April 2010
Imperial College of Science, Technology and Medicine
Clinical Sciences Centre

Characterising population variability in brain structure through models of whole-brain structural connectivity

Emma Claire Robinson

A dissertation submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
from
Imperial College London

Abstract

Models of whole-brain connectivity are valuable for understanding neurological function. This thesis seeks to develop an optimal framework for extracting models of whole-brain connectivity from clinically acquired diffusion data. We propose new approaches for studying these models. The aim is to develop techniques which can take models of brain connectivity and use them to identify biomarkers or phenotypes of disease.

The models of connectivity are extracted using a standard probabilistic tractography algorithm, modified to assess the structural integrity of tracts, through estimates of white matter anisotropy. Connections are traced between 77 regions of interest, automatically extracted by label propagation from multiple brain atlases followed by classifier fusion. The estimates of tissue integrity for each tract are input as indices in 77x77 "connectivity" matrices, extracted for large populations of clinical data. These are compared in subsequent studies.

To date, most whole-brain connectivity studies have characterised population differences using graph theory techniques. However these can be limited in their ability to pinpoint the locations of differences in the underlying neural anatomy. Therefore, this thesis proposes new techniques. These include a spectral clustering approach for comparing population differences in the clustering properties of weighted brain networks. In addition, machine learning approaches are suggested for the first time. These are particularly advantageous as they allow classification of subjects and extraction of features which best represent the differences between groups.

One limitation of the proposed approach is that errors propagate from segmentation and registration steps prior to tractography. This can cumulate in the assignment of false positive connections, where the contribution of these factors may vary across populations, causing the appearance of population differences where there are none. The final contribution of this thesis is therefore to develop a common co-ordinate space approach. This combines probabilistic models of voxel-wise diffusion for each subject into a single probabilistic model of diffusion for the population. This allows tractography to be performed only once, ensuring that there is one model of connectivity. Cross-subject differences can then be identified by mapping individual subjects' anisotropy data to this model. The approach is used to compare populations separated by age and gender.

Acknowledgements

I would like to express thanks to all those who have helped bring this thesis together. I am grateful to both of my supervisors, David Edwards and Daniel Rueckert, without whose support and advice I would not have been able to achieve as much as I have.

I would also like to thank all of my colleagues in the Departments of Computing and Hammersmith Hospital, especially: Alex Hammers for educating me in the neurobiology and helping find biological explanations for my results, Jo Hajnal for imparting an enthusiasm to understand how everything works, Paul Aljabar for his mathematical and technical support, also Fani Deligianni for being my sounding board, and for her endless patience. I must also thank Michel Valstar for all his help and advice with machine learning, and of course all of the team in Studio A.

Finally I would also like to thank all my friends and family, who thankfully have more faith in me than I have in my self. Thanks to Becks for showing me it can be done. Thanks especially to Mum, Dad, Dan, Nadine, and Rohan for their constant encouragement, support and advice.

Dedication

For Grandad George

Contents

List of Tables

List of Figures

Chapter 1

Introduction

'To understand the functioning of a network, one must know its elements and their interconnections.'
Olaf Sporns

1.1 Brain connectivity and its relationship to development, disease and ageing

The mapping of brain architecture is a key frontier of modern science. Little is known about the mechanisms which drive brain function, how they develop during childhood, decay with ageing or differ in instances of neurological damage or disease. However, it is likely that most differences are affected by changes in the underlying connective microstructure, since this forms the substrate over which signals are passed.

Diffusion magnetic resonance imaging (MRI) enables indirect visualisation of tissue microstructure via estimation of water diffusion profiles at discrete points throughout the brain. Where water diffusion is free or unconstrained, as is largely the case in cerebral spinal fluid (CSF), diffusion is shown to be predominantly isotropic. However, where diffusion is hindered by the presence of physical obstacles (cells) within the white matter (WM), or restricted inside these cellular compartments, diffusion profiles are anisotropic and directed along the projection of the fibres. The mechanisms of this restriction are not fully understood, however it is thought to be partially correlated with barriers presented by axonal membranes and nerve myelination.

Myelin is a fatty substance which coats nerves, protecting and insulating them for more efficient signalling. Studies have shown that anisotropy within the white matter of the brain increases dramatically during early stages of development and myelination [103, 117]. This process continues throughout childhood [136] into early adulthood [93]. In contrast, healthy ageing is characterised by decreased white matter anisotropy between frontal and posterior regions of the brain [172], whereas anisotropy within the splenium is relatively preserved [14, 89]. This supports theories that ageing reverses the processes seen during development [172]. In addition, decreased anisotropy in frontal white matter during ageing is correlated with reduced executive functions such as problem solving and working memory [45, 97, 172]. This demonstrates a link between WM anisotropy and functional performance.

Measurements of anisotropy in white matter have been shown to correspond to levels of structural integrity in tissues affected by neurological impairment. Infants born preterm demonstrate lower white matter anisotropy than for term born control subjects [7]. This is maintained throughout childhood [115] and is correlated with lower cognitive performance later in life [90]. Low anisotropy has also been correlated with learning disabilities such as dyslexia [120]. In addition, studies into strokes [147] suggest that measurements of diffusion anisotropy may predict the onset of ischaemic attacks. Significant differences in white matter anisotropy have also been found in patients with diseases associated with myelin degeneration, such as multiple sclerosis [40, 156] and axonal degeneration, such as Amyotrophic Lateral Sclerosis (ALS) (motor neurone disease) [178]. Also, differences have been located in cases of psychiatric conditions such as bi-polar disorders [13] and schizophrenia [156].

Tractography algorithms estimate "tracts" or pathways between regions in the brain by following the direction of maximal diffusion anisotropy through the white matter. These have been shown to reflect known neuronal pathways [41]. Tractography has allowed studies to test hypotheses relating to specific neuronal connections, for example differences in tract projections between the substantia nigra and thalamus have been used to successfully classify patients with Parkinson's disease [106]. Shape analysis of tracts has been shown to correctly identify regional differences in patients with ALS [196] and tract guided analysis of anisotropy differences in the corpus callosum have been correlated with gender [124]. In addition, anisotropy within corpus callosum tracts has been used to quantify development [58, 62].

1.2 Motivations

Whilst tractography of single tracts has significantly increased understanding of the relationship between brain connectivity and function, in many cases there are instances where the effects of a disease on the white matter are diffuse or unknown. In such circumstances, techniques based on tract specific analysis either have to make assumptions about the locations of key affected regions in the brain, or assume that the influence of the disease is the same throughout the anatomy. By contrast, mapping tractography for the whole brain [59, 66, 83] avoids the requirement for prior hypotheses by enabling comparison of all tracts across a population simultaneously.

Current techniques for analysing variability in whole brain connectivity across populations model the network of connections as a graph. Graph theory lends itself well to the mapping of brain connections as it describes the topology of structures in terms of nodes (or regions) and their connecting edges. However, to date, most studies have examined binary graph structures, where edge weights equal either 0 or 1. This ignores any properties pertaining to the structural integrity of the underlying tissue. In addition, comparisons of graphs across different subjects or populations are limited to general descriptors of brain organisation. These are not able to specify the locations of changes to the tissue microstructure nor indicate which tracts differ most across populations.

This thesis, therefore, seeks to develop new techniques for studying population differences in the structural integrity of tracts across the whole brain. A new framework is presented for extracting whole brain tractography from clinically acquired data, where estimates of anisotropy are taken along tracts and these are used to build a weighted graph. New techniques are then generated for studying population differences in the structural integrity of tracts across the whole brain. These feature a spectral clustering approach for comparing weighted graphs across populations and, for the first time, machine learning based approaches, which allow both classification of subjects and feature extraction of those tracts that differ most greatly between populations.

However, comparisons between structural connectivity networks propagated in the subjects' native co-ordinate space are subject to variations in the assignment of regions of interest between which connectivity is assessed. This can bias the results of the connectivity analysis. Therefore, a new standard space approach is presented, whereby a probabilistic white matter atlas of subjects' diffusion data is created and tractography is performed once in a common co-ordinate space. This removes the influence of segmentation variability, enabling comparisons to be made between anisotropy alone. Furthermore,

performing tractography in a common co-ordinate space enables comparisons of anisotropy along the full length of tracts.

1.3 Contributions

This thesis presents work on deriving models of whole-brain structural connectivity, from clinically acquired diffusion magnetic resonance imaging data, for the purpose of studying population differences in brain connectivity. This thesis makes four main contributions:

- A new framework for extracting models of whole brain tractography from clinically acquired data is presented. This uses a standard probabilistic tracking algorithm [24], known to perform well for clinical data, but adjusts it such that measurements of mean anisotropy are estimated along each tract. This provides an effective indicator of structural integrity through which tracts can be compared. Using this, weighted connectivity networks are formed via tractography between 77 regions of interest (ROIs), segmented via an atlas-based segmentation approach.

- Weighted structural brain connectivity networks, averaged over two age populations, are compared using graph theory techniques. In particular, spectral graph clustering is used to compare the structure of the two networks. The results reflect changes to brain organisation known to occur during ageing.

- A machine learning approach is developed for the analysis of whole brain tractography data for the first time. A principal component analysis, linear discriminate analysis approach (PCA + LDA) as well as a GentleBoost, support vector machine approach (GentleBoost-SVM) are tested. These allow classification of individual subjects as well as extraction of key features which best represent the differences between groups.

- A new framework for performing probabilistic tractography in a template space is presented. Here, probabilistic diffusion data for all subjects is registered into a common co-ordinate space from which the probability distribution for the whole population is derived at each voxel. Tractography is performed only once, generating a single model of connectivity. Then, anisotropy estimates for each subject are transformed to these tract templates, allowing connectivity matrices to be extracted for each subject. These are subsequently compared using machine learning.

The advantage is that population specific differences in the segmentation do not influence the outcome of the tractography. This approach is tested on age and gender.

1.4 Thesis overview

The rest of this thesis is laid out as follows.

- Chapter 2 provides a review of diffusion imaging and diffusion tractography techniques.
- Chapter 3 discusses image registration, segmentation and atlasing. These are presented with a view to giving a background to many of the steps involved in pre-processing of the image data prior to diffusion tractography.
- Chapter 4 describes the process of estimating tractography between 77 regions of interest. In particular, it describes how the regions are delineated, how tissue type information is incorporated to ensure tracts are only estimated between GM, the estimation of mean anisotropy along each tract and finally, how the parameters for tracking were optimised.
- Chapter 5 discusses a graph theory based approach. It discusses previous studies that have analysed networks using small world analysis and clustering approaches such as modularity. It then proposes using a combination of small world analysis and spectral clustering to compare brain networks obtained using the proposed framework, averaged over two different age populations.
- Chapter 6 presents two machine learning approaches. These allow allow classification of subjects and extraction of features which highlight the differences between groups. The approaches are compared for both anisotropy and connection probability based metrics of connectivity.
- Chapter 7 extends the framework, by applying it to a single whole brain connectivity matrix, obtained from a probabilistic white matter atlas of the population. The chapter describes how the diffusion data is registered and combined to form one data set representative of the whole population. The results are compared to those obtained via tracking in the native space.
- Chapter 8 concludes and discusses future work.

Chapter 2

Diffusion Imaging and Tractography

2.1 Introduction

Brownian motion is a passive process which characterises the random thermal motion of molecular spins in a fluid. First discovered by Robert Brown in 1828 through observation of the motion of pollen grains in water [137], it was not fully understood until the formulation of kinetic theory, whereby it was shown that particles in a solution undergo collisions with the molecules of the fluid such that they on average move from high to low concentrations [42]. This macroscopic behaviour forms a process known as diffusion and can be modelled statistically.

Diffusion between molecules of the same type is known as self-diffusion. This occurs between water molecules in human tissues. Traditional magnetic resonance techniques have for some time used water in human tissue to generate images, using the varying speed of magnetic relaxation in different tissues to provide a natural contrast. In some tissue however, such as the neural white matter and muscle tissue it is suspected that thermal water motion is preferentially restricted in directions perpendicular to the tissue fibres. In this case the diffusion profile is said to be *anisotropic* and modern diffusion imaging techniques attempt to use this directional information to probe tissue microstructure, using the application of large magnetic field gradients during the acquisition process to attenuate signal according to the rate of diffusion in the direction of the applied gradient. This is used to build up a macroscopic picture of diffusion for each voxel (volume element) in the image.

Recently diffusion tractography algorithms have started to use the information extracted from diffusion imaging to predict the existence of anatomical connections in the brain. This chapter provides an

overview of the development of this technique, describing the origins of the diffusion signal and how diffusion has typically been modelled at the voxel and global level.

2.1.1 Neural architecture and its influence on diffusion

Brain tissue is composed of two basic types of tissue: white matter and grey matter [96]. Grey matter regions are predominately made up of nerve cell bodies (soma) and their supportive glial cells, where the main neuronal components, dendrites and cell bodies, are responsible for many of the computational processes of the brain. White matter regions represent the pathways of communication and consist mainly of myelinated nerve axons. Accurate mapping of white matter connections in the brain is therefore essential for understanding brain function at a systems level.

Since the earliest diffusion studies on animals [112, 114], research has strived to understand the exact causes of anisotropic diffusion. In simple terms the nerves and white matter are all ordered axonal structures with the same micro-structural components: an axonal membrane and myelin sheath, containing microtubules and neurofilaments connected by a scaffold of tiny microfilaments. These are subject to two different generalised modes of diffusion: hindered and restricted [27]. Hindered diffusion describes diffusion around the extracellular space of the cells, such that molecules can no longer go straight from A to B but rather must take a more tortuous route increasing the diffusion time. In addition, restricted diffusion refers to the movement of intra-cellular water. This is prevented from free motion by the boundaries of cells, and is characterised by levelling off of the diffusion distance over time (figure 2.1).

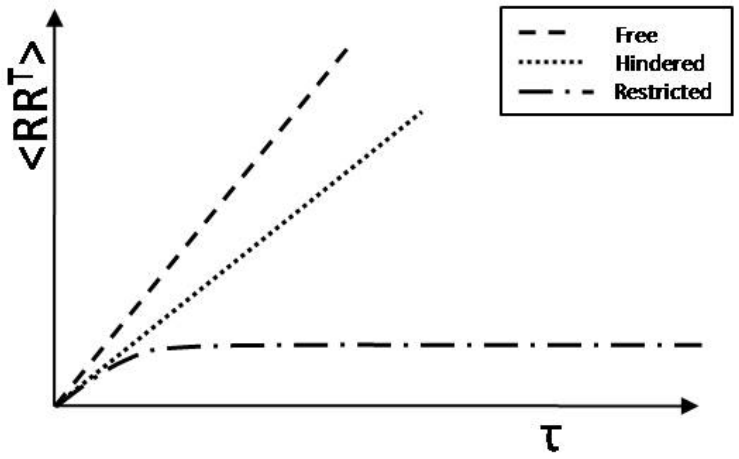


Figure 2.1: The relationship between the diffusion distance and time for free, hindered and restricted diffusion. This is described Einstein relation: $< \mathbf{R}\mathbf{R}^T >= 6D\tau$, where $< \mathbf{R}\mathbf{R}^T >$ is the mean square displacement, D is the diffusion coefficient and τ is the diffusion time (equation 2.2)

Many different factors are known to contribute to the directed diffusion seen in white matter. Studies of unmyelinated axons of the garfish [19] demonstrate that a significant barrier is provided by the axon membrane. However, contributions to anisotropy are also made by the insulating myelin sheath, which is designed to promote faster signal transmission. This is known to increase during development and degrade via certain types of neurodegenerative disease. In addition, longitudinally-orientated neurofilaments and microtubules (related to axonal transport) hinder diffusion perpendicular to nerve bundles, and also align nerves into bundles large enough to meet the resolution available to MR imaging studies. The combined effect of these different factors is complex and not fully understood. Nevertheless, a key area of recent research has been to build models which separate the contributions from free/hindered and restricted diffusion components for the purpose of achieving virtual histology. Such studies are beyond the scope of this work but further details can be found in the following papers [10, 9]. In what follows, we use simple models of diffusion, and aim to extract the orientations of major fibre bundles.

2.1.2 Modelling diffusion in tissues

Diffusion properties of tissues offer a means of elucidating details of tissue microstructure, which would otherwise exist far below the resolution of current imaging techniques. Patterns of diffusion in tissues can be modelled using an ensemble average diffusion propagator $P(\mathbf{R}|\tau)$. This represents the expected distribution of relative displacements $\mathbf{R} = \mathbf{r} - \mathbf{r_0}$ of all spins within the image volume (or voxel), where $\mathbf{r_0}$ represents the initial starting point for each spin and $\mathbf{r}$ represents its position after time τ.

In the specific case where diffusion is free, i.e. there are no barriers, diffusion can be described using a Gaussian diffusion propagator P_G:

$$P_G(\mathbf{R}, \tau) = (4\pi D\tau)^{-\frac{3}{2}} \exp(-\frac{\mathbf{R^T R}}{4D\tau}) \tag{2.1}$$

D is the diffusion coefficient:

$$D = \frac{1}{6\tau} < \mathbf{R}\mathbf{R}^T > \tag{2.2}$$

$< \mathbf{R}\mathbf{R}^T >$ represents the mean square displacement [42].

In reality, diffusion in biological tissues is subject to many of the physical barriers mentioned above and can never be completely described as a Gaussian process. Nevertheless, the Gaussian assumption

has proved very successful in the early years of diffusion imaging. The following sections describe an overview of the principles of magnetic resonance imaging and measurement of the diffusion signal.

2.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) utilises the quantum mechanical spin properties of atomic nuclei in a magnetic field. Typically MRI uses hydrogen protons in tissue to provide contrast although other substances can be used. These contain a non zero spin angular momentum which induces a magnetic moment (μ) equivalent to a microscopic magnetic dipole along their axis of rotation.

In general these axes are randomly orientated, but in the presence of a static external field ($\mathbf{B_0}$), their moments feel a torque causing them to precess about the field with frequency:

$$\omega = \gamma B_0 \tag{2.3}$$

Here ω is known as the Larmor or resonance frequency and γ is the gyromagnetic ratio, which describes the relation between spin angular momentum and its magnetic moment. This leads to a net magnetisation (magnitude M_z) of the spins in the direction of the applied field (defined by the z-axis, figure 2.2)

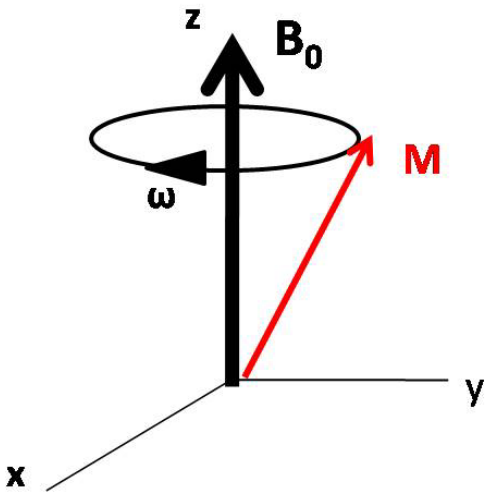
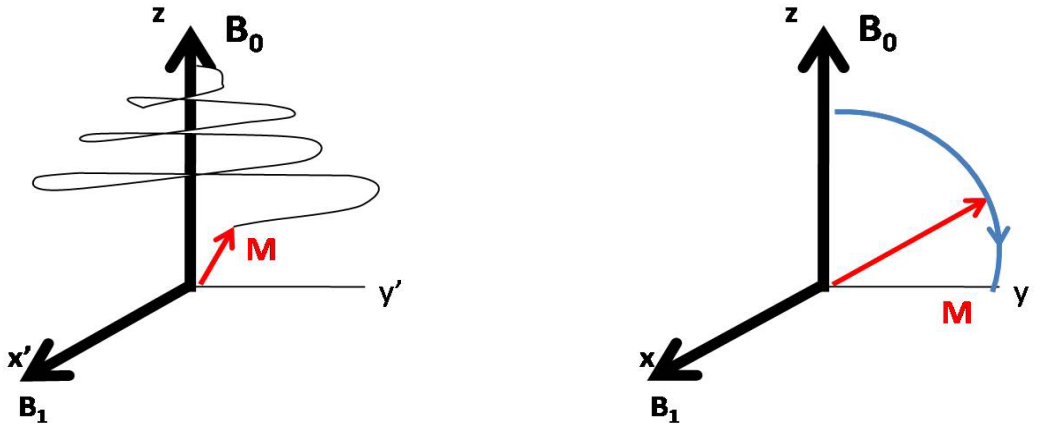


Figure 2.2: Precession of spin magnetisation around a static magnetic field $\mathbf{B_0}$

Magnetic resonance works via injection of an oscillating radiofrequency (RF) pulse at the resonant frequency, in a plane perpendicular to the static field $\mathbf{B_0}$. This generates a time varying magnetic field $\mathbf{B_1}$ in the transverse plane and causes spins to precess simultaneously around $\mathbf{B_0}$ and $\mathbf{B_1}$ (figure 2.3 a). The combined effect can be simplified by viewing the system in a rotating frame of reference ($x\prime$,$y\prime$,z) (figure 2.3 b). This rotates about the z-axis at the Larmor frequency. In this frame we can ignore the contribution to the precession from the $\mathbf{B_0}$, and assume the $\mathbf{B_1}$ field is constant along $x\prime$. The result is a rotation of the net magnetisation vector about about $\mathbf{B_1}$, where the angle of rotation is proportional to the length of the RF pulse. This leads to an increase in transverse magnetisation M_{xy}. Spins now precess in phase.



a) Laboratory frame b) Rotating frame

Figure 2.3: Simultaneous precession of spin magnetisation about a static magnetic field $\mathbf{B_0}$ and a time varying field $\mathbf{B_1}$

However, when the pulse is removed, the nuclei gradually return to their original state, releasing energy in the form of an RF pulse from which the MRI image is built up. The nuclei experience two key forms of relaxation: spin-lattice (T_1) relaxation:

$$M_z = M_{z0}(1 - exp^{(-t/T_1)}) \tag{2.4}$$

This is caused by loss of energy to the surroundings. In addition T_2 relaxation:

$$M_{xy} = M_{xy0}exp^{(-t/T_2)} \tag{2.5}$$

This is caused by loss of coherence of the transverse magnetisation as spins fall out of phase due to spin-spin interactions and inhomogeneities in the static $\mathbf{B_0}$ field.

MRI images may be built simply from visualising the contrast due to different water content in different tissues, otherwise known as proton density images. However different tissues in the body also have different T1 and T2 values. Stronger contrast can be seen by weighting image schemes relative to these relaxation times.

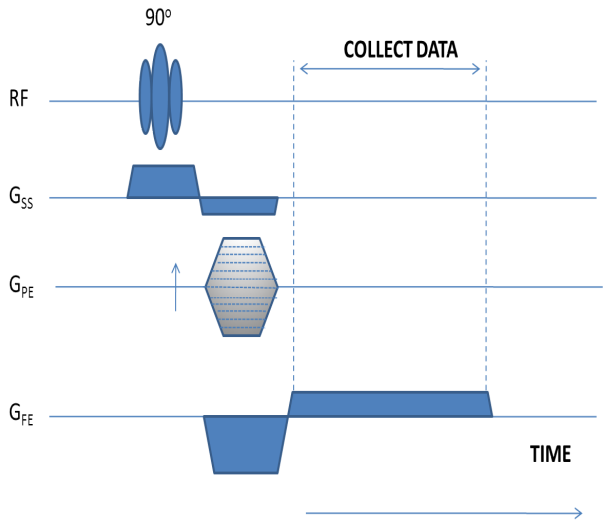
2.2.1 Spatial encoding

Images are generated by pinpointing the locations of different frequency signals within the tissue using encoding gradients. Encoding gradients cause position dependent precession of the nuclei in different sections of the body by ensuring each position is subject to a different magnetic field.

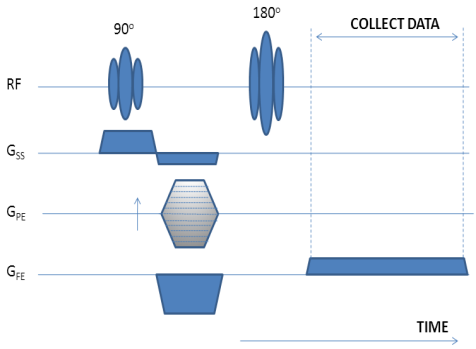
There are three encoding gradients applied in a conventional slice MRI sequence (Figure 2.4 a): first slice-select gradients (G_{SS}) are applied coincidentally with the RF pulse so as to ensure MR interactions are restricted to a two-dimensional slab in the x-y plane. Following this individual locations in each slice are identified via repetitions of frequency and phase encoding. First a phase encoding (G_{PE}) gradient is applied in an orthogonal direction to the slice select gradient. This causes spin to precess at position dependent frequencies until the gradient is turned off, whereby spins return to precess at the same rate but retain differences in phase. Then frequency encoding (G_{FE}) gradients are applied in the third direction, enforcing position dependent frequencies along the remaining dimension.

Data is measured during the application of the frequency encoding gradient but after phase encoding. The whole process is repeated with different magnitudes of phase encoding gradient (shown by dotted lines figure 2.4 a and b). This provides each position along the phase encoding axis with a unique rate of change of phase, or synthesised frequency[1]. 2-dimensional Fourier transforms of the signal can then be used to reconstruct the image.

[1]Rate of change of phase = frequency



a) Standard sequence



b) Spin echo sequence

Figure 2.4: Pulse diagrams for gradient sequence MRI imaging sequencies: a) In the conventional MR gradient echo sequence a 90° RF frequency pulse is applied simultaneously with a slice select gradient applied in the z direction. This ensures that spins are only selected for rotation within a thin slice through the object. Positions in x and y are identified through phase and frequency encoding. b) Spin echo sequences apply an additional 180° RF pulse prior to frequency encoding. This causes spins that are phase advanced to phase delayed and vice versa leading to a partial recovery of the signal following T2 decay

2.2.2 The spin echo sequence

Spin echoes (see Figure 2.4 b) provide image contrast by weighting images relative to T_2 relaxation times. In these sequences loss of phase coherence is counteracted by the application of a second 180° pulse which flips the magnetization, causing spins which are phase advanced to become delayed and vice versa. As the spins continue to de-phase according to the local field this leads to a recovery of alignment and partial recovery of signal. If the 180° pulse is applied at a time t after the original RF signal then alignment is recovered at time $2t$. Therefore by using a series of spin echo sequences and measuring the signal height at each echo time, T_2 can be approximated. The principles of spin echo form the basis of Pulsed Gradient Spin Echo (PGSE) diffusion imaging sequences.

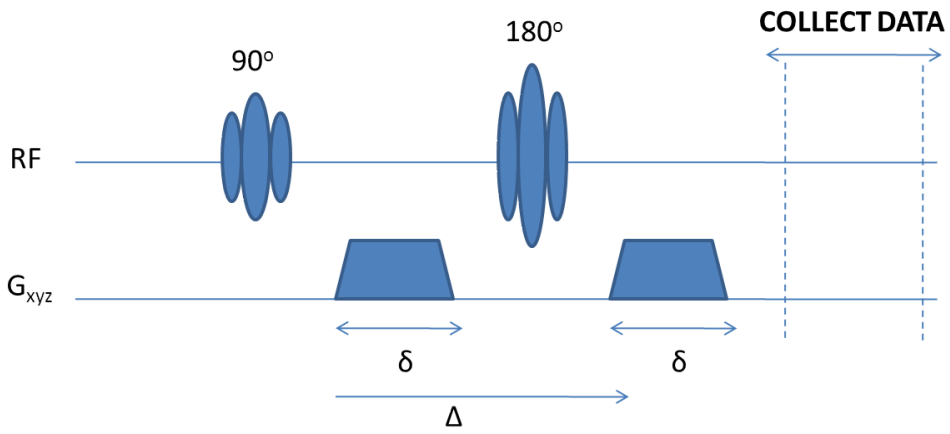


Figure 2.5: Placement of diffusion gradients during spin echo sequence

2.3 The Pulsed Gradient Spin Echo sequence

In 1965 Stejskal and Tanner [168] formulated an experiment, known as the Pulsed Gradient Spin Echo (PGSE) to extract diffusion information from the MR signal by increasing the sensitivity of a conventional spin echo sequence to its influence. They added two short strong gradient pulses, before and after the application of the 180° radio frequency (RF) pulse (figure 2.5) and solved the Torrey-Bloch equations[2] under the assumption that: a) diffusion is Gaussian and b) the gradients are applied for a sufficiently short time that it can be assumed that no diffusion occurs during them. The result is that spins which stay in the same place (and do not move due to diffusion) experience the same magnetic field as before and rephase completely. In contrast, spins which do move are subject to a difference in the position dependent magnetic field. This causes them to develop different Larmor frequencies and not rephase completely. The net results is an attenuation in the magnetic resonance signal that can be approximated as:

$$\ln[\frac{S}{S_0}] = -\gamma^2 D_{ADC} \delta^2 \mathbf{g}^2 (\Delta - \frac{\delta}{3}) = -bD_{ACD} \tag{2.6}$$

where S and S_0, represent the signal with and without diffusion weighting respectively, D_{ADC} refers to the *apparent* diffusion coefficient (ADC) (measured under Gaussian conditions), $\mathbf{g}$ is the gradient strength vector, δ is the gradient duration, and Δ is the time between gradients. Therefore, the b-value, $b = |-\gamma\delta\mathbf{g}|^2(\Delta - \frac{\delta}{3})$ relates to the strength and separation of the applied magnetic gradients.

[2]The Bloch equations describe MR. These were extended by Torrey [179] to factor out the influence of diffusion

2.4 The diffusion tensor

The Stejskal-Tanner equation (2.6) makes it possible to estimate the amount of diffusion at any voxel by comparison of the signal attenuation between a single spin echo measurement and a PGSE experiment acquired with one gradient direction:

$$D_{ADC} = -\frac{1}{b} ln[\frac{S}{S_0}] \tag{2.7}$$

However this generates no information about the directional nature of the diffusion in some tissues. Therefore in 1994, Basser et. al. [15] extended this approach by linearly relating the apparent self-diffusion in three dimensions to the measured single attenuation:

$$ln[\frac{S}{S_0}] = -\sum_{i=1}^{3}\sum_{j=1}^{3} b_{ij} D_{ij} \tag{2.8}$$

Here, b_{ij} and D_{ij} are the matrix elements of the b-matrix (derived from the size and direction of each of the applied diffusion gradients) and *apparent* diffusion tensor ($\mathbf{D}$) . Therefore the diffusion tensor can be seen as a 3x3 covariance matrix of spin displacements. The off-diagonal elements describe the degree of correlation between displacements along each of the three major axes, and must be approximated such that (where appropriate) tensors may be orientated away from the principle co-ordinate frame.

As the Gaussian approximation naturally enforces symmetry to the estimation of the tensor (equation 2.8), this requires the calculation of six independent elements of $\mathbf{D}$ (3 diagonal, 3 off diagonal) plus S_0. This requires at least 7 measurements, including 6 gradient directions and one non-diffusion weighted measurement:

$$\ln \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \end{bmatrix} = \begin{bmatrix} b^1_{xx} & 2b^1_{xy} & 2b^1_{xz} & b^1_{yy} & 2b^1_{yz} & b^1_{zz} & -1 \\ b^2_{xx} & 2b^2_{xy} & 2b^2_{xz} & b^2_{yy} & 2b^2_{yz} & b^2_{zz} & -1 \\ b^3_{xx} & 2b^3_{xy} & 2b^3_{xz} & b^3_{yy} & 2b^3_{yz} & b^3_{zz} & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b^n_{xx} & 2b^n_{xy} & 2b^n_{xz} & b^n_{yy} & 2b^n_{yz} & b^n_{zz} & -1 \end{bmatrix} \times \begin{bmatrix} D_{xx} \\ D_{xy} \\ D_{xz} \\ D_{yy} \\ D_{yz} \\ D_{zz} \\ \ln S_0 \end{bmatrix}$$

or

$$\ln \mathbf{S} = \mathbf{B}\alpha \tag{2.9}$$

Here $\ln \mathbf{S}$ is the $nm \times 1$ column vector of observed signal intensities given the model, for n directions each with m measurements. $\mathbf{B}$ is a matrix containing as many rows as measurements and seven columns (representing the 6 unique elements of the b-matrix plus one b_0 measurement) and α, which represents the set of 6 independent elements of $\mathbf{D}$, and $ln[S_0]$ (the predicted value of the signal in the absence of any diffusion).

For exactly seven measurements B is square and the problem can be solved by taking the inverse:

$$\alpha = \mathbf{B^{-1}} \ln(\mathbf{S}) \tag{2.10}$$

However, this is typically too sensitive to measurement noise. Therefore it is standard practise to image using at least 15 gradient directions (and one b_0) from which m measurements are taken. The optimal α is then found by minimising the weighted sum of squared differences between the observed and predicted measurements of the signal:

$$\alpha_{\mathbf{opt}} = (\mathbf{B^T \Sigma^{-1} B})^{-1} (\mathbf{B^T \Sigma^{-1}}) \ln(\mathbf{S}). \tag{2.11}$$

Weights ($\mathbf{\Sigma}^{-1}$) are added to take into account the fact the elements of α have higher variance for low signal and vice versa.

2.4.1 The eigensystem of the diffusion tensor

The diffusion tensor formulation provides a simple representation of diffusion with many advantages. Firstly, since the diffusion tensor is positive-definite and symmetric it is possible to decompose it into a system of eigenvalues and eigenvectors:

$$\mathbf{DE} = \mathbf{E\Lambda} \tag{2.12}$$

where, the $\mathbf{E}$ is a matrix of orthonormal eigenvectors ($\mathbf{E} = (\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3})$) where $\mathbf{e_1}$ represents the principal axis of diffusion and $\mathbf{\Lambda}$ contains the eigenvalues:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

This can be represented by a Gaussian probability distribution:

$$P_G(\mathbf{R}, \tau) = (4\pi \mathbf{D}\tau)^{-\frac{3}{2}} exp(-\frac{\mathbf{R^T R}}{4\mathbf{D}\tau}) \tag{2.13}$$

represented graphically as a diffusion ellipsoid [16], or isoprobability surface whereby the boundaries of the ellipsoid represent the mean translational displacement of spins after time τ. See figure 2.6. This allows an anisotropic representation: one that differs according to orientation.

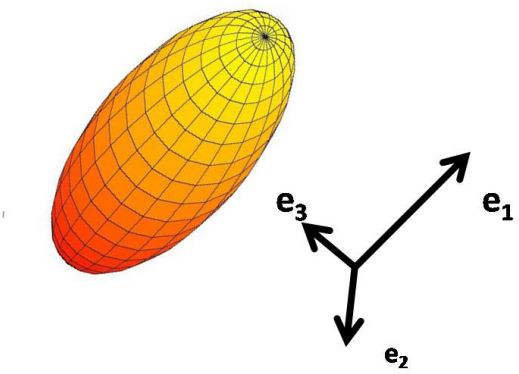


Figure 2.6: The diffusion ellipsoid

In addition, the representation also provides a number of simple, rotationally invariant scalar measures of anisotropy [17]. For example:

- **Mean diffusivity:** Taken from the trace of the tensor, $Tr(\mathbf{D}) = D_{xx} + D_{yy} + D_{zz}$, mean diffusivity is given by $Tr(\mathbf{D})/3$.
- **Relative Anisotropy:** Relative anisotropy (RA) is a normalised standard deviation representing the ratio of the anisotropic part of the diffusion tensor to the isotropic part:

$$RA = \sqrt{\frac{1}{3}} \frac{\sqrt{(\lambda_1 - <\lambda>)^2 + (\lambda_2 - <\lambda>)^2 + (\lambda_3 - <\lambda>)^2}}{<\lambda>} \tag{2.14}$$

  Here $<\lambda>$ is one third of the trace of the tensor.

- **Fractional Anisotropy:** Fractional anisotropy measures the fraction of the magnitude of the diffusion tensor that relates to anisotropic diffusion:

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - < \lambda >)^2 + (\lambda_2 - < \lambda >)^2 + (\lambda_3 - < \lambda >)^2}}{\sqrt{(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \tag{2.15}$$

Furthermore, the diffusion ellipsoid representation provides a simple and intuitive starting point for modelling pathways through the brain. This process is known as tractography and is described in the next section.

2.5 Fibre tractography

The decomposition of diffusion tensors into eigenvectors can be exploited to allow estimation of the directions of principal diffusion at many points throughout the brain. This enables the trajectory of cerebral fibre bundles to be traced non-invasively using deterministic tractography algorithms [111, 132, 133]. These have been used very successfully to model many major white matter tracts throughout the brain [188]. However, the model is reliant on clear inference on the direction of maximal diffusion and is therefore unsuited to propagating through areas of low diffusion anisotropy. In contrast, some approaches have evolved which attempt to use estimates of the uncertainty to relax the tractography criteria and generate a probabilistic map of tract trajectories [24, 53, 79, 105, 131]. In addition global techniques have been proposed. These attempt to take into account all possible trajectories [83, 85, 132, 133, 149, 195] before choosing an optimal path.

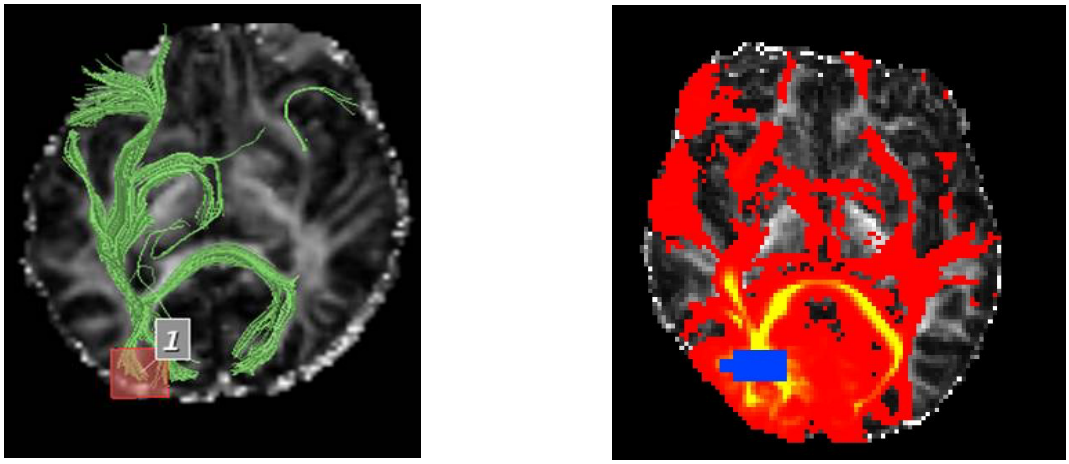
2.5.1 Streamline tractography

The basis of most common deterministic tracking approaches is the fibre assignment by continuous tracking (FACT) algorithm [111]. This first identifies a starting or seed point ($\mathbf{r_t}$) (typically at the centre of a voxel) and then propagates in the direction of principal eigenvector ($\mathbf{e_1}$) of the diffusion tensor using a discretisation step ($\triangle t$) equal or smaller than the size of a voxel:

$$\mathbf{r_{t+1}} = \mathbf{r_t} + \mathbf{e_1} \triangle t \tag{2.16}$$

The algorithm is propagated until anisotropy dips below an unacceptable threshold level or track angle changes above a pre-defined curvature limit.

The resulting tracts are known as streamlines due to the similarity of the approach to techniques used for finding paths through vector fields. Streamline algorithms are often thought of as a maximum likelihood estimate of the fibre path. This maximum likelihood assumption is reasonable in areas of high anisotropy but insufficient for low anisotropy tissue or where limited angular resolution and partial volume effects become too high and lead to uncertainty. Increased robustness is offered by using continuous tracking: this interpolates tract orientation at intermediate positions between voxels ensuring that tracts do not drift off course. In addition "Brute force" approaches can be used: these seed multiple tracts for every voxel in the tissue volume but only retain streamlines where they pass through the regions of interest (ROI). An example of streamline tractography is shown in figure 2.7 a.



a) Streamline tractography b Probabilistic tractography

Figure 2.7: Comparison of streamline and probabilistic tractography: a) streamline tracts (green) drawn using dtiquery software [152]. This employs a brute force approach which seeds streamlines for every voxel in the brain, but only retains streamlines that pass through the ROIs (shown as red box); b) probabilistic tracts (red) drawn using the approach by Behrens et al [24]. Multiple streamlines are initialised from each voxel in the ROI (blue). The results represents a distribution on the most likely trajectory of fibres from the ROI (areas of high probability are indicated by high intensity in the image

2.5.2 Probabilistic tractography

Understanding the influence of uncertainties in any fibre tracking scheme is essential if a full picture of the true pattern of connectivity is to be built up. Uncertainties enter the scheme, not only through artefacts and noise, but also by the simplification of the diffusion modelling process. In particular by making the assumption that diffusion is Gaussian, much of the complexity of the diffusion signal is ignored, causing residuals and additional uncertainty in the model parameters.

Probabilistic tractography schemes [24, 53, 79, 105, 131] attempt to represent uncertainty in the form of a distribution of tracts from each voxel. These can be obtained by stochastic modelling of the diffusion process [68]. Alternatively, the probabilistic index of connectivity (PICo) approach of Parker and Alexander [130, 131] assigns different magnitudes of uncertainty to the principal diffusion direction via estimations of both the anisotropy and the relative magnitudes and orientations of the second and third eigenvectors of the tensor at each point.

Other approaches [24, 53, 79, 105] use Bayesian inference to generate posterior distributions on the parameters of the local diffusion model. Bayes' theorem states that the posterior probability ($P(\Theta|Y,M)$ on the set of model parameters (Θ) given the data (Y) and the model (M) is proportional to the likelihood of seeing this data for the given parameters, $P(Y|\Theta,M)$, multiplied by some prior distribution which is informed by what we already know (or have assumed):

$$P(\Theta|Y,M) = \frac{P(Y|\Theta,M)P(\Theta|M)}{P(Y|M)} \tag{2.17}$$

Calculation of the likelihood requires a model of the underlying process. This could be a diffusion tensor model. However, in the approach developed by Behrens et. al. (used in this thesis) a conventional tensor representation is dropped in favour of a simple two-component partial volume model of the signal, S_i (figure 2.8). Here the first component models anisotropic diffusion in and around the axons (in a single direction) and the second represents isotropic diffusion. This ensures that all uncertainty is reflected in the principal diffusion direction, as opposed to all three eigenvectors as would be the case for the diffusion tensor. The model is parameterised as:

$$S_i = S_0(fexp(-b_i d \mathbf{v_i}^T \mathbf{R A R^T} \mathbf{v_i}) + (1-f)exp(-b_i d)) \tag{2.18}$$

where, d is the diffusivity, b_i and $\mathbf{v_i}$ are the b-value and gradient directions associated with the i-th

gradient acquisition, f is the fraction of the signal contributed by the anisotropic component and, in this instance, $\mathbf{R}$ is a rotation, which transforms the anisotropic diffusion tensor $\mathbf{A}$ to align with fibre direction (θ, ϕ):

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

In this model all parameters are assumed to have non informative priors except to ensure positivity where sensible. Noise is modelled as Gaussian and posterior distributions on the parameters of the model are estimated from the joint distribution using Markov Chain Monte Carlo (MCMC) sampling.

In all instances probabilistic tractography is performed by seeding multiple streamlines from each voxel in the seed region. These propagate by sampling from posterior distributions on θ and ϕ at points neighbouring the streamlines' current location (determined by probabilistic interpolation[3]). Streamlines only terminate where curvature exceeds a pre defined threshold. The result is a range of possible end points (or probability distribution over all possible trajectories) for each tract (figure 2.7 b). The approach displays a greater robustness to noise and uncertainty caused by fibre crossing than for conventional streamline approaches, as tracts will often still be propagated through these regions although the resulting tracts may show low connection probability. Tracts traced in this way have been shown to reflect underlying neuronal connections [23, 41].

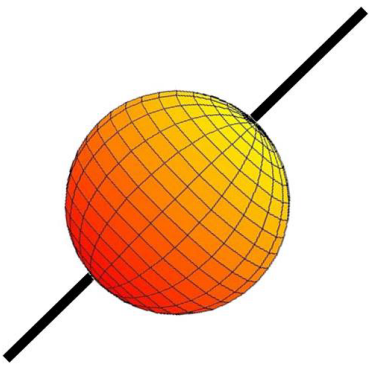


Figure 2.8: The partial volume model

[3]Discussed further in section 4.7

2.5.3 Global tractography

One problem with traditional deterministic or probabilistic tractography algorithms is that they propagate using only the information provided by the local model at each point. This means that local uncertainties lead to either terminations or deviations of the tract or pathway. In contrast global tractography models optimise over all points between regions. The fundamental idea is based on finding the shortest path or curve between regions according to some metric.

One example is the fast marching approach (FM) [132, 133]. This generates a wavefront at each seed point which unfolds fastest at points where principal eigenvectors are closely aligned. This leads to a range of possible tract end points, where the most likely are seen as those where the front moves fastest. Shortest paths are then found by following the gradient of the front arrival time.

Recent alternatives to FM have been generated via geodesic approaches [85] or approaches which optimise functionals designed to minimise curvature [149]. These take into account the naturally curved geometry of the tracts. Also, Zalesky et al [195] attempt to solve a Euclidean shortest path problem.

In contrast the approach of Iturria-Medina et al [83] solves a shortest path algorithm for a weighted graph, where each voxel, $\mathbf{r_i}$, represents a node in the graph and the neighbourhood of each node (characterised by the set of 26 contiguous voxels, j) are linked via edges ($\Delta\mathbf{r_{ij}}$) which are assigned weights (w_{ij}) according to the probability (P_{m}) that each node belongs to white matter and the probability (P_{diff}) that nervous fibres are orientated in the direction of each edge:

$$w_{ij} \equiv w_{ji} = P_{\mathrm{m}}(\mathbf{r_i})P_m(\mathbf{r_j})[P_{\mathrm{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) + P_{\mathrm{diff}}(\mathbf{r_j}, \Delta\mathbf{r_{ji}})] \tag{2.19}$$

Here P_{m} is estimated from probabilistic tissue maps obtained using tissue segmentation[4] and $P_{\mathrm{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}})$ is estimated by integrating the orientation distribution function (ODF: described in more detail in the next section), over a solid angle β around $\Delta\mathbf{r_{ij}}$ (see figure 2.9):

$$P_{\mathrm{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) = \frac{1}{Z}\int_{\beta} \psi(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) \tag{2.20}$$

[4]Described in full in the next chapter, section 3.4

Therefore, $[P_{\text{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) + P_{\text{diff}}(\mathbf{r_j}, \Delta\mathbf{r_{ji}})]$ is a measure of fibre coherence, representative of overlap between ODFs at each voxel and therefore the probability that tracts are aligned along each edge. The most probable path between regions is found by iteratively finding all paths between voxels and finding the path which maximises the sum of weights over all edges whilst penalising path curvature.

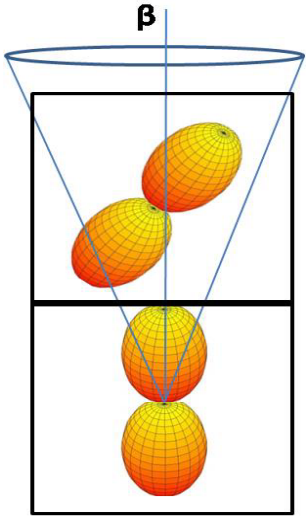


Figure 2.9: Integrating voxel ODFs over a solid angle β. This estimates the probability of diffusion between the voxel centres. The figure is adapted from [83]

2.6 Connectivity scoring

Often it is desirable to assign weights or confidence values to a tract, as this provides a means of comparing tracts between regions, modalities (for example with functional data) or across different subjects. Probabilistic tractography algorithms assess connectivity by estimating the global probability of the target (B) being connected to the seed region (A), given the data (Y). This can be calculated as the proportion of streamlines which reach B ($n_{A\rightarrow B}$) from the total number seeded (n_A):

$$P(\exists\ A \rightarrow B|Y) = \frac{n_{A\rightarrow B}}{n_A} \tag{2.21}$$

Other metrics have also been suggested based on the evolution of wavefronts in FM [132, 133] or geodesic distance between regions [85]. In addition, Zalesky et al [195] quantify paths in terms of the maximum number of direct link-disjoint paths between regions (where link-disjoint means that they share no edges). By contrast, the ConTrack algorithm proposed by Sherbondy et. al. [153] scores pathways according to their length, curvature and end points, inferring that the highest scoring pathways are those most likely to exist.

These approaches reflect the structural integrity of the tissue microstructure in a complex way. A more direct approach is to estimate measures of anisotropy along a tract, for example by taking the mean FA. In contrast, some approaches base their scoring on principles of 'information flow' [83, 195]. In particular Iturria-Medina et al [83] quantify inter-region connectivity in three ways: Anatomical Connectivity Strength (ACS), calculated by combining measures of the maximum arc weight along the path with an estimation of the tract cross section (from the number of seed and target surface voxels it crosses), Anatomical Connection Density (ACD), calculated as the fraction of the surface involved in the connection with respect to the total surface of both areas. It is estimated as the ACS relative to the total number of seed and target voxels. Also, Anatomical Connection Probability (ACP), estimated as the probability that two regions will be connected by at least one tract (equal to the maximum arc weight along the path). This approach has significantly influenced the anisotropy estimate in this thesis. This is described in full in Chapter 4.

