Improving Sensitivity and Specificity in Diffusion MRI Group Studies

Thesis for the Degree of Doctor of Philosophy

Emmanuel Vallée



[1]Oxford Centre for Functional Magnetic Resonance Imaging of the Brain

Nuffield Department of Clinical Neurosciences

University of Oxford

Wadham College

Academic advisors:

Saad Jbabdi[1], Stephen M. Smith[1]

Submitted: Michaelmas 2017

A toi maman

Tu es aimée, le reste ne fait que passer.

Acknowledgements

Thanks to all of you who supported me through these challenging and surprising years.

Thank you Steve for giving me the exceptional opportunity to be part of FMRIB. This is a keystone of my life. You have been very generous and sensitive towards me. You have taught me how to be thorough, persistent and humble.

Merci Saad, for your patience and efforts in teaching me how to decipher the unknown. Your passion and curiosity continuously contaminated my thoughts. Our meetings were a pleasure, they always made me feel better.

I want to express my gratitude towards my examiners, Stam and Demian, for spending a long time reading and correcting this thesis. I also want to say that I particularly enjoyed the fruitful discussion we had during the defence.

I am very grateful to the entire FMRIB analysis group for the friendly and intellectually stimulating environment. I could not expect a better place to discover the academic world.

Thanks to Karla, Stam and Jesper for providing me with their guidance during the intermediate examination processes. It was very nice of you to put interest on in my research and providing me with valuable and original feedback.

Also a very big thank you to Wenchuan for spending so much time with me acquiring data, to Gwen for helping me through the analysis of Alzheimer's disease data and generally for being supportive and caring, and to Mark and Janine for giving me the opportunity to attend and tutor the FSL and FMRIB course.

Thanks you Sue for being so cheerful, funny and reassuring.

A very special thank you to you Moises, for being my friend and colleague from the very first day. I know it has not always been easy for you.

Thanks to Paul, Ludo, Fidel and Jelena for the many lunches and the too few beers we shared.

None of this would have been initiated without the seed planted in my mind by Agnies, Isabelle Bloch and Olivier Commowick.

I particularly want to express my gratitude towards my friends who are my safe place for happiness, fun and emotions. Thank you Joce, Andrei, Kyoko, Ioni, Carina, Richard, Benoit, Olive, Florence, Fred, Karim, Matthieu, Clara, Gines... for your unconditional support and infinite tolerance.

A very special thank you to Francy for sharing my life during those four years and for helping me to complete this thesis. I admire your constant support and encouragement, continually willing to proofread the same pages over and over again without any complain, ever.

Above all, I would like to thank my mother, my brother, Malo, and Fred who have been a continuous source of motivation and love.

Improving Sensitivity and Specificity in Diffusion MRI Group Studies

Emmanuel Vallée
Wadham College
University of Oxford

Abstract

Diffusion MRI provides a unique opportunity to study the brain tissue architecture at a microscopic level. More specifically, it allows to infer biophysical properties of the axons in the white matter in-vivo. Microstructural parameters are widely used in multi-subject studies to track pathological processes, follow normal development and aging, or investigate behaviour. This thesis aims to identify and potentially address the limitations and pitfalls in voxelwise comparison of diffusion MRI parameters across subjects.

To allow for accurate matching of brain structures across subjects, non-linear transformation that spatially aligns the data is required. We demonstrate that using advanced registration methods, we can outperform the standard registration-projection approach both in terms of sensitivity and specificity.

The coarse resolution of the data typically causes partial volume effects that bias the diffusion parameters and potentially mislead the interpretation of a group study outcome. We provide evidence that these effects can be addressed by constraining the diffusion model parameter space, which leads to marginally lower sensitivity, but allows an accurate interpretation of the results. Additionally, we suggest that additional information inferred with a data driven approach might mitigate the loss in sensitivity.

Finally, we design an original tract-specific modelling framework that enables to estimate microstructural parameters unbiased by the presence of foreign fibre populations or tissues. We demonstrate the sensitivity of our method in a study relating microstructure and behaviour.

Contents

List of Figures **xi**

List of Tables **xv**

1 General introduction and background **1**

1.1 White matter in the human brain 2

1.2 Diffusion MRI 6

1.2.1 Diffusion 7

1.2.2 General principles of MRI 9

1.2.3 Diffusion MRI 10

1.3 Processing and analysis of diffusion MRI data 12

1.3.1 Image artefacts and pre-processing 13

1.3.2 Diffusion models 18

1.3.3 Tractography 24

1.4 Cross-subjects comparisons 25

1.4.1 Spatial normalisation 25

1.4.2 Statistical inference 27

1.5 Thesis outline 28

2 Evaluation of improved cross-subject registration **31**

2.1 Introduction 31

2.2 Materials and methods 35

2.2.1 Registration approaches 35

2.2.2 Datasets description and preprocessing 40

2.2.3 Template creation strategies 41
2.2.4 Performance evaluation 43
2.3 Results 47
2.3.1 Registration algorithm 47
2.3.2 Dataset quality 51
2.3.3 Templates 52
2.3.4 TBSS skeleton and projection step 54
2.3.5 Group study: Alzheimer's disease 58
2.4 Discussion 60
2.4.1 Overall performance 60
2.4.2 Dataset and template strategy 63
2.4.3 TBSS projection step 64
2.4.4 Group study: Alzheimer's disease 65
2.5 Conclusion 66

3 Free water: confound or biomarker? 67

3.1 Introduction 68
3.2 Materials and methods 72
3.2.1 Free Water Model 72
3.2.2 Simulations 75
3.2.3 Validation on in-vivo data 76
3.2.4 Group study: Alzheimer's disease 79
3.3 Results 81
3.3.1 Simulations 81
3.3.2 Validation on in-vivo data 89
3.3.3 Group study: Alzheimer's disease 100
3.4 Discussion 103
3.4.1 Identifiability of the FW model 103

3.4.2 Properties and relevance of FW in deep white matter 105
3.4.3 Implications for a typical group study 107
3.4.4 Limitations and future work 108
3.5 Conclusion 109

4 Machine learning approach to partial volume segmentation 111
4.1 Introduction 112
4.2 Background on machine learning techniques 114
4.2.1 Random forests 115
4.2.2 Support Vector Machines 116
4.2.3 Gaussian Process 117
4.2.4 Neural networks 118
4.3 Materials and methods 119
4.3.1 Data acquisition and pre-processing 120
4.3.2 Predicting tissues volume fraction 121
4.3.3 Comparison of the machine learning approaches 121
4.3.4 Mis-registration between T1 and diffusion space 122
4.3.5 Application to FW modelling 123
4.4 Results 123
4.4.1 Machine learning approaches optimisation and comparison . . 123
4.4.2 Mis-registration between T1 and diffusion space 135
4.4.3 Application to FW modelling 137
4.5 Discussion 140
4.5.1 Machine learning approaches 140
4.5.2 Diffusion features 141
4.5.3 Diffusion data segmentation 141
4.5.4 Application to FW modelling 142
4.5.5 Implications for the future 143

4.6 Conclusion 144

5 Tract specific micro-structural modelling 145

5.1 Introduction 146

5.2 Materials and methods 149

5.2.1 Model and spatial regularisation 149

5.2.2 Simulations 152

5.2.3 In-vivo 153

5.2.4 Group study 153

5.3 Results 155

5.3.1 Simulations 155

5.3.2 In-vivo 156

5.3.3 Group study 161

5.4 Discussion 166

5.4.1 Spatial regularisation 167

5.4.2 Specificity of the model 168

5.4.3 Sensitivity to micro-structural changes 169

5.4.4 Group study 170

5.5 Conclusion 171

6 General conclusion 173

Bibliography 177

List of Figures

1.1 Schematic representation of a neuron 3

1.2 Transmission electron micrograph of an axon and of its surrounding . 4

1.3 Brownian's motion of a particle 8

1.4 Spin-echo experiment 10

1.5 Pulsed gradients spin-echo experiment 11

1.6 Susceptibility-induced distortions 14

1.7 Eddy-currents induced distortions 16

1.8 Within volume and within slice movement artefacts 17

2.1 Pipeline for the construction of an unbiased study specific template . 42

2.2 Representation of the tracts used across the performance evaluation . 44

2.3 Spatial correlation difference as a function of the registration error . 45

2.4 Corticospinal tract overlap across subjects 46

2.5 Spatial correlation for 27 tracts on the Whitehall dataset 48

2.6 Tract overlap across subjects 50

2.7 Spatial correlation for the three datasets 51

2.8 Illustration of the templates 53

2.9 Tract overlap across subjects, with and without TBSS projection . . . 57

2.10 TBSS skeleton derived from mean FA and density maps 57

2.11 TBSS result with FNIRT, oFNIRT and DTITK on Alzheimer's disease 59

3.1 MD histogram for the Whitehall dataset 74

3.2 FA and MD mean difference (between models) with the simulated ground truth . . . 83
3.3 FA and MD root mean squared error (between models) with the simulated ground truth . . . 84
3.4 FA and MD values compared with the simulated ground truth . . . 86
3.5 RMSE for FA_{FW}, MD_{FW} and FWF . . . 87
3.6 Posterior distributions of MD_{FW} for different values of FA . . . 89
3.7 FA and MD maps for DTI, IR, our model and FWE method . . . 90
3.8 FA and MD histograms for DTI, IR, our model and FWE method . . 92
3.9 Scatter plot for the FA and MD obtained with DTI, our FW model and FWE method . . . 93
3.10 Bland and Altman plots for the FA and MD obtained with DTI, our FW model and FWE method . . . 95
3.11 FWF estimated by our model and FWE method compared to T1 CSF PVE . . . 95
3.12 Volume fraction and T1 relaxation time of the FW compartment . . . 97
3.13 Volume fraction, diffusivity and spatial distribution of the FW compartment . . . 99
3.14 TBSS results with DTI and FW model . . . 101

4.1 Random forest hyperparameter optimisation - Number of trees and features . . . 124
4.2 Random forest hyperparameter optimisation - Tree depth, number of features . . . 125
4.3 SVR hyperparameters optimisation - γ and C . . . 126
4.4 Gaussian process hyperparameters optimisation - noise level and length scale . . . 127
4.5 Neural network architecture . . . 128

4.6 Neural network optimisation loss-function 129
4.7 Illustration of the model predictions for 2 features 129
4.8 Diffusion features importance 131
4.9 Impact of the additional features on the performance 132
4.10 Comparison of the four ML approaches 132
4.11 Segmentation maps predicted with the 4 ML methods 134
4.12 Impact of the mis-registration on the RF performance 136
4.13 Prediction of T1w image intensities 137
4.14 Comparison of the FA and MD estimated by the FW model 139

5.1 Diagram of the model fitting steps 151
5.2 Glyph representation of the simulations results 157
5.3 Plot of the error for micro-structural parameters and signal predictions 158
5.4 Glyph map of the zeppelin representing the SLF and CST 159
5.5 Glyph map of the zeppelin representing the FMI 160
5.6 Plot of the micro-structural parameters and signal predictions 161
5.7 Cingulum voxels with significant correlation with "negativity" 165
5.8 Projection of the cingulum voxels showing significant correlation with "negativity" 166

List of Tables

2.1 FNIRT Parameters for original and optimal settings 36

2.2 Acquisition parameters for the three datasets 40

2.3 Average spatial correlation obtained with FNIRT4D, FNIRT, oFNIRT and DTITK on the Whitehall dataset 48

2.4 Average tract overlap across subjects obtained with FNIRT4D, FNIRT, oFNIRT and DTITK on the Whitehall dataset 50

2.5 Overall performances of the different algorithms and templates for whitehall dataset 54

2.6 Effect of the projection step on the spatial correlation 55

2.7 Average tract overlap across subjects, with and without TBSS projection 56

3.1 Acquisition parameters for the acquisition schemes 79

5.1 Behavioral measurements correlation with "negativity" 162

5.2 Tracts showing significant negative correlation with "negativity" . . . 164

1

General introduction and background

The brain is a superstar. News headlines are increasingly reporting findings on "the most complex object of the universe" [Murray, 2012]. But in spite of the outstanding progresses in scientific fields such as cellular biology, engineering, physics, and computing, the neurobiological bases of pathologies, behaviour and development remain largely unknown [Gorman, 2014]. The work presented in this thesis intends to understand and potentially address some of the pitfalls and limitations in neuroimaging studies.

Magnetic resonance imaging (MRI) is a unique tool that allows us to study brain structure and function in-vivo. MRI group studies typically aim at inferring correlations between MRI measurements and, e.g, development, behaviour, or disease at a population level. However, the complex MRI machinery comprises numerous pitfalls and approximations. In this thesis, we concentrated our efforts on the MRI processing and analysis steps that are necessary to perform group studies.

Our research focuses on diffusion MRI (dMRI), a branch of MRI that aims at retrieving micro-structural properties and localisation of the brain connections from the movement of water molecules. The main feature of diffusion MRI is that it uses a macroscopic probe to infer microscopic features of the brain tissues, e.g. the

white matter (WM). The nature of this inverse problem, the difference in scale, and the complexity of the cerebral structures bring several difficulties. The goal of our research study is to increase the sensitivity and specificity of the standard methods involved in voxel-wise cross-subject analysis of dMRI data.

In this introductory chapter, we briefly describe the brain WM function, composition, and structure. Then, we give an overview of the physical principles and methods involved in dMRI data acquisition, processing, and analysis. Finally, we describe standard approaches allowing cross-subject comparison of local micro-structure.

1.1 White matter in the human brain

The neuron The neuron is the functional unit of the nervous system (Figure 1.1). It is an excitable and conductive cell, meaning that it can react to stimuli and transmit electrical pulses. A typical neuron is composed of three main entities: the cell body, the dendrites and a unique axon. The cell body, also called soma, contains the nucleus and connects the dendrites to the axon. The dendrites are extensions of the cell with a tree-like structure. They receive messages from other cells via their synaptic connections. The out-going information is transmitted via the axon and its terminal branches which form junctions with other cells.

In the human brain, axons can form long range connections, reaching over several centimetres. To ensure a fast and reliable transmission of the electrical signal, the axons are insulated by a myelin sheath. The latter is composed of proteins and lipid tissue, created and supported by Schawnn cells in the peripheral nervous system and oligodendrocytes in the central nervous system. The myelin sheath exhibits some regular discontinuities called nodes of Ranvier. They are short, unmyelinated areas where the action potentials are regenerated, resulting in a faster transmission.

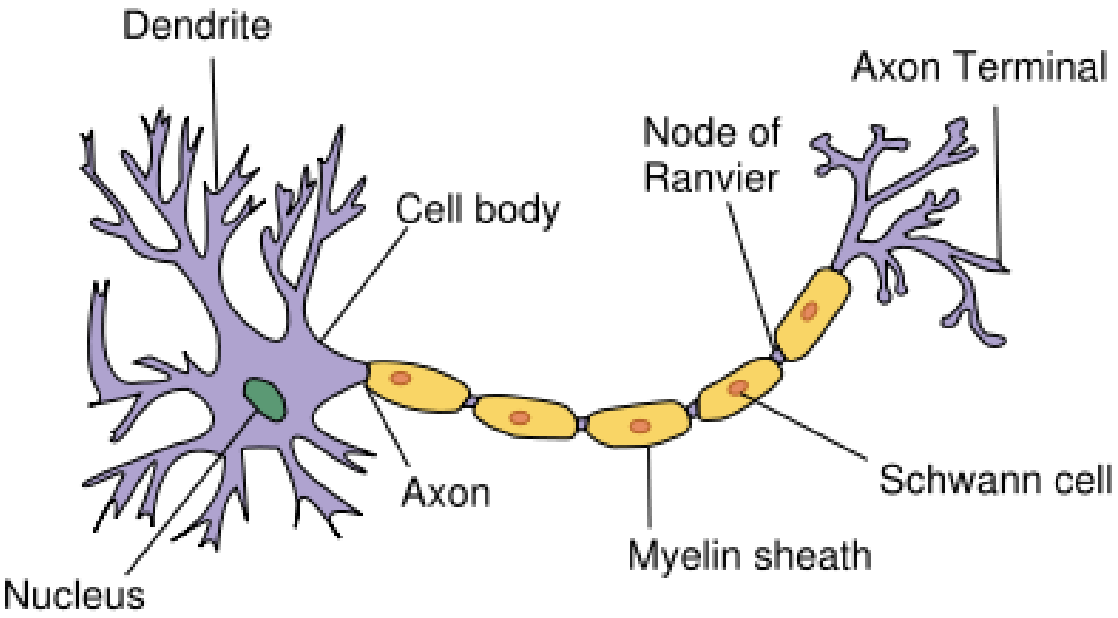


Figure 1.1: Schematic representation of a neuron. Simplified view of a neuron with a myelinated axon. Figure licensed under Creative Commons license via Wikimedia Commons.

Myelinated axons Long range myelinated axons are the main component of the brain white matter. In Figure 1.2, we show a transmission electron micrograph of a myelinated axon in its environment. The picture can grossly be decomposed into: the axon body, its myelin sheath, and the extra-axonal space.

The myelin sheath is a lipidic envelope that develops around some axons during the myelination process. Its role is to protect the axon and to ensure electrical isolation. The thickness of the myelin sheath is in the range of $0.1 - 1\mu$m, and the myelinated axons have a diameter between 1 and 15 μm. These two quantities and their ratio are of utmost importance, as they are directly related to the transmission speed of the nervous signal [Waxman and Bennett, 1972].

The extra-axonal space provides insulation, protection and support to the axons. It is mostly composed of astrocytes, oligodendrocytes and microglia. The astrocytes perform many functions, such as maintaining the ion balance, providing nutrients to the nervous tissue. The oligodendrocytes produce and maintain the integrity of the myelin sheath. The microglia are the main immune defense cells in the central

nervous system.

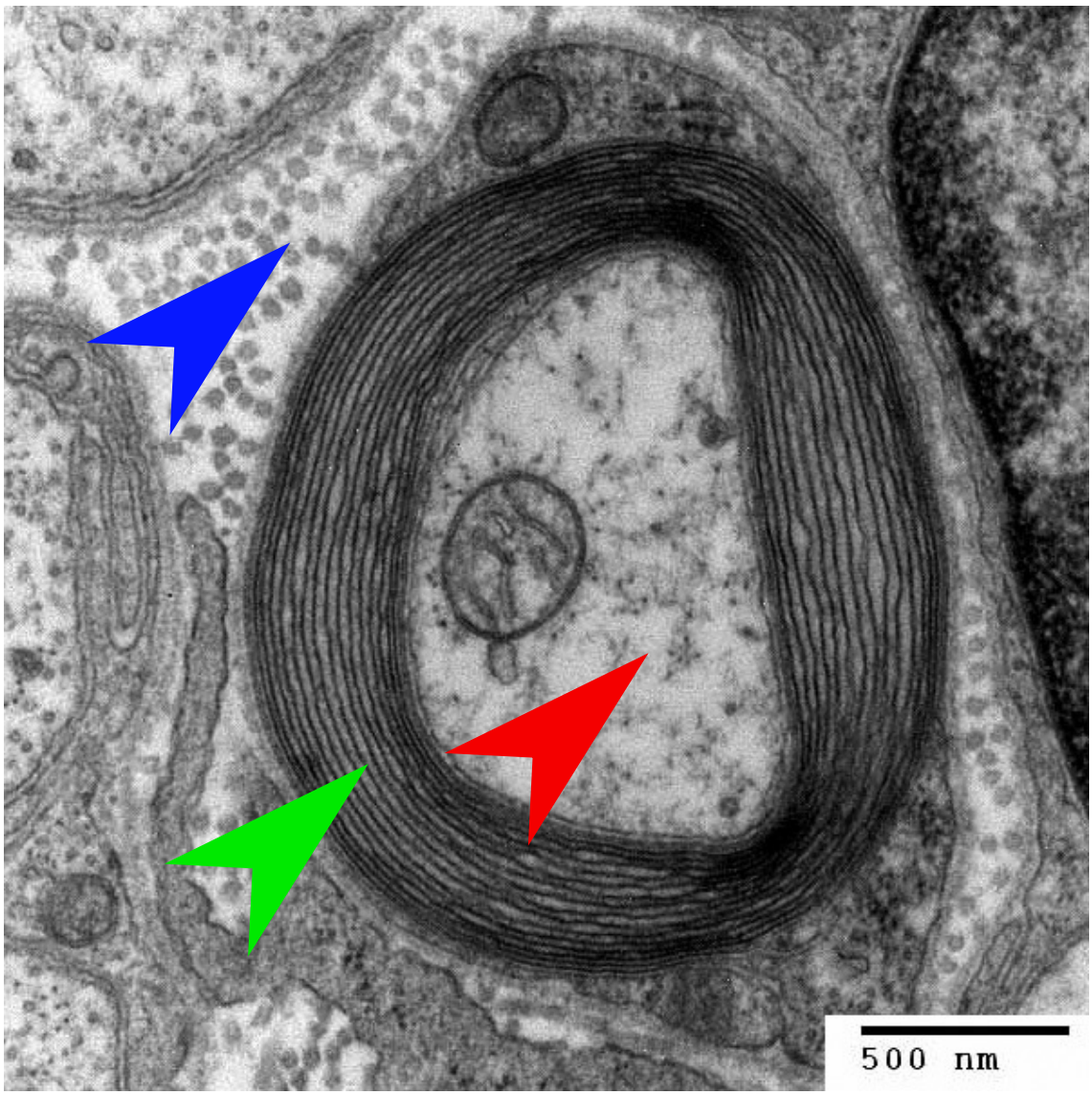


Figure 1.2: Transmission electron micrograph of an axon and of its surrounding. We can see three main compartments: the intra-axonal space (red), the myelin sheath (green) and the extra-axonal space (blue). Figure licensed under Creative Commons license via Wikimedia Commons.

White matter anatomy The human brain comprises between 80 and 100 billion neurons, located in the cortical and sub-cortical grey matter. The grey matter can be decomposed into functional areas which are strongly interconnected by axons composing the white matter. The latter presents a particular structure, as the axons are typically organised in bundles, called fasciculi. The degree of myelination of

the axons is very variable and it primarily depends on the axon diameter and fibre bundle. Anatomical dissections differentiated four main types of fibre bundles, based on their origin and location:

- brainstem fibres such as the inferior, middle, and inferior cerebellar penduncle, and medial lemniscus.
- projection fibres, connecting sub-cortical and cortical areas: acoustic radiations, anterior thalamic radiations, cortico-spinal tract, superior thalamic radiations and posterior thalamic radiations.
- association fibres connect different functional areas within the same hemisphere. The major association tracts are: superior longitudinal fasciculus, inferior longitudinal fasciculus, inferior fronto-occipital fasciculus, uncinate fasciculus, cingulum bundle and the fornix
- commisural fibres connect functional areas across hemispheres: corpus callosum, anterior commissure and posterior commissure

We do not intend to describe the anatomy of the WM tracts in detail, as it has already been extensively documented, e.g. in [Catani and Thiebaut de Schotten, 2008]. The identification of the functions supported by the tracts is an active topic of research, notably enabled by advances in functional and diffusion MRI [Supekar et al., 2010].

White matter integrity Changes in the WM micro-structure affect how information is carried through the brain, and therefore have an impact on its global functioning. The myelination process primarily occurs during the brain maturation, from the foetal phase to early adulthood, and at a different rate for each WM bundle [Lebel et al., 2008]. Changes in the WM micro-structure can also be observed in normal ageing and, interestingly, in the process of learning (e.g., to juggle) [Scholz et al., 2009].

Several pathologies have been linked to alterations of white matter microstructure. The latter can suffer from changes in the myelin sheath, axonal degeneration or changes in the extra-axonal space.

Multiple sclerosis is the most common demyelinating disease, where the immune system triggers a neuro-inflammatory response that eventually damages the myelin sheath throughout the WM [Hauser and Oksenberg, 2006]. Thanks to recent advances in neuroimaging, some localised demyelinating processes have been identified in diseases such as schizophrenia [Kubicki et al., 2005] or Alzheimer's disease [Douaud et al., 2011].

Axonal degeneration (e.g. Wallerian degeneration) is a process where a local axonal damage propagates upstream along the axon towards the cell body. This process was identified as a result of direct traumatic event (e.g., traumatic brain injury) [Povlishock, 1992], and also in diseases such as amyotrophic lateral sclerosis [Fischer and Glass, 2007] and multiple sclerosis [Peterson and Fujinami, 2007]. First, the axonal membrane breaks apart and the axon density is reduced. In later stages of the process, the microglial cells in the extra-axonal space are activated to remove the axonal and myelin debris.

The glial cells monitor their environment and change their morphology when activated. They typically respond to neuroinflammation that can be caused by e.g. traumatic injury, toxic metabolites, microbes and viruses.

We describe in the next section how diffusion MRI can potentially identify these WM changes.

1.2 Diffusion MRI

Diffusion Weighted Imaging (DWI) is an MRI modality sensitive to the mean displacement of water molecules at the microscopic scale [Le Bihan et al., 1986].

Clinically, it is widely employed for e.g. the diagnosis and localisation of strokes [Moseley et al., 1990], multiple sclerosis lesions [Larsson et al., 1992], and for pre-surgical planning [Clark et al., 2003]. In the research context, one of the main applications of dMRI is to relate a given disorder (or variable) with some changes in the micro-structural parameters in a group a subjects.

1.2.1 Diffusion

Diffusion is a natural mass transport phenomenon which results in molecular or particle movement without bulk motion. We can imagine it as an ink drop introduced into a glass of water. The drop will appear compact at the beginning and the ink particles will spread radially over time with a spherical profile. This physical phenomenon is described by Fick's first law [Fick, 1855]:

$$J = -D\nabla C \tag{1.1}$$

with J the net flow vector, D the diffusion coefficient and C the particle's concentration.

The particles tend to move to regions where the concentration is lower, in the same way heat flows from high to low temperatures. The diffusion coefficient relates to the speed at which the process occurs. It is an intrinsic property of the medium and it depends on the temperature, particle size, and micro-structural properties of the environment.

At the molecular level, diffusion is due to collisions between atoms or molecules in liquid or gaseous state. It happens even when the system is in equilibrium state, for instance in a homogeneous fluid with constant temperature and pressure. That could appear as incompatible with Fick's first law, which states that the flow is null

in this case. In fact, all the micro-movements cancel each other, resulting in a null global flow. This phenomenon was observed for the first time in 1828 by Robert Brown [Brown, 1828], who saw that pollen particles in water were moving without apparent cause.

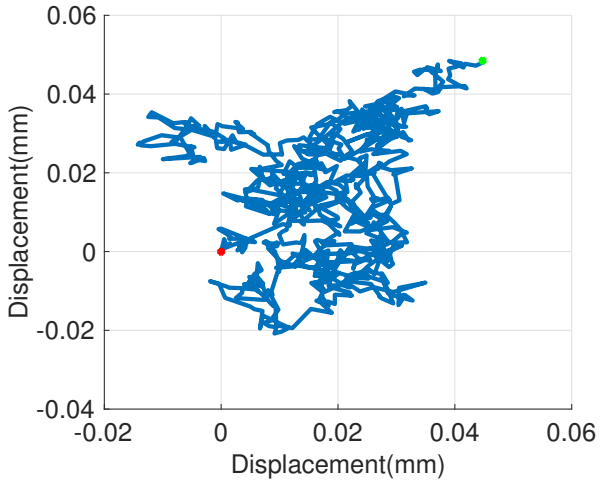


Figure 1.3: Brownian motion of a particle Simulated Brownian motion for a single particle with a diffusion coefficient of $0.003mm^2.s^{-1}$ and a diffusion time of 1 second. The red dot indicates t=0s and the green t=1s.

The movement of such a particle, illustrated in figure 1.3, was formalized by Einstein in 1905 [Einstein, 1905]. He used a probabilistic framework to describe the movement by introducing the notion of displacement distribution. In the case of free diffusion, the displacement distribution is a zero-mean Gaussian, where the width is defined by the diffusion coefficient.

With this displacement distribution, Einstein connected the Brownian movement, and Fick's first law [Fick, 1855]. The second moment of the distribution is given by:

$$< x^2 >= 2nD\Delta \tag{1.2}$$

where $< x^2 >$ is the mean squared displacement of a particle, Δ the time interval, n the number of dimensions, and D the diffusion coefficient from Equation (1.1)

When diffusion occurs in a non-constrained environment, for instance free water,

diffusion occurs in all directions with the same probability. Such diffusion process is said to have an isotropic profile. In brain tissues the diffusion is slower than in free water, because molecule movement is hindered by the presence of structures such as cell membranes. Diffusion in and around the axons is faster along the main orientation of the axons than across it, and it is in this case said to be anisotropic. Diffusion MRI exploits this key feature to estimate the diffusion profile and to indirectly infer white matter organisation and micro-structural properties.

1.2.2 General principles of MRI

Atomic nuclei with an odd number of neutrons or protons possess a magnetic moment characterising the strength and direction of a microscopic magnetic field. This magnetic moment arises from the spin, that is the angular momentum carried by the nucleus.

Nuclear magnetic resonance (NMR) describes the properties of the spins exposed to a magnetic field. When the atomic nuclei are not exposed to a magnetic field, their spins are randomly oriented. When exposed to a magnetic field, the spins precess, following a circular path perpendicular to the field. The frequency of the precession ω is proportional to the strength of the magnetic field B_0 as described by the Larmor equation:

$$\omega = \gamma B_0 \tag{1.3}$$

where γ is the gyromagnetic ratio.

NMR relies on the excitation and relaxation of the spins to infer the magnetic properties of their environment. It typically uses the resonance frequency of the hydrogen proton, by far the most abundant in the body. The protons are exposed to a constant magnetic field B_0. An electromagnetic excitation pulse tips the spins by

a certain angle. After the pulse, the spins progressively dephase due to local field inhomogeneities (T2 relaxation) and realign with B_0 (T1 relaxation). In a spin-echo experiment (illustrated in Figure 1.4), a 90° excitation pulse is followed (after a time τ) by a 180° refocusing pulse. The signal obtained after the same time τ reflects the phase coherence of the spins. If the protons have been exposed to some magnetic inhomogeneities, the phase coherence of their spins will be lower, and so will be the signal. In MRI, field gradients are employed to modify the phase coherence in a controlled way for signal localisation and hence image reconstruction.

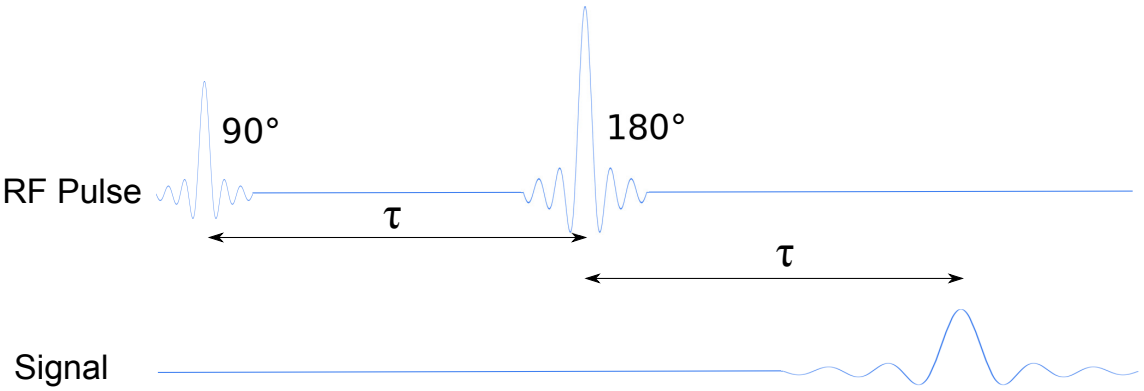


Figure 1.4: Spin-echo experiment The excitation and refocusing RF pulse are separated by a time τ. The signal acquired after the same time τ reflects the phase coherence. The time between the 90° excitation and the signal acquisition (2τ) is the echo time.

1.2.3 Diffusion MRI

The spin-echo NMR sequence was modified to measure diffusion with the use of additional gradients. In Figure 1.5, we show the pulsed-gradients spin-echo sequence introduced by Stejskal and Tanner in [Stejskal and Tanner, 1965]. This sequence, or variants of it, is the most commonly used in dMRI.

The pulsed gradients allow a relatively precise definition of the diffusion time Δ, i.e., the time between the two gradients, whereby incoherent molecular motion, such

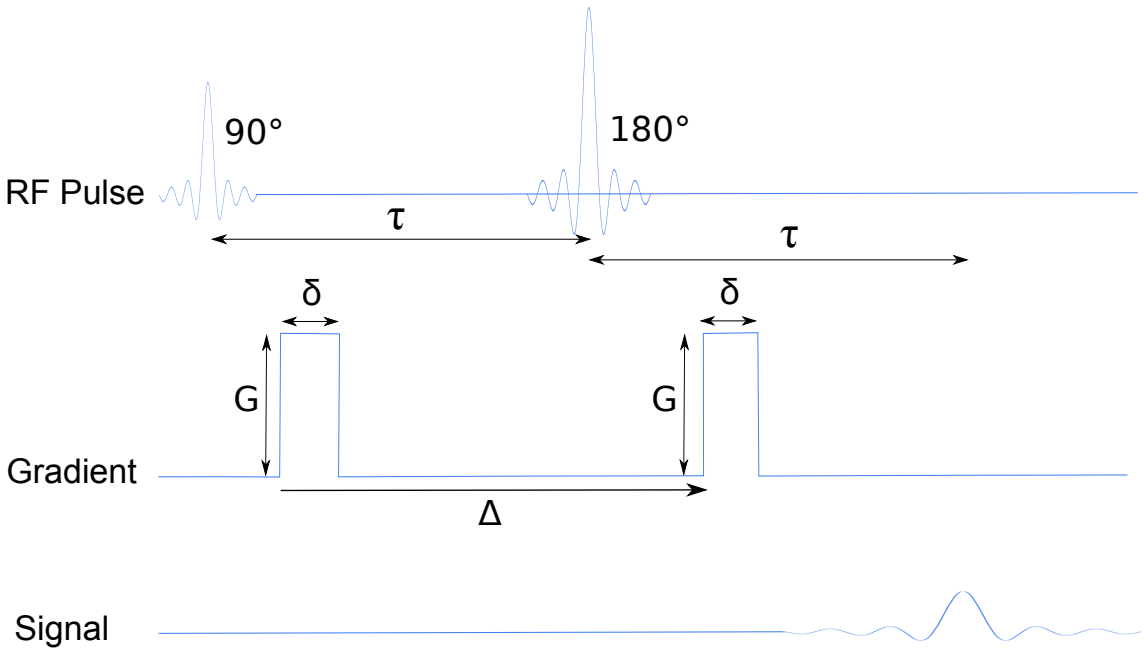


Figure 1.5: Pulse gradients spin-echo experiment The RF pulses and signal readout timing are similar to the spin-echo experiemnt descibed in Figure 1.4. Two diffusion encoding gradients of magnitude G and duration δ (encoding time) are added. The diffusion gradients are applied at a time interval Δ (diffusion time), which can vary from 10 ms to a few hundreds of milliseconds.

as diffusion, induces signal loss. The application of the gradients makes the signal sensitive to the protons displacements. If the protons moved along the diffusion gradient in the time interval Δ, then they will not recover their initial phase with the application of the second diffusion gradient, and the signal will be attenuated. The two gradients have the same amplitude G that does not vary over the short time δ (pulsed gradient). In [Stejskal and Tanner, 1965], the authors proposed to combine the gradient duration δ, strength G and interval Δ to form the b-value, given by:

$$b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3) \tag{1.4}$$

where γ is the gyro-magnetic ratio. b is often expressed in $s \times mm^{-2}$.

Rather than expressing the attenuated signal S itself, it can be convenient to manipulate the signal attenuation, that can be obtained by dividing the attenuated signals with the signal without diffusion gradients S_0: $A = \frac{S}{S_0}$). In the case where

the diffusion is Gaussian, the signal attenuation only depends on the b-value and the diffusion coefficient D:

$$A = exp(-bD) \tag{1.5}$$

Stejskal and Tanner experiments were limited to NMR , as the techniques for image reconstruction were not available at the time. The invention of echo-planar imaging (EPI) in 1977 allowed to localise the NMR signal and reconstruct a 3D image [Mansfield, 1977]. In 1985, [Le Bihan and Breton, 1985] successfully combined the localisation and diffusion gradients to form the first in-vivo diffusion MR images. In 1994, the diffusion tensor formalism was introduced in [Basser et al., 1994]. The tensor model allowed, for the first time, the quantification of diffusion anisotropy independently of the orientation of the imaged object (e.g. brain) relative to the scanner. The signal attenuation can be written as:

$$A = exp\left(-b\vec{g_i}^T D\vec{g_i}\right) \tag{1.6}$$

where $\vec{g_i}$ is the direction of the gradients, D is the diffusion tensor, and b is the b-value.

1.3 Processing and analysis of diffusion MRI data

In this section, we describe the main steps involved in the processing and analysis of diffusion data. These steps are usually performed at the single subject level.

1.3.1 Image artefacts and pre-processing

In-vivo diffusion data typically exhibit three main sources of image artefacts: susceptibility-induced distortions, eddy-currents induced distortions and artefacts caused by subject's movements.

Susceptibility-induced distortions Diffusion MRI requires the acquisition of several volumes, each volume representing the signal attenuation in a given gradient orientation. Data acquisition can be very time consuming and a fast acquisition is required. Echo-planar imaging is the most popular fast acquisition method. It allows acquisition of an entire slice with a single excitation pulse. The gain in acquisition speed however comes at the cost of reduced SNR and resolution. It also introduces image distortions along the phase encoding direction. This phenomenon is particularly noticeable at the interface between brain tissues and air cavities, where variations in the field due to magnetic susceptibility changes are more pronounced. In Figure 1.6, we show the effects of susceptibility-induced distortions on the b=0 images acquired with two opposite phase-encoding directions. The phenomenon is particularly visible in the anterior part of the brain and in the temporal lobes.

Susceptibility-induced distortion can be corrected using a fieldmap that is separately acquired along with the diffusion data. The fieldmap measures the field inhomogeneities and allows to accurately unwarp the EPI images [Jezzard and Balaban, 1995]. Alternative methods can be used to estimate such fieldmap when the latter was not acquired in the experiment [Andersson et al., 2003]. These methods compute the fieldmap that recovers the unwarped image using two images acquired with different phase-encoding directions, e.g. anterior-posterior and posterior-anterior. We can see the unwarped b=0 images as well as the estimated fieldmap in the Figure 1.6.

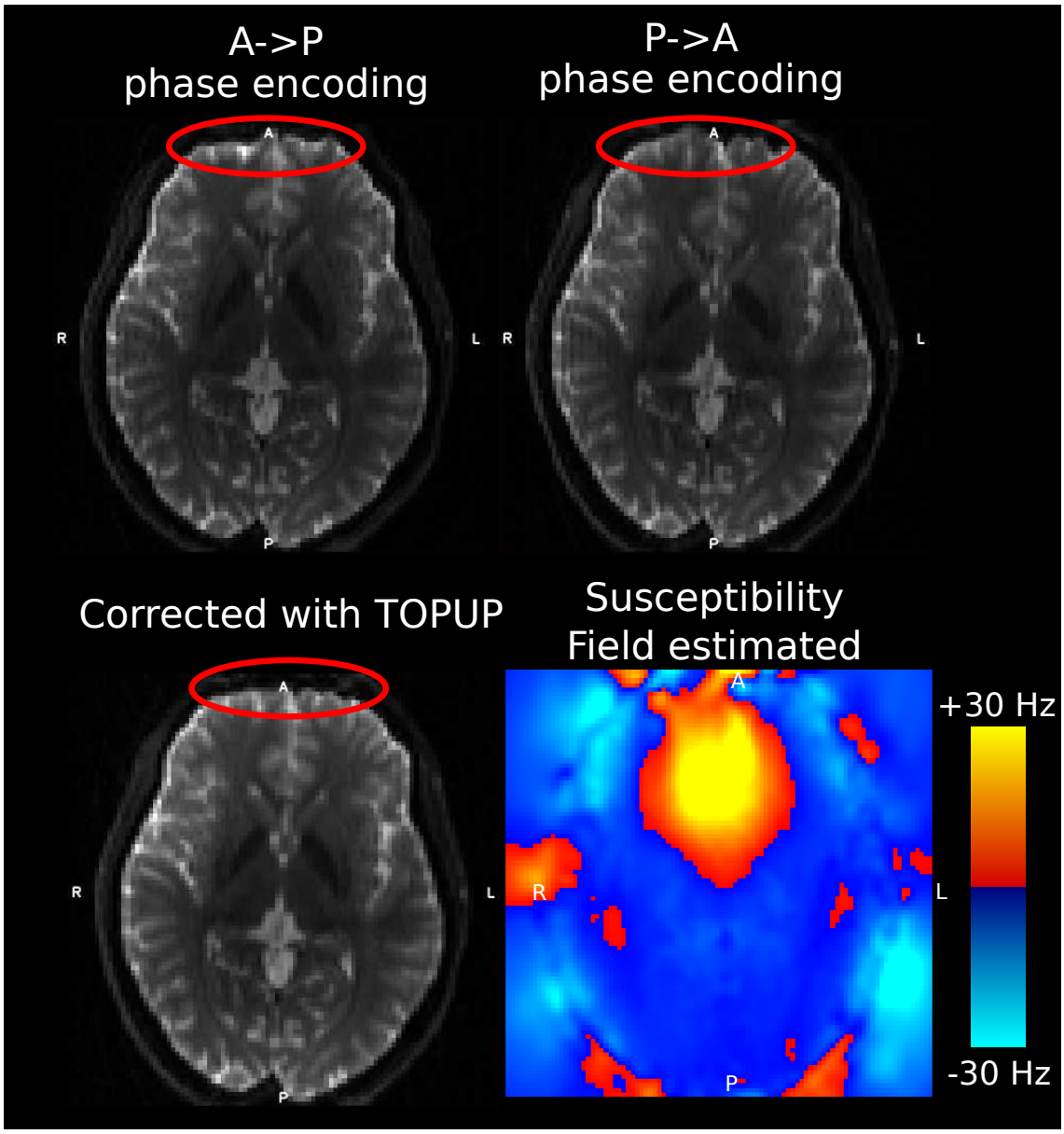


Figure 1.6: Susceptibility-induced distortions in EPI acquisition. Axial slice of a b=0 acquisition with susceptibility induced distortion (Top), and corrected image and estimated field map obtained with TOPUP.

Eddy-currents induced distortions The diffusion gradients are producing rapidly changing magnetic fields that induce eddy currents in the magnet bore. These eddy currents give rise to an off-resonance magnetic field that affects the signal localisation in a similar way that the susceptibility field does. An important difference with susceptibility-induced off-resonance field is that the eddy current induced field depends on the diffusion gradient orientation and the b-value [Jezzard et al., 1998].

In the Figure 1.7, we show simulated distortions resulting from a diffusion gradient in the x and y direction. We can observe that the diffusion gradient along x induces a shearing of the k-space trajectory, corresponding to a x translation depending on the y position. It results that the reconstructed image will present the opposite shearing: a translation along y depending on the x position. On the right of Figure 1.7, we can see how a diffusion gradient along y stretches the k-space trajectory along y and that the reconstructed image is squeezed along y. The nature and amplitude of these effects combine, depending on the direction and b-value respectively.

In order to correct for eddy-current distortions, registration based methods were proposed [Andersson and Skare, 2002]. They rely on the alignment of each diffusion-weighted volume to a reference b=0 image, where no eddy current distortions are present. These methods were recently superseded by a forward prediction approach that enables to predict the diffusion signal expected when no eddy current distortions are present [Andersson and Sotiropoulos, 2016].

Motion artefacts The acquisition of dMRI data takes between 5 minutes in a clinical context and an hour or more for advanced sequences including hundreds of gradient orientations and multiple b-values. With long acquisition times, the subject is likely to experience discomfort and the images will suffer from motion artefacts. The latter affect the estimation of the diffusion parameters and can ultimately bias the outcome of cross-subject studies, e.g. if the amount of movement is correlated with the variable of interest [Yendiki et al., 2014]. The nature of the artefacts depends on whether the movement occurs between the acquisition of two volumes (between volume), within the time of the volume acquisition, or within the diffusion encoding period (within slice).

Between volume motion refers to a rigid displacement of the head between the acquisition of two brain volumes. This type of motion can be adressed with registration approaches, similarly to eddy-current induced distortions [Andersson

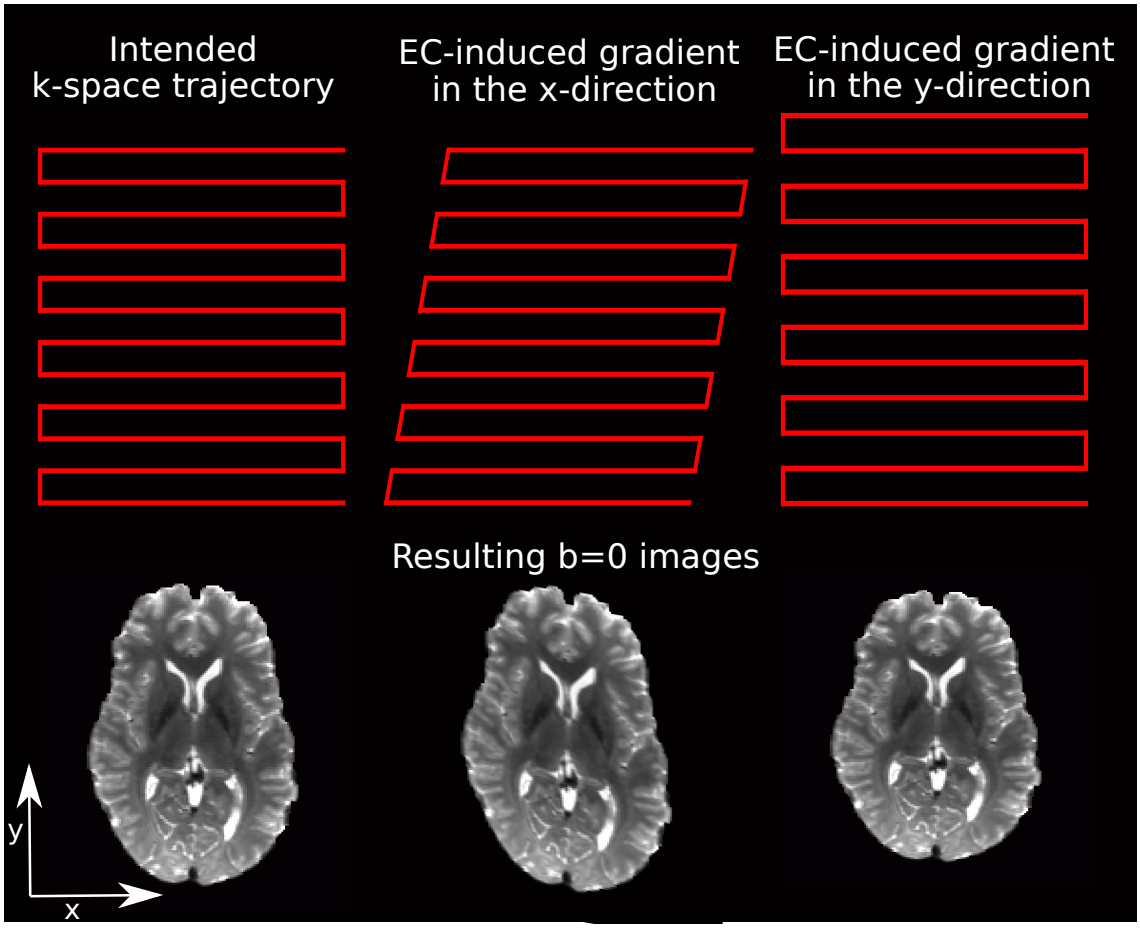


Figure 1.7: Illustration of the k-space trajectories and reconstructed b=0 images without and with Eddy currents (EC) induced distortions. On the left, we show the trajectory and image withou eddy-current induced gradient. In the middle, we show the shearing effect resulting from a eddy current induced gradient in the x direction. On the right, we show the scaling effect resulting from a eddy current induced gradient in the y direction.

and Skare, 2002]. Besides, it is important to account for the rotational movements, and correct the gradient orientation accordingly [Leemans and Jones, 2009].

Movements occurring during the acquisition will distort the reconstructed 3D image and some data might be lost. In Figure 1.8 we illustrate the effect of in-plane translation occurring within the volume (green ellipse). We can see that the slices are not aligned across the z direction (inferior-superior). The effect is particularly visible because the slices were acquired using simultaneous multi-slice acquisition [Barth et al., 2016], hence the "zig-zag" pattern. In the case of a translation along z, or a rotation that is not around z (e.g. the subject "nods"), some data will not be

acquired.

As dMRI is sensitive to microscopic movements of the water molecules, a macroscopic movement of the brain occurring between the two encoding gradients will result in a huge signal attenuation for the whole slice. This slice "dropout" is visible in Figure 1.8, where we can see black stripes on the coronal view. This effect can be detected and the data labelled as outlier in the modelling steps.

The Eddy method mentioned in the previous sub-section originally accounted only for between volume movements [Andersson and Sotiropoulos, 2016]. Using the same forward model for the diffusion signal, it was extended to account for within volume and within slice movements [Andersson et al., 2016] [Andersson et al., 2017]. The Eddy method is therefore a comprehensive framework accounting for the most prominent artefacts in dMRI.

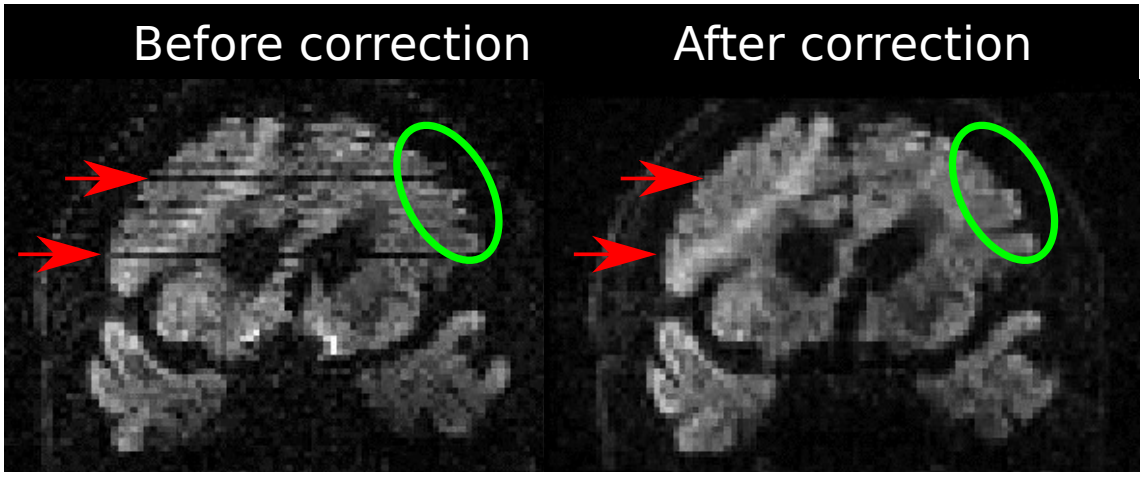


Figure 1.8: Within volume and within slice movement artefacts. Coronal slice illustrating the "zig-zag" effects of within-volume movements (green ellipse) and the signal dropout resulting from within-slice movements (red arrows). Courtesy of J.L.R. Andersson

1.3.2 Diffusion models

The multi-dimensional nature of the diffusion signal allows for a wide range of modelling approaches. The choice of the model will depend on the sequence parameters and the desired application for the acquisition.

