MR Diffusion Tractography: Methods and Applications

Tim Behrens

Hilary Term, 2004

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain

and

Department of Engineering Science

University of Oxford

Acknowledgements

There are many people who have contributed to the work in this thesis both directly and indirectly.

First, I would like to thank my DPhil supervisors, Steve Smith and Mike Brady for their continual excellent support and advice throughout the course of my DPhil - without their contributions this work would not have been possible; Paul Matthews and Steve Smith for providing a truly first class environment in which to carry out scientific research; and, again, Steve Smith for lending me his spare bed when I was homeless writing this thesis, and for generally being a very good friend.

Heidi Johansen-Berg has proved an endless source of neuroscience questions, ideas and of core research, and has an equal share in the work in the second half of my DPhil, specifically chapters 6 and 7. She is great fun both to work, and to drink with and I hope I can continue to spend time with her in both capacities.

Despite his invaluable advice in Bayesian Statistics and, specifically, MCMC, without which I would not have been able to take the route I did in the methods developement in this thesis, I still regard Mark Woolrich's major contribution to my years at FMRIB as being in the role of "Whipping Boy". His repeated failed efforts to beat me on the Golf Course, the tennis court, the squash court; even in the bowling lane and on the "iSnooker" table have filled me with an enormous sense of self-satisfaction over the past three years. Long may it continue.

Mark Jenkinson and Christian Beckmann have provided excellent advice in mathematical and physical modelling during the course of this research. They have also both

contributed hugely to my enjoyment of my time at FMRIB, and will remain long-term friends.

Huge Thank Yous should also go out the the whole image analysis group at FMRIB. In particular, Ivana Drobnjak was invaluable when I was searching for a matrix permutation algorithm.

Dave Flitney and Duncan Mortimer have provided both excellent IT support for myself and everybody using my software (it hasn't been easy given my lack of coding prowess) and first class banter over the coffee table in the morning.

Donna Lloyd, Arthur Magill and Dave Homfrey brought much amusement and, more importantly, Tequila until they deserted us for greener pastures. Nice!

Further Thank Yous should go to: Narender Ramnani for close collaboration in the cerebral peduncle work; Claudia Wheeler-Kingshott, Gareth Barker, Phil Boulby, Olga Ciccarelli, Alan Thompson and the Multiple Sclerosis Society of Great Britain for providing much of the human diffusion data used in this thesis; Marlene and Wolfgang Richter for providing the monkey data for the peduncle study; Des Higham for help and advice with the matrix permutation; and the EPSRC for funding me during my DPhil.

Above all, I would like to thank my family, and my girlfriend, Louise. My mother and father have been supportive in every way possible. I am trying to find some eloquent words to describe what they have given me but, for once, my tongue is tied. I can only hope that they already know. Gina has always been a joy to know. She will remain my closest friend. Perhaps now she will stop asking me when I will get a job!

I know that I should tell her this more often than I do and I hope that she knows it without ever having to be told: Louise is a constant source of support, love and happiness in my life, and without her I could never have finished this thesis.

Statement of intellectual contribution.

The material presented in this thesis is entirely my own work except in the following places:

Chapter 6 was contributed to in equal shares by myself and Heidi Johansen-Berg. It is not possible to subdivide this chapter into individual contributions. Every idea was discussed between us.

Chapter 7:

Section 7.1 was carried out by myself and Narender Ramnani. The work with the monkey diffusion data is mine. The Neuroscience concepts and ideas belong to Narender. This section is included largely as a demonstration of the potential of the technique and diffusion tractography in general.

Section 7.3.1 was again carried out in equal share with Heidi Johansen-Berg. Again many of the ideas were formed between us, but here there is a slightly clearer division. The developement of the methodology for connectivity-profile based parcellation is mine; the application to the delineation of SMA/preSMA belongs to Heidi.

MR Diffusion Tractography: Methods and Applications

Tim Behrens

Magdalen College
University of Oxford

Abstract

Diffusion MR imaging provides a unique insight into tissue architecture at the microscopic level. Specifically, it enables scientists to probe the orientational structure of brain white matter and, from this, to recover fibre trajectories in the living human brain. However, in general, "Diffusion Tractography" algorithms have been limited to the reconstruction of pathways in deep white matter regions and have provided only a qualitative description, or visualisation, of the resulting fibre tracts. This thesis addresses the problem of inferring on white matter connectivity in the presence of imperfect diffusion data, providing a generalisation of the commonly used "streamlining tractography" algorithm to the case when there is uncertainty in the local fibre architecture. In doing so, it recovers connectivity *distributions* providing a *quantitative* description of belief in the fibre trajectories, and enabling complete trajectories to be traced from their grey matter sources to their grey matter targets. This thesis provides the first demonstration of quantitative inference of anatomical connectivity between grey matter structures using diffusion data.

The connectivity distributions recovered from this "probabilistic tractography" provide a new and rich source of information with which to infer on anatomical connectivity in the living human brain. This information is used to provide novel neuroscientific findings relating to thalamo-cortical and cerebellar-cortical circuitry in the human brain. Specifically, this thesis provides the first topographic map of the cortical connections of the human thalamus, and demonstrates that, by examining only cortico-thalamic circuitry, functional and cytoarchitectonic subunits, or nuclei, of the thalamus may be indentified. In the cortico-cerebellar system, cortical inputs to the cerebellum in human are contrasted and compared with those in macaque, providing the first demonstration of inter-species connectional differences with diffusion data. Finally new ways are proposed in which diffusion tractography data may be used to identify functional subunits and functional sub-systems in the absence of any prior knowledge about the connectional architecture of the brain region.

Thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford

Hilary Term, 2004

Contents

List of Figures

Chapter 1

Introduction

The transfer of information between processing units in the brain provides the basis for brain function. The substrate for this information transfer is brain white matter - a network of microscopic cellular wire-like structures, known as axons, carrying electric signals from their own cell bodies onto the synapses of other cells. The function of every processing unit, or neuron, in the brain is constrained by its location in the axonal network. For this reason, a full understanding of brain function at any scale relies on the ability to draw the axonal network at that same scale.

At present, information on axonal connectivity between brain regions is very hard to establish. In the in-vivo human brain, anatomical connectivity is effectively invisible and, even post-mortem, the available methodologies reveal only very limited information (see chapters 5 and 6). By far the dominant source of information regarding connectivity in the *human* brain is indirect inference from sacrificial tracer studies in *non-human* primate. Such studies have revealed a great deal but are extremely time consuming and limited to specific brain regions. Moreover, knowledge of connectional anatomy in the non-human brain does not necessarily imply knowledge in human (again, see chapter 5).

This thesis focuses on the development and use of a new method for exploring connectivity in the living human brain. Magnetic resonance diffusion imaging uses the diffusion of water molecules as a macroscopic probe of tissue microstucture - particularly of the orientational structure of axonal fibres in brain white matter (1). Using diffusion imaging and its derivative, diffusion tensor imaging (2), it has been possible to estimate mean fibre directions in fine resolution imaging voxels throughout the brain white matter (see chapter 2) and thence to reconstruct major white matter pathways in the living human brain (3; 4; 5). Such early reports of the in-vivo measurement of anatomical connectivity raised hopes of new insights into brain function and connectional dysfunction but, as yet, diffusion tractography has been limited to the reconstruction of deep white matter pathways easily visible in post-mortem dissection.

1.1 Neural Architecture and Diffusion Imaging.

A brain cell, or neuron, consists of a cell body (soma) which processes signals received at its dendrites and transmits signals down its axon. The axon terminals synapse at other cell dendrites, passing signals to connected cells. Surrounding the axon, and providing electrical insulation, are Schwann's cells which contain a fatty substance known as Myelin. A simplified diagram of a neuron is shown in figure 1.1.

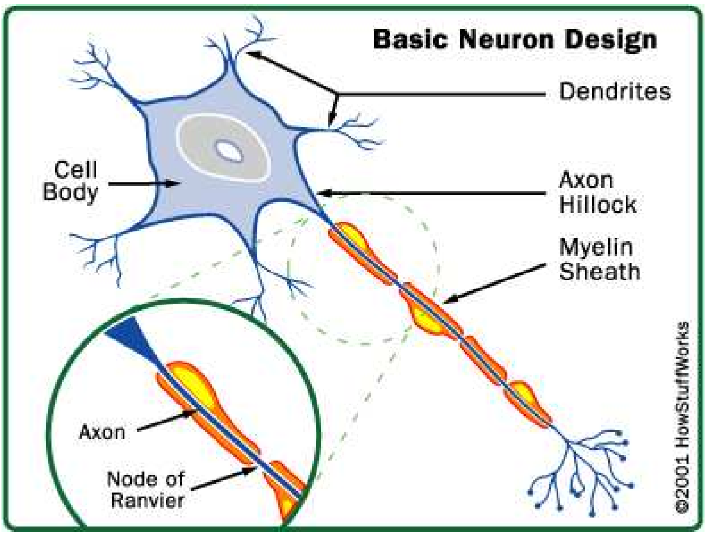


Figure 1.1: Diagram of a Neuron. Inset, the wire-like structure of the axon, along with its surrounding myelin sheath hinders diffusion perpendicular to its long axis. Taken from (6)

Diffusion imaging relies upon this structure of axons, and their myelin sheaths.

Water molecules undergoing random thermal motion (see chapter 2) in the vicinity of the axon are less likely to diffuse across the long axis of the semi-permeable axonal cell membrane and myelin sheath than they are along it. Diffusion is said to be more "hindered" perpendicular to the long axis of the fibre than parallel with it. When we attempt to trace fibre pathways using diffusion imaging data, we exploit this diffusion "anisotropy" to recover the local orientation of axons. However, we must consider the crucial question of scale. The diameter of a typical axon is in the region of $1 - 3\mu m$, whereas diffusion weighted MR data are typically acquired on a grid of the order of $2 \times 2 \times 2mm^3$. There may be hundreds of thousands of axons passing through each imaging voxel. If, in a particular voxel, the axonal fibres are oriented at random then the diffusion data will contain no orientational information. The second important feature of axons relied upon by diffusion imaging is their organisation. Axons emerging from or entering a brain area tend to organise themselves into well aligned "fibre bundles", and deep white matter consists of large bundles of aligned axonal fibres (see figure 1.2) carrying electric signals between distant areas of the brain. These fibre bundles exist at a resolution which is available to MR imaging studies, hence diffusion anisotropy on *this scale* may be measured with magnetic resonance. Figure 1.3 shows principal diffusion directions (foreground) and diffusion anisotropy (background) recovered from a diffusion tensor imaging experiment. Note that in areas of large fibre bundles, diffusion anisotropy is high and the principal diffusion directions align well with the long axis of the bundle. For these reasons, it is the large deep white matter fibre

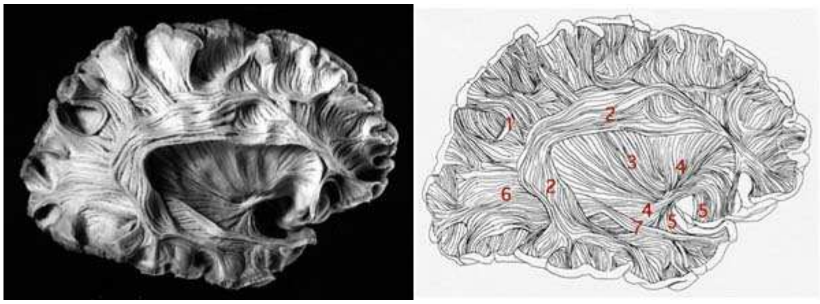


Figure 1.2: Fibre bundles in the brain. 1. Short arcuate bundles. 2. Superior longitudinal fasciculus. 3. External capsule. 4. Inferior occipitofrontal fasciculus. 5. Uncinate fasciculus. 6. Sagittal stratum. 7. Inferior longitudinal fasciculus. Taken from (7).