2.7 High Angular Resolution Diffusion Imaging (HARDI) and multiple fibre reconstruction schemes

A single fibre model is a good approximation for white matter voxels containing a single fibre population. However, at the resolution scales currently available to diffusion imaging, it is often the case that more than one fibre population will pass through each voxel. In these cases Gaussian diffusion tensors are forced to represent crossing fibre populations by an inflated oblate tensor, misrepresenting the underlying tissue microstructure and hiding the directional information provided by the diffusion signal (figure 2.10). For this reason, a number of alternative schemes have been suggested. Many approaches are built upon high angular resolution diffusion imaging (HARDI) imaging schemes [181], characterised by high b-values and often increased numbers of gradient directions. Increasing b-value, and therefore the diffusion time, increases the sensitivity of the signal to the restricted diffusion component allowing neighbouring fibre populations to be resolved. However, this is traded against a lower signal to noise ratio as signal attenuation is increased. In addition, extra gradient directions increase acquisition times beyond practical limits for many clinical requirements. For this reason there is often a trade-off between optimal reconstruction of the diffusion propagator and clinical practicality of the approach. Typically methods can be split into approaches which attempt to sample the diffusion profile directly (model-free) and those which aim to fit more sophisticated models to the data. Some

of these are discussed below:

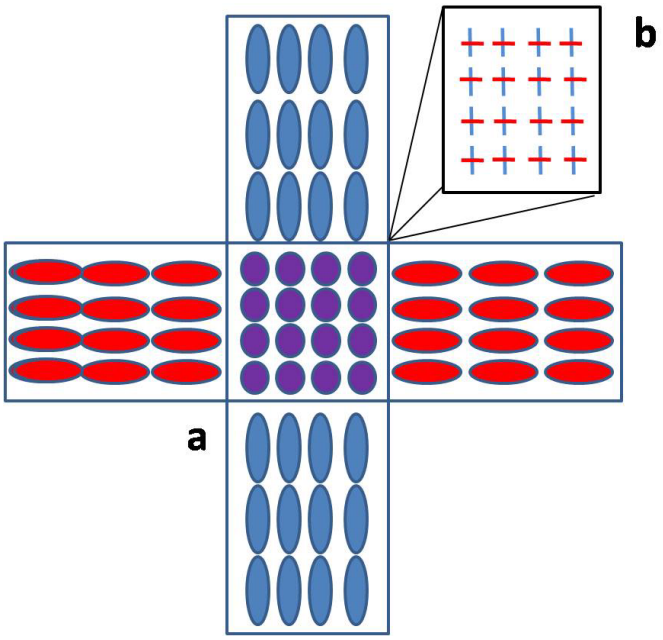


Figure 2.10: Modelling fibre crossing: a) crossing fibres are modelled by tensors with no preferred direction; b) in contrast the multiple fibres algorithm proposed by Behrens et al. [22] fits a principle fibre direction to each fibre population

2.7.1 Multiple fibre models

Gaussian mixture models [2, 181] extend the Stejskal-Tanner model under the assumption that each voxel contains n compartments each containing a different tissue type in proportion a_j , $(\sum_j a_j = 1)$, such that each component can be described by a Gaussian diffusion propagator with diffusion tensor $\mathbf{D}_j$ and there is no mixing of molecules between components. The complete ADC profile for the voxel is then described by a weighted sum:

$$\frac{S}{S_0} = \sum_{j=1}^{n} a_j exp(-\mathbf{bD_j}) \tag{2.22}$$

Unfortunately multiple tensor models are still reliant on Gaussian assumptions and require prior information regarding the number of fibre populations expected in each voxel. By contrast the second generation of the probabilistic tractography algorithm (probtrackx) proposed by Behrens et al [22] adjusts the original approach to account for multiple fibres in voxels for which there is evidence of their existence. This uses a Bayesian technique known as Automatic Relevance Determination (ARD), to fit a more complicated multiple fibre model at each voxel, but ensures (through use of shrinkage

priors on all volume fraction parameters above the first) that these variables are forced to zero where there is little evidence for them in the data. The model is then estimated as follows:

$$S_i = S_0(\sum_{j=1}^{n} f_j exp(-b_i d \mathbf{v_i^T R_j A R_j^T v_i}) + (1 - \sum_{j=1}^{n} f_j) exp(-b_i d)) \tag{2.23}$$

where n is the maximum number of fibres estimated and the remaining parameters are defined as for equation 2.18. This enables reconstruction of multiple fibre populations under clinically achievable acquisition parameters (b=1000s/mm^2; 64 gradient directions). Typically under these conditions the resolution is sufficient for two fibre populations per voxel.

Furthermore, Ozarslan et al [127] have also proposed modifying the Bloch-Torrey equations to provide a generalised solution to the Stejskal-Tanner equation which allows higher rank (l) tensors:

$$ln\frac{S}{S_0} = \sum_{i_1=1}^{3}\sum_{i_2=1}^{3}....\sum_{i_l=1}^{3} bi_1 i_2 ... i_l D_{i_1} i_2 ... i_l \tag{2.24}$$

The approach has the advantage of providing a quantitative measurement of anisotropy for multiple fibres [128], where this measure has been shown to better represent structural integrity than FA in instances of fibre crossing and merging [95].

2.7.2 Model free approaches

Diffusion Spectrum Imaging (DSI) [191] and Q-ball [182] are two forms of q-space imaging, a branch of diffusion imaging which utilises a Fourier relationship between the diffusion signal and the diffusion propagator to sample the diffusion profile directly:

$$P(\mathbf{R}, \tau) = S_0^{-1} \mathcal{F}^{-1}[S(\mathbf{q}|\tau)] \tag{2.25}$$

Here $\mathbf{q} = \gamma \mathbf{g} \delta$ is the spin displacement wave vector, proportional to the duration and magnitude of the applied magnetic field gradient.

In the case of DSI, samples of the signal are made on a Cartesian grid. Orientation structure is recovered by radially projecting the diffusion propagator $P(\mathbf{R}, \tau)$ onto a sphere by integrating over the radial co-ordinate in the diffusion space co-ordinate system. The resulting function $\psi(\mathbf{u})$ (known as the orientation distribution function, ODF) represents the probability of a spin displacing into a

differential solid angle around $\mathbf{u}$.

Unfortunately, the Cartesian sampling scheme used by DSI is highly inefficient, requiring the use of very strong magnetic gradients acquired in very many directions in order to sample the diffusion profile sufficiently finely. For this reason Tuch proposed using a HARDI approach [181, 183] whereby sampling was instead performed directly on a tessellated icosahedral sphere, significantly reducing the gradient sampling requirement. This has the specific advantage that sampling frequency is directly correlated to angular resolution allowing the sampling scheme to be tuned for optimal reconstruction. Furthermore, by using this acquisition scheme, Q-ball imaging [182, 183] effectively bypasses the intermediate reconstruction of the diffusion propagator, reconstructing the ODF directly on the sphere using the reciprocal space Funk-Radon transform (FRT). This is calculated via great circle integration:

$$\psi(\mathbf{u}) = \oint_{\mathbf{q}\perp\mathbf{u}} S(\mathbf{q}) d\mathbf{q} \tag{2.26}$$

where the value of the odf for each unit vector position $\mathbf{u}$ is given by the sum of all values of the interpolated signal around an equator perpendicular to $\mathbf{u}$. Nevertheless, whilst this vastly reduces the number of gradient directions required (typically from ~ 515 for DSI [191] to 252 for Q-ball [182]) the acquisition times are still too long for clinical practises.

Low b-value approximations of the ODF can be made using spherical harmonics. These present a smooth orthogonal basis for approximating functions on a sphere and therefore have been used to derive regularised analytic solutions to the Funk-Radon transform [39, 75] to make robust estimates of ODFs under low b-value clinical acquisition conditions. Spherical harmonics were first suggested in the context of HARDI by Frank [50] who extracted orientation information from multi-Gaussian approximated ADC profiles by fitting a 4th-order simple harmonic (SH) expansion. Truncated SH expansions were also used by Alexander et al [3] to provide a hierarchy of models for the ADC under clinical (low b-value) acquisition conditions. In addition Tournier et al [180] deconvolved a spherical harmonic expansion of the measured signal attenuation with a response function (estimated for a single fibre population in the z direction) in order to derive the fibre ODF.

Finally, the persistent angular structure approach by Jansons and Alexander [84] optimises an energy functional based on the principle of maximum entropy. This attempts to fit a function which best fits the Fourier transform of the signal attenuation on the sphere, minimising the deviance of the function from the sphere subject to the constraints posed by the data.

2.8 Summary

This chapter has provided an overview of the techniques traditionally employed in measuring and modelling the diffusion process in the brain, describing the current limitations and cutting-edge reconstruction approaches used in the field. Diffusion is directional within the white matter of the brain and the degree to which this is true has been shown to correspond to the structural integrity of the underlying microstructure. Tractography techniques can be used to model neuronal pathways and connectivity scoring methods can be used to compare tracts throughout the brain or across subjects. However, in order to understand the functional importance of these pathways it is necessary to correlate the pathways with known histology or functional data. This requires mapping of data between different co-ordinate frames and delineation of seed and target regions for tractography into recognised anatomical subunits. The techniques required to do this are described in the next chapter.

Chapter 3

Image Registration, Segmentation and Atlasing

3.1 Introduction

Before whole-brain tractography can be performed, a number of preprocessing steps are required. These delineate relevant anatomical regions of interest between which tracts are propagated, segment tissue subregions in the brain (to constrain tractography) and provide correspondences between T1 and diffusion imaging data. In addition, performing tractography in a common space also requires transformation and averaging of diffusion data. This is complicated by the fact that it is non scalar. This chapter provides an overview of the image registration, segmentation and atlasing required to perform these tasks. Specific emphasis is given to the techniques applied throughout this thesis.

3.2 Image Registration

Image registration seeks to find the optimal spatial correspondence or mapping between structures in different images by comparing image intensity values. Medical images such as MRI are sampled on a discrete grid of points, formed from 3D volumes (or voxels), whose dimensions are controlled by spatial encoding during acquisition. Optimal mapping is therefore found from the transformation $\mathbf{T} : (x_A, y_A, z_A) \mapsto (x_B, y_B, z_B)$, which aligns images A and B such that all points in A are mapped to their corresponding points in B. Specifically, in the approach used in this thesis, this is defined in

terms of the transformation which maps points in the target image (the frame of reference) to their locations in the source image. This allows image intensities in the source image to be 'pulled-back' from the target co-ordinate frame, implicitly transforming the source image to the target as required.

Medical image registration requires several components, including similarity metrics, interpolation schemes, regularisation and an optimisation approach. In order to find the optimal transformation it is necessary to measure similarity between the transformed source and target images, through comparison of image intensities at each point. For this purpose interpolation is required in order to ensure correspondence between source and target voxel grids after spatial transformation. Generally transformations between similar structures, for example brains in two different subjects are also assumed to preserve topology. This requires transformations to be smooth and diffeomorphic, and is controlled via regularisation. The key approaches and applications of image registration used in this thesis are outlined below. Additional details can be found in [69].

3.2.1 Types of transformation

There are 3 basic types of image transformation: rigid, affine and non-rigid. The type used depends on the nature of the images being registered. Where the images are of firm structures (such as bone) and images are taken from the same subject, rigid alignment is typically sufficient. However, if the same structures are registered between different subjects affine transformations must typically be used, and for more complex transformations, such as those involving soft tissues, or where structure varies significantly across subjects, non-rigid approaches become necessary. Typically each type of transformation is characterised by the degree of freedom with which the object can deform.

Rigid

Rigid transformations preserve shape, specifically distances between points remain the same before and after transformation. In 3D rigid transformations have six degrees of freedom which can be summarised as translations in the x, y and z directions $\mathbf{t} = (t_x, t_y, t_z)^T$ coupled with rotations about each axis (α, β, γ). For example a rigid transformation of the point $\mathbf{x} = (x, y, z,)^T$ can be written as:

$$\mathbf{T_{rigid}}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t} \tag{3.1}$$

where $\mathbf{R}$ is constructed from the angles of rotation.

Affine

If rotations and translations are insufficient to fully describe the transformation affine transformations may be used. These increase the degrees of freedom of the transformation to twelve and add shears and stretches to the variety of available deformations. Shears fix all points along one axis but shift other points parallel to the axis by a distance proportional to their perpendicular distance from the axis. Affine transformations preserve collinearity. This means that points that laid in a line before the transformation continue to do so, and parallel lines are preserved. The result be summarised as a linear transformation ($\mathbf{F}$) followed by a transformation $\mathbf{t}$:

$$\mathbf{T_{affine}}(\mathbf{x}) = \mathbf{F}\mathbf{x} + \mathbf{t} \tag{3.2}$$

or

$$\mathbf{T_{affine}}(\mathbf{x}) = \begin{bmatrix} f_{00} & f_{01} & f_{02} \\ f_{10} & f_{11} & f_{12} \\ f_{20} & f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} f_{04} \\ f_{14} \\ f_{24} \end{bmatrix}$$

where f_{ij} represent the twelve parameters of the transformation.

Typically affine transformations approximate alignment preceding a non-rigid step but are also used in the correction of scanner calibration errors, for example those caused by eddy currents. Eddy currents are induced by the large diffusion gradients, and create additional magnetic fields which survive after the gradient coils are turned off. They add to the spatial encoding gradients causing spins to experience different image gradients from what were programmed. This causes distortions in the reconstructed image, in the form of shears and stretches, primarily in the phase encoding direction which is encoded more gradually. Therefore, affine transformations can be used to correct for this by transforming all diffusion weighted images (taken from a single scan session) to the non diffusion weighted (or b_0) image [99].

Non-Rigid

Rigid and affine transformations are classed as global since all parameters have a global effect. Typically, affine transformations are suited to describing global differences in location, size or shape between objects but are generally insufficient for registering soft, deformable tissues (such as the breast) or mapping the same structure across different subjects. In contrast, non-rigid (or non-linear) registration algorithms significantly increase the degree of deformation of the transformation by specifying different (local) transformations at each point.

In general, non-rigid registration algorithms are constrained by a desire for the resulting transformations to be smooth and invertible. Several different approaches exist: these include techniques which model the material as an elastic [11] or fluid [37] material, where the equations that govern the elasticity or fluid properties of the deformation ensure smoothness. Alternatively some approaches model the deformation using vector fields constrained to be smooth [20]. In addition there are algorithms which constrain the deformation using sets of smooth basis functions or splines [144]. These can be used to provide a smooth approximation to the deformations of regularly spaced grids of control points. The approach, referred to as free-form deformation (FFD), is the technique used throughout this thesis. It is described in more detail below:

3.2.2 Free-form deformations

Free-form deformations rely on the manipulation of an underlying mesh of uniformly spaced control points $\Phi_{i,j,k}$. Displacements at general points in the image are then provided by convolving the control point vectors with a B-Spline kernel. This provides a smooth and continuous transformation. Furthermore, use of a B-spline kernel ensures that the displacement at any location is only dependent on control points in a local neighbourhood. This allows significant local deformations without impinging on the global transformation.

FFDs are calculated as the combination of a local and global transformation:

$$\mathbf{T}(\mathbf{x}) = \mathbf{T}_{\mathbf{global}}(\mathbf{x}) + \mathbf{T}_{\mathbf{local}}(\mathbf{x}) \tag{3.3}$$

Here $\mathbf{T_{global}}(\mathbf{x})$ is provided by a multi-resolution 12 parameter affine transformation, estimated preceding the non-rigid transformation $\mathbf{T_{local}}(\mathbf{x})$. The FFD ($\mathbf{T_{local}}(\mathbf{x})$) can then be estimated as a 3-D tensor product of the basis of 1D cubic B-splines:

$$\mathbf{T_{local}}(\mathbf{x}) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\Phi_{i+l,j+m,k+n} \tag{3.4}$$

where $n_x \times n_y \times n_z$ are the control point spacings, and therefore $i = \lfloor \frac{x}{n_x} \rfloor - 1$, $j = \lfloor \frac{y}{n_y} \rfloor - 1$, $k = \lfloor \frac{z}{n_z} \rfloor - 1$, ensure that only control points in the neighbourhood of the current location contribute to the summation. Here $u = \frac{x}{n_x} - \lfloor \frac{x}{n_x} \rfloor$, $v = \frac{y}{n_y} - \lfloor \frac{y}{n_y} \rfloor$, $w = \frac{z}{n_z} - \lfloor \frac{z}{n_z} \rfloor$, represent the positions where each of the B-splines are evaluated, and B_l represents the l-th basis function of the B-spline function:

$$\begin{aligned} B_0(u) &= (1-u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 - 3u + 1)/6 \\ B_2(u) &= (-3u^3 + 3u^2 - 3u + 1)/6 \\ B_3(u) &= u^3/6 \end{aligned}$$

The spacing of the control points controls the amount of deformation allowed: large control point spacings allow modelling of global nonrigid deformations for the alignment of large features between images, whereas small control point spacings allow highly local deformations for the alignment of fine detail. For this reason FFDs are typically optimised over several resolution levels, where the control point spacings at each new level $(l+1)$ are found by inserting new control points in between the control points of the mesh from the previous round l. Over L levels, the transformation $\mathbf{T_{local}}(\mathbf{x})$ is calculated from the sum of the transformations at each level:

$$\mathbf{T_{local}}(\mathbf{x}) = \sum_{l=1}^{L} \mathbf{T^l_{local}}(\mathbf{x}) \tag{3.5}$$

Finally, smoothness of the transformation is constrained using the following regularisation term:

$$C_{smooth} = \frac{1}{\mid \Omega \mid} \int\int\int_{\Omega} (\frac{\delta^2 \mathbf{T}}{\delta x^2})^2 + (\frac{\delta^2 \mathbf{T}}{\delta y^2})^2 + (\frac{\delta^2 \mathbf{T}}{\delta z^2})^2 + 2[(\frac{\delta^2 \mathbf{T}}{\delta xy})^2 + (\frac{\delta^2 \mathbf{T}}{\delta xz})^2 + (\frac{\delta^2 \mathbf{T}}{\delta yz})^2] \tag{3.6}$$

where Ω represents the domain of the transformation. This represents the 3-D analogue of the bending energy associated with a thin 2-D plate of metal.

Typically, registration attempts to find an optimal transformation which balances image similarity with ensuring that the transformation is smooth.

$$C = C_{similarity} + \lambda_s C_{smooth} \tag{3.7}$$

The balance of the two terms in the cost function is controlled by the parameter λ_s.

3.2.3 Similarity metrics

Similarity metrics attempt to measure registration quality by comparing intensities at equivalent points in each image. They are therefore highly sensitive to whether the images being registered are taken from the same modality or different ones as this significantly influences the relationship between the intensities. Metrics designed to compare same modality images include, sum of square differences (SSD):

$$\text{SSD} = \frac{1}{N} \sum_{\mathbf{x_A} \in \Omega} |A(\mathbf{x_A}) - B^T(\mathbf{x_A})|^2 \tag{3.8}$$

This aims to minimise the sum of squared intensity differences between voxels, $\mathbf{x_A}$, in image A and equivalent voxels in image B (N is the total number of voxels). This is the optimal approach if the two images differ only by Gaussian noise [186]. In addition, cross correlation can also be used:

$$\text{CC} = \frac{\sum_{\mathbf{x_A} \in \Omega} (A(\mathbf{x_A}) - \bar{A}).(B^T(\mathbf{x_A}) - \bar{B})}{(\sum_{\mathbf{x_A} \in \Omega} (A(\mathbf{x_A}) - \bar{A})^2 . \sum_{\mathbf{x_A} \in \Omega} (B^T(\mathbf{x_A}) - \bar{B})^2)^{1/2}} \tag{3.9}$$

if it is assumed that the relationship between the intensities in each image are linear. In this instance $\bar{A}$ and $\bar{B}$ are the mean intensities in images A and B respectively.

Both CC and SSD are limited to intra-modality registrations as they assume a simple relationship between voxel intensities in each image. Therefore, where inter-modality registration is desired, information theoretic techniques must be used. These methods examine the joint histogram formed by binning intensity values from each voxel in A against the intensity value at the corresponding voxels in image B. They incorporate measurements of entropy [150]:

$$H = -\sum_i p_i \log p_i \ , \tag{3.10}$$

where i represents all possible states of the system and p_i represents the probability of each of those states occurring.

One example of this is the joint entropy (JE). This is calculated from the joint histogram by first normalising the histogram to form the joint probability distribution function (JPDF) p_{AB}:

$$H(A,B) = -\sum_a \sum_b p_{AB}(a,b) \log p_{AB}(a,b) \tag{3.11}$$

Here a and b refer to the values of discrete bins. The more similar the images are, the lower the joint entropy (JE). Therefore optimisation is performed by minimising JE.

Unfortunately joint entropy is highly sensitive to the amount of air within the overlapping image domain, as increasing the amount of air increases the contribution to the lowest intensity bin values, thus reducing the total entropy. Mutual information ($I(A,B)$) metrics overcome this by also considering the entropy contributed by each registered image, as well as the joint entropy:

$$I(A,B) = H(A) + H(B) - H(A,B). \tag{3.12}$$

In effect mutual information can be seen as a measure of how well each image describes the other and therefore should be maximised at alignment. In reality, mutual information can still be sensitive to overlap, specifically of low intensity areas, and therefore a normalised mutual information [170] approach is more typically used:

$$\tilde{I}(A,B) = \frac{H(A) + H(B)}{H(A,B)} \tag{3.13}$$

3.2.4 Interpolation

Similarity is estimated between equivalent points in the source and target image. However, since it is unlikely that there is direct correspondence between target and source voxel grids after transformation, it is necessary to resample the transformed source image such that its voxel grid corresponds to that of the target image. This process is known as interpolation. The main types employed in this thesis are: nearest neighbour, where the intensity value of the original voxel centre nearest to the transformed point is assigned to the value of the voxel in the transformed image, and trilinear interpolation, where the interpolated voxel value in the transformed co-ordinate system is calculated as a weighted average of all eight neighbouring voxels from the original system. Weighting falls off linearly with distance.

In this thesis nearest neighbour interpolation is used for transformation of region of interest labels from template brain images to target subjects (see section 6), and for transformation of principle diffusion direction vectors during diffusion registration (see section 7). Trilinear interpolation is used in all other cases.

3.2.5 Optimisation

Optimisation is performed over several stages. First: non-rigid registration is preceded by a multi-resolution affine [169] step. This involves perturbation of the transformation parameters by a chosen step size, where the similarity metric is re-estimated at every perturbation, and the parameter with the greatest improvement is chosen and transformation updated, until no further increase in similarity is achieved. The step size is then reduced and the process repeated until a minimum step size is reached. This procedure is completed for each resolution level, where the initial step size may be linked to the resolution of the data.

Following affine registration, the parameters of the FFD (given by $\Phi_{i,j,k}$) are optimised as a function of the cost function C (eq 3.7) using gradient descent. This makes steps (of size μ) in the direction of maximum increase:

$$\Phi + \mu \frac{\nabla C}{||\nabla C||} \tag{3.14}$$

where,

$$\nabla C = \frac{\delta C}{\delta \Phi} \tag{3.15}$$

until an optimum is reached. The stopping criterion is defined as:

$$||\nabla C|| \leq \epsilon \tag{3.16}$$

where ϵ is defined as a small positive integer.

3.3 Choice of registration parameters

In summary, several parameters influence the performance of the FFD registration used in this thesis. These include:

- Similarity metrics: if images are of the same type then image similarity metrics such as SSD or CC may be used. However, it is common practise to use information theoretic metrics such as NMI for most common registration problems.
- Interpolation mode: in this thesis we use nearest neighbour or trilinear interpolation
- Control point spacing: this determines the amount of deformation possible.
- Number of resolution levels: it is common practice to reduce control point spacing over multiple resolution levels. Typically a maximum control point spacing is defined. This is halved for each resolution level.
- Blurring and downsampling: in tandem with each resolution level it is possible to blur and downsample each of the source and target images. Blurring is performed with a Gaussian kernel. Downsampling by a factor of r resizes image voxels to r times the size of their smallest dimension. It is also possible to blur and resample source and target images by different amounts. This is common practice if they have different resolutions (for example T1 and b_0) images.
- Number of iterations: this controls the maximum number of iterations during optimisation. The number of completed iterations is also controlled by stopping criterion ϵ ($\epsilon = 0.0001$).
- Length (μ) and number of steps (n_s): it is possible to define a different μ for each resolution level. In addition, if the number of steps is set > 1, this will optimise in a 'course-to-fine' fashion, where the optimisation will start with step length μ but will make n_s sub-divisions of the step length during optimisation.
- Smoothing: this is controlled by parameter λ_s. A higher value for λ_s will increase the contribution of the smoothness term in the cost function (equation 3.7). Higher values of λ_s will make the transformation more likely to be invertible, but this may come at the cost of optimal alignment.

The values of relevant parameters will be referenced at appropriate points in each chapter.

3.4 Registration of diffusion tensor data

Registration of tensors is more complex than for scalar intensity images. Specifically tensors must be realigned such that the relationship between the structure of the tensor and the surrounding tissue is preserved. The key aim of tensor realignment is for the first eigenvector to align with the direction of maximal diffusion in each voxel and the shape and size of the tensor to be preserved, since they reflect properties of the tissue micro-structure.

Many different diffusion registration algorithms have been proposed, based on variety of registration approaches [5, 4, 31, 55, 61, 129, 197, 194]. However, since reorientation significantly complicates the calculation of the registration objective function, most approaches perform registrations on scalar, or rotationally invariant features [4, 5, 55, 61, 129] prior to realignment of the tensor field. Registration is typically driven using intensity based similarity metrics. However, some studies [55, 4] also suggest tensor based metrics such as:

$$\text{Tensor difference (TD)} = ||\mathbf{D_1} - \mathbf{D_2}||_C = \sqrt{Tr((\mathbf{D_1} - \mathbf{D_2})^2)} \tag{3.17}$$

$$\text{Tensor scalar product (TSP)} = \mathbf{D_1} : \mathbf{D_2} = Tr(\mathbf{D_1} \cdot \mathbf{D_2}) \tag{3.18}$$

This is the sum of products of corresponding tensor elements.

Following spatial alignment, preservation of tensor shape is ensured by use of a rigid transformation:

$$\mathbf{D'} = \mathbf{RDR^T} \tag{3.19}$$

where $\mathbf{D}$ is the original estimate of the tensor, $\mathbf{R}$ is a rotation matrix representing a transform and $\mathbf{D'}$ is the new tensor after rotation. If the underlying transformation is rigid, this is straightforward to determine. However, when the method is extended to higher order transformations, it becomes necessary to extract rigid rotations from the global or local transformation at each point. For affine transformations $\mathbf{F}$, this can be achieved by decomposing the affine transformation matrix, $\mathbf{F}$, into a rigid rotation component, $\mathbf{R}$, and a pure deformation component, $\mathbf{U}$:

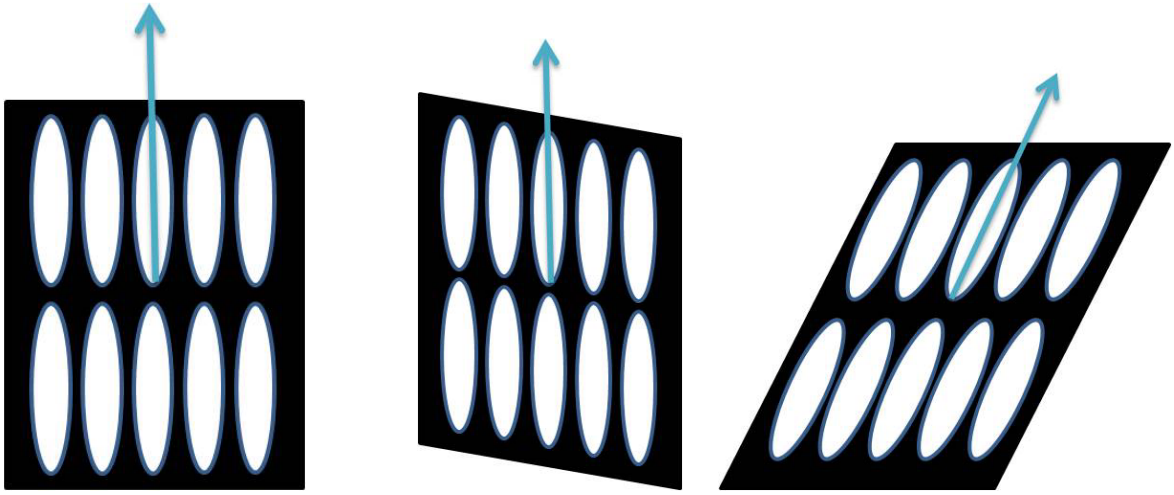
$$\mathbf{F} = \mathbf{UR} \tag{3.20}$$

This is called the finite strain method [5], and arises from the polar decomposition of $\mathbf{F}$ into the

product of a single unitary matrix $\mathbf{R^{-1}} = \mathbf{R^T}$ and a positive definite hermitian matrix $\mathbf{U}$, where $\mathbf{U}$ is unique and given by $\mathbf{U} = \sqrt{\mathbf{F^T F}}$. Therefore:

$$\mathbf{R} = \mathbf{F U^{-1}} \quad \mathbf{R} = \mathbf{F}(\mathbf{F^T F})^{-1/2} \tag{3.21}$$

However, this does not account for the affect of shears and stretches on tensor orientation (see figure 3.1).



(a) (b) (c)

Figure 3.1: Influence of affine transformations on tensor orientation: a) original image, no shear; b) vertical shear, as tensors are already pointing along the vertical axis the transformation does not affect their orientation; c) horizontal shear, this changes tensor orientation

Therefore, Alexander et al. [5] suggest an alternative, the preservation of principal directions (PPD) approach. This uses the full affine transformation $\mathbf{F}$ for the reorientation, but corrects for any deformation via normalisation. The approach works by first finding the rotation matrix $\mathbf{R_1}$ that maps the first eigenvector, $\mathbf{e_1}$, onto a unit vector $\mathbf{n_1}$ given by:

$$\mathbf{n_1} = \frac{\mathbf{F e_1}}{|\mathbf{F e_1}|} \tag{3.22}$$

where the axis and angle of this rotation are derived from the vector and scalar products of $\mathbf{e_1}$ and $\mathbf{n_1}$, and the rotation matrix is calculated from Rodrigues' rotation formula. This is followed by a secondary rotation $\mathbf{R_2}$, which is required to map $\mathbf{e_2}$ from the position it has reached after the first rotation ($\mathbf{R_1 e_2}$) into the plane spanned by $\mathbf{n_1}$ and $\mathbf{n_2}$:

$$\mathbf{n_2} = \frac{\mathbf{F e_2}}{|\mathbf{F e_2}|} \tag{3.23}$$

This retains orthogonality with $\mathbf{e_1}$. The projection of $\mathbf{n_2}$ is is calculated as: $\mathbf{P}(\mathbf{n_2}) = \mathbf{n_2} - (\mathbf{n_2}.\mathbf{n_1}).\mathbf{n_1}$. Then $\mathbf{R_2}$ is calculated about the axis defined by $\mathbf{R_1 e_1}$ and the angle is calculated from the dot product of $\mathbf{R_1 e_2}$ and $\mathbf{P}(\mathbf{n_2})/\mathbf{P}(\mathbf{n_2})$. The transformation $\mathbf{R} = \mathbf{R_2 R_1}$ completes rotation of the tensor.

For non-rigid transformations, a locally affine mode ($\mathbf{Fx} + \mathbf{b}$) can be approximated from the non-rigid transformation $\mathbf{T}(\mathbf{x})$, using the Jacobian operator $\mathbf{J}$. This represents the matrix of first order partial derivatives:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial T_x}{\partial x} & \frac{\partial T_x}{\partial y} & \frac{\partial T_x}{\partial z} \\ \frac{\partial T_y}{\partial x} & \frac{\partial T_y}{\partial y} & \frac{\partial T_y}{\partial z} \\ \frac{\partial T_z}{\partial x} & \frac{\partial T_z}{\partial y} & \frac{\partial T_z}{\partial z} \end{bmatrix}$$

The relationship can be seen from differentiation of both $\mathbf{T}(\mathbf{x})$ and $(\mathbf{Fx} + \mathbf{b})$ with respect to $\mathbf{x}$. As the FFD approach estimates transformations from the target voxels to points in the source image, the required rotation should be calculated from the inverse of the affine transformation. Following this FS or PPD can be applied as before. Note, for small deformations the FS approach is a good approximation of PPD [197].

Alternatively, tensor reorientation can be incorporated directly into the analytic objective function [31, 194, 197] of a non-rigid registration algorithm. For example, Cao et al. [31] propose calculation of the exact gradient of the PPD reorientation to drive a diffeomorphic non-rigid registration approach. Similarly, Yeo et al. [194] present an exact gradient for FS reorientation. In contrast Zhang et al. [197] propose a piecewise affine registration approach with inbuilt tensor re-orientation. The whole brain volume is split into small templates which are affinely registered as part of a global optimisation process. Image similarity is assessed in terms of a metric sensitive to the anisotropic component of the diffusion tensors only and FS reorientation is used to drive optimisation and re-orient the tensors after each affine registration. However, PPD is used to perform actual re-orientation of the tensors following spatial transformation. Typically these approaches improve alignment.

In the next sections we describe several applications which employ image registration.

3.5 Atlas Building

One specific application for co-registration of images is the use of atlases. These provide a standard representation of anatomy or function to which other images can be compared. Traditional atlases, such as the Talairach atlas [174] are based on a single segmented subject. However, these are strongly biased towards the arbitrary choice of subject and do not enable creation of models of statistical variability. Therefore, many atlases are now generated via the averaging of large numbers of subjects. The most widely used example of this is the MNI152 (Montreal Neurological Institute) brain atlas [46, 47]. This is derived from a cohort of 152 healthy young adults.

Atlas construction takes many forms: principally these are defined by the choice of reference space and spatial transformation used. Stereotaxic frameworks have the advantage of providing a standard procedure for aligning unseen anatomical images, without requirement of a single reference image. For example, the Talairach stereotaxic space [175] defines a point of origin in each image as the intersection between one line passing through the superior aspect of the anterior commissure and the inferior aspect of the posterior commissure (AC-PC line), and a second passing perpendicularly through the posterior aspect of the anterior commissure (VAC line) (figure 3.2 a). Axes are defined as x-axis: left-right; y-axis: posterior-anterior; z-axis: inferior-superior (see figure 3.2)[1]. Images are transformed by aligning AC-PC lines in all images and performing piecewise linear scaling.

Traditionally the Talairach stereotaxic space is used as a frame of reference for comparison with the Talairach anatomical atlas, formed from histological analysis of the single hemisphere of a brain post-mortem. In contrast, MNI atlases are built via averaging of multiple images, co-registered to a similar stereotaxic framework [48]. However, in this system, the AC-PC line is defined from common points along the midsagittal plane. Also the most superior and lateral points on the cortex are defined by perpendicular bisectors of this line. Images are aligned by affine transformation and averaged to generate a first estimate of the atlas. This then forms a target image in a second stage.

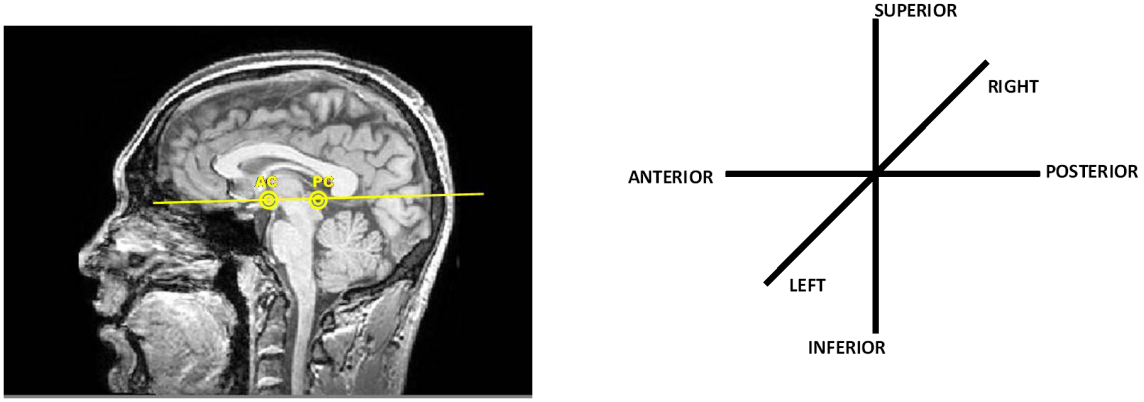
Stereotaxic approaches are limited by the requirement for manual identification of features along the AC-PC line. In addition linear transformations are often insufficient to describe the full range of possible deformations, especially during development and ageing, where the brain changes dramatically. Therefore there is now a trend towards building atlases using either small [26, 143, 146] or large [87] deformation models. Many of these use pair-wise registration between images, either registering all

[1] Other co-ordinate systems are also used. For radiological co-ordinate systems the ordering of the x-axis is swapped

images to a standardised template (such as the MNI atlas) or choosing a reference image from the pool of subjects. Unfortunately, in instances where the population is highly variable this is unlikely to lead to a representative average.

Population variability can be minimised by choosing the image most similar to all others [100]. Alternatively, Guimond et al [64] calculate average transformations and use these to create a new template for registration. Many approaches now seek to use group-wise registration removing template bias entirely. In Bhatia et al. [26] a FFD approach is used to simultaneously optimise registration of all images under the constraint that the sum of all deformation fields should be zero. Similarly, Joshi et al [87] use Large Deformation Diffeomorphic Metric Mapping (LDDMM) to both minimise squared error dissimilarity between images, and restrict total deformation [87].

Recently, some approaches have moved away from using a single template image to describe a whole population. Sabunco et al [146] address the issue of significant variability in some populations by using mixture modelling to cluster images into several template images or atlases, each of which demonstrate greater similarity to their constituent training images than for an atlas built using all images. In addition, methods have also been developed based on manifold learning [56]. For this a distance metric is proposed based on diffeomorphic transformations between images. The minimum diffeomorphic distance is then used to embed images onto a lower dimensional manifold which better characterises the geometric variability in the data. This allows creation of a local atlas for images that are nearby on the manifold.



(a) (b)

Figure 3.2: The Talairach co-ordinate system: a) T1 sagittal slice image with AC-PC line shown; b) axis defining Talairach co-ordinate system aligned with the sagittal image in (a).

3.5.1 Diffusion tensor derived atlases

Diffusion tensors do not form a vector space. This means that linear averaging does not preserve the natural properties of tensors: for example the linear averages of a set of tensors with the same determinant may result in a tensor with a larger determinant. It has been proposed that the space of tensors is better represented as a curved manifold or Riemannian symmetric space, and that tensor averaging and interpolation be performed using Riemannian metrics [49, 134]. In particular Log-Euclidean metrics proposed by Arsigny et al. have been used in the averaging of fibre tracts [34] and in diffusion tensor atlas construction [61, 129]. In this approach, interpolation and averaging are treated as weighted sums:

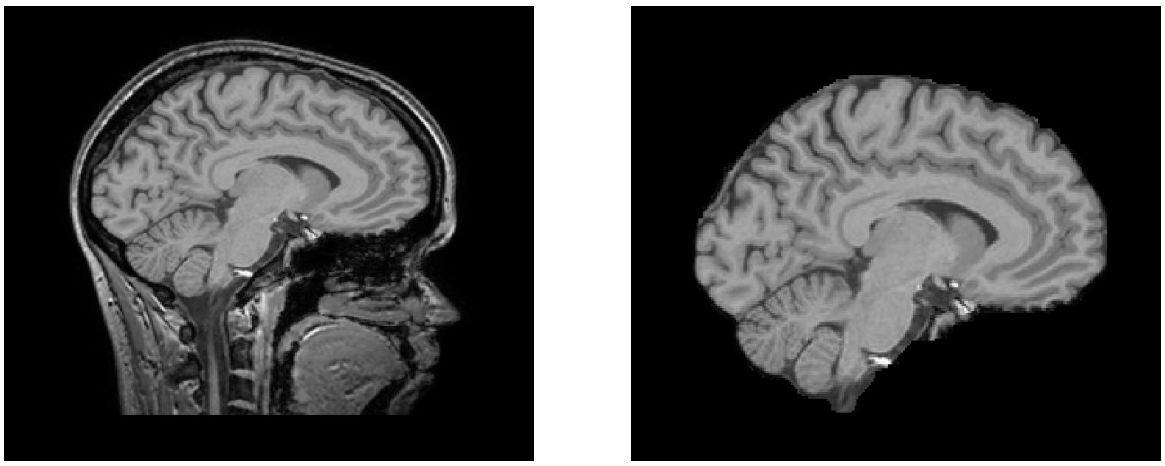
$$\hat{D} = \exp(\sum_{i=1}^{N} w_i \log(D_i)) \tag{3.24}$$

Here $\hat{D}$ is the log-Euclidean Frechet mean.[2]

Using these techniques several studies have generated atlases of diffusion images. Most use prior registration of scalar valued intensity images. These include an approach by Park et al. [129] which performs multichannel registration using all the coefficients of the diffusion tensor, and averages according to the approach of Guimond et al [64]. In addition, Goodlett et al. [61] propose an intermediate solution between using baseline images and tensor based metrics. They create a structural operator from the maximum eigenvalue of the Hessian of each FA image which is shown to exaggerate fibre bundles. They register these by optimising the sum of squared differences.

In contrast, Zhang et al. [197] use a piecewise affine approach and build tensor re-orientation into the optimisation algorithm. They avoid explicit interpolation of tensors by using a standard approach proposed by Little et al [94]. This constrains rigid structures to move subject to a user defined rigid body transformation, whilst deforming the space around them non rigidly. This enables the structures of tensors to be maintained whilst the space (provided by the affine template) is deformed to fit the target co-ordinate frame. They also propose construction of a unbiased white matter atlas using the approach of Joshi et al [87]. This maximises image similarity using tensor metrics whilst minimising the total amount of deformation used.

[2]The Frechet mean is the Riemannian equivalent of a Euclidean mean.



Original T1-weighted image | Image after skull stripping using BET

Figure 3.3: Examples of brain extraction performed using the BET tool

3.6 Segmentation

Segmentation techniques allow delineation of the brain into different tissue types or anatomical sub-regions. This allows studies to be performed which investigate variation in tissue volume, composition or shape across subjects. Currently, the gold-standard of segmentation is provided by expert manual delineations, made either post-mortem [174] or computationally [70]. However, these are time consuming and subject to inter and intra rater variability. Therefore automated approaches are increasingly being sought. Some of the main techniques are described below.

3.6.1 Tissue classification

Tissue segmentation algorithms seek to classify image voxels as white matter (WM), grey matter (GM), cerebral spinal fluid (CSF) and non brain matter (figure 3.4). One technique for the removal of non brain matter is the brain extraction tool (BET) [155] from FSL (figure 3.3). This estimates the surface of the brain by use of a surface model. The model centres a tessellated sphere on the estimated centre of gravity for the brain. This is then expanded outwards one vertex at a time until a minimum intensity threshold is reached. Expansions are constrained to be smooth. In addition, improvements can be made by taking an iterative approach where the centre of gravity is updated each time.

Tissue classification approaches typically model GM, WM and CSF intensities using Gaussian mixture models. Segmentation based on image intensities alone is possible by defining dynamic ranges for each tissue type. However, there is considerable overlap between tissue intensity distributions and intensity

ranges vary depending on the scanning protocol used. Therefore tissue types are more commonly modelled as a mixture of Gaussians.

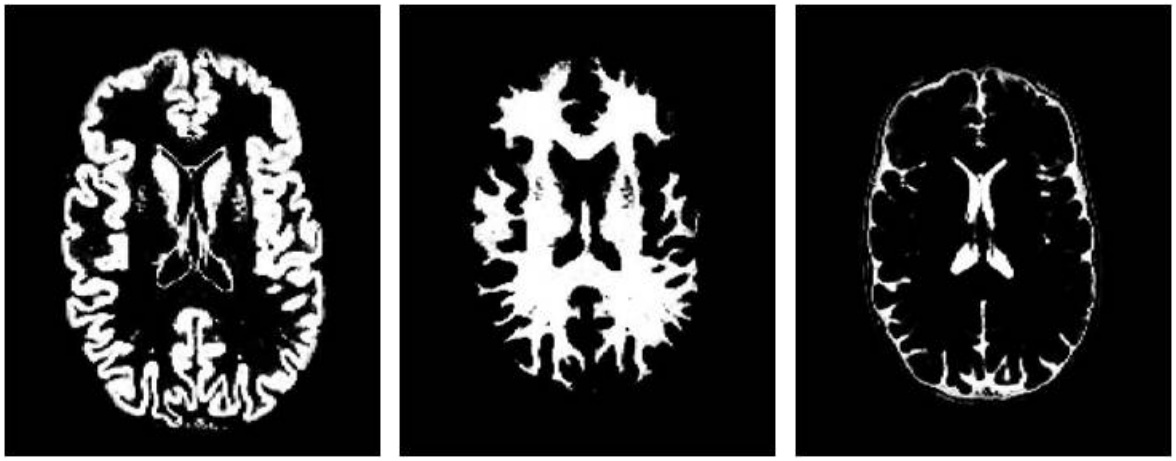
Some approaches also incorporate spatial priors. These are probabilistic tissue maps [47, 70], built by transforming multiple expert segmentations (templates) to a common (generally stereotaxic) space, and calculating the proportion of templates that agree at each voxel. In the statistical parametric mapping (SPM) [8] implementation (used in this thesis) spatial priors are combined with a tissue intensity model within a Bayesian framework. This is optimised using a expectation-maximisation approach. A similar approach is also used by Van Leemput et al. [92, 91]. However this also incorporates a Markov Random Field (MRF). MRFs assume that the states of voxels are related only to their neighbours. This is a reasonable model for tissue segmentation, as it is assumed that neighbouring voxels will generally belong to the same class.

Many approaches also incorporate corrections for intensity non-uniformity (bias) into the EM framework. Image bias is an artefact of image acquisition which modulates voxel intensities with distance from the image coils and therefore significantly influences the process of modelling tissue class distributions for the whole brain. SPM corrects for bias by including extra parameters in the objective function to account for the smooth intensity variations which result. Underpinning this is a parametric model based on multiplicative noise:

$$s_i = \frac{(y_i + n_i)}{\rho_i} \tag{3.25}$$

where each voxel's signal (s_i) is assumed to be a combination of some uncorrupted signal (y_i) and Gaussian noise (n_i), scaled by some spatially varying bias (ρ_i). This is optimised alongside estimation of tissue class memberships. Van Leemput models the bias field using polynomial basis functions [91].

An alternative model for tissue segmentation, not used in this thesis, is FAST (FSLs Automated Segmentation Tool) [198]. Here the underlying intensity distribution is still assumed to be Gaussian and initial intensities are estimated from the image histogram using a linear discriminant measure-based thresholding approach [126]. However instead of using spatial priors, the approach uses a Gaussian Hidden Markov Random Field model (GHMRF). Hidden MRFs are stochastic processes generated by MRFs whose state sequence cannot be observed directly but can be observed through a field of observations. In FAST the hidden MRF represents class labels and the observable field, the voxel intensities. Bias estimation is incorporated into the model via the approach of [192]. This also assumes multiplicative addition of bias.



(a) (b) (c)

Figure 3.4: Tissue segmentation: a) grey matter, b) white matter, and c) CSF segmentations. All performed using SPM.

3.6.2 Anatomical segmentation

Anatomical segmentation is more problematic then tissue segmentation as regional boundaries, especially within the cortex, are not accompanied by intensity change. In addition, inner grey matter structures vary widely in intensity often leading to mis-classification of their tissue type using conventional tissue classification approaches [8].

A number of alternative approaches have been proposed. These include deformable (or active contour) models [102]. These use an explicit template model of the structure and treat it like an elastic body, deforming surface points (or landmarks) according to an energy functional which adapts it to the image data, using for example the image gradient, whilst stabilising shape based on smoothness constraints. Similarly, level set approaches [36] deform an implicit shape representation provided by a distance transform. Level sets were developed to overcome some of the limitations of (explicit) deformable models which are sensitive to initialisation and susceptible to overlap or intersection of control points.

Statistical shape models are also commonly used. These model shape variation in a population by extracting the mean shape and modes of variation from the training data. Typically correspondences between shapes are made by alignment of control points or landmarks, where landmarks for each shape can then be extracted to form feature vectors for dimensionality reduction. Principal component analysis, or similar techniques can then be used to extract the major modes of variation from which test structures can be segmented by fitting the model to the new data.