The most simple approach consists of averaging the diffusivity measured along each direction to obtain the apparent diffusion coefficient. It is given by:

$$ADC = \frac{1}{n}\sum_{i}^{n} -b\ log(A_i) \tag{1.7}$$

which corresponds to Equation (1.5) averaged across the n diffusion gradients orientations. The ADC is generally used in clinical routine, e.g. to detect the localisation and extent of a stroke in the brain. It is however too simplistic especially if we are interested in the orientation and microstructure of the fibre bundles.

Diffusion Tensor The introduction of the diffusion tensor (DT) model in 1994 [Basser et al., 1994] has emerged as a major breakthrough in the study of brain micro-structure and connectivity. The DT is a 3×3 symmetric positive definite matrix that represents the 3D Gaussian displacement of the water molecules. It is written as:

$$\begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix} \tag{1.8}$$

where the diagonal elements are the diffusivities along the three axes of the scanner reference frame and the off-diagonal elements correspond to the covariance between the diffusivities along those axes.

The DT can be seen as a three-dimensional covariance matrix of displacements in a given time. The eigenvectors are defining the principal orientations of diffusion

and the eigenvalues are describing the diffusivity along and across the principal orientation. For a tensor D, we have:

$$D = (\vec{e_1}|\vec{e_2}|\vec{e_3}) \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} (\vec{e_1}|\vec{e_2}|\vec{e_3})^T \tag{1.9}$$

where λ_i are the eigenvalues of the tensor and $\vec{e_i}$ the eigenvectors

Looking back at the signal attenuation given by Equation (1.6), we can see that the system has 7 unknowns: S_0 and the 6 independent elements of the DT. It follows that solving the system uniquely requires at least 7 measurements. With only one measurement per unknown, the DT estimation would be very sensitive to the noise in the dMRI signal. Therefore, it is recommended to acquire at least 20 diffusion gradient orientations and an additional volume without diffusion encoding (S_0) [Jones, 2004].

Equation (1.6) can be linearised by taking its logarithm and the DT elements can be estimated by ordinary least squares approach or weighted least squares [Salvador et al., 2005]. However, these methods do not guarantee that the resulting tensor is positive definite, and they also rely on the erroneous assumption that the noise in dMRI is Gaussian. More robust or realistic DT estimation can be obtained, e.g., using outliers rejection techniques [Chang et al., 2005] or non-linear optimisation [Andersson, 2008].

The DT provides useful information on the microstructural environment of the fibres. The mean diffusivity (MD) is the average of the three eigenvalues of the tensor and it is therefore related to the ADC. It is commonly reported as a measure of the membrane density and microstructural organisation. The MD is sensitive to oedema, stroke, and cellularity [Sun et al., 2006].

The fractional anisotropy (FA) is a summary measure representing the degree

of anisotropy of the diffusion process. It is written as:

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{1.10}$$

FA is widely used as a biomarker reflecting fibre integrity. A decrease in FA is often associated with a reduced fibre integrity, but also fibre orientation dispersion. However, the FA is not specific to the type of change in the WM (e.g. axonal damage, demyelination, cellularity), and it is sensitive to the geometry of the underlying fibre patterns.

The axial diffusivity (AD) relates to the diffusivity along the principal DT orientation. It is sensitive to axonal damage and WM maturation [Song et al., 2003] [Ashtari et al., 2007].

The radial diffusivity (RD) is the average of the two eigenvalues associated to the eigenvectors perpendicular to the main DT orientation. It is usually considered as negatively correlated with myelin sheath integrity [Song et al., 2003].

While the diffusion tensor and its aforementioned indices prove to be very useful in typical group studies, they carry several limitations and approximations. The DT indices are typically biased by the presence of other tissues (e.g. GM or CSF) [Alexander et al., 2001a], or crossing fibres [Douaud et al., 2011]. These confounding effects can potentially mislead their interpretation with respect to a specific biophysical change.

Compartments models The DT model assumes a single diffusion process with a Gaussian profile, and therefore a monoexponential decay of the diffusion signal. While this assumption remains reasonable at average b-values ($\leq 1500 s.mm^{-2}$), it does not hold for higher b-values. It was long established that the presence of multiple compartments and cell membranes in tissues leads to a non-monoexponential signal attenuation [Kärger, 1985]. Using high b-values (up to 4000 $s.mm^{-2}$)), it was showed

that the signal decay in the human brain can be decomposed into fast and slow Gaussian diffusion [Clark and Le Bihan, 2000]. It was hypothesized that the fast diffusion mostly reflects the extra-axonal space (EAS) and that the slow diffusion can be mainly related to the intra-axonal space (IAS). Further experiments with higher b-values (up to 22 000 $s.mm^{-2}$) connected the slow component with non-Gaussian restricted diffusion within axons [Assaf and Cohen, 2000].

While the biophysical source of non-monoexponential decay remains unclear [Mulkern et al., 2009], multi-compartment models that are more representative of WM anatomy were developped. The composite hindered and restricted model of diffusion (CHARMED) [Assaf et al., 2004] [Assaf and Basser, 2005] proposed to represent EAS with a diffusion tensor and the IAS by cylinders with a fixed diameter. The volume fraction of the IAS is a quantitative and physiologically interpretable parameter representing the axonal density. The AxCaliber model extended the CHARMED model further by including another meaningful parameter, the axon diameter distribution [Assaf et al., 2008]. An important limitation of the method is that the axonal fibres orientation has to be known a priori as it must be perpendicular to the diffusion gradient direction. The ActiveAx method alleviated this limitation by modelling solely the mean of the axon diameter distribution [Alexander et al., 2010]. It was later extended to include the dispersion of the fibres, which increased the precision of the axon diameter estimation [Zhang et al., 2011]. However, these methods rely on diffusion sequences that were specifically designed for the experiments. They require a high number of gradient directions, and several b-values and diffusion times independently varying. This results in a very long acquisition time, difficult to apply in the context of a group study.

These models were therefore simplified in order to be applicable in the clinical environment. The NODDI model allows to retrieve specific microstructural parameters, such as axon density and dispersion in a clinically feasible time (less than 20 minutes) with commercially available dMRI sequences (two non-zero b-values)

[Zhang et al., 2012]. It represents the IAS with sticks following an Watson orientation distribution function (ODF). The EAS is modelled by a an axially symmetrical tensor (zeppelin) with a fixed axial diffusivity. An additional isotropic compartment with fixed diffusivity accounts for partial volume effects with the CSF. The NODDI model was later extended to represent the dispersion with a Bingham distribution [Tariq et al., 2016], similarly to the work in [Kaden et al., 2007] and in [Sotiropoulos et al., 2012].

Multiple fibres models The main limitation of the aforementioned models is their inability to represent the crossing fibres present in most of the WM voxels [Jeurissen et al., 2013]. This limitation confounds the parameters of the simple DT or the single fibre multi-compartment models mentioned in the previous section.

The multi-tensor model is a straightforward extension of the simple DT. It assumes the presence of n orientationally distinct fibre populations and models each population with a DT. The diffusion signal attenuation is expressed as a weighted sum of Equation (1.6):

$$S_i = S_0 \sum_{k=1}^{n} f_k e^{-b\vec{g}_i^T D \vec{g}_i} \tag{1.11}$$

where f_k is the signal fraction for the k^{th} fibre population. The multi-tensor model however suffers from a large number of parameters (13 parameters for 2 tensors, 20 parameters for 3 tensors) and is over-parametrised in most situations. It also poses a model selection problem, as the number of fibre populations present in a voxel is a-priori unknown.

The Ball and Stick model, originally designed to characterize only one fibre [Behrens et al., 2003], naturally extends to multiple fibres [Behrens et al., 2007]. Its lower number of parameters permits a robust estimation on most dMRI acquisition protocol. A Bayesian approach allows for the use of an automatic relevance determination (ARD) prior [Mackay, 1995]. It addresses the model selection problem by discarding fibre compartments that are not needed to explain the data (in this

case, the ARD prior pushes the signal fraction f_k to zero). The Bayesian framework estimates the posterior distribution of the model parameters which represent their uncertainty. The uncertainty of the discrete fibre orientation estimates is referred to as the uncertainty orientation distribution function (uODF).

The model-based approaches described so far inherently make assumptions on the nature of the diffusion profile (multi-Gaussian) and on the number of fibre bundles. [Tuch, 2004] proposed a non-parametric approach to map the diffusion signal directly to a diffusion orientation distribution function (dODF), using a tomographic transformation (Funk-Radon transform). The resulting dODF is represented on the sphere and it describes the probability for the water molecules to move in a given direction. The main limitation of the technique is that the dODF is a blurred function, which makes it difficult to disentangle different fibre bundles crossing at a low angle (<60 °) [Descoteaux et al., 2009].

The diffusion signal can be considered as a convolution of the underlying fibre orientation distribution (fODF) with the signal attenuation associated with a single, straight fibre population (the fibre response function - FRF). It follows that deconvolving the signal with the FRF should theoretically give the true fODF. In [Tournier et al., 2004], the FRF is computed by averaging the signal in the most anisotropic voxels (e.g. in the corpus callosum), and the diffusion signal is represented by a set of basis functions (spherical harmonics). The method became very popular and it was later improved to address some of the initial limitations, such as spurious fODF peaks [Tournier et al., 2007], tissue-type partial voluming [Jeurissen et al., 2014], or FRF estimation [Tax et al., 2014].

Multiple fibre models prove to be very useful when one is interested in the fibre orientations. However, they provide limited microstructural information, generally limited to the signal fraction of each fibre population [Behrens et al., 2007] [Raffelt et al., 2012] [Dell'Acqua et al., 2013].

1.3.3 Tractography

The DT, fODF, and uODF are extensively used to trace the fibre bundles with tractography. The latter allows us to reconstruct the WM pathways by integrating the voxel-wise fibre orientations.

Deterministic tractography consists in tracing streamlines across a vector field representing the principal directions of diffusion. Each streamline is propagated step by step, starting from a seed voxel and following the local orientation information [Mori et al., 1999]. The latter being derived from the dMRI discrete data, it is beneficial to perform a smooth interpolation between the grid points because the true underlying fibre orientation vary smoothly across voxels [Basser et al., 2000]. The local orientations can be represented by, e.g., the first eigenvector of the DT, or the fODF peaks. The propagation is stopped when some criteria are met, e.g. the FA of the DT is too low, the streamline curvature is too high, or the stop mask is reached.

The main limitation of tractography is that the errors due to imaging noise and inadequate modelling of the fibre orientation (e.g. DT cannot represent crossing fibres) are accumulated during the streamline propagation process.

Probabilistic methods allow to propagate the uncertainty associated with the measurements and modelling errors [Jones, 2003]. It allows to track fibres through regions of high uncertainty, where deterministic methods would terminate the streamline [Behrens et al., 2003]. As such, they provide an estimation of the probability of the path of least hindrance to diffusion [Parker et al., 2003]. These methods are following the same principles as deterministic tractography, except that the direction to follow is sampled at each step from a uODF representing the uncertainty associated with the noise and modelling errors.

1.4 Cross-subjects comparisons

Cross-subjects local comparison involves a few extra steps in the dMRI processing pipeline. Here, we focus on the methods performing voxel-wise comparisons across subjects. These methods, referred to as voxel-based morphometry (VBM), were originally designed to study volumetric changes in structural T1w images. VBM aims to identify macro-structural morphological changes (e.g. grey matter thinning) occurring in a population [Ashburner and Friston, 2000]. It naturally extended to the analysis of images derived from diffusion MRI, in an attempt to understand changes occurring at the microstructural level. This type of analysis, referred to as voxel-based analysis (VBA), requires processing steps similar to VBM, namely spatial normalisation and statistical inference.

1.4.1 Spatial normalisation

The spatial normalisation (or registration) step aims to match brain structures across space. This step is needed in order to obtain a voxel-wise correspondence of the WM structures across subjects.

Registration allows to transform one or several images into a reference coordinate system. In the case of MRI voxel-wise cross-subject studies, all the subject images have to be aligned with a reference image in a common space (average subject image). Ideally, each common space voxel should represent the same anatomical location across all the registered subjects. One can either use a pre-existing reference image in a common space (e.g. MNI template [Collins et al., 1994]), or build a study-specific template from the subjects [Guimond et al., 2000].

To find the transformation that best aligns two images, two steps are typically

performed. First, a linear registration computes a global matrix describing the affine transformation (i.e., scaling, rotation, translation and shearing) between the subject image and the reference image. Second, a non-linear registration finds a warp field that describes the local transformations between the subject and reference image. Non-linear registration is more challenging because it is a compromise between the similarity between images and the anatomical plausibility of the transformation.

In the case of diffusion MRI, the similarity metric between two images is generally based the FA or MD maps. We will see that another approach is to take advantage of the multi-dimensional aspect of the dMRI signal to potentially improve registration accuracy.

The specificity of spatial normalisation step is crucial in group studies. If the subjects are misaligned, a voxel in common space might represent different structures across subjects and therefore mislead the interpretation of the results. Additionally, mis-registration can impact the sensitivity of a study. For example, a pathological process might affect a very localised part of the brain. If this area is not accurately matched across subjects, it will result in a loss of statistical power, as the effect will not be seen in the mis-aligned subjects.

A typical approach to overcome the residual mis-registration is to apply a Gaussian smoothing to the registered map. This also comes with pitfalls, e.g., the choice of the smoothing kernel extent affects the spatial extent of the findings [Jones et al., 2005]. The tract-based spatial statistics approach (TBSS) was specifically designed for dMRI VBA [Smith et al., 2006]. It proposes to address the residual mis-registration with an additional projection of individuals onto a reference skeleton representing the center of the tract, thus avoiding smoothing of the data altogether.

1.4.2 Statistical inference

Once the WM voxels have been matched across subjects, we can compare the microstructural parameters derived from a given diffusion model. The inference is typically performed across subjects given some discrete or continuous variable, e.g., presence of pathology, age, IQ). In the most simple VBA, the inference is performed across the whole extent of the white matter. It was proposed in [Smith et al., 2006] to restrict the analysis to the center of the white matter, defined by a skeleton. Using tractography and deformable models, [Yushkevich et al., 2008] proposed to distinguish between WM structures and to perform the inference on a summary of the diffusion indices projected onto a skeleton.

The data being matched across subjects in a given space (e.g., WM segmentation or skeleton) can be considered as a 4D image where the 4th dimension is the subject number. A simple approach to perform voxel-wise statistics is to apply the general linear model (GLM), e.g., to test for significant differences in FA between control and patient groups in each voxel. The test produces a raw-statistic (e.g., z-statistic) image for each hypothesis tested. To decide whether the null-hypothesis is true or not, the z-statistic map is thresholded to generate a p-value. In the case of neuroimaging data, cluster-based thresholding have more sensitivity than voxel-based thresholding because the signal (e.g. reduced FA) is likely to exhibit spatial correlation. The threshold-free cluster enhancement (TFCE) is an elegant method to perform cluster-based inference without the need to choose an initial threshold to form the clusters [Smith and Nichols, 2009]. However, the p-values can be accurately generated only if the distribution of z-statistic image under the null-hypothesis is Gaussian. Unfortunately, it is rarely the case in dMRI data analysis. A solution to this problem is to perform permutation testing, a technique that enables to compute the null-distribution by randomly shuffling the order of the subjects [Winkler et al.,

2014].

Another issue is that the statistical tests are repeated across all the considered voxels, so that the usual threshold on the p-value (e.g. $p \leq 0.05$) will invariably produce many false positives findings [Bennett et al., 2009]. This multiple comparisons problem is commonly addressed by controlling the family wise error rate, that is the probability of finding one or more false positive. The Bonferroni correction controls the FWE by multiplying the p-values by the number of tests. This approach is however too conservative in the case of neuroimaging data, where the signal exhibits spatially correlation. Using permutation testing allows to control the FWE by building the null distribution of the maximum test statistic across space and test the actual statistics against it. This approach is very well suited to dMRI group studies as it is robust, accurate and sensitive.

1.5 Thesis outline

The aim of the work presented in this thesis is to identify and potentially address the limitations and pitfalls in voxel-wise comparison of diffusion parameters across subjects. The reference point of our work is TBSS, which is one of the most widely used methods in dMRI group studies [Smith et al., 2006]. We identify the spatial normalisation, tissue-type partial volume effects, and crossing fibres as sources of noise and confound in TBSS analyses. Through the four core chapters of this thesis, we evaluated how the methods involved in TBSS could be improved in order to increase the sensitivity and specificity in cross-subject comparisons.

In Chapter 2, we evaluate four registration algorithms in terms of raw performance, anatomical specificity across subjects, and sensitivity in a typical group study. Our experiments notably show that the methods using the full tensor (as opposed to FA or DWIs) in the similarity metric exhibits the best performance. Importantly,

we show that using improved registration methods alleviates the need for TBSS projection step, improves tract correspondence across subjects and leads to increased sensitivity.

In Chapter 3, we aim to address partial volume effects between white matter and CSF. Using a Bayesian setting with a prior on the white matter compartment, we are able to explicitly model the CSF contribution to the diffusion signal with a two-compartment model (FW model). Our method favourably compares with previous work in [Pasternak et al., 2009] and is directly applicable to standard diffusion data, enabling us to revisit a typical control versus patient study. We find that that most of the differences observed with the simple DTI model are due to macroscopic effects, i.e., tissue-type partial voluming. Importantly, we suggest that the FW model is not appropriate to disentangle the intra an extra-axonal micro-structural changes.

In Chapter 4, we propose a supervised machine learning approach to predict the T1w-derived tissue partial volume fractions (PVF) from the diffusion data. The motivation is the within subject mis-registration between modalities (structural T1 and diffusion), that we alleviate by learning the non-linear relationship between the PVF estimated on the structural T1w image and some diffusion features (e.g. FA, MD, S0). We explore various machine learning methods such as Gaussian processes, support vector machines, random forests and neural networks. Finally, we demonstrate that using the predicted PVF to inform the fitting of the FW model enables to reduce the bias in the tensor estimation and to avoid the presence of spurious FWF in deep white matter.

In Chapter 5, we design a framework that allows to retrieve voxel-wise tract-specific microstructural parameters. Our approach is motivated by the confounding effects of foreign fibre populations or tissues, and by the difficulty to relate a single fibre bundle to local diffusion parameters. We propose to model a each fibre or tissue population by an axially symmetric tensor. As the fitting of such model is typically ill-posed, we use a spatial regularisation scheme to constrain some of

the parameters. The method relies on tractography to identify the individual fibre bundles present in a voxel. It allows to perform a spatial regularisation in a sensitive and anatomically meaningful manner. We validate our approach through simulations and in-vivo experiments. Additionally, we demonstrate the relevance of our approach in a cross-subject study, where we look at the correlations between tract-specific micro-structural parameters and behavioural measurements.

2

Evaluation of improved cross-subject registration

Diffusion MRI group studies often involve voxel-wise comparison of brain structures, which require registration of the subjects into a common space. The popular TBSS approach combines registration with a projection onto a group skeleton to reduce the registration error. In this chapter, we quantify how improved registration methods can outperform the standard TBSS approach. Most importantly, we show that the projection step decreases the accuracy and specificity of the registration. Finally, we demonstrate that using a better registration method increases the sensitivity to WM changes in a typical group study.

2.1 Introduction

Registration is a key step when performing MRI group studies as it enables accurate matching and comparisons between subjects. It ensures that the structure represented at a given location is the same across subjects, which is essential for the sensitivity and specificity of the analysis. Registration aims to maximise the

similarity between two images. Consequently, the similarity metric and the data being compared plays an important role in the performance of the method. In diffusion MRI, the multidimensional nature of the images allows to base the similarity metric on several quantities such as the full diffusion weighted signal, the diffusion tensor, or scalar quantities derived from the latter (e.g., FA).

Most dMRI registration algorithms use scalar images (FA) to feed into the similarity metric. It allows the use of common registration methods developed for other modalities (e.g. T1 weighted images). For instance, FNIRT [Andersson et al., 2007] and ANTS [Avants et al., 2008] are registration methods developed for structural and functional images, but they can easily be adapted to diffusion data and futher optimised [de Groot et al., 2013]. The limitation of these approaches is that they only use a summary of the data, potentially missing interesting information. For example, the orientation information could help to distinguish adjacent tract with comparable FA but different orientation. That creates an interesting contrast within the white matter that does not exists in the scalar images.

Hence, it was proposed to base the registration similarity metric on the diffusion tensor images (DTI). A preliminary work showed that using the full tensor information leads to better registration [Alexander and Gee, 2000] [Ruiz-Alzola et al., 2002]. However, these methods were not reorienting the tensors, thus bringing innacuracies in the matching of the images and introducing inconsistencies in the orientation field [Alexander et al., 2001b]. It was later showed that addressing this issue leads to a better matching of the DT images, but the reorientation was not included in the optimisation scheme [Park et al., 2003]. The DTITK method proposed to use the full tensor information in the similarity metric, and also to explicitly include its reorientation in the optimisation [Zhang et al., 2006]. This algorithm ranked first in a quantitative evaluation based on a dMRI phantom of 10 registration methods, including FNIRT [Fillard et al., 2011]. Nevertheless, tensor-based registration methods rely on a model that has a limited ability to represent complex fibre

configurations.

It was proposed to use the full diffusion data, without a model, in the similarity metric, and to perform the reorientation with spherical interpolation [Tao and Miller, 2006]. The initial method was limited to linear registration, but later extended to non-linear registration [Duarte-Carvajalino et al., 2013].

In this chapter, we assess the performance level achieved by registration algorithms working with different dimensionality of the data, i.e. scalar FNIRT [Andersson et al., 2007], tensor DTITK [Zhang et al., 2007] and DWI FNIRT4D [Duarte-Carvajalino et al., 2013]. We also included an optimised version of FNIRT (oFNIRT) that estimates a non-linear transformation with increasing degrees of freedom (DoF) [de Groot et al., 2013]. The evaluation of the registration performance is not a trivial task, as it is a compromise between the similarity of two images and the plausibility of the computed warp. In the context of parameters optimisation, a registration evaluation framework based on tractography was introduced in [de Groot et al., 2013]. The latter has the advantage of being conditionally independent of the data used by the registration algorithm, thus reducing circularity in the evaluation. We make use of this performance metric throughout our evaluation which includes three datasets representing the range of image qualities that can be encountered in group studies, thus relating data quality and registration performance. Additionally, we use another performance metric that relates to the anatomical specificity of the registration method: the tract overlap across subjects.

We are also interested in the impact of using a study specific template where the registration target is iteratively built up by averaging the subjects in a common space [Guimond et al., 2000]. Simulation experiments in the context of atrophy and Alzheimer's disease suggested that using such a study specific template can improve the specificity and sensitivity in a clinical study [Shen et al., 2007] [Keihaninejad et al., 2012]. In this work, we quantify the gain in performance achieved with a study-specific template and also the impact of building the template with the tensors

instead of the FA.

In the second part of our work, we assess the impact of the TBSS projection step on registration performance and anatomical specificity.

A widely used approach to perform voxel-wise statistics among subjects is Tract-Based Spatial Statistics (TBSS) [Smith et al., 2006]. The pipeline is composed of two main steps: spatial normalisation with FNIRT and projection of the individuals onto a group mean FA skeleton in order to reduce the remaining alignment error. Recent publications evaluated to which extent the projection step is able to correct residual misalignment [Zalesky, 2011] and how it can be replaced by an improved registration algorithm [de Groot et al., 2013] [Schwarz et al., 2014] [Bach et al., 2014]. The limitations and compromises of the TBSS projection step are numerous. First, a voxel belonging to a given anatomical structure can be projected on a skeleton point located in another structure, which confuses the interpretation of the displayed result that can be driven by differences occurring in different tracts across subjects [Zalesky, 2011] [de Groot et al., 2013][Bach et al., 2014]. Furthermore, the projection step aims to correct misregistration only in the direction orthogonal to the skeleton, thus minimising the FA variance between subjects. This may not be the optimal manner to correct for the registration error, that can occur in any direction [Zalesky, 2011]. As a result, the TBSS projection might correct for the registration error in FA, but it is no guaranteed to improve the anatomical alignment across subjects.

Several recent publications have considered the replacement of the TBSS projection step by enhanced registration methods. In [de Groot et al., 2013], an optimised version of FNIRT with a high dimensional warp was proposed. The results showed that using the optimised version of FNIRT yielded better registration than a standard registration combined with the TBSS projection step. Another study compared the standard TBSS pipeline with a modified version that used ANTS registration without a projection step [Avants et al., 2008] [Schwarz et al., 2014]. They concluded that the projection step actually reduces the registration accuracy. Another concern

about the TBSS pipeline is the anatomical accuracy of the projection step. It was indeed demonstrated that using improved registration (DTITK) without a projection step gives the most anatomically correct results and that the TBSS projection step leads to anatomically incorrect projections between adjacent WM tracts [Bach et al., 2014].

We hypothesise here that an improved registration step would outperform the TBSS standard pipeline, thus avoiding the trade-offs of the projection step. We evaluate the accuracy and specificity of the projection step for the different registration methods with the two performance metrics. Finally, we introduce the improved registration algorithms in a typical clinical study, omitting the projection step and compare the results in terms of FA difference in a typical group study on Alzheimer's disease.

2.2 Materials and methods

2.2.1 Registration approaches

A registration problem consists in finding an optimal warp that maps one image onto another. The cost function defining the optimality is composed of a similarity metric and a regularisation term than ensures realistic deformation. The resulting optimisation problem can be written as:

$$argmin \sum_{i=1}^{N} (C(W_i(I), R_i) + \lambda O(W_i)) \tag{2.1}$$

where I and R and respectively the input and reference image containing N voxels, i is the i^{th} voxel, W the warp, C the cost function, λ the regularisation weight and O the

regularisation function. Of the algorithms we are comparing, FNIRT and FNIRT4D have a similar cost function (SSD), regularisation and optimisation scheme, whereas the latter are all different in DTITK.

2.2.1.1 FNIRT

FNIRT is a non-linear registration algorithm that relies on the weighted sum of squared differences (SSD) in the FA for the cost function. The warp field is represented with a set of basis functions (B-splines) and it is regularised using membrane energy [Andersson et al., 2007]. The optimisation relies on a multi-resolution scheme, from coarse to fine resolution.

De Groot et al. computed the optimal set of parameters for registering FA images with FNIRT [de Groot et al., 2013]. The optimisation was conducted on a dataset of average quality (resolution of 2mm isotropic and 60 gradient directions with a single b-values of $1000 s.mm^{-2}$). The optimal registration scheme they obtained comprises three steps, from medium to high degrees of freedom (DoF) for the calculated warp and with decreasing subsampling of the images. In the final step, the number of DoF remains the same but the warp regularisation is relaxed. We refer to FNIRT with optimal parameters as optimised FNIRT (oFNIRT). The set of parameters used for both FNIRT and oFNIRT are presented in Table 2.1.

Algorithm	**FNIRT**	**oFNIRT**$_1$	**oFNIRT**$_2$	**oFNIRT**$_3$
Subsampling scheme	8,4,2,2	8,4,2,2	1	1
Regularization	300,75,30,30	300,75,50,40	100	30
Warp field resolution (mm)	10	10	2	2
Smoothing FWHM	12,6,2,2	12,6,2,2	1	1
DOF	14 156	14 156	1 769 472	1 769 472

Table 2.1: Parameters for FNIRT original and optimal (oFNIRT) settings. The oFNIRT approach is comprised of 3 succesive steps, indicated in by i in oFNIRT$_i$

2.2.1.2 DTITK

DTITK is a piecewise-affine registration approach that computes the optimal warp from the tensor images [Zhang et al., 2007]. The similarity metric takes into account all the diffusion tensor information, including size, shape and orientation. It is based on the sum of squared differences of all the elements of the tensor via the ADC profile. The method relies on a similarity metric, called dyadic distance, that focuses on anisotropic components of the ADC profile:

$$||D_1 - D_2||_D = \sqrt{\frac{8\pi}{15}(||D_1 - D_2||_C^2) - \frac{1}{3}Tr^2(D_1 - D_2)} \tag{2.2}$$

where D_1 and D_2 are the two tensors to be compared, $||D_1 - D_2||_C^2 = \sqrt{Tr\left((D1 - D2)^2\right)}$ is the Cartesian distance between two tensors, and Tr is the trace of the tensor.

As the tensor elements are orientation dependent, reorientation needs to be accounted for while interpolating the images. The tensor reorientation is performed using a Finite Strain (FS) strategy, which is faster than the preservation of the principal directions (PPD) [Alexander et al., 2001b], and still a good approximation when the components of the Jacobian are small. Importantly, the search for optimal tensor orientation is part of the global optimisation.

In DTITK, the non-linear part of registration is composed of locally affine transformations, where regions of the images are affinely matched to the template with a smoothness constraint at region interfaces. The regularisation part of the cost function therefore expresses the discontinuities between two neighbouring regions.

The reference image is hierarchically divided into equal size regions with a number of region gradually increasing from 4 to 32. An affine transformation having 12 parameters, the finest level of registration computes a warp with $12 \times 32^3 = 393216$ DoF. One can note that the representation of the warp differs between DTITK (locally

affine) and FNIRT (basis functions). Although the two representations are in essence very close, it might still be a confound when comparing the two methods. The advantage of the basis function representation is that it provides a natural coarse to fine decomposition of the warp, which ease the multi-resolution optimisation scheme used in FNIRT [Sotiras et al., 2013]. On the other hand, the locally affine approach benefit from invertibility of the warp and computational efficiency, but the smoothness at the region interface is not guaranteed and relies heavily on regularisation.

2.2.1.3 FNIRT4D

Duarte-Carvajalino et al. proposed to use all the information contained in the DWI volumes [Duarte-Carvajalino et al., 2013]. We refer to this algorithm as FNIRT4D. Unlike FNIRT and DTITK the comparison between two images does not rely on a parametric diffusion model, and therefore does not make strong assumptions on how the signal is generated. This might be an advantage, notably when the tensor fails to represent complex fibres geometry, e.g. fibres crossing or fanning. The optimisation scheme is the same as in FNIRT, with the difference that every single DWI volume is included in the similarity metric. The key difference with FNIRT is the interpolation for the rotational component of the warp. When registering orientation independent scalar images, such as FA or MD, there is no need to account for the rotational component of the transformation. As discussed in the previous paragraph, registering tensor maps or DWIs implies a reorientation strategy. They chose to implement the finite strain strategy, as in DTITK. The difference here is the need to interpolate the full diffusion signal for each b-value instead of tensors. They proposed to use angular interpolation that transforms the original gradient direction according to the local rotational component of the transformation. The

intensity of a given DWI volume at a point r is defined by:

$$I(r,k) = \frac{\left(\sum_{i=1}^{N} w_i I(r,g_i)\right)}{\left(\sum_{i=1}^{N} w_i\right)} \tag{2.3}$$

where I is the image to interpolate, r is the voxel considered, k is the interpolated gradient direction, g_i is the original gradient and $w_i = exp(\frac{-d^2}{2\sigma^2})$, with $d = cos^{-1}|k.g_i|$ being the geodesic distance on the sphere between the original and interpolated gradient. Intuitively, we can summarise angular interpolation as a resampling of the original DWI with reoriented gradient directions. The interpolated image I' of the original image I given a transformation F represented by a 3×3 rotation matrix M and a translation t can be written as:

$$I'(r,g) = I\left(M^{-1}(r-t), \frac{M^{-1}g}{||M^{-1}g||}\right) \tag{2.4}$$

The method approximates the warp field from FNIRT with locally affine transforms, i.e. a translational (t) and rotational (M) component at each point of the warp field. It therefore minimises the following objective function between two images I and R:

$$argmin \sum_{i=1}^{K} \int_r \left[R(r, g_i^R) - I\left(M^{-1}(r-t), \frac{M^{-1}g_i^R}{||M^{-1}g_i^R||}\right)\right] \tag{2.5}$$

where i is the i^{th} gradient of the image R with K gradient orientation. We can see that the two images can have a different number of gradients because the test image I is resampled with the same gradient orientations as the reference image. It allows to register DWI acquired with different protocols, given that the two images have been normalised and corrected for bias field. An important difference with DTITK is that the reorientation of the DWI is not included in the optimisation scheme. It means that the method does not look for the optimal rotation of the underlying fibre, but only interpolates the rotational component of the already computed warp.

2.2.2 Datasets description and preprocessing

Clinical studies can rely on very different data acquisition protocols, involving various spatial and angular resolution or SNR. We decided to use three datasets representing low, medium and high quality, which we name respectively GSK, Whitehall, and HCP. The main acquisition parameters of the datasets are given in Table 2.2. For each dataset, we randomly selected 40 subjects. The impact of the registration on a group wise comparison was assessed on the GSK dataset which includes 60 patients diagnosed with Alzheimer's disease and 54 control subjects [Douaud et al., 2011]. This dataset corresponds to the lowest quality range, as described in Table 2.2.

Name (quality)	**Resolution** (mm)	**Directions**	**b-value** ($s.mm^{-2}$)
GSK (low)	2.5	2*30	900
Whitehall (medium)	2	128	1500
HCP (high)	1.25	90/270(*)	1000/2000/3000(*)

Table 2.2: Acquisition parameters for the three datasets representing low, medium and high DW-MRI quality. (*): the HCP dataset was limited to the inner shell for the DTI modelling stage as the Gaussian diffusion assumption does not hold at higher b-values.

The GSK and the Whitehall data were corrected for movement and eddy current distortions using the eddy correct algorithm, available as part of FSL [Smith et al., 2004]. For the HCP data the acquisition protocol includes specific gradient scheme and phase encoding direction (blip-up, blip-down), enabling us to use a more advanced method to correct for EPI and eddy current distortions [Andersson et al., 2003] [Andersson and Sotiropoulos, 2016].

The diffusion tensor was estimated using DTIFIT, part of FSL [Smith et al., 2004]. For the HCP dataset, we only retained the DW volumes with a b-value

of $1000 s.mm^{-2}$, as the multi-exponential decay observed at high b-values cannot be modelled correctly with a simple diffusion tensor [Clark and Le Bihan, 2000]. Additionally, the gradients were corrected for non linearity, the latter being due to the specific nature of the scanner hardware [Sotiropoulos et al., 2013].

As the evaluation pipeline relies on tractography, we used a probabilistic multi-compartment model that represents correctly crossing fibres and uncertainty on the principal fibres orientation [Jbabdi et al., 2012]. The tractography was performed with the probtrackX algorithm [Behrens et al., 2007], using 27 predefined sets of masks [de Groot et al., 2013]. An illustration of the tracts is shown in Figure 2.2.

2.2.3 Template creation strategies

The choice of the template used to normalise our dataset was driven by the nature of data used by the registration method (FA, Tensor or DWI). We hence used FMRIB58 (FA), IIT3 (tensor and FA) and study specific (FA, tensor and DWI) templates.

Varentsova et al. developed a HARDI template referred to as IIT3 [Varentsova et al., 2014]. It was computed by using DTITK to register 72 DTI datasets to a temporary template (started from IIT2 DTI template), then averaging the registered tensor images and repeating the same process for 6 iterations.

The template construction pipeline is illustrated in Figure 2.1. To create the study specific template (SST), we used a method that minimise the bias introduced by the choice of the initial reference image used for registration [Guimond et al., 2000]. First, the subjects are linearly aligned to a standard space template (FMRIB58). Each registered subject image is averaged to create the bootstrap template. Then, the subjects are linearly registered to the template, and this process is iterated 3 times. Finally, the subjects are non-linearly aligned to the template. The bias is

reduced by applying the inverse mean transform to the subjects in common space, and the process is repeated 6 times.

Additionally, we applied the calculated warps to tensor images in order to get a template created from the averaged tensor maps, as opposed to averaging FA. The tensor maps were averaged using the log-euclidean framework described in [Arsigny et al., 2006]. The mean FA image is then obtained from the average tensor map.

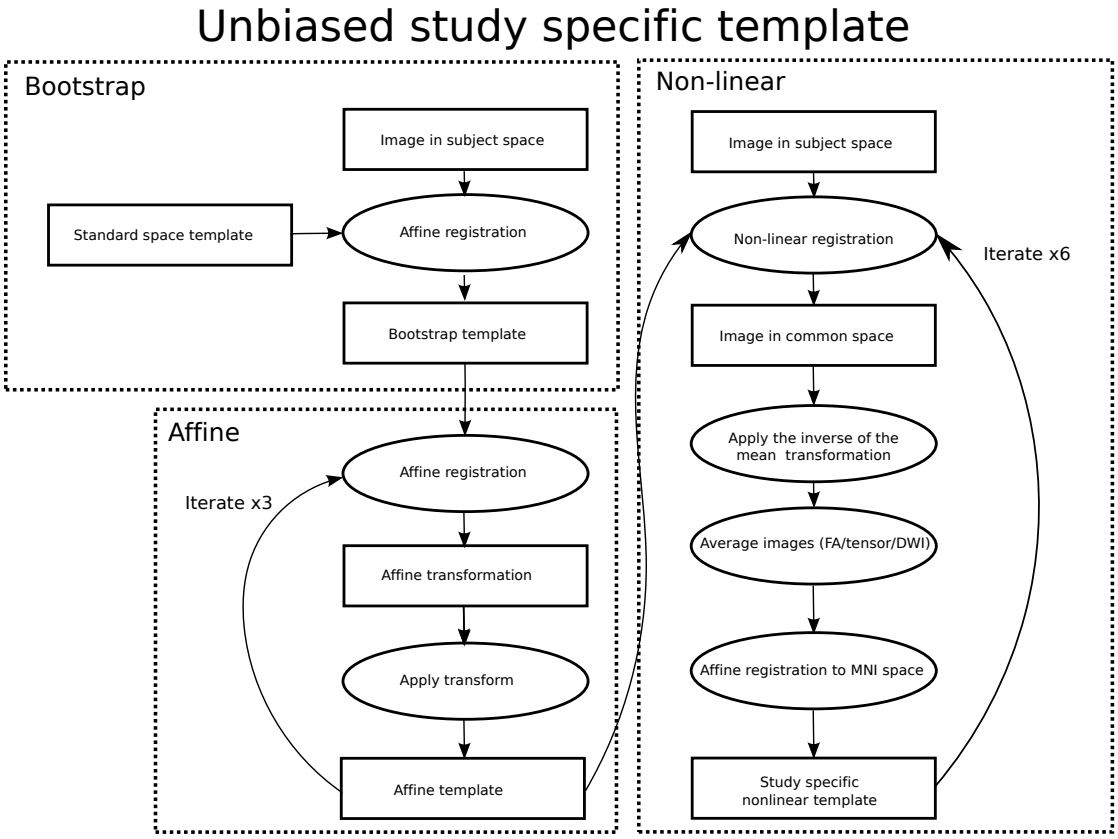


Figure 2.1: Pipeline for the construction of an unbiased study specific template

The SSTs were built using each different algorithm, reinforcing the impact of the algorithm on the final performance score.

2.2.4 Performance evaluation

The evaluation pipeline we used is described in [de Groot et al., 2013]. It makes use of the ball and stick model and probtrackx probabilistic tractography method in order to obtain tract density maps [Behrens et al., 2007]. The fibres bundles are obtained with a seed, stop, and target mask described in [de Groot et al., 2013] and defined in standard MNI space. The probabilistic tractography computes the probability of encountering a streamline belonging to the tract within the voxel. We want to outline that although we use the term "density" to refer to the tract probability, it is not the actual tract density but a normalised count of the number of streamlines belonging to a given tract in the voxel [Behrens et al., 2007]. The registration performance was computed across subjects on the warped tract density images. Normalised and non-thresholded tract density maps were warped to the common space, using a given algorithm and template. The warp density maps were then compared across tracts for each pair of subjects.

2.2.4.1 Spatial correlation

The first performance metric we used is the same as the one from the original paper, i.e., spatial correlation [de Groot et al., 2013]. It can be written as:

$$C = \frac{\sum_i J_i K_i}{\sqrt{\sum_i J_i^2}\sqrt{\sum_i K_i^2}} \tag{2.6}$$

with J_i and K_i the two subject's density map in a given voxel i.

We hence obtain a single spatial correlation measure for a given tract and for each pair of subjects. This measure is then averaged across subjects, and eventually across tracts. It provides a good estimate of the alignment of the tracts across

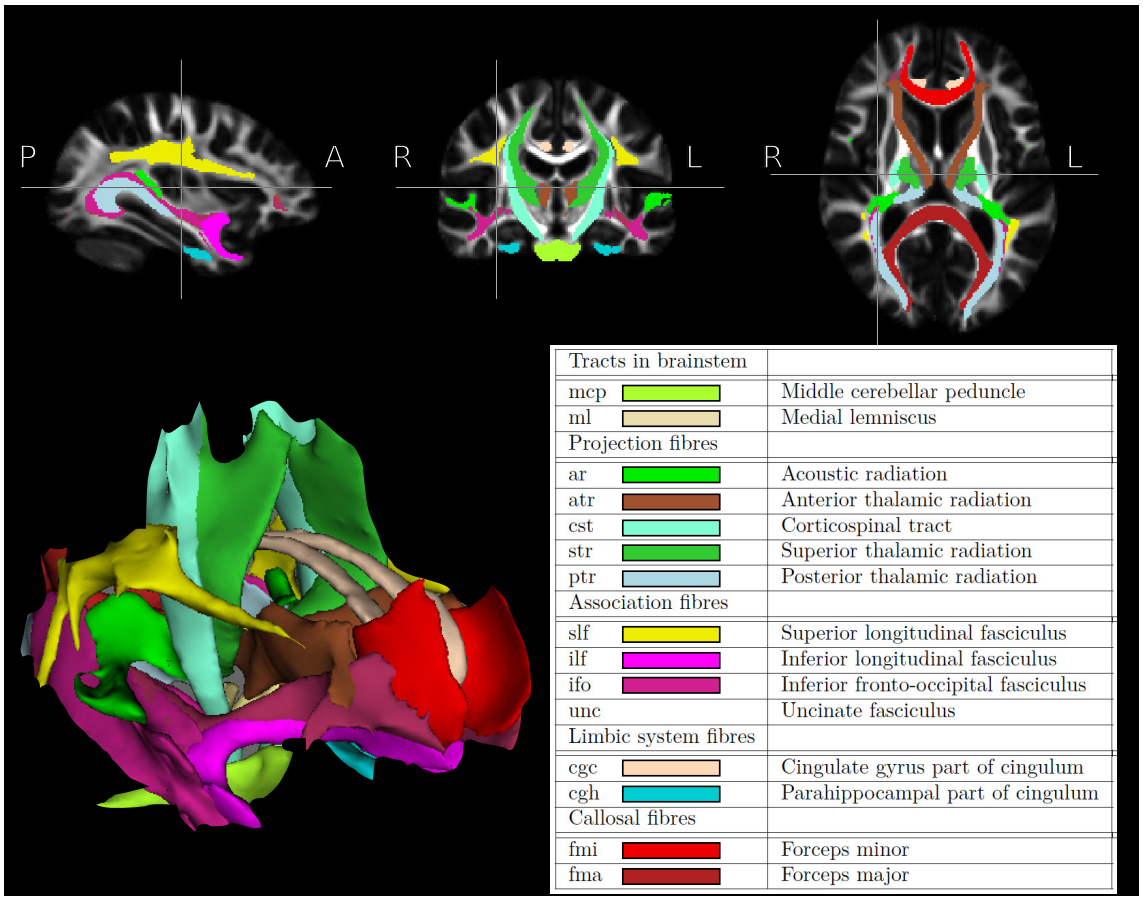


Figure 2.2: Representation of the tracts computed with the autoPtx pipeline. The tracts represented are issued from the Whitehall group averaged density maps in common space, thresholded to 0.005.

subjects and it is less correlated with the quantity optimised by the different registration algorithms, i.e. FA, tensor and DWI signal, than the FA. The correspondence between the spatial correlation and the registration error is shown in Figure 2.3. The spatial correlation metric is a global measure and is not specific to the localisation of the voxel in the tract. It means that a voxel with very low density (or streamline probability) will have the same weighting as one located in the middle of the tract.

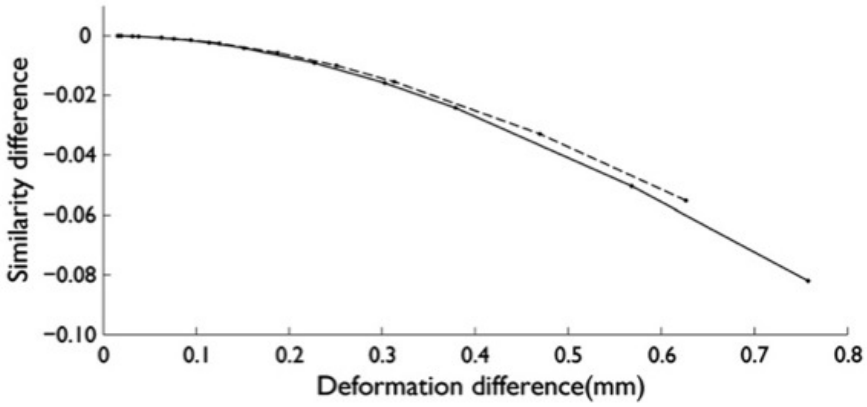


Figure 2.3: Spatial correlation difference as a function of the registration error (in mm). The solid line represents the correspondence established on a low-quality dataset, and the dashed line represents the correspondence obtained on a higher quality dataset. Figure extracted from [de Groot et al., 2013].

2.2.4.2 Tract overlap across subjects

To address the limitations of the spatial correlation, we developed another metric that relates to the tract correspondence across subjects. We compute it following these steps:

- The density maps are registered to the given template using one of the methods
- The registered density maps are thresholded to a value of 0.005
- Each voxel encodes the probability of the tract being present across subjects
- The absolute volume of overlap is given by the number of voxels where the tract is present in at least 90% of the subjects
- The relative volume is given by the absolute volume divided by the mean volume of the thresholded tract across subjects

The tract overlap across subjects measure is voxel-wise and it informs us on the coherence of the tract across subjects. This is a good measure of the specificity of the registration process, as the value of a given voxel indicates how confident one can be that the structure of interest in a voxel is present in all subjects. It is particularly interesting in the context of TBSS, as the projection step can select a voxel that is

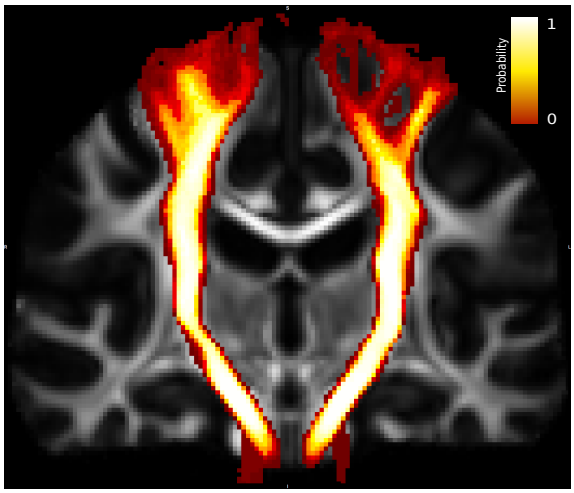


Figure 2.4: Example of the corticospinal tract overlap across subjects on the whitehall dataset. The coloured voxels represent the normalised tract overlap of the thresholded tract visitation map across the 40 subjects.

not representing the same structure across subjects. It is also more sensitive to the alignment of the shape of the tract because while the extremities can rarely overlap across all the subjects, the centre is more consistently defined. This measure can be related to the Dice coefficient, but extended to multi-class labels, where each label is a subject. It also presents the advantage of providing a voxel-wise probability, that can be used to define the region where the group statistics can be performed (e.g. voxels where more than 90% of the subjects are represented). In Figure 2.4, we illustrate the CST overlap across 40 subjects from the Whitehall dataset.

2.2.4.3 TBSS Projection step

We assessed the effect of the projection step in the context of improved registration. We are comparing the performance achieved by the different registration algorithms within the skeleton mask, with and without projection. The mask and the projection step are driven by the FA image, and we project or mask the density maps onto the skeleton. Using the similarity metrics described in the previous section, we can quantify how the skeleton masking or projection affect the registration

performance in terms of spatial correlation and tract overlap across subjects.

2.2.4.4 Group study: Alzheimer's disease

We studied the impact of spatial normalisation on a typical clinical study. We performed a preliminary TBSS analysis on 54 Alzheimer's Disease (AD) patients and 60 healthy elderly participants from the GSK dataset. The clinical data is described in [Douaud et al., 2011]. We conducted the cross-subjects comparisons on FA, using threshold-free cluster enhancement [Smith and Nichols, 2009] and permutation test [Winkler et al., 2014] to detect the clusters where FA significantly varies between control subjects and AD patients. We compared the results obtained with FNIRT, oFNIRT and DTITK using a study-specific template and masking the result with the skeleton, omitting the projection step.

2.3 Results

2.3.1 Registration algorithm

We evaluated the four algorithms on the three datasets, with a focus on the Whitehall dataset as it represents the quality most often encountered in DTI studies. The spatial correlation as defined in Eq. 2.6 ranged from 0.5 to 0.8.

Globally, DTITK outperformed FNIRT, Optimal FNIRT and FNIRT4D, regardless of the template and dataset used. On the Whitehall dataset, the spatial correlation varied widely with the tract of interest, ranging from 0.5 to 0.75. Looking at the average correlation across tracts, in Table 2.3, we see that DTITK ranks

first, followed by respectively oFNIRT, FNIRT and finally FNIRT4D. The difference between the first three methods was much smaller than between FNIRT4D and the others, as the latter extends to 0.07, which would represent a misalignment of 0.7mm, according to Figure 2.3. The average difference between FNIRT and oFNIRT was 0.019 (or 0.26mm) while the difference between oFNIRT and DTITK is 0.014 (or 0.2mm) on average.