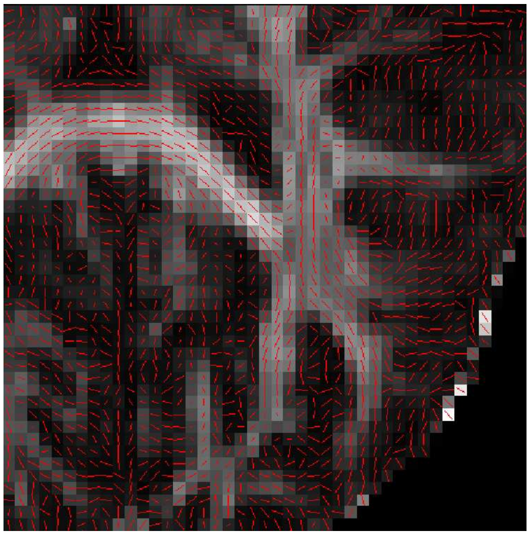


Figure 1.3: Principal diffusion directions from diffusion tensor imaging. PDDs are projected into the slice plane and overlaid on fractional anisotropy in a normal brain. Note that, in large fibre bundles, diffusion anisotropy if high (background) principal diffusion direction is aligned with the long axis of the bundle.

bundles which have been reconstructed successfully by diffusion tensor tractography.

1.2 Thesis Research

It is at this point that the research in this thesis commences. It focuses on several key areas in making diffusion based tractography a useful probe of neural connectivity and, in doing so, it provides new neuroscience results relating to thalamo-cortical and cerebellar-cortical circuitry in the human brain.

In the past, diffusion tractography studies have been limited to the goal of in-vivo *visualisation* of major fibre pathways. The resulting pictures, although often visually extremely impressive (e.g. (8; 9)) have been difficult to interpret scientifically. The lack of a quantitative measure of strength or probability of connection between two locations in the brain has made it impossible, for example, to carry out between-group comparisons to examine connectional dysfunction in disease. Moreover, the deterministic nature of the tractography algorithms to date has meant that in areas of low diffusion anisotropy, where it is not possible to decide on a single diffusion direction to pursue, it has not been possible to trace *any* connections. For this reason, it has not been possible to trace connections to their grey matter targets and hence make inference about axonal connectivity between grey matter regions in the brain.

The principal methodological contribution of this thesis is to provide a statistical framework in which to infer anatomical connectivity from diffusion imaging data. We first consider the problem of inferring on voxel-wise models of diffusion in the presence of image noise and incomplete signal modelling. The inclusion of these factors results in *uncertainty* in the recovered fibre orientation (or, in the case of more complicated models of the local architecture, uncertainty in the recovered distribution of fibre orientations). This uncertainty is represented as probability distribution functions (*pdf*s) on the local fibre architecture (in the simplest case, on the local mean fibre orientation). We then consider the problem of tracing pathways through the diffusion data field in the presence of this uncertainty. It is no longer possible to trace deterministic pathways through the data. Instead, we provide a generalisation of streamlining tractography to the case when there is uncertainty in the local fibre orientation. We estimate a *pdf* on the location of the pathway (a *connectivity distribution*), resulting in probabilities of connection between a seed location and every other voxel in the brain. We are able to *quantify* our belief in the location of a pathway, and therefore quantify our belief in the existence of axonal connections between brain regions. By removing the need to make a deterministic decision at every step in the tractography process, we are able to trace beyond regions of low diffusion anisotropy deep into grey matter structures, allowing us to provide the first demonstration of reliable inference of anatomical connectivity between human grey matter structures using diffusion data (chapter 6).

The connectivity distributions resulting from a statistical treatment of diffusion tractography provide a rich but complicated source of information. Other methodological contributions in this thesis focus on the mining and representation of this new source of data to provide means for extracting interesting results. For example, in chapter 7 we provide a method for extracting networks of strongly connected voxels from a data field, and a method for searching for “discontinuities in the connectivity field", so that neighbouring brain regions which exhibit markedly different connectivity patterns may be distinguished (See figure 1.8).

The availability of so rich a source of connectivity information along with the techniques to mine it has led to experimental contributions in this thesis.

By examining connectivity distributions seeded in the human thalamus, we provide the first topographic map of human thalamo-cortical connectivity (figure 1.6 and chapter 6). We are then able to identify functionally discrete thalamic nuclei on the basis of their cortical connections, providing an in-vivo segmentation of thalamus into its functionally and cytoarchitectonically discrete nuclei. Detailed comparison between the location of these nuclei and the location of functional MRI "hot spots" in experiments designed to activate them, along with detailed comparison of their volumes with volumes reported by post-mortem thalamic dissections provide the first quantitative attempt to validate results from diffusion tractography studies.

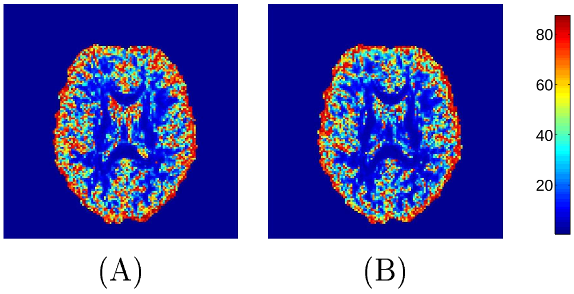
Similar techniques to those employed in the thalamus are used to investigate cortical projections to cerebellum (which pass through cerebral peduncle). Connectivity-based segmentation of the peduncle in human reveals a topology of cortical projections to peduncle which is familiar from invasive tracer studies in macaque. However, the relative contribution of fibres from the different cortical zones found in the human tractography data is dramatically different to those reported by the macaque tracer studies (figure 1.7 (A-C)). Specifically, fibres from the prefrontal cortex occupy a relatively larger territory in the human tractography data than in the tracer data from macaque, supporting the new hypothesis of a role for cerebellum in cognitive processing (e.g. (10; 11)). This result is consolidated with diffusion tractography data from macaque monkey, which shows the same topology and *the same relative contributions* as the invasive tracer studies in macaque (figure 1.7 (D)), providing further validation for diffusion tractography and, to our knowledge, the first direct comparison of diffusion tractography results in different species.

1.3 Thesis Overview

The thesis is split into the following chapters.

Chapters 2 and 3 introduce the reader to concepts fundamental to the new research in the thesis. Chapter 2 focuses on the theory behind imaging diffusion with nuclear magnetic resonance, and chapter 3 provides an introduction to learning on models with Bayesian statistics.

Chapter 4 uses Bayesian statistics to investigate the uncertainty inherent to parameter estimates in models of the diffusion and measurement processes, resulting in *probability density functions* on parameters in the diffusion tensor model and in a simple two compartment model of diffusion. These distributions are compared with empirical distributions constructed via repeated acquisition of diffusion imaging data (See figure 1.4) It goes on to propose a more complicated model of diffusion accounting for the effects on the measured NMR signal of a continuous distribution of fibre orientations within a voxel.



(A) (B)

Figure 1.4: Width of the 95% "cone of uncertainty" on fibre direction (in degrees), (A) Constructed from a Bayesian analysis of a single diffusion imaging data set. (B) Constructed from maximum likelihood analyses of repeated diffusion imaging data sets.

Chapter 5 lays down the theory for performing tractography in the presence of uncertainty in the voxel-wise parameter estimates - specifically, uncertainty in the mean underlying fibre orientation. Essentially, it derives a generalisation of streamlining tractography to the case of imperfect diffusion weighted data, resulting in a connectivty *distribution* from a seed point (see figure 1.5). It then proposes a fast algorithm for performing the multi-dimensional integrals required to compute the this distribution. The material in chapters 4 and 5 is published in Magnetic Resonance in Medicine (12).

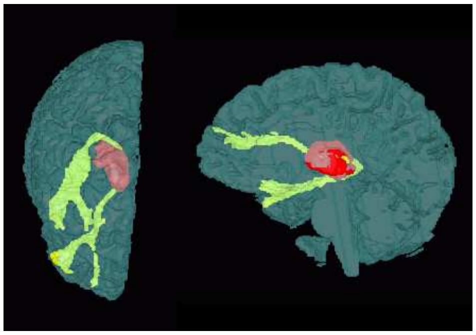
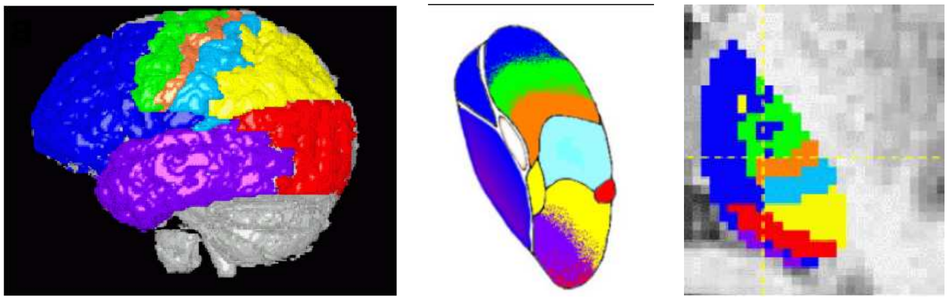


Figure 1.5: Connectivity distribution from medial dorsal thalamus. The distribution terminates in prefrontal and temporal cortices.

Chapter 6 applies the methodology in chapters 4 and 5 to map the cortical connectivity of the human thalamus. It introduces the concepts of connectivity based segmentations and connectivity defined regions and demonstrates the similarity between cortico-thalamic connections found with diffusion tractography in the human brain, and those found with sacrificial tracer studies in the non-human primate brain (figure 1.6).

Group mapping of the connectivity defined regions in figure 1.6 results in a connectivity defined probabilistic atlas of human thalamus. Locations of putative mediodorsal and ventral lateral nuclei derived from this atlas are compared with functional MRI "hot-spots" in experiments designed to activate the same nuclei. Volumes of all of the putative nuclei are compared with volumes derived from post-mortem thalamic dissection and staining. Some of the material in chapter 6 is published in Nature Neuroscience (13).

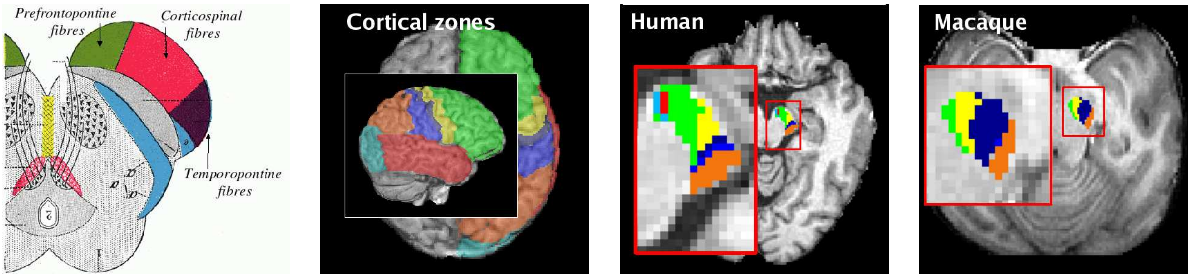
Chapter 7 presents recent developments and ongoing studies related to the research in this thesis.



(A) (B) (C)

Figure 1.6: Connectivity-based segmentation of the thalamus. (A) Division of the human cortex into principal zones of connectivity of the thalamic nuclei in macaque. (B) Schematic of a human thalamus with boundaries between nuclei marked. Overlaid in colour, the expected cortical targets of the respective thalamic nucleus derived from invasive tracer studies in macaque. (C) Most probable cortical target derived from diffusion-based connectivity distribution seeded from each voxel.

A topologic segmentation similar to that provided for thalamus in chapter 6 is provided for cerebral peduncle, to investigate projections from cortex to cerebellum (which pass through peduncle). The topology of the resulting segmentation is found to be as predicted by invasive tracer studies (figure 1.7 (A-C)), but the relative volumes of the territories occupied by the different cortical regions are dramatically different. This result is consolidated by the same experiment in macaque monkey where the relative volumes are as expected from the tracer studies (figure 1.7 (D)).



(A) (B) (C) (D)

Figure 1.7: Cortical connectivity of the cerebral peduncle: (A) Derived from invasive tracer studies in macaque, (B) Colour code for cortical zones in (C),(D). (C) derived from diffusion tractography in human. (D) Derived from diffusion tractography in macaque.