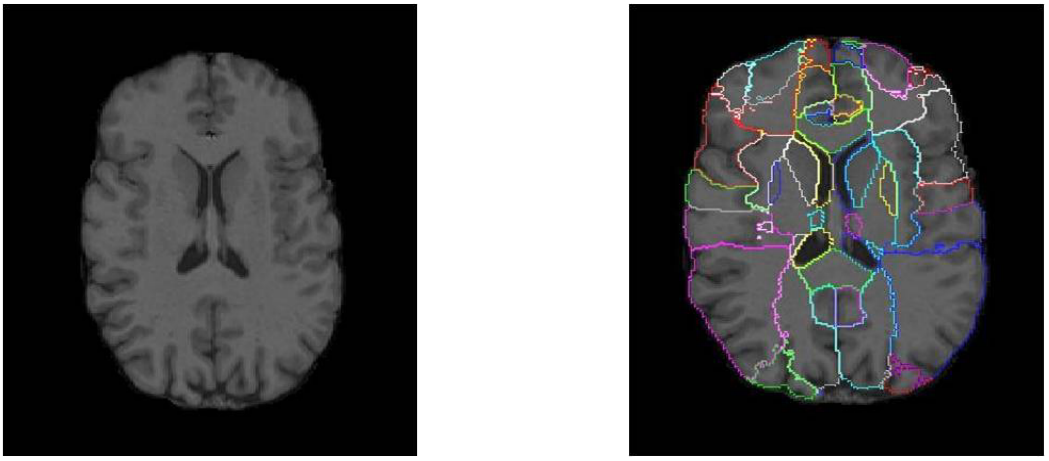


Figure 3.5: Example of anatomical segmentation performed with label propagation and fusion.

Other approaches include normalised cuts [154] or spectral graph approaches. These convert estimates of intensity similarity to a distance measure and use these to create a weighted graph, where voxels are connected by edges whose weights are inversely proportional to their similarity. These can then be used to make a partition of the image where very similar pixels are clustered together.

Nevertheless all these procedures require clear differences in image intensity between structures. This can be overcome by use of additional features as part of a classification approach such as boosting, or support vector machines (SVM) [113] (described in more detail in Chapter 6). Features may include spatial priors, tissue classification memberships and curvature estimates amongst others. These can be combined to learn the correct image structure, which can then be used to segment unseen data.

Alternatively, semi-automated procedures based on propagation and fusion of labels from manually delineated atlases [6, 63, 74, 189] are independent of image intensity. They require registration of a pre-segmented image to a target image, followed by transformation of the anatomical labels from the template image to the target (see figure 3.5). Improved accuracy is observed if several segmentations are transformed and their labels combined using weightings [189] or decision fusion rules [139]. In addition, identification of sub-sets of pre-segmented images that are most similar to the target (for example by comparison of their mutual information in a standard template space), can also improve results [6].

Atlas-based segmentation approaches tend to perform best for cortical segmentation. Therefore all anatomical segmentations in this thesis are based on this approach. However new approaches are being proposed which use tractography for the purposes of thalamic [23] and cortical segmentation purposes [138]. These are likely to contribute to future work (chapter 8).

3.7 Summary

This chapter has described the principles of image registration and related techniques of atlasing and segmentation. Image segmentation plays a fundamental part in the extraction of brain networks in this thesis as it is used to identify seed regions for tractography and to constrain the tractography to the white matter. In addition, registration is required to transform the segmentations (performed on structural T1 data) to the diffusion space where tractography is traditionally performed. This is described further in the next chapter. Later both approaches are combined to register diffusion data to a standard or atlas space in which tractography is performed. This is described in Chapter 7.

Chapter 4

Modelling Whole-Brain Structural Connectivity

4.1 Introduction

The human brain is a system of immense computational capacity and complexity: functional brain states are modulated by networks of an estimated 10^{11} neurons and 10^{16} connections. Improved understanding of how functional brain states emerge and equally how they fail in instances of brain damage or disease is reliant on an accurate mapping of this underlying substrate. However, complete characterisation of neural architecture at this scale is technically impossible. Therefore, structural descriptions of neural architecture at this time must be built over multiple resolution scales [162], from individual fibre bundles to macroscopic pathways between regions.

At the macro-scale, diffusion tractography techniques provide a non-invasive and increasingly high resolution means of reconstructing models of underlying tissue architecture. This chapter presents a framework for modelling whole brain connectivity from diffusion data acquired under clinical scanning conditions. The aim is to provide a consistent framework which will allow comparisons of whole-brain structural connectivity to be made across populations. The approach builds upon a variety of previous studies, discussed below in section 4.2. In the next chapter we present graph theory, the most commonly used tool for analysis of brain networks at this time.

4.2 Background

Studies suggest that the thickness of cortical grey-matter tissue reflects the size, density and arrangement of the cells, and is correlated for regions connected by the same neuronal bundles [116]. For this reason the first estimates of structural connectivity were made by correlating measurements of cortical thickness [71]. These estimate the closest distances between various points on the inner and outer grey-matter surfaces. Distances are then averaged over 54 anatomical regions assigned by propagation of labels from a volumetric template, and connectivity is estimated by correlating mean thickness estimates between pairs of regions and performing multiple comparison testing to test the significance of the correlation across the population. The connectivity matrix is then thresholded on the basis of significance.

Diffusion tractography provides an estimate of tracts or paths between regions of the brain and therefore provides a more direct measurement of connectivity. The first approach for modelling whole brain connectivity with tractography was proposed by Hagmann et al [67] using diffusion spectrum imaging (section 2.7.2). This requires acquisition of 515 gradient directions with a b-value of 12000s/mm^2. Orientation distribution functions (ODFs) are fit at each voxel and multiple tractography streamlines are seeded for every ODF peak. This enables high resolution reconstruction of diffusion profiles, largely overcoming problems with fibre crossing. In the original approach, Hagmann et al. [67] assess cortical connectivity between 500-4000 ROIs assigned at random on the white matter-grey matter boundary of the cortex by region growing. This is later related to 66 (subdivided) cortical labels to enable comparison with recognised anatomy [65].

Unfortunately DSI and other HARDI approaches are not viable for many studies due to the necessarily high gradient strengths and long acquisition times. Therefore other studies have looked at finding ways of estimating whole brain networks from diffusion tensor imaging. In Gong et al. [59] streamline tractography is performed on 80 subjects through a lattice of tensors fit to diffusion data acquired in 6 directions with a b-value of 1000s/mm^2. The limitations of streamline tractography are overcome in part by defining 78 sub-cortical regions within a 2mm^3 region of the grey matter tissue boundary. This ensures that tract propagation does not fail due to low diffusion anisotropy before reaching the boundary of a target. In addition, acknowledging the limitations of the data, Gong et al propose that a single network is built from the entire population of subjects by using hypothesis testing. In this way the network retains only connections that occur with high regularity across subjects. These

connections can be assumed to exist with high confidence.

In contrast, Iturria-Medina et al. [83] incorporate connectivity weights into the tractography process. In this approach orientation distribution functions, calculated analytically from diffusion tensors, are compared between voxels in order to estimate fibre coherence and hence probability of diffusion between voxels. These are combined with tissue probability maps to form edge weights between all neighbouring voxels, and tractography is performed by solving a most probable path problem which aims to find paths which maximise fibre coherence whilst penalising path curvature. Paths are traced between all voxels and grey matter voxels are clustered into subsets of anatomical regions. Finally, connection weights are assigned according to the anatomical connection strength, probability and density terms discussed previously in section 2.6. Our proposed approach is inspired by similar ideas and is discussed below.

4.3 Methods for extracting whole-brain tractography from clinical data

The aim of this chapter is to present an approach for extracting whole brain connectivity networks for large populations of clinical data acquired using low b-values and few gradient directions. This requires three key components: a) a method for automated segmentation of seed regions for tractography; b) a tractography approach suited to the data and able to contend with a range of clinical populations; and c) a means of quantifying connectivity between regions.

Region of interest segmentation is achieved via atlas-based segmentation. We use label propagation and the decision fusion approach proposed by [6]. Tractography is performed using a widely available probabilistic tractography algorithm[1] [24]. This allows tracts to be followed even through areas of significant uncertainty in tract projection. Finally, connectivity between regions is assessed via an estimate of mean tract anisotropy. This is influenced by the approach of Iturria-Medina et al. [83].

The key contributions of this chapter are the incorporation of the tract anisotropy measure into the probabilistic tractography approach, and the combination of these processes to form a fully automated approach for estimating whole-brain tractography for clinical datasets. The outline of the framework is listed below.

[1] http://www.fmrib.ox.ac.uk/fsl

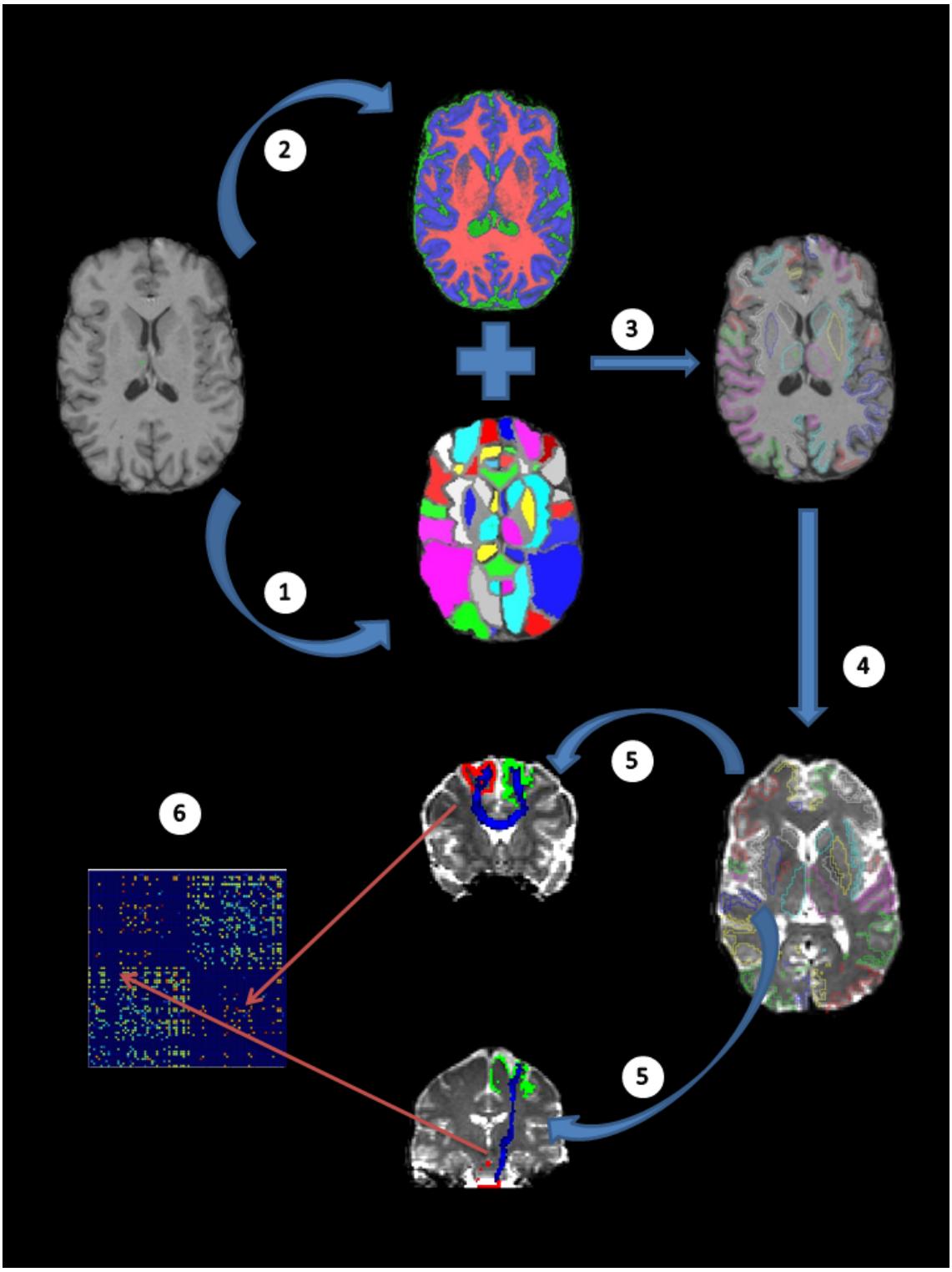


Figure 4.1: Outline of the whole-brain tractography framework

4.3.1 Outline of framework

The following summarises all processing steps taken in extraction of the structural networks. This is displayed graphically in figure 4.1:

1. 83 anatomical regions of interest (ROIs) are automatically segmented using label propagation from multiple atlases (Figure 4.1: Step 1).
2. Probabilistic tissue segmentations are derived using SPM [8] (figure 4.1: Step 2).
3. Anatomical ROIs are merged with tissue maps to retain cortical ROIs only at the white-matter-grey matter boundary. Deep GM structures are left intact. Boundaries of all regions are eroded to minimise influence of segmentation and/or registration errors (figure 4.1: Step 3).
4. ROIs are transformed to the diffusion space (figure 4.1: Step 4).
5. Tractography is performed between each pair of ROIs in turn with only direct connections being retained (Figure 4.1: Step 5). If required, tracts can be thresholded to contain only the most probable connections.
6. Results are represented as scalar values in 77×77 connection matrices. Since DTI is incapable of distinguishing between afferent and efferent biological connections symmetry is enforced (figure 4.1: Step 6).

The individual component steps are now described in more detail.

4.4 Region of interest segmentation

In previous approaches seed regions have ranged from hundreds of small ROIs placed randomly on points of the cortex, to relatively small numbers of anatomical regions propagated from an atlas template. This framework adopts the latter approach as this allows results to be linked back to functional neuroanatomy and facilitates comparisons across subjects. However, in contrast to previous approaches, multiple manually delineated templates are used and label propagation [82] and decision fusion [88] are employed. This ensures that each voxel gets classified according to the brain region it has the greatest probability of belonging to.

Experiments have shown that fusing labels from multiple templates, for example using a majority vote rule, improves the similarity between automated and manual segmentation over approaches that propagate labels from a single template. In particular, Heckeman et al. [74] compare the mean similarity index (SI) for 67 anatomical regions and demonstrate an increase from 0.754 ± 0.016 (label fusion) to 0.836 ± 0.013 (single template). Here, the SI is the ratio between the overlap and the mean volume for manual and automatic segmentations. Aljabar et al. [6] show that this is further improved if only the most similar template images were selected for fusion. Some examples of atlas segmentations, and the results of label fusion are shown in figure 4.2.

In this framework, atlas-based segmentation is achieved by non-rigid registration [144] of the T1 images of 30 atlases to each subject's T1 image. This is performed over four resolution levels, with control point spacing 20mm, 10mm, 5mm, 2mm, $\lambda_s = 0$ (no additional regularisation). The overlap of each manually segmented T1 image is compared to the target image using normalised mutual information, and the labels from the 10 best matching template images are propagated to the subject's T1 image and merged using a majority vote rule [6, 140]. This generates 83 anatomical regions per subject covering the entire brain volume (for region labels see table 4.1)

Once anatomical labels have been assigned, cortical ROIs are redefined at the white matter-grey matter interface only. This is found by comparison of anatomical labels with tissue segmentations obtained using SPM software [8]. White matter is removed because nerve pathways are known to propagate between grey matter regions. In addition, using only the boundary reduces the number of seed points and therefore the computational time required to propagate all tracts. The corpus callosum is also removed to allow connections to pass from the left hemisphere to the right hemisphere and vice versa, and 5 ROIs representing the ventricles are removed, reducing the total ROIs used to 77. In contrast, boundaries of inner grey matter structures are left intact. This is on account of recognised problems [8] with grey matter segmentation of structures such as the thalamus and putamen, where tissue intensities do not lie within the expected distribution for cortical grey matter (see figure 4.3). Finally, all ROIs are eroded at the boundaries between regions (by one voxel). This reduces the effects of partial volume segmentation and registration errors.

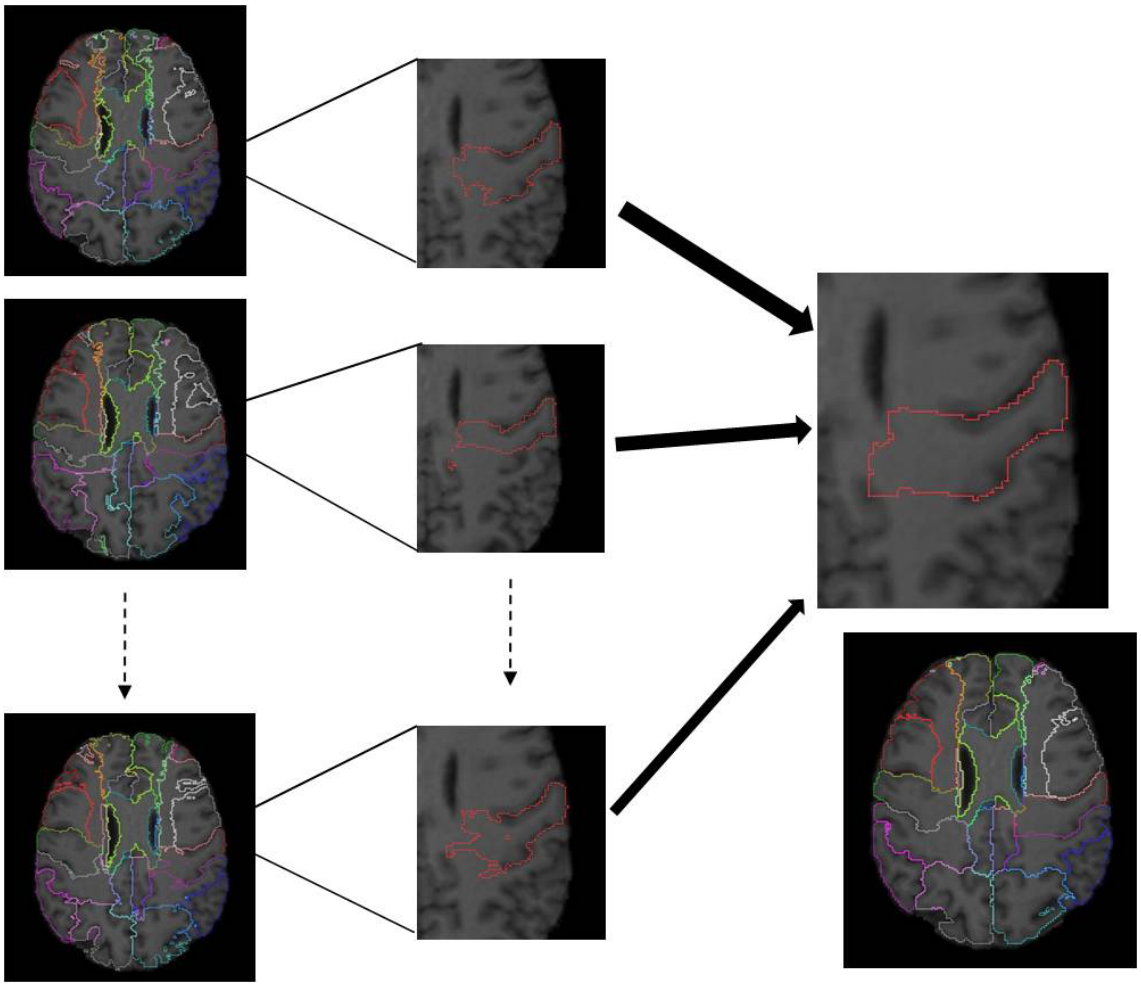


Figure 4.2: Label fusion. Multiple manual templates (shown left) are fused together using a majority vote rule to form final segmentation of the target image (bottom right). The central figures show several templates of the pre-motor cortex transformed onto the target, and the result after fusion.

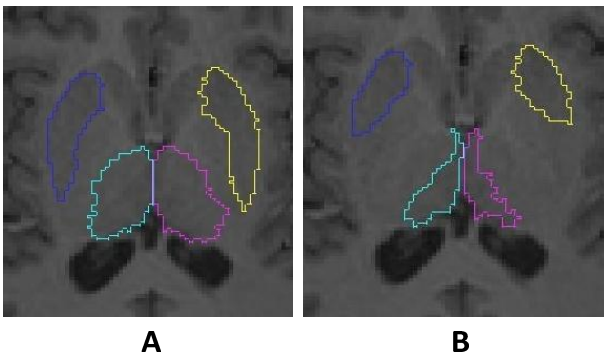


Figure 4.3: Grey matter segmentation of the putamen and thalamus by SPM: a) without tissue segmentation b) with tissue segmentation

4.4.1 Registration to diffusion space

Following segmentation, the ROIs are transformed from each subject's structural T1 image to the coordinate frame of the diffusion data in which tractography is performed. This is typically performed by registering the brain extracted T1 data to the non diffusion weighted (b_0) scans, acquired as part of the diffusion imaging protocol. However, this is complicated by the comparatively low resolution and signal to noise content of the b_0 image.

One way to overcome this is to include a multi-resolution rigid step, preceding affine and non-rigid registrations. Image registration can be improved by bias correction to both the T1 and b_0 images. Alternatively, if the T2 image is available, intermediate registration of the T1 image to the T2 image can also simplify the registration. This is because T2 scans are generated using spin echo imaging protocols and therefore have a similar image intensity profile to diffusion images. Typically, non rigid registration is performed over multiple levels, where the T1 or T2 images are blurred and downsampled more than the b_0 image at each stage. Image similarity is maximised according to measurements of normalised mutual information [170], and brain extraction is performed using the brain extraction tool (BET) from FSL [155]. In the next sections the tractography approach is presented.

4.5 Choice of tractography approach

The aim of this framework is to present an approach for extracting whole brain connectivity networks for large populations with clinical data acquired using 15 gradient directions and a b-value of 1000s/mm^2. This prohibits the use of many high angular resolution diffusion (HARDI) imaging reconstruction schemes that would otherwise allow optimal reconstruction of the data. In addition streamline tractography approaches are susceptible to failure in areas of low diffusion anisotropy. This limits their application to many clinical populations such as neonates. In contrast, probabilistic tractography algorithms enable tracts to be followed even through areas of significant uncertainty in tract projection. For this reason tractography in this framework is performed using a standard probabilistic tracking algorithm [24]. This is available as part of FSL.

There are two key differences between this approach and standard streamline tractography: first, the probabilistic approach [24, 22] does not simply fit a best estimate of the diffusion model at each voxel but rather seeks a distribution. This is done by using Bayesian inference to estimate

posterior distributions on the principal direction (θ, ϕ) of diffusion, where the posterior distributions of individual parameters are estimated from the joint distribution by Markov Chain Monte Carlo (MCMC) sampling. This enables tracts to be propagated through areas of high diffusion uncertainty. Second, rather than adopting a diffusion tensor model, it uses a two-component partial volume model (the so-called ball and stick model), which represents anisotropic diffusion as a single vector in the principal fibre direction (θ, ϕ). This has magnitude f $(0 \leq f \leq 1)$ otherwise known as the volume fraction. The isotropic diffusion is represented by a spherical component (magnitude $(1 - f)$).

Probabilistic tractography between regions (known as seed and targets) is performed conventionally by seeding multiple probabilistic streamlines from each voxel in the seed region. These sample from the posterior distribution on principal diffusion direction at each point and continue until a target is reached or the termination criterion is met. This results in a range of possible end points for each tract. Termination occurs where a curvature threshold is exceeded or tracts reach a target.

In this instance, the underlying algorithm is modified to force tracts to halt at the first target they reach. This ensures connections are only retained if the tract passes directly between the seed and target, and ensures consistency with a graph theory representation, where the distance between any two nodes (regions) must take into account the number of nodes bypassed en route. This is explained in more detail in Chapter 5. Tissue probability maps are also incorporated. This prevents tracts from passing through tissue with a less than 50% probability of being white matter, and reduces the influence of partial volume effects in the diffusion data. Therefore, tracts cannot pass across sulci (cortical folds) through the CSF. In addition, connectivity is quantified according to scalar estimates of global probability and mean tract anisotropy. The probability of a target region (B) being connected to the seed region (A), given the data (Y), can be calculated as the proportion of streamlines which reach B $(n_{A \to B})$ from the total number seeded (n_A):

$$P(\exists\ A \to B|Y) = \frac{n_{A \to B}}{n_A} \tag{4.1}$$

The mean tract anisotropy measure is described in the next section (equation 4.7)

Results are indexed within 77x77 connectivity matrices, where the columns represent seed regions and rows represent targets. Following the completion of tractography, connectivity matrices are symmetrised by setting the connection strength along any single connection to its maximal value. This is because diffusion imaging is not able to distinguish the direction that the fibres are projected.

4.6 Estimating anisotropy along tracts

Relating differences in connection probabilities to structural changes in the underlying tissue can be difficult as the global probability measure relates to the microstructure in a complex way: it is influenced by fibre crossings, region size, distance between regions, and also the proportion of streamlines seeded from each voxel that project to different target regions. This makes variations across subjects difficult to interpret. In contrast, measurements of diffusion anisotropy have been proposed to reflect changes in myelination, fibre density and packing [98] and act as effective indicators of structural integrity in several developmental [103], ageing [14, 89, 135] and neurodegenerative [40] studies.

As the underlying model of diffusion in this framework is not based on a tensor model, commonly derived measures such as FA should not be employed as it can no longer be guaranteed that tract propagation aligns with the principal eigenvector of the tensor model. Furthermore, FA is a scalar measure and therefore cannot be used to estimate anisotropy along the tract.

Instead, this framework adopts a similar approach to Iturria-Medina et al. [83], where connection weights between voxels at each step during tractography are assigned from an estimate of the probability of diffusion across them. The measure is not only related to the diffusion anisotropy at each voxel but also takes into account estimates of fibre coherence. It can therefore be used to estimate mean diffusion and hence anisotropy along tracts.

In Iturria Medina et al.[83] (and preliminary work (see publications [1]) ODFs were calculated from diffusion tensors. However, this does not correlate with the underlying model. Therefore, in this framework, the ODFs are calculated directly from all samples on the volume fraction of diffusion ($f_{kl}^{(\theta,\phi)}$), allowing direct correspondence with the underlying fibre model. This can be estimated from the partial volume model of diffusion at the Monte Carlo sampling stage as demonstrated in [105]:

$$\psi(\theta,\phi) = \frac{1}{N_s}\sum_{k=1}^{n}\sum_{l=1}^{N_s} f_{kl}^{(\theta,\phi)} \quad . \tag{4.2}$$

Here, n is the total number of fibres, N_s refers to the total number of MCMC samples and $f_{kl}^{(\theta,\phi)}$ is the volume fraction of diffusion for iteration l at orientation (θ,ϕ) for each fibre direction k. Therefore, the ODF (ψ) in the direction (θ,ϕ) reflects the probability of diffusion in that direction.

We require the proportion of the ODF which overlaps with the unit vector $\Delta\mathbf{r_{ij}}$ ($\Delta\mathbf{r_{ij}} = \mathbf{r_i} - \mathbf{r_j}$) which

joins voxel $\mathbf{r_i}$ to the centres of adjacent voxels $\mathbf{r_j}$. This is contained within the solid angle ($\beta = 4\pi/26$), and approximates the probability of exchange ($E^j_{\text{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}})$) between voxels. It is estimated from proportion of the volume fraction $f^{(\theta,\phi)}_{kl}$ within (β):

$$E^j_{\text{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) = \frac{1}{N_s}\sum_{k=1}^{n}\sum_{l=1,(\theta,\phi)\in\beta}^{N_s} f^{(\theta,\phi)}_{kl} \ ; \tag{4.3}$$

This yields 26 components per voxel which reduces to 13 due to symmetry e.g. $\psi(\theta,\phi) = -\psi(\theta,\phi)$):

$$E^j_{\text{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) = E^{-j}_{\text{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) \ ; \tag{4.4}$$

$$\sum_{j=1}^{13} E^j_{\text{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) = 1 \tag{4.5}$$

Estimates of anisotropy ($0 < a_{i\to j} < 1$) between adjacent voxels can then be determined by summing the exchanged proportions in both directions:

$$a_{i\to j} = \frac{1}{2}(E^j_{\text{diff}}(\mathbf{r_i}, \Delta\mathbf{r_{ij}}) + E^{-j}_{\text{diff}}(\mathbf{r_j}, \Delta\mathbf{r_{ji}})) \ . \tag{4.6}$$

and the mean anisotropy (ζ) along each tract T is calculated by multiplying each voxel value by the times it is sampled during tracking (w) and averaging.

$$\zeta = \frac{1}{r_{tot}}\sum_{\forall\mathbf{r}_i\in T}\sum_{j=-13}^{13} w_{i\to j}a_{i\to j} \ . \tag{4.7}$$

Here r_{tot} refers to the total number of voxels sampled. Note, we have used an average rather than the minimum value as this consistently generates higher classification rates in the statistical analysis.

4.7 Interpolation

Probabilistic tractography assumes a continuous representation of tract trajectory. However samples from the posterior distribution on the fibre directions are taken from a discrete grid. Therefore interpolation is applied such that at any position during tractography, the probability of sampling from each neighbouring voxel is weighted according to their proximity to current position of the tract. Specifically, for the x-dimension, the probability of choosing data from $\lfloor\mathbf{x}\rfloor$ is $g(\lfloor\mathbf{x}\rfloor|\mathbf{x}) = \lceil\mathbf{x}\rceil - \mathbf{x}/(\lceil\mathbf{x}\rceil - \lfloor\mathbf{x}\rfloor)$, and from $\lceil\mathbf{x}\rceil$ is $g(\lceil\mathbf{x}\rceil|\mathbf{x}) = 1 - g(\lfloor\mathbf{x}\rfloor|\mathbf{x})$. The same approach is also applied to

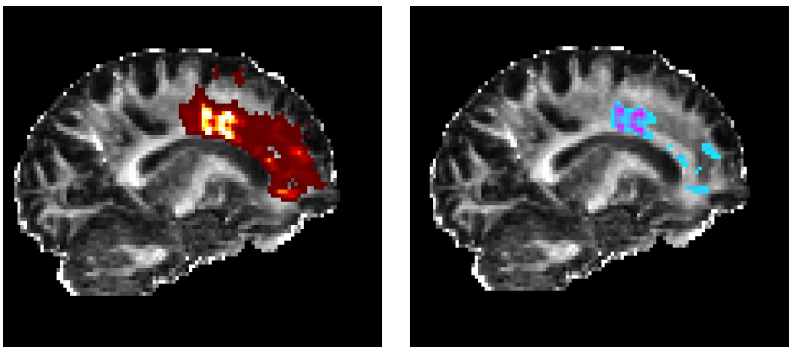
both y and z dimensions.

The goal of the mean tract anisotropy parameter (ζ) is to obtain a continuous estimate of the probability of diffusion along the tract, with the assumption that this is correlated with tract anisotropy. Therefore the parameter $a_{i \to j}$ must also be interpolated. This is achieved by fitting ODFs at tract positions for each step during the tractography. These correlate with the voxels sampled from using the probabilistic interpolation scheme, where for iteration k, $\mathbf{\Delta r_{ij}}$ is determined from the direction of projection of the tract between iteration k and $k+1$.

4.8 Thresholding

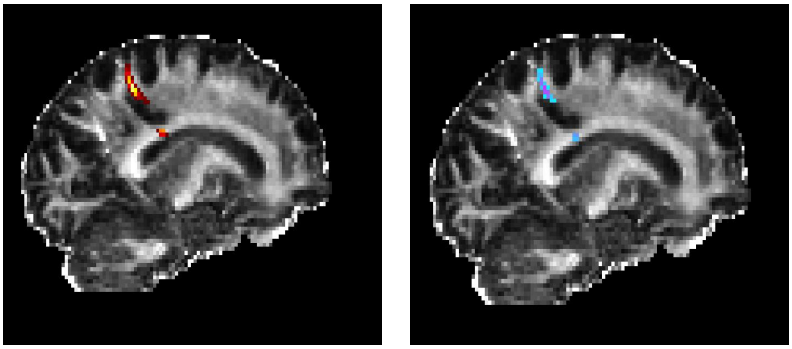
Probabilistic tractography enables tracking through areas of low certainty. Therefore it is possible that some tracts propagated in this way will represent false positives. Nevertheless, setting an absolute probabilistic threshold to prune tracts is problematic: the probability of connection is influenced by the presence of noise and fibre crossings which can lead to a lack of certainty in expected tracts; for example, tracts between frontal and posterior regions in the brain. Furthermore, probability distributions are likely to vary between seed regions and across subjects. A fixed probabilistic threshold is therefore likely to add further bias to the results. For this reason, we have not chosen to threshold connection matrices in this way for any population study performed in this thesis.

It is common practise within many probabilistic tractography studies to threshold individual tract distributions. This is performed by an adaptive approach which removes voxels from the tract if the number of times they are sampled falls below a user-defined percentage of the maximum. It is possible to incorporate this into our framework by setting a fixed threshold for all tracts. This should ensure that tracts with large number of streamlines and widely distributed trajectories are thresholded, whereas tracts with fewer projections are unaffected (figure 4.8). However, this may have a varying effect across different subjects/populations. Therefore, we use it only for the common co-ordinate space approach in chapter 7



(a) Before thresholding (b) Thresholded

Thalamus-sensory cortex



(c) Before thresholding (d) Thresholded

Contralateral motor cortex

Figure 4.4: Tract thresholding. Top: the contra-lateral superior frontal cortex tract a) without thresholding (shown in red-yellow) b) with thresholding at 5% of the maximum sampling density shown in purple-blue). Bottom: the contra-lateral sensory cortex tract a) without thresholding b) with thresholding. The different image intensities of the tracts represent changes in global probability value at each point along the tract (yellow/light blue is high intensity). Thresholding significantly reduces tract dispersion for the contra-lateral superior frontal cortex tract but has no affect on the contra-lateral sensory cortex tract. This suggests that it would be practical to use a single fixed threshold for all tracts.

4.9 Experiment to compare the performance of the model under different parameterisations

The performance of the probabilistic tractography algorithm is linked to several key parameters: a) the curvature threshold, default = 0.2 $\approx \pm 80°$. This much higher than in conventional streamline algorithms as it allows tracts to progress through areas of high uncertainty. b) The number of streamlines seeded (default = 5000). This leads to a distribution of pathways. c) The number of gradient directions (typically between 15 and 60). These influence the ability of the scan to resolve patterns of diffusion. d) The fibre model (single or multiple). This is typically correlated with the number of gradient directions.

4.9.1 Subjects

In the experiment, whole brain tractography was performed on six subjects under a variety of parameterisations. Scanning was performed on a Philips 3 Tesla system. T1 images were acquired prior to diffusion weighted imaging, and single shot echo planar DTI was acquired in 60 non collinear directions using the following parameters: TR 12000ms, TE 51ms, slice thickness 2mm, voxel size = 1.75 x 1.75 x 2mm^3, b value 1000s/mm^2. Four b_0 images were also collected. Preprocessing was carried out on all data sets according to the steps described above (figure 4.1). Whole brain tractography was then carried out multiple times for each subject. The following parameters were varied:

- Number of gradient directions: 15 gradient direction data (plus one b_0) were sub-sampled from data acquired in 60 directions. The selected samples were chosen to correspond directly with the gradient directions used in a conventional 15 gradient direction scan.
- Number of fibres fitted per voxel: one and two fibre models were fitted. This is in accord with Behrens et al [22] which suggests that for b-values of 1000s/mm^2, the maximum number of fibres that can be reliably resolved using this approach is two.
- Curvature threshold: this was run for curvature thresholds of 0.2 (default), 0.5, 0.7. These equate to maximal turning angles of approximately $\pm 80^\circ$, $\pm 60^\circ$ and $\pm 45^\circ$ respectively.

The impact of changing different parameters was examined in two ways. Firstly, via calculation of correlation coefficients (CC) between each pair of binary connection matrices, where indices were set to 1 if a connection was shown to exist and 0 otherwise. The aim was to test the choice of parameters under the assumption that at this (macroscopic) scale connections should occur consistently across all subjects. Therefore, higher agreement across subjects should indicate fewer false positive connections. Secondly, the propagation of tracts were compared for 15 and 60 gradient directions, where one fibre and two fibre models were used respectively. Difficult to track connections were examined, including connections between the substantia nigra and motor cortex, which form part of the cortico-spinal tract (identified as difficult to propagate in [22]), contra-lateral motor and sensory connections [41], which are sensitive to fibres crossing from frontal to occipital regions, and frontal and occipital lobe tracts (selected for the same reason).

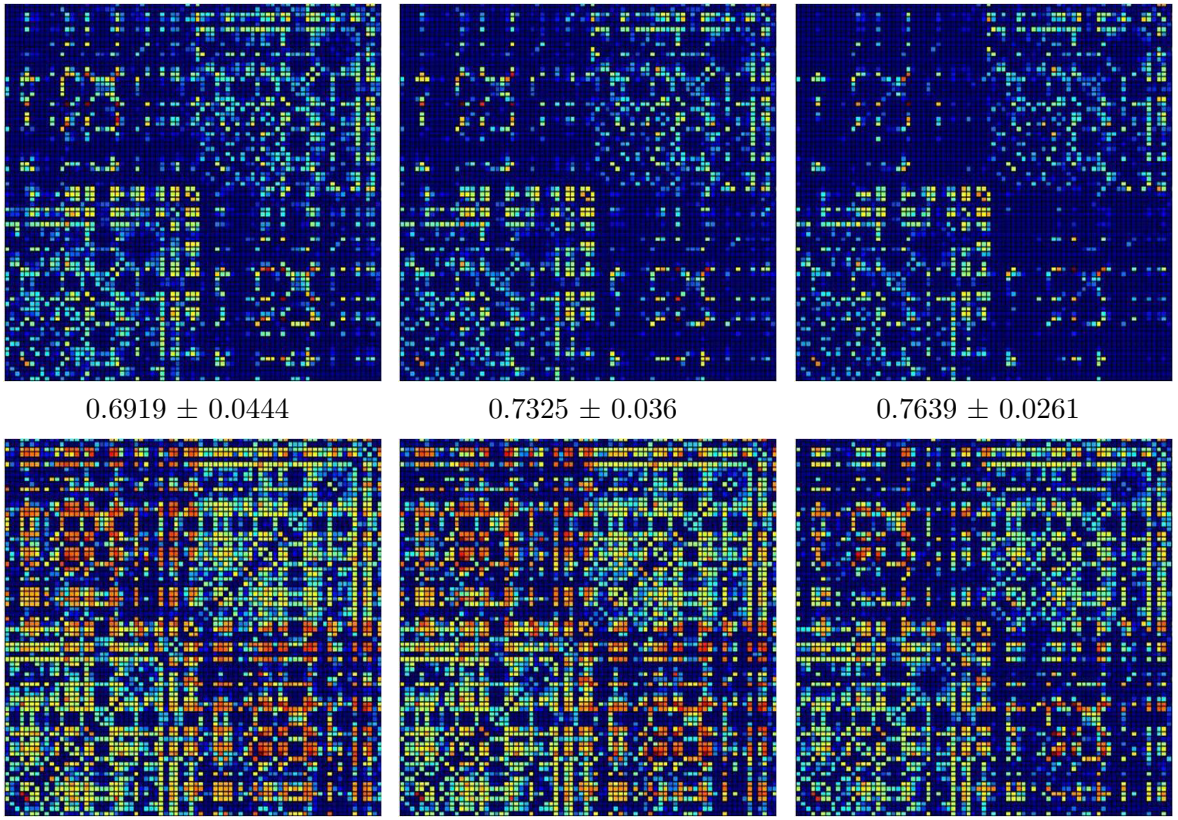


Figure 4.5: Comparison of connectivity matrix generation for 15 (top row) and 60 (bottom row gradient) directions. Results are presented for curvature thresholds (from left to right): 0.2, 0.5 and 0.7. Values represent mean correlation across all subjects for each parameterisation. Increasing curvature threshold reduces connection density and slightly increases correlation. In addition, connection density increases significantly when 60 gradient directions are used in conjunction with a 2 fibre model. This is to be expected as the 2 fibre model will significantly increase the number of possible paths each tract may take. However, as the paths represent a probability distribution on the most likely projection of tracts from each seed region, but are not thresholded due to possible bias, this may reflect and increase in false positive connections

4.9.2 Results of correlation analysis

For all parameterisations, correlation across subjects was in agreement with the results of Hagmann et al. [65], who report cross subject correlation of 0.78, and Iturria-Medina et al. [158] who report inter-subject correlation at 0.65 (weighted graphs). Mean connectivity matrices for 15 gradient directions (one fibre model) and 60 gradient directions (two fibre model) are shown in figure 4.5.

Increasing the number of gradient directions has limited impact on the density of connections for a one fibre model. Similarly fitting a two fibre model to 15 direction gradient data has little impact on mean connectivity. This reflects the fact that 15 gradient direction data does not support a two fibre model. In contrast, fitting a two fibre model with 60 gradient directions significantly increases

connection density.

The curvature threshold is correlated to connection density and in all cases increasing the curvature threshold (decreasing curvature angle) leads to an increase in cross subject correlation. The results for 60 gradient direction data are particularly dense at low curvature thresholds. However, previous studies have all reported sparse connection matrices [65]. Therefore, this may reflect an increase in the number of false positive connections.

4.9.3 Comparison of 15 and 60 direction data

It has previously been shown [22] that increasing the number of fibres fitted at each voxel (in correspondence with the number of gradient directions) improves the ability of the approach to propagate tracts through areas of fibre crossing. This is supported by the results shown in figures 4.6 and 4.7, which demonstrate that using 60 gradient direction data significantly improves the success of the framework at reconstructing contra-lateral motor and sensory, and cortical spinal tracts. Connections between frontal and occipital lobes are also propagated successfully for most 60 gradient direction data sets (figure 4.9) whereas this is not possible with 15 gradient direction data.

The success of the 60 gradient direction two fibre model approach to propagate tracts in areas of high uncertainty also leads to a significant increase in the total number of tracts propagated. This may reflect an increase in false positive connections. Therefore, it is useful to consider whether tracts can still be propagated for 60 gradient directions when a lower curvature threshold is used. Figures 4.8 and 4.9 show results for tracts propagated under a curvature threshold constrained to 0.5. These show that, in most instances (with the exception of the right motor cortex to substantia nigra, figure 4.8 a and d, and frontal-occipital tracts in figure 4.9) tract propagation is unaffected by reducing the curvature threshold.

Some of the failures may be attributed to the constrains of the framework, which prevents tracts passing through other regions between the seed and designated target. Alternatively, it may reflect mislabelling or registration errors during automatic segmentation. Nevertheless, it is worth noting that, where tracts do fail, the impact may be reduced by successful propagation of tracts in the opposite direction. Therefore, failed propagation is compensated for in many cases.

4.10 Summary

This chapter has proposed a framework for modelling whole brain connectivity from clinically acquired diffusion data using probabilistic tractography. The choice of tractography algorithm has been motivated by limitations placed by the data. Clinical data is typically acquired using few gradient directions and low b-values and therefore is unsuited to analysis using HARDI approaches as the data lacks sufficient angular resolution. In addition, simple streamline tractography approaches are unsuited to many clinical data sets as they are unable to track through areas of low diffusion anisotropy. In contrast, probabilistic tractography enables reconstruction of tracts in areas of fibre crossing [22], grey matter and low diffusion anisotropy [35].

The purpose of the framework is to provide a model of connectivity through which differences in tissue microstructure can be compared along tracts throughout the brain. However, global probability measures are influenced by region size. Therefore a parameter has been introduced to the model that is independent, and also reflects the structural integrity of the underlying microstructure. This has been interpolated in order to generate a continuous estimate of the probability of diffusion along the full length of each tract. It is hoped that this will enable more effective comparison of tracts across populations. This property is tested further in subsequent chapters.

Prior to tractography, regions of interest are defined by propagation and fusion of labels from multiple manually delineated atlases. This is an automated approach which vastly improves the speed of segmentation over manual delineation, and provides a consistent labelling system which facilitates straightforward comparisons across subjects. However, whilst label fusion has been shown to increase the accuracy of the segmentation over a single template approach, it is not as accurate as manual segmentations. As such any errors made during segmentation may be propagated through to the tractography results. This may lead to inconsistent propagation of tracts across subjects.

The main population studied throughout this thesis consists of 173 subjects acquired using 15 gradient direction diffusion data and a b value of 1000s/mm^2. Therefore the framework has been designed primarily with a single fibre model in mind. Nevertheless it has been important to test that the approach is extendible to higher resolution data sets. Therefore the framework has been tested under a variety of parameterisations. These have investigated the impact of changes to the curvature threshold, fibre model and gradient sequence.

The experiments have shown that, for both 15 and 60 gradient directions, cross subject correlation

of the derived connectivity matrices matches that of previous studies [67, 158]. However, 15 gradient direction data is unsuited for propagating tracts known to be sensitive to fibre crossing. In contrast, connectivity propagated using 60 gradient direction data is more successful. This suggests that comparative studies performed using 60 gradient data will be able to include a more comprehensive reconstruction of the brain's connections.

Connection probabilities vary across tracts due to factors such as fibre crossing, and differ between subjects. Therefore we have chosen not to threshold connectivity matrices. Instead, we have constrained tracts through use of tissue information and erosion between regions. Nevertheless, it is clear that this approach may not be appropriate for 60 gradient direction data, where connection density is very high. In this instance experiments have supported a possible reduction in curvature threshold (figures 4.8, 4.9), but suggest that this may lead to reduction in the number of tracts successfully propagated. A completely different approach may be to incorporate thresholds based on prior knowledge, similar to Jbabdi et al. [86]. Here, a connectivity prior based on functional or histological data could be used to discourage tracts from being propagated where they are not expected and aid propagation of predicted tracts where they are. This is likely to form a key part of future work.

In the next Chapter 5, the potential of the framework is tested by applying graph theory analysis, a technique commonly used in whole brain connectivity studies. A new approach for studying changes to brain organisation during healthy ageing is also presented.

Table 4.1: Region of interest labels: 38 regions per hemisphere plus brainstem

Anterior Orbital Gyrus (AOG)	Superior Frontal Gyrus (SFG)
Anterior Temporal Gyrus Lateral (ATGl)	Superior Parietal Gyrus (SPG)
Anterior Temporal Gyrus Medial (ATGm)	Superior Temporal Gyrus Posterior (STGP)
Anterior Cingulate Gyrus (ACG)	Superior Temporal Gyrus Anterior (STGA)
Posterior Cingulate Gyrus (PCG)	Amygdala (AMYG)
Frontal Lobe (Fr)	Caudate Nucleus (CAU)
Inferior Frontal Gyrus (IFG)	Cerebellum (CER)
Lateral Occipito-Temporal Gyrus (LOT)	Cuneus (CUN)
Lateral Orbital Gyrus (LOG)	Gyri Parrahippocampus (GPH)
Lingual Gyrus (LING)	Hippocampus (HIP)
Medial and Inferior Temporal Gyrus (M&IT)	Insula (INS)
Medial Orbital Gyrus (MOG)	Nucleus Accumbens (NucA)
Occipital Lobe (OCC)	Pallidum (PAL)
Parietal Lobe (PAR)	Putamen (PUT)
Post Central Gyrus (PoCG)	Sub-Callosal Area (SubC)
Posterior Orbital Gyrus (POG)	Substantia Nigra (SubN)
Posterior Temporal Gyrus (PTG)	Thalamus (THAL)
Pre Central Gyrus (PreCG)	
Pre-subgenual Frontal Gyrus (PSgF)	
Straight Gyrus (SG)	Brainstem (Br)
Subgenual Frontal (SgF)	

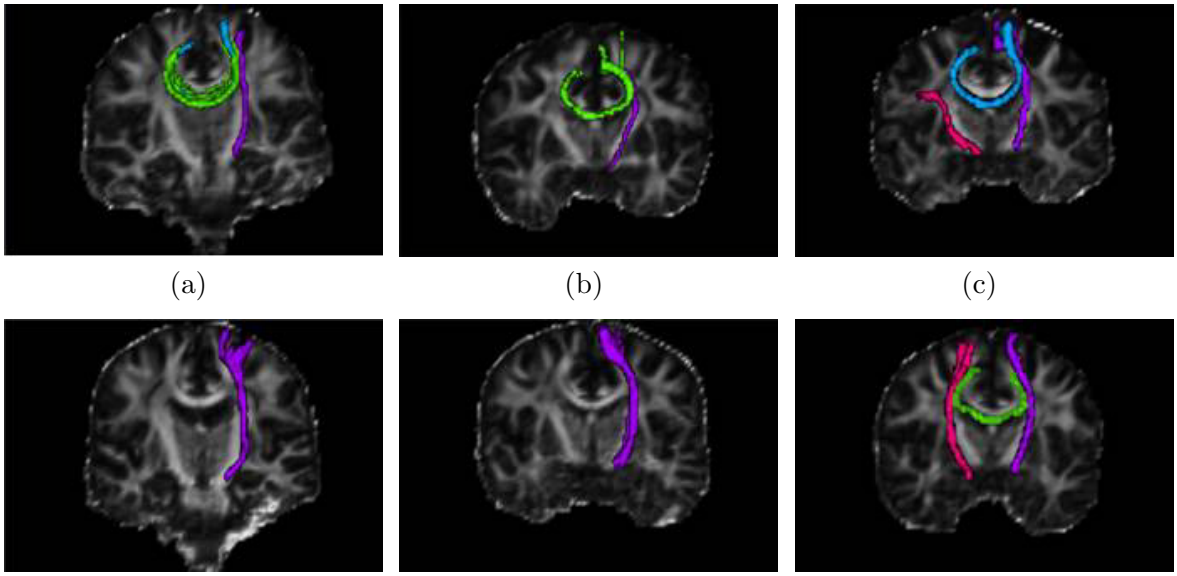


Figure 4.6: Tractography results for 15 gradient direction data: blue=contra-lateral motor tracts, green=contra-lateral somatosensory tracts; purple=motor cortex to substantia nigra (left); pink=motor cortex to substantia nigra (right). The reproducibility of tracts across subjects is variable, with the 1 fibre model specifically struggling to reproduce the tract between the right motor cortex to substantia nigra, and contra-lateral motor tracts.