Algorithm	Spatial Correlation	CV
FNIRT4D	0.5875	17.5%
FNIRT	0.6433	13.5%
oFNIRT	0.6627	12.8%
DTITK	0.6772	13.7%

Table 2.3: Average spatial correlation obtained with FNIRT4D, FNIRT, oFNIRT and DTITK on the Whitehall dataset. CV represents the coefficient of variation across tracts.

Figure 2.5 illustrates the performance obtained with the 4 methods on the 27 tracts identified with the autoPtx pipeline described in [de Groot et al., 2013]. The tracts are ordered by the maximum performance achieved, from high to low performance. We observe that FNIRT4D exhibits poor performance relatively to the

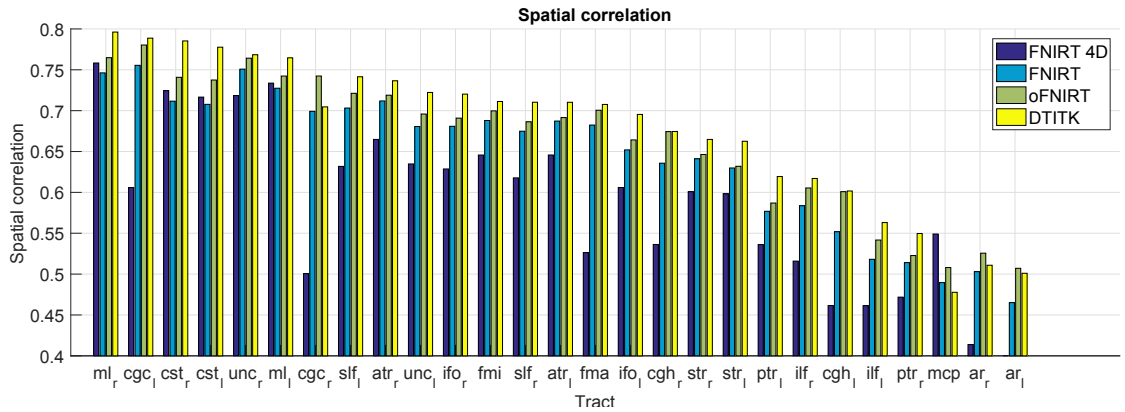


Figure 2.5: Spatial correlation obtained with FNIRT4D, FNIRT, oFNIRT and DTITK for the 27 tracts on the Whitehall dataset

other methods in all the fibre bundles, except in the cerebellum, where it ranked first. It also outperformed the standard FNIRT in the medial lemniscus and cortico-spinal, consistently for the right and the left tract.

Optimised FNIRT outperformed FNIRT for all the tracts, which confirms the work of de Groot et al. [de Groot et al., 2013].

Comparing FNIRT and DTITK, we see that the latter performs better, with the exception of the right cingulum, the acoustic radiations and the cerebellum. One can note that those are thin tracts, and that the advantage for DTITK is more important in large tracts, such as the Corticospinal tract and the superior longitudinal fasciculus. The latter advantage can be explained by the contrast within white matter which exists in tensor images, allowing DTITK to preserve the orientation coherence in the tract. The difference in performance between oFNIRT and DTITK was mostly observed in low tract density regions, i.e. at the extremities of the tracts. In these regions, we know that partial voluming is highly present and that the FA alone is not as robust to noise than in pure WM regions. It is also known that the tensor orientation is less impacted by partial voluming than FA, because the latter occurs in tissues where the diffusion is mostly isotropic (GM ans CSF). As DTITK makes use of the orientation information, it is capable of aligning voxels with low tract density. We also found a negative correlation between average tract volume and spatial correlation ($r = -0.32$ on average). It means that the larger the tract, the lower the spatial correlation. One can also note that the coefficient of variation (CV in Table 2.3) is the lowest for oFNIRT, indicating that the latter is more consistent across tracts.

The other performance metric, the tract overlap across subjects, gives different results than the spatial correlation. As we can see in Table 2.4, oFNIRT now outperforms DTITK and all the other algorithms. The average relative overlap varies from 17% (FNIRT4D) to 28% (oFNIRT). The latter number indicates that, on average, 28% of the initial tract volume overlaps across more than 90% of the subjects. Once again, the lowest coefficient of variation across tracts is reached by oFNIRT. Although the relative overlap appear to be low, it points out the ill-posed nature of the registration to match the shape and topology of the tracts across

individuals.

Algorithm	Mean relative overlap	CV
FNIRT 4D	0.174	41.4%
FNIRT classic	0.234	34.6%
oFNIRT	0.276	34%
DTITK	0.261	36%

Table 2.4: Average tract overlap across subjects obtained with FNIRT4D, FNIRT, oFNIRT and DTITK on the Whitehall dataset. CV represents the coefficient of variation across tracts.

The relative volume of overlap for the individual tracts (Figure 2.6) shows important variations between tracts. Values are ranging from 55% for the cingulum to 11% for the middle cerebellar penducle. Although oFNIRT gives the best overlap for most of the tracts, DTITK performs better for the ml, atr, str and ptr. Again, FNIRT4D performance is much lower than the three other algorithms, except for the mcp, as for the spatial correlation. We did not found correlation between average tract volume and mean relative overlap (r=-0.01 on average with a maximum r=-0.07 for oFNIRT). It means that this metric allows a performance comparison between tracts without the size confound.

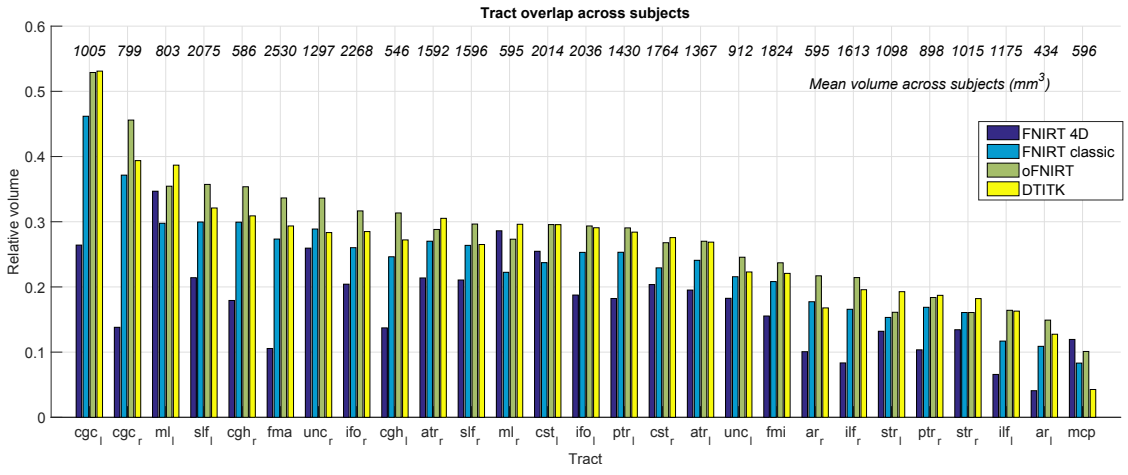


Figure 2.6: Tract overlap across subjects. Relative volume change of the tract. Results are shown for the Whitehall dataset, using a study specific template

Although the ranking between oFNIRT and DTITK changed whether we look at the spatial correlation or the tract overlap, the two performance indicators showed

consistent results when comparing FNIRT4D and FNIRT. Also, the performance ranking across tracts was consistent, with tracts such as the mcp being poorly registered and others, such as the cingulum, and the slf being well registered. Indeed, we found a good correlation between the spatial correlation and the relative tract overlap (r=0.76). Given the poor results obtained with FNIRT4D and its computational cost, we decided to exclude it from the rest of the performance comparisons.

2.3.2 Dataset quality

In Figure 2.7, we can see how FNIRT, oFNIRT and DTITK algorithms perform on the three datasets, i.e. GSK, Whitehall and HCP.

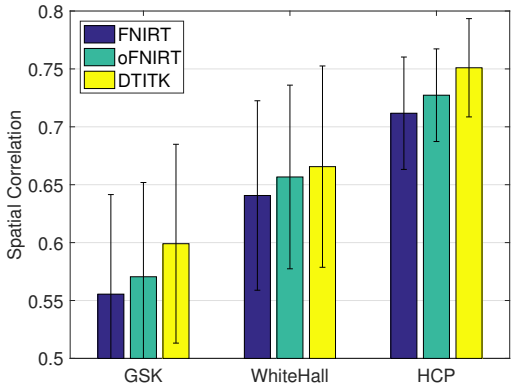


Figure 2.7: Spatial correlation for the three datasets. The error bars represent the standard deviation across tracts.

As presented in Table 2.2, the three datasets differ first by their spatial resolution. We can first see that the quality of the data greatly influences the registration accuracy. The difference across datasets is indeed higher than the difference between algorithms. For example, the difference between GSK and Whitehall dataset for oFNIRT is 0.09, that is about 0.8mm, whereas the difference between oFNIRT and DTITK on Whitehall dataset is only 0.013, or 0.3mm. We note that although the standard deviation within each set of algorithm/dataset is high, every single result

significantly differs from the others (according to pair-wise t-test), with the highest p-value being between oFNIRT and DTITK on the Whitehall dataset ($p = 0.01$, no multiple comparison correction applied).

This effect is expected as images with higher resolution contain more information to drive the registration process. The difference in performance can be further explained by the resolution at which the performance is computed. One has to remember that during the registration, the images are upsampled to the template resolution, i.e. 1mm. It means that computation of the warp partly relies on interpolation, which will have a stronger effect as the resolution of the original image gets worse.

This confirms that the extra information available at an increased resolution is very beneficial for registration accuracy. We also tested different upsampling approaches before registration, i.e. tensor upsampling. The latter did not significantly improve the registration accuracy, although the upsampled images appeared sharper.

2.3.3 Templates

Figure 2.8 shows the standard templates (FMRIB58 and IIT) as well as the study-specific templates (from 40 whitehall subjects) built with the different algorithms. Comparing FMRIB58 and FNIRT SST template, we see that the latter is significantly sharper, which translates into a smaller resolution element (RESEL) [Worsley et al., 1992]. Given that those two templates were created with the same algorithm, we can reasonably assume that the Whitehall dataset is of higher quality than the images used to build the FMRIB58 and that our iterative template creation strategy is more accurate. We also found that the iterative template construction converges after 5-6 iterations (on the HCP dataset). It means that the template sharpness (measured with

the RESEL) does not increase further and neither does the registration performance. We noted that the number of iterations needed for convergence increased with the quality of the dataset and the accuracy of the registration algorithm.

We also noticed that averaging the tensors instead of FA gives a sharper template, meaning that the tracts are better delineated and that the RESEL is significantly smaller (respectively 8.6mm and 10.6mm for tensor and FA with oFNIRT). We explain this by the fact that interpolating tensors with different orientation gives low FA, resulting in a good delineation between different bundles, and reinforcing the gradient between gray and white matter. Overall, the template built with oFNIRT and tensor averaging is the sharpest one.

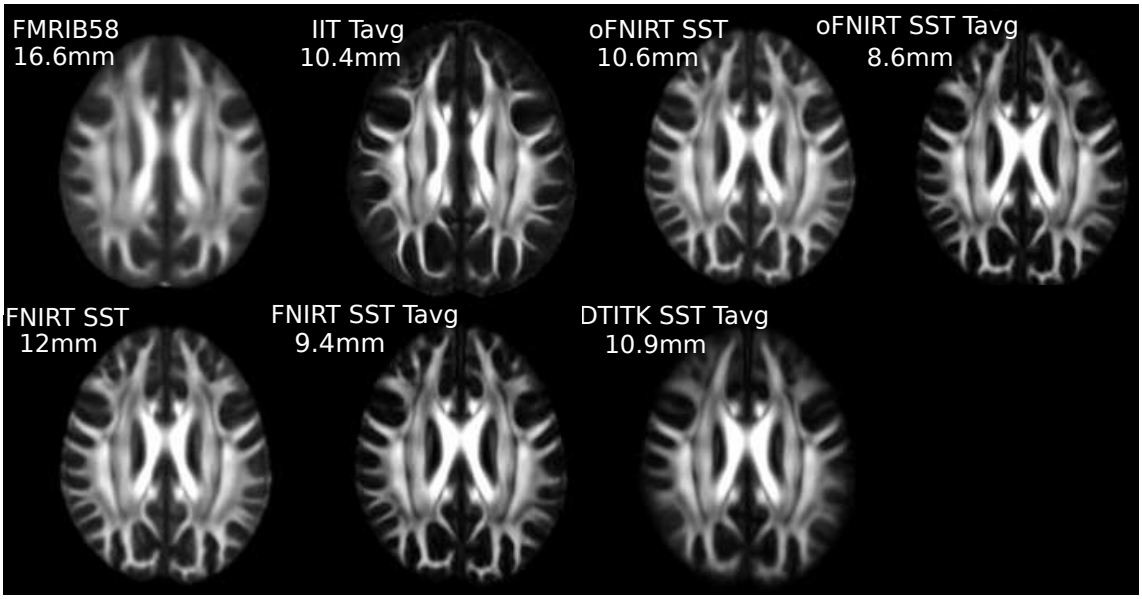


Figure 2.8: Standard and study specific templates with their actual spatial resolution (RESEL) in mm. The templates created by averaging tensors are noted Tavg.

We also assessed the impact of the template on the group-wise registration accuracy in terms of spatial correlation. The results are shown in Table 2.5. For FNIRT, the best performance is achieved with the tensor averaged SST, although the difference with FMRIB58 is not significant. This suggests that FNIRT does not take advantage of the sharper template. Nevertheless, there is a significant difference

with the IIT template, which was built using DTITK. It means that it is preferable to use a template that was built using the same registration algorithm as the one used to register the population. There is also a significant difference between SST tensor averaged and FA averaged.

Algorithm	Spatial Correlation	CV
FNIRT FMRIB58	0.64589	13.8%
FNIRT IIT	0.6354*	15.4%
FNIRT SST	0.64328*	13.5%
FNIRT SST Tavg	0.6478	13.7%
oFNIRT FMRIB58	0.65543*	13.4%
oFNIRT IIT	0.66183	14.4%
oFNIRT SST	0.6627	12.8%
oFNIRT SST Tavg	0.66574	13.3%
DTITK IIT	0.67775	13.6%
DTITK SST	0.67716	13.6%

Table 2.5: Overall performances of the different algorithms and templates for whitehall dataset. * indicates a significant difference with the best performing template for a given registration approach

With oFNIRT, the best performance is also achieved with a tensor averaged SST, and this in this case the difference is significant with the FMRIB58 template. One can also note that for both FNIRT and oFNIRT, the lowest coefficient of variation is obtained with the SST. This suggests that the latter gives more robust results.

For DTITK, there is no significant difference between IIT and SST, and one can also note that they have the same effective resolution.

2.3.4 TBSS skeleton and projection step

In this section, we evaluate how the registration performance changes when we perform the projection onto the TBSS skeleton. We can see in Table 2.6 that restricting the evaluation to the skeleton (as a mask) gives an improved performance (compared to the full extent of the tract, cf. Table 2.3). It means that the registration

algorithms perform better at the centre of the tracts. This is expected, as the anatomical variability across subjects increases as one moves away from the centre of the tract, defined here by the FA skeleton.

Algorithm	Spatial Correlation	CV
FNIRT Masked	0.69635	12.3%
FNIRT Proj	0.64018	14.3%
oFNIRT Masked	0.7087	11.8%
oFNIRT Proj	0.65195	13.9%
DTITK Masked	0.71672	13.0%
DTITK Proj	0.64694	16.2%

Table 2.6: Spatial correlation with FNIRT, oFNIRT and DTITK within the TBSS skeleton, with and without the projection step. Results obtained on the Whitehall dataset, using a study specific template

As we can see in Table 2.6, the projection step decreases the average performance of all the algorithms, with a drop of more than 0.05 in spatial correlation (compared to Table 2.3). It means that the projection step actually increases the registration error by slightly more than 0.5mm on average.

The ranking of the three algorithms is preserved, as DTITK performs better than oFNIRT, which is in turn better than FNIRT. DTITK and oFNIRT results restricted to the skeleton are better than the projected ones from FNIRT. It suggests that oFNIRT performs better than standard TBSS pipeline when projecting the density maps. Indeed, the projection step is meant to decrease the registration error only in terms of FA. Using a metric that relates more closely to the anatomy, we can see that the anatomical correspondence is lower when we apply the projection step. This shows that reducing the FA error does not necessarily improve the anatomical correspondence between subjects, and may be detrimental in a group study by reducing sensitivity.

We reach the same conclusion when we compare masking and projecting in terms of tract overlap across subjects. The latter is summarised in Table 2.7. We see that the projection step once again decreases the performance, compared to masking

within the skeleton and even compared to the results obtained without masking (cf. Table 2.4). This metric also reveals that oFNIRT outperforms DTITK both with masked and projected results. One can also note that the masked results with DTITK are lower than with FNIRT. This suggests that FNIRT and oFNIRT are performing better than DTITK in terms of shape registration and that the result is more specific.

Algorithm	Mean relative volume	CV
FNIRT masked	0.305	47.5%
oFNIRT masked	0.33	44.5%
DTITK masked	0.275	46.9%
FNIRT projected	0.195	67%
oFNIRT projected	0.229	60.2%
DTITK projected	0.23	70.8%

Table 2.7: Average tract overlap across subjects with TBSS skeleton. Average relative volume of overlap of the tract within the TBSS skeleton, with and without the projection step. Results are shown for the Whitehall dataset, using a study specific template

In Figure 2.9 we show the tract overlap across subjects for each tract, with and without the projection step. The first observation is that the order of the tracts is globally unchanged, compared to the results obtained on the whole tracts (Figure 2.6). For example, the cgc is on the top of the ranking in both cases, while the mcp, ar and str are towards the end. One can also note that the projection is beneficial in a few cases. For instance, the DTITK projected results are the best ones in the cgh. This suggests that projecting density given maximum FA is well adapted for this tract. Nevertheless, in most cases, the optimal approach remains an optimised registration (oFNIRT or DTITK) without projection step.

To illustrate why the projection step is detrimental to anatomical correspondence across subjects, we show the discrepancy between FA and tract density skeletons in Figure 2.10. In most of the voxels, the FA skeleton does not match with the the density skeleton. The projection step selects a voxel that is located at the centre of the tract as defined by FA. However, this voxel is not guaranteed to be at the centre

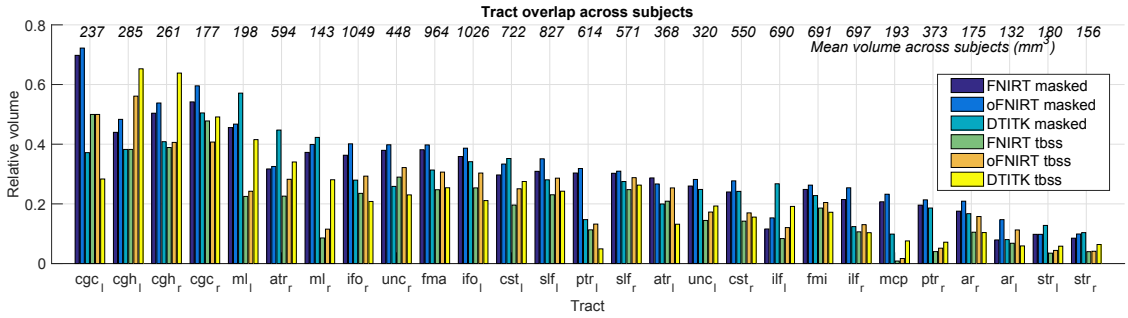


Figure 2.9: Tract overlap across subjects with TBSS skeleton Relative volume of the tract overlap within the TBSS skeleton, with and without the projection step. Results are shown for the Whitehall dataset, using a study specific template

of the tract when we defined the latter by its density. When we compute the spatial correlation on FA, the registration performance slightly increases with the projection step. By de-projecting the skeleton voxels, we observed that the projected density was on average lower than the masked density. This suggests that the projection step tends to select voxels with a lower probability for the tract being present, likely because FA is poorly correlated with the tract density in crossing fibres areas.

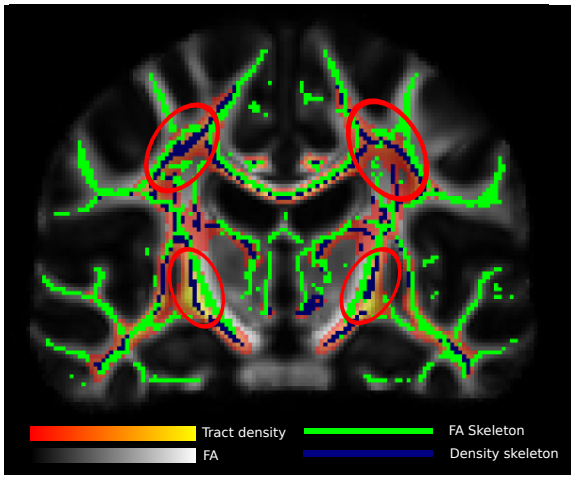


Figure 2.10: TBSS skeleton derived from mean FA and tract density maps. Mean FA image (Grey), mean density maps (yellow-red), FA derived skeleton (Green) and Density Derived skeleton (Blue). The red ellipse outline the locations where the skeleton derived from the density map differs from the one derived from the FA

Once again, this shows that the projection step may only be adequate to enhance FA registration, but might be inadequate to compare other values, such as tract density. With the assumption that the tract is better defined by the density (that is

the tract probability), we can conclude that projecting with respect to maximum FA is anatomically incorrect.

2.3.5 Group study: Alzheimer's disease

To assess the impact of registration algorithms on a group study, we performed a TBSS analysis with standard FNIRT, oFNIRT and DTITK registration algorithms. We used a study-specific template with tensor averaging as a reference, as described in section 2.2.3. Following our findings in the previous section, we did not apply the TBSS projection step and only used the FA skeleton as a mask.

The results obtained with the three algorithms show widespread FA decrease in Alzheimer's patients.

With FNIRT, significant FA decreases appeared in the anterior and middle part of the corpus callosum, and in the fornix.

oFNIRT reveals more differences (circled in yellow) in FA than FNIRT, notably in the right inferior fronto-occipital fasciculus and in the posterior part of the corpus callosum. The extent of the FA differences in the fornix was also increased with oFNIRT. Globally, the pattern of FA decrease for Alzheimer's disease patients is the same, with an increased sensitivity for oFNIRT compared to FNIRT.

DTITK appeared to bring more sensitivity to the group-wise FA comparison as the differences in FA were more widespread than with oFNIRT. For instance, DTITK revealed FA differences (circled in blue) in the whole of the corpus callosum, the left inferior fronto-occipital fasciculus and the inferior longitudinal fasciculus. This suggest that the improved performance observed with DTITK in terms of spatial correlation leads to a more sensitivity in the context of AD group study.

We note that in the previous sections, the differences in performance between FNIRT and oFNIRT were larger than between oFNIRT and DTITK. It appears

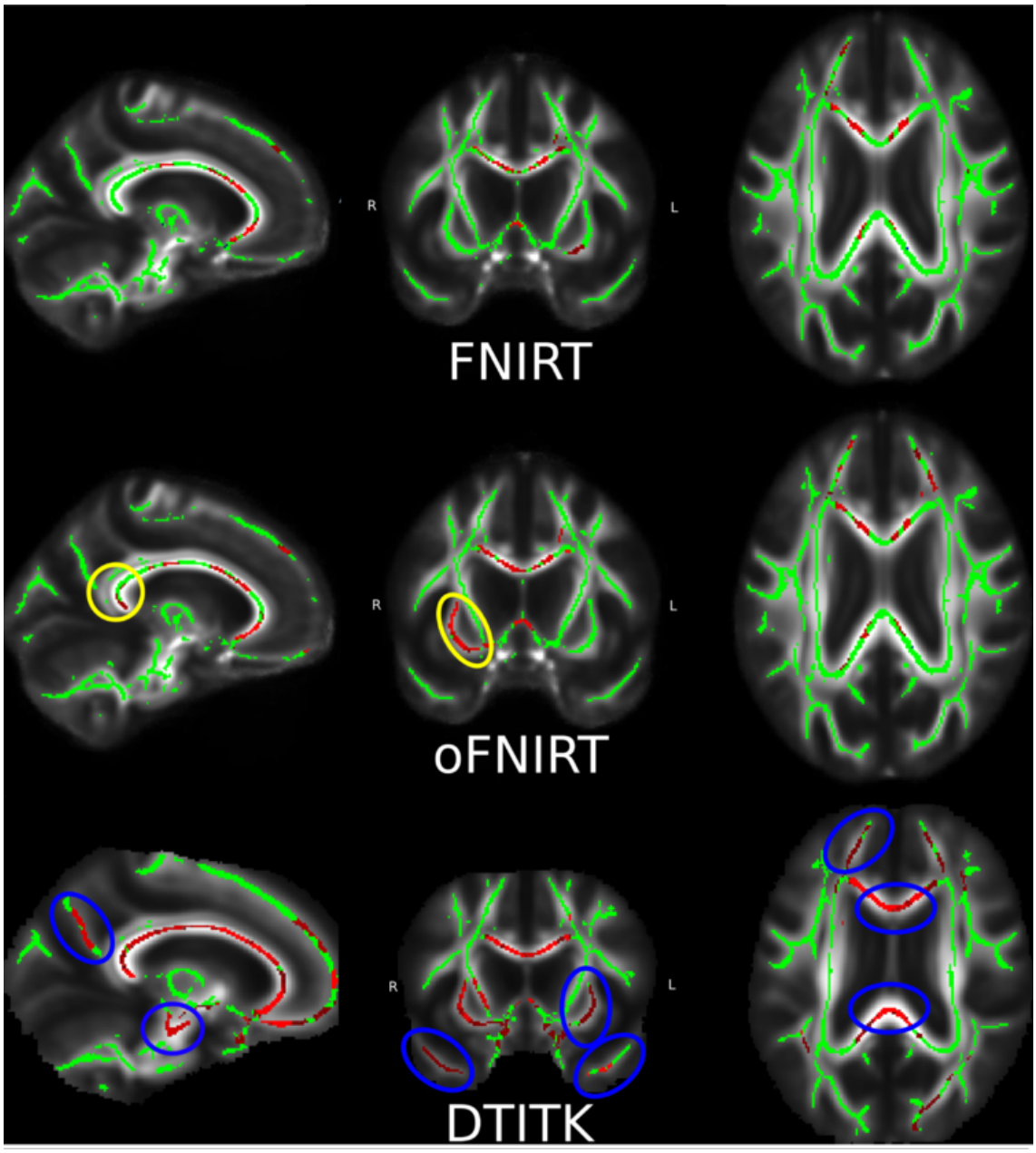


Figure 2.11: TBSS result with FNIRT, oFNIRT and DTITK on Alzheimer's disease, showing decreased FA in AD subjects (FWE corrected, $p < 0.05$). Top: FNIRT, middle: oFNIRT, bottom: DTITK. Red: FA decrease in AD patients, green TBSS skeleton. Yellow circles: Difference revealed by oFNIRT and blue circles: differences revealed by DTITK.

that in the group study, the difference between DTITK and oFNIRT is larger than between FNIRT and oFNIRT. We hypothesize that this result is related to the ability of DTITK to match voxels where tissue-type partial voluming is present by using the orientation information. Conversely, FA based registration methods such as FNIRT

would account for partial volume and smoothen the group differences related to partial volume effects.

2.4 Discussion

In this chapter, we evaluated the performance of 4 registration algorithms and template building strategies in terms of anatomical correspondence across 40 subjects for 3 datasets. The evaluation was driven by 27 anatomical structures that were delineated via probabilistic tractography. We used spatial correlation as a performance indicator, as well as the tract overlap across subjects, that is a measure of registration specificity, and is also more sensitive to the shape of the registered structures.

2.4.1 Overall performance

An important part of this work was to assess the performance and anatomical specificity of registration methods working on different dimensionality of the data. The results obtained with FNIRT4D, which uses the same optimisation approach as FNIRT, were surprisingly poor. Indeed, one could expect that using the extra information contained in the DWI instead of a summary (FA) would improve the registration performance. We suggest that the relatively poor performance of FNIRT4D can be partly due to the angular interpolation that is not explicitly part of the optimisation scheme. It is only used to resample the DWIs, and not used to find the best local rotation between two anatomical locations. As a result, the registration is mostly driven by the ADC, which has a relatively poor contrast. Conversely, with DTITK the tensor reorientation is explicitly part of the optimisation. This is the

key difference with FNIRT, oFNIRT and FNIRT4D, and might explain why DTITK performs better than FNIRT and FNIRT4D in all the cases and better than oFNIRT in some cases.

The gain in registration accuracy between FNIRT and oFNIRT relies on an optimised registration scheme that consists of three steps, with increasing degrees of freedom. Our results show that the higher Dof of oFNIRT warp is beneficial to the registration accuracy in all cases when compared to FNIRT. This is in line with a study where 14 algorithms were compared and the number of degrees of freedoms was related to the ranking of the method [Klein et al., 2009]. The medium dimensionality of DTITK compared to oFNIRT (400k vs 1.6m) might hinder its performance, but its regularisation scheme being different, we did not try to increase the number of DoF for DTITK.

Comparing oFNIRT and DTITK, we saw that the latter shows better spatial correlation. This result is in line with a study where DTITK also obtained better results than FNIRT, but oFNIRT was not included as it was not available yet [Wang et al., 2011]. In the latter publication, the evaluation was based on FA correlation with the template, that is more circular than correlating density maps. When we performed the evaluation by computing the spatial correlation on FA instead of density maps, oFNIRT actually outperformed DTITK. However, we discarded this result, as we considered such analysis too circular. Nevertheless, it can be considered as a sanity check, which confirmed that oFNIRT was the best method to reduce the FA error. One could nevertheless argue that the density maps are more correlated with the tensor than with the FA, thus introducing a bias in the analysis in favour of DTITK. This was one of the motivations to introduce a performance measure that is less circular than the spatial correlation of density maps.

The tract overlap across subject is indeed one step further away from the tensor because it involves shape correspondence of the individual tracts across subjects, and it is a global measure. It is also an interesting indicator of specificity of the

registration as it ensures that a voxel in common space represents the structure of interest across the population. The difference between spatial correlation and tract overlap resides mostly in their interpretation. The former is a better overall measure of good registration, while the latter describes very well how the overall shape of the tract can be matched across subjects. The benefit of the spatial correlation is that the actual density is involved in the metric, meaning that some properties of the tracts are taken into account. It makes an important difference with the tract overlap, where only the shape given by the thresholded density is considered. The advantage of the tract overlap across subject is that it is a conservative measure of specificity, which is very important for the interpretability of the results in the context of group studies.

The results obtained with the tract overlap across subjects were slightly different to the ones obtained with the spatial correlation. While the ranking of FNIRT4D and FNIRT with respect to oFNIRT did not change, the latter proves to be superior to DTITK for most of the structures studied. This might first be explained by the higher dimensionality of oFNIRT. Another factor to take into account is that DTITK might be less sensitive to partial volume effects. In fact, DTITK can match a voxel composed of CSF and white matter with a voxel composed of pure white matter, as long as the fibres have similar orientation. oFNIRT being based on FA, the two same voxels will not be matched.

The increase in performance brought between FNIRT and oFNIRT comes from the increased degrees of freedom coupled with a careful step by step regularisation and sampling scheme. On the other hand, DTITK relies on more information to find an optimal warp. We did not attempt to combine the two enhancements (increased DoF and tensor similarity measure), but we strongly believe that it would lead to an even better registration performance.

Finally, we conclude that DTITK shows the best overall performance when one is interested in the whole extent of the white matter. On the other hand, oFNIRT

leads to more anatomically accurate results, which can be interesting when one wants to study tracts separately, or when the effect of a given pathology is expected to impact specific tracts.

2.4.2 Dataset and template strategy

We quantified the improvement related to increased image quality. We conclude that the effect of image quality on the registration performance is two times more important that the algorithm used (0.3mm difference between algorithms and 0.8mm between GSK and Whitehall dataset). One has to consider that the huge improvement in registration performance observed with the better quality dataset might be partly due to the improved tractography. As a matter of fact, the increased number of directions and resolution result in a better signal modelling (we used bedpostX with 3 fibres for the three datasets), which in turn will provide better density maps. Nevertheless, better diffusion modelling is also an argument for pushing DWI image quality in order to improve group studies sensitivity.

We also quantified how the template used as a reference can affect the registration accuracy. First, using a study-specific template was always at least as good as using a standard template. This confirms the results of previous studies, where a similar effect was observed [Shen et al., 2007] [Keihaninejad et al., 2012] [Bach et al., 2014]. An important finding was that generating the template by averaging the registered tensor image had an even stronger effect than using study-specific templates. Averaging tensors with the appropriate metric [Arsigny et al., 2006] leads to a sharper template (as measured with the RESEL) and also better registration results, both with FNIRT and oFNIRT. The improvement is of the same order of magnitude as the improvement between FNIRT and oFNIRT or oFNIRT and DTITK. We conclude here that it is beneficial to manipulate the tensor images rather than FA all along the template

construction pipeline, even when only the FA is used in the similarity metric of the registration algorithm.

2.4.3 TBSS projection step

The second aim of this work was to assess whether the projection step is anatomically correct and if improved registration can alleviate the need for it. Looking at the spatial correlation and the tract overlap, it is clear that the projection step does not improve the registration accuracy, independently of the algorithm used. Our results suggest that the projection step is likely to select voxels with a lower density (or tract probability) than the original voxel masked by the skeleton. We argue that the tract density is a more accurate marker of the tract anatomy than FA, because FA is more sensitive to fibre dispersion, crossing and noise.

Considering the multiple drawbacks of the projection step [Zalesky, 2011] [Schwarz et al., 2014], we suggest that it should not be performed, even when using a standard registration method such as the original FNIRT.

We do not discard the TBSS pipeline completely, as masking the results with the skeleton improved the performance. This is expected, as the skeleton follows the centre of the tract, which is more consistent across subjects, whereas the extremities of the tract might show topological differences that cannot be matched. We did not explore how a projection step driven by maximum density would affect the registration performance. This might increase the anatomical correspondence, but presents the same rick of confounding tracts.

2.4.4 Group study: Alzheimer's disease

In the group study on Alzheimer's disease, we observed an increase in sensitivity to detect FA difference with oFNIRT and in a greater extent with DTITK compared to FNIRT. The improvement observed with oFNIRT can be related to another study, where FNIRT was outperformed by ANTS, which is a FA based non-linear registration algorithm with higher dimensionality than the original FNIRT [Avants et al., 2008].

DTITK prove to be more sensitive than oFNIRT to FA difference between patients and control. This increased sensitivity can be partially explained by the better registration performance. On the other hand, we also conjectured that DTITK was more sensitive to partial volume effects, as it is able to align voxels with different FA, but similar orientation. This is confirmed by our results, as the spatial correlation was higher for DTITK at the extremities of the tracts, and more importantly because the tract overlap (which excludes low-density regions, prone to partial volume effects) was higher with oFNIRT.

As we do not have ground truth available for the Alzheimer's disease study, we cannot reach a conclusion on which algorithm is better for registration in groups studies. Nevertheless, we can argue that DTITK has an advantage over oFNIRT in terms of sensitivity. On the other hand, the results obtained by oFNIRT showed that it is more specific to the tract present in a voxel, and therefore more reliable, although more conservative. Finally, we note that oFNIRT and DTITK produced very different results in spite of a limited difference in performance.

2.5 Conclusion

Given the promising results obtained by both oFNIRT and DTITK, the natural continuation of this work would be to integrate a tensor-based optimisation to oFNIRT. We additionally confirmed several ways of improving the group-wise spatial normalisation in diffusion imaging. Better quality of the acquisition is of utmost importance, and carefully building a study-specific template also proves to have a significant impact on the registration accuracy.

3

Free water: confound or biomarker?

Diffusion MRI is widely used in clinical research to infer micro-structural alterations of white matter in pathologies such as Alzheimer's disease and schizophrenia. The diffusion tensor (DT) is the most common model used to describe white matter properties. As the analyses are usually performed on a voxelwise basis, the outcome can be confounded by macroscopic partial volume effects, where a voxel carries information coming from two or more different tissue types (e.g. free water and white matter). While free water (FW) is merely a confound for voxels located at the interface of two tissues, it has been hypothesized that it could also reflect micro-structural changes occurring in the extra-cellular space.

We propose here to model the contribution of free water on a voxel-wise basis to estimate micro-structural properties of the white matter unbiased by tissue-type partial volume effects, in the context of standard clinical data. Our model uses a prior on the tensor properties, but we also show that using an extra shell with a low b-value allows relaxing this prior. We validate our model through simulations and in-vivo by comparing the results with DT fitting to an inversion recovery acquisition where FW signal is nulled.

We determine the properties of the FW compartment in deep white matter to assess

whether it can be a biomarker for extra-cellular space (ECS) alteration. Using a diffusion weighted inversion recovery sequence with multiple inversion times, and an acquisition including a dense range of b-values, we suggest that the FW in deep white matter is merely an artefact arising from the limitation of the tensor model and the constraints applied to the parameters.

We finally apply the FW model in a typical clinical study and show that most of the changes revealed by the simple DTI model can in fact be attributed to macro-structural changes caused by tissue atrophy.

We conclude that free water is a confound that can be addressed in single shell data when it refers to the CSF. However, we suggest that modelling free water along with a tensor in deep white matter is inappropriate and may reduce the sensitivity to microstructural changes.

3.1 Introduction

The diffusion tensor model and its derived measures, such as FA and MD [Basser et al., 1994], are widely employed in brain imaging studies as markers of WM microstructural changes. However, it is known that tensor-derived measures can be confounded by partial volume effects (PVE), for instance in the presence of CSF or GM. The resulting signal in a given voxel represents a weighted average of the different tissue types, designated as partial volume fraction [Le Bihan, 1995] [Alexander et al., 2001a]. This confounds the common interpretation that tensor-derived measures reflect micro-structure [Alexander et al., 2001a].

In white matter, fibre bundles passing near the ventricles, such as the corpus callosum or the fornix, are prone to CSF contamination [Concha et al., 2005]. The fornix is known to have an important role in memory function and it is suspected to be compromised in diseases such as Alzheimer's [Copenhaver et al., 2006]. Nevertheless,

it is very difficult to interpret DTI indices estimated along the fornix when the resolution is worse than 1mm [Sakaie et al., 2012] (a common resolution in clinical studies is 2-3mm isotropic). As a consequence, WM atrophy within a given voxel can translate into a larger CSF volume fraction and can be wrongly interpreted as being only caused by a change in microstructural properties of the tissue [Smith et al., 2006] [Oouchi et al., 2007]. Various indices including axial diffusivity (AD), radial diffusivity (RD) and MD increase with CSF contamination, while the FA decreases [Alexander et al., 2001a] [Pfefferbaum and Sullivan, 2003] [Metzler-Baddeley et al., 2012]. It has been estimated that more than one third of "WM voxels" (defined as containing at least 25% true WM) are affected by partial voluming at a resolution of 2mm isotropic [Roine et al., 2014]. Modelling the CSF contribution to the diffusion signal is crucial if one is to avoid biased conclusions with respect to the underlying WM microstructure alterations. Nonetheless, this issue is often overlooked in dMRI group studies.

In order to suppress CSF contribution to the diffusion signal, an Inversion Recovery (IR) pulse sequence can be introduced in the image acquisition [Papadakis et al., 2002] [Chou et al., 2005]. Yet, this approach has numerous drawbacks, such as increased acquisition time and lower signal-to-noise ratio (SNR). Composite FLAIR is a recent technique that exploits the strong attenuation of the CSF signal at b-values above 1000 $s.mm^{-2}$. The IR pulse is applied to the B0 images only, with a consequent reduction in scan time and higher SNR [Jung, 2014]. Yet, none of these methods allows an estimation of the CSF volume fraction, which might be a quantity of interest in clinical studies. Another limitation of FLAIR is that it does not attenuate edema signal and it is also questionable whether it attenuates CSF within tissue, as it may have different spin-lattice relaxation time (T_1).

A different approach, performed post-acquisition, proposes to explicitly model the CSF contribution to the signal, using a two-compartment model: one compartment accounts for the fast isotropic diffusion of the CSF and the second describes

the slow anisotropic diffusion occurring in tissues. It has been shown that this model, with the diffusivity of the isotropic compartment fixed to 3×10^{-3} $mm^2.s^{-1}$, performs well at disentangling CSF and tissue contributions when using 8 different b-values, ranging from b=10 $s.mm^{-2}$ to b=1200 $s.mm^{-2}$[Pierpaoli and Jones, 2004].

The free water elimination (FWE) method proposed to fit such a model in the context of acquisition typically encountered in group studies, i.e. single-shell data [Pasternak et al., 2009]. The term free water (FW) is more generic than CSF, as it can either relate to CSF in the ventricles and outside the brain (macroscopic partial voluming) or to fast diffusing water in the white matter (potentially biologically interpretable). The method relies on a variational framework to spatially constrain and regularize the model fitting [Pasternak et al., 2008] [Gur et al., 2008]. However, the regularization step implies smoothing of data and may reduce the sensitivity to subtle effects. More recently, it has been proposed to combine a multi-shell acquisition with the regularization process [Pasternak et al., 2012a]. The shells are limited to two low b-values (b=100, 400 $s.mm^{-2}$) and two higher ones (b=900, 1400 $s.mm^{-2}$). The tensor without free water is estimated using the two higher shells, where the free water signal is mostly attenuated, while the isotropic signal compartment accounting for fast diffusing free water is estimated using the two lower shells, where the signal from FW is not fully attenuated. As a result, the weight of the spatial regularization can be decreased because the tensor is initialized from data with limited free water contribution. This shows that using low b-value shells along with standard dMRI acquisition improves the bi-exponential fit for two-compartment models, as also suggested in [Jbabdi et al., 2012].

In order to evaluate how a clinical study is impacted by FW, Metzler-Baddeley et al. compared the tensor fit with the FWE method [Metzler-Baddeley et al., 2012]. They performed a standard tract-based spatial statistics (TBSS) [Smith et al., 2006] analysis on data from an aging study that included atrophy of the WM structures. They focused on the fornix as this is particularly prone to PVE and they have shown

that FW PVE affects the TBSS results. They compared the FWE method with brain size correction and the latter seemed to have little effect on removing confounded correlations between MD increase and ageing. However, their results with unbiased MD showed no alteration of tissue micro-structure in ageing, suggesting that the FWE method might not be sensitive enough to reveal subtle changes that are known to be present. Another TBSS study of mild cognitive impairment has shown that imperfect correction for FW contribution can result in confounded differences (in the fornix) and mask true ones (in the superior longitudinal fasciculus) [Berlot et al., 2014].

Recently, it has been suggested that the FW within deep white matter might be related to the extra-cellular space (ECS) [Pasternak et al., 2012a]. It has been hypothesized that the FW signal originates from large water pockets that can only exist in the extra-cellular space. Recent studies suggest that FW might be a biomarker for diseases where a neuro-inflammatory process is suspected to occur, e.g. in schizophrenia [Pasternak et al., 2012b] [Pasternak et al., 2015] [Pasternak et al., 2016], dementia [Maier-Hein et al., 2014], or Alzheimer's disease [Hoy et al., 2017]. The FW fraction is in this case considered as a biomarker for neuro-inflammation, and the technique is referred to as free water imaging.

In the first part of this chapter, we propose a method that estimates the FW fraction and the diffusion tensor independently in each voxel and without spatial regularization. We apply a Bayesian framework that allows the use of priors on the parameters to constrain the solution space during the model fitting. We hypothesize that such an approach, where no spatial regularisation is involved, allows to retrieve the tensor indices in a more accurate way than the FWE method. Moreover, we assess the added value of an extra shell with a low b-value where the free water signal is not fully attenuated. The model is validated through simulations and an IR-DWI acquisition where the CSF signal is nulled.

In the second part of this chapter, we investigate the properties of the FW

compartment in deep white matter so to determine whether it originates from the ECS. We estimate T1 spin-lattice relaxation time, diffusivity and volume fraction of the FW compartment in deep white matter by using IR-DWI acquisition with multiple inversion times and a DWI acquisition with a dense range of b-values.

Lastly, we apply the model to a dataset from a standard clinical study of Alzheimer's disease. We use both the diffusion and the T1 structural acquisitions to assess how the macroscopic effects of the disease impact the FW model parameters and the interpretation of the results.

3.2 Materials and methods

3.2.1 Free Water Model

The model we use is similar to the one proposed in [Pierpaoli and Jones, 2004] and also in [Pasternak et al., 2009], which adds a single isotropic compartment to the tensor model:

$$S_i = S_{fw} \exp(-b_i D_{fw}) + S_t \exp(-b_i r_i^T D r_i) \tag{3.1}$$

where S_i is the signal measured with diffusion gradient applied along orientation $\mathbf{r_i}$ with b-value b_i, S_{fw} and S_t are the baseline unattenuated signals from, respectively, free water and tissue. D_{fw} is the free water diffusion coefficient, which we set equal to $3 \times 10^{-3} mm^2 s^{-1}$ throughout the whole study, and D is the diffusion tensor. The tensor D is parametrised using its three eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ and the three polar angles (θ, ϕ, ψ). (θ, ϕ) determine the polar and azimuth angles of the principal eigenvector, while the angle ψ parametrises the angle of the secondary eigenvector on

the plane orthogonal to the principal eigenvector. The third eigenvector is uniquely determined (up to a scaling factor) given the first two eigenvectors and it is therefore not explicitly modelled.

Equation (3.1) is generally ill-conditioned in the single-shell case. Fixing free water diffusivity to a constant value helps to improve the behaviour of the model, but only if the diffusion in the tissue is anisotropic (see simulations in the Results section). If the latter is isotropic, then the three parameters S_{fw}, S_t and the trace of (D) span a two-dimensional manifold, leading to an unidentifiable model.

In order to better constrain the model, Pasternak et al. introduced a spatial regularisation term with a piece-wise smoothness constraint on the tensor D [Pasternak et al., 2009]. Here, we take a simpler approach to the problem, and instead, we assume a Gaussian prior distribution on the tensor MD, with prior mean and standard deviation μ_{MD} and σ_{MD}, truncated to positive values of MD. We show through simulations (Results section) that, whilst this prior stabilises the fit for low anisotropy cases, it does not introduce unwanted bias in the cases of medium to high anisotropic diffusion.

Priors on all other parameters in the model are uniform, with a positivity constraint for the S_F, S_T and the eigenvalues of D. The prior mean μ_{MD} was computed by averaging the MD in the white matter skeleton, on a dataset containing 40 non-pathologic subjects with age 69±5 years, acquired at 3T with 128 DW gradients at a b-value of $1500 s.mm^{-2}$, with a resolution of 2mm isotropic. The skeleton was further thresholded to FA values $ge0.3$ in order to avoid voxels where tissue-type partial volume effects occur.

The resulting MD histogram and the normal distribution fitted on the data are illustrated in Figure 3.1. We observe that the MD values are distributed around $\mu_{\mathrm{MD}} = 6.97 \times 10^{-4} mm^2.s^{-1}$ with a standard deviation $\sigma_{\mathrm{MD}} = 1.04 \times 10^{-4} mm^2.s^{-1}$. The normal distribution fitted on the data does not appear optimal, as its tail is heavier towards higher MD values and it is truncated to positive MD values. While

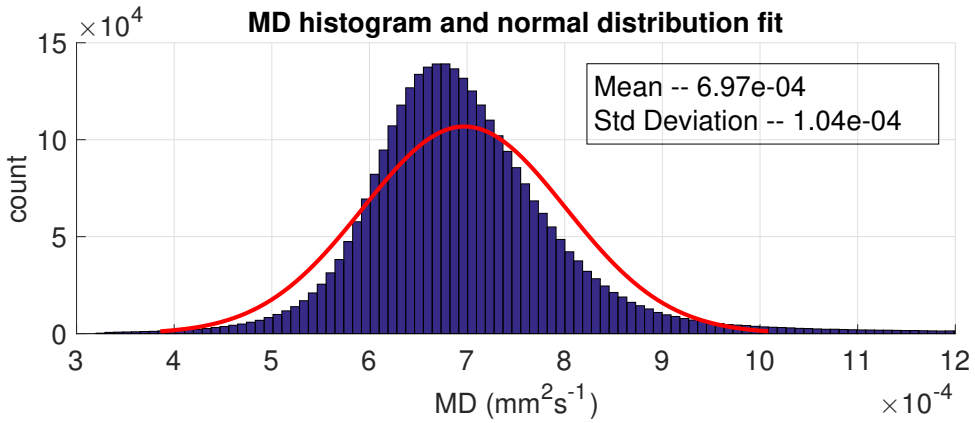


Figure 3.1: MD histogram obtained from the Whitehall dataset. The histogram shows the MD values obtained within the WM skeleton thresholded to $FA \geq 0.3$. In red, we can see the Gaussian distribution fitted on the data.

using a Gamma distribution for the MD prior might be more adequate, it would require an additional parameter that might not be consistent across datasets and it would also add computational burden to the model fitting procedure. The effect of the prior, and in particular its precision, was investigated through simulations, as shown in the next section.

Finally, the model was inferred on using an initial least-squares fitting, followed by a sampling of the posterior distribution on all model parameters. We used Markov Chain Monte-Carlo (MCMC) sampling with the Metropolis-Hastings algorithm, with a burn-in period of 1000 jumps, a sampling period of 1250 jumps, and an inter-sample interval of 25 jumps to allow for good mixing. The details of the sampling algorithm can be found in [Behrens et al., 2003], where a similar method was used to fit the Ball and Sticks model parameters. The algorithm samples directly from the posterior distribution on all the model parameters. The parameter prediction is given by the mean of the distribution while its uncertainty is reflected by the standard deviation.

