(2):513–537, 2006. 33

[27] R. P. Cabeen, M. E. Bastin, and D. Laidlaw. Estimating constrained multi-fiber diffusion MR volumes by orientation clustering. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2013. 48

[28] P. Callaghan, C. Eccles, and Y. Xia. NMR microscopy of dynamic displacements: k-space and q-space imaging. *Journal of Physics E: Scientific Instruments*, 21(8):820, 1988. 38

[29] S. L. Codd and P. T. Callaghan. Spin echo analysis of restricted diffusion under generalized gradient waveforms: planar, cylindrical, and spherical pores with wall relaxivity. *Journal of Magnetic Resonance*, 137(2):358–372, 1999. 46

[30] M. P. Coleman and V. H. Perry. Axon pathology in neurological disease: a neglected therapeutic target. *Trends in neurosciences*, 25(10):532–537, 2002. 51

[31] A. Collard, S. Bonnabel, C. Phillips, and R. Sepulchre. An anisotropy preserving metric for dti processing. *arXiv preprint arXiv:1210.2826*, 2012. 42

[32] O. Commowick, V. Arsigny, A. Isambert, J. Costa, F. Dhermain, F. Bidault, P. Bondiau, N. Ayache, and G. Malandain. An efficient locally affine framework for the smooth registration of anatomical structures. *Medical Image Analysis*, 12(4):427–441, 2008. 65

[33] L. De Fossé, S. M. Hodge, N. Makris, D. N. Kennedy, V. S. Caviness, L. McGrath, S. Steele, D. A. Ziegler, M. R. Herbert, J. A. Frazier, et al. Language-association cortex asymmetry in autism and specific language impairment. *Annals of neurology*, 56(6):757–766, 2004. 80

[34] B. Deen and K. Pelphrey. Perspective: brain scans need a rethink. *Nature*, 491(7422):S20–S20, 2012. 23

[35] J. Dhillon. Differential entropic clustering of multivariate Gaussians. In *Advances in Neural Information Processing Systems*, volume 19, page 337. The MIT Press, 2007. 57, 58

[36] B. Efron and R. Tibshirani. Improvements on cross-validation : The . 632 + bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997. 87, 88, 89, 90, 91

[37] A. M. Fjell, K. B. Walhovd, T. T. Brown, J. M. Kuperman, Y. Chung, D. J. Hagler, V. Venkatraman, J. C. Roddey, M. Erhart, C. McCabe, et al. Multimodal imaging of the self-regulating developing brain. *Proceedings of the National Academy of Sciences*, 109(48):19620–19625, 2012. 103

[38] U. Frith. *Autism: Explaining the enigma*. Wiley Online Library, 1989. 23

[39] M. L. Ganz. The costs of autism, 2006. 21

[40] W. Gao, W. Lin, Y. Chen, G. Gerig, J. Smith, V. Jewells, and J. Gilmore. Temporal and spatial development of axonal maturation and myelination of white matter in the developing brain. *American journal of neuroradiology*, 30(2):290–296, 2009. 102

[41] V. Garcia, O. Commowick, G. Malandain, et al. A robust and efficient block-matching framework for non linear registration of thoracic ct images. *Proceedings of A Grand Challenge on Pulmonary Image Registration (EMPIRE'10), held in conjunction with MICCAI*, 10, 2010. 66

[42] D. H. Geschwind and P. Levitt. Autism spectrum disorders: developmental disconnection syndromes. *Current opinion in neurobiology*, 17(1):103–111, 2007. 23

[43] E. Gibson, A. Fenster, and A. Ward. Registration accuracy: How good is good enough? A statistical power calculation incorporating image registration uncertainty. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 643–650, 2012. 78

[44] H. Gudbjartsson and S. Patz. The rician distribution of noisy mri data. *Magnetic Resonance in Medicine*, 34(6):910–914, 1995. 112

[45] A. Guimond, J. Meunier, and J. P. Thirion. Average brain models: A convergence study. *Computer vision and image understanding*, 77(2):192–210, 2000. 66