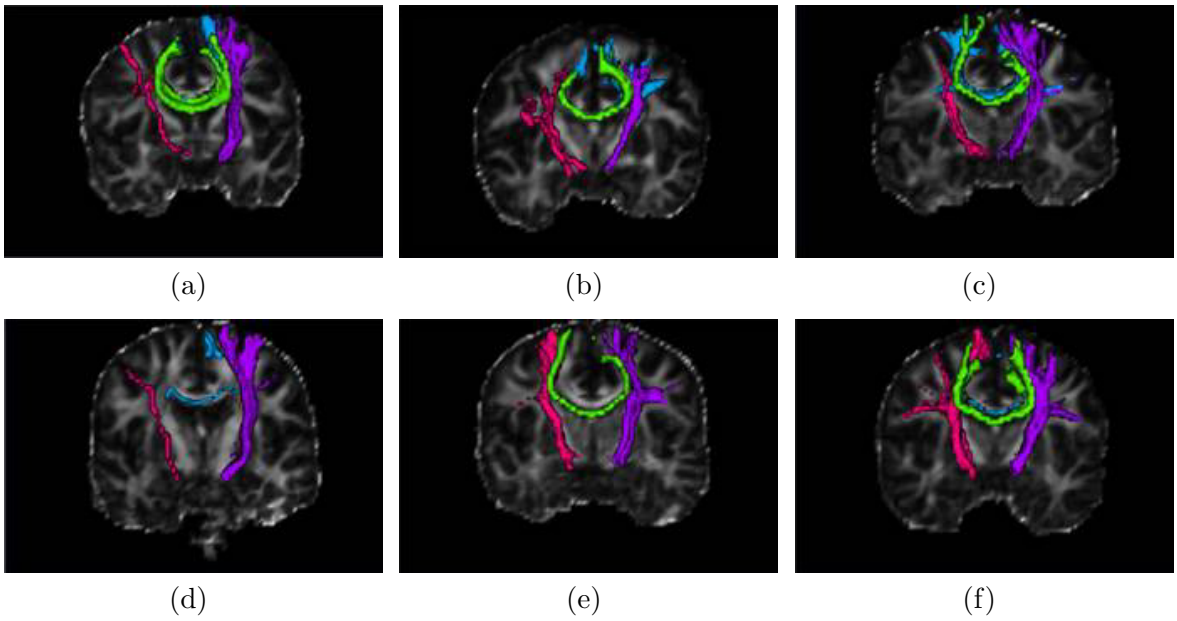


Figure 4.7: Tractography results for 60 gradient direction data (curvature threshold 0.2): blue=contra-lateral motor tracts, green=contra-lateral somatosensory tracts; purple=motor cortex to substantia nigra (left); pink=motor cortex to substantia nigra (right). In general, all tracts are propagated successfully using the 2 fibre model. Instances where tracts are not successfully propagated may reflect inaccuracies in the underlying segmentation. This is emphasised by restrictions placed by the model, which terminate tracts at the first target region they reach.

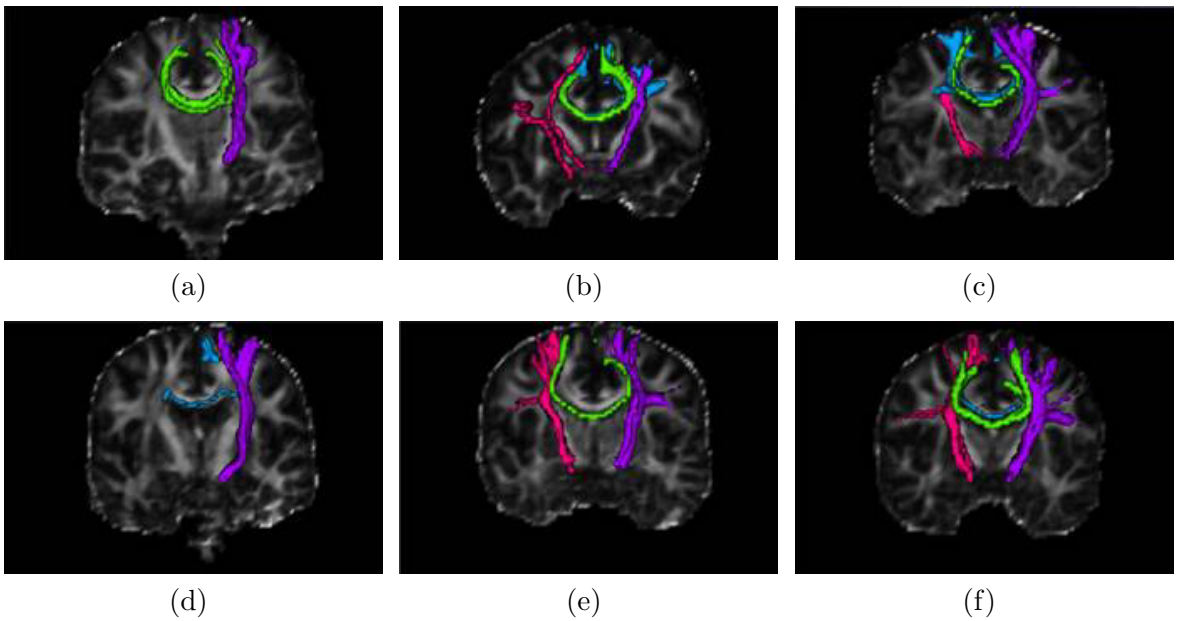
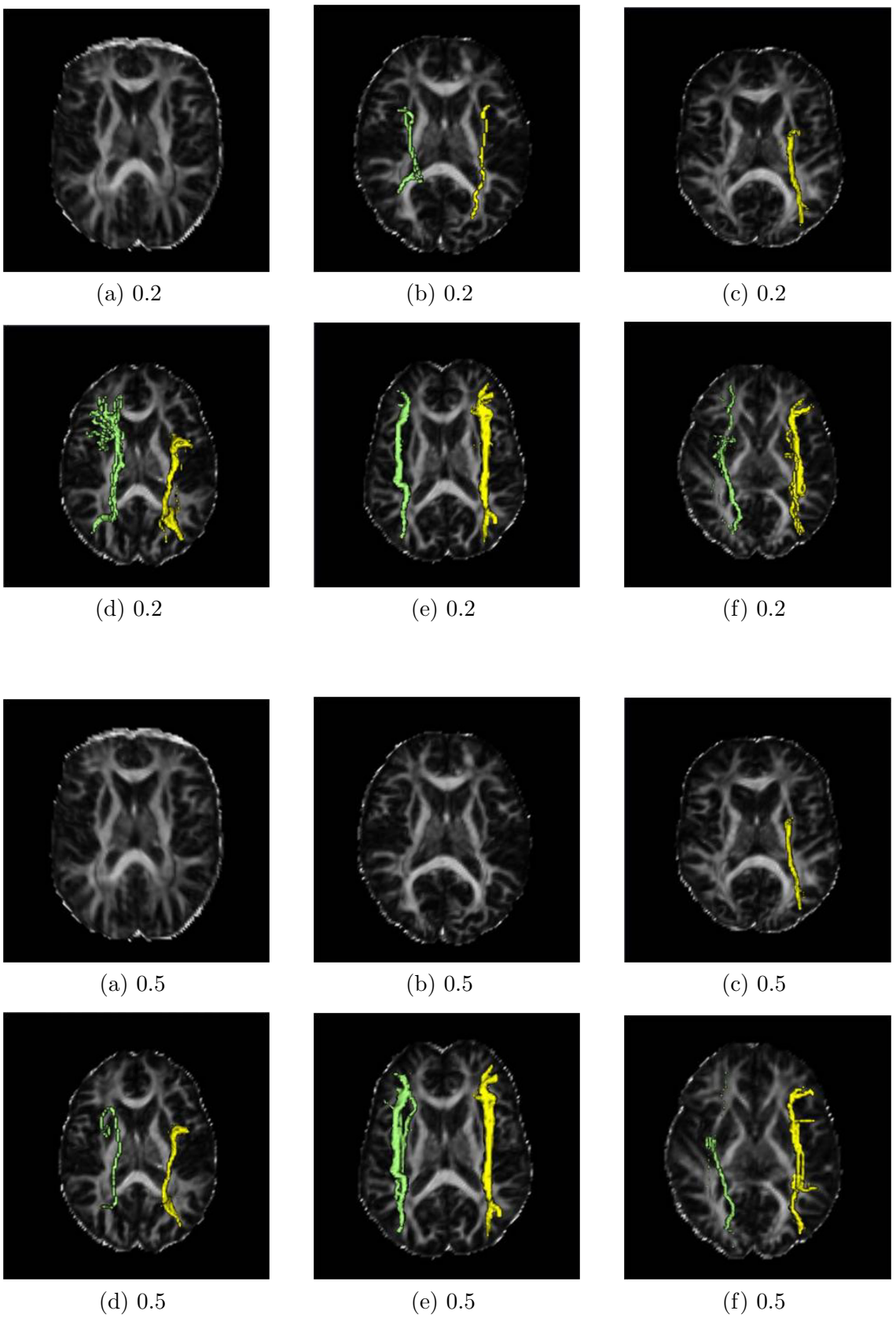


Figure 4.8: Tractography results for 60 gradient direction data (curvature threshold 0.5): blue=contra-lateral motor tracts, green=contra-lateral somatosensory tracts; purple=motor cortex to substantia nigra (left); pink=motor cortex to substantia nigra (right). Reducing the curvature threshold has minimal impact on the success rate for propagating these tracts, with the exception of a) and d) where the right motor cortex to substantia nigra tract is no longer propagated.



(a) 0.2 (b) 0.2 (c) 0.2

(d) 0.2 (e) 0.2 (f) 0.2

(a) 0.5 (b) 0.5 (c) 0.5

(d) 0.5 (e) 0.5 (f) 0.5

Figure 4.9: Tractography results for 60 gradient direction data: medial frontal gyrus to occipital lobe (top a-f curvature threshold=0.2, bottom a-f curvature threshold=0.5). Tracts between anterior and posterior regions are not propagated successfully for any of the datasets using the 1 fibre model due to sensitivity to tracts propagating laterally. By contrast, the 2-fibre model allows these tracts to be recovered in the majority of datasets. This is true for both curvature threshold settings.

Chapter 5

Graph Theoretic Analysis of Whole-Brain Networks

5.1 Introduction

Graphs are mathematical structures represented by sets of nodes and connecting edges. In this sense they lend themselves well to analysis of brain networks, where nodes in the graph may correspond to either functional or anatomical regions, and connections (where they occur) may be expressed by an edge.

At present graph theory is by far the most popular technique used for analysing models of whole-brain connectivity. This is because graph theory allows simple categorisation of the organisational structure of brains in terms of their small-world properties [190]. This terminology, described in section 5.3, has been used in many experiments to study brain evolution and disease. However, in general, studies have concentrated on studying properties of binary graphs, ignoring much of the rich variety of possible neuronal structures and corresponding connection strengths. This chapter therefore assesses graph properties for weighted graphs calculated using the framework proposed in the previous chapter.

Graph theoretic approaches also allow brains to be partitioned into modules or clusters, where clusters represent densely connected sub-sets of the graph. This has allowed whole-brain connectivity studies to partition binary brain networks into recognisable functional sub-units such as the frontal, occipital, parietal and temporal lobes [65, 77, 107, 148]. This chapter proposes comparing the clustering structure of weighted graphs for different populations. We perform spectral clustering of graphs averaged over

two populations: one aged 20-30 years and a second aged 60-90 years. The results reflect expected changes in the neurological micro-structure during healthy ageing.

5.2 Nomenclature

Graphs ($G = [V, E, W]$) are mathematical structures consisting of a set of vertices or nodes (V) and connections between nodes provided by edges (E). Edges may be assigned weights (W) and can be considered to be directed or undirected depending on whether the direction of the connection or relationship between the regions is known. As diffusion imaging techniques do not allow for distinguishing between afferent and efferent connections the brain network models described in this chapter are undirected.

The terms used in this chapter are:

Adjacency matrix: The adjacency matrix ($\mathbf{A}$) of a graph is a $|V| \times |V|$ matrix with entries $a_{ij} = 1$ if a node i connects to node j and $a_{ij} = 0$ elsewhere. The weighted adjacency matrix $\mathbf{W}$ has entries w_{ij} where weights relate to the 'strength' of connection.

In the case of neuroanatomy we should be interested in weighted graphs, since connections between regions of the brain are known to vary in physical size and strength. However, until recently, there was no consensus on how to quantify clustering for weighted graphs. Therefore, many brain connectivity studies [1, 43, 59, 67, 71, 148, 164, 185] have enforced a binary representation by forcing all existing connections to have a unit size.

Degree: The degree matrix $\mathbf{D}$ of a graph G is a diagonal matrix with entries $k_i = \sum_{j=1}^{|V|} a_{ij}$. For binary graphs this is equal to the number of edges connected to each node i. However, for weighted matrices the degree is calculated from $k_i = \sum_{j=1}^{|V|} w_{ij}$.

Clustering coefficient: The clustering coefficient (C_i) is defined for each node i by examining the connections of its k_i neighbours. Neighbours N_i of node v_i are all nodes with direct connections to it.

$$N_i = v_j : e_{ij} \in E \tag{5.1}$$

The maximum number of possible connections between neighbours in an undirected graph is $k_i(k_i - 1)/2$, as a_{ij} and a_{ji} are considered identical.

For non-weighted graphs the coefficient (C_i) is estimated from the proportion of actual connections between neighbours [190]:

$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)} \tag{5.2}$$

This can be equivalently calculated from the number of 3 sub-graphs formed from neighbours from the total possible.

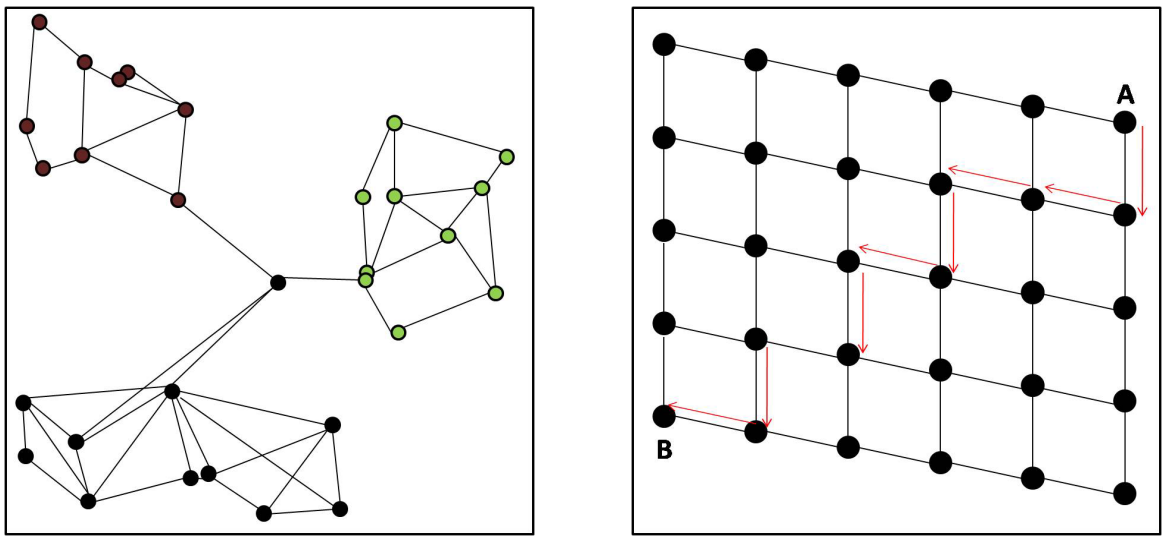
$$C_i = \frac{2}{k_i(k_i-1)} \sum_{j,k \in G; j,k \neq i} (a_{ij}.a_{jk}.a_{ki})^{\frac{1}{3}} \tag{5.3}$$

Therefore, for weighted graphs an equivalent metric has been proposed by Onnella et al. [125]. Here, C_i is calculated from the ratio of all 3 possible sub-graph weights to the total possible:

$$C_i = \frac{2}{k_i(k_i-1)} \sum_{j,k \in G; j,k \neq i} (\tilde{w_{ij}}.\tilde{w_{jk}}.\tilde{w_{ki}})^{\frac{1}{3}}, \tag{5.4}$$

Weights are normalised by the largest weight in the network ($\tilde{w_{ij}} = w_{ij}/max(\mathbf{W})$)

The clustering coefficient (C) for the whole graph is calculated by averaging ($C = 1/|V| \sum_i^{|V|} C_i$). An example of graph clustering see figure 5.1 a.



(a) (b)

Figure 5.1: Network properties: a) clustering ratios highlight graphs with cliques of highly connected nodes; b) the path length two nodes (in a binary graph) represents the smallest number of intermediate nodes between them. An example path is shown by red arrows

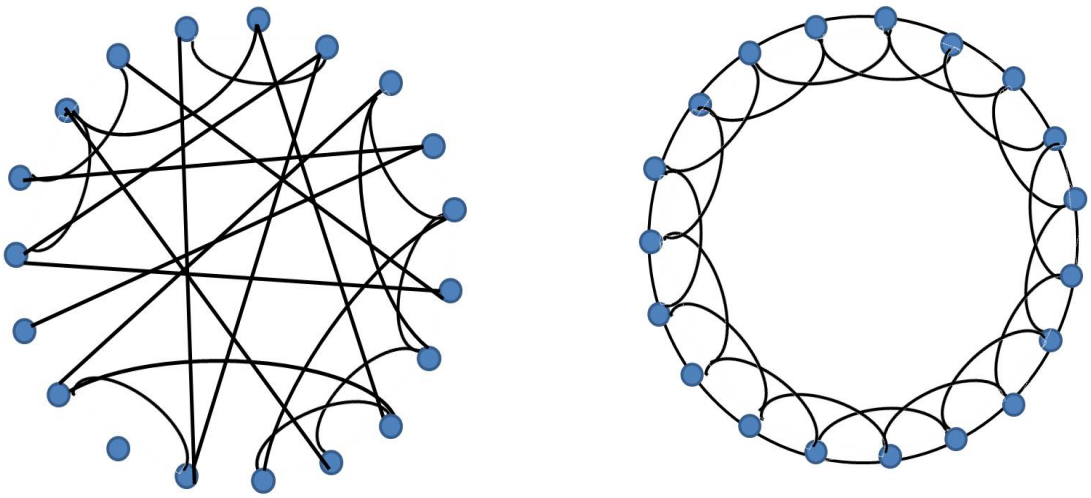
Characteristic path length: The mean (or characteristic) path length (figure 5.1b) is defined as:

$$L = \frac{1}{|V|(|V|-1)} \sum_{i,j \in G, i \neq j} d_{ij} \tag{5.5}$$

Where d is defined by the smallest number of edges needed to be traversed to connect node i with j. For weighted graphs strong connections are modelled as short distances by representing edge length in terms of the inverse of its weight, i.e. $d_{ij} = \frac{1}{w_{ij}}$.

Random graphs: Random graphs (figure 5.2 a) are characterised by relatively short path lengths and minimal clustering. Models of random graphs can be built by random assignment of edges for each node, using the Erdös-Rényi model [44], which suggests assigning edges to nodes with equal probability. Alternately, the approach of Maslov and Sneppen [101] proposes random rewiring of all the edges in the graph (by swapping pairs of edges) in such a way that preserves node density and in and out degree. This is relevant since total edge weight and degree number are known to strongly influence measures of network topology. Defining random graphs in this way enables the structure of real graphs to be compared with random equivalents

Regular lattices: Lattices (figure 5.2 b) are characterised by relatively high mean clustering, as each node is maximally connected to each of its 4 adjacent neighbours, and long path lengths, as there are no long range connections (therefore paths must pass through many intermediate nodes to cross from one side of the lattice to the other).



(a) (b)

Figure 5.2: Network diagrams: a) random network: edges are randomly assigned between nodes; b) regular lattice: each node is connected to it's four nearest neighbours

Small world graphs: Small world graphs form a sub-class of structures assumed to have similar path lengths to random graphs ($\lambda = \frac{L_{real}}{L_{rand}} \sim 1$) but higher clustering coefficients more typical of lattice networks ($\gamma = \frac{C_{real}}{C_{rand}} >> 1$). This can be summarised by a single ratio: $\sigma = \lambda/\gamma$ [81]. They have been found to characterise a wide range of complex networks ranging from power grids to human brain networks.

5.3 Small world graph analysis of brain connectivity

The small-world classification was first described by Watts and Strogatz [190]. They found that by rewiring only a few connections of a regular lattice network, they were able to model self-organising systems that are neither random nor completely regular. In particular, they demonstrated that simple redirection of a small number of connections creates 'short-cuts' between distant nodes. This significantly reduces mean path length but leaves the clustering coefficient relatively unaffected. The approach was shown to accurately characterise a variety of complex networks representing the power grid of the western United States, a graph of film actors and the brain network of the c-elgans worm.

Investigations of the small world properties of histologically recorded animal brain networks such as the cat and macaque monkey [77, 163] subsequently led to suggestions that small world properties are linked to complexity and information processing, and may be indicators of brain evolution. This was tested by dynamical models which showed that functional brain networks optimised for both complexity [161] and fractal composition [159] (hierarchical organisation) both demonstrate small world properties.

The success of the histological analysis encouraged small world analysis to be applied to the result of functional imaging studies [1, 43, 148, 164, 185]. These attempt to model patterns of neuronal activity in the form of time series signals at every voxel in the brain. Whole brain connectivity is assessed from functional data by comparing the similarity between time series via, for example, estimates of signal correlation. Typically networks are thresholded below a certain similarity level and connections are set to 0 or 1. The results obtained mirror patterns of anatomical connectivity, and many generate similar small world ratios to histological results obtained for the cat and macaque [1, 148]. For more details of functional connectivity see Appendix B.

More recently, the development of whole-brain models of structural connectivity has allowed small world properties to be assessed for anatomical brain networks. These have been generated via two types of technique: correlated measurement of cortical thickness [71], and tractography approaches [59, 67, 83] (described in more detail in section 4.2). Of these only Iturria-Medina et al. [158] assess small world graph properties for weighted graphs. They assign weights from estimates of anatomical connection probabilities (ACPs: section 2.6)

The results for several studies are summarised in table 5.1. This includes results for the complete brain network (every neuron) of the c-elgans worm: the only brain network to so far be mapped comprehensively, and histological assessments of the presence of inter-regional connections for the cat and macaque cortex. In addition, results are presented for several assessments of functional[1] and anatomical human brain connectivity, where these have generally been obtained over multiple subjects (N) and averaged. Functional studies which report connectivity between ROIs (as opposed to voxels) have averaged time series within each region before evaluating signal similarity.

Table 5.1: Small world graph properties of complex brain networks. N is the number of ROIs, γ is the clustering ratio, λ is the path length ratio and σ is the small world ratio.

Study	Species	Study type	subjects	N	γ	λ	σ
Watts & Strogatz (1998)	*c elgans*	histology	-	282	5.6	1.18	4.75
Hilgetag (2000)	cat	histology	-	55	1.77	1.06	1.67
Hilgetag (2000)	mac	histology	-	73	3.06	1.17	2.61
Sporns & Zwi (2004)	cat	histology	-	52	1.99	1.07	1.30
Sporns & Zwi (2004)	mac	histology	-	71	1.92	1.12	1.70
Achard (2006)	human	fMRI	5	90	2.37	1.09	2.18
Salvador (2005)	human	fMRI	12	90	2.08	1.09	1.91
Eguiluz (2004)	human	fMRI	7	voxel	168.54	1	168.54
van den Heuvel (2008)	human	fMRI	28	voxel	1.67	1.00	1.67
Stam (2004)	human	MEG	5	126	1.89	1.19	1.59
Hagmann (2006)	human	tractography	2	748	3.92	1.19	3.29
He (2007)	human	cortical thickness	124	54	2.36	1.15	2.04
Iturria-Medina (2008)	human	tractography	20	90	1.85	1.12	1.64
Gong (2008)	human	tractography	80	78	4.07	1.15	3.54

[1] obtained from either functional magnetic resonance (fMRI) or magneto-encephalography (MEG) measurements

All human brain connectivity studies report networks that are small world, with σ ratios predominantly within the range of 1.30-3.54. This is with the exception of the functional study by Eguiluz et al. [43], which has a much higher clustering ratio. This may be partially due to voxel-wise assessment of connectivity. Nevertheless, the same increase is not reported by a similar approach by Van den Heuvel et al. [185]. All results are similar to the histological analysis of animal brains but do not show clear evidence of an evolutionary trend. Unfortunately, direct comparison is complicated by the widely different approaches and scales of analysis used.

5.4 Small world graphs as biomarkers for disease

The relationship between small world properties, and the organisation and efficiency of the brain has led some studies to investigate whether small world graph properties may provide biomarkers for disease. In particular, several different studies have compared path lengths and clustering ratios of healthy subjects and patients with Alzheimer's disease (AD). These include: Supekar et al [173], who report results for an fMRI study which suggests that path length ratios do not vary significantly, whereas clustering ratios are increased. This allows them to classify AD patients versus controls with 71.4% accuracy. In contrast, He et al. [72] report results from networks derived from correlated measurements of cortical thickness, which suggest that the graphs of AD patients display significantly increased clustering and increased path lengths ratios. Also, Stam et. al. [166, 165] perform electro-encephalography (EEG) analysis of functional connectivity for weighted and non-weighted graphs. They report significant changes to path lengths only for un-weighted graphs [166], but show weighting graph edges by EEG signal similarity increases the sensitivity of the approach to also detect changes in clustering.

Small world analysis has also been applied to study schizophrenia, a disease that manifests in terms of disrupted thought processes, and is thought to be caused by disconnectivity of either spatial or temporal cortical co-ordination [54]. Rubinov et al. [142] report decreased path lengths and increased clustering for patients with schizophrenia versus controls. In contrast, Micheloyannis et al. [109] report increased clustering, but also increased path length. These results differ but do support the suggestion that schizophrenia may be consistent with a 'subtle randomisation' in the underlying architecture [142].

5.5 Graph clustering

As small world analysis has shown, clustering or community structure of networks is fundamental for the development of complexity and plays a key role in efficient network dynamics. Therefore a growing interest has developed in studying community structure in brain networks in more detail. To this end, some studies have attempted to 'grow' clusters from networks using, for example, optimal set analysis (OSA), which uses modified simulated annealing [77], or agglomerative hierarchical clustering, which builds a hierarchy by serially merging similar clusters [148]. More recently, interest has centred on two graph partitioning algorithms: modularity and spectral clustering. These are described below.

5.5.1 Modularity

Modularity [119] is a measure of how much the clustering structure of a network varies from what would be expected if it were random. It is equivalent to the difference between the number of edges within clusters minus the expected number of edges in an equivalent random network. This is defined as $k_i k_j / 2m$, where k_i and k_j are the degrees for vertices i and j, and m is the total number of edges in the network ($m = \frac{1}{2}\sum_i k_i$). From this the modularity matrix ($\mathbf{B}$) may be calculated as:

$$B_{ij} = a_{ij} - \frac{k_i k_j}{2m} \tag{5.6}$$

Alternatively, the adjacency a_{ij}, may be replaced by weights w_{ij} for a weighted graph approach.

The original modularity algorithms propose agglomerative approaches [118]. These grow modules from an initially empty network which builds, one edge at a time, by adding the vertex pairing which most increases the modularity. Recent methods [119] propose hierarchical partitioning. These use the eigenvector, $\mathbf{e}$, of $\mathbf{B}$ corresponding to its largest positive eigenvalue, and separate nodes into two groups according to whether the corresponding element of $\mathbf{e}$ is positive or negative. This process is then repeated on the resulting modules.

The specific advantage of modularity approaches is that they automatically provide a metric which allows the optimal number of clusters (or modules) to be found. In agglomerative approaches, this means that the algorithm converges to a solution with maximum modularity. In contrast, in the partitioning approach, when maximum modularity is reached, all nodes will be assigned to the same cluster.

5.5.2 Spectral clustering

Spectral clustering uses the properties of the graph Laplacian ($\mathbf{L} = \mathbf{D} - \mathbf{W}$) formed from the degree matrix $\mathbf{D}$ and weighted adjacency matrix $\mathbf{W}$ of the graph ($G = (V, E)$). It has been shown by Shi and Malik [154] that the second smallest eigenvector $\mathbf{e}$ of the normalised Laplacian $\mathbf{L_{norm}} = \mathbf{D^{-1}L}$ provides an analytical solution for the normalised graph cut problem for two clusters A and B, $A \bigcup B = V$:

$$\mathrm{Ncut}(A, B) = \frac{\mathrm{cut}(A, B)}{\mathrm{Vol}(A, V)} + \frac{\mathrm{cut}(A, B)}{\mathrm{Vol}(B, V)} \tag{5.7}$$

Here, $\mathrm{Vol}(A, V) = \sum_{i \in A, j \in V} w_{ij}$ is the total weight of nodes in the cluster to all other nodes in the graph, and the cut:

$$\mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \tag{5.8}$$

aims to find a partition of the graph which minimises the total edge weight shared between clusters.

The un-normalised cut approach can lead to solutions which isolate a single node. This is clearly undesirable. Therefore, normalisation balances the requirement that the total edge weights between clusters is minimised by ensuring as far as possible that clusters are of an equal size. Examples of normalised and un-normalised cuts are shown in figure 5.3.

Shi and Malik [154] showed that by relaxing any requirement for a discrete solution[2], the problem of minimising the normalised cut for two clusters is equivalent to finding the second eigenvector of the following generalised eigensystem problem:

$$(\mathbf{D} - \mathbf{W})\mathbf{e} = \lambda \mathbf{D}\mathbf{e} \tag{5.9}$$

$$\mathbf{L_{norm}}\mathbf{e} = \lambda \mathbf{e} \tag{5.10}$$

Therefore, the problem of minimising the normalised cut for two clusters is estimated by finding an optimal breaking point along $\mathbf{e}$ which minimises the value of Ncut.

The approach can be extended, such that the third eigenvector partitions the first two clusters and further additional eigenvectors will continue sub-partitioning. However, while this approach is adopted by some [187, 76], Shi and Malik [154] suggest that it will lead to a cumulation of errors on account of the relaxation of the requirement for discrete separation. Instead, they recommend a hierarchical

[2] where A and B can be completely separated

approach, where each time a new normalised graph Laplacian is estimated for each sub-partition. The stability of the cut is estimated by first permuting the splitting point to induce two different partitions $P_1(A_1, B_1)$ and $P_2(A_2, B_2)$, and calculating the ratio:

$$\frac{\delta \text{cut}(P_1, P_2)}{\delta D(P_1, P_2)} \tag{5.11}$$

where $\delta D(P_1, P_2) = \sum_{i \in (A_1/A_2)} k_i$.

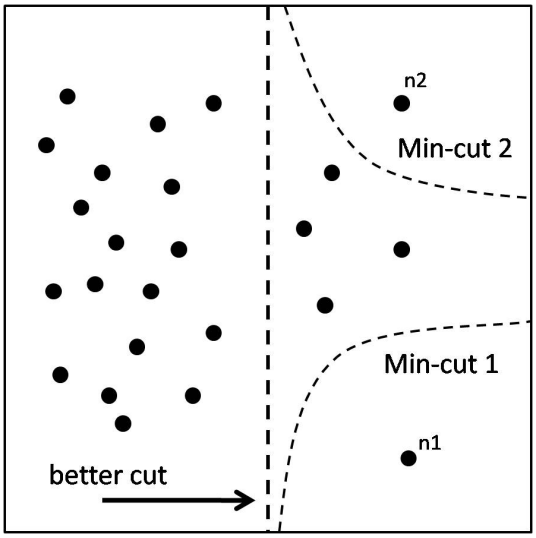


Figure 5.3: Ncut: the un-normalised min cut favours cuts which isolate single nodes of the graph. In contrast, the normalisation term ensures that if sub-sets A or B were to be small, then the cut would almost certainly be a large fraction of the total edge weight. This reduces bias towards single node clusters

5.6 Community structure in whole brain networks

Graph clustering of brain networks produces regions with recognised functional significance. This was first shown via hierarchical clustering models of the cat and macaque cortex, but has since been replicated in functional networks obtained from resting state fMRI data [148, 107]. These favour bilateral clusters (which encompass brain regions from both hemispheres) and commonly include recognised functional sub-regions of the brain such as: frontal, temporal, occipital and parietal-(pre) motor units. Similar bilateral patterns have also been observed from modular analysis of structural connectivity networks extracted from cortical thickness measurements [32]. In contrast, Hagmann et al [65] perform modular clustering on anatomical networks built from tractography, and report predominantly single hemisphere modules. This may reflect the fact that anatomical connections in the brain are predominantly short range.

Some studies have also used clustering properties to compare averaged brain graphs across populations. In particular, Meunier et al. [107] use the modular structure of functional networks to study ageing. Here, resting state functional brain networks (see appendix B) are estimated for two populations (17 subjects aged 18-33 years and 13 subjects aged 62-67 years) using wavelet correlation analysis. These are extracted over a range of thresholds and used to form binarised mean brain networks for each population. Then modules are identified in each mean network using an agglomerative approach, and the component structure of modules, as well as the role of nodes within modules are used to compare average networks for both populations. The results show that the network for the young population splits into 5 bilateral modules: a central, fronto-cingulo-parietal, posterior, ventral frontal and medial temporal module, whereas the older population forms additional modules through segregation of the large frontal-cingulo-parietal module into more local frontal and medial parietal modules. In addition the amount of inter-modular connectivity of frontal brain regions is shown to change. These changes are linked to pre-frontal dysfunction

As yet few studies have applied spectral clustering approaches to brain networks. This is primarily because most studies have represented brains using binary graphs, where edge weight does not provide much contrast[3]. However, Behrens et al. [21] use a similar technique, spectral re-ordering, to perform clustering of results obtained from the cross-correlation of voxels' connectivity profiles, obtained via probabilistic tractography. This is used to successfully cluster voxels in the medial frontal cortex.

Recently, Shen et al. [151] reported clustering results for *weighted* functional networks which demonstrate that Ncut generates more consistent results than modularity. These results were most prominent for data whose underlying cluster distributions were very uneven in size. We propose using spectral clustering to compare weighted anatomical brain regions from two different age groups. The connection weights within our proposed framework are designed to reflect the structural integrity of the underlying tissue. It is our hypothesis that changes in brain structure correlated with ageing may be identified through the clustering approach. This is tested in the next section.

[3]In this instance modularity methods offer specific advantages as they also compare the distribution of edges

5.7 Studying the effects of age using small world properties and spectral clustering

In this section spectral clustering is used to extract the community structure of weighted brain networks extracted using the framework proposed in the previous section. Results are compared for networks extracted from two age populations. In extension to previous work (see Publications, [2]) the findings are correlated with the small world graph properties of each population.

5.7.1 Subjects

Analysis was performed on 96 brains (42 subjects aged between 20-30 years and 54 subjects between 60-90 years). Scanning was performed on a Philips 3 Tesla system. T1 and T2 images were acquired prior to diffusion weighted imaging using 3D MPRAGE and dual echo weighted imaging and single shot echo planar DTI was acquired in 15 non collinear directions using the following parameters: TR 12000ms, TE 51ms, slice thickness 2mm, voxel size 1.75 x 1.75 x 2mm^3, b-value 1000s/mm^2. Data was acquired from the IXI database. Ethical approval was granted by the Thames Valley MREC.

Brain networks were extracted from the diffusion data according to the protocol laid out in Chapter 4. As the data was acquired in 15 gradient directions a single fibre model was fit at each voxel. Tractography was performed by seeding 1000 streamlines per voxel, and results were summarised in 77x77 matrices. Small world and spectral clustering analysis was performed on connectivity data quantified using mean tract anisotropy. The results are described in the following sections.

5.7.2 Small world graph analysis

Small world graph analysis was applied using the approach detailed for weighted graphs in Iturria-Medina et al [158]. Therefore, weighted clustering ratios and path length are calculated according to the equations listed above. In accordance with other weighted graph comparison studies, random graphs for each network were generated by the rewiring of edge weights [101]. This ensures that the edge weight distribution within the graph is maintained. Therefore, any detected differences between populations relates to differences in network organisation rather than overall changes in connection weights.

Table 5.2: Use of small world graph properties as an indicator of healthy ageing. N is the number of ROIS, γ is the clustering ratio, λ is the path length ratio and σ is the small world ratio

Population	Study type	subjects	N	γ	λ	σ
20-30years	tractography	42	77	1.62	1.19	1.91
60-90years	tractography	54	77	1.48	1.22	1.74

Results averaged over each population are shown in table 5.2. All brain networks are shown to be small world. In addition permutation testing of the values for each subject show that reductions in clustering, increases in path lengths and decreases in small world ratios with age are all significant ($p < 0.01$). These results mirror studies that suggest that the small world properties of functional brain networks change during childhood [110] and they reflect general changes in brain organisation during healthy ageing.

5.7.3 Hierarchical spectral clustering analysis

Weighted brain networks were averaged over each of the two age populations and hierarchical clustering analysis was performed according to the framework proposed by Shi and Malik [154]. Normalised cuts were made over four cycles of recursive partitioning where in each case the clusters were sub-partitioned into two. Clustering was stable up to and including the third cycle of recursive clustering. After this point only some clusters supported further partitioning. The threshold was set by comparing the stability of a permutation of graph labels to the stability of a random permutation using equation 5.11, and a user defined cut-off value.

Dendrograms displaying the results of the hierarchical clustering are shown in figure 5.4. Each of the components of the 8 main clusters (from the third cycle of clustering) are shown in figure 5.5 projected (in the same order) into a normalized anatomical space which corresponds to the MNI152 atlas. The results are similar to previous studies [32, 107, 108, 148] inasmuch as many of the clusters form well-known functional sub-units. In particular the limbic-striatal and parietal-pre-motor cluster (left=red, right=yellow) bears significant resemblance in composition to the frontal-cingulo-parietal module in the Meunier study [107]. Both populations also display posterior temporal (blue=left, green=right) and orbito-frontal (light green=left, light blue=right) clusters in each hemisphere. In addition the medial temporal (left=orange, right=lilac) cluster appears for both populations. However, whereas in the 20-30 years population it appears in the 3rd cycle of clustering, for the 60-90 year population it

is not seen until the 4th cycle. This is shown at the bottom of the posterior temporal partitions in dendrogram for the 60-90 year hierarchy (5.4).

Whilst there are many similarities between the two populations, there are also significant differences. Firstly, whilst the initial partition of the network for the younger population splits the brain into left and right components, this is not the case for the older population, where the first cut splits the brain into anterior and posterior regions. This is still apparent after four cycles of clustering, where the network of the older population retains a left-right frontal cluster (purple, in both figures) despite the rest of the network sub-partitioning into single hemispheric clusters. The significance of this difference was confirmed ($p = 0.0079$) by performing a t-test of the value of Ncut between this cluster and the rest of the brain for each individual subject. It is likely that this change in organisational structure is the origin of the observed changes in small world ratios as the weakening of these long range connections will inevitably lengthen average path lengths and may also weaken the cluster coefficients for hubs in these areas.

The presence of an isolated frontal cluster in the older population reflects current understanding of the mechanisms underpinning healthy ageing: that frontal white matter [60] reduces with age, and that the efficiency of long distance connections between frontal regions to the rest of the brain may also be reduced [172]. This is suspected to lead to a disconnection between anterior and posterior regions of the brain, which is further linked to age-related changes in performance [45]. It is also reflected in Meunier's [107] study of changes to frontal and parietal modules in the brain. However, in this study the frontal cluster splits off in the first cut rather than being seen as a direct partition of the frontal and parietal modules formed from the younger population's graph.

Other differences include an additional inferior frontal cluster (pink) in the elderly population which does not appear for the younger population. Figure 5.5 clearly show this to be a split from the left orbital-frontal region (shown in light green: cluster 8 for both populations). Many papers have cited general age-related changes in FA values [97] at variable rates throughout the brain. Such regional degradation may contribute to the increased lack of organisation and therefore reflect the overall loss of symmetry that we have reported.

There are still some unexpected results. For example, we would have expected the limbic-striatal and parietal pre-motor regions to further sub-partition. However, it is likely that with higher quality diffusion data (obtained at higher resolution) further subdivisions may be found. Additionally, the study

differs from functional approaches in that it suggests significant partitioning between hemispheres into predominantly left and right brain regions. In general, similar clusters are mirrored on both sides and most clusters contain regions which are adjacent to each other. This may reflect in part the difference between modularity and clustering approaches as bilateral connections may feature along several shortest paths, but are likely to be few in number and therefore not favoured by normalised cut approaches. However, the results are not dissimilar to the Hagmann study [65] which reports modularity results for streamline tractography in DSI data, and reflects the fact that the brain is predominately made of short range connections [160], and neighbouring regions have similar functions. Nevertheless it would be interesting to see whether the patterns change for 60 direction diffusion data, which has been shown to be more successful at tracing bilaterally.

5.8 Summary

This chapter has presented the first graph theory-based study of the effects of ageing on weighted anatomical brain networks. Networks are extracted from 15 gradient direction diffusion data via probabilistic tractography, and strengths of connection are estimated from measurements of the mean anisotropy along tracts. We have applied small-world graph analysis and spectral clustering. The results indicate that ageing changes the organisational structure of the brain.

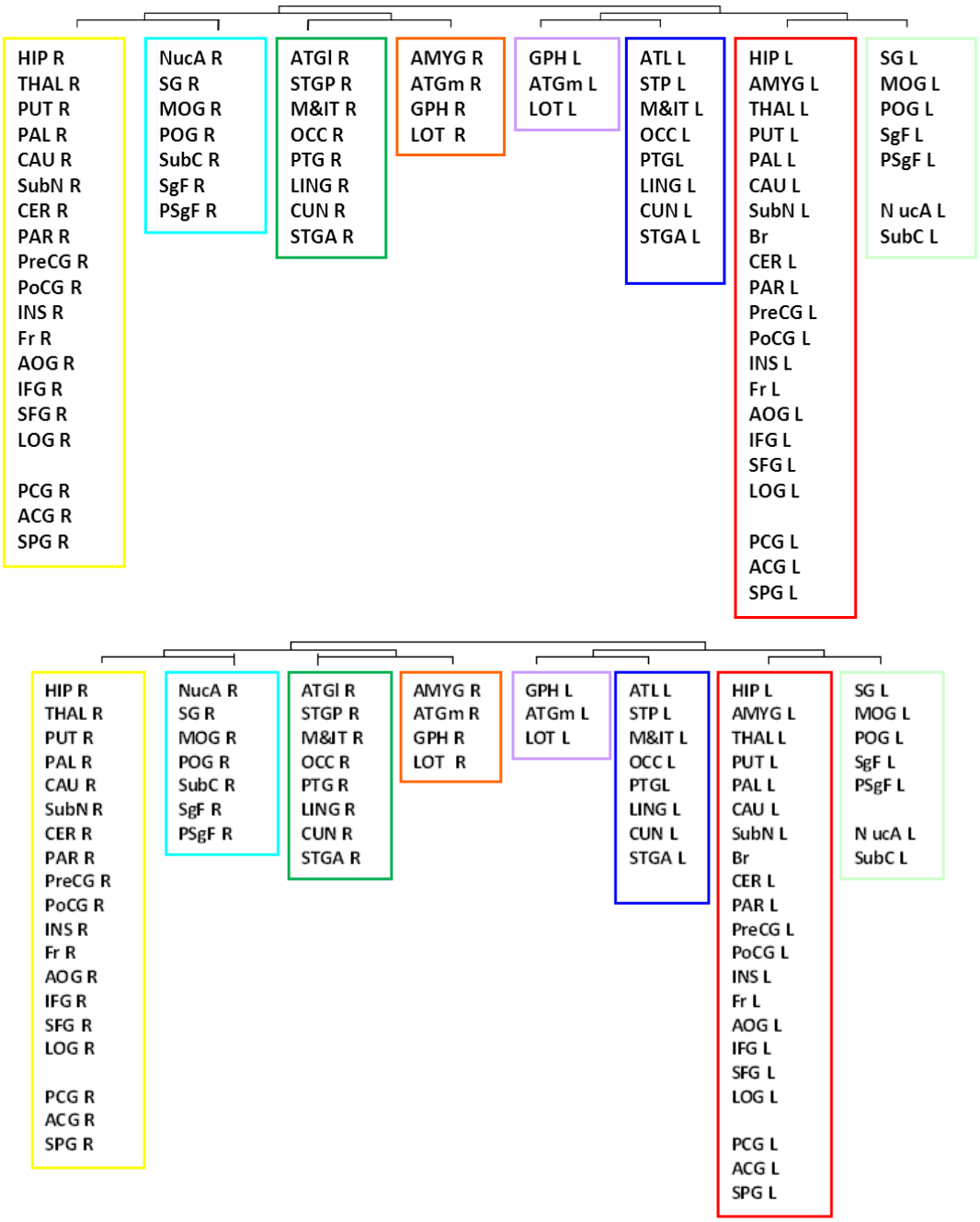
Many studies of whole brain functional and structural connectivity have used small world graph properties to describe their networks. These are characterised by short path lengths and increased clustering where this is thought to support increased complexity and efficiency. Some studies have also investigated the potential for using small world properties as biomarkers for disease. In particular, several studies have analysed brain networks derived from patients with Alzheimer's [165, 166] or schizophrenia [109, 142]. These have demonstrated decreased small world ratios when compared to control populations.

Unfortunately, while small world ratios can be used to indicate the presence of a disease, they are not able locate the positions of any organisational changes in the microstructure. Therefore, some approaches have instead investigated the community structure of networks using techniques such as hierarchical clustering [77, 148] or modularity [32, 65, 107]. These have shown that the brain can be clustered into recognised anatomical subunits, and have been used to reflect functional changes associated with ageing [107] and schizophrenia [18]. However, to date many of these approaches have

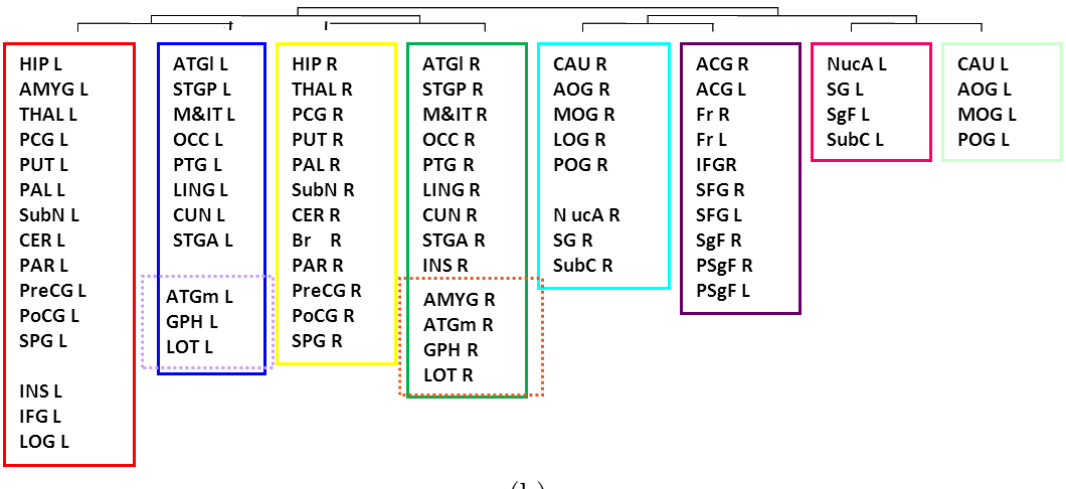
been limited to analysis of binary brain networks. This has ignored much of the rich variability in the underlying tissue micro-structure. In contrast, the proposed framework builds weighted connectivity graphs from measurements of tissue anisotropy.

The results of this chapter have demonstrated significant differences in the small world structure of weighted anatomical brain networks, extracted for two age populations. These have been correlated to specific changes in brain organisation through a cluster analysis approach. The results reflect the literature which suggests that healthy ageing is accompanied by a disconnection between anterior and posterior regions of the brain [172], and reflect results presented by a modularity study on a similar population [107]. However, a key advantage of this approach is that (unlike unweighted graph approaches) differences in clustering across populations can be directly linked to changes in the structural integrity of underlying anatomical connections. This is due to the inclusion of estimates of tissue anisotropy to weight graph edges.