Two studies are presented which rely on mining the connectivity data to reveal interesting features in the data. The first relies on the hierarchical organisation of the primate visual system which reveals itself as a characteristic structure in the connectivity matrix of the system. By permuting the matrix in search of this structure we are able to extract clusters of highly connected voxels corresponding to putative ventral and dorsal visual streams, and primary visual cortex. The second attempts to isolate clusters of voxels which exhibit similar connectivity patterns, and therefore draw boundaries between neighbouring brain areas which show differences in their connectivity pattern. In order to achieve this, the cross correlation matrices of the connectivity matrices are computed and reordered such that highly correlated "nodes" (or voxels) appear near to each other on the graph (figure 1.8 (a) and (b) left). By identifying clusters in this reordered correlation matrix, we are able to draw a previously invisible boundary between functional areas in brain (figure 1.8 (a) and (b) right). The specific example chosen is the boundary between Supplementary Motor Area (SMA) which connects primarily to motor regions and preSMA which connects primarily to frontal regions.

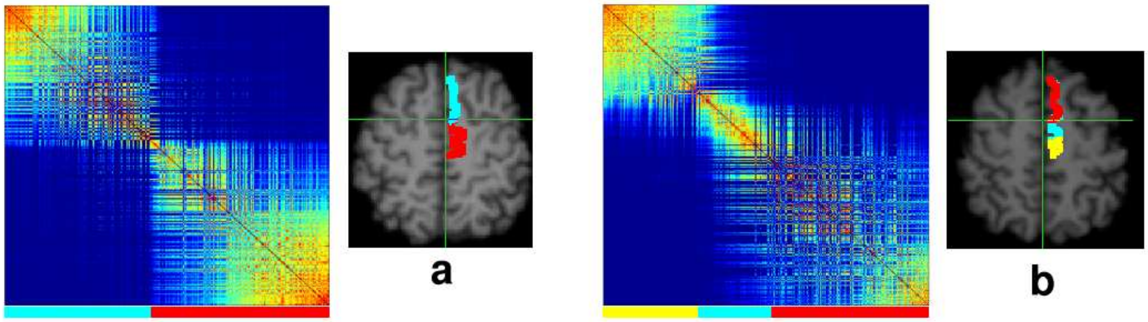


Figure 1.8: Reordering connectivity correlation matrices in SMA/preSMA. In 8/11 subjects two clusters were apparent in the reordered correlation matrices (a typical example is shown in (a)). In 3/11 subjects three clusters were apparent (a typical example is shown in (b)).

Chapter 8 Draws conclusions from the previous chapters, and discusses some open questions in the field of diffusion tractography.

Chapter 2

Measuring Molecular Diffusion with NMR

2.1 Introduction

Imagine, for a moment, that you are sitting in a sealed room staring at a stationary glass of water. At least, imagine that you are staring at what appears to be a stationary glass of water. At the molecular level, well beyond the resolution of human perception things are far from peaceful. Water molecules, and in fact the molecules of any fluid, are at constant random motion. This phenomenon, known as Brownian motion, was first observed by Robert Brown in the early 19^{th} century while he was examining pollen grains in a water solution (14), but remained unexplained for nearly 50 years until the emergence of the kinetic theory of matter late in the same century. At the macroscopic level, the effects of Brownian motion may be witnessed as fluid diffusion. If we take our imaginary glass of water and place in it a small drop of iodine we will see that, over time, the iodine diffuses through the water until the concentration is the same throughout. Each molecule of iodine is taking a random route through the water molecules but, *on average*, they are moving from areas of high iodine concentration to areas of low concentration. The statistical nature of the diffusion process renders its macroscopic effects entirely predictable. Laws governing macroscopic diffusion were laid out well before the microscopic processes were understood (15). Crucially, even without the presence of the iodine, the water molecules are undergoing diffusion in exactly the same fashion, but now the water molecules are diffusing amongst themselves. This process is known as *self-diffusion* and obeys all the same laws as diffusion. In this thesis, the terms "self-diffusion" and "diffusion" are used interchangeably to mean self-diffusion.

For scientists interested in the microstructure of tissue, the appeal of diffusion is that *microscopic* features of the medium in which diffusion is occurring will reveal themselves in the *macroscopic* properties of the diffusion process. If a water molecule undergoing random motion encounters a barrier such as a myelin wall, it may be less likely to diffuse across it than along it, and moreover, *every* molecule which encounters the wall will experience these same probabilities. The presence of the myelin wall will be visible in the macroscopic properties of diffusion in its near vicinity. The goal of diffusion imaging in biological tissue is to use the diffusion process as a macroscopic probe of tissue microstructure.

This chapter aims to introduce the reader to many of the concepts in diffusion imaging required to understand the remainder of this thesis. In doing so, it provides an outline of the theory behind measuring diffusion with NMR. For excellent detailed descriptions of this theory along with descriptions of the underlying physics of diffusion see (16) and (17).

2.2 Molecular diffusion and the ensemble average diffusion propagator

Due to the random processes at hand, we are not able to predict the path of a single molecule during the process of Brownian motion. However, if we know all of the relevant physical properties of the diffusing substance, we *are* able to make statistical predictions about the macroscopic behaviour of a large number of molecules. The most general descriptive tool available to us for this purpose is the *diffusion propagator.* The diffusion propagator describes the *distribution* of displacements of molecules in the ensemble. That is, the probability of finding a molecule at $\mathbf{r}$ given that it was at $\mathbf{r}_0$ a time τ ago:

$$p(\mathbf{r}|\mathbf{r}_0, \tau). \tag{2.1}$$

However, this distribution may change over space; it is conditioned on $\mathbf{r}_0$, the initial location of the molecule. If we wish to describe, or measure, the distribution of displacements of an *ensemble* of molecules, we may not restrict ourselves to considering an infinitesimal point in space, $\mathbf{r}_0$. We must describe or measure the *average* diffusion propagator for an ensemble of molecules in some finite volume $\mathcal{V}$.

The *ensemble average diffusion propagator* in volume $\mathcal{V}$ describes the expected distribution of vector displacements, $\mathbf{r} - \mathbf{r}_0$, of molecules in volume $\mathcal{V}$ after time τ or, equivalently, describes the average over $\mathcal{V}$ of the probability of a single molecule moving a distance $|\mathbf{r} - \mathbf{r}_0|$ along direction $\widehat{\mathbf{r} - \mathbf{r}_0}$ in time τ:

$$p(\mathbf{r} - \mathbf{r}_0|\tau). \tag{2.2}$$

The common objective of the NMR diffusion techniques presented in the remainder of this chapter is to measure, or approximate, this distribution.

2.3 Diffusion and the NMR spin echo

The possibility of measuring diffusion was hinted at in the very early days of NMR. In his 1950 paper on the NMR spin echo (18) Hahn noted that the random thermal motion of spins would reduce the amplitude of the spin echo in the presence of magnetic field inhomogeneity. Subsequently, Torrey (19) provided an adaption to the Bloch equations (the fundamental equations describing the magnetisation of spins in an NMR experiment) to account for this effect, and hence predict the attenuation of the spin echo given knowledge of the magnetic field and the local diffusion characteristics. Initially this effect was regarded as a potential artefact in a spin echo NMR experiment. However, in 1965, Stejskal and Tanner (20) used the *Pulsed gradient spin echo experiment* (PGSE) to harness this NMR "artefact" to perform quantitative measurements of molecular diffusion in a sample.

2.3.1 The Pulsed Gradient Spin Echo

The PGSE modifies Hahn's spin echo experiment with the introduction of pulsed gradients before and after the application of the 180° radio-frequency pulse (see figure 2.1), thereby explicitly creating the field inhomogeneities which had earlier been a cause of concern for Hahn.

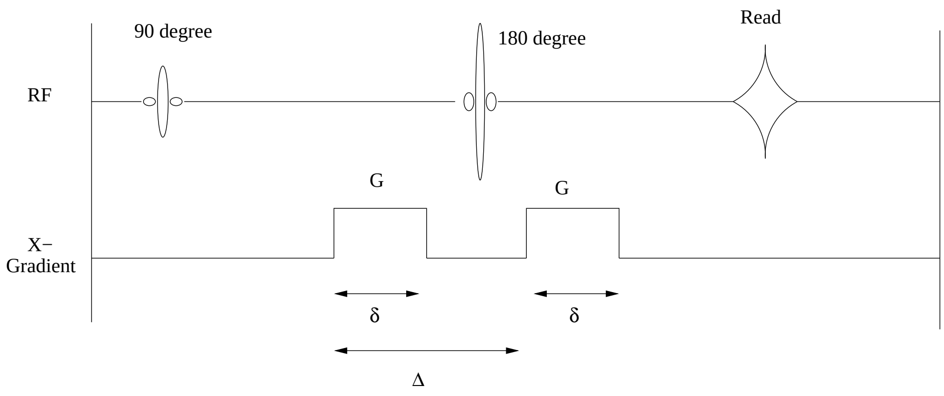


Figure 2.1: Pulse sequence diagram for Stejskal-Tanner pulse gradient spin echo experiment. For an explanation of pulse sequence diagrams see (21)

The easiest way to understand this experiment (as is often the case with NMR) is to imagine ourselves rotating at the Larmor frequency in the transverse magnetic plane (see (21)). First let us consider the spin echo experiment in the absence of diffusion encoding gradients (i.e. the top row in figure 2.1). The 90° RF pulse flips the spins into the transverse plane where, in the ideal scenario, they remain spinning at the Larmor frequency ($\omega_0 = \gamma B_0$, where γ is the Larmor constant and B_0 is the static magnetic field), but lose magnetisation due to spin-spin, or T_2, relaxation. In fact, due to local variations in B_0, the spins will each rotate at slightly different Larmor frequencies causing a net spin dephasing (figure 2.2 left). After the 180° pulse, the spins rotating faster are “behind” the slower spins, hence the spins begin to rephase (figure 2.2). The RF read out is taken at the “spin-echo”, when the spins are once again in phase, with the only attenuation of the magnetisation being due to spin-spin relaxation.

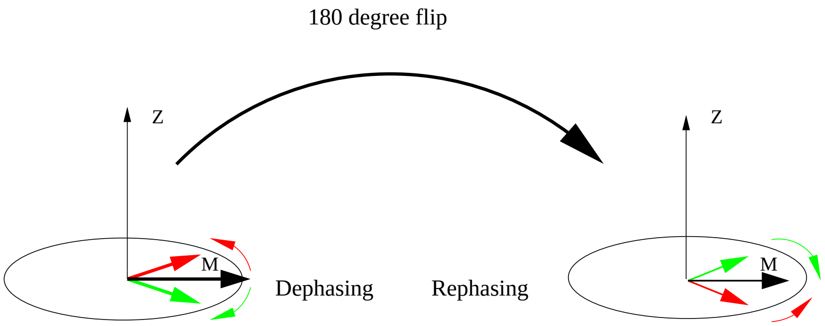


Figure 2.2: Dephasing and rephasing of the spins in a spin echo experiment viewed in a coordinate frame rotating about z at the mean Larmor frequency. Note that the sizes of the magnetisation vectors are smaller in the second diagram due to the inherent spin-spin relaxation.

If we now modify this experiment with Stejskal and Tanner’s pulsed gradients (bottom row in figure 2.1), then we are *ensuring* that the static field is inhomogeneous. The effect of this is that spins that do not move in the course of the experiment will experience the same field strength during the first and second pulses, and hence rephase exactly as before, but spins which have some component of displacement along the gradient direction (for example, due to random molecular diffusion) will experience different field strengths during the two pulses and hence not return to (net) zero phase at the time of the echo. This dephasing of the magnetisation vectors of the spins leads to an attenuation of net magnetisation measured in the NMR signal.