[46] A. K. Gupta and D. K. Nagar. *Matrix variate distributions*, volume 104. CRC Press, 2000. 134

[47] N. Hadjikhani, R. M. Joseph, J. Snyder, and H. Tager-Flusberg. Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral cortex*, 16(9):1276–1282, 2006. 22

[48] A. Y. Hardan, S. Muddasani, M. Vemulapalli, M. S. Keshavan, and N. J. Minshew. An mri study of increased cortical thickness in autism. *The American journal of psychiatry*, 163(7):1290, 2006. 22

[49] T. Heads, M. Pollock, A. Robertson, W. Sutherland, and S. Allpress. Sensory nerve pathology in amyotrophic lateral sclerosis. *Acta neuropathologica*, 82(4):316–320, 1991. 102

[50] J. D. Heiss, E. Papavassiliou, M. J. Merrill, L. Nieman, J. J. Knightly, S. Walbridge, N. A. Edwards, and E. H. Oldfield. Mechanism of dexamethasone suppression of brain tumor-associated vascular permeability in rats. involvement of the glucocorticoid receptor and vascular permeability factor. *Journal of Clinical Investigation*, 98(6):1400, 1996. 52

[51] S. Huang and D. E. Ingber. The structural and mechanical complexity of cell-growth control. *Nature cell biology*, 1(5):E131–E138, 1999. 46

[52] F. Jansen, K. Vincken, A. Algra, P. Anbeek, O. Braams, M. Nellist, B. Zonnenberg, A. Jennekens-Schinkel, A. Van Den Ouweland, D. Halley, et al. Cognitive impairment in tuberous sclerosis complex is a multifactorial condition. *Neurology*, 70(12):916–923, 2008. 22

[53] S. Jbabdi, T. E. Behrens, and S. M. Smith. Crossing fibres in tract-based spatial statistics. *Neuroimage*, 49(1):249–256, 2010. 55

[54] S. S. Jeste, M. Sahin, P. Bolton, G. B. Ploubidis, and A. Humphrey. Characterization of autism in young children with tuberous sclerosis complex. *Journal of child neurology*, 23(5):520–525, 2008. 22

[55] B. Jeurissen, A. Leemans, J.-D. Tournier, D. K. Jones, and J. Sijbers. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human Brain Mapping*, 2012. 43, 47

[56] B. Jian, B. C. Vemuri, E. Özarslan, P. R. Carney, and T. H. Mareci. A novel tensor distribution model for the diffusion-weighted mr signal. *NeuroImage*, 37(1):164–176, 2007. 134

[57] K. Jordán. *Calculus of finite differences*. AMS Bookstore, 1965. 107

[58] A. Klin, D. J. Lin, P. Gorrindo, G. Ramsay, and W. Jones. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, 459(7244):257–261, 2009. 81

[59] B. Kreher, J. Schneider, I. Mader, E. Martin, J. Hennig, and K. Il'yasov. Multitensor approach for analysis and tracking of complex fiber configurations. *Magnetic resonance in medicine*, 54(5):1216–1225, 2005. 48, 86, 102, 113

[60] M. Kubicki, H. Park, C. Westin, P. Nestor, R. Mulkern, S. Maier, M. Niznikiewicz, E. Connor, J. Levitt, M. Frumin, et al. Dti and mtr abnormalities in schizophrenia: analysis of white matter integrity. *Neuroimage*, 26(4):1109–1118, 2005. 53

[61] S. Kumar, N. S. Mattan, and J. de Vellis. Canavan disease: a white matter disorder. *Mental retardation and developmental disabilities research reviews*, 12(2):157–165, 2006. 50

[62] A.-S. Lamantia and P. Rakic. Cytological and quantitative characteristics of four cerebral commissures in the rhesus monkey. *Journal of Comparative Neurology*, 291(4):520–537, 1990. 102