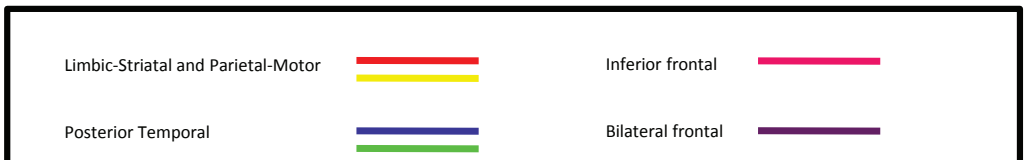
In conclusion, the results show that applying spectral graph clustering to weighted graphs can relate organisational changes in brain structure across populations to changes in the underlying microstructure, and can link changes in small world properties to the anatomy. It may be interesting to apply the technique to the other populations singled out as having disrupted organisation, for example Alzheimer's and schizophrenic populations. We believe, this combined approach may increase the effectiveness of graph metrics as biomarkers of disease.

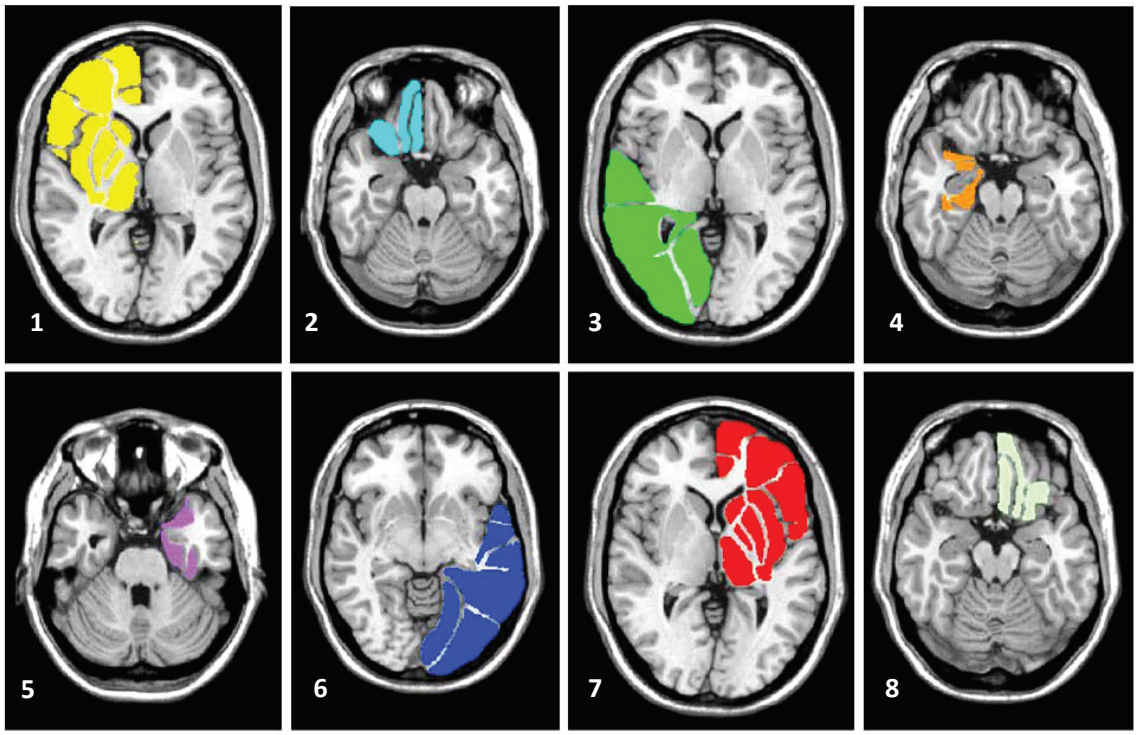

HIP R
THAL R
PUT R
PAL R
CAU R
SubN R
CER R
PAR R
PreCG R
PoCG R
INS R
Fr R
AOG R
IFG R
SFG R
LOG R
PCG R
ACG R
SPG R
NucA R
SG R
MOG R
POG R
SubC R
SgF R
PSgF R
ATGl R
STGP R
M&IT R
OCC R
PTG R
LING R
CUN R
STGA R
AMYG R
ATGm R
GPH R
LOT R
GPH L
ATGm L
LOT L
ATL L
STP L
M&IT L
OCC L
PTGL
LING L
CUN L
STGA L
HIP L
AMYG L
THAL L
PUT L
PAL L
CAU L
SubN L
Br
CER L
PAR L
PreCG L
PoCG L
INS L
Fr L
AOG L
IFG L
SFG L
LOG L
PCG L
ACG L
SPG L
SG L
MOG L
POG L
SgF L
PSgF L
N ucA L
SubC L

(a)


HIP L
AMYG L
THAL L
PCG L
PUT L
PAL L
SubN L
CER L
PAR L
PreCG L
PoCG L
SPG L
INS L
IFG L
LOG L
ATGl L
STGP L
M&IT L
OCC L
PTG L
LING L
CUN L
STGA L
ATGm L
GPH L
LOT L
HIP R
THAL R
PCG R
PUT R
PAL R
SubN R
CER R
Br R
PAR R
PreCG R
PoCG R
SPG R
ATGl R
STGP R
M&IT R
OCC R
PTG R
LING R
CUN R
STGA R
INS R
AMYG R
ATGm R
GPH R
LOT R
CAU R
AOG R
MOG R
LOG R
POG R
N ucA R
SG R
SubC R
ACG R
ACG L
Fr R
Fr L
IFGR
SFG R
SFG L
SgF R
PSgF R
PSgF L
NucA L
SG L
SgF L
SubC L
CAU L
AOG L
MOG L
POG L

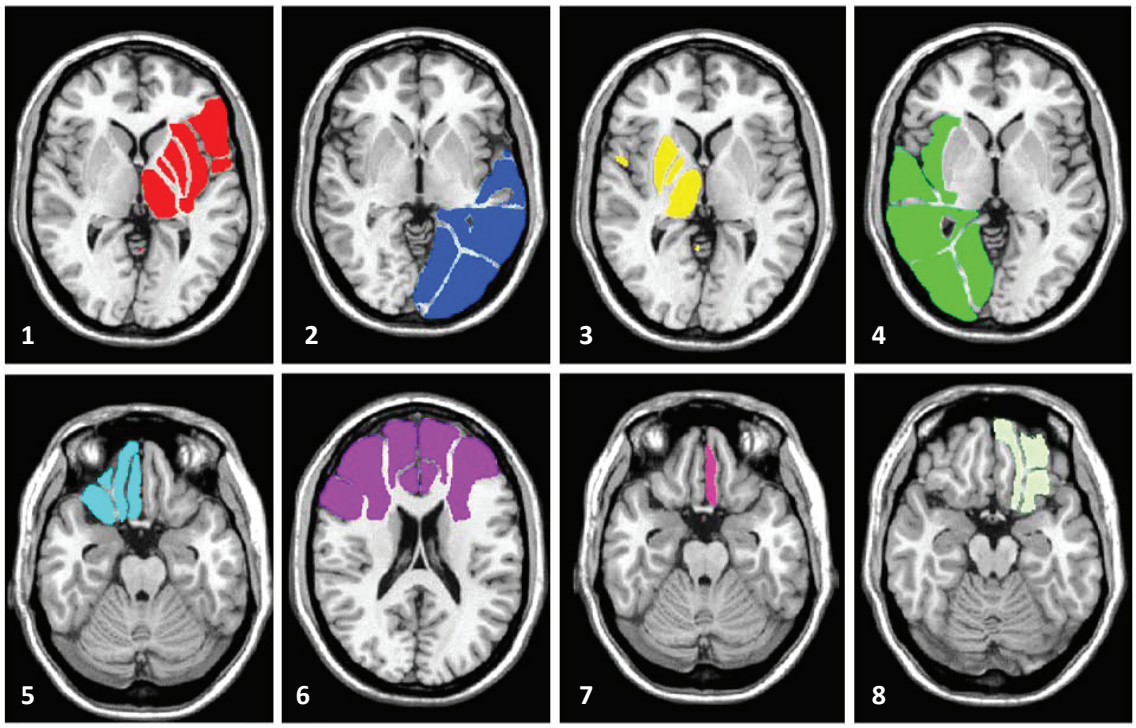
(b)


Limbic-Striatal and Parietal-Motor
Inferior frontal
Posterior Temporal
Bilateral frontal



20-30yrs

(a)



60-90yrs

(b)

Figure 5.5: Projection of the cluster hierarchy on a T1 axial slice, for a) the 20-30 years b) 60-90 years populations

Chapter 6

Multivariate Statistical Analysis of Structural Brain Networks

6.1 Introduction

Graph theoretic approaches are well suited to studying general properties of brain organisation and have shown some potential as biomarkers for disease. However, they are unable to specify the locations of any differences in the underlying microstructure. In contrast, statistical approaches allow direct comparison of features of interest across large sets of data. This has allowed neuroimaging studies to make voxel-wise comparisons of properties such as anisotropy or deformation across large populations. We therefore adapt the statistical approach for analysis of whole-brain connectivity matrices. Our aim is to associate differences in white matter tissue microstructure with the functional grey matter regions to which they are connected to. Two machine learning approaches are proposed: 1) a principal component (PCA), maximum uncertainty linear discriminant analysis (MLDA) approach [177, 176]; and 2) a non-linear support vector machine (SVM) approach with GentleBoost feature extraction. Both approaches classify subjects and extract features which represent the key differences between groups.

The next section presents an overview of current techniques for identifying statistical differences in the brain's neuronal white-matter. After this, two different machine learning approaches are presented. Both are applied to a population of 90 subjects separated into two age groups (20-30 years and 60-90 years) to test the approach on an already well understood phenomena, normal ageing.

6.2 Background

One commonly used approach for cross-subject statistical comparison of properties throughout the brain is voxel based morphometry (VBM). In the past, this has primarily been used for comparing T1 images for changes in white/grey-matter and CSF [8]. However it is now increasingly being used for analysis of diffusion data [129, 73, 193].

VBM works by first transforming all subjects' images to a common template, and smoothing them to reduce the impact of misalignment and partial volume. Statistical comparisons of tissue intensity are then made at equivalent voxels throughout the brain using univariate hypothesis testing. This investigates whether the observed differences between populations occur with significantly greater regularity than would be expected by random chance.

One key problem with VBM approaches is that misalignment of structures can lead to suggestions of tissue volume changes where there are none, and smoothing can disproportionately exaggerate the presence of any existing differences. In addition, VBM of diffusion data is traditionally performed via alignment of structural T1 images which may warp white matter structures too far. Therefore alternate approaches have been proposed based on registration of diffusion data using alignment of FA images themselves [73], or multichannel registration using all the coefficients of the diffusion tensor [129]. In addition, Whitcher et al [193] propose a multivariate hypothesis testing approach, where all of the coefficients of the diffusion tensor are compared simultaneously. This allows detection of pathologies which also affect diffusion orientation.

A similar approach to VBM is tensor based morphometry (TBM). This compares the Jacobian determinants of the transformations required to map subjects to the co-ordinate frame of the template. It has recently been adapted for diffusion data analysis by Zhang et al. [196] who use re-alignment of diffusion tensors to decompose the deformation into components which describe volume changes in parallel to the principal eigenvectors of the tensors, and perpendicular to them. This overcomes in part some of the problems of VBM, in particular, partial volume, which leads to difficulty in interpreting whether effects are due to changes in tissue integrity or tract volume.

Other studies only investigate voxels within sub-structures or regions of interest (ROIs), such as tracts. This overcomes some of the limitations of the aforementioned morphometric approaches by alleviating the requirement for perfect alignment of the whole brain. Nevertheless it is necessary to generate voxel-wise correspondences along the lengths of tracts. Approaches have been proposed by Courage et

al. [34] who transform propagated streamlines into a common co-ordinate frame and perform tensor interpolation to establish corresponding points on each tract. In addition, Goodlett et al [62] extend this approach by parameterising FA as a function of arc length along each fibre. They fit B-splines in order to regularise and average multiple fibre bundles into a single vector representation per subject, perform principal component analysis to identify the major modes of variance within the data set, and test for significant differences across populations using hypothesis testing.

In contrast, tract-based spatial statistics (TBSS) avoids the requirement for perfect non-rigid alignment of brain images by restricting comparisons to the centre of major white matter tracts only. In TBSS FA data is registered to a common template space using a constrained non-rigid registration, which strikes a balance between optimal alignment of structures and ensuring that tracts are not warped. Then, to minimise the influence of imperfect alignment at regional boundaries and to remove the need for smoothing, a FA skeleton is fit to the centre of all major white matter tracts in the mean FA image. This is projected onto each subject's individual FA to ensure that the values represent the maximum FA at the centre of each tract. Finally statistical permutation tests are performed with the option of threshold-free cluster enhancement [157] pre-processing. This seeks to enhance the contribution of clusters of similar values whilst reducing the impact of noise. The outcome is a set of voxels on the FA skeleton which show significant differences for each population.

In the approach proposed in this chapter, the issue of registration to a common template is avoided by performing multivariate statistical analysis of whole-brain connectivity matrices, derived using the framework proposed in Chapter 4. Therefore the examples compared are not voxels but feature vectors representative of mean anisotropy, or global probability values for all tracts. The issue of registration is re-addressed in Chapter 7, where we consider an atlas-based approach.

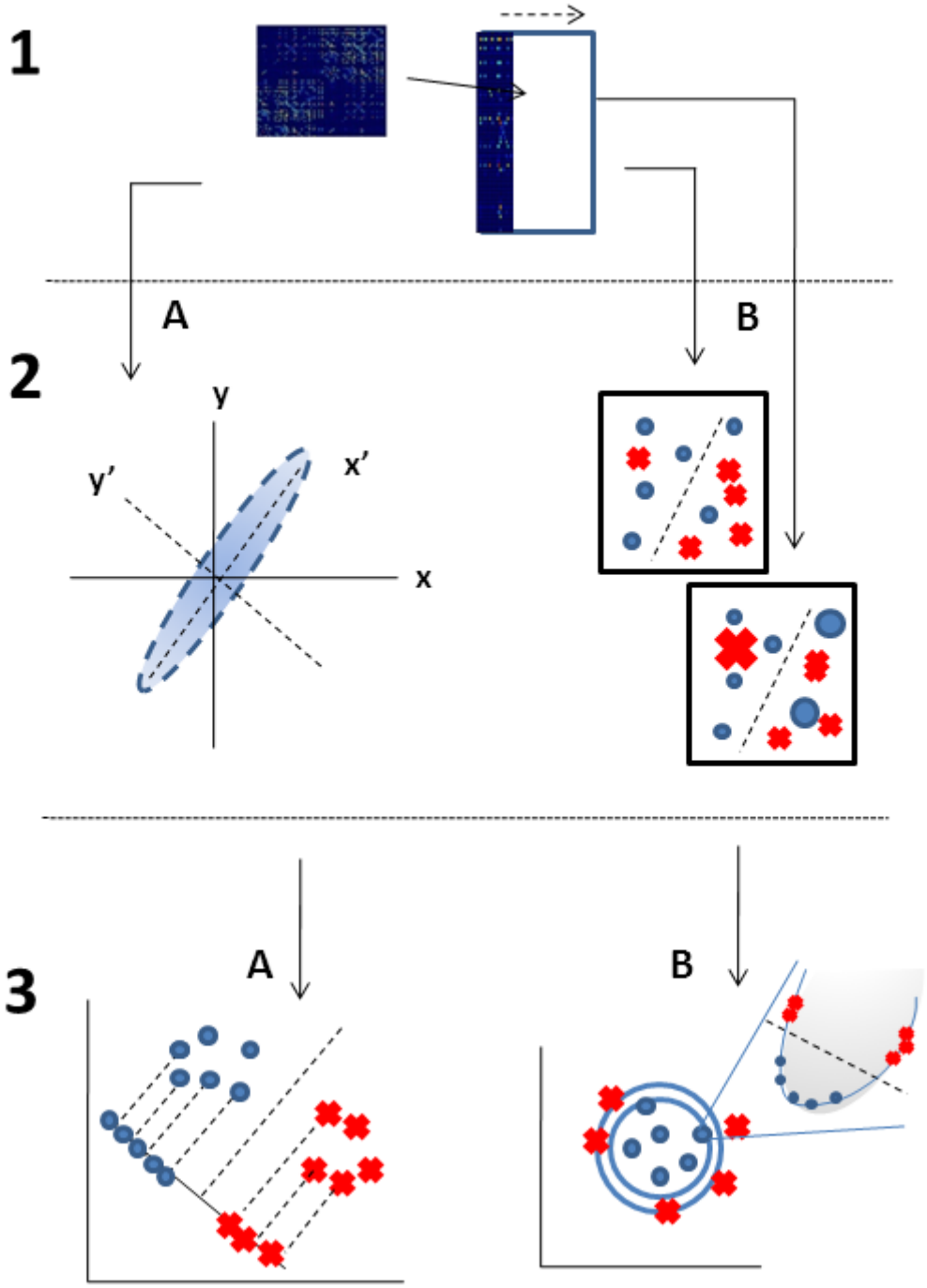


Figure 6.1: Machine learning framework: 1) feature vectors are formed by concatenation of all rows of each connectivity. 2) Dimension reduction is performed by either a) PCA or b) GentleBoost feature extraction. PCA rotates the data to a basis which better reflects the variance in the data (shown by major axes of variation of the ellipse). Projecting the data matrix onto the PCA eigenvectors reduces the dimension of problem. In contrast, GentleBoost serially trains weak classifiers on individual features from the data set. Examples that are misclassified are re-weighted so that they are given higher precedence in the next round. A sub-set of features is collected over multiple rounds of boosting. 3) Classification is performed using either a) MLDA or b) non-linear SVMs. MLDA finds a projection that maximises between class separation. SVMs map the data to a higher dimensional space where it is separable

6.3 Machine learning approaches

This section presents two approaches for multivariate statistical analysis of whole-brain connectivity networks. Both approaches use state of the art machine learning techniques. The aim is to classify examples correctly using the information provided by the feature vectors, and identify those features (in this case tracts) which differ most across the populations. Our approach consists of three stages (figure 6.1):

1. A feature vector for each subject is generated by concatenating the rows of each connectivity matrix. These are then mean centred (by subtracting the average vector from each feature vector) and combined as columns in a large data matrix (figure 6.1 1).

2. Dimensionality reduction is performed by either PCA or GentleBoost feature extraction (figure 6.1 2).

3. Classification is performed either by either a MLDA [177, 176] or non-linear SVM approach. The aim is find a hyperplane through the data which separates the populations.

The machine learning approaches are described in more detail in the next section. We consider results for two-class problems where data vectors, $\mathbf{x_i} \in \mathbf{X}$ have corresponding labels $y_i \in \{-1, +1\}$.

6.3.1 PCA+MLDA

Dimensionality reduction with PCA

Principal components are calculated via eigen-decomposition of the covariance matrix ($\mathbf{S}$):

$$\mathbf{S} = \frac{1}{N-1}(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T \tag{6.1}$$

where $\mathbf{X}$, is formed by concatenating all feature vectors such that each subject is represented by one column of the matrix. It is assumed that all N patterns are linearly independent and therefore the rank of the covariance matrix is $m = N - 1$ ($m < n$). The m eigenvectors of S which correspond to non zero eigenvalues, form the principal components.

MLDA classification and feature extraction

Following eigenvector calculation, the data is projected onto the set of principal components, reducing the dimension of the problem to m and rotating the data to a basis which better reflects the variance. In this new basis, MLDA attempts to separate groups of samples by projecting the output onto a single dimension which maximises the ratio between the between-class scatter matrix ($\mathbf{S_b}$) and the within class scatter matrix ($\mathbf{S_w}$):

$$\mathbf{S_b} = \sum_{i=1}^{g} N_i(\bar{\mathbf{x}}_\mathbf{i} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{i} - \bar{\mathbf{x}})^T \tag{6.2}$$

$$\mathbf{S_w} = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (\mathbf{x_{i,j}} - \bar{\mathbf{x}})(\mathbf{x_{i,j}} - \bar{\mathbf{x}})^T \tag{6.3}$$

Here $g = 2$ refers to the numbers of classes and $\mathbf{x_{i,j}}$ are formed from the m-dimensional rows of the data matrix after projection onto principal components and class separation. Note $\bar{\mathbf{x}}_\mathbf{i}$ is the group mean vector and $\bar{\mathbf{x}}$ is the mean vector over all samples.

In cases where the dimension of the problem is close to the number of examples, the optimal solution is found by maximising the following ratio:

$$\frac{\mathbf{P^T S_b P}}{\mathbf{P^T S_w P}} \ , \tag{6.4}$$

where $\mathbf{P}$ is the projection matrix. This reduces to an eigensystem problem:

$$(\mathbf{S_w^{-1} S_b})\mathbf{P} = \mathbf{P\Lambda} \tag{6.5}$$

where, for two classes there will be at most one non zero eigenvalue. The corresponding eigenvector forms the discriminant vector $\mathbf{w}$.

In circumstances such as this, where $\mathbf{S_w}$ has $(N-2)$ or less linearly independent eigenvectors $\mathbf{S_w}$ is often singular or unstable. In this instance regularisation methods are sought which offer alternatives to estimation of the covariance matrix by sampling. The maximum uncertainty (MLDA) approach [177, 176], calculates a modified within class scatter matrix $\mathbf{S_w^*}$ using covariance selection, where all non-dominant or small eigenvalues are increased but large eigenvalues are left constant. In this way covariance information is retained but unreliable eigenvalues are replaced to allow matrix inversion.

The modified within class scatter matrix $\mathbf{S}_\mathbf{w}^*$ is found by first finding all N_{eig} eigenvalues $\mathbf{\Lambda}$ and vectors $\mathbf{E}$ of the pooled covariance matrix $\mathbf{S_P}$:

$$\mathbf{S_P} = \frac{\mathbf{S_w}}{N-g} \tag{6.6}$$

Then calculating the mean eigenvalue $\bar{\lambda} = trace(\mathbf{S_P})/N_{eig}$ and forming a new matrix of eigenvalues based on the largest values:

$$\mathbf{\Lambda}^* = diag[max(\lambda_1, \bar{\lambda}).....,max(\lambda_n, \bar{\lambda})]. \tag{6.7}$$

This is used to form a modified within-class scatter matrix:

$$\mathbf{S}_\mathbf{w}^* = \mathbf{S}_\mathbf{P}^*(N-g) = (\mathbf{\Phi}\mathbf{\Lambda}^*\mathbf{\Phi}^\mathbf{T})(N-g) \tag{6.8}$$

which replaces $\mathbf{S_w}$ in equation 6.4 resulting in a modified projection matrix $\mathbf{P}*$, and corresponding discriminant vector $\mathbf{w}*$. The classifier f_{mlda} can then be estimated by setting a threshold (c) on the projection such that:

$$f_{mlda}(\mathbf{x_i}) = \begin{cases} +1 & \text{if } \mathbf{w}^* \cdot \mathbf{x_i} > c \\ -1 & \text{if } \mathbf{w}^* \cdot \mathbf{x_i} < c \end{cases} \tag{6.9}$$

Here the $\mathbf{x_i}$ are modified feature vectors calculated as a result of projecting the data onto the set of principal component eigenvectors.

An advantage of this simple linear approach is that, unlike non-linear kernel methods, the most discriminating hyperplane can be easily mapped back to the domain of the connectivity matrix. This enables selection of those features (or connections) that differ most between groups. Features are extracted in reverse order of the original process. First, matrices indicative of the extrema of each class are obtained by selecting points furthest away from the hyperplane. Then, by 1) multiplying the point with the transpose of the linear discriminant vector; 2) multiplying the output with the transpose of the PCA eigenvector matrix and 3) finding the difference between the results for each class, it is possible to identify those features that discriminate best between groups. Features are tested for significance. To exclude the influence of any population based differences in the ROI labelling we remove contributions from subjects for which no connections appear and test for significant differences in anisotropy (or global probability) values alone.

6.3.2 SVMs with GentleBoosting

Simultaneous dimension reduction and feature extraction

Boosting is a technique which greatly improves the performance of classifiers by serially training weak learners on weighted data [51]. In this approach, all data sets are initially given a uniform weighting. Then a weak learner, for example a decision tree, or stump[1], is trialled on all dimensions of the data. The best preforming dimension forms the first feature. Examples are then re-weighted such that miss-classified examples are given higher weight and thus precedence in the next round. This process is repeated over several rounds until all training vectors are correctly classified.

AdaBoost [51] is the original boosting algorithm. This aims the find a weak classifier $f_t(\mathbf{x_i}) \in \{+1, -1\}$ at each step which minimises the training error with respect to a distribution, D_t. Initially uniform, this is updated at each round from the error rate (ε_t) of the previous round of boosting (t):

$$D_{t+1}(i) = \frac{D_t(i)e^{-\alpha_t.y_i.f_t(\mathbf{x_i})}}{Z_t} \tag{6.10}$$

Here,

$$\alpha_t = \frac{1}{2}ln\frac{1-\varepsilon_t}{\varepsilon_t} \tag{6.11}$$

and Z_t is a normalisation constant. This is chosen to ensure $D_{t=1}$ remains a probability distribution.

The log ratio can be numerically unstable, which makes AdaBoost very sensitive to noisy data and outliers [52]. In contrast, GentleBoost [52] has been shown to be more robust. This is because it re-weights examples directly from the weighted class probabilities (P_w) rather than using a log ratio.

$$f_t(\mathbf{x_i}) = P_w(y_i = 1|\mathbf{x_i}) - P_w(y_i = -1|\mathbf{x_i}) \tag{6.12}$$

Weights are adjusted as:

$$D_{t=1}(i) \leftarrow D_t(i)e^{-y_i f_t(x_i)} \tag{6.13}$$

The weak learner f_t is calculated from least squares weighted regression.

[1] A stump is a one dimensional decision trees defined simply by a feature, a Boolean and a threshold which decides whether examples above the threshold are positive or negative

In this study boosting is not used as a classifier, rather it is used for simultaneous feature extraction and dimensionality reduction. The top features, selected over several rounds, are then passed to a non-linear support vector machine for classification. This has been shown to improve classification rates. However, performing boosting prior to classification has the advantage that a series of correlated features are selected from the data. These, can be used to point to patterns of differences between groups[2].

Support vector machines

The purpose of SVMs [30] is learning a classifier which maximises the separation between classes:

$$f_{svm}(\mathbf{x}) = \langle \mathbf{x} \cdot \mathbf{w} \rangle + b \tag{6.14}$$

For the linearly separable case this problem can be formulated by fitting the hyperplane:

$$\mathbf{x} \cdot \mathbf{w} + b = 0 \tag{6.15}$$

and parallel hyperplanes:

$$H_1 : \mathbf{w} \cdot \mathbf{x} + b = 1 \qquad \text{for } y_i = +1 \tag{6.16}$$

$$H_2 : \mathbf{w} \cdot \mathbf{x} + b = -1 \qquad \text{for } y_i = -1 \tag{6.17}$$

between which no examples are found. Here $\mathbf{w}$ is the vector (normal to the plane) that maximises the margin. Points which fall on H_1 and H_2 are the support vectors (figure 6.2).

This is solved as a quadratic optimisation problem:

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x_i} \cdot \mathbf{x_j} \tag{6.18}$$

subject to the constraints:

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x_i} \tag{6.19}$$

[2]The GentleBoost-SVM algorithm was provided by Michel Valstar [184] and is based on the open source SVM software package "Spider" (http://www.kyb.tuebingen.mpg.de/bs/people/spider/)

$$\sum_i \alpha_i y_i = 0 \tag{6.20}$$

where α_i are positive Lagrangian multipliers.

The specific advantage of this representation is that the problem is formulated as dot products between vectors. This means that for non-linearly separable data it is possible to map the data to a higher dimensional space H, $\Phi : \mathbb{R}^d \mapsto H$, where the data can still be separated by a hyperplane. This changes the training algorithm to functions of the form $\Phi(\mathbf{x_i}) \cdot \Phi(\mathbf{x_j})$, which can be conveniently rewritten as kernel function $K(\mathbf{x_i}, \mathbf{x_j}) = \Phi(\mathbf{x_i}) \cdot \Phi(\mathbf{x_j})$ The kernel used here is the radial basis function:

$$K(\mathbf{x_i}, \mathbf{x_j}) = e^{-||\mathbf{x_i}-\mathbf{x_j}||^2/2\sigma^2} \tag{6.21}$$

where σ is the standard deviation.

For data which is non-separable, an optimal solution can be found by setting the parameter ξ such that:

$$\mathbf{w} \cdot \mathbf{x} + b \geq 1 - \xi_i \text{ for } y_i = +1 \tag{6.22}$$

$$\mathbf{w} \cdot \mathbf{x} + b \leq 1 + \xi_i \text{ for } y_i = -1 \tag{6.23}$$

which fixes the cost of errors in the objective function. This and σ are optimised during training. More details on SVMs are provided in [30].

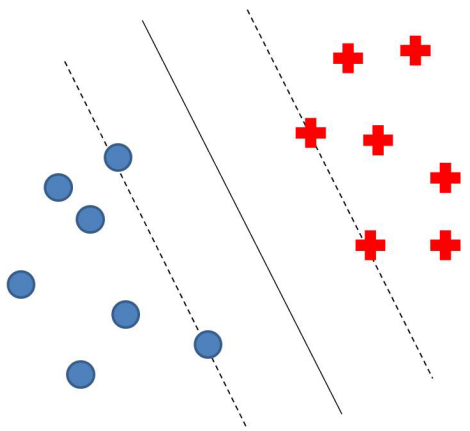


Figure 6.2: Support vector machines aim to maximise the separation between classes by defining a separating hyperplane and parallel hyperplanes between which no examples are should found. The figure shows an example for a linear SVM

6.3.3 Cross validation

For both classifiers performance is assessed by K-fold cross validation. This tests the predictive performance of classifier on unseen data. Here the population is randomly separated into K groups. One group of (test) subjects is then removed from the population and the remaining $K-1$ groups data are used for training. The classifiers learnt during the training stage (given by equations 6.9 and 6.14) are used to test the remaining subjects. The proportion of correctly classified test subjects provides a classification rate for each fold. This is repeated for each of the K groups and the mean and the standard deviation over all folds are reported. In this study 10-fold classification is used as this generates a more robust prediction of classification accuracy than leave one out classification, which removes a single test subject from the population each time.

6.4 Experiment

6.4.1 Subjects

Connectivity analysis was performed on 90 adult brains separated into two age groups a) 20-30 years (40 subjects) and b) 60-90 years (50 subjects). The data was acquired using the same protocol as for section 5.7.1. This is also taken from the same data set used in the published study (see Publications, [3]) however the population used is different (90 rather than 96 subjects). This is because some brain images showed significant deformation due to ageing, which prohibited their inclusion in the atlas-based study[3] (next chapter 7). We therefore left them out for consistency. This did not change the age range of the population (20-90).

6.4.2 Whole brain inter-regional tractography

Whole-brain tractography results were extracted following the framework proposed in chapter 4 for both mean anisotropy and global probability measures. Examples of whole brain tractography output for the mean anisotropy and global probability measures are shown in Figure 6.3 a and b.

In contrast to previous studies, tractography was run using the default parameters (curvature threshold 0.2, streamlines 5000). The aim was to better reconstruct tracts subject to crossing. However, as this

[3]This requires non-rigid registration of all brain images to a single template image selected from the population

might also increase the number of false positive connections, we also performed hypothesis test-based thresholding using a one sided sign test with Bonferri correction (for multiple comparisons). We tested against the null assumption that mean probability for connection is zero ($p < 0.01$). If the connection was not significant it was set to zero for all subjects. This is equivalent to the method proposed in Gong et al [59]. Following thresholding, mean correlation across subjects was improved from 0.70 ± 0.01 to 0.76 ± 0.04.

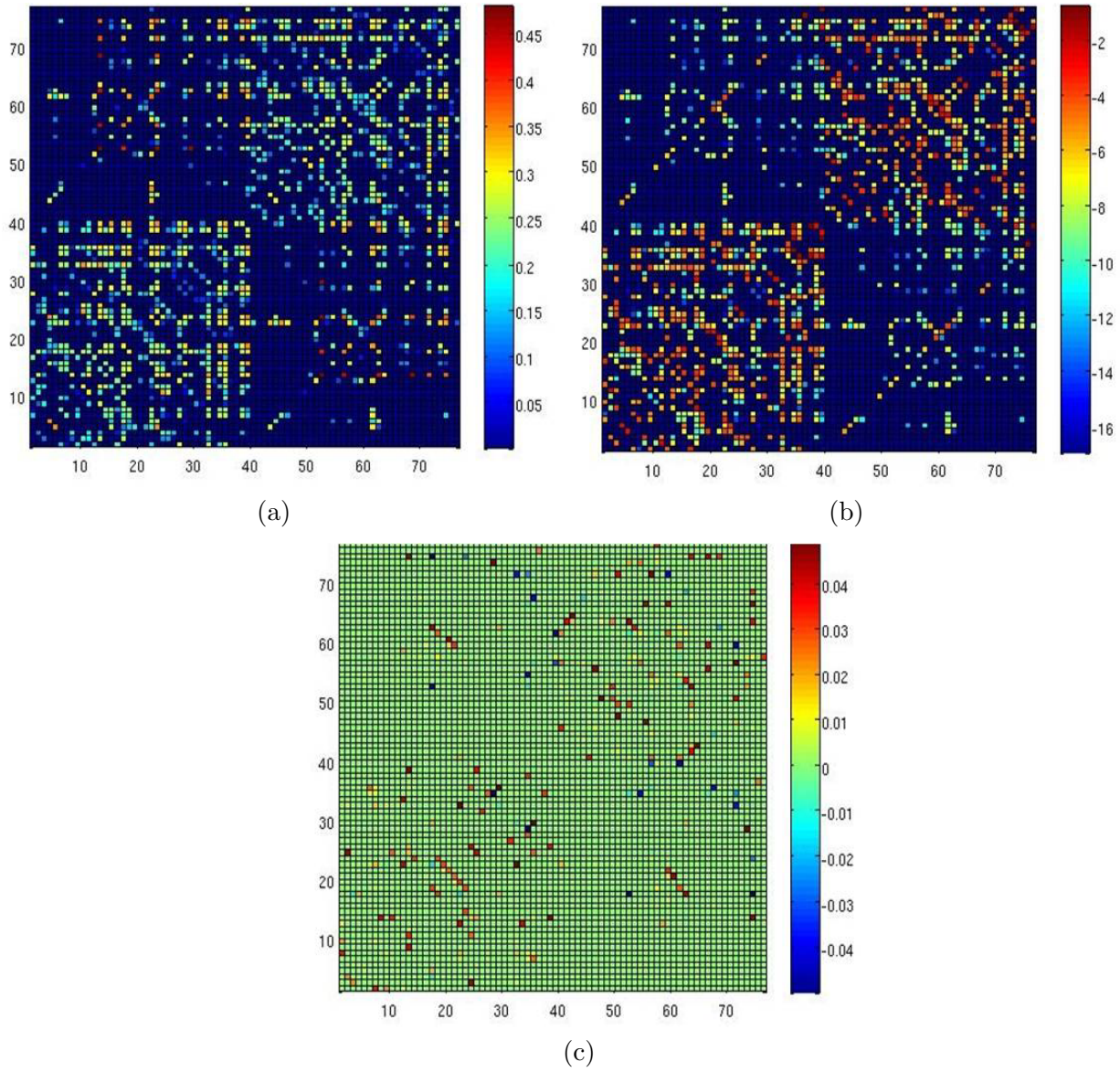


Figure 6.3: Connectivity matrices: a) mean tract anisotropy matrix (averaged over the population); b) the logarithm of the global probability matrix (averaged over the population); c) t-test matrix showing significant differences between the populations

Mean weights (estimated across all non zero connections for each population) were higher for the 20-30 years age group with values of 27.24 ± 0.01 compared to 26.81 ± 0.01 for the 60-90 years groups. The difference is small but was found to be significant ($p = 0.04$). In addition t-tests were performed for all non zero connections to detect which tracts differ significantly in mean anisotropy between populations. The results are pictured in figure 6.3. We rescale the global probability matrix, by taking the log, to improve visualisation of the connections.

6.5 Results: PCA+MLDA

6.5.1 Classification

Classification results for 10-fold cross-validation are shown in Table 6.1. In addition, plots are shown in figure 6.4,where test subjects are projected onto the MLDA plane (calculated during training) and the distance of the projected point from the hyperplane is plotted against age for each subject. In both instances, the separation from the hyperplane increases slightly with age for the older population.

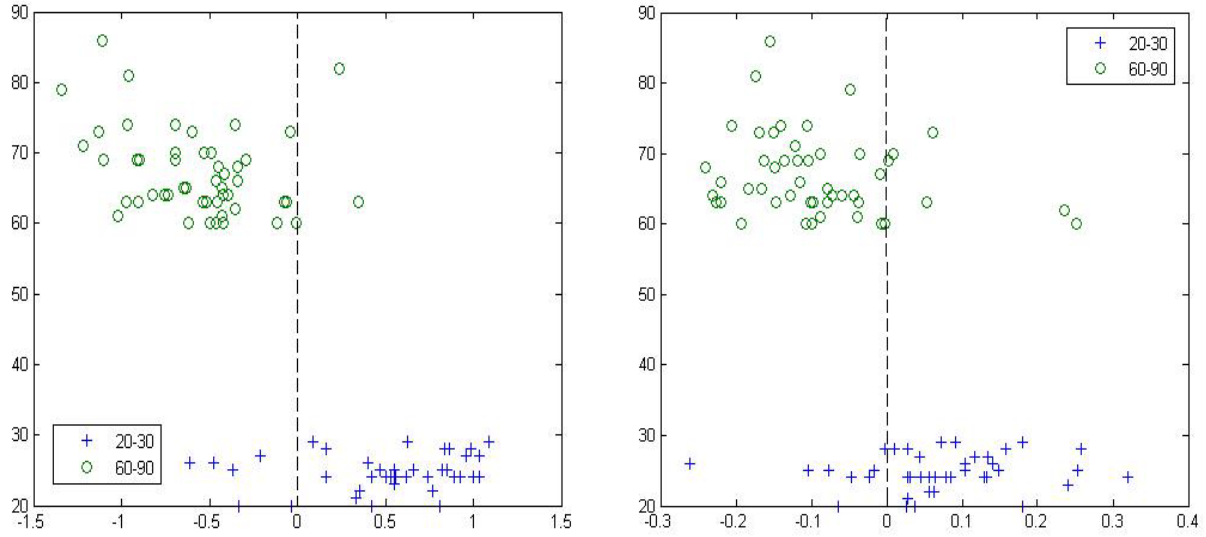
Classification rates are good for both connectivity measures[4]. This suggests that the approach still retains high predictive power when 10% of the population is removed, which indicates that it would make robust predictions for unseen data. In addition results are improved compared to previous work (see Publications, [1, 3]). This suggests that thresholding improves sensitivity of the approach.

The influence of gender was ruled out by splitting the population into male and female populations and repeating classification results for the mean anisotropy measure. There were 53 female subjects (37 aged 60-90 years) and 39 male subjects (18 aged 60-90 years). Classification results remained high (90.33 ± 13.92% for female subjects and 84.17 ± 21.65 % for male). The slight reduction in classification rates and increase in variance is likely due to the reduction in the number of training subjects.

Table 6.1: Classification results for the PCA+MLDA approach

	Classification rate (%)	Sensitivity (%)	Specificity (%)
Mean anisotropy	91.11 ± 7.03	87.33 ± 14.62	96.57 ± 7.35
Connection probability	83.33 ± 13.09	83.50 ± 15.04	84.07 ± 15.00

[4] the relatively high variance is representative of the small population size for each fold (< 10 test subjects)



(a) (b)

Figure 6.4: Visualisation of MLDA cross validation results: a) for mean anisotropy measure and b) global probability measure. The x-axis represents projected distance from the hyperplane (positive for 20-30 years age group); y-axis represents age.

6.5.2 Feature extraction

The most significant discriminant features for each connectivity measure are shown in Tables 6.2 and 6.3. For visualisation purposes tracts were also transformed from each subject to a standard space using: 1) non-rigid registration of each subject's FA image to the FMRIB58_FA_1mm brain template (available from FSL distribution 4.1) and 2) affine transformation to a normalised anatomical space (corresponding to the MNI152 atlas).

In each case features are represented by tracts drawn for mean anisotropy measurements. This is so as to inform correspondences and to assert whether differences in global probability measures reflect differences in anisotropy within the feature. Tracts are averaged across groups and subtracted in order to show the difference between populations. Figures show 3D renderings of the features projected onto anatomical space together with their respective seed (light blue) and target (lilac) regions. Red voxels show where voxels along the pathway have, on average, higher mean anisotropy for the young population. Blue voxels are, on average, stronger for the older population.

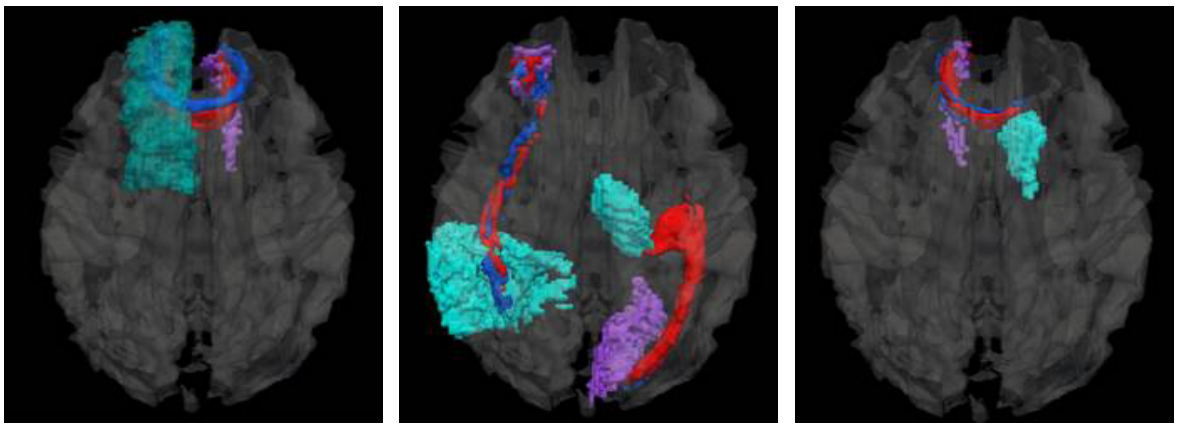
Mean anisotropy results

The results for the 20-30 years population, shown in table 6.2 are broadly described by three groups: 1) Anterior-posterior connections, these include thalamus to lingual gyrus, putamen to lingual gyrus, and posterior temporal gyrus to anterior orbital gyrus (Figure 6.5 b). 2) Basal ganglia connections, these pass between the caudate and medial orbital gyrus (Figure 6.5 c). 3) Frontal/frontal lateral connections, these include feature numbers: 1,2,6,8 and 9. Feature 1, the right superior frontal gyrus to left medial orbital gyrus tract, is shown in figure 6.5 a.

These findings agree with results of established techniques such as tract based spatial statistics analysis (TBSS) [89], voxel based morphometry (VBM) [25, 60], diffusion tracking [38, 172] and histology [29]. In particular VBM studies support a decrease in white matter volume within pre-frontal regions which may reflect the frontal/frontal lateral features seen here (Figure 6.5 a). VBM studies have also detected WM decrease within the posterior limb of the internal capsule, this projects from the thalamus to the occipital lobe [60], and is represented by the feature between the thalamus and lingual gyrus (Figure 6.5 b). In addition, other studies report graduated anisotropy decrease between frontal and posterior regions [38, 172]. This is believed to be significantly related to age related declines in executive functions such as problem solving and working memory [45, 97, 172] and exacerbated by the fact that associative regions are more metabolically active and therefore vulnerable to accumulation of damage. Some studies have also linked cognitive decline in ageing to degeneration of frontal-striatal connections involving the caudate [141] (Figure 6.5 c).

In contrast, the results for the 60-90 years population (table 6.2, bottom) are collectively characterised by features that pass through the splenium of the corpus callosum (Figure 6.6). This is supported by several studies which report relative preservation of FA in the splenium, thought to be facilitated by the larger size and earlier myelination of posterior fibres relative to anterior ones [14, 29]. Also Kochunov et al. [89] perform TBSS of the corpus callosum and forceps, reporting either positive or no correlation with age for the voxels of the splenium and the callosal body. However, these results suggest an unexpected increase in anisotropy. This may be explained by studies that report continued increase in FA in frontal brain regions until the mid to late twenties [104].

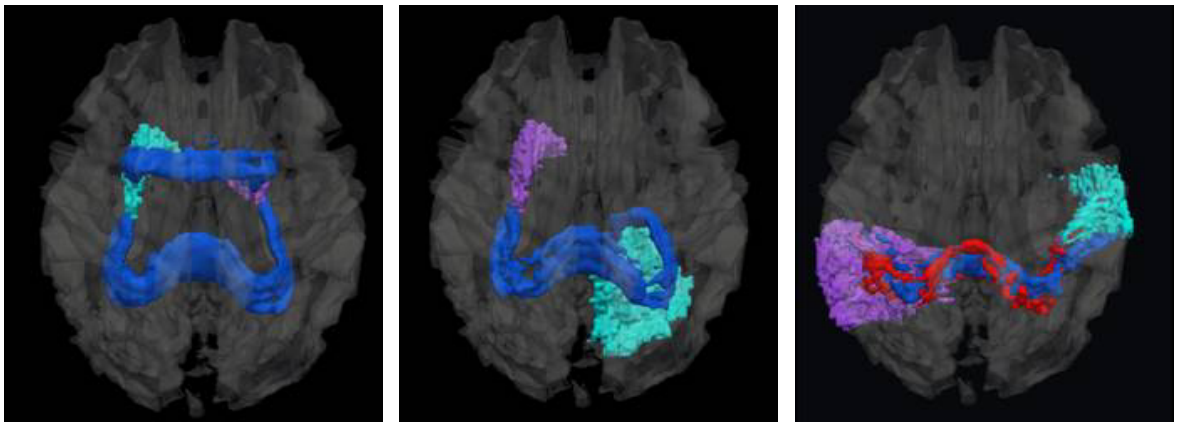
The features for the 60-90 years population include multiple connections featuring the putamen, thalamus, cerebellum and temporal lobe ROIs. This is supported by the results of Good et al. [60] who perform VBM of the white matter and report relative preservation of the cerebellum. In addition,



(a) (b) (c)

Figure 6.5: Mean anisotropy features from PCA-MLDA analysis for the 20-30 year population. These are three-dimensional renderings showing the difference in strength between some of the connections which feature as being significantly stronger in the younger population. Red voxels are where mean connection strength across populations is higher for the 20-30 year age group. Blue voxels are where mean connection strength across populations is higher for the 60-90 year age group. Regions 1 from table 6.2 are coloured light blue, regions 2 are lilac. The connections are: a) Superior Frontal R & Medial Orbital b) Thalamus L & Lingual L and Posterior Temporal R & Anterior Orbital R c) Caudate L & Medial orbital R

Davis et al. [38] report relative preservation of temporal lobe white matter through tractography, and Bergfield et al [25] report relative preservation of the thalamus and putamen (also in a VBM study). However, it is difficult to associate many of the selected connections with recognised white matter fibre projections. This may reflect the fact the white matter voxels can contain many different fibre bundles.



(a) (b) (c)

Figure 6.6: Mean anisotropy features from PCA-MLDA analysis for the 60-90 year population. Colours are the same as for figure 3. All pass through the splenium of the corpus callosum. a) Putamen R & Pallidum L; b) Cerebellum L & Putamen R c) Superior Temporal L & Posterior Temporal R

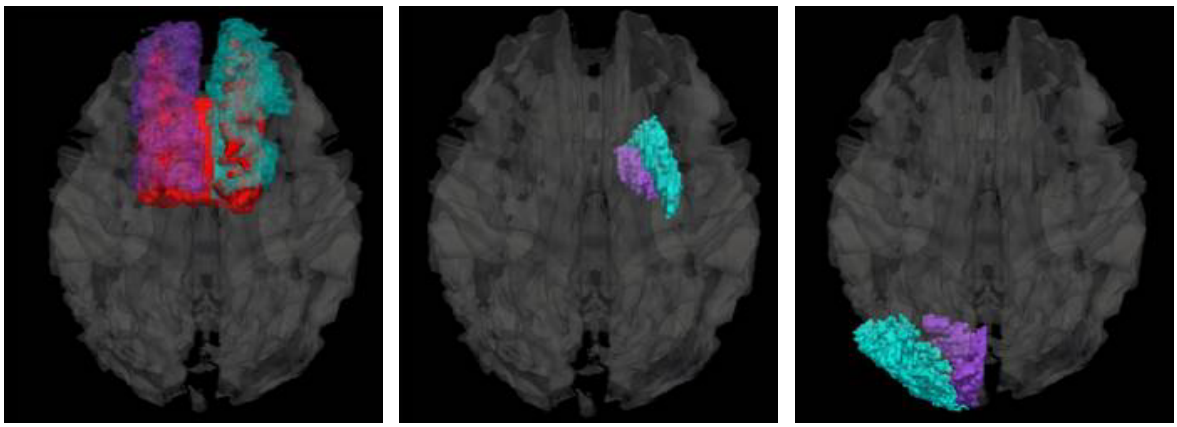
The same pattern of features are reflected in the female and male only analysis. For male subjects, the most significant features discriminating the young population from the old were the pre-central gyrus (right) to post-central gyrus (left), and superior frontal gyrus (right) to subgenual frontal (left). Discriminating old from young was the caudate (right) to putamen (left). For women the most significant features discriminating young from old were the superior frontal gyrus (right) to medial orbital gyrus (left), and contra-lateral medial orbital connections. For older subjects the most discriminating features were the cerebellum (left) to putamen (right), and putamen (right) to pallidum (left). There was a reduction in the significance of the findings. This may be due to fewer training samples.