Unlike the work in [Pierpaoli and Jones, 2004] and [Pasternak et al., 2009] we do not use a volume fraction parameter combined with a single baseline signal. Instead, we use parameters for the unattenuated signals from each compartment separately (S_F and S_T) with no explicit volume fraction. Although both formalisms lead to

the same fits (they are equivalent up to re-parametrisation), we chose the above parameters because using volume fractions assumes equal T_1 and T_2 relaxation times for the two compartments, and therefore the resulting volume fractions may have significant biases. We hence infer the free water volume fraction using correction factors that account for T_1 and T_2 relaxation (assuming pulsed-gradient spin echo measurements). This corrected value of free water fraction is given by:

$$FWF = \frac{\hat{S}_{fw}}{\hat{S}_{fw} + \hat{S}_T}, \tag{3.2}$$

where $S_X = \hat{S}_X(1 - \exp[-\mathrm{TR}/T_{1,X}])\exp[-\mathrm{TE}/T_{2,X}]$. In this formula, we fix the free water and tissue relaxation times to their typically measured values at 3T: $(T_{1,fw}, T_{2,fw}) = (3000, 1000)$ms, and $(T_{1,T}, T_{2,T}) = (800, 80)$ms [Wansapura et al., 1999].

Regarding the implementation of the fitting, the robustness of MCMC sampling for the estimation of our model parameters comes at the cost of a high computation time. Fitting the FW model on a standard DWI acquisition takes on average 40 core hours with our C++ implementation, whereas the FWE method [Pasternak et al., 2009] takes 15 minutes. Parallelization of the code, remarkably using GPU could reduce the time to a few minutes, in the same way that the ball and stick model fitting is implemented on GPU [Hernández et al., 2013].

3.2.2 Simulations

The goal of the simulations was on the one hand to test the identifiability of the model, and on the other hand to assess the added value of acquiring a small b-value shell besides the conventional $b = 1000s.mm^{-2}$ shell (on top of the b=0 baseline signal). Simulations were implemented in Matlab (Mathworks copyright)

and the Metropolis-Hastings algorithm was implemented in C++. We assessed the model accuracy at various levels of FWF, with 500 generated tensors for each set of simulations. To ensure realistic FA and mode (MO) [Ennis and Kindlmann, 2006] distributions of the tensors, we selected the tensors in deep white matter with an Inversion Recovery DWI acquisition. The tensors were then sampled with a normal MD distribution with $(\mu_{\mathrm{MD}}, \sigma_{\mathrm{MD}})$ =(0.7,0.14) $\times 10^{-3} mm^2.s^{-1}$. The diffusion weighted signal was simulated according to Equation (3.1). Rician noise was added to the signal, with the SNR of 30 defined in relation to the mean $b = 0 s.mm^{-2}$ signal. Additionally, we assessed the behaviour of the model fitting for a SNR ranging from 10 to 40. We simulated 72 directions for the single-shell case (with b=1000$s.mm^{-2}$), and 12+60 directions for the two-shell data (with b=300, 1000 $s.mm^{-2}$). The goodness of the fit was evaluated in terms of mean difference and root mean squared error (RMSE) between the true and estimated FA, MD and FWF.

To assess the impact of the MD prior on the MD estimates, we performed simulations where FA was considered known and the full posterior distribution of all other parameters was calculated using a grid search approach. The posterior was then marginalised over all parameters excluding MD to obtain a marginal posterior distribution of MD for each FA value (30 values ranging from zero to one). The true MD was set away from the prior and with a value of $1 \times 10^{-3} mm^2 s^{-1}$, where the prior mean value was $0.7 \cdot 10^{-3} mm^2.s^{-1}$.

3.2.3 Validation on in-vivo data

We carried out the in-vivo experiments on 4 diffusion acquisition schemes that are summarised in Table 3.1. The acquisition parameters were carefully chosen to fit the needs of the FW model validation and the data was acquired specifically for our experiments.

3.2.3.1 Free water at the tissues interface

We added an Inversion-Recovery (IR) pulse to a diffusion sequence to validate the free water model on real data. The inversion time was chosen such that the free water signal from the whole brain was nulled (assuming that $T1_{FW}$ = 3000ms on average). This provided a measure of the true underlying tensor in the absence of partial voluming with free water, which can be compared to the tensor estimated from our model on standard PGSE data with no inversion pulse.

The model was fitted to a single-shell DWI acquisition as well as to the same DWI acquisition with an additional shell with a low b-value (b=300$s.mm^{-2}$). The sequence parameters for the standard DWI (DTI) acquisition were TR=9600ms, TE=76ms and b=1000$s.mm^{-2}$, with 60 non-collinear diffusion gradient directions in addition to 6 b=0 images. For the two-shell acquisition, we added 12 directions weighted with a b-value of 300$s.mm^{-2}$ to the single-shell acquisition. The voxel size was 2mm isotropic, for a total scan time of 12 minutes. The IR sequence parameters (IR-DTI) were TR=20.6s, TI=2300ms, b=1000$s.mm^{-2}$, with 60 non-collinear diffusion gradient directions in addition to 6 b=0 images. The resolution was 2mm isotropic with a scan time of 25 minutes. A structural T_1-weighted image was also acquired with a resolution of 1mm isotropic.

Image pre-processing consisted of a correction for EPI distortions using TOPUP [Andersson et al., 2003] and a correction for movement and eddy current distortions using EDDY [Andersson and Sotiropoulos, 2016]. The model fitting was performed using our Metropolis-Hastings procedure with one compartment for the IR-DWI, and one and two compartments for the standard DWI acquisition. FA maps from the IR-DWI and the standard DWI acquisitions were co-registered to halfway space using FLIRT, so that the same amount of spline interpolation was applied to each image. The structural T1 image was segmented with FAST [Zhang et al., 2001] in

order to locate the voxels affected by macro-structural partial voluming and then registered to the FA maps using FLIRT boundary based registration (BBR) [Greve and Fischl, 2009].

3.2.3.2 Free water in deep white matter

In order to assess the properties of the FW compartment both at tissue interface and in deep white matter, two additional experiments were performed. In the first experiment, an IR-DWI sequence with multiple inversion times (TIs) was employed to estimate the spin-lattice relaxation time (T1) of the FW compartment (IR-DTI 9TIs). The IR-DWI sequence consisted of 9 acquisitions with the TIs uniformly sampled between 900 and 2400ms. We used a diffusion scheme optimized for the FW model fitting [Hoy et al., 2014], consisting of 4 non-weighted B0 and 28 DWIs (b=500 and 1500 $s.mm^{-2}$) with respectively 8 and 20 non-collinear diffusion gradient directions. The sequence parameters were TR=12900ms and TE=62ms, with an isotropic resolution of 2mm, for a total scan time of 80 minutes. The FW model equation for the IR-DWI acquisition is derived from Equation (3.1), with S_{fw} and S_t weighted by their inversion recovery relaxation time:

$$\begin{aligned} S_i = S_{fw}\left(1 - 2exp\left(-\frac{TI}{T1_{fw}}\right) + exp\left(\frac{TR}{T1_{fw}}\right)\right) exp(-b_i D_{fw}) \\ + S_t\left(1 - 2exp\left(-\frac{TI}{T1_t}\right) + exp\left(\frac{TR}{T1_t}\right)\right) exp(-b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i) \end{aligned} \tag{3.3}$$

$T1_{fw}$, $T1_t$ are the relaxation time of the FW and tissue, TI is the inversion time and TR the repetition time. We have added a term accounting for the relaxation time because a TR equal to 12900ms does not allow for a full recovery of the longitudinal magnetization.

In the second experiment, we acquired a DWI with 10 shells with the b-values ranging from $100 s.mm^{-2}$ to $1000 s.mm^{-2}$ and 26 non-collinear diffusion gradient

directions for each shell. The sequence parameters were TR=10100ms and TE=72ms, with a resolution of 2mm isotropic, for a total scan time of 50 minutes. The dense range of b-values allowed an estimation of the diffusivity in the FW compartment. We were also interested in precisely estimating the FW fraction in deep white matter, and assessing its relevance in deep white matter. To achieve this, we applied the automatic relevance determination (ARD) method to the FW and tissue compartment. This method ensures that the parameters subject to ARD take the value zero if they do not contribute to the likelihood [Behrens et al., 2007].

Dataset	**b-values**	**# directions**	**TI (ms)**
DTI	300,1000	12,60	N/A
IR-DTI	1000	60	2300
IR-DTI 9TIs	500,1500	8,20	900:2400
DTI 10 shells	100:1000	20*10	N/A

Table 3.1: Acquisition parameters for the four acquisition schemes used in the experiments

3.2.4 Group study: Alzheimer's disease

In order to evaluate the effect of FW modelling on a typical group study, we performed a TBSS analysis using both a standard tensor fit and (separately) the free water model to estimate the tensor and the associated indices. We also made use of the T1 partial volume segmentation of the tissues to perform a voxel based morphometry analysis to quantify tissue volumetric changes across groups. The dataset was chosen for its relatively modest resolution of 2.5mm isotropic, which makes partial volume effects more problematic. It is notable though that this spatial resolution is fairly common in dMRI studies.

From the original study, described in [Douaud et al., 2011], we included 114 participants, of whom 61 were controls and 53 were probable AD patients without a

vascular component. DWI data were acquired on a 3T Siemens MR scanner using echo-planar imaging, with TR=7000ms, TE=89ms, resolution= 2.5mm isotropic, 30 non-collinear orientations with b-value=$900 s.mm^{-2}$ and 6 b=0 images. In order to increase the SNR, scanning was repeated twice and both scans were corrected for head motion and eddy currents using an affine registration before averaging. The acquisition protocol also included structural T_1-weighted images obtained with an MPRAGE sequence (TE=5.3ms, TI=1000ms, TR=2150 ms, at a resolution of 1.1mm isotropic).

The diffusion tensor was calculated using the standard linear fit as implemented in FDT, part of FSL [Smith et al., 2004], and with the free water model, using default parameters ($D_F = 3^{-3} mm^2.s^{-1}$ and priors $\mu_{\text{MD}}, \sigma_{\text{MD}}$=0.7,0.1 $\times 10^{-3} mm^2.s^{-1}$). The structural T_1-weighted images were segmented using FAST [Zhang et al., 2001]. The resulting WM partial volume estimates were then registered to diffusion images using FLIRT BBR [Greve and Fischl, 2009].

The ventricles of many of the participants were particularly large due to their relatively advanced age. We therefore created a study specific template using the method described in [Guimond et al., 2000] and in the Chapter 2. The registration was performed using an optimized version of FNIRT that improves the registration of individual tracts across subjects as demonstrated in Chapter 2 and in [de Groot et al., 2013]. The template creation and subject registration were performed on FA maps derived from the simple DTI model, denoted as FA_{DTI}. The obtained warps were applied to tensor images, and the tensors were reoriented with preservation of principal direction method, as described in [Alexander et al., 2001b]. We used the same warps to transform the WM partial volume estimates images, obtained from T_1-weighted images and modulated by the Jacobian determinant of the non-linear part of the warp.

Unlike in the standard TBSS pipeline, individual values of FA and other indices were not projected onto the skeleton as the improved registration removes the need for

this step, as shown in [de Groot et al., 2013]. The skeleton was simply used as a mask for comparison between control and patient groups. The TBSS skeleton was computed from the mean FA_{DTI} map with FA values thresholded to 0.2. The statistics were computed on the FA_{DTI}, and the MD_{DTI} estimated with standard DTI and the FA_{FW}, MD_{FW} and FWF estimated with our model, using permutations testing via the randomise tool [Winkler et al., 2014]. The significance threshold was set at $p \leq 0.05$, with family-wise error correction for multiple comparisons across voxels, and using threshold-free cluster enhancement (TFCE) [Smith and Nichols, 2009].

Finally, we performed a voxel based morphometry analysis (VBM) for both CSF and WM to locate the regions that were affected by CSF or WM volume changes. The T_1-weighted images were registered to the associated b=0 images in diffusion space using FLIRT BBR [Greve and Fischl, 2009] and segmented using FAST [Zhang et al., 2001] to obtain the CSF and WM partial volume segmentation in diffusion space. It is important to emphasize that the alignment between diffusion and structural images is imperfect, especially in the frontal lobe, because the diffusion images suffer from EPI distortions. Nevertheless, we considered that the VBM and the TBSS analysis were conducted in the same space, and the VBM statistics were performed within the TBSS skeleton.

3.3 Results

3.3.1 Simulations

Figure 3.2 shows the mean difference between the model estimates of FA and MD and the true underlying values for various FWF. The comparison includes a tensor fit with single-compartment and a two-compartment fit with our FW model

using one and two-shell simulated data, with and without a prior on MD. On the left, we can observe that the FA_{DTI} estimated with the simple tensor significantly drops with increasing FWF. In contrast, the FA estimated with the FW model (FA_{FW}) is closer to the ground truth, in both single and two-shell data, with or without prior. Looking at the single-shell fit, we can notice a significant reduction in the mean difference when a prior is used. For small values of FWF (< 0.1), the FA_{FW} is slightly overestimated with and without prior, while for values of FWF higher than 0.1, the FA_{FW} tends to be underestimated. When a prior is used, FA_{FW} slowly decreases with increasing FWF, while with no prior, the FA_{FW} quickly drops. In two-shell data, the added value of using a prior is not as substantial as in one-shell data. For values of FWF lower than 0.6, there is no appreciable difference between the estimated FA with and without prior. For values of FWF higher than 0.6 the FA without prior tends to drop more dramatically than with prior.

On the right hand side of Figure 3.2, we can observe how the different models and data affect the MD estimation. In the single-compartment fit, the MD_{DTI} linearly increases with the FWF, from 0 up to 1.7 $\times 10^{-3} mm^2.s^{-1}$. The MD_{FW}, estimated by applying our FW model, remains closer to the ground truth for single and two-shell data, with or without prior. When the prior is used, then the mean difference is close to 0. The MD_{FW} is increasingly overestimated when the prior is not set, especially in the single-shell data. With two shells and no prior, the MD_{FW} is correctly recovered for a FWF up to 0.6.

In the Figure 3.3, we plot the root mean squared error (RMSE) between the FW model predictions and the simulated ground truth, with the noise level varying from 10 to 40. Plotting the RMSE highlights the precision of the estimates, as opposed to Figure 3.2, where we focused on the bias.

Looking at the FA, we first observe that the noise level has a limited impact on the precision of the estimates, for a FWF < 0.5. For higher FWF, the error is marginally higher with low SNR. Interestingly, we see that the addition of the second

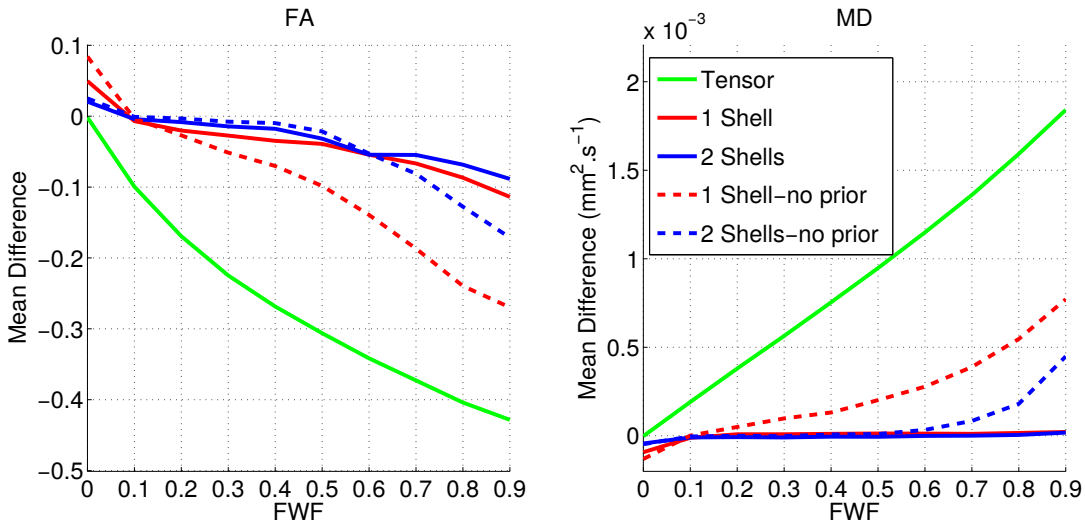


Figure 3.2: FA (left) and MD (right) mean difference with simulated ground. The difference is expressed as a function of FWF. We show the results with tensor model (green) and FW model with one (red) and two shells (blue), with (solid) and without prior (dashed)

shell ($b = 300s.mm^{-2}$) is mostly beneficial when the SNR is ≥ 30. Conversely, the use of the prior on the MD is mostly beneficial when the SNR is ≤ 20. We also note that the precision is comparable on the 2 shells data whether or not the prior is used when the SNR is ≥ 30 and the FWF is < 0.8.

Similarly, the precision of the MD estimates is marginally impacted by the SNR, and the prior has a larger influence when the SNR is < 30 and when only one shell is acquired.

Figure 3.4 shows the estimated FA and MD as a function of the true FA and MD when fitting the model to the single or two-shell data, for FWF ranging from 0 to 0.7. This plot allows to visualise the effect of the prior on the FA and MD estimated by the FW model. We can observe that the precision of the FA_{FW} decreases with the FWF, whereas with the single tensor the dispersion remains low, at the cost of a higher bias. Using two-shell data increases the precision of the estimated FA_{FW}.

The estimated MD shows a different pattern. As the FWF increases, the slope of the MD_{DTI} remains constant, whereas the slope of MD_{FW} gradually becomes

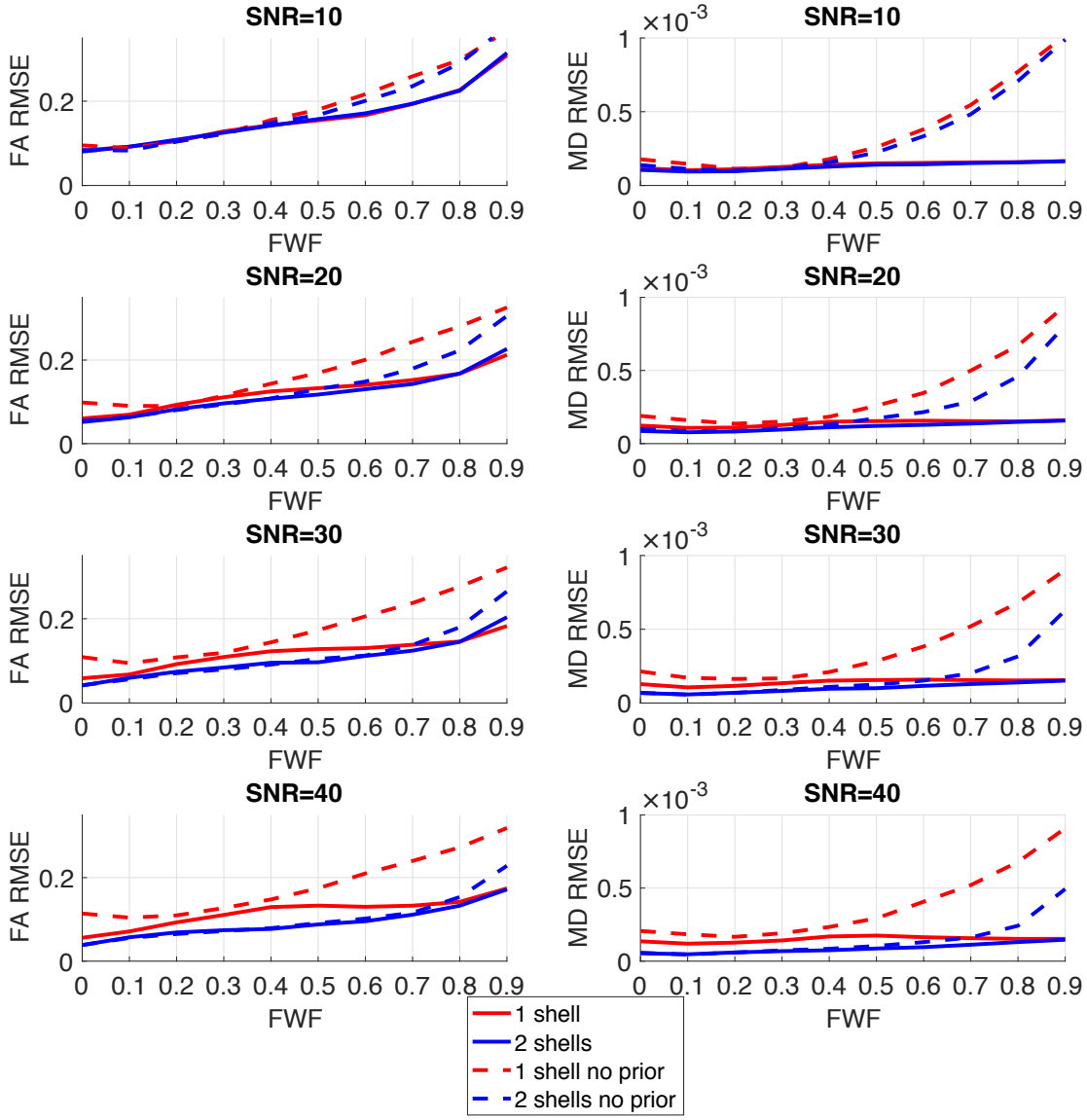


Figure 3.3: FA (left) and MD (right) root mean squared error with simulated ground truth. The difference is expressed as a function of FWF. We show the results with the FW model with one (red) and two shells (blue), with (solid) and without prior (dashed). The results are shown for four SNR levels, ranging from 10 to 40.

close to 0. The benefit of a second shell is noticeable on the MD_{FW} estimates as the slope remains closer to the ground truth than in the single shell case for all the examined FWF values. This suggests that as the FWF increases, the MD estimation will be mostly driven by the prior. The effect of the prior is also visible when the FWF is equal to 0. Indeed, in the single shell case, we can see that the MD_{FW} is

underestimated when the true MD is high, whereas this effect is no longer present in the two shells case.

Figure 3.5 shows the root mean squared error (RMSE) of the model parameter estimates (FA, MD and FWF) as a function of FA and FWF, with one and two shell data. Using the RMSE rather than the mean error allows to visualize the precision of our estimates, as it relates to the slope and dispersion of the estimates in Figure 3.4.

The top row of Figure 3.5 shows that the FA_{FW} estimation is more accurate for high values of true FA (top row). The benefit of the second shell appears when both the FA and the FWF are low < 0.4. With the single shell data, the error reaches the highest value for high values of FWF and FA (FWF> 0.7 and FA< 0.4). In both the cases, the RMSE of FA_{FW} decreases when the true underlying FA increases.

The error in MD_{FW} estimation appears to be high (10^{-4}) with the single shell data for any value of FWF when the FA is low, with a peak around FWF=0.2. Nevertheless, it decreases with increasing FA. With the two shells data, the MD_{FW} estimation is more accurate, and it does not exhibit the peak for low FWF(< 0.2) that was observed on single shell data.

Estimation of FWF appears to be inaccurate for low values of true FWF (< 0.3) and low FA, in the single shell case. As for the other estimates, the RMSE decreases with increasing FA, in both the single and two-shell case. The error distributions of MD_{FW} and FWF appears very similar. This result was expected, as the MD and FWF were suspected to be the most correlated parameters of the model. This correlation was confirmed by observing the samples of the FWF and MD parameters generated with the Metropolis-Hastings fitting method.

Overall, the model is accurate for the FA_{FW} estimation when the FWF is below 0.9. The MD_{FW} estimation is less accurate in the single shell case for all the FWF range, but the RMSE decreases with increasing true FA. FWF estimation accuracy follows the same pattern as the MD_{FW}. This means that an error in MD estimation

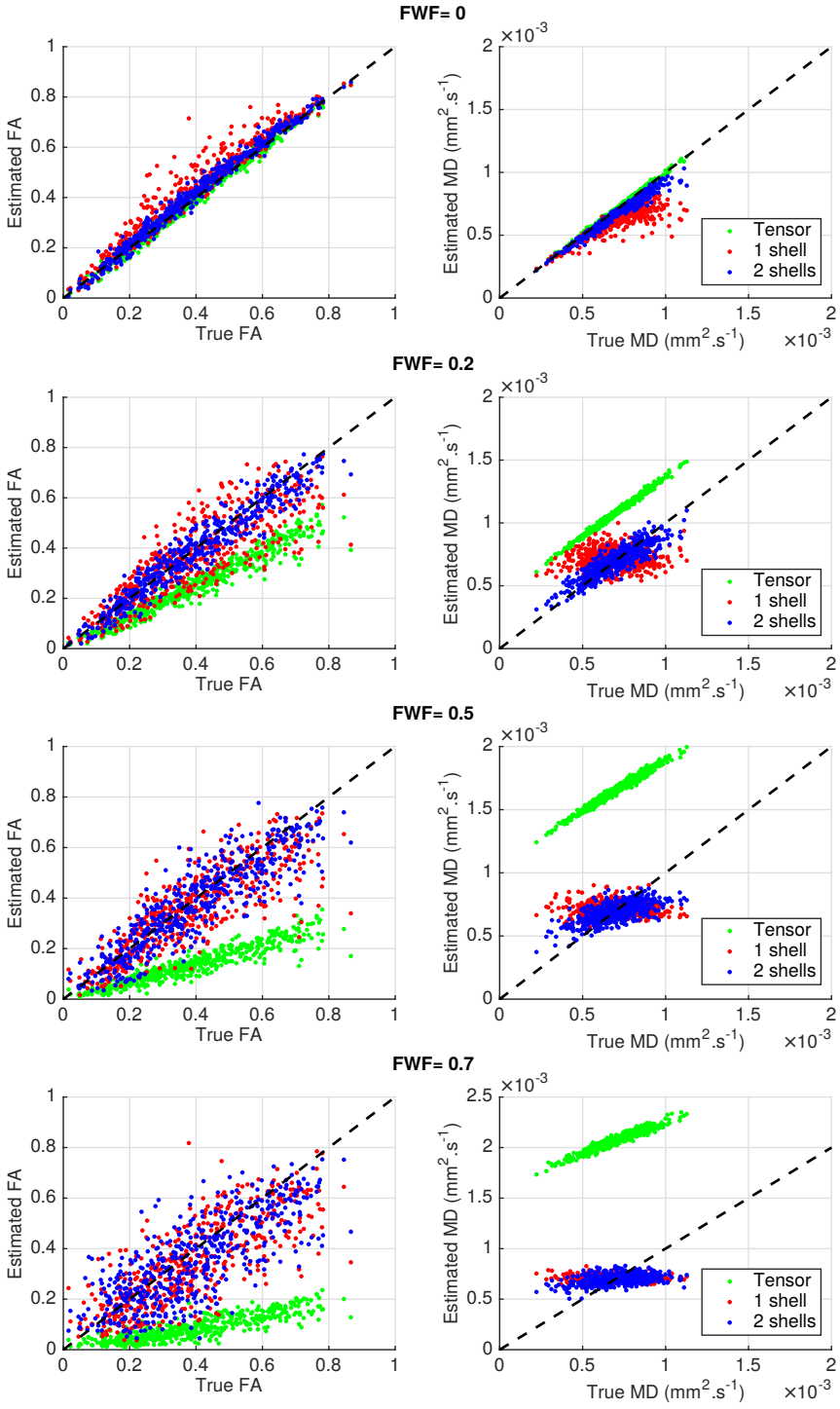


Figure 3.4: Estimated FA and MD values against the true values. Results obtained using single-compartment tensor estimation (green) and FW model tensor estimation with one (red) and two shells (blue), with a prior on the MD, and with a FWF ranging from 0 to 0.7.

will translate into an error in FWF estimation. FWF and MD_{FW} estimations will therefore be increasingly biased by the prior on MD for decreasing FA. Adding a second shell greatly diminishes this bias, and allows a good estimation of the MD_{FW} and the FWF.

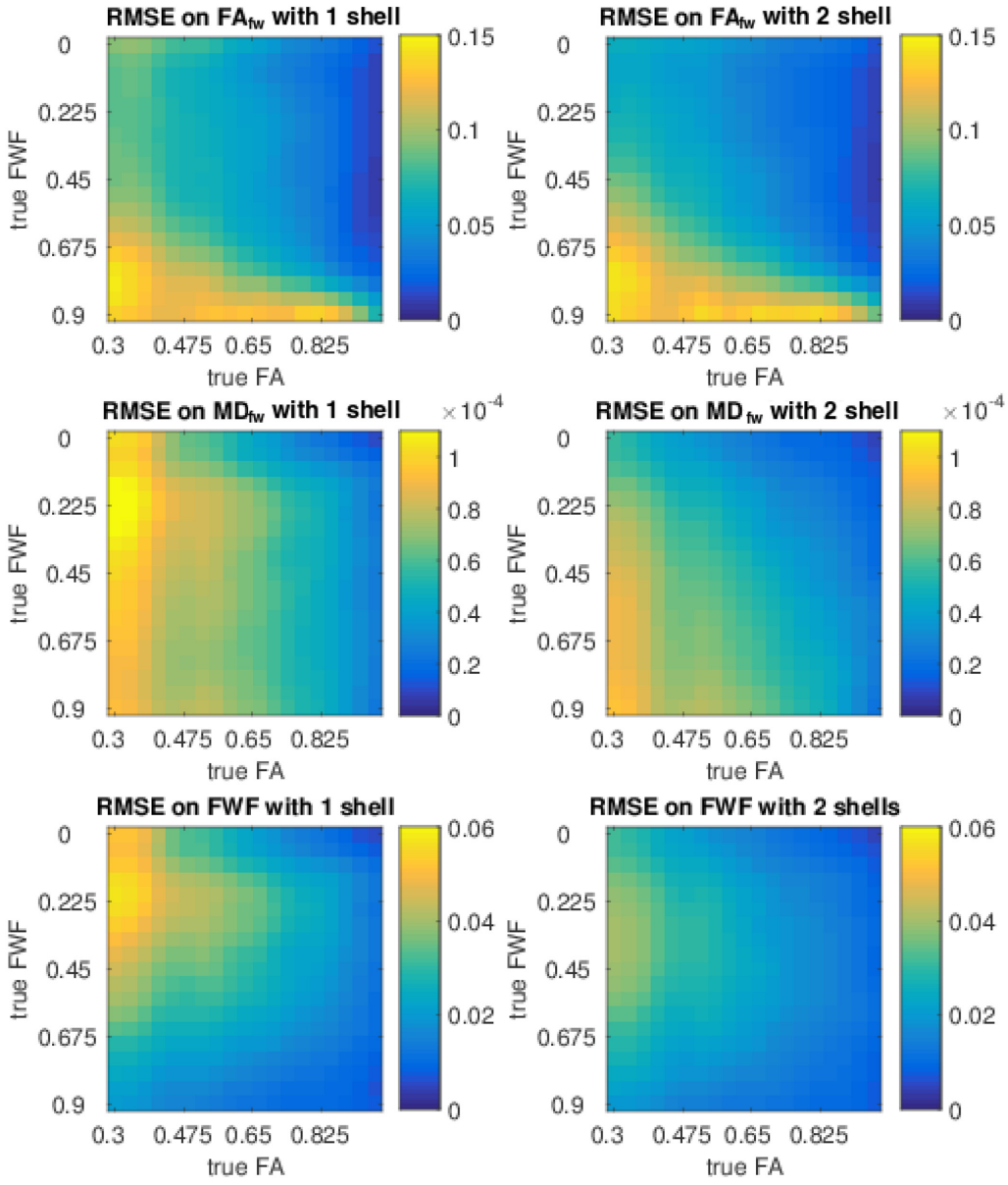


Figure 3.5: Root Mean Squared Error for FA_{FW}, MD_{FW} and FWF estimate on single and 2 shells data. The results are obtained with the prior set on the MD.

To asses further how the prior on MD affects the recovery of the true underlying

MD, we present in Figure 3.6 the posterior distribution of MD estimates for any fixed FA value. In the single-shell data, high standard deviation (i.e., uncertainty) on the prior gives estimates that are not biased, i.e. the mean of the MD posterior is close to the true MD, but the uncertainty is very high, especially at low FA values. As the standard deviation of the MD prior ($\sigma_{MDprior}$) decreases, the uncertainty decreases, but the bias increases. The effect is particularly evident with the lowest $\sigma_{MDprior}$, as the uncertainty remains almost constant, the mean of the MD posterior is equal to the prior with low FA, and gradually approaches the true MD when FA increases. In other words, the model is more identifiable at higher anisotropy. With the two-shell data, the uncertainty remains low, even with a high value for $\sigma_{MDprior}$. The bias gradually increases as $\sigma_{MDprior}$ decreases, but it remains lower than in the single-shell case. The benefit of the second shell is that the model identifiability remains good over a wide range of anisotropies, even when the weight of the prior decreases, leading to relatively unbiased MD_{FW} estimation. This result is in agreement with the behaviour of our model in Figure 3.4, where MD estimation is more precise and less biased with the second shell. It also supports the results in Figure 3.5, where the MD_{FW} estimation appears to be more accurate for high FA values.

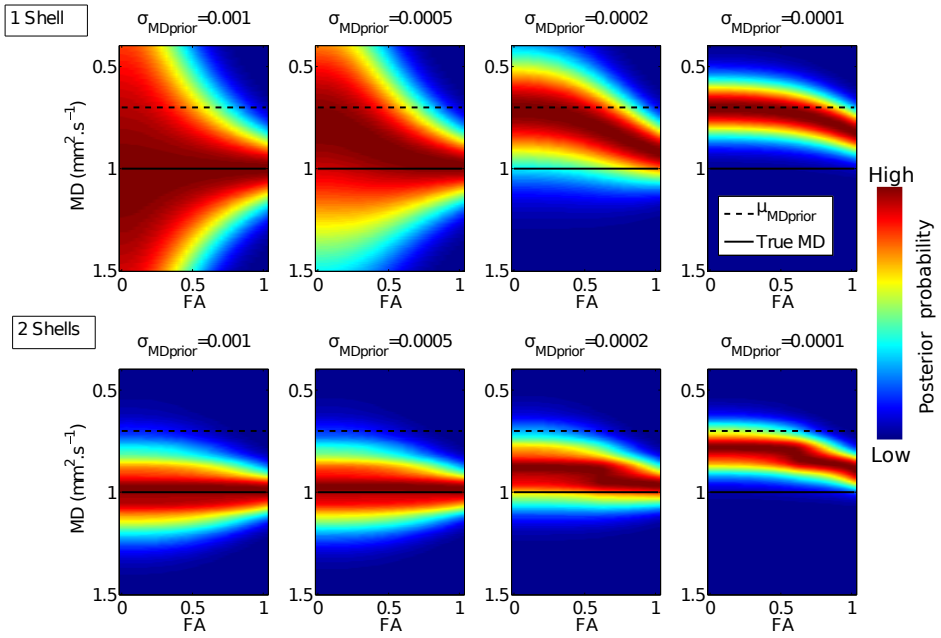


Figure 3.6: Posterior distributions of MD (given by each column) plotted for different values of true FA. We show the results for several different standard deviations on the MD prior ($\sigma_{MDprior} = 0.001 \sim 0.0001 s.mm^{-2}$) and with one or two shells with FWF set to 0.5. The dotted line represents the value of the mean MD prior ($\mu_{MDprior}$) and the solid line represents the true MD.

3.3.2 Validation on in-vivo data

3.3.2.1 Free water at the tissues interface

In the first part of this section, we show how the FW model can retrieve the unbiased FA and MD in voxels at the interface of the CSF and white matter. We compare our method and Pasternak's FWE to an inversion recovery acquisition where the CSF signal was nulled.

Figure 3.7 (top row) shows an axial view of the FA maps obtained with standard

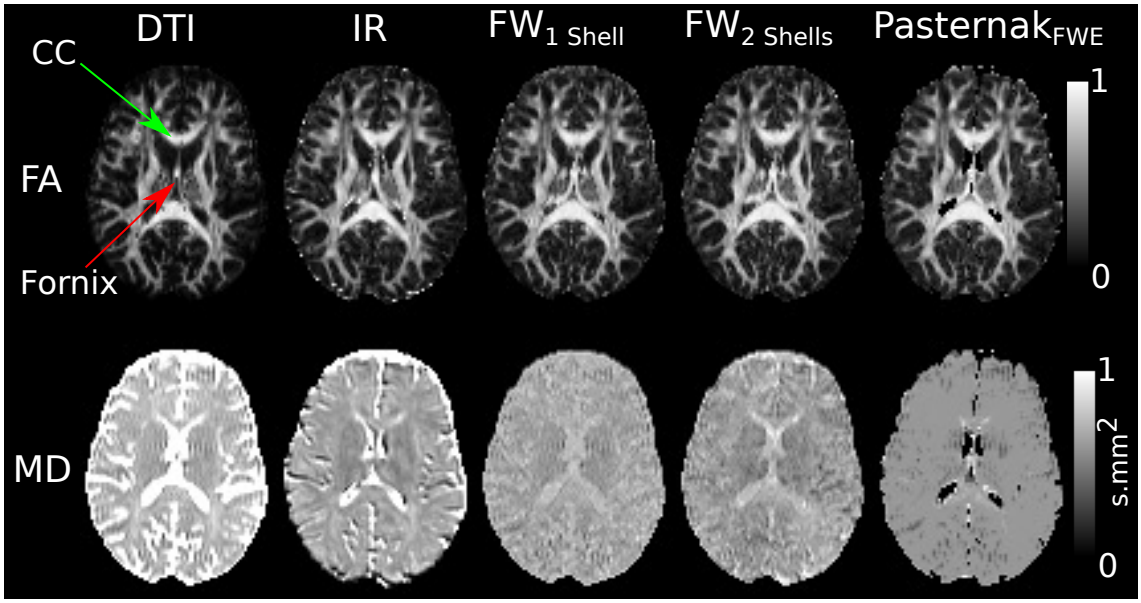


Figure 3.7: FA and MD maps for DTI, IR, our model and FWE method. FA (top row) and MD (bottom row) maps obtained with the simple tensor model, IR acquisition, FW model with one and two shells and FWE method with one shell. The green arrow indicates the corpus callosum and the red arrow shows the fornix. Both tracts are highly affected by partial volume effects.

DTI, IR, FW model (with one and two shells) and the FWE method (with one shell). Partial volume effects are particularly visible in the fornix, as it is thinner and has lower FA values with the standard DTI compared to IR. Generally, the FA obtained with the FW model and the FWE method is similar to the one computed on the IR data. For example, the fornix appears with a higher FA than with the DTI acquisition, and the corpus callosum is broader.

Looking at the MD maps in Figure 3.7 (bottom row), we can see that the MD in the middle of the ventricle is high on the standard DTI map, with a value of 0.0034 mm^2s^{-1}, close to the theoretical diffusivity of free water at body temperature (0.003 mm^2s^{-1}). Surprisingly, it also appears to be high in the ventricles on the IR acquisition. Compared with non IR acquisition we measured a CSF signal drop of 90% on average in the ventricles. In the meantime, the MD was only reduced from 0.0034 mm^2s^{-1} to 0.0020 mm^2s^{-1}, suggesting that MD is still sensitive to the CSF signal remaining in the IR acquisition. However, in areas where partial volume

effects are expected, i.e. between WM and the ventricles, the MD in the IR image is comparable to that in the tissues. With our FW model the MD follows the prior in the ventricles, because of the very small amount of tissue present. In the MD map computed with the FWE method, the MD is set to 0 in the middle of the ventricles. Interestingly in partial volume areas the MD is higher with the FWE method than with ours, the latter giving values similar to the MD in tissues. This suggests that some CSF contamination is still occurring with the FWE method.

In areas where partial voluming is not expected to occur (i.e. in deep white matter), the MD appears very homogeneous with the FWE method, possibly reflecting the effect of the spatial regularization. With the FW model, the MD appears less homogeneous in both single and two-shell case, and some structural details are visible, notably with lower MD values in the internal capsule and higher values in the genu of the corpus callosum. Also, grey matter has high MD values in both IR and FW model fits, whereas there is no contrast in the FWE MD map. These contrasts and structures are more visible in the FW model fit with two-shell than in the single-shell case. This is in agreement with our simulation results that indicated more accurate MD estimates in the two shell case.

Figure 3.8 illustrates the histograms of the FA and MD estimated within the WM (including partial volume areas), delineated by the co-registered segmentation obtained with FAST on the corresponding T1w images [Zhang et al., 2001].

The FA histogram peaks just over 0.4 for FA_{DTI}, whereas the FA_{IR} mode is located around 0.5. The histogram obtained with our FW model on single shell data shows good correspondence with the IR one. Both our model on 2 shell data and FWE method histograms are slightly shifted towards higher FA values, therefore estimating a higher FA than IR on average.

The MD_{DTI} histogram is shifted towards higher values compared to MD_{IR}, and the modes are respectively $7 \cdot 10^{-4}$ and $6 \cdot 10^{-4} mm^2.s^{-1}$. The MD histogram estimated with our model on the single shell data is located between the IR and

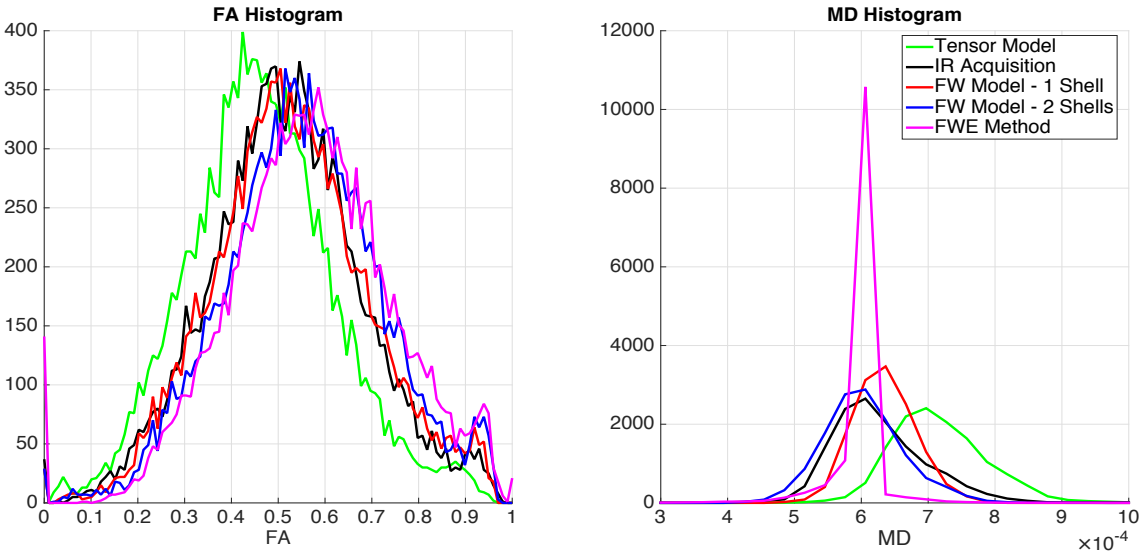


Figure 3.8: FA and MD histograms for DTI, IR, our model and FWE method in the WM. Left: FA histogram with tensor fit on inversion recovery (green) and standard acquisition (black), with our FW model on one (red) and two shells data (blue) and FWE method (pink) - Right: Corresponding MD histograms

single-compartment, with a flat mode around $6.4 \cdot 10^{-4}$ that is shifted towards the MD prior ($7 \cdot 10^{-4}$). The mode and shape of the MD_{FW} histogram obtained with our method on the 2 shells data presents similar shape and peak as the MD_{IR} histogram. The MD_{FWE} histogram presents a very different profile as the mode is located at $6 \cdot 10^{-4}$, with a very narrow distribution of values, in comparison to the four other histograms.

Quantitative comparisons between DTI, our FW model and the FWE method are presented in Figure 3.9. The voxels represented in the figure are located at the interface of the CSF and the white matter, as defined by the T1 segmentation (WM PVF>0.5 and CSF PVF>0). The IR-DTI acquisition is considered as a ground truth without CSF contamination.

On the left column, we see how the FA estimated by the different models compares with the FA estimated on the IR acquisition. In the DTI case, the correlation in FA between DTI and IR-DTI is high (r=0.74), but we can see that it is underestimated for any FA value, in accordance with the simulation results in Figure 3.4. That is reflected by the root mean squared error (RMSE) which is equal

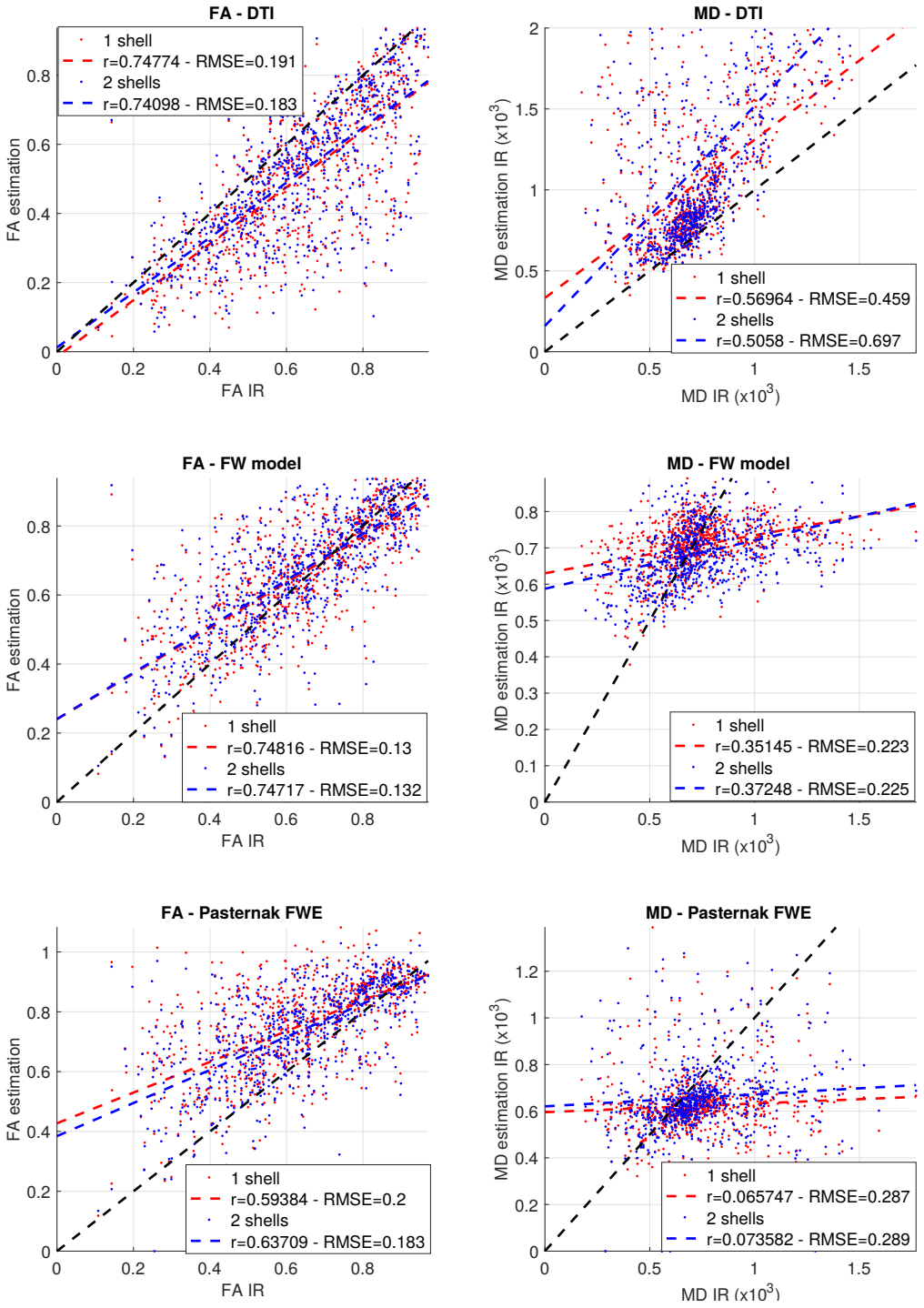


Figure 3.9: Scatter plot for the FA and MD obtained with DTI, our FW model and FWE method. Left: FA estimated by simple DTI model (Top), our FW method (Middle), and the FWE method (Bottom), with one (red) and two (blue) shells. Right: Results for the MD.

to 0.19 and 0.18 for single and two shells data respectively. With our method, the correlation in FA is also high (r=0.75), and the RMSE is significantly lower (0.13)

than with the simple DTI model. Interestingly, the additional shell does not give better results regarding FA estimation with our method. With the FWE method, the correlation in FA is lower than with our method (0.59 and 0.64 for single and two-shells respectively), and the RMSE is roughly equal to the DTI case (0.20 and 0.18). Both our method and FWE tend to overestimate FA, with a higher bias with the FWE approach.

Regarding the MD estimation (right column in Figure 3.9), we see that the simple DTI correlates well with IR, but it is biased toward higher MD values. As a result, the RMSE is high for both single and two shells case (respectively 4.6 and $7\times10^{-4}mm^2.s^{-1}$). The MD estimated by our model does not correlate as well as the MD estimated by the simple DTI (r=0.56 and 0.51 for DTI and r=0.35 and 0.37 for the FW model). Nevertheless, it is significantly less biased and the RMSE is 2 to 3 times lower than the with the simple DTI. We also note that fitting our model on two shells data gives less biased estimates than on single shell data, as the slope of the least-square fit is steeper than with one shell. This results in a slightly higher correlation with two shells and a wider range of estimated MD values. This indicates that when there is additional data (i.e. second shell) to support the fitting, the MD prior has less importance during the fitting procedure, in accordance with the simulation results in Figure 3.6. Looking as the MD estimated by Pasternak's FWE method, we see that although the RMSE in MD is lower than in the DTI case, it is worse than the one obtained with our approach. More importantly, the correlation between FWE and IR-DTI is almost zero, both in one and two shells case. This is in line with the very homogeneous FWE MD maps shown in Figure 3.7.