[63] W. W. Lewis, M. Sahin, B. Scherrer, J. M. Peters, R. O. Suarez, V. K. Vogel-Farley, S. S. Jeste, M. C. Gregas, S. P. Prabhu, C. A. Nelson, and S. K. Warfield. Impaired language pathways in tuberous sclerosis complex patients with autism spectrum disorders. *Cerebral Cortex*, 23(7):1526–1532, 2013. 66, 80, 127, 129

[64] D. Liang, S. Bhatta, V. Gerzanich, and J. M. Simard. Cytotoxic edema: mechanisms of pathological cell swelling. *Neurosurgical focus*, 22(5):E2, 2007. 51, 52

[65] N. F. Lori, T. E. Conturo, and D. Le Bihan. Definition of displacement probability and diffusion time in q-space magnetic resonance measurements that use finite-duration diffusion-encoding gradients. *Journal of Magnetic Resonance*, 165(2):185–195, 2003. 38

[66] J. G. Malcolm, O. Michailovich, S. Bouix, C.-F. Westin, M. E. Shenton, and Y. Rathi. A filtered approach to neural tractography using the watson directional function. *Medical Image Analysis*, 14(1):58, 2010. 102

[67] M. Martín-Fernández, E. Muñoz-Moreno, L. Cammoun, J.-P. Thiran, C.-F. Westin, and C. Alberola-López. Sequential anisotropic multichannel wiener filtering with rician bias correction applied to 3d regularization of dwi data. *Medical Image Analysis*, 13(1):19–35, 2009. 119

[68] K. L. Miller and *et al.* Diffusion imaging of whole, post-mortem human brains on a clinical MRI scanner. *Neuroimage*, 57(1):167–181, 2011. 86

[69] N. J. Minshew and D. L. Williams. The new neurobiology of autism: cortex, connectivity, and neuronal organization. *Archives of neurology*, 64(7):945, 2007. 23

[70] A. Molinaro, R. Simon, and R. Pfeiffer. Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, 21(15):3301–3307, 2005. 89

[71] R. J. Muirhead. *Aspects of multivariate statistical theory*, volume 197. Wiley. com, 2009. 134

[72] R. V. Mulkern, H. Gudbjartsson, C.-F. Westin, H. P. Zengingonul, W. Gartner, C. R. Guttmann, R. L. Robertson, W. Kyriakos, R. Schwartz, D. Holtzman, et al. Multi-component apparent diffusion coefficients in human brain^2. *NMR in Biomedicine*, 12:51–62, 1999. 46

[73] N. Muller and M. Schwarz. Schizophrenia as an inflammation-mediated dysbalance of glutamatergic neurotransmission. *Neurotoxicity research*, 10(2):131–148, 2006. 53

[74] C. Neuman. Spin echo of spins diffusing in a bounded medium. *The Journal of Chemical Physics*, 60:4508, 1974. 46

[75] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25, 2001. 67, 68

[76] T. Niendorf, R. M. Dijkhuizen, D. G. Norris, M. van Lookeren Campagne, and K. Nicolay. Biexponential diffusion attenuation in various states of brain tissue: Implications for diffusion-weighted imaging. *Magnetic Resonance in Medicine*, 36(6):847–857, 1996. 45, 46

[77] M. Nilsson, D. van Westen, F. Ståhlberg, P. C. Sundgren, and J. Lätt. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. *Magnetic Resonance Materials in Physics, Biology and Medicine*, pages 1–26, 2013. 47

[78] S. Ourselin, A. Roche, S. Prima, and N. Ayache. Block matching: A general framework to improve robustness of rigid registration of medical images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000*, pages 557–566. Springer, 2000. 65

[79] E. Panagiotaki, T. Schneider, B. Siow, M. G. Hall, M. F. Lythgoe, and D. C. Alexander. Compartment models of the diffusion mr signal in brain white matter: a taxonomy and comparison. *Neuroimage*, 59(3):2241–2254, 2012. 45, 85

[80] O. Pasternak, M. Shenton, and C.-F. Westin. Estimation of extracellular volume from regularized multi-shell diffusion MRI. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 305–312, 2012. 45, 68, 102, 103, 107