Table 6.2: Table of PCA-MLDA features for mean tract anisotropy analysis, in order of significance

	Region 1	Region 2	Significance
Features stronger in young population			
1	Superior Frontal R	Medial Orbital L	4.6510^{-10}
2	Medial Orbital L	Medial Orbital R	1.7210^{-09}
3	Thalamus L	Lingual L	7.2510^{-04}
4	Putamen L	Lingual L	1.7010^{-03}
5	Pre-central R	Post-central R	2.9010^{-03}
6	Anterior Orbital L	Medial Orbital R	4.3010^{-03}
7	Caudate L	Medial Orbital R	2.0210^{-02}
8	Straight Gyrus R	Subgenual Frontal L	2.2110^{-02}
9	Superior Frontal R	Subgenual Frontal L	2.6610^{-02}
10	Posterior Temporal R	Anterior Orbital R	3.2810^{-02}
Features stronger in old population			
1	Putamen R	Pallidum L	1.4710^{-06}
2	Cerebellum L	Putamen R	1.5410^{-05}
3	Insula L	Putamen R	2.6010^{-03}
4	Insula L	Thalamus R	4.5010^{-03}
5	Superior Temporal L	Posterior Temporal R	6.3010^{-03}
6	Insula L	Straight Gyrus L	1.1510^{-02}
7	Posterior Temporal R	Putamen L	1.2210^{-02}
8	Cerebellum R	Posterior Cingulate R	1.5410^{-02}
9	Insula R	Pallidum L	2.1810^{-02}
10	Cerebellum R	Putamen L	3.9010^{-02}

Global probability results

The analysis of the global probability measure reflects its sensitivity to changes in seed and target region size. This is exacerbated by the PCA step which rotates the data to a basis which reflects the largest variance, and is reflected by a large number of features representing connections from adjacent ROIs, for example connections between the left putamen and pallidum (Figure 6.7 b and feature 4: 20-30 year group), and connections between the occipital lobe and cuneus for both hemispheres (Figure 6.7c and features 3 and 7: 60-90 year population). In addition a number of features for the 60-90 year age group represented connections within frontal regions, which goes against the results from the mean anisotropy analysis and the literature [60, 89, 135]. Grey matter volume is expected to decrease not increase in this age group [60, 25], especially within frontal regions. Therefore this may reflect a reduction in volume of other parts of the seed region, in which case the proportion of the seed region projecting to the target may increase even if the actual streamline total stays the same.



(a) (b) (c)

Figure 6.7: Global probability features from PCA-MLDA analysis: Some features reflect expected patterns of white matter degeneration, for example a) Superior Frontal L & Superior Frontal R. However other features demonstrate probability measures sensitivity to region size: b) Putamen L & Pallidum L) c) Occipital R & Cuneus R

Table 6.3: Table of PCA-MLDA features for global connection probability analysis, in order of significance

	Region 1	Region 2	Significance
	Features stronger in young population		
1	Superior Frontal L	Superior Frontal R	8.4210^{-10}
2	Pallidum R	Substantia Nigra R	8.5410^{-06}
3	Precentral R	Postcentral R	1.0510^{-05}
4	Putamen L	Pallidum L	5.2610^{-04}
5	Medial Frontal R	Caudate R	4.8010^{-03}
6	Pallidum L	Substantia Nigra L	7.1010^{-03}
7	Putamen R	Pallidum R	1.0410^{-02}
8	Cerebellum R	Substantia Nigra R	1.0710^{-02}
9	Superior temporal Posterior L	Superior temporal Anterior L	1.8510^{-02}
10	Superior Frontal L	Substantia Nigra L	4.9410^{-02}
	Features stronger in old population		
1	Medial Frontal L	Superior Frontal L	9.4010^{-05}
2	Posterior Cingulate	Superior Parietal	7.7610^{-04}
3	Occipital Lobe R	Cuneus R	3.8010^{-03}
4	Brainstem	Substantia Nigra R	4.2010^{-03}
5	Putamen L	Superior Frontal L	4.310^{-03}
6	Medial Frontal R	Superior Frontal R	4.4010^{-03}
7	Occipital Lobe L	Cuneus L	5.610^{-03}
8	Putamen R	Superior Frontal L	8.5010^{-03}
9	Brainstem	Thalamus R	8.7010^{-03}
10	Straight L	Subgenual Frontal L	3.4410^{-02}

6.6 Results: GentleBoost+SVM

6.6.1 Classification

The results for cross-validation are shown in Table 6.4. Classification rates are lower for probability. This may reflect the reduced sensitivity of GentleBoost to the largest sources of variance in the data.

Table 6.4: Classification results for the GentleBoost-SVM approach

	Classification rate (%)	Sensitivity (%)	Specificity (%)
Mean anisotropy	90.00 ± 9.73	85.00 ± 17.48	94.00 ± 9.66
Connection probability	78.89 ± 8.20	77.50 ± 18.45	80.00 ± 13.33

6.6.2 Feature extraction

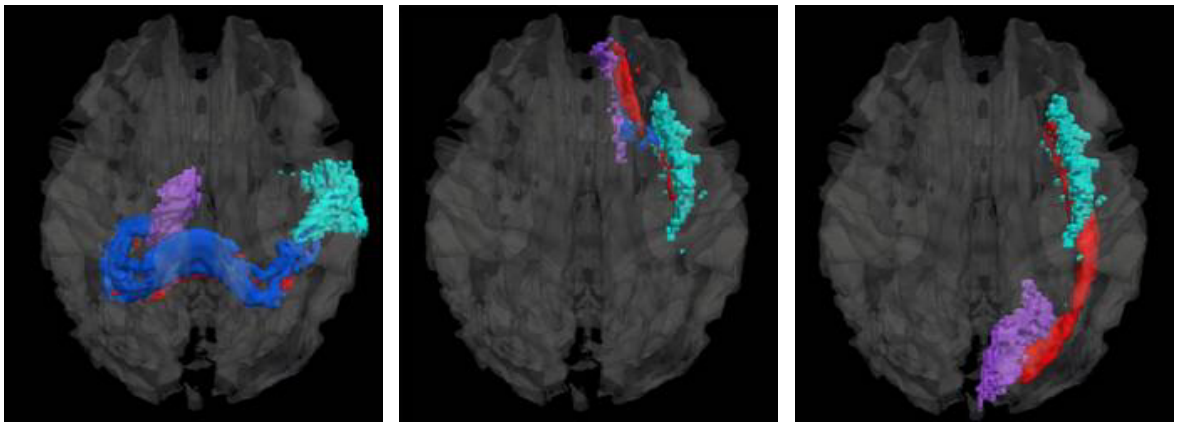
Features selected from the whole population by GentleBoost show much greater agreement between the connectivity measures (table 6.5). In addition the significance of the selected features for anisotropy is much higher than for PCA+MLDA. This reflects the fact that the GentleBoost approach is less sensitive to the main sources of variance in the data, caused by variability in tract propagation and segmentation results across the population. This means that the features selected from the mean anisotropy measure more often reflect tracts that are propagated across the full population, and that features selected for the global probability measure are less sensitive to large changes in region size. This may be reflected by the lower classification rates and significance for the global probability measure.

The results of the feature extraction follow the same pattern as the mean anisotropy results from the PCA-MLDA analysis. Connections that are stronger in the old population predominantly reflect connections across the splenium of the corpus callosum, and feature ROIs such as the cerebellum, thalamus, putamen and temporal lobe. These include an anisotropy feature between the left superior temporal gyrus and right thalamus (figure 6.8 a) and a probability feature between the right thalamus R and left superior frontal gyrus (figure 6.9 b).

Most of the features selected from the anisotropy analysis reflect connections that are stronger in the 20-30 yr population. This possibly reflects the fact that there are likely to be far more connections with higher anisotropy in this population. Several of these represent contra-lateral frontal, and anterior-posterior connections (figures 6.8 b,c). These include connections between the insula and the medial orbital and lingual gyri. In addition several basal-ganglia-frontal connections are featured, including multiple connections from the caudate and putamen (table 6.5).

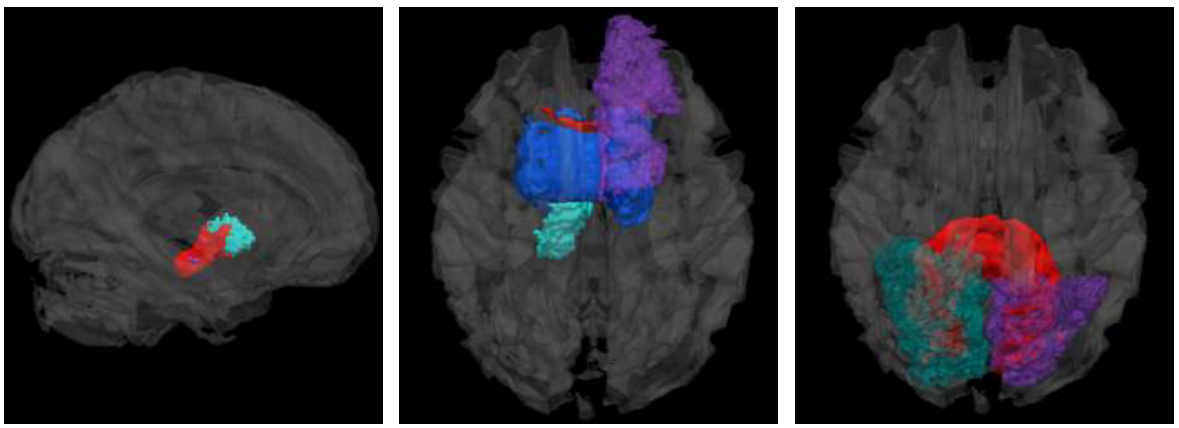
The features from the global probability matrices are more balanced across the populations. There are fewer contra-lateral frontal or anterior-posterior connections: only feature 9 (table 6.5) passes through the genu. However, there are several connections from basal ganglia regions (feature numbers, 1 and 3-8, table 6.5), including one between the right pallidum to right substantia nigra (figure 6.9 a), which may support the results from previous work which suggest connections from the substantia nigra degenerate even during healthy ageing (see Publications [3]). In addition, one feature is picked up contra-laterally between parietal lobes (figures 6.9 c). This reflects studies which show grey matter in the parietal lobe undergoes relatively accelerated decline [60].

Some connections are picked up by both PCA-MLDA and GentleBoost SVM approaches: these include the mean anisotropy features: superior frontal gyrus (right) to medial orbital gyrus (left), and putamen (right) to pallidum (left). In addition, global probability features: cerebellum (left) to putamen (right), insula (left) to putamen (left) and pallidum (right) to substantia nigra (right, also shown in Figure 6.9 a). Several features also reflect results from the previous study (see Publications [3]) including left insula to lingual gyrus connections, and superior temporal gyrus posterior (left) to thalamus (right), and putamen (left) to thalamus (right).



(a) (b) (c)

Figure 6.8: GentleBoost feature extraction for mean anisotropy data: a) Superior Temporal Posterior L & Thalamus R; b) Insula L & Medial Orbital L; c) Insula L & Lingual L



(a) (b) (c)

Figure 6.9: GentleBoost feature extraction for global probability data: a) Pallidum R & Substantia Nigra R b) Thalamus R & Superior Frontal L c) Superior Parietal R & Superior Parietal L

Table 6.5: Table of GentleBoost features, in order of selection

	region 1	Region 2	Significance	Greater in:
	Features from the mean anisotropy matrices			
1	Superior Temporal Posterior L	Thalamus R	5.1310^{-2}	60-90 years
2	Superior Frontal R	Medial Orbital L	8.4210^{-10}	20-30 years
3	Caudate L	Medial Orbital L	8.0210^{-13}	20-30 years
4	Putamen R	Pallidum L	1.4710^{-6}	20-30 years
5	Putamen R	Pre-central R	9.4010^{-3}	20-30 years
6	Insula L	Medial Orbital L	2.0310^{-7}	20-30 years
7	Cerebellum L	Putamen R	1.5410^{-05}	20-30 years
8	Insula L	Lingual L	3.2010^{-7}	20-30 years
9	Caudate R	Thalamus L	0.28	20-30 years
10	Insula L	Pre-central R	3.3710^{-6}	20-30 years
	Features from the global probability matrices			
1	Putamen L	Thalamus R	4.0010^{-3}	60-90 years
2	Thalamus R	Superior Frontal L	6.010^{-3}	60-90 years
3	Pallidum R	Substantia Nigra R	8.5410^{-6}	20-30 years
4	Anterior Temporal Lobe Medial R	Putamen R	3.0510^{-4}	60-90 years
5	Thalamus L	Pallidum R	0.3710^{-6}	60-90 years
6	Insula R	Putamen L	5.4910^{-2}	60-90 years
7	Nucleus Accumbens R	Medial Orbital R	6.810^{-3}	20-30 years
8	Putamen R	Pre-central R	2.2210^{-2}	20-30 years
9	Insula L	Straight R	0.45	20-30 years
10	Superior Parietal R	Superior Parietal L	1.0210^{-7}	20-30 years

6.7 Summary

This chapter proposes a new approach to identifying group-wise differences in the whole-brain network of macroscopic neural tracts, by combining a clinically applicable protocol for extracting models of connectivity with machine learning. Two machine learning approaches are proposed: a PCA-MLDA approach, which characterises the main sources of variance in the data, and a GentleBoost-SVM approach, which extracts combinations of features to allow optimal classification. Both techniques are used to successfully classify images from two different age groups. Classification rates are shown to be good, with an optimal classification rate of 91.11 $\pm$ 7.03% for mean anisotropy and 83.33 $\pm$ 13.09 for global probability. In addition both approaches were used to identify the tracts which differ most

significantly between the groups.

In general, both approaches generate findings which agree with results of established techniques such as tract based spatial statistics analysis (TBSS) [89], voxel based morphometry (VBM) [60], and diffusion tracking [172]. The results show good correspondence across approaches and compare well to our previous study (see Publications [3]). Many features are extracted by both machine learning approaches, and there is a common pattern of results. These suggest that connectivity measures are significantly stronger for the younger population within frontal brain regions, and this effect is graduated between anterior and posterior regions of the brain. In contrast, connections with the splenium are relatively preserved with age. However, although many ROIs are featured repeatedly, there is some variability in the specific connections selected. Furthermore, in contrast to previous results there are relatively few connections featuring the substantia nigra. This may reflect the removal of several subjects due to large deformation. Segmentation of small structures such as the substantia nigra was likely to have been unreliable in these subjects.

There are some advantages to our approach: in particular, in comparison to graph based approaches such as small world metrics, this is the first whole brain connectivity study which is able to pinpoint population differences in the underlying connective matrix. Also, unlike most diffusion tracking studies it does not require manual selection of specific regions of interest. It instead models connectivity for the whole brain, This could be particularly useful in instances where the impact or origin of the disease in the connective structure of brain is not fully understood (for example schizophrenia). Also, contrary to voxel based methods, the approach does not require cross-subject registration of diffusion images, and unlike TBSS the approach is not restricted to the centre of major white matter tracts and is able to distinguish between multiple connections within large fibre bundles such as the corpus callosum. Finally, it can also identify the end points of the connection and represent connectivity strength for the full path as opposed to a single focal region. This may be particularly important where the impact of the disease is dependent upon origin of the connections, for example in neurodegenerative diseases such as Parkinson's.

There are some limitations, for example it is possible that other variables may influence the results. In this case the influence of gender was ruled out as very similar results were obtained when using male and female populations. In particular, classification rates were maintained and extracted features were consistent with the general pattern of results for the full population. Nevertheless, ongoing work suggests that age may impact on gender classification.

A potential weakness of the PCA+MLDA approach is highlighted by the results from global probability matrices. These include multiple neighbouring regions and reflect the fact the PCA rotates the data to a basis which represents the variance in the data. In contrast the features extracted by the GentleBoost approach show greater similarity across measures. This is because GentleBoost is not sensitive to the largest variance within the data, but rather picks a combination of features that optimise classification. As such GentleBoost is potentially more reliable where the data is confounded by nuisance variables. Nevertheless, the reliability of the global probability measure may be improved by introducing a correction measure to reduce sensitivity to distance and correct for total GM volume.

The data also places some restrictions of the approach. Using 15 gradient direction diffusion data limits the angular resolution and therefore increases the uncertainty with regards to tract direction, creating the need to use a single fibre probabilistic tracking model. However at the resolution of diffusion data several axons can pass through each voxel and diverge or cross, leading to false positive connections. This problem could be partially alleviated by using 32 or more direction data which may be practical for some studies.

False positives are also contributed to by cumulation of errors during registration and segmentation. This is minimised as far as possible through optimisation of the registration process, erosion at the boundaries between regions (to reduce the effect of mislabelling), use of population based thresholding, and inclusion of constraints applied to the tracking to prevent tracts passing through CSF or grey matter. Nevertheless, the issue of propagation of errors from segmentation and registration is further addressed in the next chapter via a standard space approach. This removes inter-subject labelling variability and should reduce uncertainty in the data provided registration is complete.

In conclusion, this study has developed a practical method for the description and classification of the whole brain network of macroscopic neural connections. The approach allows classification of the differences between large groups of subjects. In addition features extracted from group analysis of patient/control populations could highlight vulnerable tracts for more detailed analysis. However, interpretation of specific features should be made cautiously on account of the potential for false positives.

Chapter 7

Probabilistic White Matter and Fibre Tract Atlas Construction

7.1 Introduction

Anisotropic diffusion in the brain is caused by the presence of physical barriers, formed predominantly from the axon membrane and myelin sheath. These restrict molecular motion perpendicular to tracts, and are known to increase during brain development, but decrease during ageing or in instances of disease. Measurements of diffusion anisotropy from images across multiple subjects can be used to study changes. However, this requires the establishment of correspondences across subjects. In the previous chapter this was achieved by comparing mean estimates of tract anisotropy in a machine learning framework. This avoids the requirement for registration of diffusion images between subjects. However, it is sensitive to segmentation variability and inconsistencies in tract propagation across subjects. In addition, it is unable to quantify anisotropic variations along the length of tract.

In contrast, spatial normalisation of whole populations of diffusion data, via transformation to a common co-ordinate space, allows comparisons of diffusion properties without bias from macro-structural differences in tissue volume. In addition, performing tractography in a common co-ordinate system, and mapping the results back to each subject's native space, enables study of the morphological and structural variability of tracts. This chapter therefore proposes a framework which builds a probabilistic white matter atlas from diffusion data acquired across a population. The next section provides more detail on diffusion data registration and discusses other techniques that have performed fibre tractography in a common co-ordinate space.

7.2 Background

Registration of diffusion data is made complicated by the fact that the diffusion profile at each voxel is three-dimensional and represents the orientation of local tissue microstructure. Therefore, in addition to spatial transformation, diffusion profiles or fibre bundles must be re-orientated in such a way, that they re-align with the underlying tissue.

Techniques for registration and re-orientation of diffusion tensor images were first proposed by Alexander et al [5, 55]. These extract the rotational components of affine transformations by either Finite Strain (FS) or preservation of principal direction (PPD) approaches (discussed in more detail in section 3.4). For non-rigid transformations, locally affine transformations can be estimated from the Jacobian of the transformation at each voxel. The resulting rotation matrices ($\mathbf{R}$) are then used to re-orient the tensors ($\mathbf{D}' = \mathbf{R}^{\mathbf{T}}\mathbf{D}\mathbf{R}$) :

Originally tensor re-alignment was performed following registration of scalar images such as FA, using traditional image similarity metrics, or diffusion tensor images, using tensor based similarity metrics. More recently, some approaches have also suggested incorporating the tensor re-orientation directly into the registration algorithm, for example, Zhang et al [197] propose a piecewise affine approach which optimises the algorithm according to a similarity measure estimated from the apparent diffusion coefficients (ADC) at each voxel. This has been shown to improve correspondences between fibre tracts propagated in each subject's native space and those propagated in an atlas generated by affine registration.

Registration of diffusion data allows diffusion properties to be compared at equivalent voxels across multiple subjects. Some groups have extended this to studying properties along major white matter tracts. For example Davis et al. [38] follow the approach of Corouge [34] by fitting B-spline curves to tracts propagated in native space via interpolation and averaging of tensors. They then parametrise FA as a function of arc length to spatially normalise tracts across subjects and allow investigation of age related FA change throughout the brain. Goodlett et al. [62] follow a similar approach, but follow tensor realignment by propagating tracts in a common co-ordinate frame using streamline tractography. They also fit B-splines, but use principal component analysis to identify the major modes of variance within the data set. Following this, they test for any significant difference across populations using hypothesis testing.

Other groups instead align and cluster fibre bundles, for example Ziyan et al. [199] propagate tracts

in native space, and then align fibres in a common coordinate space using a polyaffine framework. In contrast, O'Donnell et al. [121, 122] create a fibre bundle atlas by embedding and clustering streamlines from multiple subjects using a spectral clustering approach. In addition, Oh et al. [124] combine region of interest segmentation and tract propagation to provide point to point correspondence of tracts through the corpus callosum in a group of subjects. These approaches have also been used in population studies to assign equivalent points along tracts for FA comparison [122, 124] .

A potential limitation of these approaches is their reliance on streamline tractography which is sensitive to uncertainty and fibre crossing. In addition, white matter atlas approaches based on the averaging of tensors may smooth out details of population variability. One alternative to streamline tractography has been proposed, which performs atlas-guided probabilistic tractography by using the distribution of subjects' principal eigenvectors after tensor alignment to represent uncertainty [33]. Nevertheless, this does not incorporate any information regarding the suitability of the tensor model to represent the diffusion profile at each voxel. In contrast, this chapter proposes a framework which builds a probabilistic white matter atlas from probabilistic representations of diffusion, acquired across a population. It uses this to perform fibre tractography once for the whole population. The next section provides an overview of the approach

7.3 Overview

The framework aims to:

1. Generate a probabilistic white matter atlas for a population of subjects. Voxels in the atlas hold samples from the population distribution on the principle diffusion direction.

2. Perform whole-brain probabilistic tractography through the white matter atlas once. Then use this to generate a single set of tracts for the whole population.

3. Estimate mean anisotropy along all tracts, for each subject, and use this to build connectivity matrices. These are studied using the machine learning frameworks proposed in chapter 6.

4. Evaluate differences in anisotropy across the population, by transforming fractional anisotropy (FA) images for each subject onto the templates provided by each atlas tract. Then voxelwise comparisons are performed.

These steps are outlined in figure 7.1

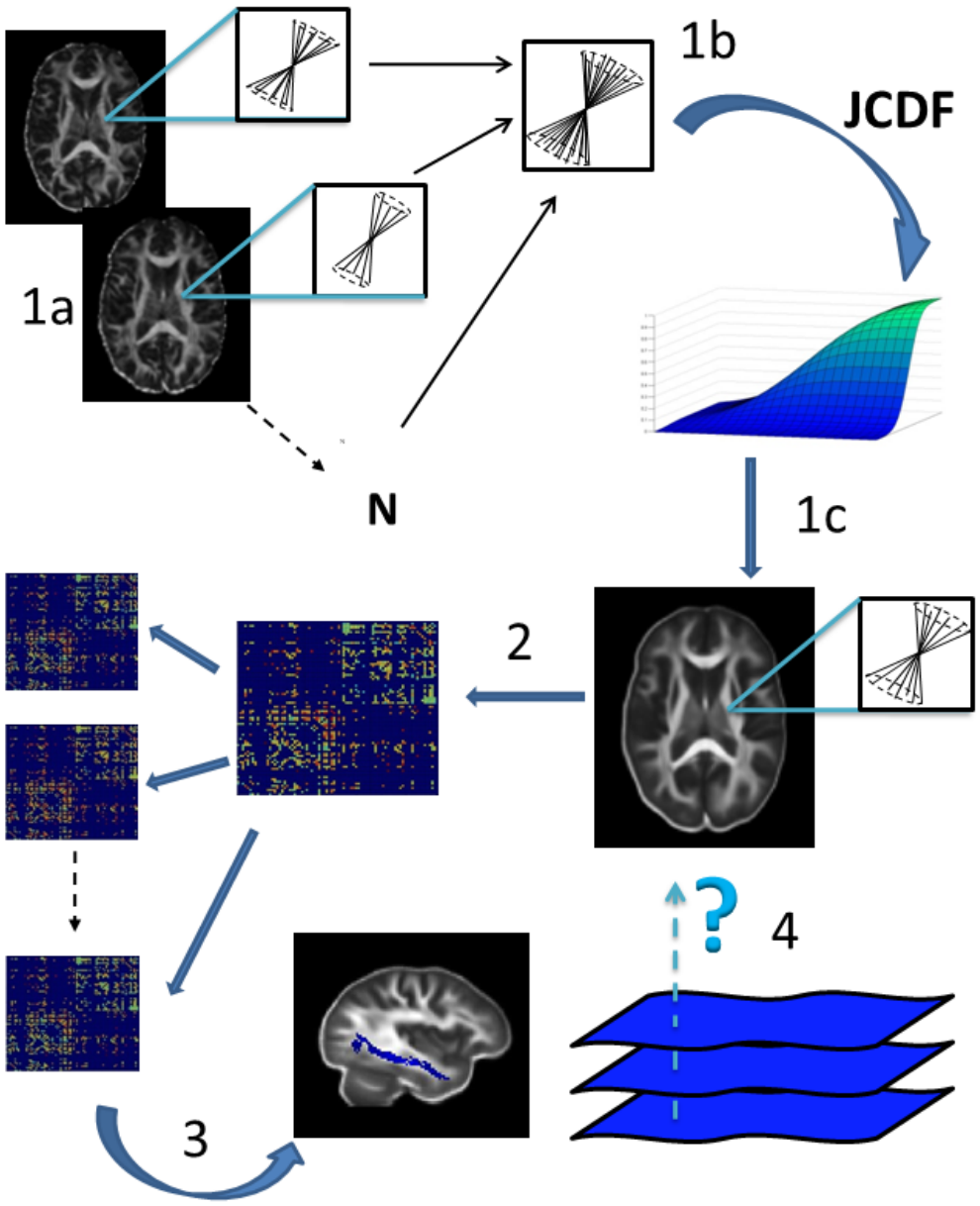


Figure 7.1: The probabilistic white matter atlas: construction and analysis: 1) Construction of the white matter atlas a) FA maps for multiple subjects are registered to a common co-ordinate frame where voxels now represent equivalent points in each brain. b) The posterior distributions on the principle diffusion direction for all subjects are re-oriented and combined at each voxel. c) These are used to form a joint cumulative distribution function (JCDF, in fact a discrete distribution) across the population. Samples are drawn from the inverse JCDF to form a new representative distribution for the population. 2) Whole-brain fibre tractography is performed through the atlas. 3) Connectivity matrices for each subject are formed by transforming each subject's FA data to the template of each tract, and then calculating mean and max FA. Machine learning analysis is used to identify features which differ across populations. 4) These features are analysed in more detail through voxel-wise comparisons of FA via permutation testing

7.4 Pre-processing

In this framework, spatial transformation is obtained by first performing non-rigid registration of rotationally invariant images (T1 or FA) to a reference subject. Registration is performed using free form deformations [144]. The following sections describe how the reference subject is selected from the population, and report creation of a T1 anatomical atlas of the population with corresponding ROI labels. These are later transformed to the co-ordinate frame of the white matter atlas and used to generate seeds for tractography.

7.4.1 Selection of a reference subject

The choice of reference subject for registration strongly influences the quality of the spatial alignment. Therefore, it is important to choose an image that is most similar to the rest of the population. As the objective is to align diffusion data, specifically white matter tracts, we assess similarity by comparing subjects' FA images. These are high intensity at the centre of white matter tracts and very low intensity elsewhere. To assess similarity all subjects' FA images are first mapped to the FMRIB57_FA_1mm atlas (provided as part of FSL)[1]. This enables comparisons to be made in a template space, which prevents the need for exhaustive registration between all pairs of images (prohibitive for large cohorts). Then, as all images are of the same type (from the same scanner), images are compared through estimates of cross correlation. The image with highest mean cross correlation with all other images is used as the reference.

7.4.2 Region of interest and tissue segmentation

Segmentation is performed on a T1 average image of all subjects. This is obtained by non-rigid registration of the T1 images of all subjects, to the T1 image of the referenced subject. This is followed by spatial normalisation. Registration is parametrised so as to allow significant deformation of the images. This ensures optimal alignment of all anatomical structures. The registration is optimised

[1] we use the non-rigid registration parametrisation proposed for tract-based spatial statistics (TBSS). This has one resolution level, 20mm control point spacing and uses cross correlation as a similarity metric

over four resolution levels with successive control point spacings of 20mm, 10mm, 5mm and 2.5mm. There is no additional regularisation ($\lambda_s = 0$).

Each subject's T1 image is associated with a set of 83 anatomical labels. These are obtained by automated atlas-based segmentation and label propagation following the approach described in section 4.4. The labels are then propagated to the co-ordinate frame of the T1 atlas using the transformations defined in the previous stage. They are fused using a majority vote rule. However, probabilities are assigned to each label according to the number of subjects that agree. These are later used to modulate the number of streamlines seeded from each seed voxel. In addition, tissue segmentations are also obtained for the T1 average image using SPM [8]. These are fused with the anatomical labels to retain 77 ROIs at the white matter grey matter boundary. White matter tissue probabilities are also used to constrain tractography.

7.5 White matter atlas construction

The spatial transformation and re-orientation of the diffusion data is calculated from transformations estimated between of all images to the reference subject. The specific transformations used are selected in a testing stage (described below in section 7.5.4). The following sections describe re-orientation:

7.5.1 Local diffusion model

The probabilistic atlas builds on the single fibre probabilistic tractography approach of Behrens et al [24], where instead of using a tensor model, the local model of diffusion fit at each voxel is a two component partial volume (ball and stick) model. Anisotropic diffusion is represented as a single vector in the principal fibre direction (θ, ϕ) and isotropic diffusion is represented by a spherical component. The posterior distributions on the key parameters of the diffusion mode (θ, ϕ and f) are estimated using Bayesian inference, where samples from the distribution are drawn using Markov Chain Monte Carlo (MCMC) sampling. Typically 50 samples are drawn per voxel. The re-orientation strategy transforms each sample on the principal fibre direction (θ, ϕ) independently.

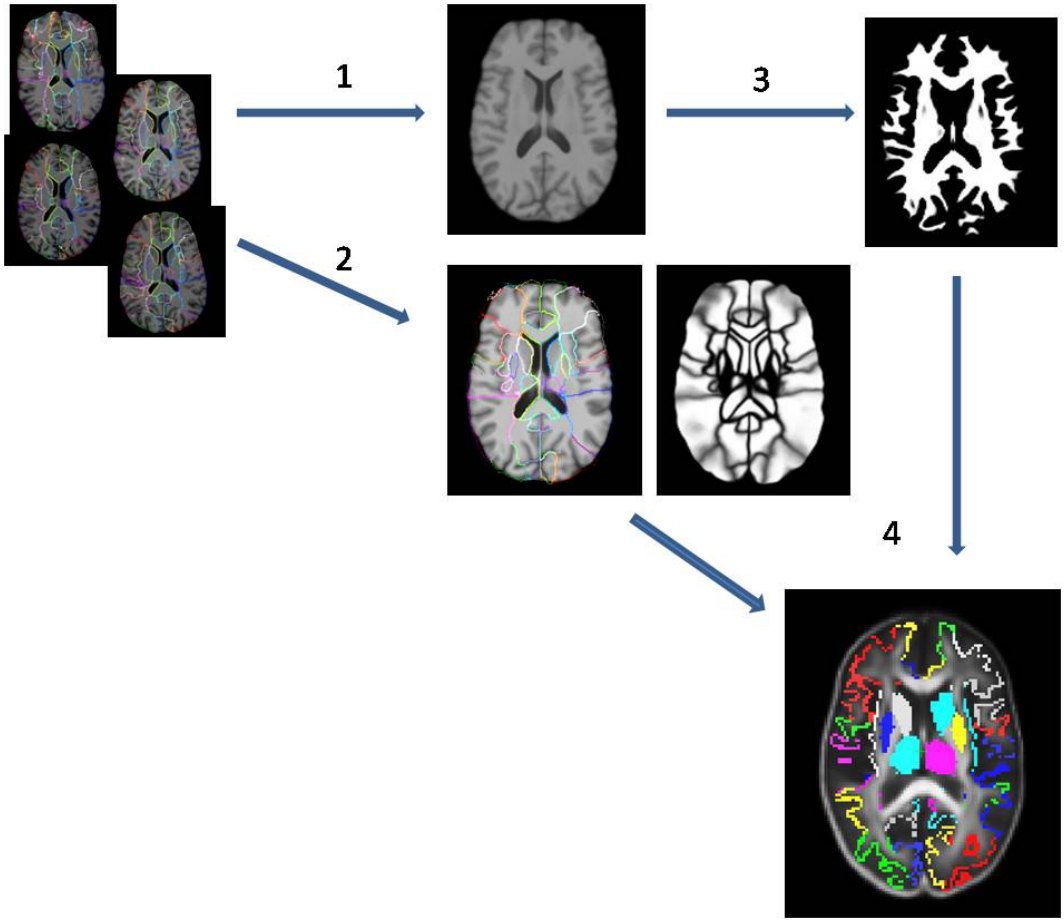


Figure 7.2: Probabilistic region of interest segmentation: 1) A T1 average image (atlas) is created by transforming all subjects' T1 images to a reference image and then, performing spatial normalisation. 2) The labels from all subjects are propagated to the T1 atlas and fused. These are associated with probabilities, calculated from the proportion of subjects that agree on the most popular label. 3) A white matter segmentation is also calculated from the T1 average image using SMP [8]. 4) Meanwhile, a FA average image is also estimated (see section 7.5.4). 5) Finally, labels are propagated from the T1 atlas to the FA atlas, and fused with the tissue segmentation to retain ROIs only at the WM-GM boundary

7.5.2 Re-orientation

Local affine transformations are extracted from the non-rigid transformation by calculating the Jacobian of the transformation at each point as described in section 3.4. Then spatial transformations of all samples on the principal fibre direction are performed using the finite strain (FS) approach for tensor reorientation suggested by Alexander et al [5, 55]. This aims to decompose affine transformations ($\mathbf{F}$) into pure deformation ($\mathbf{U}$) and rotational ($\mathbf{R}$) components:

$$\mathbf{F} = \mathbf{UR} \ , \quad \mathbf{R} = \mathbf{F}(\mathbf{F^T F})^{-\frac{1}{2}} \tag{7.1}$$

The approach is favoured over a preservation of principal directions (PPD) [5, 55] approach. This is because it ensures that the posterior distributions at each voxel are not modified by the transformation. In contrast, PPD re-orientation (which takes into account re-orientation due to shears and stretches) varies depending on the original orientation of the structure [5, 55], which may alter the representation of uncertainty. Interpolation is applied using a nearest neighbour approach as it is also not clear how to average the distributions.

7.5.3 Sampling

Once spatial alignment and reorientation of all samples is complete, each voxel $\nu_{\theta,\phi}$ in the template is represented by samples from joint distributions on θ and ϕ from all (N) subjects in the population:

$$\nu_{\theta,\phi} = \begin{bmatrix} P_{11}(\theta,\phi) \\ P_{12}(\theta,\phi) \\ \vdots \\ P_{NS}(\theta,\phi) \end{bmatrix}$$

where S is the total number of samples taken from each distribution.

It is possible to recover a representative distribution for the entire population by drawing samples evenly from $\nu_{\theta,\phi}$. This is achieved by sampling from the inverse of the discrete joint cumulative distribution function (JCDF) formed from these samples. A large number of samples per voxel are drawn in order to ensure a fair representation of the entire population.

7.5.4 Choice of registration approach

This framework follows the approach of many diffusion registration studies by estimating tensor re-orientation after the images have been spatially aligned. Most studies perform registration of images derived from diffusion data, such as FA. This is because diffusion data is sensitive to white matter structure and orientation. As such, approaches can be optimised to take into account white matter morphology [62, 197], or be constrained to prevent deformation from one fibre structure to another (as for TBSS). Nevertheless, registration of structural images such as T1 or T2 is aided by higher resolution and increased structural detail. In addition, T1 and T2 registration is based on alignment

of sulci and gyri which are thought to correlate with white matter connectivity [71]. Therefore, this section trials three registration approaches. These include registration of FA images using two different parametrisations, and an approach based on T1 registration.

The aim of this chapter is to use an atlas based approach to study ageing using the same population described in the previous chapter. Therefore. the three approaches were evaluated on 20 subjects aged between 20 and 68 (median 29) in order to test the robustness of the approach over the full age range. It was found that the high resolution of the T1 images meant that the number of samples saved lead to prohibitive memory requirements. Therefore these were resampled to a lower resolution.

Optimal performance is constrained by the choice of re-orientation strategy which restricts the amount of deformation. It was found that non rigid registration [144] of T1 images enabled optimal alignment of image structures. However, this proved to be unsuitable for registration of the diffusion images, as it first requires transformation of the diffusion data back to each subject's native T1 image. This requires the transformation to be invertible, which requires the registrations to be re-run with additional regularisation ($\lambda_s > 0$) to ensure that the transformations are smooth. Furthermore, re-orientation using the FS strategy, with simple nearest neighbour interpolation, is unable to compensate for large deformations between each subject's T1 image and the reference. This leads to greatly increased uncertainty at all voxels within the atlas. Therefore T1 registration was considered unsuitable for this approach.

FA registration was trialled using two approaches. The first uses a parameterisation proposed for TBSS. This has only one resolution level, control point spacing of 20mm and uses cross correlation as a similarity metric. The purpose of this is to align central structures without warping the white matter. It is not concerned with alignment of minor white matter pathways. Therefore, we also trial a second parameterisation. This starts with the same parameters but decreases control point spacing over an additional resolution level. It is hoped that this will strike a balance between misalignment, and not deforming structures so far that they warp, or deform further than FS re-orientation can cope with.

7.5.5 Validation

Several tracts were propagated in each atlas, these included: a) thalamus to motor cortex, b) thalamus to sensory cortex, c) thalamus to superior frontal cortex, d) brainstem to motor cortex. These were

traced bilaterally. In addition tracts were propagated across the splenium and genu of the corpus callosum. In each instance, atlases were generated by taking 1000 samples (all possible) and propagating 100000 streamlines per seed voxel. Voxels containing less than 10% of the maximum samples per tract were thresholded. Tractography in each subject's native space was also thresholded at this level.

The three atlases were assessed by comparing the mean shortest distance between these tracks propagated in the atlases, and the same tracts propagated in each subject's native space. This follows the validation approach used in Zhang et. al. [57, 197]. However, in this instance, ROIs were not constrained to be consistent. Instead, each ROI was selected using the segmentation procedure detailed in section 7.4.2. This enables correspondences to be drawn with the expected trajectory of tracts, generated by the previous framework.

Mean shortest distance was assessed by comparing all voxels on each tract bundle, rather than comparing distances between individual streamlines. This is because individual streamlines are not saved by this approach. Rather, a single intensity image is saved representing the distribution of all pathways. Therefore, the mean shortest distance (d_s) between two fibre bundles $\mathcal{F}$ and $\mathcal{G}$ is assessed as:

$$d_s = \frac{1}{|\mathcal{F}| + |\mathcal{G}|}\left(\sum_{v_F \in \mathcal{F}} \min d(v_F, \mathcal{G}) + \sum_{v_G \in \mathcal{G}} \min d(v_G, \mathcal{F})\right) \tag{7.2}$$

where $|\mathcal{F}|$ and $|\mathcal{G}|$ are the total number of voxels in $\mathcal{F}$ and $\mathcal{G}$ and $\min d(v_F, \mathcal{G})$ is the minimum distance between each voxel in fibre $\mathcal{F}$ and any point on fibre $\mathcal{G}$.

The results in figure 7.3 a show mean shortest distances for the tract between the right thalamus to the right sensory cortex for all three approaches. This demonstrates that, as expected, T1 registration is not appropriate under the reorientation scheme used and leads to unreliable tract propagation. In contrast, figures 7.3 b and c show that results for the two FA registration schemes are comparable to the results presented by Zhang et al [197]. This is especially encouraging considering that the age range of subjects is much greater, and ROIs are subject specific.

Some exceptions are seen for tracts between the thalamus and superior frontal cortex (figure 7.4). This may be due to the large span of the superior frontal gyrus, where it can be seen that the tract with greatest dissimilarity (purple) follows a completely different projection from the atlas (blue) and the majority of the remaining subjects (example shown in orange). Alternatively, such large population differences might suggest errors in the native space segmentation of these subjects.

The results averaged over all subjects and all tracts are shown in figures 7.5 a and b, respectively. Overall the less constrained FA approach outperforms the TBSS parameterisation with an average mean shortest distance over all subjects for all tracts of 3.54 ± 1.33mm compared to 3.90 ± 1.66mm. This parameterisation is therefore used throughout the rest of the chapter.

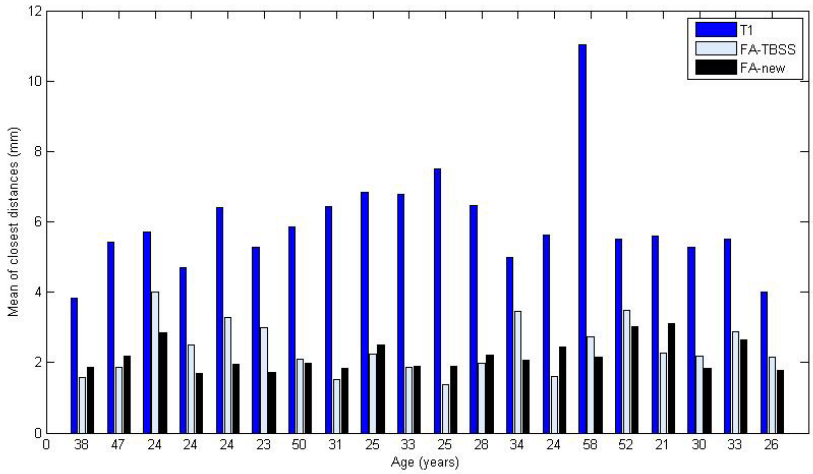
7.6 Fibre tract atlas construction

7.6.1 Spatial normalisation of FA images

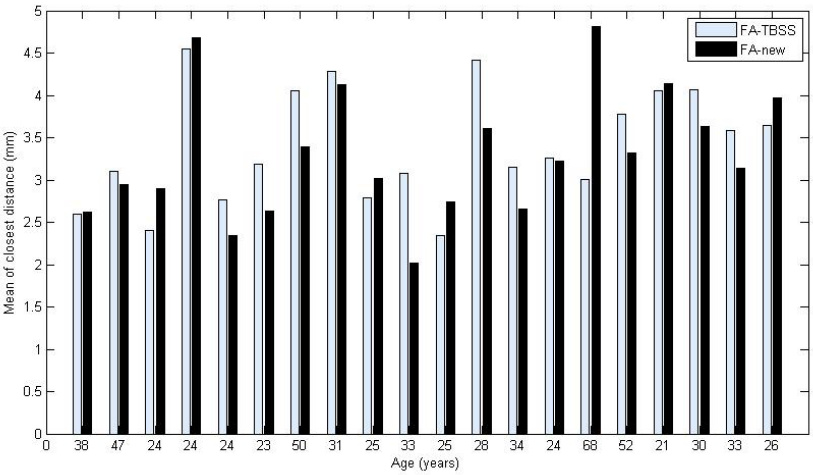
All subject's FA images are transformed to the co-ordinate frame of the reference subject using the registration parameterisation chosen in the validation step. These are then spatially normalised to form an average FA image for the population. Anatomical and tissue segmentations are then transformed to the FA average image, via non rigid registration of the T1 average image to the FA average image, using normalised mutual information as a similarity metric. These are used to constrain tractography.

7.6.2 Tractography

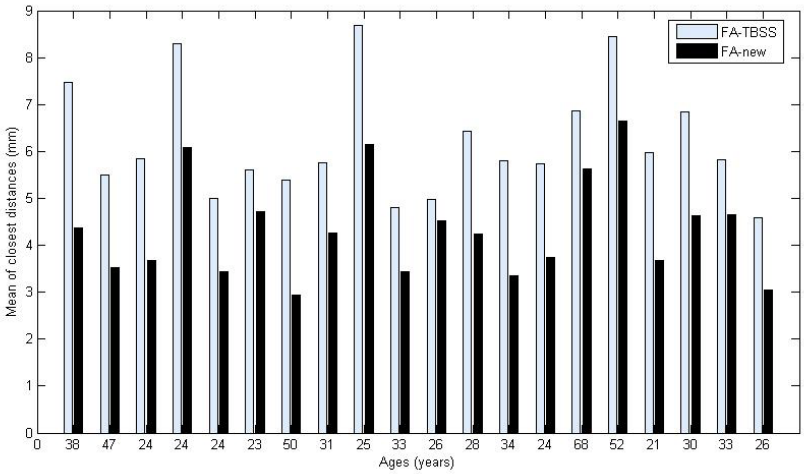
Probabilistic tractography is performed from all 77 seed regions propagated from the T1 atlas template. As before, tissue segmentations are used to constrain tractography from passing into (non ROI) grey matter or CSF, and tracts are terminated at the first target ROI they reach. The curvature threshold is set at 0.2, and correspondence with native space tractography is maintained by keeping the same ratio between the maximum number of streamlines per voxel and samples as proposed by the default in FSL (50 samples, 5000 streamlines). However, the actual number of streamlines propagated from each seed voxel is found by multiplying the maximum number of streamlines by the probability that all subjects agree on the label at that voxel (see section 7.4.2). In addition, tracts are thresholded to remove voxels which are sampled less the 10% of the maximum number of times for each tract. This removes some contribution from errant trajectories without thresholding tracts beyond the point which they resemble fibres.