The Bland and Altman plots in Figure 3.10 illustrate the difference between the DTI, FW and FWE estimated on the single shell DTI and the IR-DTI acquisition. The plots represents the same voxels as in Figure 3.9. We observe that the FA is underestimated in the DTI case and that the bias increases with the FA. Conversely, the FA is overestimated with our method and the FWE, and the bias appears higher

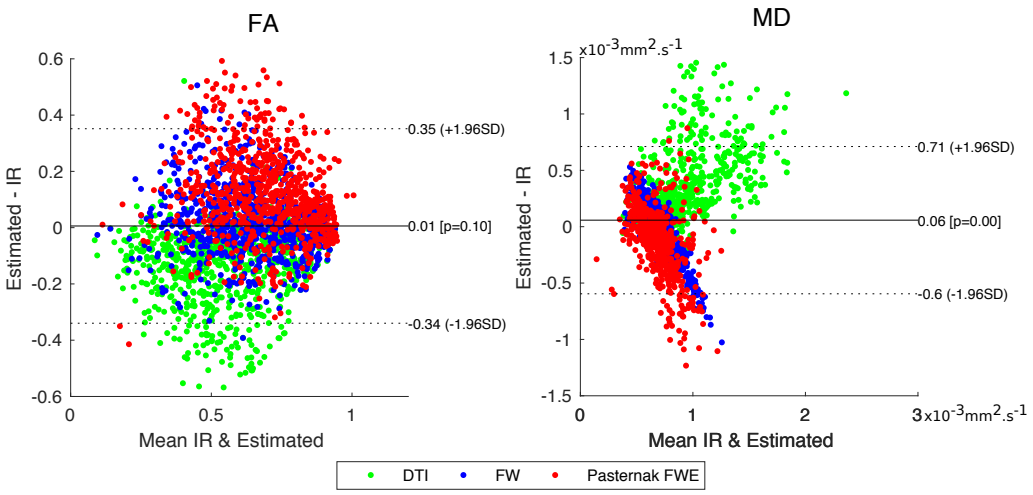


Figure 3.10: Bland and Altman plots for the FA and MD obtained with DTI, our FW model and FWE method Left: FA estimated by simple DTI model (Green), our FW method (Blue), and the FWE method (Red), with one shell. Right: Results for the MD.

with the latter. Looking at the MD, we observe that it is overestimated in the DTI case, with an increasing bias for higher MD values. The bias with the FW and FWE methods is negatively correlated with the MD. We note that the MD estimation with the FW and FWE methods is the most accurate for lower MD values ($\leq 0.7 \times 10^{-3} mm^2.s^{-1}$.)

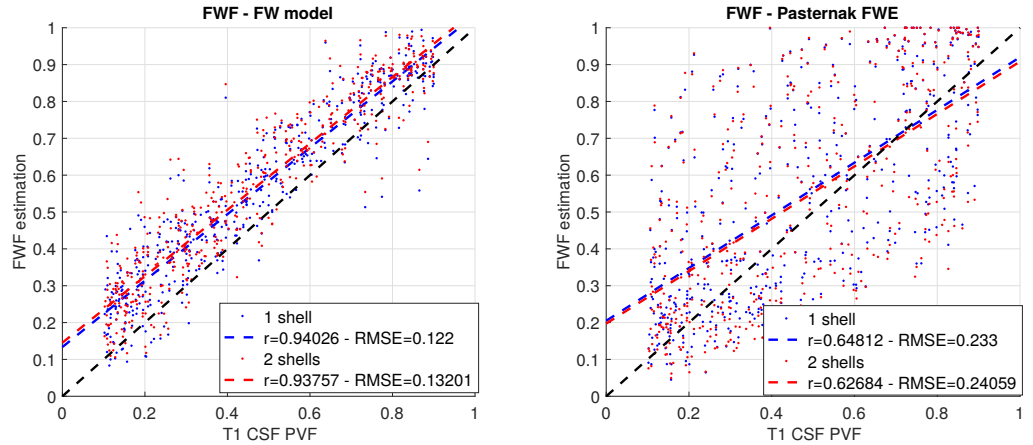


Figure 3.11: FWF estimated by our model and FWE method compared to T1 CSF PVE Left: FWF estimated on DWI with one (blue) and two (red) shells compared to the CSF volume fraction estimated from the T1. Right: Results with the FWE method.

In Figure 3.11, we see how the FWF estimates correlate with the CSF partial

volume estimates obtained from the T1 segmentation. On the left, we see that our model gives a high correlation (r=0.94) with both single and two shell data. We observe a linear relationship between our FWF estimates and the T1 CSF partial volume estimates. However, it appears that our model overestimates the FWF compared to the T1 CSF partial volume segmentation. The FWE method FWF estimates do not exhibit such a strong correlation with the T1 CSF PVE (r=0.64) and the RMSE is twice that of our method. Surprisingly, the addition of a second shell has a very little negative effect on the FWF estimation with both methods.

3.3.2.2 Free water in deep white matter

Our in-vivo results focused so far on voxels located at the interface of the WM and CSF (macroscopic partial voluming). In the second part of our in-vivo experiments, we investigate the properties of the FWF in deep white matter (microscopic partial voluming). It was recently hypothesised that the FW compartment in deep WM represents the ECS and that the tissue compartment (tensor) represents the intra-cellular space [Pasternak et al., 2012b]. We investigate the relevance of the FW compartment in deep WM by measuring some of its properties (T1 relaxation time, volume fraction, and diffusivity) and their spatial distribution.

The IR-DWI experiment with multiple inversion times and two shells allowed us to measure the T1 relaxation time and volume fraction of the WM and FW compartment, shown in Figure 3.12. We focused on the voxels in deep white matter defined by the eroded T1 WM segmentation, so that the selected voxels are not affected by neibourghing partial volume areas. We also included voxels in pure CSF for reference (in red). In deep white matter, the FW compartment is represented in green and the tissue compartment in blue. Our first observation is that the FW accounts for 16% on average ($\sigma = 4\%$) of the signal in deep white matter, which is in

line with the value previously reported in the literature [Pasternak et al., 2012a].

An important finding of this experiment is that the T1 of the FW in deep WM ($\mu = 770ms, \sigma = 303ms$) is not close to the T1 of the FW compartment for voxels at the interface of CSF and WM ($\mu = 2900ms, \sigma = 805ms$). Actually, the T1 of the FW in deep WM is very close to the T1 of the tissue compartment ($\mu = 830ms, \sigma = 54ms$). This suggests that either the FW and tissue compartments are not distinguishable, or that they have very similar spin-lattice relaxation times.

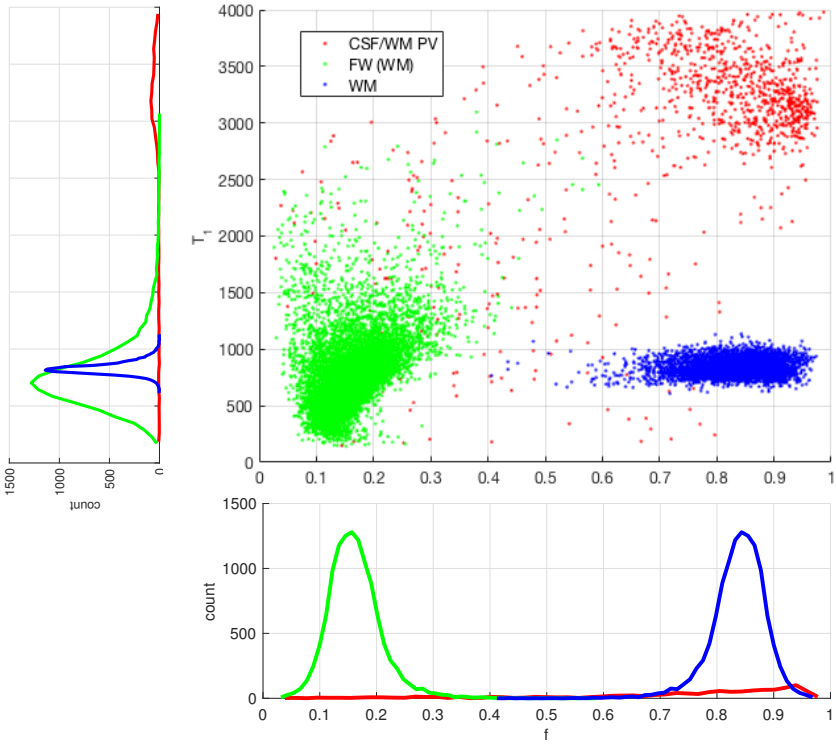


Figure 3.12: Volume fraction (f) and T1 relaxation time for the FW compartment Datapoints represent the voxels at the interface of CSF and WM (red), FW compartment in deep WM (Green) and tissue compartment in deep WM (Blue), with associated histograms.

In the DWI experiment with 10 shells, we explore further the properties of the FW compartment, namely its diffusivity, volume fraction and spatial distribution. The dense range of b-value allows for a good discrimination between slow and fast

diffusing components, so we were able to fit the FW model without placing a prior on MD. Additionally, the use of ARD allows to determine if the FW compartment is supported by the data. In Figure 3.13, we plot the diffusivity and volume fraction of the FW and tissue compartments (respectively in green and blue) in deep white matter and the FW compartment in the ventricles (red). The latter exhibits properties that are expected from the CSF in the ventricles, i.e. a fast diffusivity and a volume fraction close to 1.

Focusing on the FW compartment properties in voxels located in deep white matter, we can see that they form three clusters. The majority of the voxels (63%) have a FWF below 1%, and the diffusivity is in this case irrelevant as it just reflects the mean of the parameter space. These voxels are represented by the green peak in the histogram in Figure 3.13 (bottom), in orange on the diffusivity map Figure 3.13 (right), and in dark blue on the FWF map in Figure 3.13 (top). In the smallest cluster (5% of the voxels), the diffusivity of the FW is comparable with the tissue one, around 0.0006 mm^2s^{-1} and its volume fraction is about 0.3. These voxels are represented in red on the diffusivity map Figure 3.13 (right), and in light blue on the FWF map in Figure 3.13 (top). They are very likely to represent cases where the model fitting mistakenly assigns white matter fibres with low diffusivity to the FW compartment, because no prior on the FW diffusivity was set.

The voxels of most interest in Figure 3.13 are outlined with the red ellipse. They represent 32% of all the WM voxels and have a high diffusivity with a FWF of 18% on average. Their diffusivity is 3 $\times 10^{-3} mm^2s^{-1}$ on average, which what we expect from the free water. When we look at the spatial distribution of these voxels on the diffusivity and FWF maps, we see that they concentrate in regions where the fibre geometry is complex. In these crossing fibres areas, the diffusion tensor fails to capture the geometric complexity. Therefore, a FW compartment with a non-zero value is fitted to explain the residuals of a simple diffusion tensor. The FW compartment is therefore very unlikely to represent the extra-cellular space, as it

would mean that the FW is only present in crossing fibres areas.

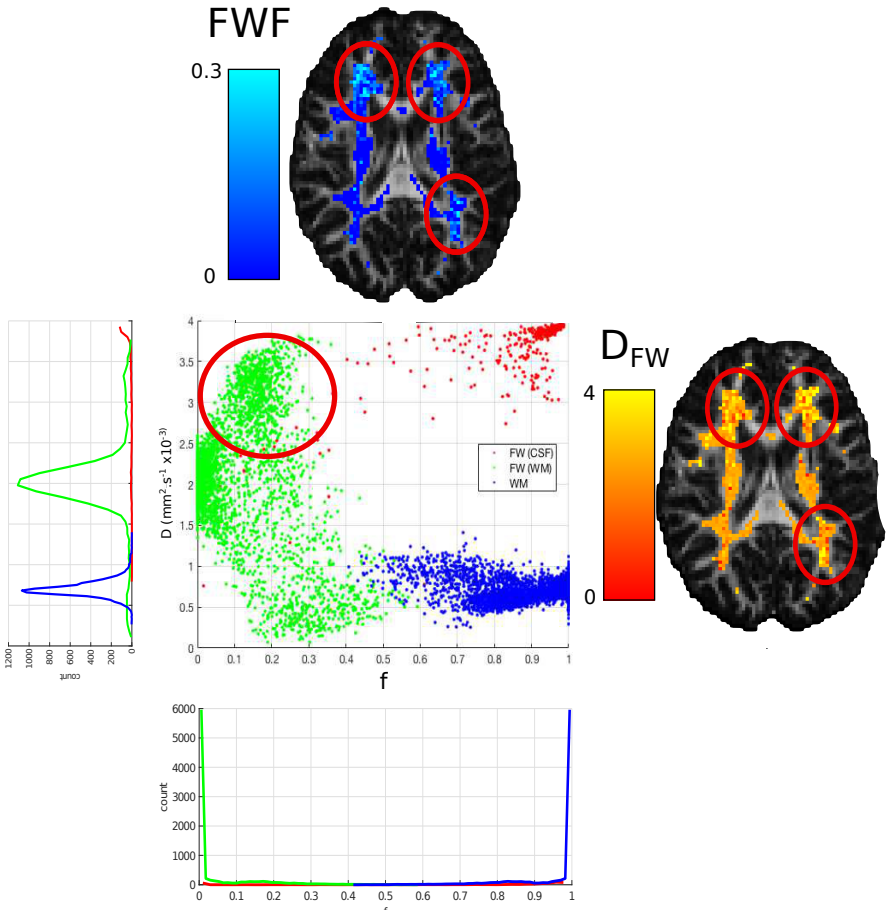


Figure 3.13: Volume fraction (f) and diffusivity (D) for the FW compartment. Datapoints are representing voxels in the ventricles (red), FW compartment in deep WM (Green) and tissue compartment in deep WM (Blue) with associated histograms. Top: FWF map overlayed on the FA map. Right: FW compartment diffusivity map overlayed on the FA map. The red ellipses show the voxels with high diffusivity and non-zero FWF.

Based on its spatial distribution, we conclude that the FW compartment in deep white matter does not represent the extra-cellular space. It appears that its presence is rather the result of the inability of the diffusion tensor to model crossing fibres appropriately and of the bias introduced by the MD prior.

3.3.3 Group study: Alzheimer's disease

The results from the TBSS analysis of Alzheimer's disease data with FA_{DTI} and FA_{FW} are presented in Figure 3.14 (top left row). The analysis shows widespread FA_{DTI} decrease in AD patients. The corpus callosum and the fornix appear to be impacted by the disease, with FA_{DTI} being significantly lower in patients for a large part of the two tracts. The fornix is located near the ventricles and is potentially impacted by CSF partial voluming. Other WM structures that are surrounded by GM also show significant decrease in FA_{DTI}, e.g. the inferior fronto-occipital fasciculus. We also observe a trend of increased FA_{DTI} in the centrum semi-ovale (not shown), where the cortico-spinal tract crosses the superior longitudinal fasciculus. This difference did not survive multiple comparison correction, which is consistent with a previous study performed on the same dataset [Douaud et al., 2011].

Results with FA_{FW} (top right) exhibit a very different pattern as almost none the decreases observed in FA_{DTI} are observed with FA_{FW}. We also see some significant decreases in FA_{FW} for the AD patients that were not present with FA_{DTI}, notably in the cerebellum (yellow ellipse) and the posterior thalamic radiations/forceps major (red ellipses). In these regions, there was a trend for decreased FA_{DTI} as well as, somewhat counter-intuitively, decreased MD_{DTI}. This suggests that our model might be more sensitive to micro-structural changes occurring in areas of partial voluming with grey matter. Interestingly, we also see a significant FA_{FW} increase in the extrnal capsule (green ellipse) in the AD patients, which corresponds to the area where FA_{DTI} shows a trend towards significant increase in AD. Again, our model shows more sensitivity to this alteration that can be explained by the selective degeneration of one of the two tracts crossing in this region (SLF compared with the relatively spared CST [Douaud et al., 2011]), thus increasing the observed FA. As stated in the in-vivo results section, these results have to be taken

with caution because the diffusion tensor is inappropriate to model the complex fibre geometry present in crossing fibres areas. This is particularly true in the centrum semi-ovale, which is at the intersection of the SLF and the CST.

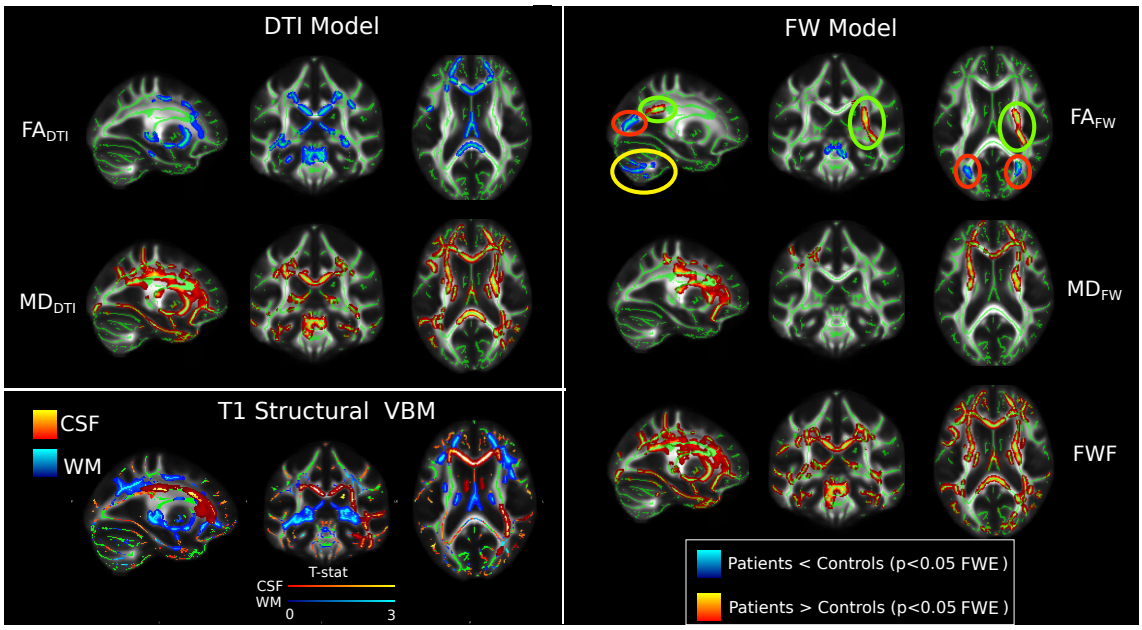


Figure 3.14: TBSS results obtained on with the simple DTI model (top left), the FW model (right), and T1 structural VBM (bottom left). The regions represented in blue exhibit a significant decrease in AD patients for the given parameter (FA, MD or FWF) and the red/yellow color represent a significant increase in AD patients ($p < 0.05$ corrected for multiple comparisons with FWE). On the bottom left, we show the results of a VBM analysis conducted in diffusion space, within the TBSS skeleton. The red/yellow region represent a significant increase in CSF volume and the blue regions represent a significant WM volume decrease in AD patients ($p < 0.05$ corrected for multiple comparisons with FWE). The thin colored lines represent the t-stat for CSF hypertrophy (red/yellow) and WM atrophy (blue).

We observe a widespread increase in MD_{DTI} in AD patients, partially overlapping with the areas of decreased FA_{DTI} in AD patients. Differences in MD_{FW} occur in the same areas, but mostly in the anterior part of the WM. It is worth noting that although our model found less widespread difference in MD_{FW} compared to MD_{DTI} between the two groups, we cannot necessarily attribute this result to FWF alone. As shown in our simulations, the fitting method applied to single shell data can suffer both from bias (towards the prior mean) and increased uncertainty in MD, particularly when anisotropy is low. Both of these effects can result in reduced

sensitivity to MD changes in low FA voxels. Therefore, the changes in MD observed only with the simple DTI model might appear in the FWF.

Indeed, we can see that the areas of FWF increase overlap with the regions where an increased MD_{DTI} or a decreased FA_{DTI} was observed. For instance, in the fornix the increased FWF in AD patients seems to explain the decreased FA_{DTI} and increased MD_{DTI}, which were likely caused by the contribution of FW.

Interestingly, there is also increased FWF in the centrum semi-ovale, where there was a significant increase in FA_{FW}. As we mentioned previously, the FW model is inadequate in such crossing fibres regions. In this particular area, we argue that the increased FWF in AD patient might reflect a micro-strcutural change in the secondary fibres, as it was also hypothesizes in [Douaud et al., 2011]. However, the isotropic nature of the FW compartment does not correctly model a single fibre. As a result, the diffusion tensor indices modelling the primary fibre will be affected by the inadequate FW compartment, and its FA_{FW} will very likely be overestimated.

In Figure 3.14, we can observe the results of the VBM analysis conducted in diffusion space within the TBSS skeleton. One can note that the regions where a significant WM atrophy or CSF hypertrophy occurred in AD patients are largely overlapping with regions where an increased FWF was observed. This suggests that the FWF is an accurate descriptor of macroscopic changes induced by the pathology.

Generally, the differences observed with TBSS were consistent, as a decreased FA_{DTI} and increased MD_{DTI} will very likely give an increased FWF, but there still might still be a difference in FA_{FW} or MD_{FW}. Overall, there was no region where a difference observed with the simple DTI model disappears when using the FW model. As the FW model accounts for partial voluming, it brings more specificity to the tensor indices that are supposed to describe the white matter micro-structure. Our Alzheimer's disease study clearly shows that the FA_{DTI} decrease was primarily describing a macro-structural effect, i.e., tract atrophy and/or ventricles hypertrophy.

In the anterior part of the corpus callosum, the VBM did not reveal significant atrophy of the tract in AD patients. However, both the MD_{DTI} and the FWF showed a significant increase in AD patients. This suggests that either the FWF and MD_{DTI} are more sensitive than the VBM (unlikely), or that the MD_{DTI} and FWF captured some micro-structural effect in AD patients. In the latter case, we claim that whether the difference appears in MD_{FW} or FWF is a matter of constrains on the MD_{FW}.

Our model also unveiled both increases and decreases in FA_{FW} that were not visible with FA_{DTI}, and we can argue with more confidence that we are describing a true micro-structure alteration in this case.

3.4 Discussion

The contribution of this work is a partial volume model including free water in conventional diffusion MRI, which enables us to estimate the isotropic tensor accounting for free water contribution to the signal. We assessed the nature and relevance of a FW compartment modelled along with a diffusion tensor in deep WM.

3.4.1 Identifiability of the FW model

Our simulations show that FA and MD are reasonably well recovered by our model, compared to the tensor model. The novelties of our method, compared to the work presented by Pasternak et al. [Pasternak et al., 2009] reside in two main points. First, the nature of the fitting (MCMC) ensures the tensor remains positive definite and its eigenvalues positive. Additionally, we obtain a distribution for all the

parameters that relates to the uncertainty of the parameter estimates. This extra information could be useful for the statistics, and also for probabilistic tractography [Behrens et al., 2003]. Second, and most importantly, we are using a Bayesian framework that allows to set priors on the model parameters. The key difference with the FWE method [Pasternak et al., 2009], which uses spatial regularization, is that the Bayesian framework allows to balance the fitting between likelihood and prior information, depending on the level of regularization needed in a given voxel.

As we have shown in Figure 3.5 and 3.6, the amount of regularization needed depends on the underlying FA of the tissue, with higher FA giving higher model identifiability. Because the FW compartment is isotropic by definition, an anisotropic attenuation, even with one-shell data, will help to separate the contribution of each compartment to the signal. In this case, the model fitting will be primarily guided by the likelihood, whereas in areas of low anisotropy, e.g. in gray matter, it will follow the prior. As the MD and the FWF are the most correlated parameters of the model, the bias on MD estimates also affects the FWF estimation, especially at low FWF regime. This observation questions the interpretability of the FW compartment in deep white matter, where the estimated FWF is typically below 20% [Pasternak et al., 2012a].

The prior is particularly important when fitting the model on single shell data, but it comes at the cost of narrowing the range of MD estimates when the FWF increases, as the data does not support the fitting any more (Figure 3.4). Adding an extra shell with a relatively small b-value, where the free water signal is not completely attenuated, allows to relax the prior on MD and gives more precise and accurate estimates of FA and MD. We chose $300 s.mm^{-2}$ as a b-value because at this weighting the signal still contains a significant contribution from the FW (about 40%). In a recent work, Hoy et al. suggest that using two shells with a weighting of 500 and 1500 $s.mm^{-2}$ would be close to the optimal weighting for the fit of a model similar to ours [Hoy et al., 2014].

The in-vivo results showed that both FWE method and our FW model give FA maps that are similar to those obtained with an inversion recovery acquisition that attempts to null free water signal. When we compare our method with the FWE, we obtain similar FA maps, but without the need for spatial regularization. The FWE method proves to greatly constrain the range of MD estimates, as these did not exhibit any correlation with the IR acquisition. Also, our method produces FWF estimates that are very close to the ones obtained from the CSF partial volume segmentation. The signal attenuation correction described in Equation (3.2) proves to be efficient as the relation between T1w CSF PVF and FWF is linear. This linear relationship suggests that the T1w CSF partial volume estimates could be used to drive the FW model fitting, if the T1w and diffusion images could be perfectly aligned. We acknowledge that the in-vivo experiments were conducted on only one subject, and reproducing the results remains to be done in future work. However, we believe that the trends are sufficiently strong to support our conclusions.

3.4.2 Properties and relevance of FW in deep white matter

In previous work "free water" was found to be present in deep white matter [Pasternak et al., 2012a]. It was suggested that it arises from the extra-cellular space (ECS) between axons, and that it can constitute a biomarker specific to ECS alteration such as neuro-inflammation [Pasternak et al., 2012b] [Maier-Hein et al., 2014] [Pasternak et al., 2015] [Pasternak et al., 2016]. While we also found the presence of a significant FWF in deep white matter, our results question its origin and relevance.

The properties of the ECS in white matter have been investigated in the past [Syková and Nicholson, 2008] and it was established that its volume fraction is typically around 20% in healthy WM tissues, which corresponds well to the

volume fraction estimated by the FWE method [Pasternak et al., 2012a] and in our experiments (16%). The interstitial fluid is the main component of the ECS [Syková and Nicholson, 2008], and it is expected to share properties similar to the CSF, i.e. a long spin-lattice relaxation time (T1). If the FW compartment was reflecting the ECS, its T1 relaxation time should be longer than the T1 of the tissue compartment, because an expansion of the extracellular space, e.g. following neuroinflammation, is associated with an increased T1 relaxation time at the voxel level [Sahraian et al., 2010].

In our experiment on the DWI data with 10 shells, we explored further the nature and relevance of the FW compartment. In our FW model (Equation (3.1)) and the FWE one [Pasternak et al., 2009], the diffusivity of the FW is fixed to $0.003 mm^2.s^{-1}$, which is that of CSF in the ventricles. Although fixing the value of the FW diffusivity in the ECS might just cause a shift in the estimated FWF, we were interested in measuring the actual diffusivity of the FW in our model. Using a dense range of b-values, we found that in areas where the FW compartment had a non-zero value, its estimated diffusivity was the same as expected from free water. This is surprising because the diffusion of water molecules in the ECS is hindered by the extra-cellular matrix so their diffusivity is expected to be lower (around 0.0012 $mm^2.s^{-1}$) than in the ventricles [Syková and Nicholson, 2008].

Both experiments show that the FW compartment is unlikely to arise from the extra-cellular space, but rather it is generally the result of bias introduced by the prior on MD and the inability of the diffusion tensor to represent crossing fibres. It is interesting to compare the FW model with NODDI, where a FW compartment is modelled along with an extra-cellular compartment represented by an axially symmetric tensor with fixed diffusivity [Zhang et al., 2012]. The NODDI model also reveals a significant FWF in deep white matter, which in this case is not biologically plausible as the extra-cellular space is explicitely modelled by another compartment. Both the original NODDI paper and a recent study point outs that the FWF in deep

WM is in this case the result of constraints set on the intrinsic diffusivity and lack of data [Reisert et al., 2017].

3.4.3 Implications for a typical group study

In this chapter, the aim of the Alzheimer's disease group study is not to revisit the neuroscientific results per se, but rather to show that modelling FW makes a difference in terms of the results in a typical controls-versus-patients study. We conclude that FW is a confound for FA and MD when using a single tensor model.

Many of the differences in FA_{DTI} between the patient and control groups can be attributed to free water contamination. On the other hand, significant differences in FA_{FW} between the healthy elderly participants and the AD patients could be seen in regions that were not showing any significant differences in FA_{DTI}. Namely, there was a significant increase of FA_{FW} in the centrum semi-ovale of the AD patients – presumably related to the selective degeneration of the SLF compared with the motor-related tracts [Douaud et al., 2011] – which could only be seen subthreshold in FA_{DTI}, suggesting that the FW model used here might be more specific to detect orientation differences between populations. Sensory WM tracts (Figure 3.14) also showed a significant decrease of FA_{FW} in the AD patients, raising the interesting possibility that the FW method is also more sensitive to subtle micro-structural damage not yet detected using conventional DTI (albeit existing subthreshold with a trend for decreased FA_{DTI}), as these sensory tracts are relatively spared compared with associative tracts at this early stage of the disease.

The interpretation of the FWF as a biomarker for microstructural changes is largely compromised in our in-vivo results, as we showed that it was merely a result of the bias introduced by the prior and by the inability of the diffusion tensor to represent crossing fibres. In the context of a clinical study, we observed that some of

the changes in MD_{DTI} were not reflected in MD_{FW}, but were still present in FWF. The VBM results suggest that many of these areas were affected by macro-structural partial voluming, i.e. WM atrophy and/or CSF hypertrophy. We can reasonably argue that the increased FWF and MD_{DTI} observed in these areas predominantly reflect macro-structural changes, i.e. that the FW is only a confound. However, there were still cases where there was a difference in MD_{DTI} and FWF, without changes in MD_{FW} and without significant macrostructural changes revealed by the VBM. It is very likely that this case points out the limitation our approach, where true microstructural changes will rather be revealed by the FWF rather than MD_{FW}. Given our simulations and in-vivo experiments, we can argue with confidence that whether the difference is observed in FWF or MD_{FW} does not relate to the extra or intra-cellular (respectively) origin of the underlying process, but only to the constraints on MD_{FW}. A conservative interpretation of the results is therefore that there was no microstructural change in this specific area.

Overall, we argue that accounting for FW in the context of group studies allows for improved specificity at the cost of a possible loss of sensitivity to changes in MD. This is especially true when the disease studied involves white matter atrophy, as it is in the case of Alzheimer's disease.

3.4.4 Limitations and future work

An important limitation of our approach resides in the usage of the tensor model to represent the tissue compartment. The motivation of our choice was the wide usage of the tensor and its indices (e.g., FA and MD) in clinical studies. With more than 4000 entries in Pubmed for the keywords "DTI" and "study", we believe that our approach is of great interest to revisit previous studies or analyse current DTI datasets. One must be aware that we overcome some limitations inherent to the

tensor model, e.g., its inability to represent crossing fibres. While such limitations might interfere with a good estimation of the FWF, we argue that including a FW compartment in complex fibre orientation models is still of value, especially if we want to interpret orientation-independent parameters (e.g. Generalised Anisotropy, kurtosis [Jensen et al., 2005], or partial volume fractions). Also, we noted the importance of using a second shell with a low b-value (even with just a low number of directions), as this would certainly make such models more sensitive to pathological changes and also allow to disentangle macro-structural changes from micro-structural ones, with limited assumptions or better fibre characterization [Fieremans et al., 2011] [Zhang et al., 2012].

Relaxing the constraints on the model will be the focus of our future work, but in the context of standard clinical acquisitions. We plan to take advantage of the structural T1w images that are almost systematically acquired along with the DWI in clinical studies. Using a machine learning approach, one could leverage the registration issue between modalities by learning the different tissues volume fractions (namely GM, WM and CSF) from the structural T1 and infer them directly from the diffusion data. The estimated partial volume fractions can then be used as priors while fitting a multi-compartment model to the diffusion data. Crucially, this approach would also allow to identify "pure WM" voxels, and in that case, to perform the fitting with a FWF set to 0 and without prior on the MD.

3.5 Conclusion

In this chapter, we have proposed a method for modelling free-water contribution to diffusion MRI signal that we validate through simulation and in-vivo experiments. We have shown that our method is able to retrieve DT indices unbiased by the presence of CSF, and it is applicable to standard diffusion MRI acquisition with a

single shell. The results reveal that our approach is more specific and possibly more sensitive to changes in FA. Importantly, we have shown that the FW modelled along with a DT should not be considered as a biomarker for microstructural change in the extra-cellular space, but only as a confound reflecting white matter atrophy.

4

Machine learning approach to partial volume segmentation

Tissue-type partial volume modelling is generally an ill-posed problem in single-shell dMRI. On the other hand, T1w images, typically acquired along with the diffusion data, allow for an accurate estimation of the tissue partial volume fractions (PVFs). We propose in this Chapter to use a data driven approach to predict the T1w-derived PVFs from the diffusion data. The aim is to alleviate the within subject mis-registration between the two modalities.

We explore various machine learning methods such as Gaussian processes, support vector machines, random forests and neural networks. Our experiments show that the random forests approach is the most accurate and scalable method for predicting the tissue partial volume fraction. We demonstrate that such predictions can be used to inform the fitting of the FW model to retrieve a diffusion tensor that is not biased by partial volume effects or constraints.

4.1 Introduction

We have seen in Chapter 3 that inferring the CSF volume fraction from diffusion MRI data is not a trivial task. We found that using a second shell with a low b-value (300 $mm^2.s^{-1}$) or setting a prior on the tensor MD helps to stabilize the fitting and thus to retrieve tensor indices that are unbiased by CSF contamination. However, such an additional shell might not be always available. In these situations, the prior on the MD introduces a bias in the MD estimation, thus reducing its sensitivity to WM alteration (e.g., in pathology). Additionally, the voxel-wise fitting of the FW model estimates a non-zero CSF volume fraction in deep WM, where it is expected to be zero.

A T1w image is usually acquired along with the diffusion sequence in a typical group study. T1w images generally have high resolution, good SNR and a high contrast between tissues. This allows for an accurate and robust estimation of the tissue partial volume fractions (PVFs). Several methods, such as Gaussian mixture models (GMM) [Zhang et al., 2001] [Ashburner and Friston, 2005] or surface based segmentation [Dale et al., 1999] can be used to retrieve the PVFs of the tissues (WM, GM and CSF).

We propose here to use the CSF PVF estimated from the T1w image as a prior in order to fit the FW model on the diffusion data. We hypothesize that the extra information can improve the stability of the fitting and allow to relax the constraint on the MD. A similar approach was previously proposed by [Eaton-Rosen et al., 2016].

However the crucial issue of mis-registration between the T1w image and the diffusion data was not addressed. For voxelwise PVF estimates from the T1w image to be useful to the PVF modelling in the diffusion data, alignment needs to be accurate to the sub-voxel level. [Eaton-Rosen et al., 2016] applied their method

to the Human Connectome Project dataset, which presents very high quality and undistorted diffusion data, allowing for a good co-registration. In a typical group study, the error in the co-registration between the T1w and the diffusion data is expected to be significant because the two images do not show the same contrast (e.g. diffusion data has a poor WM/GM contrast) and have a different resolution (typically 1mm for T1w and 2mm for DWI).

Other issues limiting the accuracy of the T1w-dMRI co-registration are the susceptibility-induced distortions associated with the echo planar imaging (EPI) and the eddy current induced distortions. The latter can be adressed using a forward prediction approach that enables to predict the diffusion signal expected when no eddy current distortions are present [Andersson and Sotiropoulos, 2016]. For typical phase encodings along the Anterior-Posterior direction, susceptibility-induced distortions are largely affecting the corpus callosum, a structure where CSF contamination is present. Susceptibility-induced distortions can be adressed using a separately acquired fieldmap that allows to accurately unwarp the EPI images [Jezzard and Balaban, 1995]. Alternative methods can be used to estimate such fieldmap when the latter was not acquired in the experiment, but they require the acquisition of images with opposite phase-ecoding directions [Andersson et al., 2003]. However, in many dMRI group studies, no fieldmap or opposite phase encoded images are available. Consequently, using directly co-registered T1w-derived tissue PVFs in a diffusion model may result in significant deviations in the estimation of the tensor FA or MD.

In order to avoid the co-registration issue, we propose to infer the PVFs of tissues directly from the diffusion data. This might be achieved by using the same approach as for T1 segmentation, by using some diffusion features (e.g. FA, MD) in a multivariate GMM [Kumazawa et al., 2010]. Such methods unfortunately prove to be inaccurate, as the variables might not follow a normal distribution (e.g. the FA) and the relation between the variable and the PVFs is non-linear. Another approach

using constrained spherical deconvolution (CSD) proposed to estimate the tissue PVFs based on their response function [Jeurissen et al., 2014]. This method prove to be reliable, but it requires a multi-shell acquisition and relies on the registered T1w segmentation to identify the tissue types.

We propose to use a data-driven approach to learn a mapping between diffusion features and the tissues' PVFs estimated from the T1w images. Supervised machine learning methods have already been successfully applied in diffusion MRI. For example, it was shown that the amount of data needed to fit the NODDI model [Zhang et al., 2012] can be reduced by a factor of 12 when using a neural network approach [Golkov et al., 2016]. In another study, the same model was successfully fitted on single shell data by applying a random forest regression method [Alexander et al., 2017].

In this chapter, we compare 4 supervised machine learning approaches in their ability to predict the tissue PVFs. We use a range of diffusion features estimated from models applicable to low b-value single-shell data. We also evaluate the robustness of our approach to the error in registration between the T1w and diffusion images. Finally, we use the CSF PVF predicted by our method as a prior while fitting the FW model and assess how the additional information relaxes the constraints on the diffusion tensor parameters.

4.2 Background on machine learning techniques

Predicting the T1w-derived tissue PVFs from the diffusion data is a regression problem which can be formulated as $Y = f(X, \beta) + \varepsilon$, where Y is the quantity of interest (i.e., the tissue PVFs), f the regression model, X the diffusion features, β the regression model parameters and ε the residuals. Amongst a wide range of methods that address the problem, we chose supervised machine learning methods

because no explicit model can map the single-shell diffusion data to tissues PVFs. The methods we are comparing are: random forests (RF), support vector machines (SVM), Gaussian processes (GP), and neural networks (NN).

4.2.1 Random forests

The random forest method relies on binary regression trees where each node contains a linear transformation of the features and each leaf gives the quantity of interest [Breiman, 1984]. A regression tree splits a non-linear regression problem into smaller problems which can be solved linearly. A tree is iteratively built by splitting the data at each node. The splitting decision is typically based on the mean squared error, as a node finds the best fit for a subset of the input data. An important parameter of the method is the maximum tree depth. Shallow trees will fail to represent the data correctly. Deeper trees will explain the data better, but they are prone to over-fitting, i.e. they have low bias but high variance and they do not generalize well to new data.

A random forest is a collection of trees, where each tree is built on subset of the data drawn at random [Breiman, 2001]. It is an ensemble method, which means that each tree is a weak learner. An important parameter is the number of features to use for each split, which controls the proportion of features considered when computing the decision criteria. As a result, the trees are weakly correlated between each other. The output of the forest is obtained by averaging the predictions of all the trees. Consequently, averaging the trees will increase the precision of the predictions, without increasing the bias. It follows that the trees constituting the forest can be grown very deep as long as the number of trees is sufficient.

4.2.2 Support Vector Machines

SVMs are well-known machine learning approaches that were introduced by Vapnik et al. [Vapnik, 1995]. SVM was originally designed for classification problems (SVC), and it was rapidly extended to regression (SVR) [Drucker et al., 1997]. In the case of classification, a SVC aims to maximize the margin between a hyperplane h and the closest data-points from the different classes. The support vectors w are simply the coordinates of the data-points x that are being considered for fitting the hyperplane, defined by $h(x) = w^T x + w_0$ and constrained by $w.x - w_0 >= 1$ if $y_i = 1$ and $w.x - w_0 <= 1$ if $y_i = -1$, where w_0 is the intercept of the hyperplane and y_i the i^{th} class prediction.

In the case where the data is not linearly separable, the problem can projected into a higher dimensional space, with the aim of finding a linear boundary between the classes. In practice, a non-linear transformation ϕ is applied to the data-points in feature space: $h(x) = w^T \phi(x) + w_0$. To avoid a costly computation of the optimal hyperplane in a high dimensional space, the mapping is implicitly embedded in a kernel K. The computation is made simpler by writing the kernel as a mapping of the features from the original feature space to the higher dimensional space. The kernel is then written as: $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. The trick is that an explicit representation for ϕ is not needed as long as $\langle ., . \rangle$ is a proper inner product. Widely used kernels are the polynomial kernel $K(x_i, x_j) = (x_i^T.x_j + 1)^d$ and the Gaussian kernel (or radial basis function, RBF) $K(x_i, x_j) = exp(-\gamma ||x_i - x_j||^2)$, with $\gamma = \frac{1}{2\sigma^2}$. γ defines the radius of influence for a single training example, with low values for γ corresponding to a large radius and high values for a small radius. Thus, large γ values lead to models with high bias and low variance.

However, it is in general impossible to find a hyperplane, even in a higher dimensional space, due to noise and mis-labelled datapoints. The soft margin

technique was introduced to allow mis-classification of some data-points [Cortes and Vapnik, 1995]. The technique finds a hyperplane that minimises the number of mis-classified data-points with the introduction of slack variable ξ, which allows to relax the constraints on the support vectors so that $\xi = max(0, 1 - h(x))$. Training a SVC can be summarised as follows: $min\left(\frac{1}{2}||w||^2 + C\sum_{k=1}^{p}\xi_k\right)$. The constant $C > 0$ is the regularization parameter which trades off misclassification of training examples against simplicity of the hyperplane. Low C values makes the decision surface smooth, while a high C values aim to classify all training examples correctly.

With regards to regression, the aim is to find a function that minimises the distance between the data-points and the hyperplane. The method remains the same as for SVC, with the exception that the hyperplane is constrained by $y - w.x - w_0 <= \varepsilon$ and $w.x + w_0 - y <= \varepsilon$, where y is the predicted continuous output. By introducing the parameter ε, the hyperplane has data-points on either side, but the distance between these points and the hyperplane should not be larger than ε. The hyper-parameters of the soft-margin ε-SVR with an RBF kernel are hence C, γ and ε.

4.2.3 Gaussian Process

A Gaussian Process (GP) performs a probabilistic interpolation of the training data-points x and associates an uncertainty to the predictions y. In GP regression, the data is represented as a sample from a multivariate Gaussian distribution. The relationship between one observation and another is the covariance matrix, modelled by the covariance function, assuming that the mean of the GP is zero.

A popular choice for the covariance function is the Gaussian, or RBF kernel: $K_{RBF}(x_i, x_j) = exp\left(\frac{||x_i - x_j||^2}{2l^2}\right)$, where l is the length scale hyper-parameter. The length-scale defines the period of variation for the predictions in the features space. In the case where the mean of the GP is not zero, it can be modelled with a constant

C which is also a hyper-parameter. Additionally, it is useful to estimate the noise in the data, with a white noise kernel given by $K_{noise}(x_i, x_j) = \sigma\delta_{x_i,x_j}$, where δ is the Kronecker delta.

In this case, there are 3 hyper-parameters to be estimated: C, l and σ. A particularity of GP is that their set of hyper-parameters θ can be estimated on the training data only, without cross-validation. This is done by maximizing the log-marginal likelihood of the observations y, given θ and the data x:

$$p(y|\theta, x) = -\frac{1}{2}y^T K^{-1} y - \frac{1}{2}log|K| - \frac{n}{2}log(2\pi) \tag{4.1}$$

where K is the Kernel, n the number of observations, and θ the set of hyper-parameters.

The GP needs to compute the full covariance matrix over all the feature space, which can be computationally very expensive when the number of samples exceeds a few tens of thousand. The optimisation of the log-marginal likelihood is done with a quasi-Newton approach, usually optimised for memory consumption, e.g., L-BFGS [Byrd et al., 1995].

4.2.4 Neural networks

Neural Networks (NN) are a relatively old machine learning approach that took inspiration from the biological neural networks of animal brains. A class of neural networks is the multi-layer perceptron, where a series of layers maps an input vector (the features x) to some continuous quantity, or class y. Each layer is constituted of a finite number of units that performs a transformation f of its inputs. The layers are fully connected, meaning that each unit from a given layer has an incoming connection with each unit of the previous layer, and an outgoing connection with

all the units of the next layer. Each connection between the units is weighted by a scalar w (a separate w for each connection). Additionally, a bias vector b is added to the output, allowing to shift the response function. The output o_u for a given input vector x_u of a unit u can be written as: $o_u = f(\sum_{u=1}^{n} w_u.x_u) + b_u$, where n is the size of the input vector.

The activation functions f of the layers are crucial to the performance of the NN, notably to its ability to address non-linear problems. Examples of common activation functions are the rectified linear unit (ReLU) [Nair and Hinton, 2010], hyperbolic tangent (tanh), sigmoid and softmax.

The parameters of the network are the weights for each connection between units. This gives rise to very high-dimensional optimisation problems, usually solved with a stochastic gradient descent approach (SGD). To make the latter efficient, the optimal weights are computed layer by layer starting from the last layer, and the error is propagated up to the first layer. This technique is called back-propagation.

A major problem of the NN is that their high number of parameters gives rise to very complex models which makes the method prone to over-fitting, i.e., the model does not generalise well to new samples. Amongst many approaches to control the over-fitting, a simple and elegant one is to randomly discard a proportion of the units for each layer [Hinton et al., 2012]. The proportion of units is defined by the dropout rate, which is gradually increased from the first to the last layer. The forward propagation and the optimisation through back-propagation are performed without taking into account the units which are randomly discarded at each iteration of the optimisation.

4.3 Materials and methods

4.3.1 Data acquisition and pre-processing

We ran our experiments on two similar datasets for two separate subjects, one male 31 years old and a female 32 years old. Data was acquired on a 3T Siemens Verio scanner. Diffusion data consisted of a standard DWI acquisition with TR=9600ms, TE=76ms and b=1000$s.mm^{-2}$, with 60 non-collinear diffusion gradient directions in addition to 6 non-weighted b=0 images. Half of the b=0 images were acquired with an anterior-posterior phase encoding direction and the other half with the opposite direction. The voxel size was 2mm isotropic, with a parallel imaging acceleration factor of 2, for a total scan time of 12 minutes. The structural $T1w$ image was acquired at a resolution of 1mm isotropic for a scan time of 6 minutes.

Another set of diffusion data was also acquired, consisting of an Inversion Recovery (IR) sequence where the CSF signal was nulled. The acquisition parameters were: TR=20.6s, TI=2300ms, b=1000$s.mm^{-2}$, with 60 non-collinear diffusion gradient directions in addition to 6 non-weighted b=0 images. This acquisition allows retrieving tensor indices that are not contaminated by the CSF contribution to the signal.

The diffusion data was preprocessed to correct for susceptibility-induced distortions with TOPUP [Andersson et al., 2003] and for movement and eddy currents distortions with EDDY [Andersson and Sotiropoulos, 2016]. The diffusion features consisted of 17 parameters resulting from: the DWIs (mean b0, mean b1000), the diffusion tensor (FA, MD, tensor mode and the 3 tensor eigenvalues), Ball&Stick model [Behrens et al., 2007] (Volume fractions of the 3 fibres, total fibre volume fraction, isotropic diffusivity and the dispersion of the 3 fibres orientations) and finally probabilistic tractography (visitation maps for full brain tractography). In order to add some information from the neighbouring voxels, we added a smoothed version of each feature map (σ = 6mm) as well as its gradient, for a total of 51

features.

The tissue PVFs were estimated on the T1w image using FAST [Zhang et al., 2001]. These PVFs were then registered to the diffusion data with FLIRT BBR [Greve and Fischl, 2009].

4.3.2 Predicting tissues volume fraction

Our method relies on voxel-wise regression, that is to learn a mapping between the features derived from the diffusion data and the corresponding tissue PVFs computed on the T1w image. A supervised machine learning (ML) approach requires a training and a testing dataset. In our case, training is performed on one subject (male) and testing on another subject (female). Reversing this order produced the same results. In the case where only one subject is available, the train and test data can be split between hemispheres.

4.3.3 Comparison of the machine learning approaches

We compare the performance of 4 machine learning approaches: random forest (RF), support vector machine regression (SVR), Gaussian process (GP) and neural networks (NN). The methods were implemented in Python using the scikit-learn library [Pedregosa et al., 2011]. Their hyper-parameters were optimised using grid search for the RF and SVR, gradient descent for the Gaussian process and an empirical manner for the NN. The optimisation of the 4 ML approaches (RF, SVR, GP and NN) was based on the correlation between the predictions and the T1 segmentation, except for the GP, where the parameters are optimised only on the training set.

Each of the 4 methods was optimised on the same training set, comprising 20000 samples.

The performance of each ML approach was similarly assessed with the correlation between their PVF predictions and the PVF estimated from the T1, for each tissue (WM, GM and CSF). The correlation was computed on the full brain of the test subject. It ensured that the performance measure includes false positives, e.g. non-zero white matter volume fraction in the CSF as well as false negatives, e.g. zero WM volume fraction in the middle of the white matter. The correlation coefficient is not expected to reach 1 because of residual mis-registration between the DWI and T1 images.

Features importance was assessed with the RF, according to the out of bag error (OOB). The latter is the prediction error introduced when a given feature is removed. The advantage is that it is computed during the training phase, without additional cost. Empirical evidence showed that it is as accurate as estimating the error on a separate testing dataset [Breiman, 1996]. Although some of these features are not mutually independent, we chose to include all of them to maximise the available information.

4.3.4 Mis-registration between T1 and diffusion space

The main motivations to use a ML approach is the mis-registration between T1 and diffusion space. Therefore, we evaluated how the registration error affects the performance of the learning process. We introduced a registration error between structural and diffusion data on the learning dataset, and computed the performance of the segmentation on the test dataset was impacted.

To illustrate further how our approach can alleviate the mis-registration issue, we trained our RF model to predict the absolute T1 intensities at their native

resolution (1mm). We up-sampled the diffusion features to the resolution of 1mm, and predicted the T1 intensities with the RF approach. This application is just an example showing the potential of our ML approach. We chose the RF because it is very fast, an important factor when the training is performed on 1mm voxels.

4.3.5 Application to FW modelling

In this experiment, we fitted the FW model model described in Chapter 3, Equation 3.1. We used the predicted CSF volume fraction as a prior for the FW compartment. The mean of the prior is equal to the volume fraction predicted, with a correction factor to account for the different T1 and T2 relaxation times of the tissues, according to Equation 3.2 in Chapter 3. The standard deviation of the tissue PVF prior is set to 0.1, which approximately corresponds to the standard deviation of the error between our model prediction and the T1 segmentation. We also relaxed the prior on the MD, increasing its standard deviation to 0.5 $\times 10 -^{-3} mm^2.s^{-1}$.

4.4 Results

4.4.1 Machine learning approaches optimisation and comparison

4.4.1.1 Optimisation of hyper-parameters

Random forest For the RF, the optimisation of the hyper-parameters was conducted using a 3-fold cross validation scheme. We looked at the impact of the number of trees, the minimum samples per leaf and the number of features to be considered when looking for the best split. Figure 4.1 shows a fast increase in the performance with the number of trees, up to 25 trees. For more than 25 trees, the correlation slowly increase up to 100 trees. It remains constant for a number of trees greater than 100. The maximum number of features to be considered for a split has a strong impact on the performance. The optimal number is located between 80% and 100%.