[81] O. Pasternak, N. Sochen, Y. Gur, N. Intrator, and Y. Assaf. Free water elimination and mapping from diffusion MRI. *Magnetic Resonance in Medicine*, 62(3):717–730, 2009. 102, 107

[82] O. Pasternak, C.-F. Westin, S. Bouix, L. J. Seidman, J. M. Goldstein, T.-U. W. Woo, T. L. Petryshen, R. I. Mesholam-Gately, R. W. McCarley, R. Kikinis, et al. Excessive extracellular volume reveals a neurodegenerative pattern in schizophrenia onset. *The Journal of Neuroscience*, 32(48):17365–17372, 2012. 45, 53, 68, 126

[83] J. Peters, M. Sahin, V. Vogel-Farley, S. Jeste, C. Nelson, M. Gregas, S. Prabhu, B. Scherrer, and S. K. Warfield. Loss of white matter microstructural integrity is associated with adverse neurological outcome in tuberous sclerosis complex. *Academic Radiology*, 19(1):17–25, 2012. 66, 80, 127

[84] J. M. Peters, M. Taquet, A. K. Prohl, B. Scherrer, A. M. van Eeghen, S. P. Prabhu, M. Sahin, and S. K. Warfield. Diffusion tensor imaging and related techniques in tuberous sclerosis complex: Review and future directions. *Future Neurology*, 8(5), 2013. 22

[85] J. M. Peters, M. Taquet, C. Vega, S. S. Jeste, I. Sanchez Fernandez, J. Tan, C. A. Nelson, M. Sahin, and S. K. Warfield. Brain functional networks in syndromic and non-syndromic autism: a graph theoretical study of EEG connectivity. *BMC medicine*, 11(1):54, 2013. 23, 80

[86] M. J. Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, UK*, 2009. 112, 146

[87] J. D. Power, K. A. Barnes, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *Neuroimage*, 59(3):2142–2154, 2012. 23

[88] J. Radua, M. L. Phillips, T. Russell, N. Lawrence, N. Marshall, S. Kalidindi, W. El-Hage, C. McDonald, V. Giampietro, M. J. Brammer, et al. Neural response to specific components of fearful faces in healthy and schizophrenic adults. *Neuroimage*, 49(1):939–946, 2010. 80

[89] Y. Rathi, M. Kubicki, S. Bouix, C.-F. Westin, J. Goldstein, L. Seidman, R. Mesholam-Gately, R. W. McCarley, and M. E. Shenton. Statistical analysis of fiber bundles using multi-tensor tractography: application to first-episode schizophrenia. *Magnetic resonance imaging*, 29(4):507–515, 2011. 55

[90] T. G. Reese, O. Heid, R. M. Weisskoff, and V. J. Wedeen. Reduction of eddy-current-induced distortion in diffusion mri using a twice-refocused spin echo. *Magn Reson Med*, 49(1):177–182, 2003. 69, 116

[91] K. Ridler, E. Bullmore, P. De Vries, J. Suckling, G. Barker, S. Meara, S. Williams, and P. Bolton. Widespread anatomical abnormalities of grey and white matter structure in tuberous sclerosis. *Psychological medicine*, 31(08):1437–1446, 2001. 78, 126

[92] G. Rippon, J. Brock, C. Brown, and J. Boucher. Disordered connectivity in the autistic brain: Challenges for the 'new psychophysiology'. *International Journal of Psychophysiology*, 63(2):164–172, 2007. 23

[93] E. Roach, F. J. DiMario, R. S. Kandt, and H. Northrup. Tuberous sclerosis consensus conference: recommendations for diagnostic evaluation. *Journal of child neurology*, 14(6):401–407, 1999. 157

[94] G. A. Rosenberg. Ischemic brain edema. *Progress in cardiovascular diseases*, 42(3):209–216, 1999. 51

[95] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010. 24

[96] J. Ruiz-Alzola, C.-F. Westin, S. K. Warfield, C. Alberola, S. Maier, and R. Kikinis. Nonrigid registration of 3D tensor medical data. *Medical Image Analysis*, 6(2):143–161, 2002. 60