(a)

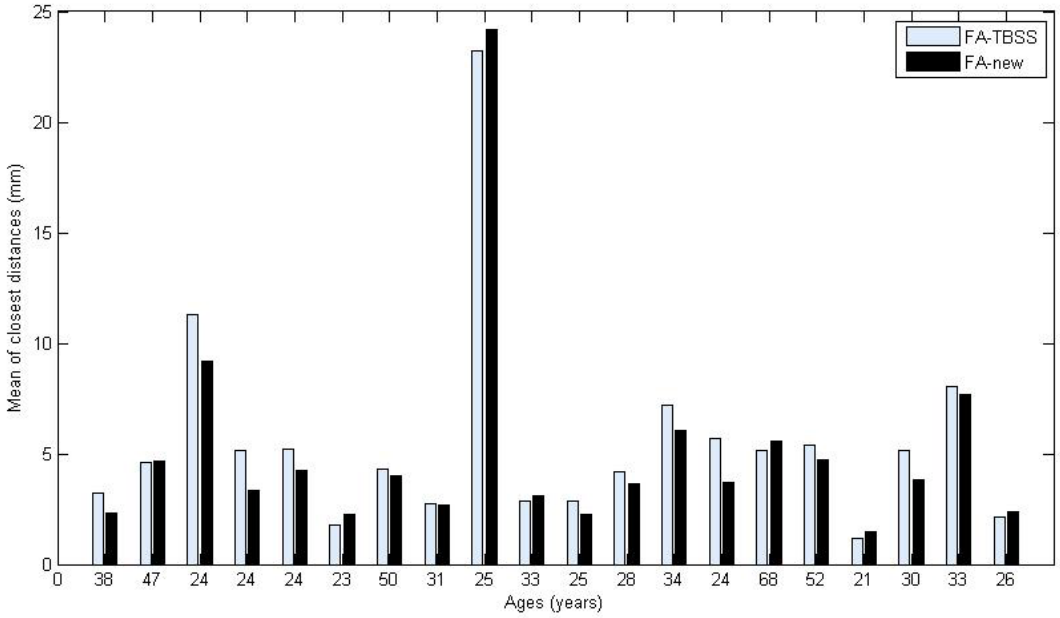


(b)

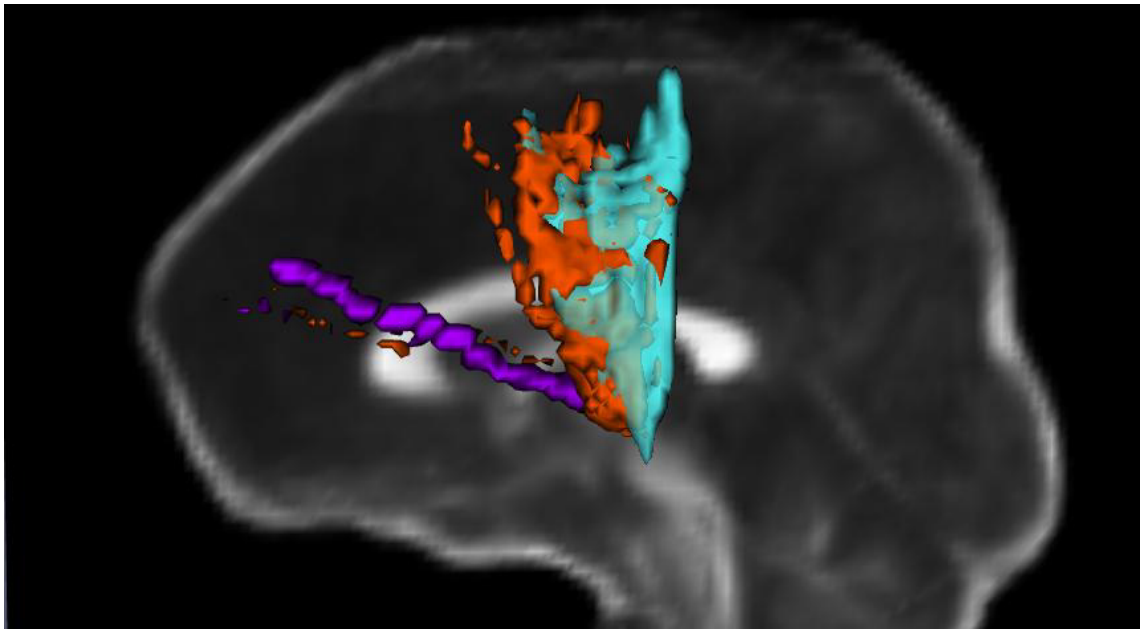


(c)

Figure 7.3: Atlas validation: comparison of tracts between a) right thalamus and right sensory cortex, b) contra-lateral superior parietal gyri and c) the brainstem and left motor cortex. The top graph compares atlases generated by T1 registration, and for both FA registration protocols. Comparing the mean of closest distances (equation 7.2) between atlas tracts and tracts propagated in native space (transformed to the atlas space) shows that the T1 atlas performs much worse than the FA atlases. In general the newly proposed FA protocol produces tracts that are closest to those propagated in native space.



(a)



(b)

Figure 7.4: a) Some large dissimilarities are seen for tracts propagated between the thalamus and superior frontal gyrus. b) examination of tract trajectories suggests some native subjects' tracts (purple) follow completely different trajectories to the atlas (blue). This corresponds to the subject whose mean of closest distances measure (d_s) is 24mm for this tract. The result may reflect inaccuracies in this subjects initial segmentation, which further emphasises our motivation for the common co-ordinate space approach.

7.7 Extraction of individual subject connectivity matrices

The aim of the approach is to study variability in tract integrity across populations. Therefore, following construction of the fibre tract atlas, anisotropy images from individual subjects are mapped

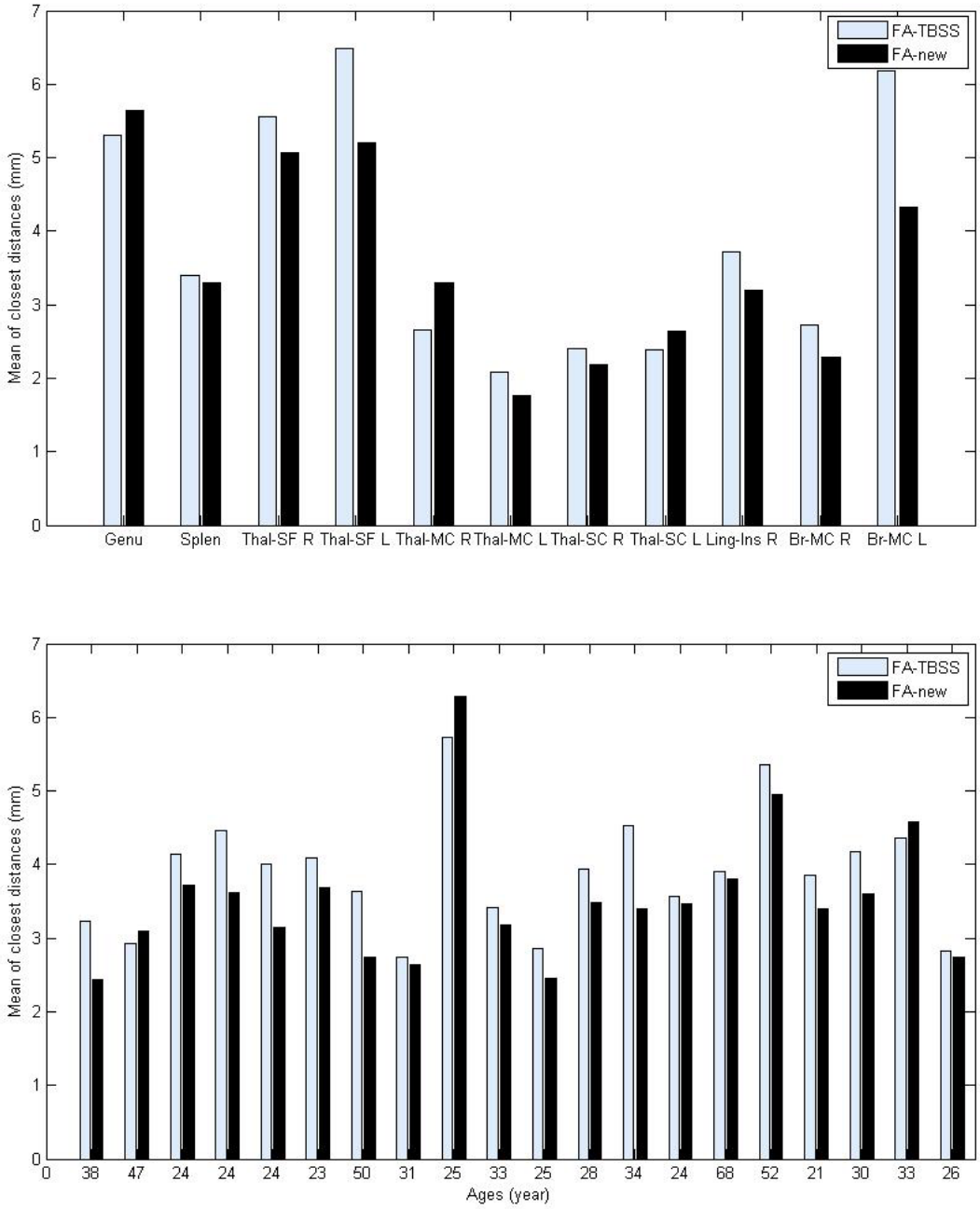


Figure 7.5: Atlas validation: mean results. Top: Comparisons of the mean of closest distances for each tract averaged over all subjects: splen=splenium; Thal=thalamus; SF=superior frontal, MC= motor cortex; SC=sensory cortex; Ling=lingual gyrus; Ins=Insula; Br=Brainstem. In general performance is good, with all tracts, on average, being less than 6mm from their native space counterparts. Bottom: Comparisons of the mean of closest distances for each subject averaged over all tracts. Most subjects, with the exception of the subject discussed in figure 7.4 show a mean similarity between atlas and native space tracts of < 5mm. The average over all subjects is 3.5mm (for the new protocol). This is comparable with the results of Zhang et al [197]

onto templates formed from the propagated tracts. These are then used to generate tracts for each subject. The advantage is that all subjects' tracts are aligned with the co-ordinate frame of the reference subject. This allows voxel-wise comparisons of anisotropy to be made across subjects.

In this study, we use FA images, rather than the direction dependent anisotropy measure presented in section 4.6. This is because individual streamline projections are not retained during tractography, and therefore the direction of propagation is unknown[2]. Connectivity matrices for each subject are estimated by calculating mean and maximum FA for each tract. Mean FA values are weighted to account for the number of times each voxel is sampled during tractography. However, the minimum is not used due to its potential sensitivity to registration error.

Registration errors also require restrictions to be put in place to ensure voxels are only compared where there is a significant chance that all subjects agree that the voxel belongs to the white matter. This is assessed by transforming all subject's white matter tissue maps to the template space. All voxels that have less than a 90% chance of belonging to all subjects' white matter are discarded from subsequent analysis. In addition, if tracts are reduced to less than 10 voxels after thresholding, these are removed. This reduces the impact that thresholding artefacts may have on the feature extraction.

7.8 Cross subject comparison

Connectivity matrices are built for each subject by assessing mean and maximum FA for all tracts. Then symmetry is enforced and these matrices are concatenated into feature vectors for subsequent machine learning analysis. This extends on the work of our conference papers (see Publications, [4, 5]), allowing classification of subjects and extraction of discriminating features. Analysis is performed using the GentleBoost, support vector machine (SVM) approach with 10 fold cross validation. This is because GentleBoost should be less sensitive than principal component analysis to any variances caused by misalignment of diffusion data.

The features selected from the machine learning analysis are examined for voxel-wise differences in FA using permutation testing. This tests whether FA values at each voxel in the tract template are significantly different across populations. In each instance 10000 permutations are performed. Significance is assessed using threshold free cluster enhancement [157] and results are corrected for multi comparisons unless otherwise stated.

[2]It would not be possible to save all streams as memory requirements would be too prohibitive

7.9 Experiments

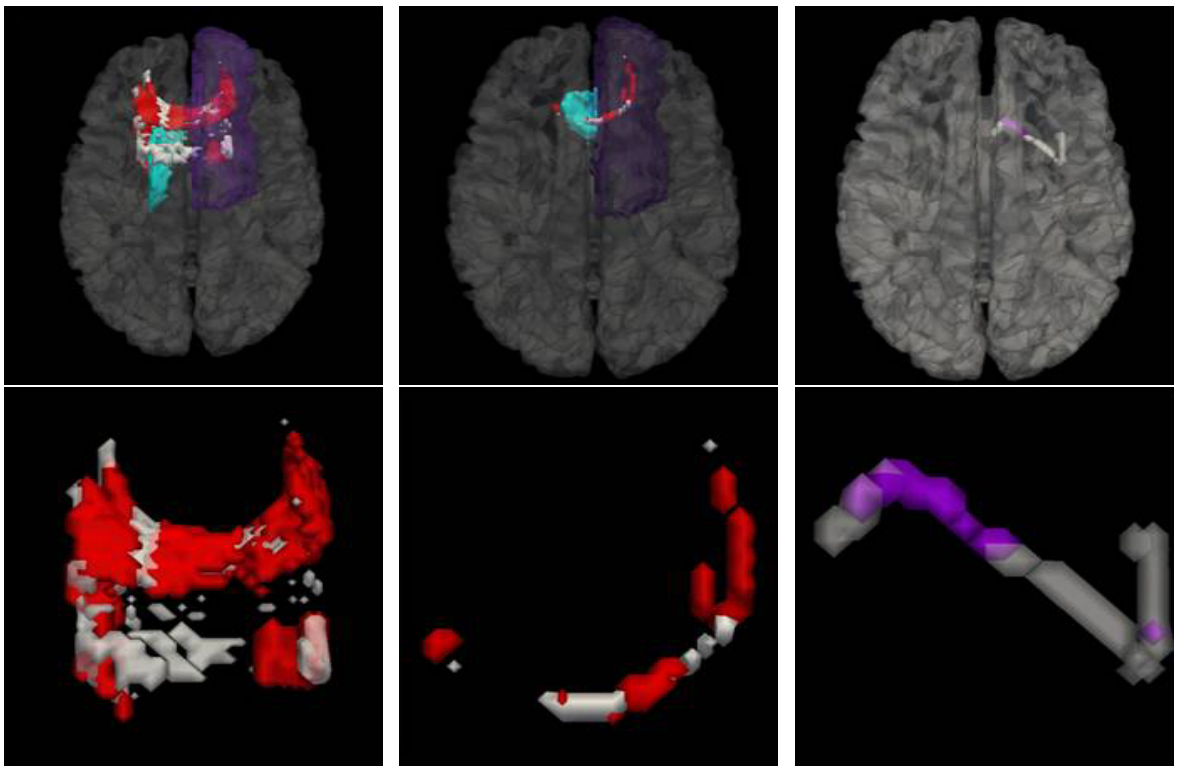
Experiments were performed to test the suitability of this atlas approach for detecting white matter differences across populations. The first experiment tests the approach on the same age population used in the previous chapter 6. The second study examines the effects of gender in a population of 117 adults under 60 (53 female).

7.9.1 Age study

An atlas was constructed from the total population of 173 subjects ranging from 20-86 years (87 female) taken from the IXI database and scanned using the protocol described in the previous chapter. Subjects from the whole age range were used, for the purpose of minimising the required deformation to the template. Subjects were eliminated from the original population if the initial registration to the reference image was judged unsuitable, due to warping or gross misalignment. This retained 169 subjects where the age range was retained. White matter and fibre tract atlases were obtained according to the protocols described above.

Following fibre atlas construction, 40 subjects aged 20-30 years and 50 subjects aged 60-86 years, were selected for the purpose of studying the effects of age. Mean and max FA connectivity matrices were extracted as proposed in section 7.7. These were then concatenated into feature vectors and analysed using GentleBoost feature extraction followed by SVM classification.

It was found that use of mean FA increased sensitivity of the feature extraction to very short tracts. This may reflect the fact that for long tracts, the presence of several different significant clusters may neutralise any tract-wide effect. Classification rates are slightly lower than for the native space approach: classification rate for maximum FA = $88.89 \pm 7.41\%$ (sensitivity = $87.50 \pm 1.68\%$, specificity = 89.50 ± 11.17) whereas, classification rate for mean = $88.89 \pm 11.71\%$ (sensitivity = $85.00 \pm 24.15\%$, specificity = 92.17 ± 13.01). This may reflect the imperfect registration, which is particularly sensitive to the changes in brain structure associated with ageing. Nevertheless, the significance of all features is improved. These are shown in tables 7.1 and 7.2 for each measure.



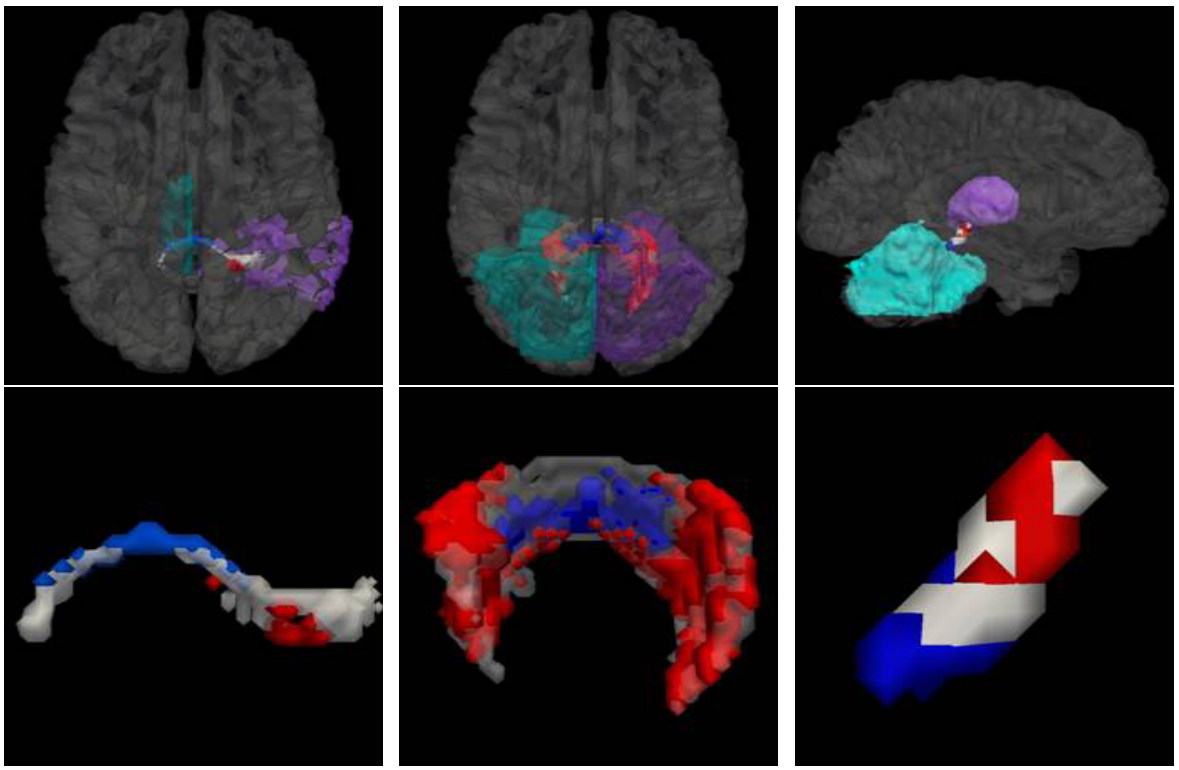
(a) (b) (c)

Figure 7.6: Results of permutation tests on tracts extracted by analysis of max FA. Features stronger in 20-30 years population (voxels shown to be significantly stronger in 20-30 years are highlighted in red): a) Caudate R & Superior Frontal L; b) Superior Frontal R & Subgenual Frontal L; c) Feature 1 between the left anterior temporal lobe to left caudate (purple) shown on the unthresholded tract (white). This suggests the results are sensitive to the degree of thresholding

Differences in maximum FA across the populations

Max FA generates comparable features to the native space study detailed in the previous chapter. The top features selected as stronger for the young population include several features traversing the genu of the corpus callosum. These include features numbers: 2,6,8, 10, 13 and 14. In addition, permutation tests of the tract between the right caudate and left superior frontal gyrus (feature 8: Figure 7.8 a), show as expected that FA is stronger for the young population (shown by red voxels) along most of the length of the tract (green).

Many of the remaining features selected for the 20-30 years population represent tracts within the frontal lobe, for example bilateral superior frontal gyrus to subgenual frontal gyrus features (Figure 7.6 b). There are also several connections between frontal regions and the basal ganglia (features 4, 7 and 8). Unfortunately, no features are selected by this approach between anterior and posterior brain regions. In addition, the approach may still be sensitive to misalignment. This is seen from



(a) (b) (c)

Figure 7.7: Features stronger in 60-90 years population (voxels shown to be significantly stronger in 60-90 years are highlighted in blue): a) Posterior Cingulate R & Posterior Temporal Lobe L b) Superior Parietal L & Superior Parietal R c) Cerebellum R & Thalamus R

selection of a feature from the left anterior temporal lobe to left caudate, which is heavily thresholded on account of its uncertain alignment with white matter (shown in figure 7.6 in purple superimposed upon the original tract in white).

For the 60-90 year old population, several features through the splenium are selected as having higher maximum FA. These include feature numbers: 3 (figure 7.7 c), 9, 11, and 15 (figure 7.7 b, also selected by the native space approach, section 6.5). Permutation tests of: a) the contra-lateral superior parietal gyrus tract, and b) the tract between the left posterior temporal lobe and the right superior parietal gyrus, show that FA is relatively preserved within the centre of the splenium, whereas other parts of the tract are by contrast lower in this population. This demonstrates the advantage of using maximum FA to highlight local variations.

Table 7.1: Table of maximum FA GentleBoost features from age population in order of selection

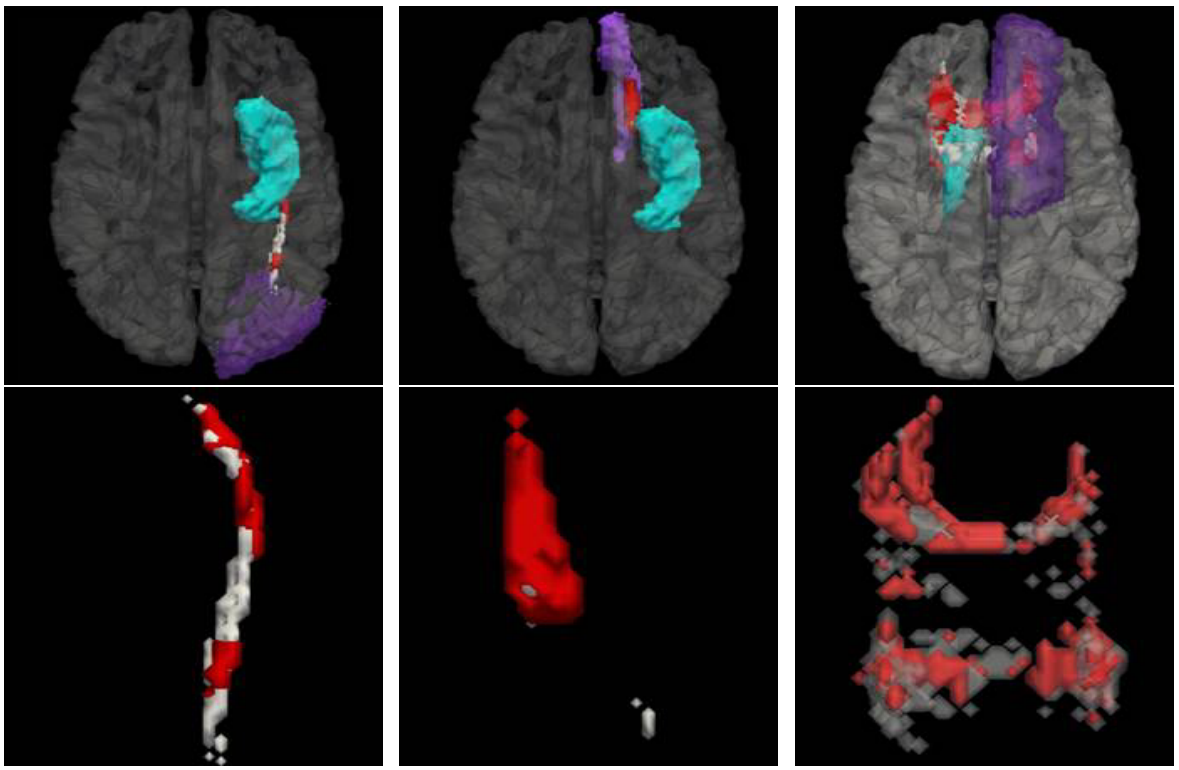
Feature	Region 1	Region 2	Significance	Greater in:
1	Anterior Temporal L	Caudate L	5.5210^{-11}	20-30 years
2	Superior Frontal R	Subgenual Frontal L	1.4210^{-11}	20-30 years
3	Posterior Cingulate R	Posterior Temporal L	1.4710^{-6}	60-90 years
4	Putamen L	Subgenual Frontal L	2.0710^{-4}	20-30 years
5	Cerebellum R	Thalamus R	1.5610^{-5}	60-90 years
6	Superior Frontal L	Subgenual Frontal R	8.7310^{-7}	20-30 years
7	Putamen L	Medial Orbital L	1.8910^{-10}	20-30 years
8	Caudate R	Superior Frontal L	6.9110^{-7}	20-30 years
9	Posterior Temporal L	Superior Parietal R	1.3010^{-3}	20-30 years
10	Straight Gyrus L	Subgenual Frontal R	1.1410^{-5}	20-30 years
11	Thalamus L	Superior Parietal R	4.3810^{-7}	60-90 years
12	Anterior Orbital L	Superior Frontal L	1.2210^{-9}	20-30 years
13	Medial Orbital R	Pre-Subgenual Frontal L	1.0710^{-7}	20-30 years
14	Posterior Orbital L	Subgenual Frontal L	1.4710^{-6}	20-30 years
15	Superior Parietal L	Superior Parietal R	1.9010^{-2}	60-90 years

Unlike the previous chapter, there are few features passing through the splenium of the corpus callosum to the thalamus or basal ganglia. In addition, whilst the feature selected between the right cerebellum and the thalamus reflect the pattern of previous results (figure 7.7 c), the results of the permutation tests show that significant increases in FA occur at different points along the tract for both populations.

Differences in mean FA across the populations

Table 7.2 shows the top ten features selected from the approach when mean FA is compared. In addition the 20th feature is shown. This is because it is the only feature selected by either approach which represents a tract between anterior and posterior regions of the brain. It is studied in more detail in Figure 7.8 a. Permutation testing demonstrates that FA is stronger for the 20-30 year age population predominantly towards more anterior regions of the tract.

The majority of the features selected from analysis of mean anisotropy represent frontal brain connections (Figure 7.8 b) or connections through the genu (Figure 7.8 c). Several features agree with the features extracted by GentleBoost for mean connectivity in the previous chapter (Table 6.5). These include contra-lateral superior frontal tracts (feature 5: Figure 7.8 c), tracts between the superior



(a) (b) (c)

Figure 7.8: Results of permutation tests on tracts extracted by analysis of mean FA: a) Insula L & Occipital L Caudate R; b) Insula L & Medial Orbital L; c) Superior Frontal L & Superior Frontal R

frontal gyrus and medial orbital gyrus through the genu (features 3 and 14, also picked up by PCA-MLDA), and the tract between the right superior frontal gyrus to the left pre-subgenual frontal gyrus (feature 6 table 6.2, also a PCA-MLDA feature). There are no significant features selected for the 60-90 year age group.

7.9.2 Gender study

For the gender study a separate atlas was constructed from a subsection of 117 (53 female: mean age 37.81 ± 11.89, and 64 men mean age 36.73 ± 11.12) subjects from the population that are under 60 years old. As before the atlas was constructed with 1000 samples per voxel and a maximum of 100000 streamlines were seeded. Voxels were thresholded from the tract image if their intensities reflect less than 10% of the maximum number of samples. Machine learning and permutation testing were performed as before.

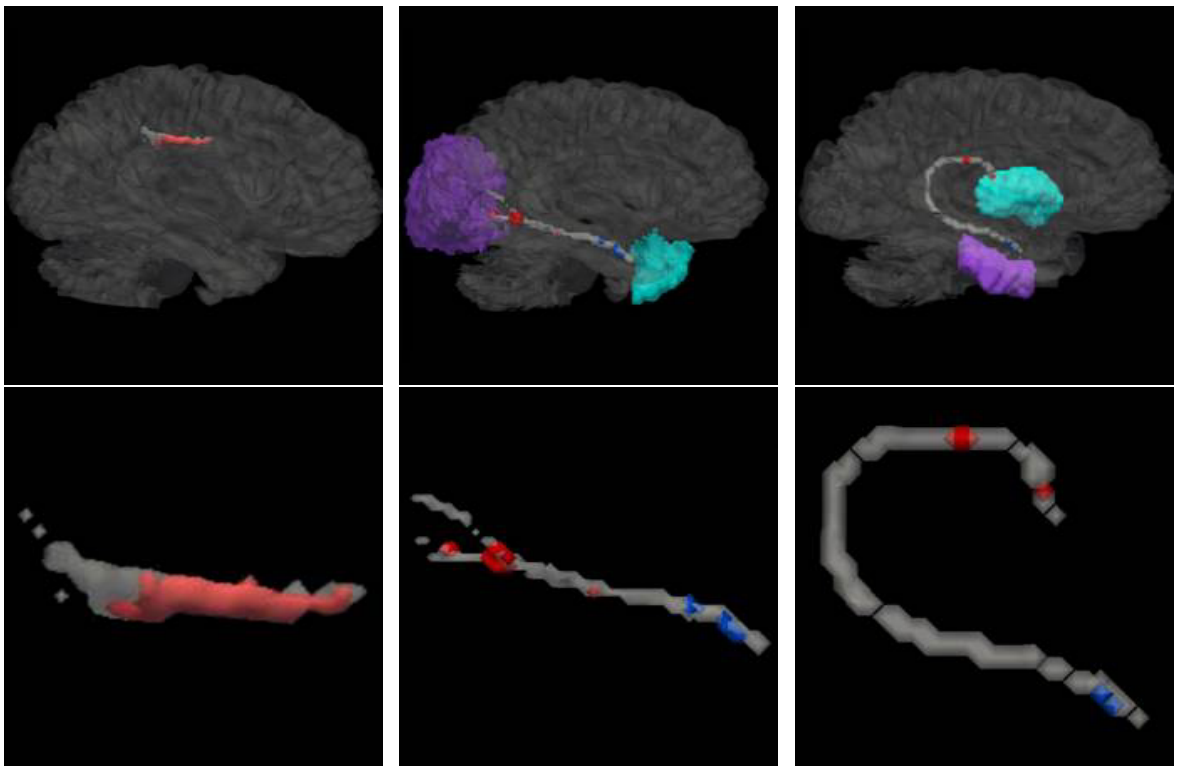
Although GentleBoost is not as sensitive as PCA to the major sources of variance in the data (in particular age) it is still unable to completely separate age effects. This may contribute to very low classification rates for both max and mean FA connectivity matrices: classification rate for mean FA

Table 7.2: Table of mean FA GentleBoost features from age population atlas, in order of selection

Feature	Region 1	Region 2	Significance	Greater in:
1	Anterior Orbital L	Superior Frontal L	9.3310^{-11}	20-30 years
2	Anterior Cingulate R	Middle Frontal	4.2110^{-14}	20-30 years
3	Superior Frontal R	Medial Orbital L	1.7310^{-9}	20-30 years
4	Insula L	Medial Orbital L	2.4110^{-13}	60-90 years
5	Cerebellum L	Substantia Nigra L	0.21	60-90 years
6	Superior Frontal L	Superior Frontal R	1.4310^{-13}	20-30 years
7	Superior Frontal L	Subgenual Frontal R	1.7710^{-10}	20-30 years
8	Anterior Cingulate R	Anterior Orbital R	1.6010^{-13}	20-30 years
9	Anterior Cingulate R	Pallidum R	1.9110^{-2}	20-30 years
10	Superior Frontal R	Pre-subgenual Frontal L	5.8010^{-9}	20-30 years
11	Straight Gyrus L	Subgenual Frontal R	1.1410^{-5}	20-30 years
12	Posterior Cingulate L	Superior Parietal L	2.0010^{-2}	20-30 years
13	Anterior Cingulate R	Putamen R	2.0610^{-2}	20-30 years
14	Superior Frontal L	Medial Orbital R	7.4810^{-8}	20-30 years
15	Straight Gyrus L	Anterior Orbital L	5.5910^{-5}	60-90 years
⋮	⋮	⋮	⋮	⋮
20	Insula L	Occipital L	0.14	20-30 years

$= 62.25 \pm 11.69\%$ (sensitivity for classifying women $= 65.15 \pm 17.04\%$, specificity $= 61.7 \pm 16.90\%$) classification rate for maximum FA $= 60.05 \pm 10.06\%$ (sensitivity $= 65.59 \pm 20.86\%$, specificity $= 54.39 \pm 16.90\%$). Nevertheless, attempts to remove age sensitivity by rotating the data to a basis which specifically reflects gender variance using partial least squares regression were also unsuccessful, which suggests that there may be no consistent pattern of differences between genders. This view is supported by other studies which have also found no significant white matter differences across genders during healthy ageing [171].

Even though the results suggest that there may be no consistent pattern of results across genders this does not rule out the possibility that some individual features may be significantly different. Therefore, to increase sensitivity to these differences, a paired t-test of all connections within the brain was performed. The most significant connections indicated from this approach using max and mean FA are shown in tables 7.3 and 7.4. The results are supported by voxel-wise permutation tests on the tracts corrected for the confound effect of age.



(a) (b) (c)

Figure 7.9: Results of permutation tests on tracts extracted by analysis of max FA for gender study (voxels shown to be significantly stronger in men are highlighted in red, and in women are highlighted in blue). FA within the cingulate is significantly higher in men a) Anterior Cingulate L & Superior Parietal L. However many of the features do not retain significance after correction for the family-wise error rate, for example b) Anterior Temporal Lobe Medial L & Occipital Lobe L ($p < 0.05$ uncorrected); c) Lateral Occipito-Temporal L & Putamen L ($p < 0.05$ uncorrected)

Differences in maximum FA across the populations

We found, as for the results of the age study, that the features selected from maximum FA were more sensitive to long tracts. These predict significant differences in maximum FA for several tracts propagated from the left anterior temporal lobe (Features 1,9), and lateral occipitotemporal lobe (Features, 5 and 7). This is supported by white matter VBM studies [80], which predict significant gender differences in WM within the left anterior temporal lobe and bilateral deep temporal regions. Unfortunately, the results are not supported by permutation tests, which show no areas of significant difference in FA after correction for multiple comparisons. Therefore, figure 7.9 b and c show uncorrected p-values.

Other results are supported by permutation tests. In particular, permutation tests of tracts within the cingulate do show large clusters of significantly different FA that remain after correction for multiple comparisons. This reflects several connections selected from the cingulate (Features 3,4,6,8) and

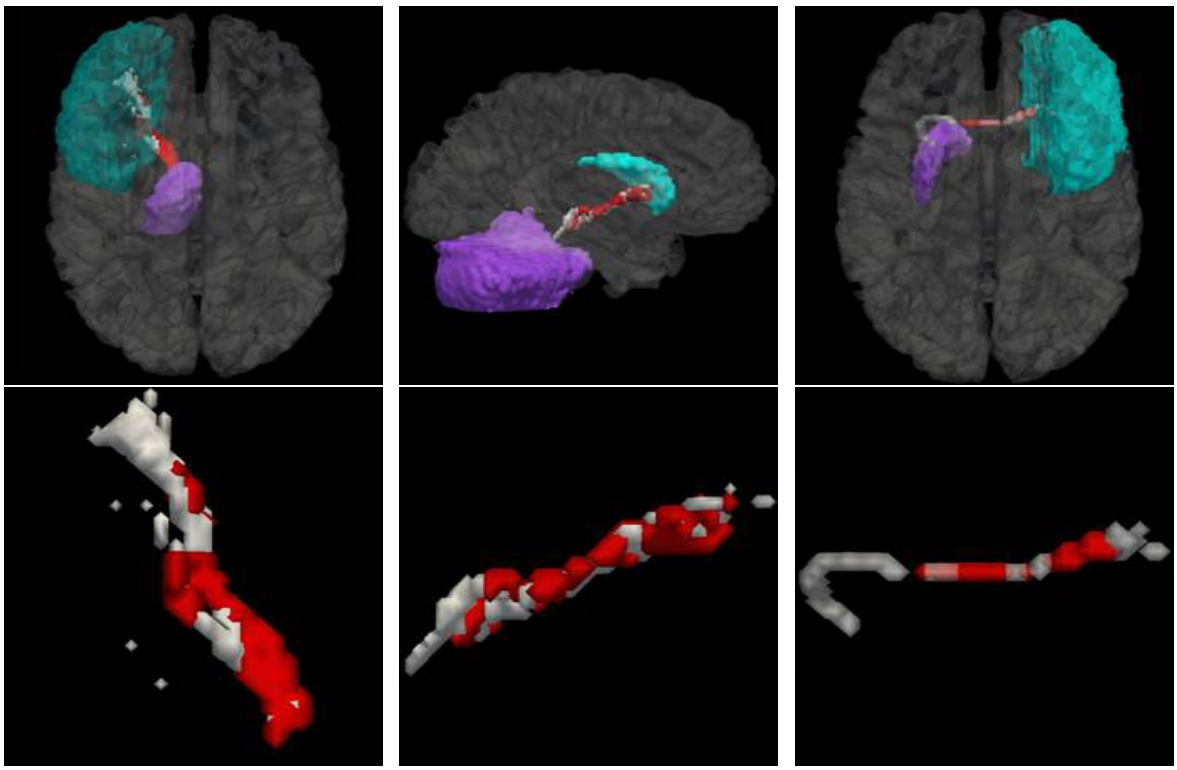
Table 7.3: Table of results of t-tests on maximum FA for the gender population, in order of significance

Feature	Region 1	Region 2	Significance	Greater in:
1	Cerebellum L	Pallidum L	2.0010^{-4}	Male
2	Anterior Temporal L	Thalamus L	4.0010^{-4}	Male
3	Anterior Cingulate L	Postcentral L	5.0010^{-4}	Male
4	Posterior Cingulate L	Postcentral L	5.0010^{-4}	Male
5	Lateral Occipito-Temporal L	Putamen L	5.0010^{-4}	Male
6	Anterior Cingulate L	Postcentral L	7.0010^{-4}	Male
7	Lateral Occipito-Temporal L	Thalamus L	7.0010^{-4}	Male
8	Anterior Cingulate L	Superior Parietal L	8.0010^{-4}	Male
9	Anterior Temporal Lobe L	Occipital Lobe L	1.0010^{-3}	Male
10	Postcentral R	Posterior Orbital R	1.1010^{-3}	Male
11	Caudate L	Pallidum L	1.1010^{-3}	Male
12	Thalamus R	Lateral Orbital R	1.1510^{-3}	Male
13	Superior Frontal L	Superior Parietal L	2.110^{-3}	Male
14	Superior Temporal (post.) R	Insula R	2.410^{-3}	Male
15	Insula L	Straight Gyrus R	8.510^{-2}	Male

the anterior limb of the internal capsule, which project between the caudate and putamen (features 1,11,12). In addition, Hsu et al [80] also report significant differences in FA within the cingulate, and left side anterior limb of the internal capsule.

Differences in mean FA across the populations

Results from the analysis of mean FA also reflect many connections that pass through the left anterior limb of the internal capsule. These include features 3, 4, 7, 14 and 15. In addition, permutation tests on the tract between the cerebellum and the caudate (figure 7.10 b) demonstrate that these reflect significantly higher FA for men throughout the length of the tract. This pattern is also reflected on the right side for a tract between the middle frontal gyrus and thalamus (figure 7.10 a). In addition, several tracts featuring the cingulate and the body of the corpus callosum (figure 7.10 c) are also selected. This has also been shown to be significantly stronger in men by Oh et al. [124] in a tractography guided statistical approach. No features are found to be stronger in women.



(a) (b) (c)

Figure 7.10: Results of permutation tests on tracts extracted by analysis of mean FA for gender study: a) Middle Frontal R & Thalamus R; b) Cerebellum L & Caudate L. FA across the body of the corpus callosum is also higher in men c) Middle Frontal L & Putamen R

Table 7.4: Table of results of t-tests on mean FA for the gender population, in order of significance

Feature	Region 1	Region 2	Significance	Greater in:
1	Middle Frontal R	Thalamus R	$< 1.0010^{-5}$	Male
2	Caudate R	Thalamus R	$< 1.0010^{-5}$	Male
3	Pallidum L	Subgenual Frontal L	$< 1.0010^{-5}$	Male
4	Cerebellum L	Caudate L	1.0010^{-4}	Male
5	Middle Frontal L	Inferior Frontal R	1.0010^{-4}	Male
6	Caudate R	Inferior Frontal R	1.0010^{-4}	Male
7	Putamen L	Subgenual Frontal L	1.0010^{-4}	Male
8	Anterior Cingulate R	Posterior Cingulate R	4.0010^{-4}	Male
9	Middle Frontal L	Putamen R	4.0010^{-4}	Male
10	Anterior Cingulate R	Inferior Frontal R	5.0010^{-4}	Male
11	Middle Frontal L	Caudate R	7.0010^{-4}	Male
12	Middle Frontal R	Caudate R	8.0010^{-4}	Male
13	Thalamus R	Inferior Frontal R	8.0010^{-4}	Male
14	Caudate L	Pallidum L	9.0010^{-4}	Male
15	Brainstem	Caudate L	1.110^{-3}	Male

7.10 Summary

This chapter has presented a probabilistic atlas based approach for studying differences in whole brain connectivity across populations of subjects. The approach uses a probabilistic white matter atlas constructed by transformation, re-alignment and sampling from posterior distributions on the principal direction of diffusion across a population of subjects. This has been used to generate a fibre tract atlas, where samples are taken evenly from the posterior distribution for the population, allowing propagation of tracts through areas of known fibre crossing and into regions of grey matter such as the thalamus. Tractography is performed only once, and therefore the tracts propagated represent a consensus across all subjects. The approach has been used successfully to identify differences in whole brain tractography for age and gender population. In addition, this has been used to investigate voxel-wise differences in FA along the full length of tracts.

The results for the age population study show comparable results to that of the previous chapter, inasmuch as whilst classification rates are slightly reduced (maximum classification rate = 88.89 ± 7.41) the results are not biased by variability in the success rate of tract propagation across the population. This leads to higher significance for the selected features for both mean and max FA.

Several features are picked up by both approaches including contra-lateral superior frontal tracts, and tracts between the superior frontal and medial orbital gyrus. In addition, all features agree with the pattern of results from the previous chapter and with the literature [38, 60, 89, 172]. This suggests that diffusion anisotropy is higher in the 20-30 year population especially within the frontal lobe and that this is graduated between frontal and posterior regions of the brain. In contrast, tracts within posterior regions (and specifically those that project through the splenium) are relatively preserved.

By contrast machine learning analysis of max and mean FA connectivity matrices was not able to find sufficient patterns of differences to allow successful classification of subjects by gender. This may be due to the strong confounding affect of age on the dataset which may mask any specific gender differences. Alternatively, previous studies have suggested that there is no consistent pattern of differences between genders [171]. Nevertheless, the results of subsequent paired t-tests do suggest that there may be some significant differences retained between individual connections. These agree with other VBM studies of diffusion which report significant differences within the cingulate, left anterior temporal lobe, and left anterior limb of the internal capsule [80]. In addition, Oh et. al. [124] compare diffusion properties along tracts within the corpus callosum and report significantly increased

FA throughout its body for men. This is also supported by these results.

There are advantages to this approach. First, by transforming samples from the posterior distribution on the principle diffusion direction rather than whole tensors, the atlas retains estimates of the uncertainty of the model fit at each voxel. This allows tracts to be propagated through areas of fibre crossing and into the grey matter. In addition, transforming samples in this way allows information regarding the population distribution at each voxel to be retained. This may allow for probabilistic comparisons of different tract morphologies. This is shown in figure 7.4, where it is shown that one subjects' tract follows a completely different trajectory to that of the atlas and the majority of the population.

Performing tractography once in a common template space, and projecting FA values from subjects' native image onto the propagated tracts, also allows the construction of connectivity matrices which are consistent across the population. This removes the influence of segmentation variability and reduces the impact of errors during the propagation of labels, allowing comparisons to be made between measurements of diffusion anisotropy alone. In addition, spatial normalisation allows voxel-wise comparisons of the properties of diffusion to be made along the full length of the tract using permutation testing. This enables detection of graduated differences in diffusion (for example between anterior and posterior regions of the brain during ageing: figure 7.8 a) or significant differences for both populations within the same tract (figure 7.7 b, and figures 7.9 b & c).

There are some limitations. Specifically, the choice of re-orientation strategy limits the amount of deformation that can be used. This means that alignment away from major white matter bundles is unreliable and FA values at these points are inconsistent. This requires a strong threshold to be put in place restricting comparisons to be made only where greater than 90% of the population agree the tract is propagated through white matter. The approach is also biased by the choice of reference subject. This encourages tracts within the atlas to be propagated along the same trajectory as the template subject. However, unbiased approaches such as those proposed by Joshi et al [26, 87] would take a significant amount of time to optimise on large data sets. In addition, such approaches would be reliant on optimal alignment. This would require re-addressing of the re-alignment scheme here.

In conclusion, we have proposed a common co-ordinate space approach for studying population differences in whole-brain tractography. This is independent of inter-subject variations in region of interest segmentation, which increases our confidence that the results of the feature extraction reflect changes

in tissue microstructure. This is supported by increased significance of the features compared to our native space approach (chapter 6). Nevertheless, the approach is sensitive to errors in alignment across subjects. This is reflected by a reduction is classification rates.

Future work will incorporate ODFs into the scheme by fitting B-splines to the centre of atlas tracts and using the probabilistic interpolation scheme used in tractography to fit subject data. This will offer advantages over the current approach, which compares only scalar values of FA, as it will incorporate estimates of anisotropy along the direction of projection of each tract. In addition, due of the limited number of diffusion directions for this data set, a single fibre model of diffusion was fit at each voxel. However, the proposed approach may be extended to multiple fibre data because the underlying model, used by the probabilistic tractography approach, saves each fibre direction separately. This would allow them to be treated independently provided that different fibre populations can be correctly grouped across all subjects [145]. Finally work will be conducted to improve alignment, by finding a way to incorporate PPD reorientation. Our long term aim is to use this approach on in clinical populations, specifically preterm infants.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

This thesis has proposed new techniques for studying the population variability in the brain structure through analysis of brain connectivity. In particular, new approaches for comparing whole-brain tractography data have been presented. This allows subjects to be classified into groups, and enables specific differences in the connective microstructure to be identified via comparisons of tract anisotropy.

The analysis uses a new framework for extracting whole-brain tractography from clinical data. This delineates anatomical regions in the brain using an atlas-based segmentation approach, and uses a widely available probabilistic tractography algorithm to assess connectivity between regions. It also proposes a new term for the estimation of the mean anisotropy along the length of each probabilistic tract. This is inspired by the graph-based tracking approach of Iturria-Medina et al. [83], and allows comparisons to be drawn with properties that reflect the integrity of tissue microstructure.

Currently, the most popular techniques for analysing whole-brain connectivity data are based on graph theory. Several studies have shown that brain networks estimated from functional and structural imaging data have displayed small world properties [1, 43, 59, 67, 71, 148, 158, 185], indicative of efficient organisation and complexity [161]. Some studies have also shown that brain connectivity displays specific clustering or modular structure [65, 107]. These techniques have been used to compare and contrast different populations. However, many studies have been restricted to the analysis of binary graph structures. In addition, small world graph comparisons of different populations have not been able to link differences to changes in the microstructure.

In contrast, this thesis has proposed an approach based on spectral graph clustering of weighted networks. This has been used to analyse the clustering properties of averaged brain networks across two age populations. The findings have been linked to anatomical changes reported in the literature. In addition, an entirely different approach based on machine learning has also been proposed. This has been tested using two frameworks: a principal component, linear discriminant analysis approach, as well as a GentleBoost, support vector machine approach. The specific advantage of this technique is that subjects can both be classified into groups, and that differences in tract structure can be located. The approach has been used to successfully investigate changes in brain structure during ageing. We now intend to apply it to less well-understood clinical cohorts. The aim is that it will be able to highlight new tracts as potential bio-markers, or phenotypes of disease.

One limitation of the proposed framework is its sensitivity to inter-subject variability in the tractography results. This is contributed to by individual differences in anatomical segmentation and registration. Therefore the final contribution of this thesis has been to extend the approach to generate a probabilistic fibre tract atlas. Tractography is performed only once and fractional anisotropy estimates from each subject are projected onto each of the template tracts for comparison. This has ensured that tracts are compared on the basis of anisotropy estimates alone. The approach has been used to study age and gender. Work still needs to be done to incorporate the proposed measure of mean tract anisotropy into the atlas framework. This will form a key part of future work.

8.2 Future Work

The techniques proposed in this thesis have shown the potential for improving understanding of how brain connectivity varies across populations. Nevertheless there are still a number of outstanding questions, for example how to reduce the impact of false positives on the connectivity analysis, how to implement the proposed tract anisotropy measure into the atlas-based approach, and how to correlate these results with the findings of functional imaging studies.

8.2.1 Reduction of false positives

The impact of false positives and the issue of thresholding has been a key concern of this thesis. In the presented work, the choice was taken not to threshold individual tracts. This is on account of the

problems associated with choosing a consistent threshold across all tracts and all subjects. Instead, attempts were made to reduce the impact of false positives through use of tissue type information to constrain the tractography from passing across sulci. In addition, a population-based thresholding approach was introduced (section 6.4.2), which removed connections which appeared infrequently across the population. Nevertheless, the presence of false positives remains a limitation. In future this may be reduced by constraining the results of the tractography using a connectivity prior as proposed in Jbabdi et al [86], formed from histological or functional data. Alternatively, a population-based connectivity prior could be formed from thresholded tractography results obtained from the atlas, or tracts themselves might be constrained by a population based mask of tract projections [12].

8.2.2 Improving the atlas-based approach

At present the atlas-based approach is limited in two keys ways. First, the use of finite strain re-orientation limits the degree to which brains can be deformed during alignment. We believe improvements may be obtained by incorporating a preservation of principle directions approach. However, care will need to be taken to ensure that the re-orientation strategy does not change the representation of uncertainty at each voxel.

Second, since the projections of individual streamlines are not saved during tractography, it has not been possible to implement the ODF-based estimate of anisotropy, used in the native space approach. Instead, anisotropy was estimated from FA averaged over all voxels in the tract. In future, the ODF measure may be re-introduced by first fitting B-splines to the centre of each tract following the approach of Goodlett et al. [62]. Then ODFs may be fitted to these mean tracts using probabilistic interpolation.

8.2.3 Research into preterm brain development

In this thesis we have demonstrated that the proposed techniques are able to successfully detect differences in tissue microstructure by applying them to a phenomenon already well understood, ageing. However, a key test for these techniques will be how successfully they can be applied to clinical populations. In particular, we are interested in applying the techniques to preterm infant populations.