Using the Bloch-Torrey equations, Stejskal and Tanner were able to relate this attenuation directly to the diffusion coefficient (20), thus providing a complete description of the diffusion process in the case of isotropic Gaussian or *free* diffusion. Later in the same year (22), Stejskal provided a more general formalism to relate the net magnetisation measured in the PGSE experiment to any generic diffusion propagator. This relationship is explained in detail in (17), and is outlined here.

2.3.2 Relationship with the diffusion propagator

If we assume that the spins do not move during the application of diffusion encoding gradients (an assumption known as the "infinitely narrow pulse condition"), then the phase accrued by each spin relative to the phase $\gamma B_0 \delta$ during each pulse will be:

$$\phi = \gamma \int_0^\delta \mathbf{g}^T \mathbf{r} \; dt = \gamma \delta \mathbf{g}^T \mathbf{r} \tag{2.3}$$

where γ is the Larmor constant, δ is the length of time over which the gradient is applied (figure 2.1), $\mathbf{g}$ is the gradient strength vector and $\mathbf{r}$ is the position of the spin in question.

The phase accrued by each spin during the course of the experiment is then (still relative to $\gamma B_0 \delta$)

$$\phi = \gamma \delta \mathbf{g}^T (\mathbf{r}_2 - \mathbf{r}_1) \tag{2.4}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the positions of the spin during the first and second gradient pulses.

The signal, $S(\tau_\epsilon)$, measured from the PGSE experiment is simply the linear superposition of the magnetisations of N spins each with phase ϕ_k.

$$S(\tau_\epsilon) = \sum_{k=1}^{N} m(\tau_\epsilon) e^{i\phi_k} = S_0(\tau_\epsilon) \left[\frac{1}{N} \sum_{k=1}^{N} e^{i\phi_k} \right] \tag{2.5}$$

where N is the number of spins, $m(\tau_\epsilon)$ is the modulus of the magnetisation vector after echo time τ_ϵ (see figure 2.2) and $S_0(\tau_\epsilon) = N m(\tau_\epsilon)$ is the spin echo signal in the absence of any diffusion encoding gradients.

The term in square brackets in equation 2.5 is the expected value, or ensemble average, of $e^{i\phi}$ which may be rewritten in terms of the probability density function on the spin phases as

$$\frac{1}{N} \sum_{k=1}^{N} e^{i\phi_k} \approx \langle e^{i\phi} \rangle = \int e^{i\phi} p(\phi|\tau) d\phi \tag{2.6}$$

for diffusion time τ. Substituting in equation 2.4 and defining the wave vector $\mathbf{q} = \gamma \delta g$

and $\mathbf{R} = \mathbf{r}_2 - \mathbf{r}_1$ gives

$$\frac{S(\mathbf{q}, \tau, \tau_\epsilon)}{S_0(\tau_\epsilon)} = \langle e^{i\phi} \rangle = \int e^{i\mathbf{q}^T\mathbf{R}} p(\mathbf{q}^T\mathbf{R}|\tau) d\mathbf{q}^T\mathbf{R}. \tag{2.7}$$

However, noting the linearity between spin displacement $\mathbf{R}$ and phase $\phi = \mathbf{q}^T\mathbf{R}$, for constant $\mathbf{q}$, we may replace the expectation over $\mathbf{q}^T\mathbf{R}$ with an expectation over $\mathbf{R}$:

$$\begin{aligned} \frac{S(\mathbf{q}, \tau, \tau_\epsilon)}{S_0(\tau_\epsilon)} &= \int e^{i\mathbf{q}^T\mathbf{R}} p(\mathbf{R}|\tau) d\mathbf{R}. && (2.8) \\ &= \mathcal{F}_3 \left[p(\mathbf{R}|\tau) \right]. && (2.9) \end{aligned}$$

Where, $\mathcal{F}_3$ is the 3D Fourier transform with respect to spin displacement $\mathbf{R}$. The signal we expect to measure as a function of wave-vector, $\mathbf{q}$, is linearly related to the Fourier reciprocal of the ensemble average diffusion propagator. By taking measurements at many points in $\mathbf{q}$-space and simply performing an inverse Fourier transform, it is possible to reconstruct the ensemble average diffusion propagator.

2.3.3 Free Diffusion

In the case where diffusion is free (i.e. where there are no barriers to restrict or hinder molecular displacement and no molecular absorption), solution of Fick's laws, which govern the macroscopic behaviour of diffusion (15), leads to an isotropic Gaussian diffusion propagator.

$$p(\mathbf{R}|\tau) = (4\pi D\tau)^{-3/2} \exp\left(-\frac{\mathbf{R}^T\mathbf{R}}{4D\tau} \right) \tag{2.10}$$

where D is the diffusion coefficient in Fick's laws of diffusion and subsequently in the Einstein relation (23)

$$D = \frac{1}{6\tau} \langle \mathbf{R}^T\mathbf{R} \rangle, \tag{2.11}$$

relating the diffusion time, τ, to the mean square spin displacement, $\langle \mathbf{R}^T\mathbf{R} \rangle$, in a freely diffusing medium.

If we image free diffusion with the PGSE experiment, then equation 2.9 predicts the signal attenuation as follows:

$$\begin{aligned} \frac{S(\mathbf{q},\tau)}{S_0} &= \mathcal{F}_3\left[p(\mathbf{R}|\tau)\right] && (2.12) \\ &= \exp\left(-D\tau|\mathbf{q}|^2\right). && (2.13) \end{aligned}$$

Returning to figure 2.1, the effective diffusion time, τ, after correction for diffusion during the gradient pulses (see (20)) is $\Delta - \delta/3$. Replacing $\mathbf{q}$ with its constituent variables now gives:

$$\frac{S(\mathbf{q},\tau)}{S_0} = \exp\big(-|\gamma\mathbf{g}\delta|^2(\Delta - \delta/3)D\big), \tag{2.14}$$

which is the original relationship derived by Stejskal and Tanner. Introducing LeBihan's b-value,

$$b = |\gamma\mathbf{g}\delta|^2(\Delta - \delta/3), \tag{2.15}$$

reveals the most widely used equation in diffusion imaging:

$$\frac{S(b)}{S_0} = e^{-bD}. \tag{2.16}$$

Under the assumption of free diffusion, this relationship enables us to characterise completely the macroscopic diffusion of the medium in question with a single spin echo experiment (to measure S_0) and a single PGSE experiment (to measure $S(b)$). We may simply compute

$$D = -\frac{1}{b}\log_e\frac{S(b)}{S_0} \tag{2.17}$$

2.3.4 The Apparent Diffusion Coefficient

In biological tissue, the assumption of free diffusion does not hold. Barriers to diffusion such as semi-permeable cell membranes or myelinated axonal sheaths will hinder diffusion such that the diffusion propagator is non-Gaussian (see section 2.6). The true form of the diffusion propagator has been characterised in many such biological samples by acquiring at a fine resolution in $\mathbf{q}$-space and performing the inverse Fourier transform (e.g. (24; 25)). However, with the arrival of diffusion *imaging*, the change from measuring a single diffusion propagator for the entire sample to measuring a separate propagator for each imaging voxel necessitated the assumption of *free* diffusion (diffusion *weighted* imaging). Diffusion *tensor* imaging relaxes this assumption to allow for anisotropic diffusion but still relies on the the Gaussianity of the propagator (see section 2.5). Without this assumption of Gaussianity the first diffusion imaging experiments would not have been feasible. However its result is that the quantitative measures of diffusion - the diffusion coefficient or tensor - will differ between experiments with different diffusion times or gradient strengths. For this reason, the recovered measures are termed the *Apparent* Self Diffusion Coefficient (ADC) and the *Apparent* Self Diffusion Tensor.

2.4 Imaging Diffusion

The theory related in the section above addresses the problem of measuring the diffusion propagator as an ensemble average over the entire sample excited by the initial RF pulse. However, in 1973, almost a decade after Stejskal and Tanner first measured diffusion with NMR, American chemist Paul Lauterbur (26) and English Physicists Peter Mansfield and Peter Grannell (27) showed independently that, also by applying magnetic gradients on top of the static magnetic field and therefore also by ensuring that spins at different locations in the sample rotated at different Larmor frequencies, it was possible to use the NMR phenomenon to form two dimensional images of the experimental sample. Shortly afterwards, Garroway, Grannell and Mansfield (28) devised a scheme for selectively exciting a single two dimensional plane of a three dimensional sample, allowing for non-invasive three dimensional imaging of biological samples with NMR.

In the mid 1980s (29; 30), it became apparent that it was possible to use the pulsed gradients of Stejskal and Tanner in the same experiment as the imaging gradients of Lauterbur and Mansfield to attenuate the signal in *each imaging voxel* by the effects of diffusion in the same way as the signal from the entire sample was attenuated in section 2.3. The importance of *Diffusion-weighted imaging* first became clear when it was noted that in animal models of stroke, the local diffusion characteristics around the infarction changed dramatically within minutes of the ischemia (specfically, the diffusion coefficient was reduced by up to 60%)(31) (see figure 2.3) whereas other markers available to MRI changed only very slightly and hours after the event.

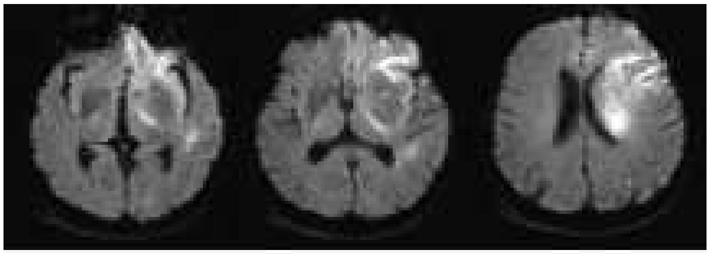


Figure 2.3: Diffusion weighted images of sub-acute stroke in the vascular territory of the left middle cerebral artery (taken from (32))

2.5 Diffusion Tensor Imaging

The reconstruction of the diffusion coefficient under the assumption of free diffusion, as is carried out in *diffusion weighted imaging*, deliberately masks the effects of any orientational structure in the diffusion propagator. The spins in the voxel are assumed to diffuse along each direction with the identical probability profile. The first and, as yet, most significant step to relax this assumption was taken by Peter Basser in 1994 (2) with the introduction of *Diffusion Tensor Imaging* (DTI). DTI does not rely on the assumption of free diffusion but, in its place, assumes that the diffusion propagator takes on some generic zero-mean 3D Gaussian profile:

$$P(\mathbf{R}|\tau) = \frac{1}{\sqrt{|\mathbf{D}|(4\pi\tau)^3}} \exp\left(-\frac{\mathbf{R}^T\mathbf{D}^{-1}\mathbf{R}}{4\tau}\right) \tag{2.18}$$

where $\mathbf{D}$ is the symmetric positive definite Apparent Self-Diffusion Tensor. The diffusion coefficient $D_{\mathbf{R}}$ in any single direction $\mathbf{R}$ may be computed as:

$$D_{\mathbf{R}} = \mathbf{R}^T \mathbf{D} \mathbf{R} \tag{2.19}$$

and the covariance matrix, $\mathbf{C}$, of the Gaussian propagator is related to the diffusion tensor according to:

$$\mathbf{C} = 2\mathbf{D}\tau. \tag{2.20}$$

The key effect of moving to a generic Gaussian propagator is the ability to characterise diffusion anisotropy. That is, DTI may characterise diffusion processes in which the diffusion characteristics are different in different directions.

2.5.1 Estimating the Diffusion Tensor

Referring back to sections 2.3.2 and 2.3.3, it follows that the expected strength of the spin echo in the case of generic Gaussian diffusion is

$$\frac{S(b, \hat{\mathbf{u}})}{S_0} = \exp\left(-b\hat{\mathbf{u}}^T \mathbf{D} \hat{\mathbf{u}}\right) \tag{2.21}$$

where b is the b-value in equation 2.15 and $\hat{\mathbf{u}}$ is the unit vector along which the diffusion encoding gradients were applied. Taking logs and rewriting leads to

$$\log\left(\frac{S(b, \hat{\mathbf{u}})}{S_0}\right) = -\sum_{i=1}^{3}\sum_{j=1}^{3} \mathbf{b}_{ij}\mathbf{D}_{ij} \tag{2.22}$$

where $\mathbf{b}_{ij}$ is the ij^{th} element of the $\mathbf{b}$-matrix; $\mathbf{b} = b\mathbf{u}\mathbf{u}^T$.