[97] M. Rutter. Incidence of autism spectrum disorders: Changes over time and their meaning. *Acta Paediatrica*, 94(1):2–15, 2005. 21

[98] B. Scherrer, A. Schwartzman, M. Taquet, S. P. Prabhu, M. Sahin, A. Akhondi-Asl, and S. K. Warfield. Characterizing the DIstribution of Anisotropic MicrO-structural eNvironments with Diffusion-weighted imaging (DIAMOND). In *Proc. of the 16th Int Conf Med Image Comput Comput Assist Interv (MICCAI)*, page to appear, 2013. 47, 48, 85, 131, 134, 136, 138, 146

[99] B. Scherrer, M. Taquet, O. Afacan, and S. K. Warfield. Does the signal arising from a single fascicle significantly deviates from a monoexponential decay with a clinical scanner? In *Proc. the 21th International Symposium on Magnetic Resonance in Medicine (ISMRM)*, 2013. 46

[100] B. Scherrer, M. Taquet, and S. K. Warfield. Reliable selection of the number of fascicles in diffusion images by estimation of the generalization error. In *Information Processing in Medical Imaging, 23rd biennial International Conference on*, pages 742–753. Springer, 2013. 30, 113, 115

[101] B. Scherrer and S. K. Warfield. Why multiple b-values are required for multi-tensor models. evaluation with a constrained log-euclidean model. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 1389–1392. IEEE, 2010. 102

[102] B. Scherrer and S. K. Warfield. Parametric representation of multiple white matter fascicles from cube and sphere diffusion MRI. *PLoS one*, 7(11):e48232, 2012. 35, 45, 48, 69, 85, 86, 93, 102, 103, 107, 112, 113, 115

[103] T. Schultz. Learning a reliable estimate of the number of fiber directions in diffusion MRI. In *Med Image Comput and Comput Assis Interv*, volume 15(Pt 3) of *Lecture Notes in Computer Science*, pages 493–500, Nice, France, 2012. Springer. 113

[104] T. Schultz. Towards resolving fiber crossings with higher order tensor inpainting. In *New Developments in the Visualization and Processing of Tensor Fields*, pages 253–265. Springer, 2012. 107

[105] T. Schultz, C.-F. Westin, and G. Kindlmann. Multi-diffusion-tensor fitting via spherical deconvolution: a unifying framework. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 674–681. Springer, 2010. 48, 86, 102

[106] A. Schwartzman. *Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data*. PhD thesis, Stanford University, 2006. 41, 134

[107] A. Schwartzman, W. F. Mascarenhas, and J. E. Taylor. Inference for eigenvalues and eigenvectors of gaussian symmetric matrices. *The Annals of Statistics*, pages 2886–2919, 2008. 109

[108] K. K. Seunarine and D. C. Alexander. Chapter 4 - multiple fibers: Beyond the diffusion tensor. In H. Johansen-Berg and T. E. Behrens, editors, *Diffusion MRI*, pages 55–72. Academic Press, San Diego, 2009. 101

[109] J. Shao. Linear Model Selection by Cross-Validation. *Journal of the American Statistical Association*, 88(442):486–494, 1993. 86

[110] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000. 58

[111] G. Simon, A. Lendasse, V. Wertz, M. Verleysen, et al. Fast approximation of the bootstrap for model selection. In *Esann*, pages 475–480, 2003. 99

[112] L. Smeeth, C. Cook, E. Fombonne, L. Heavey, L. C. Rodrigues, P. G. Smith, and A. J. Hall. Rate of first recorded diagnosis of autism and other pervasive developmental disorders in united kingdom general practice, 1988 to 2001. *BMC medicine*, 2(1):39, 2004. 21

[113] S. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. Nichols, C. Mackay, K. Watkins, O. Ciccarelli, M. Cader, P. Matthews, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505, 2006. 55, 128