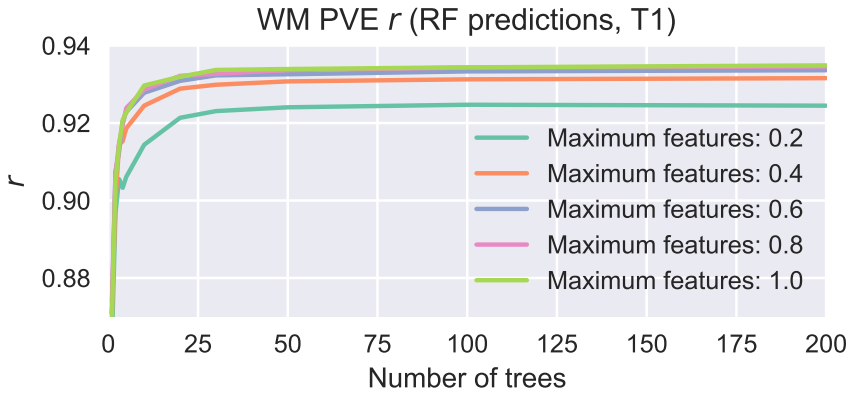


Figure 4.1: Random forest hyperparameter optimisation - Number of trees and features. Correlation between the T1 WM PVEs and the RF predictions as a function of the number of trees, for various maximum number of features

Figure 4.2 illustrates the correlation between the RF prediction and the T1 WM PVFs as a function of the maximum number of features to be considered for a split and the minimum number of samples per leaf. The highest correlation is obtained for almost fully-grown trees (minimum samples per leaf=2) and 90% of the features to be considered for each split.

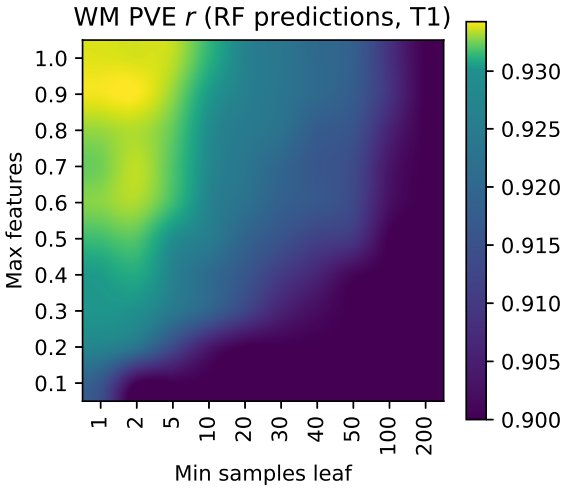


Figure 4.2: Random forest hyperparameter optimisation - Tree depth, number of features Correlation between the T1 WM PVFs and the RF predictions as a function of the minimum number of samples per leaf and maximum number of features

Support vector regression The hyper-parameters optimisation for the SVR was conducted using a 3-fold cross-validation scheme. We found that the ε parameter had limited impact on the performance, with an optimal value of 0.01. We ran a grid search on the γ and C parameters. Figure 4.3 illustrates the correlation coefficient for the WM PVE as a function of (γ, C). The best performance was obtained with a value of 1 for C and a value of 0.1 for γ. We can observe that the two parameters are negatively correlated. When γ is low, the optimal performance is obtained with a high value of C. This illustrates that the high bias low variance obtained with a high γ does not need much regularisation, i.e. the values of C can remain high. Conversely, when γ is low, it produces a hyperplane with high variance and low bias, that benefits from the regularisation corresponding to a high value of C.

Gaussian process For the GP, the method allows to optimize the hyper-parameters during the training phase, by minimising the negative log-likelihood of the model with a gradient descent scheme. We used a RBF kernel for the covariance function together

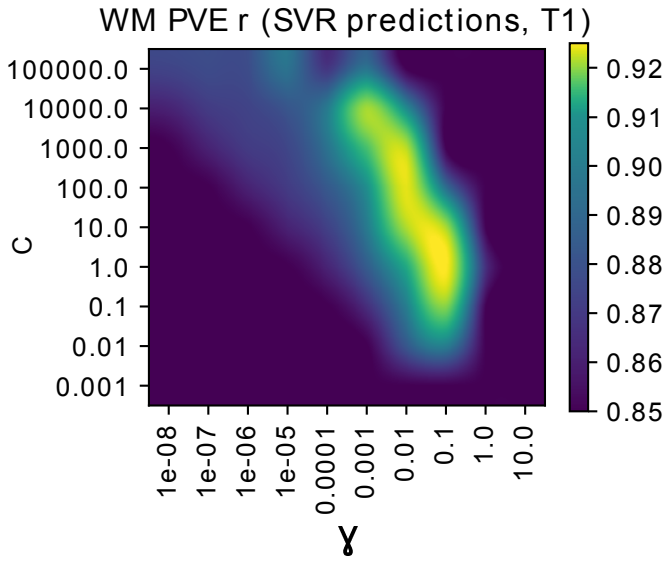


Figure 4.3: SVR hyperparameters optimisation - γ and C. Correlation between the T1 WM PVFs and the SVR predictions as a function of γ and C

with a white noise kernel and a constant kernel. Figure 4.4 shows the landscape of the optimisation problem for the hyper-parameters of the covariance function, i.e. the length-scale of the RBF kernel and the noise level. The first observation is that the optimisation problem appears to be convex, with a global optimum. Both the noise level and the length-scale have an important role in minimising the log-likelihood. The maximum likelihood is obtained for a length-scale of 3.6. We only have one length scale for all the feature space; in other words, the kernel is isotropic in the features space. We found that using an anisotropic RBF kernel (i.e. one length-scale per feature) did not increase the log-likelihood or the prediction accuracy.

We note that the RBF kernel is infinitely differentiable, and it results that the associated GP has its mean square derivatives differentiable for all orders and it is thus very smooth. As our features might not live in a infinitely differentiable space, we additionally tried to replace the RBF kernel with a Matérn kernel which allows to control the smoothness of the covariance function [Rasmussen and Williams, 2006]. However, using a Matérn kernel that was set to be only once or only twice

differentiable did not improve the log-likelihood or the prediction accuracy.

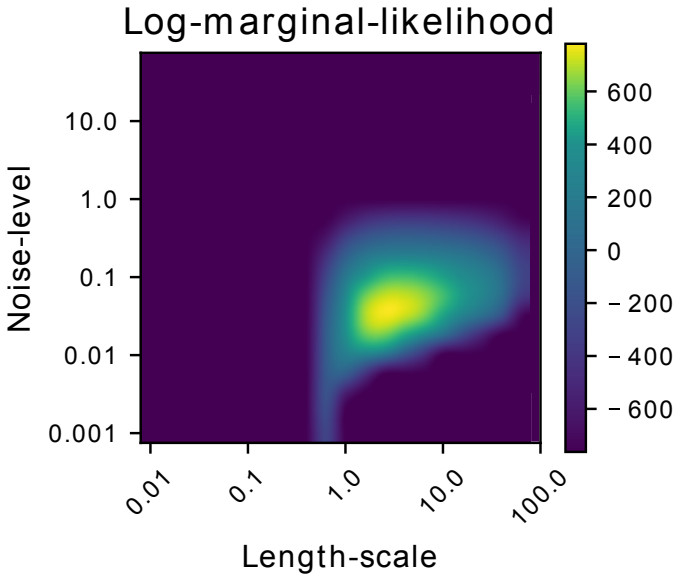


Figure 4.4: Gaussian process hyperparameters optimisation - noise level and length scale Negative marginal log-likelihood as a function of Gaussian process the RBF isotropic length-scale and the noise level

Neural network For the NN, we found the optimal settings empirically, by sequentially trying different activation functions, number of units per layer, number of layers, drop-out rate and learning rate. Our optimal network is illustrated in Figure 4.5. We found that 3 hidden layers, with 200 units per layers were enough to capture the complexity of the relationship between the features and the tissues PVEs. The rectified linear activation function showed the best performance for the hidden layer. The output unit activation function is the sigmoid function, which ensures that the predictions are bounded between 0 and 1, while also showing good overall performance. The regularisation was performed with a dropout rate of 0.1, 0.2 and 0.3 for the respective hidden layers 1, 2 and 3. The optimisation of the weight was conducted with the usual back-propagation method via a stochastic gradient descent with Nesterov momentum. A learning rate of 0.002 and a momentum of 0.9 appeared

to ensure robustness as well as reasonably fast convergence.

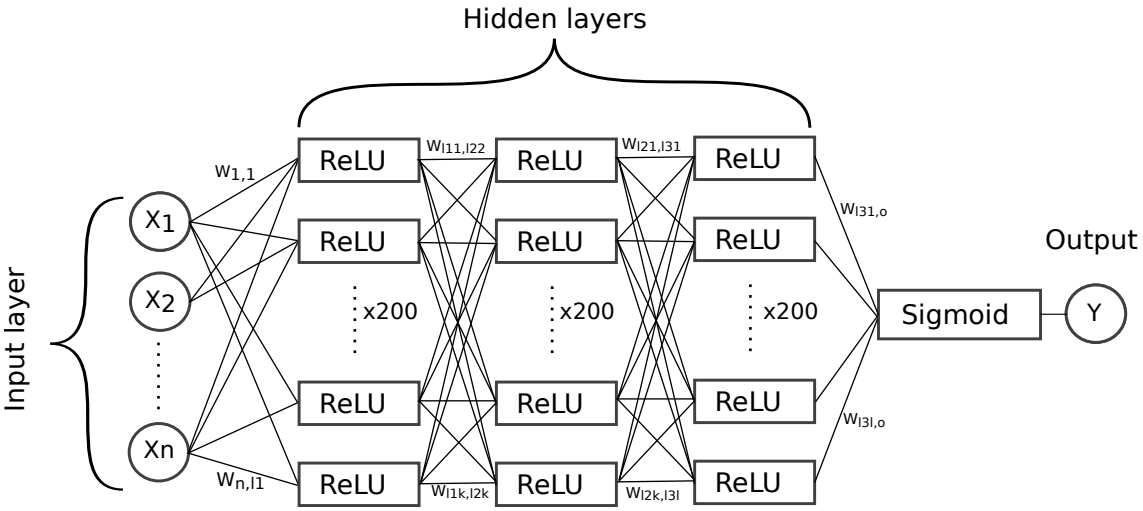


Figure 4.5: Architecture of the neural network. Each box represents a unit with the ReLU activation function, n is the number of input features, and w is the weight of each connection

In figure 4.6, we can observe the learning curve of the neural network. As expected, the loss function rapidly drops at the beginning of the optimisation, up until 25 epochs. Then, the slope gradually decreases, and finally stabilises after 800 epochs. We see that the training loss is smaller than the validation one, but the ratio remains high, at 0.96 towards the last epochs. This means that the overfitting is controlled and that the NN generalises well to unseen data samples (validation set).

Comparison of the prediction functions Figure 4.7 shows an overview of the functions that maps 2 diffusion features to WM partial volume fraction. To be able to visualise the function, we restricted the diffusion features to MD and FA. The methods were trained on the data illustrated in the top left corner of the Figure (T1 Segmentation). As expected, high WM PVE corresponds to a high FA, and low WM PVE is found when the FA is low and the MD is high.

We can see that a tree divides the feature space into discrete areas, each area having the same WM PV prediction. The random forest predictions illustrate well

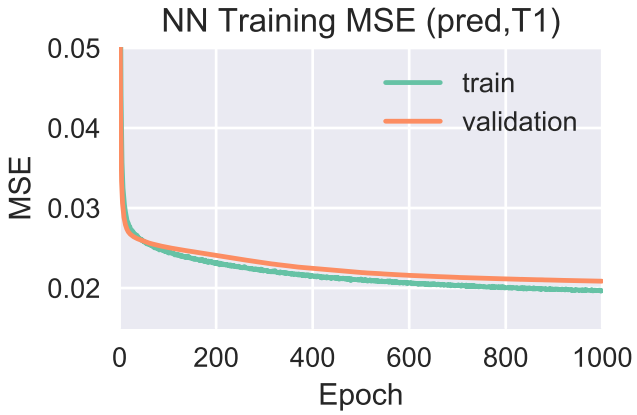


Figure 4.6: Neural network optimisation loss-function Mean squared error on the training (green) and validation (orange) dataset as a function of the number of epoch.

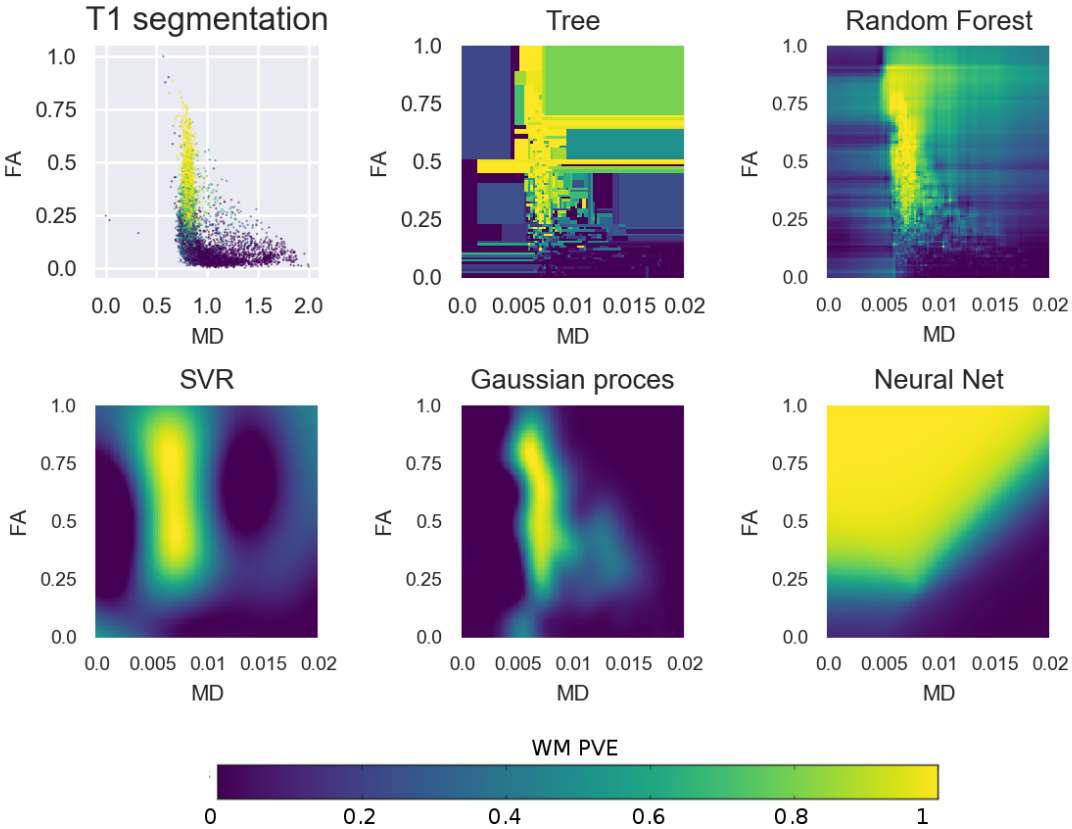


Figure 4.7: Illustration of the model WM PVF predictions for 2 features. The different model predictions for the WM partial volume fraction in a 2D feature space (FA,MD). The top left plot shows the T1 segmentation WM PVE on which the models were trained.

how it is built by averaging different trees. The space still appears discrete, but a smooth pattern emerges.

The SVR function appears smooth and it also give high WM PVF for high FA values. It is also reaching a maximum value of 1. However, we observe that the SVR extrapolates when both the FA and the MD is high (FA>0.5 and MD>1.5). These values are unlikely to be found in-vivo, and they indeed do not appear in the training set. In this case, there is no good prediction to be made, but there is also no reason to give high predictions for the WM.

The GP produces predictions that are less smooth, but it does not exhibit the extrapolation pattern observed with the SVR.

The NN gives a very smooth function, with a very different pattern compared to the other methods. The function appears much simpler than with the other methods, but one should consider that only 2 features were used in this example. Additionally, some parts of the plot had very little data originally, e.g., high FA and high MD.

4.4.1.2 Diffusion features

We show in Figure 4.8 the features' importance in terms of OOB error when predicting the 3 tissue types with a random forest. We can observe that the most important features relate to anisotropy of the signal. Indeed, the total volume fraction of the fibres (fsum) and the fraction of the first fibre of the ball and stick model rank first and third, while the tensor FA is slightly less important. The tractography visitation maps prove to be very important as well, likely because this helps to characterize deep white matter in crossing fibres areas. Additionally, the tractography visitation map brings accurate super-resolution information, as was demonstrated in tract density imaging [Calamante et al., 2010]. The raw diffusion features, i.e. the mean b0 and mean b1000 also contribute to the model accuracy, as

they provide a good contrast between tissues. The remaining features still have some importance in the prediction accuracy, with the exception of the volume fractions of the second and third fibres.

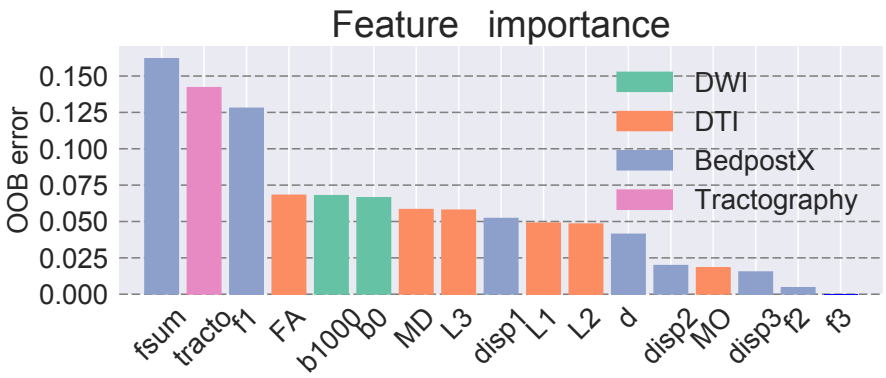


Figure 4.8: Diffusion features importance. Diffusion features ranked by their out of bag error estimated with the random forest for WM PVF prediction.

The addition of spatial information in the set of features has a positive effect on the prediction accuracy of all the tissues' PVF. In Figure 4.9 (Top), we can see an example of the gradient and smoothing ($\sigma = 6mm$) of the FA map. At the bottom of Figure 4.9, we observe the impact of the additional features on the performance of the RF regression for the three tissue types. Adding the gradient of the image for each feature, we notice an increase in performance for all the tissues. Adding the smoothed feature maps increases the performance even further for the CSF and GM predictions. Finally, we added both the gradient and the smoothed features, which unsurprisingly gives the best performance.

4.4.1.3 Comparison

In figure 4.10, we show the performance obtained with the 4 ML approaches for the 3 tissue types (CSF, GM and WM). Overall, the RF exhibits the highest correlation between the predictions and the T1 segmentation, for the 3 tissue types

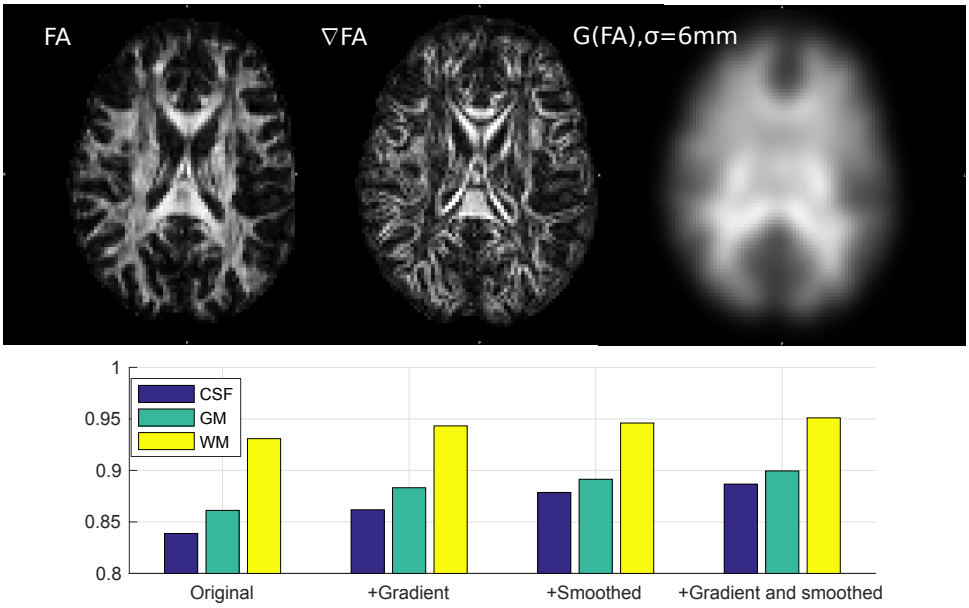


Figure 4.9: Impact of the additional features on the performance. Example of edge and smoothed version of FA (top) and related improvement in the PVF predictions correlation with T1 PVFs

(r =0.83, 0.88 and 0.93 for the CSF, GM and WM respectively). It is closely followed by the NN. The SVR and GP perform less well. In addition to its high accuracy, the random forest method was also the fastest. The training phase was performed in a few minutes, whereas the other methods required a few hours.

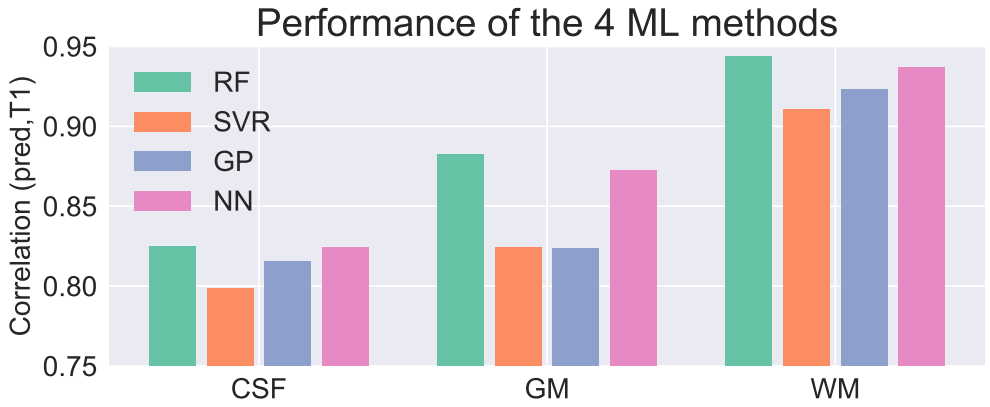


Figure 4.10: Comparison of the four ML approaches. Correlation between T1w-derived and predicted (from diffusion data) partial volume segmentation using random forest (RF), Gaussian process (GP) and support vector regression (SVR)

Figure 4.11 shows the segmentation maps predicted by the 4 methods, as well

as the T1w image segmentation. For a given voxel, the tissue with the highest PVF is represented. Visually, the RF and the NN give segmentation maps that are very similar to the T1 segmentation. The CSF predictions appear accurate for the RF and NN, whereas the GP and SVR gives low CSF PVFs in the ventricle, which quite never reaches 1. Both GM and WM PVF predictions appears noisier with GP and SVR than with RF and NN. Looking at crossing fibres areas, we see that the NN is the only method that predicts a WM PVF of 1. The RF still gives high WM PVF prediction in the crossing areas (> 0.95), and the SVR give lower volume fraction (> 0.7), while the GP gives noisy estimates in most part of the WM.

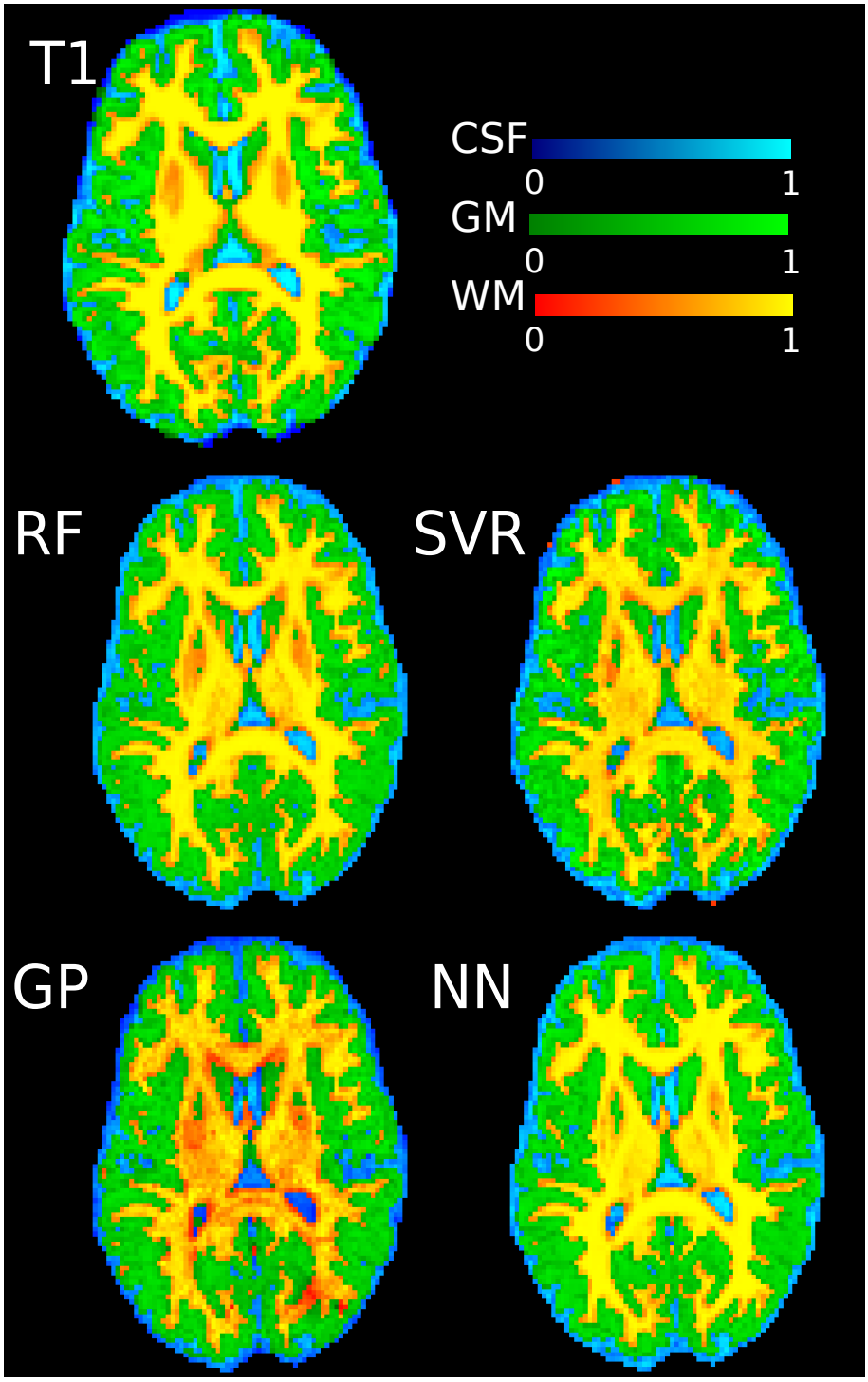


Figure 4.11: Segmentation maps predicted with the 4 ML methods. Tissues' partial volume fraction obtained from the T1 (Top) and the 4 machine learning methods. All the segmentations are displayed in the diffusion space

Looking at the T1 image segmentation, we can see that the sub-cortical GM

structures are a mixture of GM and WM. It is a known limitation of the GMM approach, as the intensities of sub-cortical structures are between those of cortical GM and "pure" WM. Notably, the caudate nucleus is identified as mostly composed of GM, whereas the putamen and thalamus are identified as mostly WM. We can observe a similar trend in the ML predictions. We tried to include a sub-cortical segmentation [Patenaude et al., 2011] in order to have an accurate sub-cortical GM T1w segmentation. The results (not shown) indicated that none of the ML method were able to recover the sub-cortical GM PVFs correctly. Therefore, we did not use the FIRST T1 segmentation, and conceded that the subcortical GM is partly composed of WM.

4.4.2 Mis-registration between T1 and diffusion space

Figure 4.12 shows the correlation and MSE obtained on the test dataset as a function of the mis-registration (translation along the x, y and z axes) artificially introduced in the training dataset. To compute the scores, we used the RF regression. The plots show that the correlation is at its maximum when there is no mis-registration, reaching 0.943, with an MSE of 0.017. The correlation gradually decreases (and conversely, the MSE increases) as the registration error increases. Interestingly, we see that a mis-registration along the x axis (left-right orientation) has more impact on the performance than in the y and z axes.

Here, the key result is that the method is very robust to misregistration of the training dataset. Indeed, the correlation and the MSE obtained on the test dataset stays almost constant when the training error remains under 2mm, which corresponds to the voxel size of our diffusion data.

The robustness of our approach validates its applicability to datasets where EPI distortions are significant and cannot be corrected.

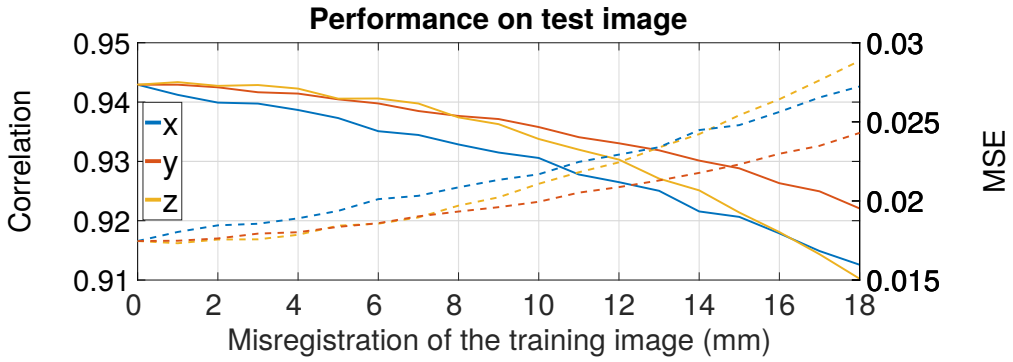


Figure 4.12: Impact of the mis-registration on the RF performance. Correlation and MSE between predicted and ground truth PVF on the test data as a function of the misregistration between diffusion and structural space on the training data.

We performed an additional experiment to predict the absolute T1 intensities at their native resolution (1mm), using RF and the diffusion features up-sampled to 1mm. In the Figure 4.13, the RF predictions are shown together with the T1w original image in the same space. We observe that the predictions look very similar to the original T1w image, and the correlation coefficient between the two images is high ($r = 0.948$). The contrast between grey and white matter is excellent on the original T1w image, as expected from this modality. The WM/GM contrast is also very good in the predictions. This is an interesting result, given that the prediction is obtained from a single-shell acquisition.

The main outcome of this experiment is the quality of the predictions in terms of resolution. Although the diffusion features were derived from data with a resolution of 2mm, the predictions at 1mm appear excellent. The predicted T1w is slightly blurrier than the original one, but the actual resolution is very close to the original T1w. In term of the resolution element (RESEL) [Worsley et al., 1992] the original T1w linear RESEL is 17.2mm and the predicted T1w RESEL is 19.4mm. The latter is lower than the RESEL obtained with spline interpolation, which is in this case 22.5mm. Importantly, this result shows that our method takes advantage of the sub-voxel "super-resolution" information contained in the diffusion data.

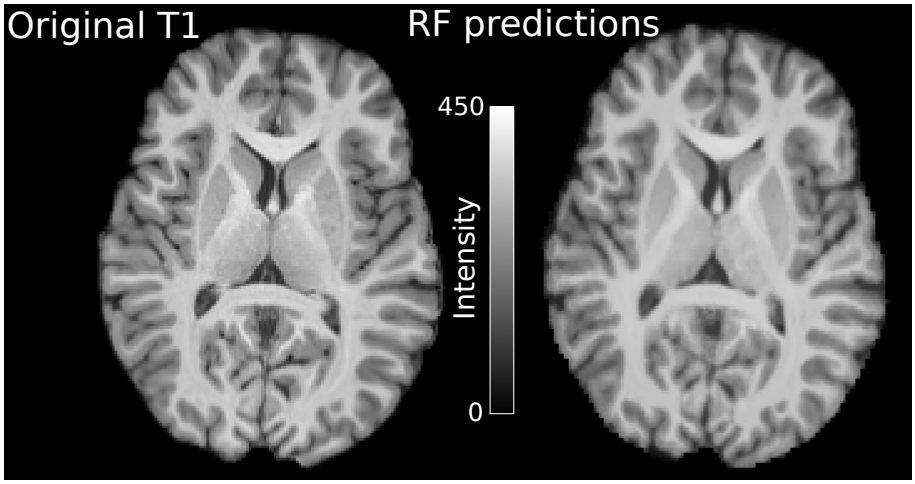


Figure 4.13: Prediction of T1w image intensities. Original T1w and T1w image predicted with the random forest from upsampled diffusion data

4.4.3 Application to FW modelling

We show here how the prior (derived from the RF predictions) on the CSF compartment affects the fitting of the FW model to the diffusion data. We used the RF predictions, as we previously showed that they are the most accurate. As a reference, we also included the results obtained with the T1-derived CSF PVFs, which are prone to registration error. Finally, we show the results obtained with the same model, without the prior on the CSF volume fraction.

In Figure 4.14, we show the FA and MD obtained in the regions where partial voluming between the CSF and the WM occurs. These regions are defined by a WM PVF ≥ 0.5, a CSF PVF ≥ 0.1, and a GM PVF $= 0$. We compare our fitting with an inversion recovery acquisition where the CSF signal was nulled (see Chapter 3). We indicated the correlation as well as the MSE, the latter being more sensitive to biased estimates.

Looking at the FA (right), the best correlation and RMSE between the two-

compartment model and the IR fit are obtained when we use the RF predictions (respectively $r = 0.66$ and $RMSE = 0.15$). Not using the prior on the CSF volume fraction decreases the accuracy of the FA estimates ($r = 0.62$ and $RMSE = 0.17$). Finally, using the T1 CSF partial volume estimates prove to be ineffective in improving the fitting accuracy ($r = 0.57$ and $RMSE = 0.19$).

The MD estimation is also more accurate and less biased when we use the RF PVEs as a prior. The correlation coefficient slightly increases to $r = 0.51$, compared to $r = 0.49$ without the PVE prior. The RMSE is also down to $0.27mm{\cdot}s^{-1}$, from $0.34mm{\cdot}s^{-1}$ when we do not set a PVE prior. When we use the T1 PVE as prior, the RMSE is $0.31mm{\cdot}s^{-1}$, while the correlation is relatively low ($r = 0.35$).

Importantly, when we used the RF CSF prior, the CSF PVF estimated in the deep WM was constant with a value of 0. When the prior is not used, the model estimates a non-zero FWF (on average 0.1) in crossing fibres areas, which affects the tissue MD and FA.

These results indicate that the fitting benefits from the extra information inferred with the RF approach. Importantly, we showed that the PVFs estimated from the T1 cannot be directly used when fitting the model on the diffusion data.

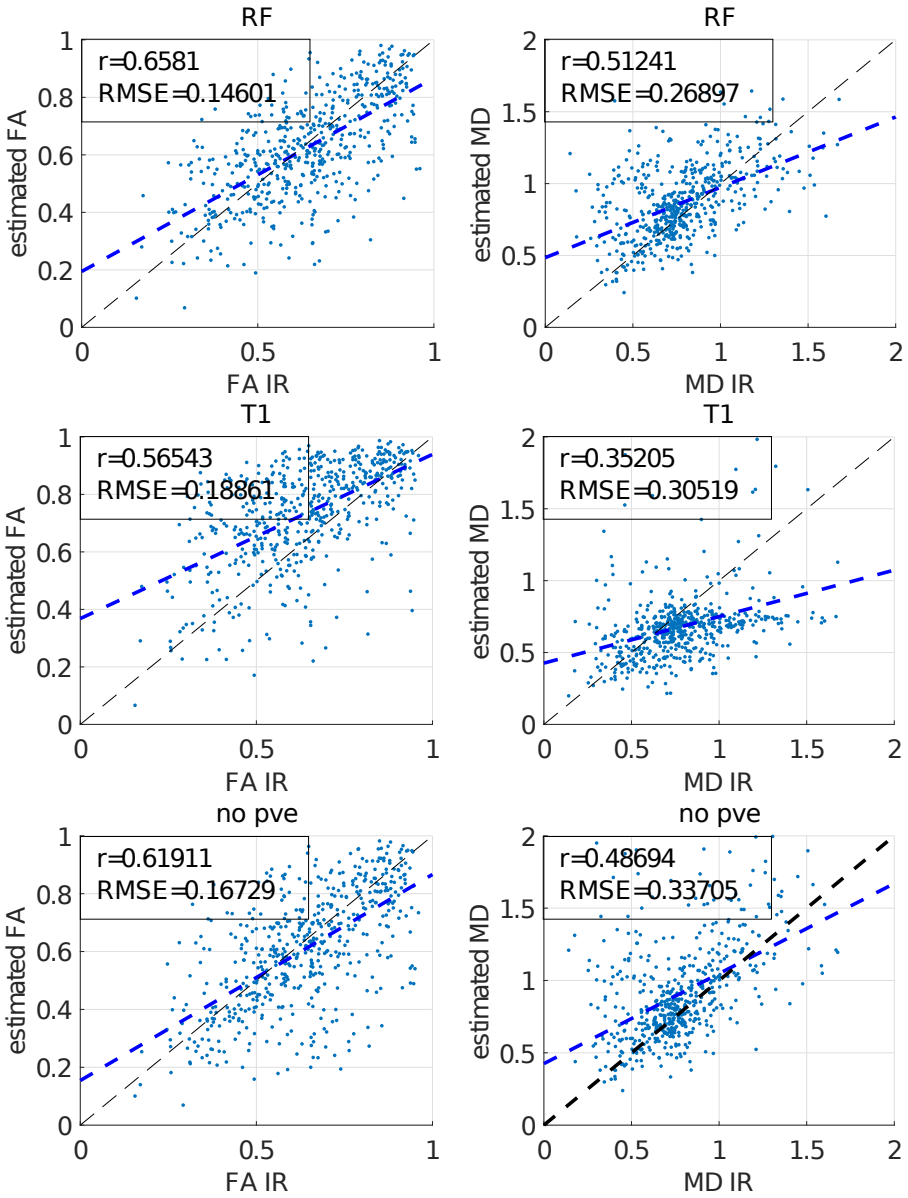


Figure 4.14: Comparison of the FA and MD obtained with the FW model and with the inversion recovery acquisition. Scatter plot of the FA (Left) and (MD) parameters obtained on the inversion recovery acquisition against FA and MD obtained with the FW model with different CSF tissue priors (Top: RF predictions, middle: T1 segmentation, bottom: no tissue prior

4.5 Discussion

4.5.1 Machine learning approaches

Amongst the 4 machine learning techniques we evaluated, we found that the random forest gives the most accurate predictions. It is closely followed by the NN method, while the SVR and GP are both under-performing. This result is not unexpected, as the RF approach is known to be very efficient, scalable, and effective for a wide range of supervised learning problems. In fact, several studies involving supervised ML in the context of diffusion MRI modelling successfully used random forests [Alexander et al., 2017] [Nedjati-Gilani et al., 2017]. The neural network approach has been shown to be promising as well, although it requires a more tedious tuning and is more computationally demanding than the RF. On the other hand, the NN approach may be more adequate to address highly non-linear problems, e.g. mapping directly the diffusion signal to NODDI model parameters [Golkov et al., 2016]. Additionally, the NN can be naturally extended to use the spatial neighbourhood with the convolutional neural networks (CNN). We tried to adapt an existing 3D CNN method that was originally designed for lesion segmentation [Kamnitsas et al., 2017]. Unfortunately, we faced implementation issues, as the optimisation of the weights in the context of regression proved to be difficult. Therefore, the method did not outperform our conventional NN. Additionally, such a spatial approach could potentially learn the mis-registration and/or the EPI distortion during the training phase.

4.5.2 Diffusion features

In this work, we employed diffusion features derived from well known models (diffusion tensor and ball and stick), as well as the tractography visitation map. One of the motivations was to make the method applicable to acquisitions with different parameters (e.g. number and orientation of diffusion gradients, resolution, number of b-values). Another motivation was that the features are rotationally independent, and therefore they do not require the training and testing datasets to be aligned.

Other studies also used features derived from an explicit diffusion model. In [Nedjati-Gilani et al., 2017], the diffusion tensor indices and quantities derived from the 4th order spherical harmonics were used. In their study, they inferred micro-structural properties derived from a multi-compartment permeability model [Lasič et al., 2011] that usually requires double diffusion encoding [Shemesh et al., 2016]. [Alexander et al., 2017] used the 6 independent elements of the diffusion tensor. They also included the neighbouring voxels to form cubic patches. We did try the same approach for our problem, but we found that adding the smoothed version and the gradient of the feature maps yielded better performance.

A model free approach was used in [Golkov et al., 2016]. The signal intensities from each diffusion gradient were directly used as features in order to infer NODDI and DKI parameters. While their approach appears to be promising, it is restricted to the case where the same diffusion protocol is used for an entire study.

4.5.3 Diffusion data segmentation

We successfully applied our ML approaches to partial volume segmentation. The segmentation maps appeared homogeneous, even though no spatial regularisation was

used. However, it heavily relies on the accuracy of the T1 segmentation, which was inaccurate, notably in the sub-cortical grey matter structures. Using a sub-cortical GM segmentation tool such as FIRST can provide good segmentations. Our method was unfortunately unable to properly segment those sub-cortical structures, as a significant WM PVF was still found.

Using a geometric regularisation might be of interest in our case, as it would explicitely model the grey-white matter boundary. We implemented a simple patch regularisation that fits a plane in each voxel and its 27 voxels neighbourhood and recomputes a voxel-wise volume fraction according to the fitted plane. Unfortunately, it showed inaccuracies in the sub-cortical structures, and it did not improve the partial volume estimates at the tissue interfaces. A more advanced geometric model, such as active shapes could be implemented in the future. In a similar way, a concurrent registration-segmentation method was proposed in [Esteban et al., 2016]. They used a set of active surfaces that were fitted to the multivariate diffusion data (FA and MD) as well as on the T1w image. The method fits the surfaces concurrently with the co-registration of the diffusion and T1w images, and it achieves sub-voxel accuracy. Adapting this approach to the PVFs or the synthetic T1 intensities predicted by our ML framework could potentially greatly improve our results.

4.5.4 Application to FW modelling

The first motivation of our work was to improve the fitting of a two-compartment model that accounts for CSF contamination. By adding an informative prior on the CSF volume fraction, we were able to prevent the model from fitting a CSF compartment in deep white matter. As a result, the micro-structural properties estimated in the tissue are not biased either by the MD prior or by the presence of a spurious CSF compartment.

Secondly, the CSF PVF prior enables to relax the prior on the MD. We showed that the extra information provided by the RF reduces the bias on the MD estimated by the model in areas where partial volume effects are present.

We note that our work could easily be applicable to a more complex diffusion model with a CSF compartment, such as NODDI [Zhang et al., 2012]. It also naturally extends to models with a GM compartment, such as multi-tissue spherical deconvolution [Jeurissen et al., 2014].

4.5.5 Implications for the future

The analysis of diffusion data was until recently exclusively based on an explicit model that often require advanced diffusion acquisition protocols. Our results showed that a machine learning approach can provide extra-information that is useful for solving ill-posed problems. A corollary is that the diffusion data contains redundant information. It was indeed demonstrated that a machine learning approach (neural networks) can reduce the amount of data needed to fit NODDI model by a factor of 12 [Golkov et al., 2016]. In another study, it was shown that using an ML based approach can infer NODDI micro-structural parameters more accurately than a direct fitting on the data [Reisert et al., 2017]. They showed that their method is robust to mesoscopic changes of the white matter (i.e., the geometric padding of the fibres).

We also demostrated the ability of ML learning approaches to achieve super-resolution, i.e., up-sampling the data in a more accurate way that standard interpolation. This was already shown in [Alexander et al., 2017], where the diffusion data is accurately up-sampled using RF. The same experiment was performed in [Tanno et al., 2017], except that RF were replaced by CNN, which allowed to push the super-resolution accuracy one step further. Such advances in the anatomical ac-

curacy of diffusion MRI could be greatly beneficial to tractography application, when thin pathways are to be retrieved. As tractography often relies on the anatomical parcellation of the GM, an accurate correspondence between diffusion and T1w space is desirable. It was shown in [Lorio et al., 2016] that the T1w image is affected by micro-structural properties of the tissues, and that it leads to an inaccurate WM/GM boundary. Our method to derive a synthetic T1w image directly from the diffusion data might provide a useful reference.

4.6 Conclusion

In this Chapter, we showed the potential of ML approaches to retrieve quantities from the diffusion that cannot be derived with an explicit model. On top of ensuring anatomical accuracy, our method also provides useful information that allows to retrieve unbiased tensor indices throughout the entire WM.

5

Tract specific micro-structural modelling

Although diffusion MRI provides quantitative measurements of water diffusion, the entangled nature of the fibres and the presence of multiple tissue confounds the estimation of biophysically meaningful information. Furthermore, the link between global fibre bundle characteristics and local microstructure models remains difficult to establish.

In this Chapter, we design a method that allows us to retrieve tract specific microstructural parameters and to perform cross-subject comparisons of these parameters. Our method proposes to identify the white matter fibre bundle via probabilistic tractography and to perform spatial regularisation in order to fit a simple, yet over-parametrised diffusion model. We validate our approach on simulated data and in-vivo high-quality acquisition from the HCP dataset. The results suggest that our regularisation scheme efficiently stabilises the fitting and achieves a good compromise between precision and bias.

Finally, we apply our framework on 100 subjects issued from the HCP dataset and establish a link between behavioural measurements and WM microstructure in well identified WM fibre bundles.

5.1 Introduction

The quantitative nature of the measurements in dMRI theoretically allows to infer biophysically interpretable parameters of the fibres constituting the white matter. Nevertheless, the entangled nature of the fibre bundles and their environment often confounds the parameters derived from micro-structural models. Notably, the prevalence of crossing fibres in the WM voxels has been estimated to be between 60 and 90% [Jeurissen et al., 2013], and voxels with tissue-type partial voluming between 35 and 50% [Roine et al., 2014].

These mesoscopic and macroscopic effects are entangled in the simple diffusion tensor case, affecting its derived indices such as fractional anisotropy (FA) or mean diffusivity (MD). Advances in dMRI acquisition, especially the ability to acquire more orientations and higher b-values, allow us to model complex fibre architecture more accurately. These models aim to represent the signal fraction and orientation dispersion of the fibre populations. A common representation of crossing fibres is the fibre orientation distribution (fODF), which estimates the orientation distribution of the fibres by deconvolving the diffusion signal with the average signal obtained from a single fibre [Tournier et al., 2004]. Although they prove to be very efficient for tractography applications, they are of limited interest when one is interested in the micro-structural properties of the white matter. In an attempt to quantify the white matter properties, it was proposed to use the fODF amplitude as a quantitative measure of apparent fibre density (AFD) [Raffelt et al., 2012]. The latter claims that the amplitude of the fODF peak relates to the intra-axonal volume fraction, i.e. the fibre density. This concept was expanded to the analysis of the shape of the fODF, in order to derive a measure not only related to the fibre density, but also to the radial diffusion properties (Hindrance Modulated Orientational Anisotropy - HMOA) [Dell'Acqua et al., 2013]. The AFD and HMOA both provide a orientation-specific

index, but their biophysical interpretation remains unspecific (e.g. myelination, axonal degeneration, atrophy).

Another approach was proposed in [Kaden et al., 2015], where the mesoscopic effects of fibre dispersion and crossing are integrated over all orientations. The technique was applied to multi-shell data in order to retrieve the average parallel and perpendicular diffusivity. It was later extended to a multi-compartment model that separates intra and extra-axonal diffusion [Kaden et al., 2016]. While this technique is appealing in areas where the fibre populations present similar properties, it does not allow to disentangle the properties between two crossing tracts. In this case, the parallel and perpendicular diffusivity will represent a weighted average of the two tracts.

In order to derive microstructural properties in the context of crossing fibres, a multi-tensor model can be used. This is a simple generalisation of the DTI model, that represents the diffusivity profile as a mixture of Gaussians. Unlike the simple DTI, the multi-tensor model cannot be expressed as a linear function of the measurements, and therefore requires non-linear optimisation. Nevertheless, the multi-tensor model proves to be an ill-posed problem, and the optimisation is hindered by the presence of local minima [Scherrer and Warfield, 2010] and cross-talk effects between fibres.

A typical approach to stabilizing the fitting is to take advantage of the spatial correlation across voxels with spatial regularisation. A simple regularisation scheme is the total variation, which aims at minimising the norm of the gradient of the tensor elements (eigenvalues, eigenvectors, and fibre volume fraction) [Pasternak et al., 2008]. The limitation of this approach is that the regularisation applies to parameters that are not expected to vary smoothly along the tracts (e.g. the volume fractions). Most importantly, the regularisation might enforce similarity between compartments which are not representing the same tract or tissue across voxels. In an effort to address this issue, it was proposed to perform the model fitting and tractography simultaneously [Malcolm et al., 2010] [Reisert et al., 2014]. Such approaches are

promising and they were also extended to more complex fibre representations such as intra and extra-cellular space [Reddy and Rathi, 2016] or the axonal diameter distribution [Girard et al., 2016].

The application of such multi-fibre representations is ambiguous in the context of group studies. It is indeed important that the fibre compartments in a given voxel are associated with the same respective tract across subjects. [Jbabdi et al., 2010] proposed to label the fibre compartments obtained with an fODF model (Ball and Sticks [Behrens et al., 2007]) according to their orientation. The labels are propagated across voxels sharing similar orientations and then matched across subjects. A recent approach proposed to perform the smoothing and statistical inference step along tracts identified with a probabilistic tractography atlas [Raffelt et al., 2015].

In this chapter, we propose to combine spatial regularisation and fibre labelling and to use this information during the fitting and group inference steps. Our purpose is two-fold. On the one hand, we aim to perform spatial regularisation that only applies to fibre compartments labelled as belonging to an individual tract. The key hypothesis here is that the micro-structural parameters of interest vary smoothly along a given tract. On the other hand, labelling the voxel compartments allows to perform tract-specific voxelwise comparisons across subjects.

We show through simulations and in-vivo experiments that our approach allows to recover the micro-structural properties unbiased by the presence of fibres or tissues that do not belong to the tract of interest. Additionally, we applied our method to a study with 100 subjects' data from the Human Connectome Project (HCP) dataset [Sotiropoulos et al., 2013]. We looked at potential correlations between the micro-structural properties of 35 tracts and a set of behavioural measurements.