Equation 2.22 has 7 unknowns: The six independent elements of the $\mathbf{D}$, and S_0. Hence, in order to solve for $\mathbf{D}$ uniquely, we must take at least 7 measurements. Writing these n measurements in matrix form yields:

$$\log \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_n \end{bmatrix} = - \begin{bmatrix} b^1_{xx} & 2b^1_{xy} & 2b^1_{xz} & b^1_{yy} & 2b^1_{yz} & b^1_{zz} & -1 \\ b^2_{xx} & 2b^2_{xy} & 2b^2_{xz} & b^2_{yy} & 2b^2_{yz} & b^2_{zz} & -1 \\ b^3_{xx} & 2b^3_{xy} & 2b^3_{xz} & b^3_{yy} & 2b^3_{yz} & b^3_{zz} & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b^n_{xx} & 2b^n_{xy} & 2b^n_{xz} & b^n_{yy} & 2b^n_{yz} & b^n_{zz} & -1 \end{bmatrix} \begin{bmatrix} \mathbf{D}_{xx} \\ \mathbf{D}_{xy} \\ \mathbf{D}_{xz} \\ \mathbf{D}_{yy} \\ \mathbf{D}_{yz} \\ \mathbf{D}_{zz} \\ \log S_0 \end{bmatrix}, \tag{2.23}$$

or

$$\log \mathbf{S} = \mathbf{A}\alpha. \tag{2.24}$$

If we take 7 measurements, $S_1 - S_7$, such that $\mathbf{A}$ is of full rank, we may simply invert $\mathbf{A}$ and compute

$$\alpha = \mathbf{A}^{-1} \log \mathbf{S} \tag{2.25}$$

However, with only a single measurement per parameter the recovered diffusion tensor is very sensitive to experimental error. For this reason, experimenters tend to take measurements along many different diffusion encoding directions using at least two different b-values. For example, a typical DTI experiment might, in each imaging voxel, take 6 measurements with no diffusion weighting ($b = 0smm^{-2}$) and 60 measurements with b-value, $b \approx 1000smm^{-2}$ along diffusion encoding directions evenly spaced on a unit sphere (for more details and explanation, see (33)). In this case, it is impossible to choose a diffusion tensor such that the NMR signal predicted by the Gaussian model (equation 2.21) matches exactly with the signal measured in the experiment. The system of equations laid out in equation 2.23 is over-constrained but can be solved in a least squares sense by writing the parameter $\chi^2(\alpha)$:

$$\chi^{\mathbf{2}}(\alpha) = (\log \mathbf{S} - \mathbf{A}\alpha)^T \mathbf{\Sigma}^{-1} (\log \mathbf{S} - \mathbf{A}\alpha), \tag{2.26}$$

which is the weighted sum of squared differences between observed and predicted echo intensities. The diagonal elements of Σ are σ^2/S_i^2 (where σ is the standard deviation of the image noise) and, to a first order approximation, correct for the distortion introduced by the logarithmic transformation of $S(b, \hat{\mathbf{u}})/S(0)$ (see (2) for further explanation). Minimising $\chi^2(\alpha)$ with respect to each of the seven parameters in α yields an optimal solution for α:

$$\alpha_{opt} = (\mathbf{A}^T \mathbf{\Sigma}^{-1} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{\Sigma}^{-1}) \log \mathbf{S}, \tag{2.27}$$

2.5.2 Properties of the Diffusion Tensor

The eigensystem of the diffusion tensor. For each estimated $\mathbf{D}$ it is possible to construct a local orthogonal coordinate system along which the components of diffusion are decoupled. There are then three independent apparent diffusion coefficients (one for each orthogonal axis) known as the principle diffusivities. Because $\mathbf{D}$ is positive definite and symmetric, its three eigenvectors ϵ_{1-3} are orthogonal and form the principle diffusion directions. The principle diffusivities are the components of diffusion along these directions and correspond to the eigenvalues λ_{1-3} of $\mathbf{D}$, so:

$$\mathbf{D} = (\epsilon_1 | \epsilon_2 | \epsilon_3) \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} (\epsilon_1 | \epsilon_2 | \epsilon_3)^T. \tag{2.28}$$

Isoprobability Surfaces. The 3D Gaussian diffusion propagator may be visualised by drawing surfaces of equal probability of $\mathbf{R}$ for any diffusion time τ. This can be achieved by setting the exponent in equation 2.18 to be a constant, for example:

$$\frac{\mathbf{R}^T \mathbf{D}^{-1} \mathbf{R}}{2\tau} = 1. \tag{2.29}$$

If this equation is transformed into the coordinate set of $\mathbf{D}$ by

$$\mathbf{R}' = (\epsilon_1|\epsilon_2|\epsilon_3)^T\mathbf{R}, \tag{2.30}$$

then from equations 2.29 and 2.28

$$\frac{\mathbf{R}'^T\Lambda^{-1}\mathbf{R}'}{2\tau} = 1, \tag{2.31}$$

where Λ is the matrix of eigenvalues. Equation 2.31 can be expanded to give

$$\left(\frac{x'}{\sqrt{2\lambda_1\tau}}\right)^2 + \left(\frac{y'}{\sqrt{2\lambda_2\tau}}\right)^2 + \left(\frac{z'}{\sqrt{2\lambda_3\tau}}\right)^2 = 1. \tag{2.32}$$

The solutions of equation 2.32 lie on an ellipsoid with major axes equal to the mean diffusion distances ($\sqrt{2\lambda_i\tau}$) along the principle diffusion directions.

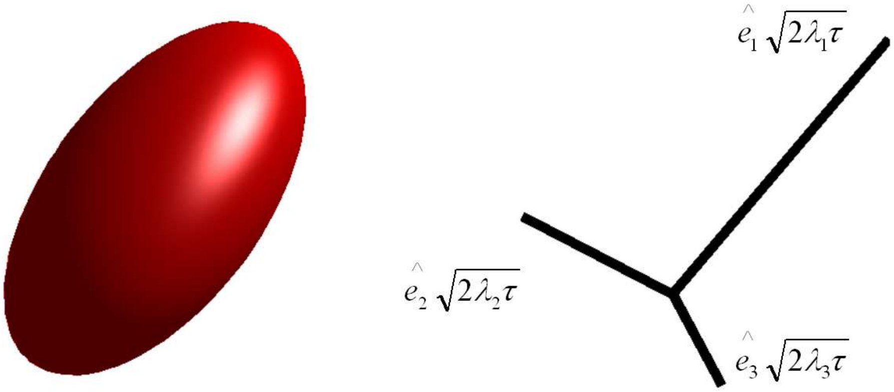


Figure 2.4: Isoprobability surface for Gaussian diffusion. Left: Ellipsoidal Diffusion isoprobability surface. Right: Corresponding orthonormal basis set.

Scalar Properties The parameterisation of the diffusion propagator with a 3D Gaussian allows for various rotationally invariant scalar properties to be extracted. Most notably, properties relating to the overall diffusivity, and to the anisotropy of the diffusion within a voxel.

Trace The trace of the diffusion tensor is a measure of the overall diffusion in a voxel, independent of the gradient direction. It can be computed from the original tensor, or from the eigenspectrum.

$$Tr(\mathbf{D}) = \sum_{i=1}^{3} \mathbf{D}_{ii} = \sum_{i=1}^{3} \lambda_i \tag{2.33}$$

Anisotropy Measures Diffusion *anisotropy* is the degree to which diffusion in a voxel is preferred in one direction over others. Various different measures of anisotropy have been proposed. For example, eigenvalue ratio A_{er}, the normalised eigenvalue ratio A_{ner} (34),the volume ratio A_{vr} (35).

$$A_{er} = \tfrac{\lambda_1}{\lambda_3} \quad A_{ner} = \tfrac{\lambda_1 - \lambda_3}{\lambda_1} \quad A_{vr} = \tfrac{\lambda_1 \lambda_2 \lambda_3}{\bar{\lambda}^3}, \tag{2.34}$$

where $\bar{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$. However the most commonly used anisotropy index is the Fractional Anisotropy (35).

$$A_{fa} = \sqrt{\frac{3}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \sum_{i=1}^{3} \sqrt{(\lambda_i - \bar{\lambda})^2}. \tag{2.35}$$

2.6 Imaging Diffusion in the Brain

We have already seen in section 2.4 that the sensitivity of the echo magnitude to the effects of stroke may be increased dramatically by encoding for diffusion along only a single gradient direction. The possibility of measuring diffusion in *several* directions in the same living brain brought with it the chance to measure diffusion anisotropy. As early as 1990, before the advent of DTI, Moseley et al. (1) measured diffusion coefficients parallel with, and perpendicular to, axons in cat white matter and made the startling observation that the measured diffusion was repeatably anisotropic. That is, diffusion was faster along the long axonal axis than in either of the perpendicular directions.

The biophysical explanation of this diffusion anisotropy is still not fully understood.

Early reports (e.g. (36)) suggested that the intracellular diffusion across the direction of the fibre was restricted by the myelin fatty sheaths which surround the axons. More recently, some anisotropy has been reported in the white matter of neonates who have not yet developed myelin (37) but the anisotropy increased as the axons became myelinated. Suggestions have also been made that a large proportion of the measured diffusion is extracellular. It is likely that diffusion anisotropy is influenced by a number of micro structural features: The integrity of axonal cell membranes, the amount and integrity of myelin around the axons, the coherence of axonal orientation, and the number and size of axons. An excellent review of this topic can be found in (38).

The advent of diffusion tensor imaging allowed for rotationally invariant quantitative measures of diffusion anisotropy to be taken at every imaging voxel (see section 2.5.2). For example, figure 2.5 shows a slice through a map of Fractional Anisotropy of a human brain. These measures alone have proven to be a strong marker for white matter integrity in neurological disorders such as Multiple Sclerosis (39), Amyotrophic Lateral Sclerosis (40), Schizophrenia (41; 42; 43) and Dyslexia (44) to name but a few. (For a review of the clinical applications of DTI, see (45)). However, possibly the

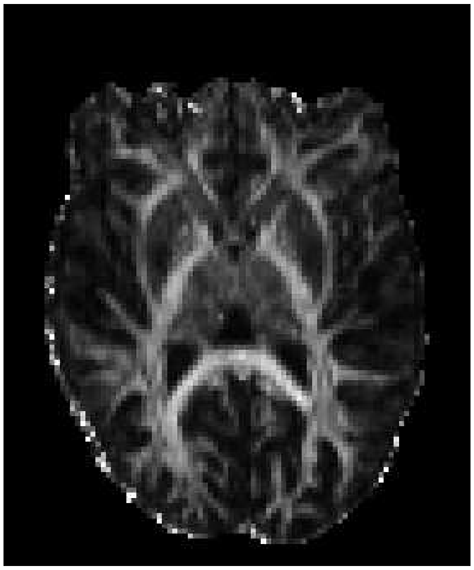


Figure 2.5: Slice through Fractional anisotropy map of a human brain.

most exciting feature of diffusion imaging, and the underlying basis for the research in this thesis is the ability to use diffusion anisotropy to measure the mean orientation of axons in a voxel. Figure 2.6 shows *principal diffusion directions* (PDDs) reconstructed from DTI overlaid on a fractional anisotropy map from the same experiment. Even cursory inspection of this figure immediately reveals the cause for the excitement. The hope is that, by simply following the directions of fastest diffusion, it may be possible to reconstruct white matter pathways in the living human brain and, from there, to estimate anatomical connectivity in vivo for the first time.