[114] S. M. Smith and T. E. Nichols. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, 44(1):83–98, 2009. 126

[115] S.-K. Song, S.-W. Sun, W.-K. Ju, S.-J. Lin, A. H. Cross, and A. H. Neufeld. Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *Neuroimage*, 20(3):1714–1722, 2003. 50

[116] S.-K. Song, S.-W. Sun, M. J. Ramsbottom, C. Chang, J. Russell, and A. H. Cross. Dysmyelination revealed through mri as increased radial (but unchanged axial) diffusion of water. *Neuroimage*, 17(3):1429–1436, 2002. 38, 50

[117] A. Stamm, P. Perez, and C. Barillot. Diffusion directions imaging (DDI). *INRIA Rapport de recherche*, 2011. 58

[118] G. J. Stanisz, G. A. Wright, R. M. Henkelman, and A. Szafer. An analytical model of restricted diffusion in bovine optic nerve. *Magnetic Resonance in Medicine*, 37(1):103–111, 1997. 45

[119] E. Stejskal and J. Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1):288, 1965. 34

[120] W. J. Streit. Microglial senescence: does the brain's immune system have an expiration date? *Trends in neurosciences*, 29(9):506–510, 2006. 52

[121] R. O. Suarez, O. Commowick, S. P. Prabhu, and S. K. Warfield. Automated delineation of white matter fiber tracts with a multiple region-of-interest approach. *Neuroimage*, 59(4):3690–3700, 2012. 80

[122] M. Taquet, B. Macq, and S. K. Warfield. Spatially adaptive log-Euclidean polyaffine registration based on sparse matches. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 590–597, 2011. 76

[123] M. Taquet, B. Macq, and S. K. Warfield. A generalized correlation coefficient: Application to DTI and multi-fiber DTI. In *Mathematical Methods in Biomedical Image Analysis (MMBIA), 2012 IEEE Workshop on*, pages 9–14. IEEE, 2012. 30, 56

[124] M. Taquet and J. M. Peters. Le réseau cérébral fonctionnel des enfants atteints d'autisme: Redondance et déconnexion. *médecine/sciences*, 29(6-7):567–569, 2013. 23

[125] M. Taquet, B. Scherrer, C. Benjamin, S. Prabhu, B. Macq, and S. K. Warfield. Interpolating multi-fiber models by Gaussian mixture simplification. In *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pages 928–931. IEEE, 2012. 30, 56, 57, 111

[126] M. Taquet, B. Scherrer, N. Boumal, B. Macq, and S. K. Warfield. Estimation of a multi-fascicle model from single b-value data with a population-informed prior. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pages 695–702. Springer, 2013. 30, 93, 115, 126

[127] M. Taquet, B. Scherrer, O. Commowick, J. Peters, M. Sahin, B. Macq, and S. Warfield. Registration and analysis of white matter group differences with a multi-fiber model. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, volume 7512 of *Lecture Notes in Computer Science*, pages 313–320. Springer Berlin Heidelberg, 2012. 30, 56, 111

[128] M. Taquet, B. Scherrer, O. Commowick, J. Peters, M. Sahin, B. Macq, and S. Warfield. A mathematical framework for the registration and analysis of multi-fascicle models for population studies of the brain microstructure. *Medical Imaging, IEEE Transactions on*, 33(2):504–517, Feb 2014. 123

[129] M. Taquet, B. Scherrer, B. Macq, and S. K. Warfield. Multi-Fascicle Model Reconstruction from Acquisitions at a Single b-value with a Population-Informed Prior. In *Proc. the 21st International Symposium on Magnetic Resonance in Medicine (ISMRM)*, 2013. 30

[130] J. Tournier, F. Calamante, A. Connelly, et al. Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4):1459–1472, 2007. 124

[131] J. Tournier, S. Mori, and A. Leemans. Diffusion tensor imaging and beyond. *Magnetic Resonance in Medicine*, 2011. 42, 43, 44