Preterm birth has long been associated with cerebral white matter injury and neurological impairment. Our aim is to identify genetic markers and neurological phenotypes indicative of hindered development. However, application of these techniques to neonates will require careful adjustment of the protocol. In particular, registration between neonatal brain images is more complex due to rapidly changing brain structure and the frequent presence of motion artefacts. This may be improved by use of an additional rigid registration step prior to affine and non-rigid registration. Nevertheless, a key question remains over how to initialise the tractography. Adult brain templates are not suited to anatomical segmentation of neonatal brain images as brain size and morphology are very different. One option may be to implement a voxel-wise approach. However, this will vastly increase the dimension of the feature vectors for machine learning analysis, which may consequently reduce the stability of the PCA-MLDA approach. An alternative is to derive new segmentations. These could be obtained from functional data [65] or by connectivity based parcellation [23, 138].

8.2.4 Connectivity based parcellation of the cortex

The underlying connective matrix of the brain underpins brain function and influences cortical folding. Therefore, equivalent functional regions can be identified across subjects through comparison of their connectivity profiles. This has been demonstrated by connectivity-based parcellation of the thalamus [23], where parcellations have been shown to correspond with histology. In addition, Roca et. al, [138] perform cortical segmentation via clustering of whole-brain connectivity matrices, and derive clusters which reflect recognisable functional sub-units [138].

Connectivity based parcellation may allow clusters of voxels to be extracted as input as seeds for tractography. However, one limitation is that it cannot consistently reproduce clusters across subjects. This issue may be irrelevant if parcellation and subsequent tractography is performed in an atlas space. Alternatively, it may be possible to incorporate atlas-based approaches in a combined label-fusion, connectivity based parcellation approach (see Publications, [6]). This would use atlas-based segmentations to initialise a probabilistic tracking step, from which voxels could subsequently be re-classified on the basis of their connectivity profiles. This might allow adult atlases to be used as a starting point for neonatal brain segmentation.

8.2.5 Drawing correspondences with functional imaging studies

Another key goal of neuroimaging is to better understand the structural correlates of functional connectivity. In Deligianni et al (see Publications, [7]) we used a predictive model based on PCA and canonical correlation analysis to infer functional connections from models of anatomical connectivity, where these were constructed using the framework proposed in Chapter 4. Similar collaborations are likely to play a central role in future research into the causes and outcomes of preterm birth.

8.3 Summary

This thesis has proposed a new direction for whole-brain connectivity analysis. It is clear that there are still a number of open questions and limitations to be overcome, before the technique can be reliably used to identify clinical biomarkers. Nevertheless, it is hoped that this technique can be used to improve understanding of how brain connections change during development and disease. The next step will be to implement the approach on a study of preterm brain development.

Appendix A

List of symbols and abbreviations

A.0.1 Abbreviations

Common abbreviations, in order of appearance:

Abbreviation	Meaning
MRI	Magnetic Resonance Imaging
GM	Grey matter
WM	White matter
CSF	Cerebral Spinal Fluid
voxel	volumetric pixel, 3D image element
ADC	Apparent Diffusion Coefficient
FA	Fractional Anistropy
DSI	Diffusion Spectrum Imaging
ODF	Orientation Distribution function
FFD	Free Form Deformations
SSD	Sum of Squared Differences
CC	Cross correlation
NMI	Normalised Mutual Information
ROI	Region of Interest
SPM	Statistical Parametric Mapping (software)
FSL	FMRIBs software library (software)
PCA	Principal component analysis
MLDA	Maximum uncertainty Linear Discriminant Analysis
SVM	Support Vector Machine

A.0.2 Symbols

Where symbols have more than one meaning their use will be made clear in the text. Common symbols, in order of appearance are:

Symbol	Meaning
$\mathbf{r}$ or $\mathbf{x}$	Position vector
$\mathbf{R}$	Relative displacement vector (Chapter 2 only) / Rotation
D	Diffusion coefficient
$\mathbf{D}$	Diffusion Tensor / Degree matrix (Chapter 5 only)
S	The magnetic resonance signal with diffusion weighting
S_0	The magnetic resonance signal with diffusion weighting
b	The b-value: relates to strength and separation of the appled magnetic field gradients
b_0	A non diffusion weighted image
$\mathbf{E}$	Matrix of eigenvectors $\mathbf{e}$
$\mathbf{\Lambda}$,	Matrix of eigenvalues λ
$\mathbf{e}$	Eigenvector
λ	Eigenvalue / path length ratio (Chapter 5 only)
(θ, ϕ)	Orientation of a sample from the distribution of diffusion direction at each voxel, in spherical polar co-ordinates
f or $f^{(\theta,\phi)}$	Volume fraction of a sample from the distribution of diffusion direction at each voxel, in spherical polar co-ordinates
ψ	Orientation distribution function (ODF)
$\mathbf{T}$	Transformation
$\mathbf{F}$	Affine transformation
$\mathbf{J}$	Jacobian of a transformation
$G = [N, E, W]$	Graph with nodes (N), edges (E) and weights (W)
$\mathbf{A}$, entries a_{ij}	Adjacency matrix
$\mathbf{W}$ entries w_{ij}	Weighted adjacency matrix
γ	Clustering ratio
σ	Small world ratio
$\mathbf{X}$	Data matrix
$y = \{-1, 1\}$	Labels

Appendix B

Measuring functional connectivity

Functional connectivity studies estimate correlated patterns of neuronal activity in the brain. In this thesis two types of functional imaging study are discussed. These are functional magnetic resonance imaging (fMRI) and electro/magneto-encephalography (EEG/MEG).

B.0.3 Functional MRI

Neurons require oxygen and glucose from the blood supply in order to fire. Functional MRI, uses this principle to estimate neuronal activity indirectly by measuring the blood oxygen level dependence (BOLD) signal [123]. This changes the magnetic resonance signal, specifically in terms of T2 relaxation. Therefore, by measuring the signal for each voxel, at regular time intervals, a time course for the BOLD signal can be recorded. Functional connectivity can then estimated by simply measuring the (partial) correlation between time signals at different positions in the brain [43, 148]. Alternatively, some approaches measure signal similarity over a range of frequencies using, for example, wavelet transformations [1].

B.0.4 Encephalography

MEG (and EEG) measure currents flowing in the dendrites of neurons during synaptic transmission. They therefore provide a more direct estimate of neuronal connectivity than fMRI, and typically have better temporal resolution. However, measurements are are restricted to discrete points on the surface of the brain.

Connectivity is measured by estimating the degree of coupling between signals. Two key estimates of signal similarity used in EEG/MEG analysis for brain connectivity include the synchronisation likelihood, which measures the correlation between two time signals, taking into account linear and non-linear interdependences [167], and phase lag index which is a measure of the asymmetry of the distribution of phase differences between two signals [165]

B.1 The resting state

Traditionally, functional imaging studies have attempted to record patterns of activity induced under particular tasks, then compare the results to activity during a control task. The difference between the two signals is used to identify points in the brain that are required for performance of that specific task. Nevertheless, the difference between the two signals, typically characterises less than 5% of the BOLD signal. Therefore, resting state imaging [28] aims to measure what happens when no external stimulus is presented. The approach has been shown to consistently identify certain clusters of brain regions that show correlated brain activity. Of these, the default mode network (consisting of the posterior cingulate cortex, precuneus, parietal cortex, medial temporal cortex and medial frontal cortex) has generated particular attention. This is because as it appears to be active during rest but predominantly inactive during task.

Functional connectivity of the brain at rest shows similar patterns of connectivity to anatomical brain networks. In addition, approaches based on wavelet analysis [1] or MEG [164] suggest that, whilst brains are 'small-world' over a range of frequency bands and thus levels of organisation, small world ratios are most salient for networks extracted from low frequency signals, such as those seen at rest. This suggests that resting state networks are most correlated with anatomical connectivity. This hypothesis is supported in a study by Honey et al [78], who replicated patterns of resting state function from low frequency dynamical models built from histological data from the macaque.

Publications

[1] E.C. Robinson, M. Valstar, A. Hammers, A. Ericson, A.D. Edwards, and D. Rueckert. Multivariate statistical analysis of whole brain structural networks obtained using probabilistic tractography. In *MICCAI*, pages 523–531. LNCS, 2008.

[2] E.C. Robinson, P. Aljabar, and D. Rueckert. Spectral clustering of brain networks. In *MICCAI Workshop: Diffusion modelling and the fibre cup*, 2009.

[3] E.C. Robinson, A. Hammers, A. Ericsson, A.D. Edwards, and D. Rueckert. Identifying population differences in whole-brain structural networks: a machine learning approach. *NeuroImage*, 50:910–919, 2010.

[4] E.C. Robinson, D. Rueckert, A. Hammers, and A.D. Edwards. Probabilistic white matter and fibre tract atlas construction. In *IEEE International Symposium on Biomedical Imaging*, 2010.

[5] E.C. Robinson, D. Rueckert, A. Hammers, and A.D. Edwards. A probabilistic white matter atlas approach to assessing age related changes in the brain. In *International Society for Magnetic Resonance Medicine*, 2010.

[6] E.C. Robinson, A.D. Edwards, and D. Rueckert. A combined label fusion and connectivity based parcellation approach for anatomical brain segmentation. In *submitted to MICCAI*, 2010.

[7] F. Deligianni, E.C. Robinson, C.F. Beckmann, D. Sharp, A.D. Edwards, and D. Rueckert. Inference of functional connectivity from structural brain connectivity. In *IEEE International Symposium on Biomedical Imaging*, 2010.

Bibliography

[1] Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. A resilient, low frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience 26(1)* (2006), 63–72.

[2] Alexander, A., Hasan, K., Lazar, M., Tsuruda, J., and Parker, D. Analysis of partial volume effects in diffusion-tensor MRI. *Magn. Reson. Med. 45* (2001), 770–780.

[3] Alexander, D., Barker, G., and Arridge, S. Detection and modelling of non-gaussian apparent diffusion coefficient profiles in human brain data. *Magn. Reson. Med. 48* (2002), 331–340.

[4] Alexander, D., Gee, J., and Bajcsy, R. Elastic matching of diffusion tensor MRI. *Comput. Vision Image Understanding - Special issue Anal. Volumetric Images 7(2)* (2000), 233–250.

[5] Alexander, D., Pierpaoli, C., Basser, P., and Gee, J. Spatial transformation of diffusion tensor magnetic resonance images. *IEEE Trans. Med. Imaging 20(11)* (2001), 1131–1139.

[6] Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., and Rueckert, D. Classifier selection strategies for label fusion using large atlas databases. In *Medical Imaging Computing and Computer assisted Intervention (MICCAI)* (2007), Lecture Notes in Computer Science, pp. 523–531.

[7] Anjari, M., Srinivasan, L., Allsop, J. M., Hajnal, J., Rutherford, M., Edwards, A., and Counsell, S. Diffusion tensor imaging with tract-based spatial statistics reveals local white matter abnormalities in preterm infants. *NeuroImage 35* (2007), 1021–1027.

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

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

[10] Assaf, Y., Freidlin, R., Rohde, G., and Basser, R. New modelling and experimental framework to characterise hindered and restricted water diffusion inbrain white matter. *Magn. Reson. Med. 52* (2004), 965–978.

[11] Bajcsy, R., and Kovacic, S. Multiresolution elastic matching. *Comput. Vision, Graphics and Image Process 46* (1989), 1–21.

[12] Ball, G., Boardman, J., Rueckert, D., Merchant, N., T.Arichi, Edwards, A., and Counsell, S. Population-mapped probabilistic tractography - a new tool for studying white matter architecture in neonates. In *Conference of Paediatric Academic Societies* (2010).

[13] Barnea-Goraly, N., Chang, K., Karchemskiy, A., Howe, M., and Reiss, A. Limbic and corpus callosum aberrations in adolescents with bipolar disorder: a tract-based statistics. *Biological Psychiatry 66(3)* (2009), 238 – 244.

[14] Bartzokis, G., Cummings, J., Sultzer, D., Henderson, V., Nuechterlein, K., and Mintz, J. White matter structural integrity in healthy ageing adults and patients with alzheimer's disease: a magnetic resonance imaging study. *Arch. Neurol 50* (2003), 393–398.

[15] Basser, P., Mattiello, J., and LeBihan, D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. 103* (1994), 247–254.

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

[17] Basser, P., and Pierpaoli, C. Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI. *J. Magn. Reson. B 111* (1996), 209–219.

[18] Bassett, D., Bullmore, E., Verchinski, B., Mattay, V., Weinberger, D., and Meyer-Lindenberg, A. Hierarchical organisation of human cortical networks in health and schizophrenia. *The Journal of Neuroscience 28(37)* (2008), 9239–9248–2381.

[19] Beaulieu, C., and Allen, P. Determinants of anisotropic water diffusion in nerves. *Magn Res Med 31* (1994), 394–400.

[20] Beg, M., Miller, M., Trouvé, A., and Younes, L. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision 61(2)* (2005), 139–157.

[21] Behrens, T. *MR diffusion tractography: methods and applications*. PhD, University of Oxford, 2004.

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

[23] Behrens, T., Johansen-berg, H., Woolrich, M., Smith, S., Wheeler-Kingshott, C., Boulby, P., Barker, G., Sillery, E., Skeehan, K., Cicarelli, O., Thompson, A., Brady, J., and Matthews, P. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci. 6(7)* (2003), 750–757.

[24] Behrens, T., Woolrich, M., Jenkinson, M., Johansen-berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., and Smith, S. Characterization and propagation of uncertainty in diffusion weighted MR imaging. *Magn. Reson. Med. 50* (2003), 1077–1088.

[25] Bergfield, K. L., Hanson, K. D., Chen, K., Teipel, S., Hampal, H., Ropoport, S., Moeller, J., and Alexander, G. Age-related networks of regional covariance in MRI grey matter: Reproducible multivariate patterns in healthy ageing. *NeuroImage 49* (2010), 1750–1759.

[26] Bhatia, K., Hajnal, J., Puri, B., Edwards, A., and Rueckert, D. Consistent group-wise non-rigid registration for atlas construction. In *IEEE International Symposium on Biomedical Imaging.* (2004), IEEE.

[27] Bihan, D. L. Molecular diffusion, tissue microdynamics and microstructure. *NMR Biomed 8* (1995), 375–386.

[28] Biswal, B. Functional connectivity in the motor cortex of the resting human brain using echo-plnar MRI. *Magn. Reson. Imag. 34* (1995), 537–541.

[29] Bjartmar, C., Hildebrand, C., and K. Loinder, K. Morphological heterogeneity of rat oligodendrocytes: electron microscopic studies on serial sections. *Glia 11* (1994), 235–244.

[30] Burges, C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2* (1998), 121–167.

[31] Cao, Y., Miller, M., Mori, S., Winslow, R., and Younes, L. Diffeomorphic matching of diffusion tensor images. In *In: Int. Conf. Comput. Vis. Pattern Recognit. Workshop Math. Methods. Biomed. Image Anal.* (2006).

[32] Chen, Z., He, Y., Rosa-Neto, P., Germann, J., and Evans, A. Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cerebral Cortex 18* (2008), 2374–2381.

[33] Cook, P., Zhang, H., Awate, S., and Gee, J. Atlas-guided probabilistic diffusion-tensor fiber tractography. In *IEEE International Symposium on Biomedical Imaging* (2008), pp. 951–955.

[34] Corouge, I., Fletcher, P., Joshi, S., Gilmore, J., and Gerig, G. Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. In *Medical Imaging Computing and Computer assisted Intervention (MICCAI)* (10 2005), Lecture Notes in Computer Science.

[35] Counsell, S., Dyet, L., Larkman, D., Nunes, R., Boardman, J., Allsop, J., Fitzpatrick, J., Srinivasan, L., Cowan, F., Hajnal, J., M.A.Rutherford, and Edwards, A. Thalamo-cortical connectivity in children born preterm mapped using probabilistic magnetic resonance tractography. *NeuroImage 34* (2007), 896–904.

[36] Cremers, D., Rousson, M., and Deriche, R. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *International Journal of Computer Vision 72(2)* (2007), 195–215.

[37] Crum, W., Tanner, C., and Hawkes, D. Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. *Phys. Med. Biol. 50* (2005), 5153–5174.

[38] Davies, S. W., Dennis, N., N.G.Buchler, White, L., Madden, D., and Cabeza, R. Assessing the effects of age on long white matter tracts using diffusion tensor tractography. *NeuroImage 46* (2009), 530–541.

[39] Descoteaux, M., Angelino, E., Fitzgibbons, S., and Deriche, R. Regularized, fast and robust analytical Q-ball imaging. *Magn. Reson. Med 58* (2007), 497–510.

[40] Dineen, R., Vilisaar, J., Hlinka, H., Bradshaw, C., Morgan, P., Constantinescu, C., and Auer, D. Disconnection as a mechanism for cognitive dysfunction in multiple sclerosis. *Brain 132* (2009), 239–249.

[41] Dyrby, T., Sogaard, L., Parker, G., Alexander, D., Lind, N., Baaré, W., Hay-Schmidt, A., Erikson, N., Pakkenberg, B., Paulson, O., and Jelsing, J. Validation of in vitro probabilistic tractography. *NeuroImage 37* (2007), 1267–1277.

[42] Einstein, A. *Investigations on the theory of Brownian motion.* Dover, New York, 1956.

[43] Equíluz, V., Chialvo, D., Cecchi, G., Baliki, M., and Apkarian, A. Scale-free brain functional networks. *PRL 94* (2005), 018102.

[44] Erdös, P., and Rényi, A. *Publ. Math. Debrecen 6* (1959), 290.

[45] Esposito, C., Kirkby, B., Horn, J. V., Ellmore, T., and Berman, K. Context-dependent, neural system-specific neurophysiological concomitants of ageing: mapping PET correlates during cognitive activation. *Brain 122(5)* (1999), 963–979.

[46] Evans, A., Collins, D., and Milner, B. An MRI-based stereotaxic atlas from 250 young normal subjects. *Soc. NeuroSci. Abstr. 18* (1992), 408.

[47] Evans, A., Kamber, M., Collins, D., and Macdonald, D. An MRI-based probabilistic atlas of neuroanatomy. In *Magnetic Resonance Scanning and Epilepsy*, S. Shorvon, D. Fish, F. Andermann, and G. Bydder, Eds., vol. 24 of *NATO ASI Ser, Ser A: Life Sci.* Plenum Press, Boston, 1994, pp. 263–274.

[48] Evans, A., Marrett, S., Neelin, P., Collins, D., Worsely, K., Dai, W., Milot, S., Meyer, E., and Bub, D. Anatomical mapping of functional activation in stereotactic coordinate space. *NeuroImage 1(1)* (1992), 43–53.

[49] Fletcher, P. T., and Joshi, S. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *In: ECCV Workshops CVAMIA and MMBIA. (2004) 87-98* (2004), Springer-Verlag, pp. 87–98.

[50] Frank, L. Characterisation of anisotropy in high angular resolution diffusion-weighted MRI. *Magn. Reson. Med 47* (2002), 1083–1099.

[51] Freund, Y., and Schapire, R. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences 55(1)* (1997), 119139.

[52] Friedman, J., Hastie, T., and Tibshirani, R. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics 28* (2000), 337–407.

[53] Friman, O., Farnebäck, G., and Westin, C. A bayesian approach for stochastic white matter tractography. *IEEE Trans. Med. Imaging 25* (2006), 965–978.

[54] Friston, K., and Frith, C. Schizophrenia: a disconnection sydrome? *Clin NeuroSci 3* (1995), 89–97.

[55] Gee, J., Alexander, D., Rivera, M., and Duda, J. Non-rigid registration of diffusion tensor MR images. In *IEEE International Symposium on Biomedical Imaging,* (2002), ProceedingsIEEE, pp. 477–480.

[56] Gerber, S., Tasdizen, R., Joshi, S., and Whitaker, R. On the manifold structure of the space of brain images. In *Medical Imaging Computing and Computer assisted Intervention (MICCAI)* (2009), Lecture Notes in Computer Science.

[57] Gerig, G., Gouttard, S., and Corouge, I. Analysis of brain white matter via fiber tract modelling. In *IEEE EMBS* (2004), pp. 4421–4424.

[58] Gilmore, J., Lin, W., Courage, I., Vetsa, Y., Smith, J., Kang, C., Gu, G., Hamer, R., Lieberman, J., and Gerig, G. Early postnatal development of corpus callosum and corticospinal white matter assessed with quantitative tractography. *Am. J. Neuroradiol 28* (2007), 1789–1795.

[59] Gong, G., He, Y., Conchas, L., Lebel, C., Gross, D., Evans, A., and Beaulieu, C. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cerebral Cortex 19* (2009), 524–536.

[60] Good, C., Johnsrude, I., Ashburner, J., Henson, R., Friston, K., and Frackowiak, R. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage 14* (2001), 21–36.

[61] Goodlett, C., Davis, B., Jean, R., Gilmore, J., and Gerig, G. Improved correspondence for DTI population studies via unbiased atlas building. In *Medical Imaging Computing and Computer assisted Intervention (MICCAI)* (2006), Lecture Notes in Computer Science, pp. 260–267.

[62] Goodlett, C., Fletcher, P., Gilmore, J., and G.Gerig. Group analysis of DTI fibre tract statistics with application to neurodevelopment. *NeuroImage 45* (2009), S133–S142.

[63] Gousias, I., Rueckert, D., Heckemann, R., Dyet, L., Boardman, J., Edwards, A., and Hammers, A. Automatic segmentation of brain MRIs of 2-years olds with 83 regions of interest. *NeuroImage 40* (2008), 672–684.

[64] Guimond, A., Meunier, J., and Thiron, J.-P. Average brain models: a convergence study. *Comput. Vis. Image Understand 77* (2000), 192–210.

[65] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., and Wedeen, V. J. Mapping the structural core of human cerebral cortex. *PLoS Biology 6(7)* (2008), e159.

[66] Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V., Meuli, R., and Thiran, J.-P. Imaging the brain neuronal network with diffusion MRI: a way to understand neuronal architecture. In *ISMRM* (2006), p. 436.

[67] Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V. J., Meuli, R., and Thiran, J.-P. Mapping human whole-brain structural networks with diffusion MRI. *PLoS one 7* (2007), e597.

[68] Hagmann, P., Thiran, J.-P., Vandergeynst, P., Clarke, S., Maeder, P., and Meuli, R. Mapping of human brain connectivity: statistical fibre track and virtual dissection. *NeuroImage 19* (2003), 545–554.

[69] Hajnal, J., Hill, D., and Hawkes, D. *Medial Image Registration.* CRC Press, London, 2001.

[70] Hammers, A., Allom, R., Koepp, M., Free, S., Myers, R., Lemieux, L., Mitchell, T., Brookes, D., and Duncan, J. Three-dimensional maximum probability atlas with particular reference to the temporal lobe. *Hum. Brain. Mapp 19(4)* (2003), 224–247.

[71] He, Y., Chen, Z., and Evans, A. Small world anatomical networks in the brain revealed by cortical thickness from MR. *Cerebr. Cortex. 17* (2007), 2407–2419.

[72] He, Y., Chen, Z., and Evans, A. Structural insights into aberrant topological patterns of large scale cortical networks in alzheimer's disease. *The journal of neuroscience 28(18)* (2008), 4756–4766.

[73] Hecke, W., Sijbers, J., D'Agnostino, E., Maes, F., Backer, S. D., Vandervliet, E., Parizel, P., and Leemans, A. On the construction of an inter-subject diffusion tensor magnetic resonance atlas of the healthy human brain. *NeuroImage 43* (2008), 69–80.

[74] Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., and Hammers, A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage 33* (2006), 115–126.

[75] Hess, C., Mukherjee, P., Han, E., Xu, D., and Vigneron, D. B. Q-ball reconstruction of multimodel fiber orientations using the spherical harmonic basis. *Magn. Reson. Med 56* (2006), 104–117.

[76] Highman, D., Kalna, G., and Kibble, M. Spectral clustering and its use in bioinformatics. *Journal of Computational and Applied Mathematics 204* (2007), 25–37.

[77] Hilgetag, C., Burns, G., O'Neill, M., Scannell, J., and Young, M. Anatomical connectivity defines organisation of clusters of connectivity in the macaque monkey and the cat. *Phil. Trans. R. Soc. Lond. B 355* (2000), 91–110.

[78] Honey, C., Kötter, R., Breakspear, M., and Sporns, O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *PNAS 104* (2007), 10240–10245.

[79] Hosey, T., Williams, G., and Ansorge, R. Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magn. Reson. Med 54* (2005), 1480–1489.

[80] Hsu, J.-L., Leemans, A., Bai, C.-H., Lee, C.-H., Tsai, Y.-F., Chiu, H.-C., and Chen, W.-H. Gender differences and age-related white matter changes of the human brain: A diffusion tensor imaging study. *NeuroImage 39* (2008), 566–577.

[81] Humphries, M., Gurney, K., and Prescott, T. The brain stem reticular formation is a small-world not scale-free network. *Proc. Biol. Sci 273* (2002), 503–511.

[82] Iosifescu, D., Shenton, M., Warfield, S., Kikinis, R., Dengler, J., Jolesz, F., and Mccarley, R. An automated registration algorithm for measuring MRI subcortical structures. *NeuroImage 6* (1997), 13–25.

[83] Iturria-Medina, Y., Canales-Rodríguez, E., Melie-García, L., Valdes-Hernández, P., Martínez-Montes, E., Alemán-Gómez, Y., and Sánchez-Bornot, J. Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *NeuroImage 36* (2007), 645–660.

[84] Jansons, K., and Alexander, D. Persistent angular structure: new insights from diffusion magnetic resonance imaging data. *Inverse Problems 19* (2003), 1031–1046.

[85] Jbabdi, S., Bellec, P., Toro, R., Daunizeau, J., Pélégrini-Issac, M., and Benali, H. Accurate anisotropic fast marching for diffusion-based geodesic tractography. *Journal of Biomedical Imaging 2008* (2008), 1–12.

[86] Jbabdi, S., Woolrich, M., Andersson, J., and Behrens, T. A Bayesian framework for global tractography. *NeuroImage 37* (2007), 116–129.

[87] Joshi, S., Davis, B., Jomier, M., and G.Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage 23* (2004), S151–S160.

[88] Kittler, J., Hatef, M., Duin, M., R.P.W., and Matas, J. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell. 20(3)* (1998), 226–239.

[89] Kochunov, P., Thompson, P., Lancaster, J., Bartzokis, G., Smith, S., Coyle, T., Royall, D., Laird, A., and Fox, P. Relationship between white matter fractional anisotropy and other indices of cerebral health in normal ageing: tract-based spatial statistics study of ageing. *NeuroImage 35* (2007), 478–487.

[90] Kontis, D., Catani, M., Cuddy, M., Walshe, M., Nosarti, C., Jones, D., Wyatt, J., Rifkin, L., Murray, R., and Alin, M. Diffusion tensor MRI of the corpus callosum and cognitive function in adults born preterm. *Cogn. NeuroSci. and Neuropsyc. 20* (2009), 424–428.

[91] Leemput, K. V., Maes, F., Vandermeulen, D., and Suetens, P. Automated model-based bias field correction of MR images of the brain. *IEEE Trans. Med. Imaging 18(10)* (1999), 885–896.

[92] Leemput, K. V., Maes, F., Vandermeulen, D., and Suetens, P. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging 18(10)* (1999), 897–908.

[93] Lenroot, R., Gogtay, N., Greenstein, D., Wells, E., Wallace, G., Clasen, L., Blumenthal, J., Lerch, J., Zijdenbos, A., Evans, A., Thompson, P., and Giedd, J. Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *NeuroImage. 36* (2007), 1065–1073.

[94] Little, J., Hill, D., and Hawkes, D. Deformations incorporating rigid structures. *Comput. Vis. Image Understand 66* (1997), 223–232.

[95] L.Minanti, Banasik, T., Brzezinski, J., Madelli, M., Bizzi, A., Bruzzone, M., Konopka, M., and Jasinski, A. Elevating tensor rank increases anisotropy in brain areas associated with intra-voxel orientational heterogeneity (IVOH): a generalised DTI (GDTI) study. *NMR Biomed 21* (2008), 2–14.

[96] Lori, N., Akbudak, E., Shimony, J., Cull, T., Snydor, A., Guillory, R., and Conturo., T. Diffusion tensor fiber tracking of human brain connectivity: acquisition methods, reliability analysis and biological results. *NMR Biomed 15* (2002), 293–515.

[97] Madden, D., Whiting, W., Huettel, S., White, L., Macfall, J., and J.M. Diffusion tensor imaging of adult age differences in cerebral white matter: relation to response time. *NeuroImage 21* (2004), 1174–1181.

[98] Mädler, B., Drabyc, S., Kolind, S., Whittal, K., and Mackay, A. Is diffusion anisotropy an accurate monitor of myelination? correlation of multicomponent t_2 relaxation and diffusion tensor anisotropy in human brain. *Magn. Reson. Imag. 26* (2008), 874–888.

[99] Mangin, J.-F., Poupon, C., Clark, C., Bihan, D. L., and Bloch, I. Eddy-current distortion correction and robust tensor estimation for MR diffusion imaging. In *Medical Imaging Computing and Computer assisted Intervention (MICCAI)* (2001), Lecture Notes in Computer Science.

[100] Marsland, S., Twining, C., and Taylor, C. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *Medical Imaging Computing and Computer assisted Intervention (MICCAI)* (2003), Lecture Notes in Computer Science, pp. 771–779.

[101] Maslov, S., and Sneppen, K. Specficity and stability in topology of protein networks. *Science 296* (2002), 910–913.

[102] McInerney, T., and Terzopoulos, D. Deformable models in medical image analysis: a survey. *Medical Image analysis 1(2)* (1996), 91–108.

[103] Mckinstry, R., Mathur, A., Miller, J., Ozcan, A., Snyder, A., Scheft, G., Almli, R., Shiran, S., Conturo, T., and Neil, J. Radial organization of developing preterm human cerebral cortex revealed by non-invasive water diffusion anisotropy MRI. *Cerebral Cortex 12* (2002), 1237–1243.

[104] McLaughlin, N., Paul, R., Grieve, S., Williams, L., Laidlaw, D., DiCarlo, M., Clark, C., Whelihan, W., Cohen, R., Whitford, T., and Gordon, E. Diffusion tensor imaging of the corpus callosum: a cross-sectional study across the lifespan. *Int. J. Devl. Neuroscience 25* (2007), 215–221.

[105] Melie-García, L., Canales-Rodríguez, E., Alemán-Gómez, Y., Lin, C., Iturria-Medina, Y., and Valdes-Hernández, P. A Bayesian framework to identify principal intravoxel diffusion profiles based on diffusion-weighted MR imaging. *NeuroImage 42* (2008), 750–770.

[106] Menke, R., Scjolz, J., Miller, K., Deoni, S., Jbadi, S., Matthews, P., and Zarei, M. MRI characteristics of the substantia nigra in parkinson's disease: A combined quantitative T1 and DTI study. *NeuroImage 47* (2009), 435–441.

[107] Meunier, D., Achard, S., Morcom, A., and Bullmore, E. Age related changes in modular organisation of human brain functional networks. *NeuroImage 44(3)* (2009), 715–723.

[108] Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D., and Bullmore, E. T. Hierarchical modularity in human brain functional networks. *frontiers in Neuroinformatics 3(37)* (2008), 1–12.

[109] Micheloyannis, S., Pachou, E., Stam, C., Breakspear, M., Bitsios, P., Vourkas, M., Erimaki, S., and Zervakis, M. Small-world networks an disturbed functional connectivity in schizophrenia. *Schizophr. Res 87* (2006), 60–66.

[110] Micheloyannis, S., Vourkas, M., Tsirka, V., Karakonstantaki, E., Kanatsouli, K., and Stam, C. The influence of ageing on complex brain networks: a graph theoretical analysis. *Human Brain Mapping 30* (2009), 200–208.

[111] Mori, S., Crain, B., Chacko, V., and van Zijl, P. Three dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol. 45* (1999), 265–269.

[112] Mori, S., Itoh, R., Zhang, J., Kaufmann, W., Zijl, P. V., Solaiyappan, M., and Yarowsky, P. Diffusion tensor imaging of developing mouse brain. *Magn Res Med 46* (2001), 18–23.

[113] Morra, J., Tu, Z., Apostolova, L., Green, A., Toga, A., and Thompson, P. Comparison of AdaBoost and Support Vector Machines for detecting Alzheimer's disease through automated hippocampal segmentation. *IEEE Trans Med Imaging 29(1)* (2010), 30–43.

[114] Moseley, M., Cohen, Y., Kucharczyk, J., Asgari, H., Wendland, M., Tsuruda, J., and Norman, D. Diffusion weighted MR imaging of anisotropic water diffusion in cat nervous system: technical, anatomic and pathologic considerations. *Radiology 14* (1990), 865–873.

[115] Nagy, Z., Westerberg, H., Skare, S., Anderson, J., Lilja, A., Flodmark, O., Fernell, E., Holmberg, K., Bohm, B., Forssberg, H., Lagercrantz, H., and Klingberg, T. Preterm children have disturbances of white matter at 11 years of age as shown by diffusion tensor imaging. *Pediatr Res 54(5)* (2003), 672–679.

[116] Narr, K., Bilder, R., Toga, A., Woods, R., Rex, D., Szeszko, P., Robinson, D., Sevy, S., Gunduz-Bruce, H., Wang, Y.-P., DeLuca, H., and Thompson, P. Mapping cortical thickness and gray matter concentraction in first episode schizophrenia. *Cerebral Cortex 15* (2005), 708–719.

[117] Neil, J., Shiran, S., Mckinstry, R., Schefft, G., Snyder, A., Almli, C., Akbudak, E., Aronovitz, J., Miller, J., Lee, B., and Conturo., T. Normal brain in human newborns diffusion coefficient and diffusion anisotropy measured by using diffusion tensor MR imaging. *Radiology 209* (1998), 55–66.

[118] Newman, M. Fast algorithm for detecting community structure in networks. *Phys. Rev. E. 69* (2004), 066133.

[119] Newman, M. Modularity and community structure in networks. *PNAS 103* (2006), 8557–8582.

[120] Odegard, T., Farris, E., Ring, J., R.McColl, and Black, J. Brain connectivity in non-reading impaired children and children diagnosed with developmental dyslexia. *Neuropsychologia 8-9* (2009), 1972 – 1977.

[121] O'Donnell, L., and Westin, C.-F. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. Med. Imaging 26(11)* (2007), 1562–1575.

[122] O'Donnell, L., Westin, C.-F., and Golby, A. Tract-based morphometry for white-matter group analysis. *NeuroImage 45* (2009), 832–844.

[123] Ogawa, S., Lee, T., Nayak, A., and Glynn, P. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn. Reson. Imag. 14* (1990), 6878.

[124] Oh, J., Song, I. C., Lee, J., Kang, H., Park, K., Kang, E., and Lee, D. Tractography-guided statistics (TGIS) in diffusion tensor imaging for the detection of gender difference of fiber integrity in the midsagittal and parasagittal corpora callosa. *NeuroImage 36* (2007), 606–616.

[125] Onnella, J.-P., Saramäki, J., Kertész, J., and Kaski, K. Intensity and coherence of motifs in weighted complex networks. *PHYSICAL REVIEW E 71* (2005), 065103.

[126] Otsu, N. A threshold selection method from a grey-level histogram. *IEEE Trans, Syst. Man. Cybern 9* (1979), 62–66.

[127] Ozarslan, E., and Mareci, T. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magn. Reson. Med. 50* (2003), 955–965.

[128] Ozarslan, E., Vermuri, B., and Mareci, T. Generalized scalar measures for diffusion MRI using trace, variance and entropy. *Magn. Reson. Med. 56* (2005), 395–410.

[129] Park, H.-J., Kubicki, M., Shenton, M. E., Guimond, A., McCarley, R. W., Maier, S., Kikinis, R., Jolesz, F., and Westin, C.-F. Spatial normalisation of diffusion tensor MRI using multiple channels. *NeuroImage 20* (2003), 1995–2009.

[130] Parker, G., and Alexander, D. Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Phil. Trans. R. Soc. B 360* (2005), 893–902.

[131] Parker, G., Haroon, H., and Wheeler-Kingshott, C. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *J. Magn. Reson. Imaging. 18* (2002), 242–254.

[132] Parker, G., Wheeler-Kingshott, C., and Barker, G. Distributed anatomical brain connectivity derived from diffusion tensor imaging. *Lecture Notes in Computer Science 2082* (2001), 106–120.

[133] Parker, G., Wheeler-Kingshott, C., and Barker, G. Estimating distributed anatomical brain connectivity using Fast Marching methods and diffusion tensor imaging. *IEEE Trans. Med. Imaging 21* (2002), 505–12.

[134] Pennec, X., Fillard, P., and Ayache, N. A riemannian framework for tensor computing. *International Journal of Computer Vision 66(1)* (2006), 41–66.

[135] Pfefferbaum, A., Adalsteinsson, E., and Sullivan, E. Frontal circuitry degradation marks healthy adult ageing. evidence from diffusion tensor imaging. *NeuroImage 26(3)* (2005), 891–899.

[136] Qiu, D., Tan, L.-H., Zhou, K., and Khong, P.-L. Diffusion tensor imaging of normal white matter maturation from late childhood to young adulthood: Voxel-wise evaluation of mean diffusivity, fractional anisotropy, radial and axial diffusivities and correlation with reading development. *NeuroImage 41* (2008), 223–232.

[137] R.Brown. 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. *Philosph. Mag 4* (1828), 161.

[138] Roca, P., Riviére, D., Guevara, P., Poupon, C., and Mangin, F.-F. Tractography based parcellation of the cortex using a spatially-informed dimension reduction of the connectivity matrix. In *Medical Imaging Computing and Computer assisted Intervention (MICCAI)* (2009), Lecture Notes in Computer Science.

[139] Rohlfing, T., Russakoff, D., and Maurer, C. Expectation maximisation strategies for multi-atlas multi-label segmentation. *Inf. Process. Med. Imaging 18* (2003), 210–221.

[140] Rohlfling, T., Brandt, R., Menzel, R., and Jr, C. M. Evaluation of atlas selection strategies for atlas-based segmentation with application to confocal microscopy of bee brains. *NeuroImage 21* (2004), 1428–1442.

[141] Rubin, D. Frontal-striatal circuits in cognitive ageing: evidence for caudate involvement. *Aging, Neuropsychology, and Cognition 6(4)* (1999), 241–259.

[142] Rubinov, M., Knock, S., Stam, C., Micheloyannis, S., Harris, A., Williams, L., and Breakspear, M. Small world properties of non-linear brain activity in schizophrenia. *Human Brain Mapping 30* (2009), 403–416.

[143] Rueckert, D., Frangi, A., and Schnabel, J. Automatic construction of 3-D statistical deformation models of the brain using non-rigid registration. *IEEE Trans. Med. Imaging 22(8)* (2003), 1014–1025.

[144] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., and Hawkes, D. Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Med. Imaging 18(8)* (1999), 712–721.

[145] S. Jbabdi, T. b., and Smith, S. Crossing fibres in tract-based spatial statistics. *NeuroImage 49* (2010), 249–256.

[146] Sabunco, M., Balci, S., and Golland, M. S. P. Image-driven population analysis through mixture modelling. *IEEE Trans. Med. Imaging* (2009).

[147] Sakai, K., Yamada, K., Nagakane, Y., Mori, S., Nakagawa, M., and Nishimura, T. Diffusion tensor imaging may help the determination of time at onset of cerebral ischaemia. *J. Neurol Neurosurg Psychiatry 80* (2009), 986–990.

[148] Salvador, R., Suckling, J., Schwarzbauer, C., and Bullmore, E. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Phil. Trans. B. 360* (2005), 937–946.

[149] Sebastiani, G., Pasquale, F. D., and Barone, P. Quantifying human brain connectivity from diffusion tensor MRI. *J Math Imaging Vis 25* (2006), 227–244.

[150] Shannon, C. The mathematical theory of communication (parts 1 and 2). *Bell Systems Tech. J 27* (1948), 379–423 and 623–656.

[151] Shen, X., Papademetrix, X., and Constable, R. Graph-theory based parcellation of functional sub-units in the brain from resting state data. *NeuroImage in press* (2010).

[152] Sherbondy, A., Akers, D., Mackenzie, R., Dougherty, R., and Wandell, B. Exploring connectivity of the brain's white matter with dynamic queries. *IEEE Transactions on Visualization and Computer Graphics 11*, 4 (2005), 419–430.

[153] Sherbondy, A., Dougherty, R., Ben-Shacher, M., Napel, M., and Wandell, B. ConTrack: finding the most likely pathways between brain regions using diffusion tractography. *Journal of Vision 8(9)* (2008), 1–16.

[154] Shi, J., and Malik, J. Normalised cuts and image segmentation. Tech. Rep. UCB/CSD-97-940, University of California, Berkeley, 200.

[155] Smith, S. Fast robust automated brain extraction. *Hum. Brain. Mapp 17* (2002), 143–155.

[156] Smith, S., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C., Watkins, K., Ciccarelli, O., Cader, M., Matthews, P., and Behrens, T. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage 31* (2006), 1487–1505.

[157] Smith, S., and Nichols, T. E. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage 44* (2009), 83–98.

[158] Soterxo, Y. I.-M. R., Canales-Rodríguez, E., Alemán-Gómez, Y., and Melie-García, L. Studying the human brain anatomical network via diffusion-weighted MRI and graph theory. *NeuroImage 40* (2008), 1064–1076.

[159] Sporns, O. Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *BioSystems 85* (2006), 55–64.

[160] Sporns, O., Chialvo, D., Kaiser, M., and Hilgetag, C. Organization, development and function of complex brain networks. *Trends in Cog. Sci. 8(9)* (2004), 418–425.

[161] Sporns, O., Tononi, G., and Edelman, G. Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex 10* (2000), 127–141.

[162] Sporns, O., Tononi, G., and Kötter, R. The human connectome: a structural description of the human brain. *PLoS Computational Biology 1(4)* (2005), e42.

[163] Sporns, O., and Zwi, J. The small world of the cerebral cortex. *Neuroinformatics 2* (2004), 145–162.

[164] Stam, C. Functional connectivity patterns of human magneto-encephalographic readings: a 'small-world' network? *Neurosci Lett 355* (2004), 25–28.

[165] Stam, C., de Haan, W., Daffersthofer, A., Jones, B., Manshanden, I., van Cappellen van Walsum, A., Montez, T., Verbunt, J., de Munck, J., van Dijk, B., Berendse, H., and Scheltens, P. Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. *Brain 132* (2009), 213–224.

[166] Stam, C., Jones, B., Nolte, G., Breakspear, M., and Scheltens, P. Small-world networks and functional connectivity in Alzheimer's disease. *Cerebral Cortex 17* (2007), 92–99.

[167] Stam, C., and van Dijk, B. Synchronisation likelihood: an unbiased measure of generalised synchronisation in multivariate data sets. *Physica D. 163* (2002), 236–241.

[168] Stejskal, E., and Tanner, J. Spin diffusion measurements: spin echoes in the prescience of a time-dependent field gradient. *J. Chem. Phys. 42* (1965), 288–292.

[169] Studholme, C., Hill, D., and Hawkes, D. Automated 3D registration of MR and PET brain images by multi-resolution optimisation of voxel similarity measures. *Med. Phys. 24(1)* (1997), 25–35.

[170] Studholme, C., Hill, D., and Hawkes, D. An overlap invariant entropy measure of 3d medical image alignment. *Pattern recognition 32* (1999), 71–86.

[171] Sullivan, E., Adalsteinsson, E., Hedehus, M., Ju, C., Mosely, M., Lim, K., and Pfefferbaum, A. Equivalent disruption of regional white matter microstructure in ageing in healthy men and women. *NeuroReport 12* (2001), 99–104.

[172] Sullivan, M., Jones, D., Summers, P., Morris, R., Williams, S., and Markus, H. Evidence for cortical "disconnection" as a mechanism as a mechanism of age-related cognitive decline. *Neurology 57* (2001), 632–638.

[173] Supekar, K., Menon, V., Rubin, D., Musen, M., and Greicius, M. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Comput. Biol. 4 e1000100 4* (2008), e1000100.

[174] Talairach, J., Szikla, G., and Tournoux, P. *Atlas d'anatomie stereotaxique du telencephale.* Masson, Paris, 1967.

[175] Talairach, J., and Tournoux, P. *Co-planar stereotactic atlas of the human brain: 3 dimensional proportional system: an approach to cerebral imaging.* George Thieme Verlag, Stuttgart, 1999.

[176] Thomaz, C. E., Boardman, J., Counsell, S., Hill, D., Hajnal, J., Edwards, A., Rutherford, M., Gillies, D., and Rueckert, D. A multivariate statistical analysis of the developing human brain in preterm infants. *Image and Vision Computing 25* (2007), 981–994.

[177] Thomaz, C. E., Duran, F., Busatto, G., Gillies, D., and Rueckert, D. Multivariate statistical differences of MRI samples of the human brain. *J. Math. Imaging Vis. 29* (2007), 95–106.

[178] Toosey, A., Werring, D., Orrell, R., Howard, R., King, M., Barker, G., Miller, D., and Thompson, A. Diffusion tensor imaging detects corticospinal tract involvement at

multiple levels in amytrophic lateral sclerosis. *J. Neurol. Neurosurg. Phsychiatry 74* (2003), 1250–1257.

[179] Torrey, H. Bloch equations with diffusion terms. *Phys. Rev. 104* (1956), 563–565.

[180] Tournier, J., Calamante, F., Gadian, D., and Connelly, A. Diffusion weighted magnetic resonance imaging fiber tracking using a front evolution algorithm. *NeuroImage 20* (2003), 276–288.

[181] Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., and Weeden, V. J. High angular diffusion imaging of intravoxel white matter fiber heterogeneity. *Magn. Reson. Med 48* (2002), 577–582.

[182] Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., and Weeden, V. J. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine 48* (2002), 577–582.

[183] Tuch, D. S., Reese, T. G., Wiegell, M. R., and Weeden, V. J. Diffusion MRI of complex neural architecture. *Neuron 40* (2003), 885–895.

[184] Valstar, M., and Pantic, M. Fully automatic facial action unit detection and temporal analysis. In *CVPR* (2006), pp. 149–.

[185] van den Heuvel, M., Stam, C., Boersma, M., and Pol, H. H. Small-world and scale free organisation of voxel-based resting state functional connectivity in the human brain. *NeuroImage* (2008).

[186] Viola., P. *Alignment by maximization of mutual information.* PhD thesis, Massachusetts Institute of Technology, 1995.

[187] von Luxburg, U. A tutorial on spectral clustering. *Stat Computing 17* (2007), 395–416.

[188] Wakana, S., Jiang, H., Nagae-Poetscher, L., Zijl, P. V., and Mori, S. Fiber tract-based atlas of human white matter anatomy. *Radiology 230* (2004), 77–87.

[189] Warfield, S., Zou, K., and Wells, W. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging 23(7)* (2004), 903–921.

[190] Watss, D., and Strogatz, S. Collective dynamics of 'small-world' networks. *Nature 393* (1998), 440–442.

[191] Wedeen, V., Hagmann, P., Tseng, W.-Y. I., Reese, T., and Weisskoff, R. Mapping complex tissue architecture with dffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine 54* (2005), 1377–1886.

[192] Wells, W., Grimson, E., Kikinis, R., and Jolesz, F. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imaging 15* (1996), 429–442.

[193] Whitcher, B., Wisco, J., Hadjikhani, N., and Tuch, D. Statistical group comparison of diffusion tensors via multivariate hypothesis testing. *Magn. Reson. Imag. 57* (2007), 1065–1074.

[194] Yeo, B., Vercuteren, T., Fillard, P., Peyrat, J.-M., Pennec, X., Golland, P., Ayache, N., and Catz, O. DT-REFinD: Diffusion tensor registration with exact finite-strain differential. *IEEE Trans. Med. Imaging 28(12)* (2009), 1914–1928.

[195] Zalesky, A., and Fornito, A. A DTI-derived measure of cortico-cortico connectivity. *IEEE Trans Med Imaging 28(7)* (2009), 1023–1036.

[196] Zhang, H., Awate, S., Das, S., Woo, J., Melhem, E., Gee, J., and Yushkevich, P. A tract-specific framework for white matter morphometry combining macroscopic and microscopic tract features. In *Medical Imaging Computing and Computer assisted Intervention ()* (2009), Lecture Notes in Computer Science.

[197] Zhang, H., Yushkevich, P., Alexander, D., and Gee, J. Deformable registration of diffusion tensor MR images with explicit orientation optimisation. *Med. Imag. Anal. 10* (2006), 764–785.

[198] Zhang, Y., Brady, M., and Smith, S. Segmentation of brain MR images through a hidden markov random field model and the expectation maximisation algorithm. *IEEE Trans. Med. Imaging 20* (2001), 45–57.

[199] Ziyan, U., Sabuncu, M., O'Donnell, L., and Westin, C.-F. Nonlinear registration of diffusion MR images based on fiber bundles. In *Medical Imaging Computing and Computer assisted Intervention (MICCAI)* (2007), Lecture Notes in Computer Science, pp. 351–538.