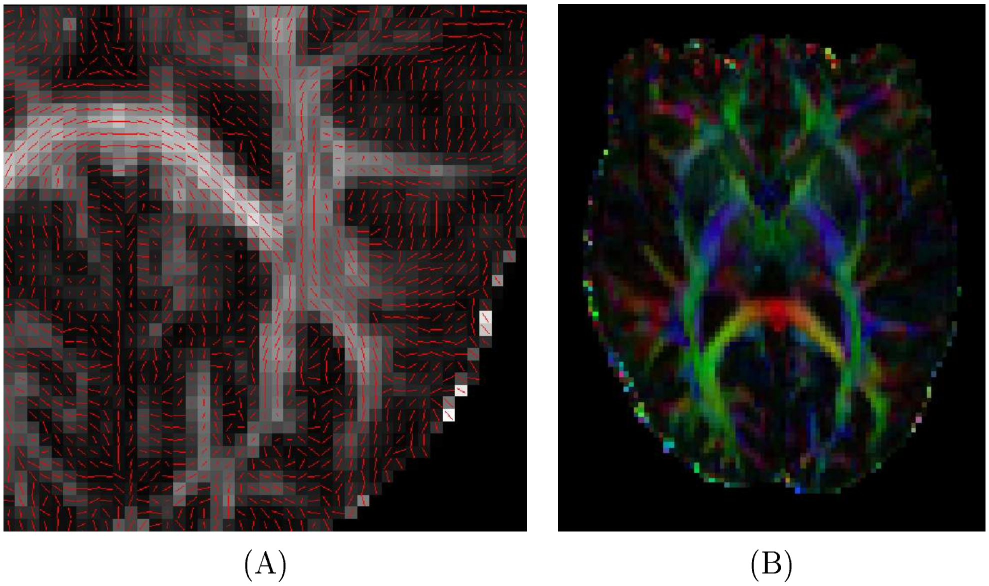


Figure 2.6: Principal diffusion directions from diffusion tensor imaging. (A) PDDs are projected into the slice plane and overlaid on fractional anisotropy in a normal brain. (B) PDDs are encoded xyz corresponding to red, green, blue and the intensity set by the fractional anisotropy.

Chapter 3

Bayesian Learning

3.1 Introduction

Learning is the process of adjusting a belief or understanding on the basis of new information. In this chapter we present a brief overview of how learning may occur in a Bayesian Framework, with particular reference to the techniques employed in this thesis. For a more extensive and detailed account of the topics covered in this chapter and many more, we refer the reader to (46; 47) and, for an excellent description of approximate and variational techniques, to (48) from where the next two paragraphs are essentially pinched.

3.2 Modelling physical systems - The generative likelihood.

When we encounter a complicated phenomenon, we seek to explain it by simplification. We break it down into pieces which we can understand, select those pieces that we believe to be important, then reassemble them to provide a simplified description. These building blocks, and their relationship in the overall design, represent a *model* of the original phenomenon; a simplification which enables us to understand its nature and, crucially, to make predictions about its behaviour in different situations.

Such models are of vital importance when we try to learn from data. Our real interest lies not in the data itself; but in specific features of the system which generated the data. In these circumstances, our model $\mathcal{M}$ encodes our assumptions about the overall *structure* of this system; but typically also contains parameters, Θ, relating to unknowns in the model which, when varied, vary the behaviour of the model. The model dictates how these parameters combine with one another to generate predictions of the data.

However, these predictions may not be deterministic. In the process of building the model, those blocks that were retained were those which were understood and regarded to be important contributors to the model's behaviour. Hence, many processes which contribute to the observed data are absent in the model. In statistical learning parlance, these processes are referred to as *noise* in the system, and are included in the model as *random variables*. That is, variables for which we may not make a deterministic prediction of any single realisation, but we may make predictions about the distribution from which that realisation was drawn.

The presence of noise in the system suggests that we should not be surprised if our model predicts data which does not match exactly the data which we observe, but we still should be surprised if the observed data is very different from the model's prediction. In fact, the distribution of the noise forms a key part of the model by telling us exactly how surprised we should be when we observe differences between predicted and observed data. We now have a model which predicts a distribution of possible realisations of the data for each choice of parameter set. This distribution is known as the *Generative Likelihood*: The probability of observing data, $\mathbf{Y}$, given the model, $\mathcal{M}$, and the true values of the model parameters, Θ,

$$\mathcal{P}(\mathbf{Y}|\Theta, \mathcal{M}). \tag{3.1}$$

This distribution provides a complete probabilistic description of the model. The process of learning on the model refers to the adjustment of our belief in the values of the model parameters, Θ, when confronted with new observations of the data $\mathbf{Y}$.

3.3 Maximum Likelihood Estimation

It seems as though the problem is solved! If we are interested in the values of the model parameters, we may simply pick those which make it most likely that the observed data was generated by our model. We may achieve this by choosing Θ to maximise the generative likelihood distribution in Equation 3.1.

$$\Theta_{ML} = \arg\max_{\Theta} \mathcal{P}(Y|\Theta, \mathcal{M}). \tag{3.2}$$

This strategy is known as maximum likelihood estimation, and it results in a *point estimate* for the parameter set Θ. That is, if we were to draw a n dimensional graph of our parameter space with each axis representing one parameter $\{\theta_1, \theta_2, \ldots \theta_n \in \Theta\}$, then the maximum likelihood estimate of Θ would exist only at a single point on the graph. At first, this seems entirely rational. If our model is good, then the "true" values of the parameters exist only at a single point on the graph. Surely, it may be supposed, when we estimate these parameters from the data, we should be aiming to estimate this single point. It is at this point that Bayesian learning techniques make their first conceptual leap towards the need for extremely expensive computers! The argument goes as follows; since we can never estimate exactly the true location of this point and, perhaps more importantly, since we cannot know how close to the truth our best estimate is, then we should not be trying to find an "estimate" of the parameters at all but rather a representation of our own belief in the parameter values.

3.4 Learning with densities - an application of Bayes' Law.

3.4.1 Probabilities and Densities

The most convenient mathematical tool for the representation of belief is the Probability Distribution or, equivalently in continuous space, the *Probability Density Function* (*pdf*).

If we have a random variable x in discrete space Ω_x^d, for example the result of a single trial of the roll of a die in the space $\{1, 2, 3, 4, 5, 6\}$, then x will take each value x_i in Ω_x with probability $p(x = x_i)$. When considered together, these probabilities form the *probability distribution* $p(x)$.

Now consider that x exists in continuous space Ω_x^c. The probability of x taking on any particular value, X, is infinitesimally small. In this case the distribution, or *probability density function*, $p(x)$ on x is defined such that the probability of the result of a trial lying between two values X_1 and X_2 is

$$p(X_1 \leq x \leq X_2) = \int_{X_1}^{X_2} p(x)dx \tag{3.3}$$

Crucial features of any distributed variable, whether continuous or discrete, are that it must take some value at each trial and it cannot take on more than a single value in any one trial. Hence the sum of all probabilities in the space must equal one.

$$\sum_{\Omega_x^d} p(x) = 1 \quad \text{Discrete} \tag{3.4}$$

$$\int_{\Omega_x^c} p(x)dx = 1 \quad \text{Continuous} \tag{3.5}$$

Probability distributions on random variables are truly frequentist objects. They represent the distribution of results if the same trial were carried out an infinite number of times. However, the Bayesian School of mathematics proposes a different interpretation of probabilities. Bayesians maintain that a probability encodes *belief* - whether it be in the predicted outcome of some future trial, or in the value of some hidden parameter. As such, probabilities and their density functions become the ideal tool with which to represent and manipulate beliefs when attempting to learn from data.

3.4.2 Joint and Conditional Densities

The aim of Bayesian Learning strategies is to reach a point where we may ask questions such as: For some hypervolume $\mathcal{V}$ in parameter space Ω_Θ , "What is our belief given the data that the true value of Θ is in $\mathcal{V}$?" or in one dimension: For any (θ_0, θ_1) "What is our belief that θ lies between θ_0 and θ_1?". However, in order to achieve this aim we require rules which will transform a *pdf* on the data given the parameters (the generative likelihood in section 3.2) into a *pdf* on the parameters of interest given the data.

Consider an experiment which results in two separate measurements, for example the rolling of a pair of dice where the measurements are the results of the two separate rolls. In order to describe the result of this experiment we need to use a vector of length two, $\mathbf{x} = [x_1 \; x_2]^T$, and in order to predict this result, we need a probability distribution which governs this vector: The *joint* probability distribution on $[x_1 \; x_2]^T$, or in continuous space the *joint pdf.* This joint distribution sets out the probability of **every** possible result of the experiment, so in the example given here

$$p(x_1, x_2) \equiv [p(x_1 = 1, x_2 = 1), p(x_1 = 1, x_2 = 2) \cdots p(x_1 = 6, x_2 = 6)]\,. \tag{3.6}$$

In this case, it is reasonable to suppose that the measurements have no effect on each other. They are said to be "independent", and computing the joint distribution is trivial.

$$p(x_1, x_2) = p(x_1)p(x_2) \tag{3.7}$$

So, for example, the probability that $x_1 = 4$ *and* $x_2 = 6$ is simply the probability that $x_1 = 4$, ($\frac{1}{6}$ in the unbiased case), multiplied by the probability that $x_2 = 6$ (again $\frac{1}{6}$)

giving $\frac{1}{36}$. In fact, we may say that two variables are independent *if and only if (iff)* equation 3.7 holds.

Now consider exactly the same experiment, but with slightly different measurements. This time, our first measurement will be the result of rolling die 1 and our second the sum of the results of the two dice rolls. It is clear that these two measurements are not independent. For example, if roll 1 results in a 6, then the sum of the two rolls cannot be less than 7. So how do we form the joint distribution on the two measurements?

The general solution to this problem was found independently by Bayes (49), and Laplace (50) and relies on the notion of *conditional probability*, the probability of an event B conditioned on (or *given*) the result of event A, $p(B|A)$. In our case A and B are measurements x_1 and x_2 and $p(x_2|x_1)$ is trivially computed. For example $p(x_2 = 5|x_1 = 2) = \frac{1}{6}$ as this would rely on rolling a 3 with our second die.

What Bayes and Laplace stated was the following:

$$p(A, B) = p(B|A)p(A) = p(A|B)p(B) \tag{3.8}$$

That is, the chance of seeing both events A and B is equal to the chance of seeing event B given that you've seen event A multiplied by the chance of seeing event A (and vice-versa). This simple and conceptually elegant statement has far reaching implications for machine learning, as the second equality may be rewritten into Bayes' law; the fundamental equation governing the adjustment of belief in the presence of new information.

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \tag{3.9}$$

3.4.3 Learning with Bayes' Law

Imagine that we are trying to discover whether there was a royal battle on British soil in 1066. At first, it seems pretty unlikely. Since the beginning of the English monarchy in AD 802 there have only been 23 battles involving the monarchy in 1202 years. the probability of one of these falling in 1066 is an incredibly small $\frac{23}{1202}$ - this probability is our *prior belief*, $p(A)$. However, searching deep into our memory, we recall that 1066 was the first year of the reign of William I. This information does not tell us directly what we want to know, but it is nevertheless useful. Battles often result in new kings. In fact, of the said 23 battles, 7 resulted in a change in the monarchy - we can say that the probability of a new king appearing given that there was a battle in that year is a much more healthy looking $\frac{7}{23}$ - this probability is the generative likelihood we see in section 3.2 or $p(B|A)$, the likelihood of observing the data (a new king) given what we wanted to know was true (there *was* a battle in 1066). In order to update our belief, we need one final piece of information. How likely is it that a new king would turn up at random, battle or no battle - For reasons to be explained later in this section, Bayesians call this probability, $p(B)$, the *marginal likelihood* of the data or equivalently the *evidence* for the model. It tells us how useful our data is; if new kings popped up year in year out, it wouldn't be very informative to see one in 1066, but if they only appeared after bloody battles, then seeing one in 1066 would tell us for sure that there was a battle in that year. The truth is somewhere between the two. There have in fact been 67 monarchs in the 1202 years of the monarchy, giving us a marginal likelihood of $\frac{67}{1202}$.