5.2 Materials and methods

5.2.1 Model and spatial regularisation

In our method, the labelling of the fibre orientations within voxels is based on probabilistic tractography. The fibre orientations are first estimated using an fODF model (Ball and Stick [Behrens et al., 2003]) and then a multi-tensor model is fitted, with a spatial regularisation that solely constrains the tensor aligned with the tract of interest. Several steps are therefore required prior to performing model fitting.

Description of the pre-fitting steps The first step consists in fitting an fODF model independently in each voxel. We chose the Ball & Stick model, with three fibres and a Gamma distribution on the diffusivity [Jbabdi et al., 2012]. We also set an automatic relevance determination (ARD) prior on the fibres' volume fractions to avoid false positive orientations [Behrens et al., 2007]. The estimation was performed using Markov Chain Monte-Carlo (MCMC) sampling and the parameters of the algorithm were set to: a burn-in period of 1000 jumps and a sampling period of 1250 jumps with a sample drawn every 25 jumps. The GPU implementation of the method allowed for a fast and accurate fitting of the fODF parameters [Hernández et al., 2013].

In the second step, probabilistic tractography is performed using the probtrackx algorithm, also implemented on GPUs [Behrens et al., 2007] [Hernandez-Fernandez et al., 2016]. In this step, the tract has to be identified by a set of masks. In our in-vivo and group study experiments, the autoPtx package [de Groot et al., 2013] allowed us to delineate 35 fibre bundles that are represented in Chapter 2, Figure

2.2. Probabilistic tractography provides two important pieces of information. First, the normalised visitation map gives the probability of encountering the tract of interest in a given voxel, and thresholding this map provides a spatial extent of the tract. Second, computation of the voxel-wise tract mean orientation allows for the identification of the fibre compartment representing the tract.

In the third step, we compute the dot product between the mean orientation of the tract and each fODF peak. The peak with the closest orientation to the tract orientation is labelled as belonging to the tract.

In the fourth step, we fit an axially symmetric tensor, or zeppelin, for each estimated fibre orientation. The model can be written as:

$$S_i = S_0 \sum_k^n \left(f_k e^{-b_i((Dpar_k - Dperp_k)cos(\theta_{k,i})^2 + Dperp_k)} \right), with \sum_k^n (f_k) = 1 \qquad (5.1)$$

where S_0 is the unattenuated signal, S_i is the attenuated signal for the i^{th} gradient, n is the number of compartments, k the k^{th} compartment, b_i is the b-value for the i^{th} gradient, $\theta_{k,i}$ the angle between the k^{th} fibre orientation and the i^{th} gradient orientation, $Dpar_k$ the parallel diffusivity of the k^{th} compartment and $Dperp_k$ the perpendicular diffusivity of the k^{th} compartment.

The orientations are fixed to the fODF peaks estimated with the Ball and Sticks models, and the volume fractions and diffusivities are unknown.

Spatial regularisation We observed that the multi-zeppelin model benefits from regularisation, even in the context of multi-shell acquisition. We performed a spatial regularisation based on the total variation (TV). The TV regularisation aims to minimise the norm of the gradient of a given parameter map. We chose to use TV regularisation because it enforces spatial regularity of the parameters while preserving potential physiological discontinuities induces by, e.g., a pathology [Tschumperle and Deriche, 2003]. As our hypothesis is that the fibre properties (i.e., Dpar and Dperp) vary smoothly along the tract, the regularisation is only applied to the diffusivities

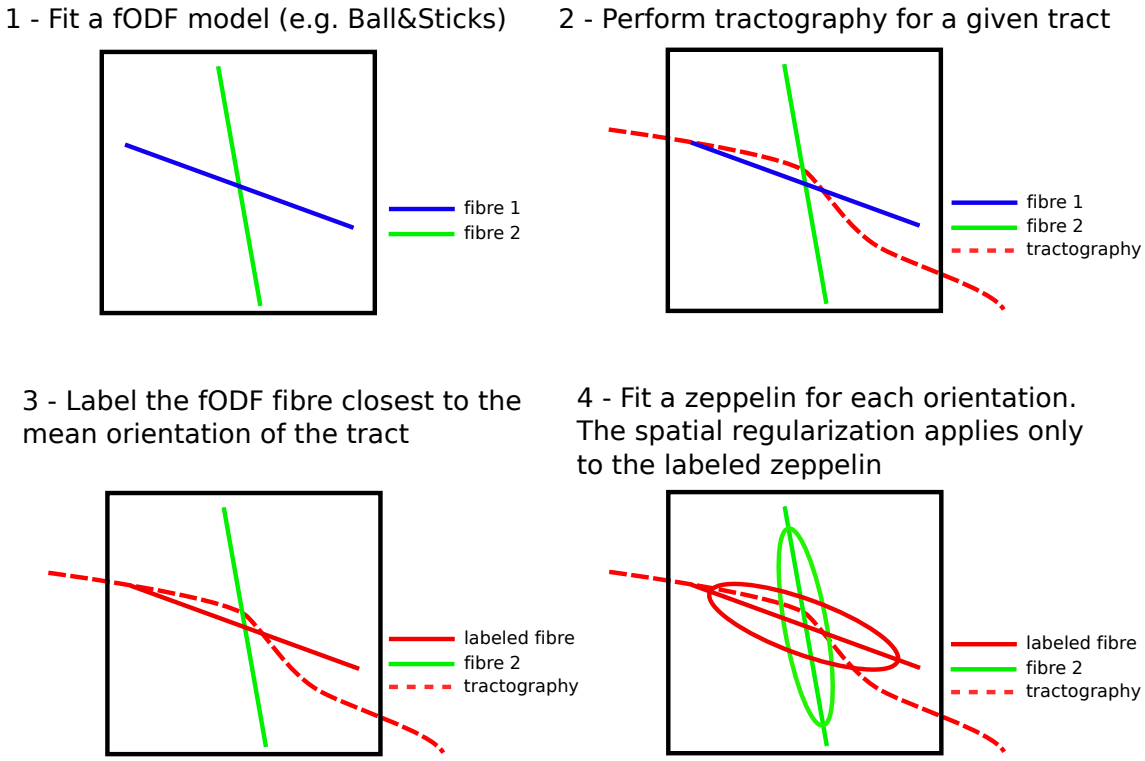


Figure 5.1: Schematic representation at the voxel level of the steps leading to fitting the spatially regularised model. The fibres' orientation distribution function is fitted (step 1). Then we perform a probabilistic tractography for a given tract (step 2). In step 3, we use the mean orientation of the tract to label the fibre of interest. Finally, a zeppelin is fitted for each fibre orientation, and only the labelled fibre is constrained by the spatial regularisation (step 4).

of the fibre compartment that was labelled as belonging to the tract. It is important to note that the other compartments are unconstrained and can therefore represent any other tract or tissue present in the voxel. The TV regularisation term can be written as:

$$TV = \sum_{\Omega} ||\nabla(Dpar_t)|| + \sum_{\Omega} ||\nabla(Dperp_t)|| \tag{5.2}$$

where Ω is the domain of integration for the TV regularisation (tract probability map thresholded to 0.5%) and $Dpar_t$, $Dperp_t$ are the parallel and perpendicular diffusivity of the zeppelin associated to the tract of interest. The cost function of the global optimisation problem is composed of the data fidelity term, i.e. the sum of squared differences between the data and our model predictions (Equation (5.1)),

and the regularisation term (Equation (5.2)).

$$min\left[(\sum_{\Omega}\sum_{i}\left((S_i-\hat{S}_i)^2\right)+\alpha TV\right] \tag{5.3}$$

where S_i is the signal predicted by the model for the i^{th} gradient orientation, $\hat{S}_i$ the measured signal for the i^{th} gradient orientation, and α is the regularisation weight parameter. Optimisation was performed with a quasi-Newton gradient descent scheme, with positivity constrains on zeppelin eigenvalues. We employed the L-BFGS method [Byrd et al., 1995] implemented in MATLAB to limit the memory consumption due to the high dimensionality of the parameter space. Furthermore, we used the analytical derivatives of Equation (5.3) to ensure fast convergence.

5.2.2 Simulations

We performed simulations on a 20*20*10 voxels volume comprising three intersecting tracts. The properties of the 3 tracts were respectively Dpar=1.6, 1.7, 1.9 $\times 10^{-3} mm^2.s^{-1}$, Dperp=0.3, 0.2, 0.3 $\times 10^{-3} mm^2.s^{-1}$, FA=0.79, 0.87, 0.83, and MD=0.73, 0.7, 0.83 $\times 10^{-3} mm^2.s^{-1}$. The signal was simulated using a diffusion encoding scheme similar the to HCP dataset [Sotiropoulos et al., 2013] and an SNR of 20.

Simulations allowed to determine the effect of the regularisation weight α and to estimate how the tract parameters Dpar and Dperp could be retrieved accurately and without bias. The model was fitted on the three tracts with the regularisation weight α ranging from 0 to 2. The three tracts were fitted independently, following the steps described in Figure 5.1. Finally, we computed the difference between the micro-structural parameters of the labelled zeppelin and the ground truth.

5.2.3 In-vivo

Our in-vivo experiments were conducted on data from one subject in the HCP dataset. The dMRI data was pre-processed for susceptibility and eddy current induced distortions [Andersson et al., 2003] [Glasser et al., 2013] [Andersson and Sotiropoulos, 2016]. We then applied our method, following the steps described in Figure 5.1. We performed tractography for 35 tracts delineated with the set of masks available in the autoPtx tool [de Groot et al., 2013], and presented in Chapter 2, Figure 2.2.

Similarly to our simulations, we varied the regularisation weight α from 0 to 10. In order to evaluate the accuracy of Dpar and Dper estimation, we considered the tract areas where neither tissue-type partial voluming nor fibre crossing was present as a baseline for unconfounded indices. We considered that a voxel did not contain crossing fibres when the ratio between the signal fraction of the first fibre and the total signal fraction of the three fibres was greater than $\frac{2}{3}$. The voxels without tissue-type partial voluming were identified on the co-registered T1w acquisition with FAST segmentation [Zhang et al., 2001].

5.2.4 Group study

We assessed the relevance and applicability of our method in the context of a group study. We chose the dMRI HCP dataset to correlate micro-structural changes with behavioural and psychological measurements.

To limit the computational cost, our study was restricted to 100 healthy male individuals from the dataset. Each subject went through a psychological and be-

havioural assessment. The latter yielded a set of 670 measurements, composing our design matrix. This matrix and the imaging data were deconfounded for age, handedness, weight and intracranial volume. We then demeaned and standardised the matrix to perform a PCA using singular value decomposition in order to reduce the dimensionality of the design matrix. As the main goal of the group study was to demonstrate the applicability and sensitivity of our approach, we restricted the statistical analysis to the first PCA component.

All subjects data were pre-processed by the HCP consortium [Glasser et al., 2013] and registered to a study-specific template in the MNI space following the method described in Chapter 2, Figure 2.1.

We fitted our model to the 35 tracts delineated with autoPtx, following the pipeline described in the previous section and in Figure 5.1. Given the in-vivo results for an individual subject, we set the level of regularisation to $\alpha = 0.4$. For each tract, we performed cross-subjects statistical testing on voxels where the tract was present in more that 90% of the subjects in MNI space (see Chapter 2, Table 2.4 and Figure 2.6). Although this reduces the spatial extent of the analysis, it should improve power and interpretability. The statistics were performed on the Dpar, Dperp, FA and MD derived from our approach and the simple DTI model.

We used permutation-based inference, as implemented in Randomise [Winkler et al., 2014]. The multiple comparisons correction was performed with family-wise error (FWE) and TFCE [Smith and Nichols, 2009]. The results were declared as significant for a p-value ≤ 0.05.

Additionally, we projected the voxels significantly correlated to our covariate of interest onto the GM represented as a surface. To this aim, we first averaged the connectivity matrices for 10 subjects in standard space (32k GM vertices $\times$ 200k WM voxels). Then, we thresholded the resulting matrix so that an element was marked as 1 if the target voxel was reached by at least 0.1% of the total number of streamlines from a given vertex. The projection was obtained via a multiplication

of the connectivity matrix with the WM significant voxels ROI. The results were normalised by the number of streamlines crossing the whole tract ROI. We finally obtained a surface where each vertex represents the ratio of streamlines crossing the significant voxels, for each tract.

5.3 Results

5.3.1 Simulations

Figure 5.2 shows the results of the fitting to the three simulated tracts. We can see the ground truth represented as zeppelins in the top left corner of Figure 5.2. The zeppelins are colour-coded with respect to their principal direction of diffusion. The two rectilinear tracts intersecting each other at 90 degrees are represented in yellow, while the circular tract colours represent all the possible orientations in the plane of the figure.

In the DTI case (top right of the figure), we observe that anisotropy is reduced in the crossing fibres areas, which are present in most of the voxels. In the centre of the image, the diffusion tensor presents a planar profile, its two first eigenvalues are similar, and the FA is low (0.4). We observe that the estimation of the principal direction of diffusion (first eigenvector) is also affected by the crossing fibres.

At the bottom left of Figure 5.2, we can see the result obtained with our multi-tensor model, without spatial regularisation. We only show the zeppelins that were labelled as representing each of the three tracts. The first observation is that the high FA of each fibre is retrieved in most cases. Also, the principal direction of diffusion is properly estimated. However, the fitting appears to be inaccurate in some voxels where two or more tracts intersect. In those cases, the perpendicular

diffusivity is overestimated.

At the bottom right of the figure, we can observe the result of the fitting with spatial regularisation, with $\alpha = 0.2$. In this case, the zeppelins present homogeneous perpendicular and parallel diffusivities.

The orientation of the zeppelin was estimated by the fODF fitting and we notice that it can slightly deviate from the ground truth when three fibres are present. We will quantify this effect in the in-vivo experiments.

In Figure 5.3, we illustrate the mean and standard deviation of the error between our model and the ground truth, averaged across the three tracts, with an increasing regularisation weight (from α=0 to 2).

For $\alpha = 0$ (no regularisation), there is a bias in the Dpar and Dperp estimation; they are both overestimated. This does not seem to affect the FA, but MD is overestimated.

With increasing α, we can observe that the error in Dpar and Dperp approaches zero, and that the standard deviation of the error decreases. In the range $\alpha = 0.1-0.5$ there is no bias in the estimated Dperp, while Dpar is slightly overestimated, resulting in an overestimated FA. For values of $\alpha > 0.5$, the bias in Dpar and Dperp gradually increases, and the precision remains fairly constant.

Looking at the MSE between the predicted signal and the data, we observe that for α between 0.1 and 0.5, spatial regularisation yields higher precision and lower bias than when no regularisation is applied.

5.3.2 In-vivo

This section presents the results obtained with our method on one HCP subject for 35 tracts identified by autoPtx.

In Figure 5.4, we illustrate the fitting obtained on the left superior longitudinal

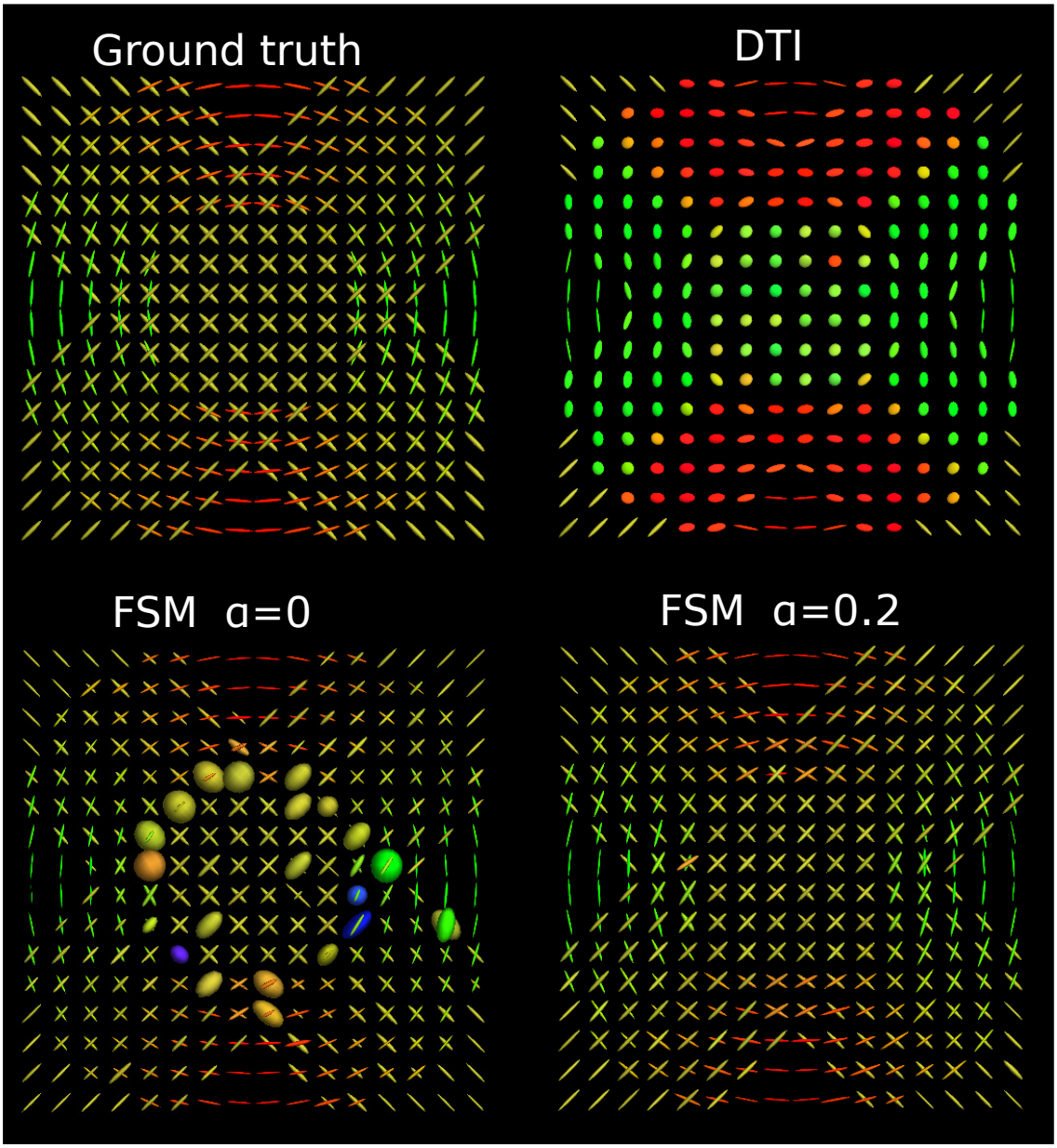


Figure 5.2: Map of the simulations results. Glyph representation of the simulated three tracts (top left), with the simple DTI model (top right), our fibre specific modelling (FSM) method without spatial regularisation ($\alpha = 0$, bottom left) and FSM with spatial regularisation ($\alpha = 0.2$, bottom right).

fasciculus (SLF - yellow) and left corticospinal tract (CST - light blue). Those two tracts intersect almost orthogonally. All voxels were fitted with three zeppelins, but only the zeppelins that were identified as belonging to the tract of interest are illustrated. Without spatial regularisation ($\alpha = 0$), the zeppelins show an irregular pattern, and the perpendicular and axial diffusivity are overestimated. When we

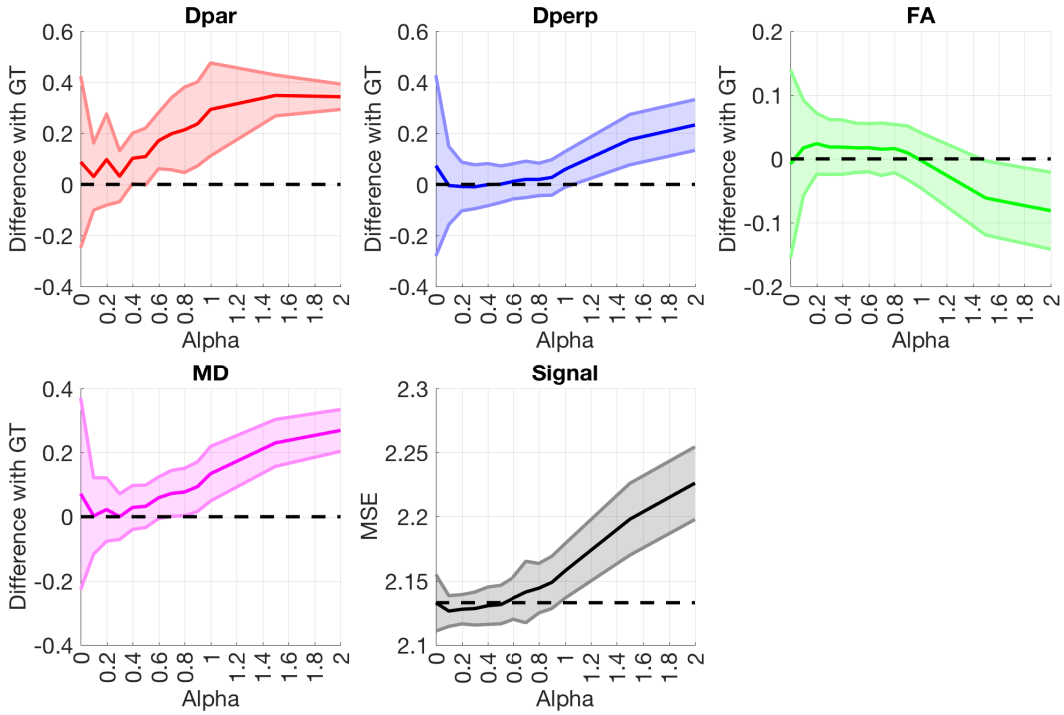


Figure 5.3: Plot of the error for micro-structural parameters and signal predictions. Mean difference and standard deviation of the error for micro-structural parameters as a function of the regularisation weight. On the right, the mean squared error between the model predictions and the simulated signal is plotted (black).

apply spatial regularisation (e.g. $\alpha = 0.5$), the zeppelin profile is more regular and less impacted by the fibre crossing. We note that although only two tracts are represented in the figure, the model accounts for any other tract or tissue present in the voxel.

Figure 5.5 shows that our model is able to deal with tissue-type partial voluming efficiently when spatial regularisation is used. The forceps minor is illustrated in the figure. Without spatial regularisation, the zeppelins show an isotropic profile at the interface of the tract and the ventricles, and also in areas where crossing fibres are present. With the spatial regularisation ($\alpha = 0.4$), this effect no longer occurs and the zeppelins appear homogeneous along the tract.

In Figure 5.6, we illustrate how the spatial regularisation affects the recovery of the micro-structural parameters. As explained in the methods section, the baseline is defined by areas without crossing fibres and without tissue-type partial voluming.

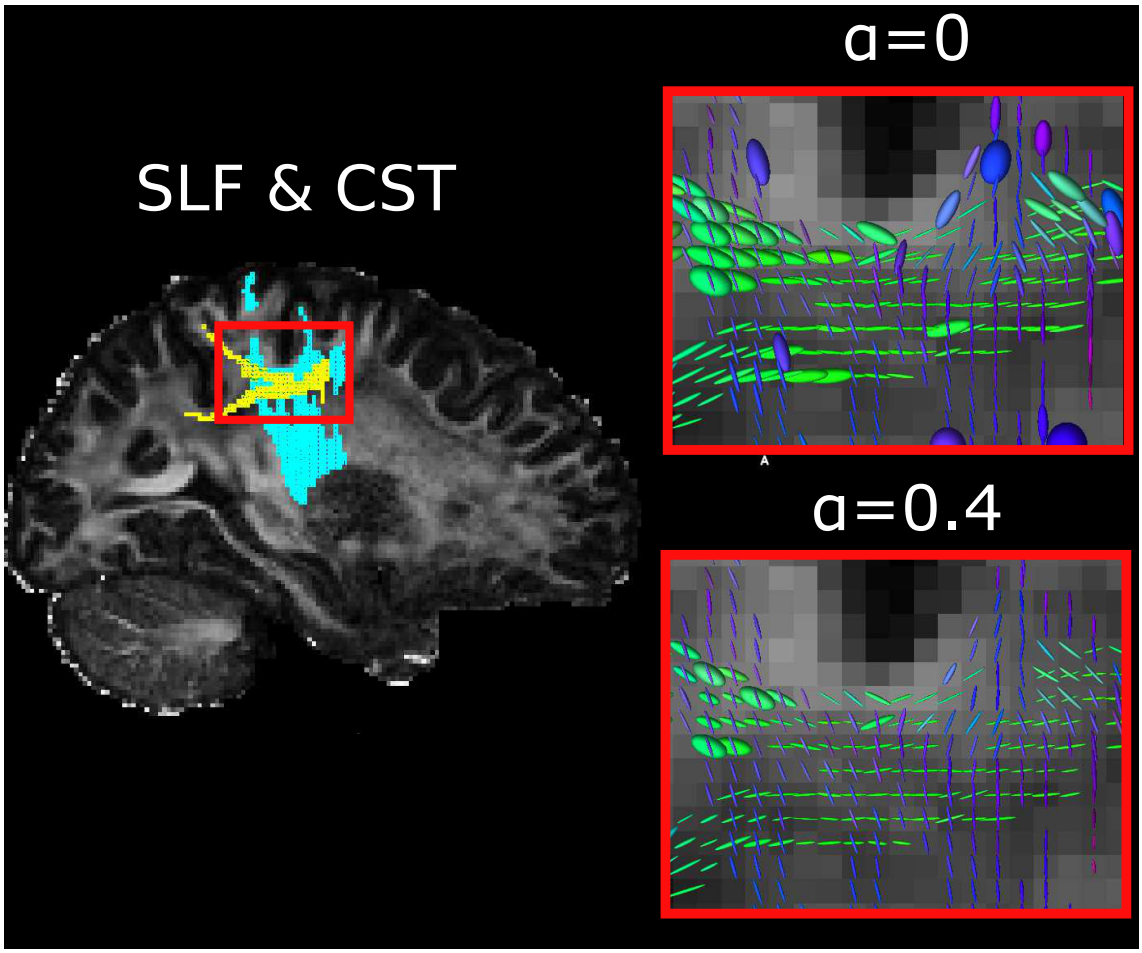


Figure 5.4: Glyph map of the zeppelin representing the superior longitudinal fasciculus (SLF) and the cortico-spinal tract (CST). Sagittal view representing the SLF (yellow) and CST (light blue), and a zoomed view of the glyph map for the zeppelin representing each tract. We display the results obtained with spatial regularisation ($\alpha = 0.4$) and without ($\alpha = 0$)

As expected from Figure 5.4 and 5.5, the parallel and perpendicular diffusivities are overestimated when we do not apply any spatial regularisation. As a result, the zeppelins have a lower FA and higher MD. This bias decreases as the amount of spatial regularisation increases, and stabilises for values of α between 0.3 and 0.7. When $\alpha > 0.7$, the bias gradually increases, i.e., the Dpar, Dperp and MD are overestimated while the FA is underestimated. The MSE between the predicted signal and the data remains constant for $\alpha < 0.6$, and then sharply increases. This indicates that although the spatial regularisation does not provide a better fit, it does not reduce the likelihood either, as long as the amount of regularisation α remains

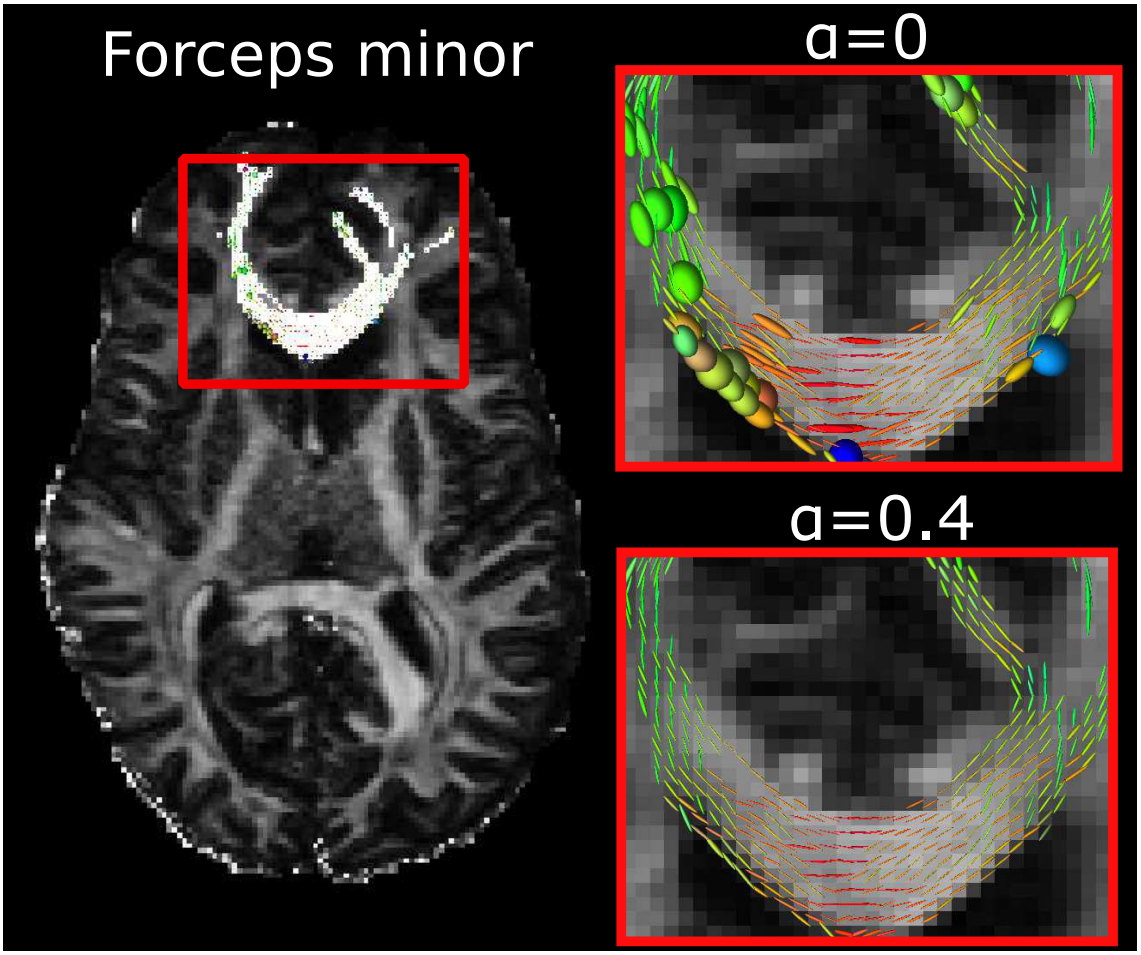


Figure 5.5: Glyph map of the zeppelin representing the forceps minor (FMI) . Axial slice representing the FMI (white), and a zoomed view of the glyph map for the zeppelin representing each tract. We display the results obtained with spatial regularisation $\alpha = 0.4$ and without $\alpha = 0$.

lower than 0.6. As in the simulations, it seems that there is a sweet spot (around $\alpha = 0.4$) for the amount of regularisation needed to retrieve the tract indices along the tract.

An important observation is the variability of the micro-structural indices across tracts, represented by the shaded area in Figure 5.6. In voxels unaffected by crossing fibres or tissue-type partial voluming, the standard deviation across tracts for Dpar and Dperp were respectively $\sigma_{\overline{Dpar}} = 0.078$ and $\sigma_{\overline{Dperp}} = 0.054 mm^2.s^{-1}$. Across the whole extent of the tracts, these values were $\sigma_{\overline{Dpar}} = 0.067$ and $\sigma_{\overline{Dperp}} = 0.104 mm^2.s^{-1}$ with $\alpha = 0$ and $\sigma_{\overline{Dpar}} = 0.079$ and $\sigma_{\overline{Dperp}} = 0.059 mm^2.s^{-1}$ with

$\alpha = 0.4$. This suggests that the spatial regularisation preserves variability of the micro-structural parameters across tracts.

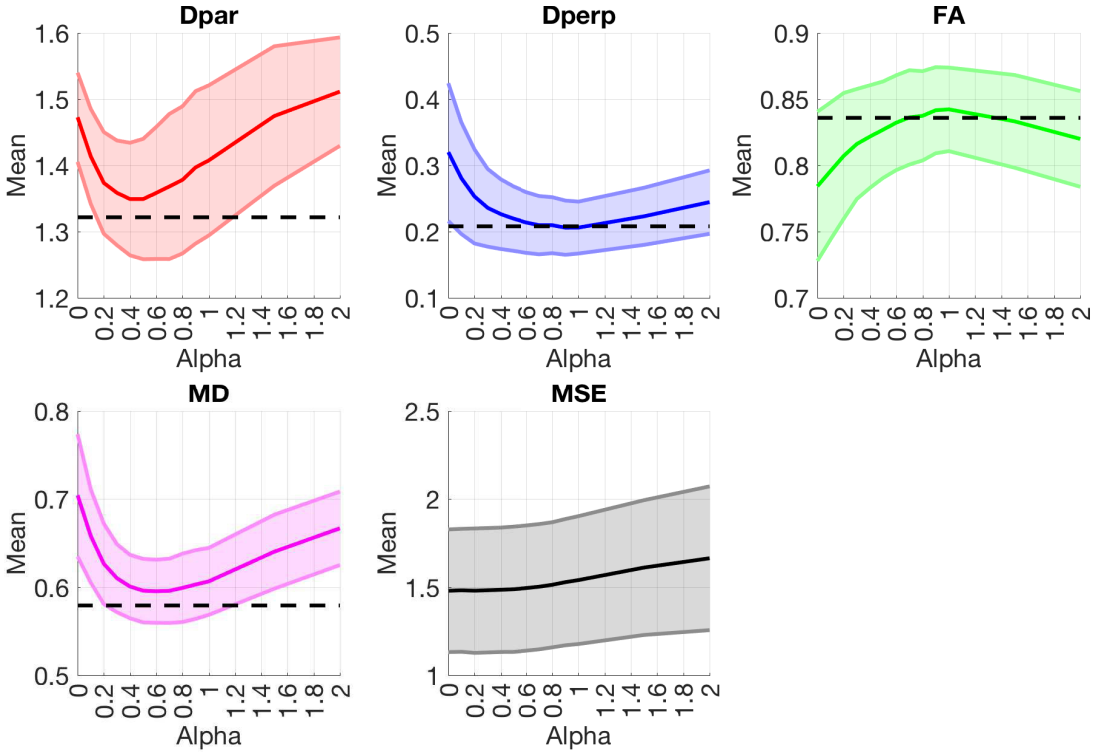


Figure 5.6: Plot of the mean and standard deviation for the micro-structural parameters and signal predictions MSE. The solid line represents the value of the parameter averaged across voxels and then across tracts. The dotted line represents the mean value of the parameter in areas where no crossing fibres or tissue-type partial voluming are present. The shaded area represent the standard deviation of the parameter across tracts.

5.3.3 Group study

Behavioural measurements In Table 5.1, we report the measurements with the ten strongest positive and negative correlations with the first PCA mode (variance explained: 16%). We observe that the measurements exhibiting a positive correlation are related to a decreased quality of life and mental health: depression, problems, and anxiety. Conversely, the measurements exhibiting a negative correlation are

related to a good mental health and quality of life, e.g., emotional and social support, life satisfaction, friendship. Therefore, we can argue that the first PCA component correlates with a negative emotional phenotype. We chose to keep only this first PCA component, as the other ones were harder to interpret and did not exhibit such a clear positive-negative behavioural trend. For conciseness, we will refer to the first PCA mode as "negativity".

Behavioural measurement	**Correlation with 1st PCA mode**
ASR Total Problems	0.91546
ASR Internalizing problems	0.84145
ASR Critical Items	0.79616
DSM Depressive Problems	0.7947
ASR Externalizing problems	0.75265
ASR Other Problems	0.73942
ASR Anxious/Depressed	0.72353
ASR Aggressive Behavior	0.71158
ASR Attention problems	0.70975
DSM Anxiety	0.70067
SSAGA Years of education	-0.30812
SSAGA Income	-0.32048
NIH Friendship	-0.37743
NIH Self-Efficacy	-0.39695
SSAGA Relationship	-0.39777
NIH Mean Purpose	-0.42145
NIH Positive Affect	-0.42944
NIH Life satisfaction	-0.46385
NIH Social support	-0.47901
NIH Emotional support	-0.56465

Table 5.1: Ten strongest positive and negative correlations of behavioural measurements with the first PCA mode ("negativity"). ASR: Achenbach Adult Self-Report for Ages 18-59 ([Achenbach, 2009], DSM: Diagnostic and Statistical Manual of Mental Disorders, SSAGA: Semi-Structured Assessment for the Genetics of Alcoholism, NIH: NIH Toolbox

Voxel-wise statistics In Table 5.2, we report how many voxels exhibited a significant negative correlation with "negativity". We included the results obtained with our approach as well as the results obtained with the simple DTI model. Only the tracts where at least one micro-structural parameter showed a negative significant

correlation with "negativity" are shown (not corrected for multiple comparisons across tracts).

With the fibre specific modelling approach (FSM), we found that 13 tracts had at least one voxel where one of the micro-structural parameters correlated with "negativity". The most sensitive parameter appeared to be the parallel diffusivity, as a total of 1058 voxels across 10 tracts exhibited a significant negative correlation with "negativity". The mean diffusivity was also negatively correlated with "negativity" in 879 voxels across 10 tracts. The tracts where the Dpar was increased corresponded to the tracts where the MD was also increased, except in the right anterior thalamic radiation and the inferior longitudinal fasciculus. The FA negatively correlated with "negativity" in only 4 tracts, with a total of 534 voxels. The perpendicular diffusivity showed a significant negative correlation in 5 tracts, for a total of 189 voxels. This last result is somehow surprising as an increase in perpendicular diffusivity is usually related to lack of myelination and decreased FA.

When looking at the results obtained with the simple DTI model, we first observe that the total number of voxels exhibiting a significant negative correlation with "negativity" is lower for all the micro-structural parameters, except for the FA. Overall, the results obtained with the DTI and FSM methods are consistent, although the former appears to be less sensitive.

The effect size for the tracts where at least one voxel was declared as significant is also shown in Table 5.2. It is expressed as the absolute correlation between "negativity" and the micro-structural parameter averaged across the voxels declared as significant. While this value cannot be interpreted as such (it is a circular result), it is valuable to compare the sensitivity of the FSM and DTI approaches. Generally, the effect size is higher with the FSM method, with the exception of PTR r for all DTI indices and with the CGC for the DTI FA. The highest effect size is observed with the MD obtained with the FSM method in the right cingulum (r=0.52) and in the right acoustic radiations (r=0.55).

	FSM				DTI			
Tract	**FA**	**MD**	**Dpar**	**Dperp**	**FA**	**MD**	**Dpar**	**Dperp**
AR l	0	29/0.55	6/0.49	76/0.55	0	2/0.45	1/0.42	2/0.45
AR r	0	0	0	0	0	0	0	2/0.36
ATR r	0	29/0.45	0	16/0.50	0	0	0	0
CGC l	421/0.43	304/0.46	436/0.49	0	543/0.48	0	176/0.42	0
CGC r	59/0.41	39/0.42	138/0.44	0	98/0.44	0	101/0.42	0
CGH l	0	0	0	0	26/0.39	0	0	0
CGH r	45/0.41	32/0.42	34/0.38	0	0	0	0	0
Cing l	3/0.47	88/0.42	116/0.44	0	177/0.46	0	0	0
Cing r	0	287/0.52	304/0.49	0	31/0.43	0	7/0.36	0
CST l	1/0.35	0	0	0	0	0	0	0
CST r	0	0	0	0	8/0.41	0	0	0
ILF r	0	0	2/0.40	0	0	0	0	0
MCP	0	25/0.41	9/0.39	11/0.47	0	0	0	0
ML r	0	0	0	0	19/0.40	0	72/0.49	0
PTR r	0	18/0.42	4/0.39	2/0.36	0	103/0.45	68/0.42	109/0.46
SLF2 l	0	0	0	62/0.39	0	0	0	0
SLF r	0	4/0.46	9/0.50	0	0	0	0	0
Total	534	879	1058	189	902	106	425	113

Table 5.2: Tracts showing significant negative correlation with "negativity". The first number indicates the number of voxels exhibiting a significant negative correlation with "negativity" and the second number is the effect size. The effect size (absolute correlation between the tract micro-structural parameter and "negativity") is shown for tracts where at least one voxel was declared as significant. The results were corrected for multiple comparisons within tracts using FWE and TFCE with a threshold $p < 0.05$.

In Figure 5.7, we display the voxels where the Dpar showed significant negative correlation with "negativity". We can see on this sagittal slice that the voxels with the lowest p-values are located in the central-posterior part of the cingulum, at the level of the junction with the white matter junction leading to the precuneus. Comparing the results obtained with FSM and the DTI method, we can observe that the former reveals more changes in micro-structural parameters. Indeed, the spatial extent of significant correlations is greater with FSM, as previously shown in Table 5.2. We also note that the significant correlations observed with the DTI model overlap with the ones observed with FSM. The voxels comprising the cingulum are marginally affected by crossing fibres, but the thin nature of the tract makes it prone to partial

voluming with the grey matter. The unconfounded micro-structural indices obtained with the FSM approach prove to be more sensitive to the white matter variations in this study.

We suggest that the results obtained with our FSM approach are more specific than those obtained with the DTI approach. In the FSM case, all micro-structural parameters are guaranteed to belong to the tract of interest. Additionally the microstructural parameters are not confounded by the presence of other tissue compartments (e.g., CSF and GM) or adjacent tracts.

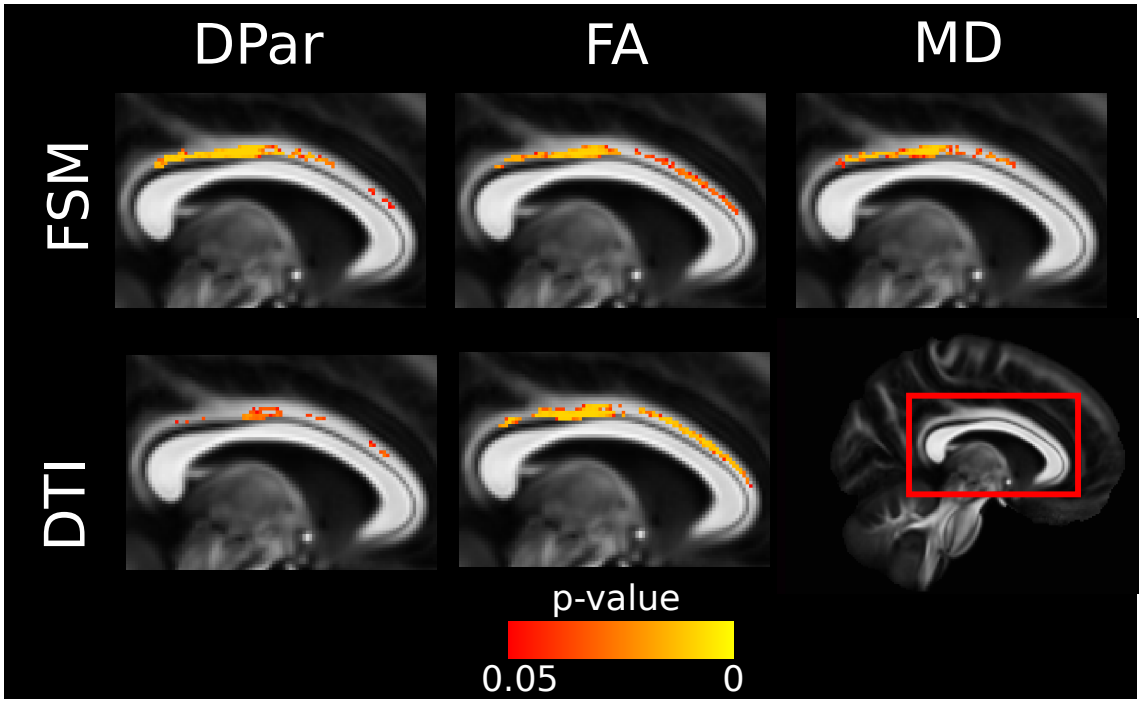


Figure 5.7: Left cingulum voxels showing significant correlation with "negativity". Left cingulum voxels exhibiting significant negative correlation between "negativity" and microstructural indicices (Dpar, FA, MD). We compare the results obtained with our FSM approach (top) and with the simple DTI (bottom). The results for MD DTI are not displayed as no voxel showed significant correlation. The p-values are corrected for multiple comparison with FWE and TFCE.

Projection of the results on the GM surface Using the connectivity matrix between the grey and white matter voxels, we projected the voxels showing significant correlation on the grey matter surface. In Figure 5.8, we show the proportion of the

left CGC streamlines passing through a voxel showing significant negative correlation between Dpar and "negativity". With FSM, the grey matter region most affected were located in the parahippocampal gyrus, the precuneus gyrus, and the frontal pole. The results obtained with the DTI model show a greatly reduced extent of the results projected on the grey matter surface. In this case, only a few vertices of the precuneus and parahippocampal gyri were affected by a change in Dpar.

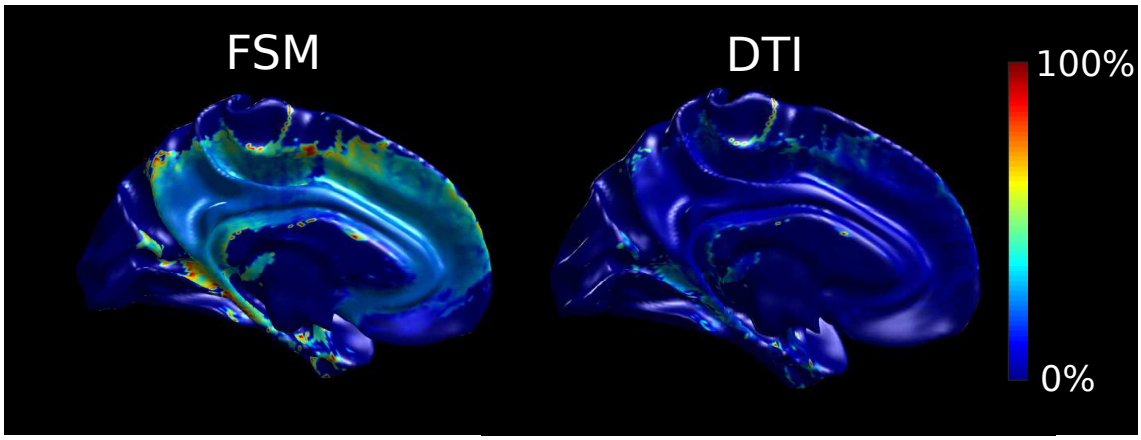


Figure 5.8: Projection of the cingulum voxels affected by a significant decrease in Dpar with "negativity". The colour bar represents the proportion of the left CGC streamlines passing through at least one voxel showing a significant negative correlation between Dpar and "negativity".

5.4 Discussion

The interpretation of population differences with scalar measures derived from the simple DTI model is challenging in regions with crossing fibres or tissue-type partial volume effects [Pierpaoli et al., 2001] [Alexander et al., 2001a] [Douaud et al., 2011]. In this Chapter, we described a method that allows us to retrieve micro-structural parameters that are tract-specific and unconfounded by the presence of crossing fibres and tissue-type partial voluming. Additionally, labelling the fibre compartments ensures that the parameters represent the same tract while performing cross-subject comparisons.

5.4.1 Spatial regularisation

We have demonstrated through simulations and in-vivo experiments the accuracy and precision of the fitting as a function of the regularisation parameter α. We identified an optimal value of α that allows the estimation of the Dpar and Dperp. Interestingly, we observed that the MSE between the model predictions and the data did not increase with the spatial regularisation. In the simulations, we found that spatial regularisation with the optimal value of α leads to a lower MSE. This means that the anatomically informed constraints on the parameters actually helps the optimisation method to find a better optimum in the parameter space. Simultaneous tractography and microstructural parameters estimation methods were also recently proposed, e.g. in [Reisert et al., 2014], [Daducci et al., 2015] and [Girard et al., 2015]. One of the major limitation of these global reconstruction methods is their reliance on the assumption that all the WM bundles can be reconstructed [Daducci et al., 2016], while in our method, the fitting is performed independently for each bundle. Yet, both approaches produce very promising results, and could potentially be leveraged to fix the parameter estimation degeneracy observed in multi-compartment models [Jelescu et al., 2016].

In the simulations, we used a fixed SNR to generate the signal. One could expect that the optimal value of α would vary with respect to the noise level. It would also be interesting to assess the applicability of our method to diffusion data with a limited number of orientations or with only one shell. While such experiments are beyond the scope of this Chapter, we expect our method to be applicable to data with lower quality than the HCP dataset.

A limitation of our approach is its strong dependence on the orientation estimated in the first step of the method (fODF estimation). The orientation of

the zeppelin representing the tract of interest is indeed fixed by the initial fODF orientation. It could be beneficial to rather fix it to the orientation of the tract defined by the tractography, which is spatially regularised by definition. We avoided this approach for two reasons. First, it would require fitting the orientations of the second and third zeppelins during the global optimisation scheme, which would be computationally expensive and would require further constrains to avoid overfitting when the fibre orientations become very close to each other. Second, we found in the simulation and in-vivo experiments that the estimation of the fibres orientation with the Ball and Stick model proves to be robust and accurate when we compare it to the orientations estimated by the tractography.

5.4.2 Specificity of the model

The use of the orientation and tract probability derived from tractography allows us to confidently identify the fibre representing a given fibre bundle. As such, we suggest that the micro-structural parameters derived with the FSM approach are specific to the tract of interest.

The tract-specific nature of our approach also allows us to retrieve the micro-structural parameters (Dpar and Dperp) with a limited bias induced by the presence of other tracts or tissues. We have shown in our in-vivo experiments that our method retrieves the micro-structural parameter values found in regions where the signal is not affected by other tracts or tissues. As such, one can interpret those parameters correctly, as well as the FA and the MD [Wheeler-Kingshott and Cercignani, 2009].

In our in-vivo experiments, we found that fibre properties vary across tracts. This questions the validity of spherical deconvolution approaches. The latter heavily relies on the calibration of the fibre response function to estimate the fODF [Tax et al., 2014]. The assumption of its homogeneity throughout the white matter affects

the amplitude of the fODF and can produce spurious peaks [Parker et al., 2013].