[132] D. Tuch, T. Reese, M. Wiegell, N. Makris, J. Belliveau, and V. Wedeen. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48(4):577–582, 2002. 48, 58, 102, 104, 119, 124

[133] A. Unterberg, J. Stover, B. Kress, and K. Kiening. Edema and brain trauma. *Neuroscience*, 129(4):1019–1027, 2004. 51

[134] P. Vangelderen, D. DesPres, P. Vanzijl, and C. Moonen. Evaluation of restricted diffusion in cylinders. phosphocreatine in rabbit leg muscle. *Journal of Magnetic Resonance, Series B*, 103(3):255–260, 1994. 46

[135] D. L. Vargas, C. Nascimbene, C. Krishnan, A. W. Zimmerman, and C. A. Pardo. Neuroglial activation and neuroinflammation in the brain of patients with autism. *Annals of neurology*, 57(1):67–81, 2005. 53, 81, 127

[136] S. B. Vos, D. K. Jones, B. Jeurissen, M. A. Viergever, and A. Leemans. The influence of complex white matter architecture on the mean diffusivity in diffusion tensor mri of the human brain. *NeuroImage*, 59(3):2208–2216, 2012. 43

[137] Y. Wang, Q. Wang, J. P. Haldar, F.-C. Yeh, M. Xie, P. Sun, T.-W. Tu, K. Trinkaus, R. S. Klein, A. H. Cross, et al. Quantification of increased cellularity during inflammatory demyelination. *Brain*, 134(12):3590–3601, 2011. 48, 51, 53, 126, 127

[138] G. L. Wenk et al. Neuropathologic changes in alzheimer's disease. *Journal of Clinical Psychiatry*, 64:7–10, 2003. 53

[139] K. P. Whittall, A. L. Mackay, D. A. Graeb, R. A. Nugent, D. K. Li, and D. W. Paty. In vivo measurement of t2 distributions and water contents in normal human brain. *Magnetic Resonance in Medicine*, 37(1):34–43, 1997. 35

[140] S. Wills, M. Cabanlit, J. Bennett, P. Ashwood, D. Amaral, and J. VAN DE WATER. Autoantibodies in autism spectrum disorders (asd). *Annals of the New York Academy of Sciences*, 1107(1):79–91, 2007. 53

[141] V. Wong and P.-L. Khong. Tuberous sclerosis complex: correlation of magnetic resonance imaging (mri) findings with comorbidities. *Journal of child neurology*, 21(2):99–105, 2006. 22

[142] M. Yoichi Katayama and T. Kawamata. Edema fluid accumulation within necrotic brain tissue as a cause of the mass effect of cerebral contusion in head trauma patients. In *Brain Edema XII*, pages 323–327. Springer, 2003. 52

[143] W. Zhan and Y. Yang. How accurately can the diffusion profiles indicate multiple fiber orientations? a study on general fiber crossings in diffusion mri. *Journal of magnetic resonance (San Diego, Calif.: 1997)*, 183(2):193, 2006. 102

[144] H. Zhang, B. Avants, P. Yushkevich, J. Woo, S. Wang, L. McCluskey, L. Elman, E. Melhem, and J. Gee. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. *Medical Imaging, IEEE Transactions on*, 26(11):1585–1597, 2007. 60

[145] H. Zhang, P. L. Hubbard, G. J. Parker, and D. C. Alexander. Axon diameter mapping in the presence of orientation dispersion with diffusion mri. *Neuroimage*, 56(3):1301–1315, 2011. 46

[146] H. Zhang, T. Schneider, C. A. Wheeler-Kingshott, and D. C. Alexander. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage*, 2012. 47, 58, 85

[147] X. Zhu, Y. Gur, W. Wang, and P. T. Fletcher. Model selection and estimation of multi-compartment models in diffusion mri with a rician noise model. In *Information Processing in Medical Imaging*, pages 644–655. Springer, 2013. 86

[148] B. Zikopoulos and H. Barbas. Changes in prefrontal axons may disrupt the network in autism. *The Journal of Neuroscience*, 30(44):14595–14609, 2010. 102