We are now in a position to update our belief in the battle of Hastings. Before we remembered about William's accession to the throne, we regarded it as incredibly unlikely that there was a battle in 1066 $p(A) = \frac{23}{1022}$, but with the additional information, albeit indirect, we are able to update our belief using Bayes' equation (3.9) to:

$$p(\text{Battle of Hastings} \mid \text{William's accession}) = \frac{\frac{7}{23} \times \frac{23}{1202}}{\frac{67}{1202}} = \frac{7}{67}. \tag{3.10}$$

Any new information we discover (e.g. that the new king wasn't the son of the previous monarch or that his surname was "The Conqueror"), may be treated in exactly the same fashion. We may continue to update our belief given every new piece of evidence, direct or circumstantial.

This example, although slightly silly, highlights a crucial point in machine learning, and particularly in Bayesian learning. We do not need to observe the parameters of interest directly. So long as we have, in the generative likelihood, a model for how our parameters will affect the observed data we may learn from the most circumstantial of observations.

3.4.4 Marginalisation

We saw in the previous section how Bayes' equation may help us to adjust our belief in an event. In machine learning it is more commonly used to update our belief in the values of model parameters, and as such, is more often written as follows:

$$p(\Theta|Y,\mathcal{M}) = \frac{p(Y|\Theta,\mathcal{M})p(\Theta|\mathcal{M})}{p(Y|\mathcal{M})} \tag{3.11}$$

In exactly the same fashion as with Harold and William (but now, in general, using continuous *pdfs*) we may update our belief from our *prior* expectations of the model parameter values $p(\Theta|\mathcal{M})$ using the data and its modelled relationship with the parameters $p(Y|\Theta,\mathcal{M})$ to give us our *posterior* distribution on the parameters $p(\Theta|Y,\mathcal{M})$. This posterior distribution represents our belief in the values of the model parameters *after we have considered the data* $\mathbf{Y}$.

However, yet again we face the problem of being short of one piece of information, and this time we cannot turn to our history books to find it out. We need to know the probability of seeing this data given the model no matter what values the parameters may take, $p(Y|\mathcal{M})$. Fortunately, where the history books have let us down, the texts on statistics come to our rescue with a useful feature of probability distributions:

$$p(B) = \int_{\Omega_B} p(A, B) dA \tag{3.12}$$

That is, if we have a joint distribution $p(A, B)$, then we may recover the distribution on B regardless of the value taken by A, $P(B)$, by simply integrating over all possible values of A. This process is known as *marginalisation* and the resulting distribution $p(B)$ is known as the *marginal* distribution on B. This result also holds when A and B are *sets* of parameters. In this case $p(B)$ is the *joint marginal* distribution on parameter set B.

We can tackle the problem of calculating our final piece of information, $p(Y|\mathcal{M})$, in exactly this fashion.

$$p(Y|\mathcal{M}) = \int_{\Omega_\Theta} p(Y, \Theta|\mathcal{M}) d\Theta = \int_{\Omega_\Theta} p(Y|\Theta, \mathcal{M}) p(\Theta|\mathcal{M}) d\Theta \tag{3.13}$$

Note that the integrand here is exactly the same as the numerator in Bayes' equation (3.11), ensuring that the posterior distribution integrates to unity in accordance with equation 3.4.

Unfortunately, calculating this integral is seldom straightforward. In general, there is no analytical solution to equation 3.13, and parameter space Ω_Θ is often very high dimensional, making standard numerical integration techniques computationally infeasible.

To make matters worse, the joint posterior distribution $p(\Theta|Y, \mathcal{M})$ is often not the distribution which we are really interested in. Models will often have "nuisance" or "latent" parameters, which add to the behaviour of the model, but which are not themselves of interest. In this case, we would like to calculate the joint marginal posterior distribution on the parameters of interest.

$$p(\Theta_{int}|Y\mathcal{M}) = \int_{\Omega_{\Theta_{latent}}} p(\Theta|Y\mathcal{M}) d\Theta_{latent} \tag{3.14}$$

again, an integral which is seldom tractable analytically.

The problem of estimation in a Bayesian framework is centred on the computation or approximation of the integrals in equations 3.13 and 3.14.

3.5 Estimation in a Bayesian Framework

A large number of methods have been proposed for solving this problem. A detailed explanation of many of them can be found in (46; 47; 51). Here, we concentrate only on the method used in this thesis, Markov Chain Monte Carlo sampling (MCMC). The description below draws heavily on that given in (52).

3.5.1 Sampling Techniques and MCMC

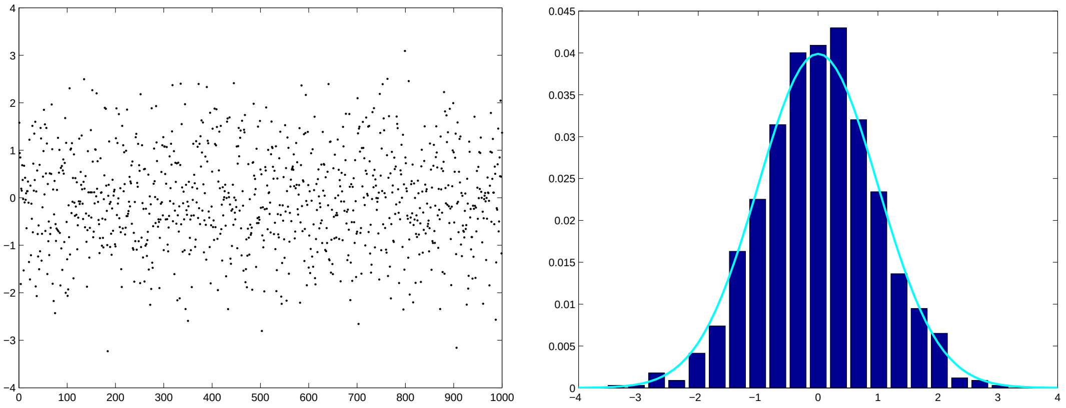
Sampling techniques provide a means for the *implicit* computation of the integrals in the previous section (equations 3.13 and 3.14). They allow us to draw samples of parameters Θ from the posterior distribution $p(\Theta|Y, \mathcal{M})$ without ever explicitly evaluating the normalising constant $p(Y|\mathcal{M})$. In common nomenclature, the techniques presented below allow us to draw samples from a distribution $p(x)$ by only ever evaluating $p^*(x)$, where

$$p(x) = \frac{p^*(x)}{Z} \quad . \tag{3.15}$$

These samples may be histogram-binned to build up the posterior distribution of interest (figure 3.1). However, drawing samples from a generic distribution is not a trivial process, particularly when an analytic form of the distribution in not available.

Rejection Sampling. Rejection sampling overcomes this problem by defining a proposal density $q(x)$ from which we *can* draw samples. If we can multiply $q(x)$ by some constant c such that $cq(x)$ envelopes x:

$$cq(x) > p^*(x) \quad \forall x, \tag{3.16}$$



(A) (B)

Figure 3.1: Building up distributions by sampling: (A) 1000 samples from a Gaussian distribution. (B) Histogram binning of the same samples.

then we may generate samples with uniform probability from the union of the two grey areas in figure 3.2. If we then reject samples with uniform probability from the light grey area, then the remaining samples will be independently uniformally distributed in the dark grey area and hence distributed in proportion to $p^*(x)$ according to $p(x)$. In order to achieve this, we draw a random sample, x, from $q(x)$. We then

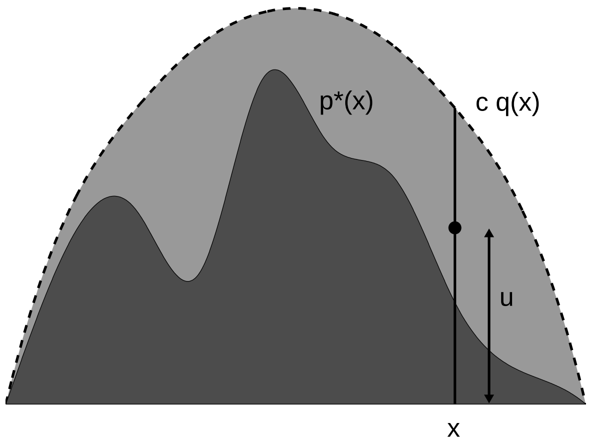


Figure 3.2: Schematic demonstrating the principal of area based rejection sampling. Samples from the proposal density are uniformally distributed in the union of the two grey areas and rejected with uniform probability in the light grey area. Accepted samples are then uniformally distributed within the dark grey area.

evaluate $cq(x)$ and draw a second random variable, u, from a uniform distribution in the range $[0, cq(x)]$. If $u > p^*(x)$ then x is attributed to the light grey area and rejected.

Otherwise, it is attributed to the dark grey area and accepted.

Rejection sampling is only efficient if the envelope function is a good approximation to $p^*(x)$. In figure 3.2 the ratio of rejected to accepted samples is the same as the ratio of light to dark grey areas; we accept only $\frac{1}{c}$ of the proposed samples. Even when sampling from a single parameter as here, it can be hard to find an enveloping function which can be sampled from and which gives a good acceptance ratio. When the dimensionality of the sampling becomes large, this problem may be exacerbated to the stage that hundreds of thousands of samples are drawn before a single one is accepted. To overcome this problem, sampling strategies were developed which concentrate the sampling on areas of high probability. Such strategies are known as Markov Chain Monte Carlo (MCMC) sampling techniques, and the two used in this thesis are Metropolis-Hastings (MH) and Gibbs sampling.

MCMC by Metropolis-Hastings. The key insight of Markovian sampling techniques is to allow the proposal density for the next sample to depend upon the value of the last sample, or in Markovian parlance "The current state of the Markov Chain". $q(x)$ becomes $q(x'; x^{(t)})$, where x' is the new proposal and $x^{(t)}$ is the current state.

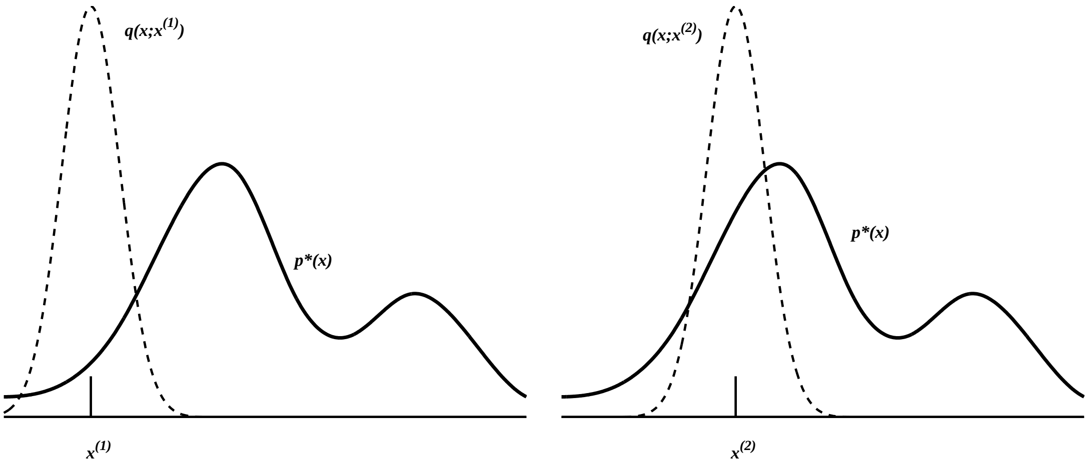


Figure 3.3: Successive jumps of the MH algorithm in one dimension.

The MH proposal density may have any form although, in this thesis, Gaussian distributions centred on the current state are always chosen. Figure 3.3 shows the proposal densities for two successive jumps of the Markov chain. A new state, x' is proposed from the proposal density and accepted with probability:

$$p_{accept} = \min\left(1, \frac{p^*(x')q(x';x^{(t)})}{p^*(x^{(t)})q(x^{(t)};x')}\right) \tag{3.17}$$

If the step is accepted, the chain moves to x', $x^{(t+1)} = x'$, otherwise the chain stays at $x^{(t)}$, $x^{(t+1)} = x^{(t)}$. Note that a rejected step results in the recurrence of the previous step in the list of accepted samples.