A zeppelin represents the signal generated by a straight fibre bundle quite well, but it assumes that the diffusivities of the extra and intra-axonal space are the same. However, it is known that the dMRI signal presents a non-monoexponential decay [Assaf and Cohen, 1998]. Our choice of a single compartment (zeppelin) to represent the signal from a single fibre bundle was motivated by its simplicity and low computational cost. We suggest that our framework could be extended to more complex representation of fibres microstructure parameters, e.g. the intra-cellular space (NODDI) [Zhang et al., 2012] or the axon diameter distribution (AxCaliber) [Alexander et al., 2010].

5.4.3 Sensitivity to micro-structural changes

Spatial regularisation is often associated with a smoothing of the parameter maps, thus possibly reducing the sensitivity to localised changes of the white matter. In our simulations, the tract properties were fixed across space, and therefore we could not evaluate how the spatial regularisation would affect the sensitivity of the model. We suggest that performing simulations that include a pathological process would enable to tune our model, and notably the spatial regularisation parameter, more accurately.

We observed in the in-vivo experiments that the spatial regularisation allows variations of the diffusion parameters both along the tract of interest and across all the tracts. The absence of ground truth prevented us from any quantification of the potential loss of sensitivity caused by the spatial regularisation. Still, we found in the group study that our FSM approach was more sensitive to changes in the white matter micro-structure than the simple DTI model, whether or not we applied spatial smoothing to the latter.

5.4.4 Group study

The clinical relevance of the group study results was not of primary importance in this work. The chapter was primarily aimed at demonstrating the applicability of our approach in the context of a neuroscience imaging experiment. Importantly, the reasonable computational requirements of our approach (between 10 and 20 minutes per tract and per subject) allowed its application on a dataset of average size (100 subjects in our case) with high angular and spatial resolution. We voluntarily restricted our sample size to 100 subjects, but a more efficient implementation of the fitting method (e.g. in C++) could allow for its application to larger dataset, e.g., the 1200 HCP subjects, or the 100 000 subjects in the Biobank initiative [Miller et al., 2016].

From a neuroscience point of view, we obtained some potentially interesting results, relating "negativity" to white matter micro-structure. In spite of the reduced power due to the restricted number of subjects (100 over 1200), we found a significant negative correlation between "negativity" and axial diffusivity in the left cingulum bundle. The latter is at the core of the limbic system, which supports a variety of functions including behaviour, motivation, and emotions. A microstructural alteration of the cingulum bundle has been identified in previous studies on major depressive syndrome [Kieseppä et al., 2010] [Cullen et al., 2010]. Our findings can also be related to a recent resting-state fMRI study on 461 HCP subjects, relating behavior and functional connectivity [Smith et al., 2015]. It revealed a strong correlation between a specific pattern of functional connectivity and a "positive-negative" mode of the behavioural measurements, the latter resembling to our "negativity" mode.

5.5 Conclusion

In this Chapter, we designed a method that allows for a truly tract-specific micro-structural parameter estimation in any white matter pathway identifiable by tractography. Additionally, we demonstrated the relevance and sensitivity of our approach in a group-study performed on a state of the art dataset. Future work could benefit from our framework, e.g., to integrate microstructure modelling and connectivity.

6

General conclusion

This thesis was focused on the theme of voxelwise cross-subject comparison of diffusion MRI data. Our contribution was to explore new approaches addressing some pitfalls in dMRI group studies, such as registration, tissue type partial-volume effects and crossing fibres. In the chapters 2-5, we assessed how the suggested improvements affect the specificity and sensitivity of the methods to white matter changes.

Spatial normalisation In the first part of our work, we found that improved registration has an impact on the results of a typical group study, and that it likely provides a more sensitive detection of WM changes occurring in Alzheimer's patients. Additionally, it increases the cross-subject tract overlap, which is important when comparing diffusion data in a tract specific fashion, as we did in Chapter 5.

We have shown that image resolution has a significant influence on registration accuracy. However, the registration methods we described in this thesis operate at the voxel level. Alternatives using continuous representations of individual tracts were previously proposed to drive the registration process [Leemans et al., 2006] [Wassermann et al., 2011]. Such approaches may be particularly powerful with low resolution data.

A potential future avenue could be integration with other MRI modalities, e.g.

high resolution T1. This may be beneficial not only for improving registration, but may also enable multi-modal group inference by having a common space across subjects and across modalities. Another potential route of multimodal registration is to combine diffusion and functional MRI. It has been recently shown that driving cortical registration using fMRI data (using surface-based multimodal registration as proposed in [Robinson et al., 2014]) improves overlap of functional areas across subjects. However, such cortical registrations are currently incompatible with volumetric registrations of the white matter. An interesting future avenue of research would be to combine surface-based (and fMRI-driven) with volumetric (structurally driven) registrations, to ensure for example that when functional areas are matched, their corresponding anatomical connections are also matched across subjects.

Multi-compartment models In Chapter 3, we demonstrated that using a Bayesian framework with an informative prior allows for an accurate fitting of the ill-posed FW model. Quantitative comparison of our approach with the pre-existing reference method [Pasternak et al., 2009] shows that the latter introduces a significantly higher bias in the estimates of the tensor parameters.

Importantly, we found that the FW compartment modelled along with a DT is unlikely associated with the extra-cellular space. This result is in contradiction with previous studies where the FW compartment was hypothesised to represent extra-cellular space, e.g. in [Oestreich et al., 2016] [Pasternak et al., 2016] [Hoy et al., 2017]. Our claim is supported by [Zhang et al., 2012] [Reisert et al., 2017], where it was also suggested that a non-zero FWF in deep WM likely arises out of constraints on the model parameters or insufficient data. We therefore recommend to discard the FW compartment in the deep WM by setting an ARD prior or anatomical priors, (e.g. derived from the data-driven segmentation described in Chapter 4).

Our work on the FW model illustrates how spurious correlation between model parameters can bias their estimation. This problem is unfortunately affecting most of

the multi-compartment models [Jelescu et al., 2016], and it is often overlooked when interpreting the outcome of group studies in terms of micro-structural changes. We believe that addressing the degeneracy of biophysical model parameter estimation is currently one of the biggest challenges in the dMRI field. In an attempt to address this issue, several innovative approaches have been recently explored, e.g. combining diffusion and relaxometry [Veraart et al., 2017], multiple or time varying diffusion encoding gradients [Özarslan and Basser, 2008] [Westin et al., 2016], or merging tractography with micro-structure modelling [Daducci et al., 2015].

Data driven methods in diffusion MRI In Chapter 4, we explored the ability of data-driven methods to predict tissue volume fractions (WM, GM and CSF) directly from the diffusion data. The predicted volume fractions were found to be more helpful than the co-registered T1 volume fractions in fitting the FW model. This lead to a more accurate estimation of the DT indices and eliminated the FW compartment from deep white matter. Another application of this work is to address the mis-registration across MR modalities, potentially helping with multi-modal integration.

Interest in machine learning approaches has recently raised within the dMRI community. These methods are able to learn representations of the dMRI data that cannot be explicitly modelled. Several applications of these have already been proposed, such as reduction in the acquisition time [Golkov et al., 2016] [Alexander et al., 2017], disentangling meso and micro-structure [Reisert et al., 2017] and achieving super-resolution [Tanno et al., 2017]. These studies still rely on supervised learning, limiting their application to cases where the ground truth is available or where the signal can be simulated with a forward model. However, the outstanding advances in machine learning techniques might enable us to explore further the relation between the diffusion signal and microstructure.

Tract-specific microstructural modelling In Chapter 5, we presented an approach to deriving tract-specific micro-structural measurements that are unconfounded by the presence of crossing fibres or non-WM tissues. We have addressed the ill-posed nature of the model with a regularisation performed in a tract-wise fashion. On the one hand, this ensures that the model parameters are not over-constrained as the regularisation relies on the assumption of a smooth variation of the microstructural parameters along the tract. It follows that our approach reaches a good compromise between precision and bias of the parameter estimation. On the other hand, the voxel-wise labelling of the fibre compartments allows to match the tracts of interest across subjects, thus ensuring specificity of the results.

Chapter 5 constitutes a good summary of our efforts to enhance dMRI group studies as it combines improved registration methods, partial volume modelling and tractography. In the group study conducted on the HCP data, we showed that our method provides results that are both more specific and sensitive to WM changes than the standard DTI approach.

We suggest that our method would benefit from an extension to multi-compartment models representing, e.g., the extra and intra-cellular spaces or axon diameter distributions [Girard et al., 2017]. Finally, we believe that merging tractography and microstructural parameters estimation is very promising as the two problems can benefit from each other. The integration of these two major applications of dMRI could potentially lead to quantitative connectomes reflecting the strength of the connections across the brain [Craddock et al., 2013] [Sotiropoulos and Zalesky, 2017].

Bibliography

[Achenbach, 2009] Achenbach, T. M. (2009). *Achenbach system of empirically based assessment (ASEBA): Development, findings, theory, and applications*. University of Vermont, Research Center of Children, Youth & Families.

[Alexander et al., 2001a] Alexander, A. L., Hasan, K. M., Lazar, M., Tsuruda, J. S., and Parker, D. L. (2001a). Analysis of partial volume effects in diffusion-tensor MRI. *Magnetic Resonance in Medicine*, 45(5):770–780.

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

[Alexander and Gee, 2000] Alexander, D. C. and Gee, J. C. (2000). Elastic Matching of Diffusion Tensor Images. *Computer Vision and Image Understanding*, 77(2):233–250.

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

[Alexander et al., 2017] Alexander, D. C., Zikic, D., Ghosh, A., Tanno, R., Wottschel, V., Zhang, J., Kaden, E., Dyrby, T. B., Sotiropoulos, S. N., Zhang, H., and Criminisi, A. (2017). Image quality transfer and applications in diffusion MRI. *NeuroImage*, 152:283–298.

[Andersson, 2008] Andersson, J. L. (2008). Maximum a posteriori estimation of diffusion tensor parameters using a Rician noise model: Why, how and but. *NeuroImage*, 42(4):1340–1356.

[Andersson et al., 2007] Andersson, J. L., Jenkinson, M., Smith, S., et al. (2007). Non-linear registration, aka Spatial normalisation FMRIB technical report. *FMRIB Analysis Group of the University of Oxford*.

[Andersson and Sotiropoulos, 2016] Andersson, J. L. and Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage*, 125:1063–1078.

[Andersson et al., 2017] Andersson, J. L. R., Graham, M. S., Drobnjak, I., Zhang, H., Filippini, N., and Bastiani, M. (2017). Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *NeuroImage*, 152(Supplement C):450–466.

[Andersson et al., 2016] Andersson, J. L. R., Graham, M. S., Zsoldos, E., and Sotiropoulos, S. N. (2016). Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*, 141(Supplement C):556–572.

[Andersson and Skare, 2002] Andersson, J. L. R. and Skare, S. (2002). A Model-Based Method for Retrospective Correction of Geometric Distortions in Diffusion-Weighted EPI. *NeuroImage*, 16(1):177–199.

[Andersson et al., 2003] Andersson, J. L. R., Skare, S., and Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage*, 20(2):870–888.

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

[Ashburner and Friston, 2000] Ashburner, J. and Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage*, 11(6):805–821.

[Ashburner and Friston, 2005] Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3):839–851.

[Ashtari et al., 2007] Ashtari, M., Cervellione, K. L., Hasan, K. M., Wu, J., McIlree, C., Kester, H., Ardekani, B. A., Roofeh, D., Szeszko, P. R., and Kumra, S. (2007). White matter development during late adolescence in healthy males: A cross-sectional diffusion tensor imaging study. *NeuroImage*, 35(2):501–510.

[Assaf and Basser, 2005] Assaf, Y. and Basser, P. J. (2005). Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage*, 27(1):48–58.

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

[Assaf and Cohen, 1998] Assaf, Y. and Cohen, Y. (1998). Non-mono-exponential attenuation of water and N-acetyl aspartate signals due to diffusion in brain tissue. *Journal of Magnetic Resonance (San Diego, Calif.: 1997)*, 131(1):69–85.

[Assaf and Cohen, 2000] Assaf, Y. and Cohen, Y. (2000). Assignment of the water slow-diffusing component in the central nervous system using q-space diffusion MRS: Implications for fiber tract imaging. *Magnetic Resonance in Medicine*, 43(2):191–199.

[Assaf et al., 2004] Assaf, Y., Freidlin, R. Z., Rohde, G. K., and Basser, P. J. (2004). New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine*, 52(5):965–978.

[Avants et al., 2008] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41.

[Bach et al., 2014] Bach, M., Laun, F. B., Leemans, A., Tax, C. M. W., Biessels, G. J., Stieltjes, B., and Maier-Hein, K. H. (2014). Methodological considerations on tract-based spatial statistics (TBSS). *NeuroImage*, 100:358–369.

[Barth et al., 2016] Barth, M., Breuer, F., Koopmans, P. J., Norris, D. G., and Poser, B. A. (2016). Simultaneous multislice (SMS) imaging techniques. *Magnetic Resonance in Medicine*, 75(1):63–81.

[Basser et al., 1994] Basser, P., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267.

[Basser et al., 2000] Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44(4):625–632.

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

[Behrens et al., 2007] Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*, 34(1):144–155.

[Bennett et al., 2009] Bennett, C. M., Miller, M. B., and Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *NeuroImage; Amsterdam*, 47:S125.

[Berlot et al., 2014] Berlot, R., Metzler-Baddeley, C., Jones, D. K., and O'Sullivan, M. J. (2014). CSF contamination contributes to apparent microstructural alterations in mild cognitive impairment. *NeuroImage*, 92:27–35.

[Breiman, 1984] Breiman, L. (1984). *Classification and regression trees*. Wadsworth International Group.

[Breiman, 1996] Breiman, L. (1996). *Out-of-bag estimation*. Technical report, Statistics Department, University of California Berkeley, Berkeley CA 94708, 1996b. 33, 34.

[Breiman, 2001] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

[Brown, 1828] Brown, R. (1828). A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine*, 4(21):161–173.

[Byrd et al., 1995] Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

[Calamante et al., 2010] Calamante, F., Tournier, J.-D., Jackson, G. D., and Connelly, A. (2010). Track-density imaging (TDI): Super-resolution white matter imaging using whole-brain track-density mapping. *NeuroImage*, 53(4):1233–1243.

[Catani and Thiebaut de Schotten, 2008] Catani, M. and Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, 44(8):1105–1132.

[Chang et al., 2005] Chang, L.-C., Jones, D. K., and Pierpaoli, C. (2005). RESTORE: Robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine*, 53(5):1088–1095.

[Chou et al., 2005] Chou, M.-C., Lin, Y.-R., Huang, T.-Y., Wang, C.-Y., Chung, H.-W., Juan, C.-J., and Chen, C.-Y. (2005). FLAIR Diffusion-Tensor MR Tractography: Comparison of Fiber Tracking with Conventional Imaging. *American Journal of Neuroradiology*, 26(3):591–597.

[Clark et al., 2003] Clark, C. A., Barrick, T. R., Murphy, M. M., and Bell, B. A. (2003). White matter fiber tracking in patients with space-occupying lesions of the brain: a new technique for neurosurgical planning? *NeuroImage*, 20(3):1601–1608.

[Clark and Le Bihan, 2000] Clark, C. A. and Le Bihan, D. (2000). Water diffusion compartmentation and anisotropy at high b values in the human brain. *Magnetic Resonance in Medicine*, 44(6):852–859.

[Collins et al., 1994] Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. (1994). Automatic 3d intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, 18(2):192–205.

[Concha et al., 2005] Concha, L., Gross, D. W., and Beaulieu, C. (2005). Diffusion Tensor Tractography of the Limbic System. *American Journal of Neuroradiology*, 26(9):2267–2274.

[Copenhaver et al., 2006] Copenhaver, B. R., Rabin, L. A., Saykin, A. J., Roth, R. M., Wishart, H. A., Flashman, L. A., Santulli, R. B., McHugh, T. L., and Mamourian, A. C. (2006). The fornix and mammillary bodies in older adults with Alzheimer's disease, mild cognitive impairment, and cognitive complaints: A volumetric MRI study. *Psychiatry Research: Neuroimaging*, 147(2–3):93–103.

[Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

[Craddock et al., 2013] Craddock, R. C., Jbabdi, S., Yan, C.-G., Vogelstein, J. T., Castellanos, F. X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., and Milham, M. P. (2013). Imaging human connectomes at the macroscale. *Nature Methods*, 10(6):524–539.

[Cullen et al., 2010] Cullen, K. R., Klimes-Dougan, B., Muetzel, R., Mueller, B. A., Camchong, J., Houri, A., Kurma, S., and Lim, K. O. (2010). Altered White Matter Microstructure in Adolescents with Major Depression: A Preliminary Study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49(2):173–83.e1.

[Daducci et al., 2016] Daducci, A., Dal Palú, A., Descoteaux, M., and Thiran, J.-P. (2016). Microstructure Informed Tractography: Pitfalls and Open Challenges. *Frontiers in Neuroscience*, 10.

[Daducci et al., 2015] Daducci, A., Palù, A. D., Lemkaddem, A., and Thiran, J. P. (2015). COMMIT: Convex Optimization Modeling for Microstructure Informed Tractography. *IEEE Transactions on Medical Imaging*, 34(1):246–257.

[Dale et al., 1999] Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, 9(2):179–194.

[de Groot et al., 2013] de Groot, M., Vernooij, M. W., Klein, S., Ikram, M. A., Vos, F. M., Smith, S. M., Niessen, W. J., and Andersson, J. L. R. (2013). Improving alignment in Tract-based spatial statistics: Evaluation and optimization of image registration. *NeuroImage*, 76:400–411.

[Dell'Acqua et al., 2013] Dell'Acqua, F., Simmons, A., Williams, S. C., and Catani, M. (2013). Can spherical deconvolution provide more information than fiber orientations? Hindrance modulated orientational anisotropy, a true-tract specific index to characterize white matter diffusion. *Human Brain Mapping*, 34(10):2464–2483.

[Descoteaux et al., 2009] Descoteaux, M., Deriche, R., Knosche, T. R., and Anwander, A. (2009). Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions. *IEEE Transactions on Medical Imaging*, 28(2):269–286.

[Douaud et al., 2011] Douaud, G., Jbabdi, S., Behrens, T. E. J., Menke, R. A., Gass, A., Monsch, A. U., Rao, A., Whitcher, B., Kindlmann, G., Matthews, P. M., and Smith, S. (2011). DTI measures in crossing-fibre areas: Increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer's disease. *NeuroImage*, 55(3):880–890.

[Drucker et al., 1997] Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., and Vapnik, V. (1997). Support vector regression machines. In *Advances in neural information processing systems*, pages 155–161.

[Duarte-Carvajalino et al., 2013] Duarte-Carvajalino, J. M., Sapiro, G., Harel, N., and Lenglet, C. (2013). A Framework for Linear and Non-Linear Registration of Diffusion-Weighted MRIs Using Angular Interpolation. *Frontiers in Neuroscience*, 7.

[Eaton-Rosen et al., 2016] Eaton-Rosen, Z., Melbourne, A., Cardoso, M. J., Marlow, N., and Ourselin, S. (2016). Beyond the Resolution Limit: Diffusion Parameter Estimation in Partial Volume. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Lecture Notes in Computer Science, pages 605–612. Springer, Cham.

[Einstein, 1905] Einstein, A. (1905). Uber die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560.

[Ennis and Kindlmann, 2006] Ennis, D. B. and Kindlmann, G. (2006). Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magnetic Resonance in Medicine*, 55(1):136–146.

[Esteban et al., 2016] Esteban, O., Zosso, D., Daducci, A., Bach-Cuadra, M., Ledesma-Carbayo, M. J., Thiran, J.-P., and Santos, A. (2016). Surface-driven registration method for the structure-informed segmentation of diffusion MR images. *NeuroImage*, 139:450–461.

[Fick, 1855] Fick, A. (1855). Ueber Diffusion. *Annalen der Physik*, 170(1):59–86.

[Fieremans et al., 2011] Fieremans, E., Jensen, J. H., and Helpern, J. A. (2011). White matter characterization with diffusional kurtosis imaging. *NeuroImage*, 58(1):177–188.

[Fillard et al., 2011] Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., Yo, T., Mangin, J.-F., and Poupon, C. (2011). Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. *NeuroImage*, 56(1):220–234.

[Fischer and Glass, 2007] Fischer, L. R. and Glass, J. D. (2007). Axonal degeneration in motor neuron disease. *Neuro-Degenerative Diseases*, 4(6):431–442.

[Girard et al., 2017] Girard, G., Daducci, A., Petit, L., Thiran, J.-P., Whittingstall, K., Deriche, R., Wassermann, D., and Descoteaux, M. (2017). AxTract: Toward microstructure informed tractography. *Human Brain Mapping*, 38(11):5485–5500.

[Girard et al., 2016] Girard, G., Daducci, A., Whittingstall, K., Deriche, R., Wassermann, D., and Descoteaux, M. (2016). Microstructure driven tractography in the human brain. In *Organization for Human Brain Mapping (OHBM).*, Proceedings of: Human Brain Mapping (HBM)., Geneva, Switzerland.

[Girard et al., 2015] Girard, G., Fick, R., Descoteaux, M., Deriche, R., and Wassermann, D. (2015). AxTract: Microstructure-Driven Tractography Based on the Ensemble Average Propagator. In *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, pages 675–686. Springer, Cham.

[Glasser et al., 2013] Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., Jenkinson, M., and WU-Minn HCP Consortium (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–124.

[Golkov et al., 2016] Golkov, V., Dosovitskiy, A., Sperl, J. I., Menzel, M. I., Czisch, M., Sämann, P., Brox, T., and Cremers, D. (2016). q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. *IEEE Transactions on Medical Imaging*, 35(5):1344–1351.

[Gorman, 2014] Gorman, J. (2014). Learning How Little We Know About the Brain. *The New York Times*.

[Greve and Fischl, 2009] Greve, D. N. and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1):63–72.

[Guimond et al., 2000] Guimond, A., Meunier, J., and Thirion, J.-P. (2000). Average Brain Models: A Convergence Study. *Computer Vision and Image Understanding*, 77(2):192–210.

[Gur et al., 2008] Gur, Y., Pasternak, O., and Sochen, N. (2008). Fast GL(n)-Invariant Framework for Tensors Regularization. *International Journal of Computer Vision*, 85(3):211–222.

[Hauser and Oksenberg, 2006] Hauser, S. L. and Oksenberg, J. R. (2006). The Neurobiology of Multiple Sclerosis: Genes, Inflammation, and Neurodegeneration. *Neuron*, 52(1):61–76.

[Hernandez-Fernandez et al., 2016] Hernandez-Fernandez, M., Reguly, I., Giles, M., Jbabdi, S., Smith, S., and Sotiropoulos, S. (2016). A fast and flexible toolbox for tracking brain connections in diffusion MRI datasets using GPUs. In *Proceedings of the 22nd Annual Meeting of the Organization for Human Brain Mapping (OHBM), Geneva*.

[Hernández et al., 2013] Hernández, M., Guerrero, G. D., Cecilia, J. M., García, J. M., Inuggi, A., Jbabdi, S., Behrens, T. E. J., and Sotiropoulos, S. N. (2013). Accelerating Fibre Orientation Estimation from Diffusion Weighted Magnetic Resonance Imaging Using GPUs. *PLoS ONE*, 8(4):e61892.

[Hinton et al., 2012] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580 [cs]*. arXiv: 1207.0580.

[Hoy et al., 2014] Hoy, A. R., Koay, C. G., Kecskemeti, S. R., and Alexander, A. L. (2014). Optimization of a free water elimination two-compartment model for diffusion tensor imaging. *NeuroImage*, 103:323–333.

[Hoy et al., 2017] Hoy, A. R., Ly, M., Carlsson, C. M., Okonkwo, O. C., Zetterberg, H., Blennow, K., Sager, M. A., Asthana, S., Johnson, S. C., Alexander, A. L., and Bendlin, B. B. (2017). Microstructural white matter alterations in preclinical Alzheimer's disease detected using free water elimination diffusion tensor imaging. *PLOS ONE*, 12(3):e0173982.

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

[Jbabdi et al., 2012] Jbabdi, S., Sotiropoulos, S. N., Savio, A. M., Graña, M., and Behrens, T. E. J. (2012). Model-based analysis of multishell diffusion MR data for tractography: How to get over fitting problems. *Magnetic Resonance in Medicine*, 68(6):1846–1855.

[Jelescu et al., 2016] Jelescu, I. O., Veraart, J., Fieremans, E., and Novikov, D. S. (2016). Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue. *NMR in biomedicine*, 29(1):33–47.

[Jensen et al., 2005] Jensen, J. H., Helpern, J. A., Ramani, A., Lu, H., and Kaczynski, K. (2005). Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 53(6):1432–1440.

[Jeurissen et al., 2013] Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., and Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human Brain Mapping*, 34(11):2747–2766.

[Jeurissen et al., 2014] Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., and Sijbers, J. (2014). Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage*, 103:411–426.

[Jezzard and Balaban, 1995] Jezzard, P. and Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, 34(1):65–73.

[Jezzard et al., 1998] Jezzard, P., Barnett, A. S., and Pierpaoli, C. (1998). Characterization of and correction for eddy current artifacts in echo planar diffusion imaging. *Magnetic Resonance in Medicine*, 39(5):801–812.

[Jones, 2003] Jones, D. K. (2003). Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magnetic Resonance in Medicine*, 49(1):7–12.

[Jones, 2004] Jones, D. K. (2004). The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: A Monte Carlo study†. *Magnetic Resonance in Medicine*, 51(4):807–815.

[Jones et al., 2005] Jones, D. K., Symms, M. R., Cercignani, M., and Howard, R. J. (2005). The effect of filter size on VBM analyses of DT-MRI data. *NeuroImage*, 26(2):546–554.

[Jung, 2014] Jung, K.-J. (2014). Removal of partial volume averaging with free water in MR diffusion tensor imaging using inversion recovery for b0 image. *Magnetic Resonance Imaging*, 32(6):619–624.

[Kaden et al., 2016] Kaden, E., Kelm, N. D., Carson, R. P., Does, M. D., and Alexander, D. C. (2016). Multi-compartment microscopic diffusion imaging. *NeuroImage*, 139:346–359.

[Kaden et al., 2007] Kaden, E., Knösche, T. R., and Anwander, A. (2007). Parametric spherical deconvolution: inferring anatomical connectivity using diffusion MR imaging. *NeuroImage*, 37(2):474–488.

[Kaden et al., 2015] Kaden, E., Kruggel, F., and Alexander, D. C. (2015). Quantitative mapping of the per-axon diffusion coefficients in brain white matter. *Magnetic Resonance in Medicine*, pages n/a–n/a.

[Kamnitsas et al., 2017] Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., and Glocker, B. (2017). Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78.

[Keihaninejad et al., 2012] Keihaninejad, S., Ryan, N. S., Malone, I. B., Modat, M., Cash, D., Ridgway, G. R., Zhang, H., Fox, N. C., and Ourselin, S. (2012). The Importance of Group-Wise Registration in Tract Based Spatial Statistics Study of Neurodegeneration: A Simulation Study in Alzheimer's Disease. *PLOS ONE*, 7(11):e45996.

[Kieseppä et al., 2010] Kieseppä, T., Eerola, M., Mäntylä, R., Neuvonen, T., Poutanen, V.-P., Luoma, K., Tuulio-Henriksson, A., Jylhä, P., Mantere, O., Melartin, T., Rytsälä, H., Vuorilehto, M., and Isometsä, E. (2010). Major depressive disorder and white matter abnormalities: a diffusion tensor imaging study with tract-based spatial statistics. *Journal of Affective Disorders*, 120(1-3):240–244.

[Klein et al., 2009] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R. P., Mann, J. J., and Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802.

[Kubicki et al., 2005] Kubicki, M., McCarley, R. W., and Shenton, M. E. (2005). Evidence for white matter abnormalities in schizophrenia. *Current opinion in psychiatry*, 18(2):121–134.

[Kumazawa et al., 2010] Kumazawa, S., Yoshiura, T., Honda, H., Toyofuku, F., and Higashida, Y. (2010). Partial volume estimation and segmentation of brain tissue based on diffusion tensor MRI. *Medical Physics*, 37(4):1482–1490.

[Kärger, 1985] Kärger, J. (1985). NMR self-diffusion studies in heterogeneous systems. *Advances in Colloid and Interface Science*, 23(Supplement C):129–148.

[Larsson et al., 1992] Larsson, H. B. W., Thomsen, C., Frederiksen, J., Stubgaard, M., and Henriksen, O. (1992). In vivo magnetic resonance diffusion measurement in the brain of patients with multiple sclerosis. *Magnetic Resonance Imaging*, 10(1):7–12.

[Lasič et al., 2011] Lasič, S., Nilsson, M., Lätt, J., Ståhlberg, F., and Topgaard, D. (2011). Apparent exchange rate mapping with diffusion MRI. *Magnetic Resonance in Medicine*, 66(2):356–365.

[Le Bihan and Breton, 1985] Le Bihan, D. and Breton, E. (1985). Imagerie de diffusion in-vivo par résonance magnétique nucléaire. *Comptes-Rendus de l'Académie des Sciences*, 93(5):27–34.

[Le Bihan et al., 1986] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., and Laval-Jeantet, M. (1986). MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407.

[Le Bihan, 1995] Le Bihan, D. L. (1995). Molecular diffusion, tissue microdynamics and microstructure. *NMR in Biomedicine*, 8(7):375–386.

[Lebel et al., 2008] Lebel, C., Walker, L., Leemans, A., Phillips, L., and Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, 40(3):1044–1055.

[Leemans and Jones, 2009] Leemans, A. and Jones, D. K. (2009). The B-matrix must be rotated when correcting for subject motion in DTI data. *Magnetic Resonance in Medicine*, 61(6):1336–1349.

[Leemans et al., 2006] Leemans, A., Sijbers, J., De Backer, S., Vandervliet, E., and Parizel, P. (2006). Multiscale white matter fiber tract coregistration: A new feature-based approach to align diffusion tensor data. *Magnetic Resonance in Medicine*, 55(6):1414–1423.

[Lorio et al., 2016] Lorio, S., Kherif, F., Ruef, A., Melie-Garcia, L., Frackowiak, R., Ashburner, J., Helms, G., Lutti, A., and Draganski, B. (2016). Neurobiological origin of spurious brain morphological changes: A quantitative MRI study. *Human Brain Mapping*, 37(5):1801–1815.

[Mackay, 1995] Mackay, D. J. C. (1995). Probable networks and plausible predictions — a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6(3):469–505.

[Maier-Hein et al., 2014] Maier-Hein, K. H., Westin, C.-F., Shenton, M. E., Weiner, M. W., Raj, A., Thomann, P., Kikinis, R., Stieltjes, B., and Pasternak, O. (2014). Widespread white matter degeneration preceding the onset of dementia. *Alzheimer's & Dementia*.

[Malcolm et al., 2010] Malcolm, J. G., Shenton, M. E., and Rathi, Y. (2010). Filtered multi-tensor tractography. *IEEE transactions on medical imaging*, 29(9):1664–1675.

[Mansfield, 1977] Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55.

[Metzler-Baddeley et al., 2012] Metzler-Baddeley, C., O'Sullivan, M. J., Bells, S., Pasternak, O., and Jones, D. K. (2012). How and how not to correct for CSF-contamination in diffusion MRI. *NeuroImage*, 59(2):1394–1403.

[Miller et al., 2016] Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., Andersson, J. L. R., Griffanti, L., Douaud, G., Okell, T. W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P. M., and Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, 19(11):1523–1536.

[Mori et al., 1999] Mori, S., Crain, B. J., Chacko, V. P., and Van Zijl, P. C. M. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2):265–269.

[Moseley et al., 1990] Moseley, M. E., Cohen, Y., Mintorovitch, J., Chileuitt, L., Shimizu, H., Kucharczyk, J., Wendland, M. F., and Weinstein, P. R. (1990). Early detection of regional cerebral ischemia in cats: Comparison of diffusion- and T2-weighted MRI and spectroscopy. *Magnetic Resonance in Medicine*, 14(2):330–346.

[Mulkern et al., 2009] Mulkern, R. V., Haker, S. J., and Maier, S. E. (2009). On high b diffusion imaging in the human brain: ruminations and experimental insights. *Magnetic Resonance Imaging*, 27(8):1151–1162.

[Murray, 2012] Murray, R. (2012). The 'most complex thing in the universe'. *BBC News*.

[Nair and Hinton, 2010] Nair, V. and Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, pages 807–814, USA. Omnipress.

[Nedjati-Gilani et al., 2017] Nedjati-Gilani, G. L., Schneider, T., Hall, M. G., Cawley, N., Hill, I., Ciccarelli, O., Drobnjak, I., Wheeler-Kingshott, C. A. M. G., and Alexander, D. C. (2017). Machine learning based compartment models with permeability for white matter microstructure imaging. *NeuroImage*, 150:119–135.

[Oestreich et al., 2016] Oestreich, L. K. L., Pasternak, O., Shenton, M. E., Kubicki, M., Gong, X., McCarthy-Jones, S., and Whitford, T. J. (2016). Abnormal white matter microstructure and increased extracellular free-water in the cingulum bundle associated with delusions in chronic schizophrenia. *NeuroImage: Clinical*, 12:405–414.

[Oouchi et al., 2007] Oouchi, H., Yamada, K., Sakai, K., Kizu, O., Kubota, T., Ito, H., and Nishimura, T. (2007). Diffusion Anisotropy Measurement of Brain White Matter Is Affected by Voxel Size: Underestimation Occurs in Areas with Crossing Fibers. *American Journal of Neuroradiology*, 28(6):1102–1106.

[Papadakis et al., 2002] Papadakis, N. G., Martin, K. M., Mustafa, M. H., Wilkinson, I. D., Griffiths, P. D., Huang, C. L.-H., and Woodruff, P. W. (2002). Study of the effect of CSF suppression on white matter diffusion anisotropy mapping of healthy human brain. *Magnetic Resonance in Medicine*, 48(2):394–398.

[Park et al., 2003] Park, H.-J., Kubicki, M., Shenton, M. E., Guimond, A., McCarley, R. W., Maier, S. E., Kikinis, R., Jolesz, F. A., and Westin, C.-F. (2003). Spatial normalization of diffusion tensor MRI using multiple channels. *NeuroImage*, 20(4):1995–2009.

[Parker et al., 2013] Parker, G. D., Marshall, D., Rosin, P. L., Drage, N., Richmond, S., and Jones, D. K. (2013). A pitfall in the reconstruction of fibre ODFs using spherical deconvolution of diffusion MRI data. *NeuroImage*, 65(Supplement C):433–448.

[Parker et al., 2003] Parker, G. J. M., Haroon, H. A., and Wheeler-Kingshott, C. A. M. (2003). A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *Journal of magnetic resonance imaging: JMRI*, 18(2):242–254.

[Pasternak et al., 2008] Pasternak, O., Assaf, Y., Intrator, N., and Sochen, N. (2008). Variational multiple-tensor fitting of fiber-ambiguous diffusion-weighted magnetic resonance imaging voxels. *Magnetic Resonance Imaging*, 26(8):1133–1144.

[Pasternak et al., 2016] Pasternak, O., Kubicki, M., and Shenton, M. E. (2016). In vivo imaging of neuroinflammation in schizophrenia. *Schizophrenia Research*, 173(3):200–212.

[Pasternak et al., 2012a] Pasternak, O., Shenton, M. E., and Westin, C.-F. (2012a). Estimation of Extracellular Volume from Regularized Multi-shell Diffusion MRI. In Ayache, N., Delingette, H., Golland, P., and Mori, K., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, number 7511 in Lecture Notes in Computer Science, pages 305–312. Springer Berlin Heidelberg.

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

[Pasternak et al., 2012b] Pasternak, O., Westin, C.-F., Bouix, S., Seidman, L. J., Goldstein, J. M., Woo, T.-U. W., Petryshen, T. L., Mesholam-Gately, R. I., McCarley, R. W., Kikinis, R., Shenton, M. E., and Kubicki, M. (2012b). Excessive Extracellular Volume Reveals a Neurodegenerative Pattern in Schizophrenia Onset. *The Journal of Neuroscience*, 32(48):17365–17372.

[Pasternak et al., 2015] Pasternak, O., Westin, C.-F., Dahlben, B., Bouix, S., and Kubicki, M. (2015). The extent of diffusion MRI markers of neuroinflammation and white matter deterioration in chronic schizophrenia. *Schizophrenia Research*.

[Patenaude et al., 2011] Patenaude, B., Smith, S. M., Kennedy, D. N., and Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*, 56(3):907–922.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830.

[Peterson and Fujinami, 2007] Peterson, L. K. and Fujinami, R. S. (2007). Inflammation, Demyelination, Neurodegeneration and Neuroprotection in the Pathogenesis of Multiple Sclerosis. *Journal of neuroimmunology*, 184(1-2):37–44.

[Pfefferbaum and Sullivan, 2003] Pfefferbaum, A. and Sullivan, E. V. (2003). Increased brain white matter diffusivity in normal adult aging: relationship to anisotropy and partial voluming. *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 49(5):953–961.

[Pierpaoli et al., 2001] Pierpaoli, C., Barnett, A., Pajevic, S., Chen, R., Penix, L., Virta, A., and Basser, P. (2001). Water Diffusion Changes in Wallerian Degeneration and Their Dependence on White Matter Architecture. *NeuroImage*, 13(6):1174–1185.

[Pierpaoli and Jones, 2004] Pierpaoli, C. and Jones, D. (2004). Removing CSF contamination in brain DT-MRIs by using a two-compartment tensor model. In *ISMRM*, page 1215.

[Povlishock, 1992] Povlishock, J. T. (1992). Traumatically induced axonal injury: pathogenesis and pathobiological implications. *Brain pathology (Zurich, Switzerland)*, 2(1):1–12.

[Raffelt et al., 2012] Raffelt, D., Tournier, J.-D., Rose, S., Ridgway, G. R., Henderson, R., Crozier, S., Salvado, O., and Connelly, A. (2012). Apparent Fibre Density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. *NeuroImage*, 59(4):3976–3994.

[Raffelt et al., 2015] Raffelt, D. A., Smith, R. E., Ridgway, G. R., Tournier, J.-D., Vaughan, D. N., Rose, S., Henderson, R., and Connelly, A. (2015). Connectivity-based fixel enhancement: Whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres. *NeuroImage*, 117:40–55.

[Rasmussen and Williams, 2006] Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*, volume 1. MIT press Cambridge.

[Reddy and Rathi, 2016] Reddy, C. P. and Rathi, Y. (2016). Joint Multi-Fiber NODDI Parameter Estimation and Tractography Using the Unscented Information Filter. *Frontiers in Neuroscience*, 10.

[Reisert et al., 2017] Reisert, M., Kellner, E., Dhital, B., Hennig, J., and Kiselev, V. G. (2017). Disentangling micro from mesostructure by diffusion MRI: A Bayesian approach. *NeuroImage*, 147:964–975.

[Reisert et al., 2014] Reisert, M., Kiselev, V. G., Dihtal, B., Kellner, E., and Novikov, D. S. (2014). MesoFT: Unifying Diffusion Modelling and Fiber Tracking. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, pages 201–208. Springer International Publishing. DOI: 10.1007/978-3-319-10443-0_26.

[Robinson et al., 2014] Robinson, E. C., Jbabdi, S., Glasser, M. F., Andersson, J., Burgess, G. C., Harms, M. P., Smith, S. M., Van Essen, D. C., and Jenkinson, M. (2014). MSM: A new flexible framework for Multimodal Surface Matching. *NeuroImage*, 100:414–426.

[Roine et al., 2014] Roine, T., Jeurissen, B., Perrone, D., Aelterman, J., Leemans, A., Philips, W., and Sijbers, J. (2014). Isotropic non-white matter partial volume effects in constrained spherical deconvolution. *Frontiers in Neuroinformatics*, 8.

[Ruiz-Alzola et al., 2002] Ruiz-Alzola, J., Westin, C. F., Warfield, S. K., Alberola, C., Maier, S., and Kikinis, R. (2002). Nonrigid registration of 3d tensor medical data. *Medical Image Analysis*, 6(2):143–161.

[Sahraian et al., 2010] Sahraian, M. A., Radue, E.-W., Haller, S., and Kappos, L. (2010). Black holes in multiple sclerosis: definition, evolution, and clinical correlations. *Acta Neurologica Scandinavica*, 122(1):1–8.

[Sakaie et al., 2012] Sakaie, K., Lowe, M., Koenig, K., Bermel, R., Stone, L., and Phillips, M. (2012). Spatial Resolution Requirements for Diffusion Tensor Imaging of Fornix.

[Salvador et al., 2005] Salvador, R., Peña, A., Menon, D. K., Carpenter, T. A., Pickard, J. D., and Bullmore, E. T. (2005). Formal characterization and extension of the linearized diffusion tensor model. *Human Brain Mapping*, 24(2):144–155.

[Scherrer and Warfield, 2010] Scherrer, B. and Warfield, S. K. (2010). Why multiple b-values are required for multi-tensor models. evaluation with a constrained log-euclidean model. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1389–1392.

[Scholz et al., 2009] Scholz, J., Klein, M. C., Behrens, T. E., and Johansen-Berg, H. (2009). Training induces changes in white matter architecture. *Nature neuroscience*, 12(11):1370–1371.

[Schwarz et al., 2014] Schwarz, C. G., Reid, R. I., Gunter, J. L., Senjem, M. L., Przybelski, S. A., Zuk, S. M., Whitwell, J. L., Vemuri, P., Josephs, K. A., Kantarci, K., Thompson, P. M., Petersen, R. C., and Jack Jr., C. R. (2014). Improved

DTI registration allows voxel-based analysis that outperforms Tract-Based Spatial Statistics. *NeuroImage*, 94:65–78.

[Shemesh et al., 2016] Shemesh, N., Jespersen, S. N., Alexander, D. C., Cohen, Y., Drobnjak, I., Dyrby, T. B., Finsterbusch, J., Koch, M. A., Kuder, T., Laun, F., Lawrenz, M., Lundell, H., Mitra, P. P., Nilsson, M., Özarslan, E., Topgaard, D., and Westin, C.-F. (2016). Conventions and nomenclature for double diffusion encoding NMR and MRI. *Magnetic Resonance in Medicine*, 75(1):82–87.

[Shen et al., 2007] Shen, S., Szameitat, A. J., and Sterr, A. (2007). VBM lesion detection depends on the normalization template: a study using simulated atrophy. *Magnetic Resonance Imaging*, 25(10):1385–1396.

[Smith et al., 2006] Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., and Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31(4):1487–1505.

[Smith et al., 2004] Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., and Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, Supplement 1:S208–S219.

[Smith and Nichols, 2009] Smith, S. M. and Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1):83–98.

[Smith et al., 2015] Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E. J., Glasser, M. F., Ugurbil, K., Barch, D. M., Van Essen, D. C., and

Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, 18(11):1565–1567.

[Song et al., 2003] Song, S.-K., Sun, S.-W., Ju, W.-K., Lin, S.-J., Cross, A. H., and Neufeld, A. H. (2003). Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *NeuroImage*, 20(3):1714–1722.

[Sotiras et al., 2013] Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable Medical Image Registration: A Survey. *IEEE transactions on medical imaging*, 32(7):1153–1190.

[Sotiropoulos et al., 2012] Sotiropoulos, S. N., Behrens, T. E., and Jbabdi, S. (2012). Ball and Rackets: Inferring Fibre Fanning from Diffusion-weighted MRI. *Neuroimage*, 60(2):1412–1425.

[Sotiropoulos et al., 2013] Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., Glasser, M. F., Hernandez, M., Sapiro, G., Jenkinson, M., Feinberg, D. A., Yacoub, E., Lenglet, C., Van Essen, D. C., Ugurbil, K., and Behrens, T. E. (2013). Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage*, 80:125–143.

[Sotiropoulos and Zalesky, 2017] Sotiropoulos, S. N. and Zalesky, A. (2017). Building connectomes using diffusion MRI: why, how and but. *NMR in biomedicine*.

[Stejskal and Tanner, 1965] Stejskal, E. O. and Tanner, J. E. (1965). Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*, 42(1):288–292.

[Sun et al., 2006] Sun, S.-W., Liang, H.-F., Le, T. Q., Armstrong, R. C., Cross, A. H., and Song, S.-K. (2006). Differential sensitivity of in vivo and ex vivo diffusion tensor

imaging to evolving optic nerve injury in mice with retinal ischemia. *NeuroImage*, 32(3):1195–1204.

[Supekar et al., 2010] Supekar, K., Uddin, L. Q., Prater, K., Amin, H., Greicius, M. D., and Menon, V. (2010). Development of functional and structural connectivity within the default mode network in young children. *NeuroImage*, 52(1):290–301.

[Syková and Nicholson, 2008] Syková, E. and Nicholson, C. (2008). Diffusion in Brain Extracellular Space. *Physiological Reviews*, 88(4):1277–1340.

[Tanno et al., 2017] Tanno, R., Worrall, D. E., Ghosh, A., Kaden, E., Sotiropoulos, S. N., Criminisi, A., and Alexander, D. C. (2017). Bayesian Image Quality Transfer with CNNs: Exploring Uncertainty in dMRI Super-Resolution. *arXiv:1705.00664 [cs]*. arXiv: 1705.00664.

[Tao and Miller, 2006] Tao, X. and Miller, J. V. (2006). A Method for Registering Diffusion Weighted Magnetic Resonance Images. In Larsen, R., Nielsen, M., and Sporring, J., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, number 4191 in Lecture Notes in Computer Science, pages 594–602. Springer Berlin Heidelberg.

[Tariq et al., 2016] Tariq, M., Schneider, T., Alexander, D. C., Gandini Wheeler-Kingshott, C. A., and Zhang, H. (2016). Bingham-NODDI: Mapping anisotropic orientation dispersion of neurites using diffusion MRI. *NeuroImage*, 133:207–223.

[Tax et al., 2014] Tax, C. M. W., Jeurissen, B., Vos, S. B., Viergever, M. A., and Leemans, A. (2014). Recursive calibration of the fiber response function for spherical deconvolution of diffusion MRI data. *NeuroImage*, 86:67–80.

[Tournier et al., 2007] Tournier, J.-D., Calamante, F., and Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI:

non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4):1459–1472.

[Tournier et al., 2004] Tournier, J. D., Calamante, F., Gadian, D. G., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23(3):1176–1185.

[Tschumperle and Deriche, 2003] Tschumperle, D. and Deriche, R. (2003). Variational frameworks for DT-MRI estimation, regularization and visualization. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 116–121 vol.1.

[Tuch, 2004] Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6):1358–1372.

[Vapnik, 1995] Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA.

[Varentsova et al., 2014] Varentsova, A., Zhang, S., and Arfanakis, K. (2014). Development of a high angular resolution diffusion imaging human brain template. *NeuroImage*, 91:177–186.

[Veraart et al., 2017] Veraart, J., Novikov, D. S., and Fieremans, E. (2017). TE dependent Diffusion Imaging (TEdDI) distinguishes between compartmental T2 relaxation times. *NeuroImage*.

[Wang et al., 2011] Wang, Y., Gupta, A., Liu, Z., Zhang, H., Escolar, M. L., Gilmore, J. H., Gouttard, S., Fillard, P., Maltbie, E., Gerig, G., and Styner, M. (2011). DTI registration in atlas based fiber analysis of infantile Krabbe disease. *NeuroImage*, 55(4):1577–1586.

[Wansapura et al., 1999] Wansapura, J. P., Holland, S. K., Dunn, R. S., and Ball, W. S. (1999). NMR relaxation times in the human brain at 3.0 tesla. *Journal of Magnetic Resonance Imaging*, 9(4):531–538.

[Wassermann et al., 2011] Wassermann, D., Rathi, Y., Bouix, S., Kubicki, M., Kikinis, R., Shenton, M., and Westin, C.-F. (2011). White Matter Bundle Registration and Population Analysis Based on Gaussian Processes. In *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, pages 320–332. Springer, Berlin, Heidelberg.

[Waxman and Bennett, 1972] Waxman, S. G. and Bennett, M. V. L. (1972). Relative Conduction Velocities of Small Myelinated and Non-myelinated Fibres in the Central Nervous System. *Nature*, 238(85):217–219.

[Westin et al., 2016] Westin, C.-F., Knutsson, H., Pasternak, O., Szczepankiewicz, F., Özarslan, E., van Westen, D., Mattisson, C., Bogren, M., O'Donnell, L. J., Kubicki, M., Topgaard, D., and Nilsson, M. (2016). Q-space trajectory imaging for multidimensional diffusion MRI of the human brain. *NeuroImage*, 135:345–362.

[Wheeler-Kingshott and Cercignani, 2009] Wheeler-Kingshott, C. A. and Cercignani, M. (2009). About "axial" and "radial" diffusivities. *Magnetic Resonance in Medicine*, 61(5):1255–1260.

[Winkler et al., 2014] Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, 92:381–397.

[Worsley et al., 1992] Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918.

[Yendiki et al., 2014] Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., and Fischl, B. (2014). Spurious group differences due to head motion in a diffusion MRI study. *NeuroImage*, 88(Supplement C):79–90.

[Yushkevich et al., 2008] Yushkevich, P. A., Zhang, H., Simon, T. J., and Gee, J. C. (2008). Structure-specific statistical mapping of white matter tracts. *NeuroImage*, 41(2):448–461.

[Zalesky, 2011] Zalesky, A. (2011). Moderating registration misalignment in voxelwise comparisons of DTI data: a performance evaluation of skeleton projection. *Magnetic Resonance Imaging*, 29(1):111–125.

[Zhang et al., 2007] Zhang, H., Avants, B., Yushkevich, P., Woo, J., Wang, S., McCluskey, L., Elman, L., Melhem, E., and Gee, J. (2007). High-Dimensional Spatial Normalization of Diffusion Tensor Images Improves the Detection of White Matter Differences: An Example Study Using Amyotrophic Lateral Sclerosis. *IEEE Transactions on Medical Imaging*, 26(11):1585–1597.

[Zhang et al., 2011] Zhang, H., Hubbard, P. L., Parker, G. J. M., and Alexander, D. C. (2011). Axon diameter mapping in the presence of orientation dispersion with diffusion MRI. *NeuroImage*, 56(3):1301–1315.

[Zhang et al., 2012] Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., and Alexander, D. C. (2012). NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage*, 61(4):1000–1016.

[Zhang et al., 2006] Zhang, H., Yushkevich, P. A., Alexander, D. C., and Gee, J. C. (2006). Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical Image Analysis*, 10(5):764–785.

[Zhang et al., 2001] Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57.

[Özarslan and Basser, 2008] Özarslan, E. and Basser, P. J. (2008). Microscopic anisotropy revealed by NMR double pulsed field gradient experiments with arbitrary timing parameters. *The Journal of Chemical Physics*, 128(15):154511.