If $q(x';x^{(t)})$ is chosen as a positive function, i.e:

$$q(x';x^{(t)}) > 0 \quad \forall x' \quad , \tag{3.18}$$

it can be shown (53) that the distribution of accepted samples will converge on $\frac{p^*(x)}{Z}$, the distribution of interest.

Note that, unlike in rejection sampling, the accepted states in any Markovian sampling technique are not independent; they are autocorrelated. However, also unlike in rejection sampling, the sampling efficiency need not depend on the dimensionality of the sampling problem. In high dimensional cases, either the proposal density may be narrowed until a satisfactory acceptance rate is obtained or, in fact, the different dimensions may be sampled from successively, each with their own proposal density, effectively reducing the multi-dimensional problem to a series of one dimensional problems. This second strategy is referred to as Metropolis-Hastings by single component jumps (47).

Gibbs Sampling In multi-dimensional problems, if it is possible from the *conditional* density on a parameter or parameter set; i.e. if it is possible to sample from $p(x|.)$, where . represents the data, the model and every parameter in the model with the exception of x, then the choice of proposal density in the MH algorithm may be avoided completely. Gibbs sampling is simply MH single component sampling where the proposal density for each parameter x_i is the conditional density on x_i. Under these conditions, substitution into equation 3.17 reveals that *every proposed sample* is accepted.

Returning to Bayesian nomenclature, we may derive the full conditional distribution on any parameter from the joint distribution of all parameters and data, $p(\Theta, Y|\mathcal{M}) = p(Y|\Theta, \mathcal{M})p(\Theta|\mathcal{M})$, the numerator in Bayes' equation. The full conditional for component i is given by:

$$P(\theta_i|\theta_{-i}, Y, \mathcal{M}) = \frac{P(Y, \Theta|\mathcal{M})}{P(Y, \theta_{-i}|\mathcal{M})} \tag{3.19}$$

where, θ_{-i} refers to all parameters with the exception of θ_i, and because the denominator does not depend on θ_i this becomes:

$$P(\theta_i|\theta_{-i}, Y, \mathcal{M}) \propto P(Y, \theta, \mathcal{M}) \tag{3.20}$$

Hence, to obtain the full conditional, we may write down the full joint distribution $P(\theta, Y, \mathcal{M})$ as a product of one or more factors and pick out the terms which involve θ_i. If it is possible to sample from the unnormalised full conditional $p^*(\theta_i|\theta_{-i}, Y, \mathcal{M})$ then Gibbs sampling is also possible. In fact it is common practice to choose prior distributions which are *conjugate* to the generative likelihood (see (46)) to ensure that it is possible to sample from the full conditional.

Chapter 4

Uncertainty in Diffusion Imaging

Uncertainty and its representation have an important role to play in any situation where the goal is to infer useful information from noisy data. In Diffusion Weighted MRI (DWMRI) scientists attempt to infer information about, for example, diffusion anisotropy or underlying fibre tract direction, by fitting models of the diffusion and measurement processes to DWMRI data (e.g. (35; 38)). In this scheme there is uncertainty caused both by the noise and artifacts present in any MR scan, but also by the incomplete modelling of the diffusion signal. That is, the true diffusion signal is more complicated than we choose to model. This additional complexity in the diffusion signal appears as residuals when we fit a simple model to the data, causing additional uncertainty in the model parameters. All of the uncertainty in these parameters may be represented in the form of probability density functions (*pdfs*).

In this chapter, we show how the Bayesian estimation techniques described in chapter 3 may be applied to models of the molecular diffusion process, as measured by diffusion weighted MRI. The approaches described here are applicable to and useful with *any* model which might be used in diffusion imaging. In fact, even in seemingly "model-free" environments such as q-space imaging (see chapter 2), there is uncertainty associated with the recovered estimates, and Bayesian approaches could usefully be applied to represent and understand its effect.

The distributions on parameters in a diffusion model are of great significance when making inference on the basis of these parameters. Inference may be at a group level; for example there have been studies showing reduced anisotropy in groups of Multiple Sclerosis patients, in comparison with groups of normal subjects (e.g. (39)). However, inference may also be within a single subject. There have been many recent papers (e.g. (3; 5; 4)) describing techniques for using parameters from a diffusion tensor fit to follow major white matter pathways in the brain. In chapter 5 we will show how the distributions derived in this chapter may be used to examine the effect of *local* uncertainty on the location of these *global* pathways.[1]

4.1 Models of Diffusion

In this section we present 3 models of the local diffusion process. The first is the familiar diffusion tensor model (2), which models the local diffusion as a 3 dimensional Gaussian. Then we choose two different models which attempt to model *underlying fibre structure* in a voxel and, from this, predict the diffusion weighted signal. The first of these is a simple partial volume model allowing for a single fibre direction mixed with an isotropically diffusing compartment in a voxel. The second is a parametrised model of the transfer function between a distribution of fibre orientations in a voxel and the measured diffusion weighted signal. We infer on the first two of these models, using MCMC to estimate the posterior distributions on parameters of interest. We present detailed results from a single white matter voxel showing recovered distributions from both models. We go on to present a validation study, comparing distributions throughout a slice with those recovered from empirical measurements of uncertainty (55).

---

[1] A preliminary version of the material in this and the following chapter has been published in *Magnetic Resonance in Medicine*(12)

4.1.1 The Diffusion Tensor Model

The diffusion tensor has often been used to model local diffusion within a voxel (e.g. (2; 56; 57)). The assumption made is that local diffusion may be characterised with a 3 Dimensional Gaussian distribution (2), whose covariance matrix is proportional to the diffusion tensor, $\mathbf{D}$. The resulting diffusion weighted signal, μ_i along a gradient direction $\mathbf{r}_i$, with b-value b_i is modelled as:

$$\mu_i = S_0 \exp\left(-b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i\right), \tag{4.1}$$

where S_0 is the signal with no diffusion gradients applied. $\mathbf{D}$, the diffusion tensor is:

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} \tag{4.2}$$

When performing point estimation of the parameters in the diffusion tensor model, it has been convenient to choose the free parameters in the model to be the 6 independent elements of the tensor, $D_{xx} - D_{zz}$, and the signal strength when no diffusion gradients are applied, S_0. This parametrisation allows estimation to take the form of a simple least squares fit to the log data. When sampling, however, our choice of parametrisation is far less constrained by our estimation technique. The parameters of real interest in the tensor are the three eigenvalues, and the three angles defining the shape and orientation of the tensor. By choosing *these* as the free parameters in the model , not only do we give ourselves immediate access to the posterior *pdf*s on the parameters of real interest, but we also allow ourselves the freedom to apply constraints or add information exactly where we would like to. As a simple example, as will be seen later, a sensible choice of prior distribution on the eigenvalues makes it easy to constrain them to be positive. So the Diffusion Tensor is now parametrised as follows:

$$\mathbf{D} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\mathbf{T}}, \tag{4.3}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \tag{4.4}$$

and $\mathbf{V}$ rotates $\mathbf{\Lambda}$ to (θ, ϕ, ψ), such that the tensor is first rotated so that its principal eigenvector aligns with (θ, ϕ) in spherical polar coordinates, and then rotated by ψ *around* its principal eigenvector[2].

The noise is modelled separately for each voxel as independently identically distributed (*iid*) Gaussian. with a mean of zero and standard deviation across acquisitions of σ. The probability of seeing the data at each voxel $\mathbf{Y}$ given the model, M, and any realisation of parameter set, $\omega = (\theta, \phi, \psi, \lambda_1, \lambda_2, \lambda_3, S_0, \sigma)$ may now be written as:

$$\begin{aligned} \mathcal{P}(\mathbf{Y}|\omega, M) &= \prod_{i=1}^{n} \mathcal{P}(y_i|\omega, M) \\ \mathcal{P}(y_i|\omega, M) &\sim \mathcal{N}(\mu_i, \sigma) \end{aligned} \tag{4.5}$$

where n is the number of acquisitions, and y_i and μ_i are the measured and predicted values of the i^{th} acquisition respectively. (Note that throughout this chapter, i will be used to index acquisition number).

$$\mu_i = S_0 \exp -b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i. \tag{4.6}$$

Thus, the model at each voxel has 8 free parameters each of which is subject to a prior distribution. Priors are chosen to be non-informative, with the exception of

---

[2] Note: This may seem an odd way to span the angular space. The reason we chose to define these angles, is that it allows us to sample directly from the principal diffusion direction (θ,ϕ)

ensuring positivity where sensible[3].

$$
\begin{aligned}
\mathcal{P}(\theta,\phi,\psi) &\propto \sin(\theta) \\
\mathcal{P}(S_0) &\sim \mathcal{U}(0,\infty) \\
\mathcal{P}(\lambda_1) = \mathcal{P}(\lambda_2) = \mathcal{P}(\lambda_3) &\sim \Gamma(a_\lambda, b_\lambda) \\
\mathcal{P}(\frac{1}{\sigma^2}) &\sim \Gamma(a_\sigma, b_\sigma) \qquad (4.7)
\end{aligned}
$$

Parameters a and b in the Gamma distributions are chosen to give these priors a suitably high variance such that they have little effect on the posterior distributions except for where we ensure positivity. Note that the non-informative prior in angle space is proportional to $sin(\theta)$ ensuring that every elemental area on the surface of the sphere, $\delta A = sin(\theta)\delta\theta\delta\phi$ has the same prior probability.

4.1.2 A Simple Partial Volume Model.

Here we take a slightly different approach to modelling in DWMRI. Instead of modelling the diffusion shape directly, we attempt to build a model of the underlying fibre structure which predicts the diffusion shape, and hence the MR measurements. The simplest such model of fibre structure is to assume that all fibres pass through a voxel in the same direction. Assuming no diffusion-diffusion exchange, this leads to a simple two compartment partial volume model. The first compartment models diffusion in and around the axons, with diffusion only in the fibre direction. The second models the diffusion of free water in the voxel as isotropic. One consequence of this model is that the diffusivity (and hence the restriction to water diffusion) in all directions perpendicular to the fibre axis is constrained to be the same. This is very different to the Diffusion Tensor model, where any ellipsoidal diffusion shape may be modelled.

---

[3] A description of the Γ distribution may be found in the appendix at the end of this chapter

The predicted diffusion signal is

$$\begin{aligned} \mu_i &= S_0((1-f)\exp(-b_i d) \\ &\quad +f\exp(-b_i d \mathbf{r_i}^T \mathbf{R}\mathbf{A}\mathbf{R}^T \mathbf{r_i})) \end{aligned} \tag{4.8}$$

where d is the diffusivity, b_i and $\mathbf{r_i}$ are the b-value and gradient direction associated with the i^{th} acquisition, f and $\mathbf{RAR}^T$ are the fraction of signal contributed by, and anisotropic diffusion tensor along, the fibre direction (θ, ϕ). That is $\mathbf{A}$ is fixed as:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4.9}$$

and $\mathbf{R}$ rotates $\mathbf{A}$ to (θ, ϕ):

Again noise is modelled as *iid* Gaussian:

$$\begin{aligned} \mathcal{P}(\mathbf{Y}|\omega, M) &= \prod_{i=1}^{n} \mathcal{P}(y_i|\omega, M) \\ \mathcal{P}(y_i|\omega, M) &\sim \mathcal{N}(\mu_i, \sigma), \end{aligned} \tag{4.10}$$

where the parameter set ω now has 6 free parameters $(\sigma, S_0, d, f, \theta, \phi)$. Each of these parameters is subject to a prior distribution, which is chosen to be non-informative except for where we ensure positivity:

$$\begin{aligned} \mathcal{P}(\theta, \phi) &\propto \sin(\theta) \\ \mathcal{P}(S_0) &\sim \mathcal{U}(0, \infty) \\ \mathcal{P}(f) &\sim \mathcal{U}(0, 1) \\ \mathcal{P}(d) &\sim \Gamma(a_d, b_d) \\ \mathcal{P}(\frac{1}{\sigma^2}) &\sim \Gamma(a_\sigma, b_\sigma). \end{aligned} \tag{4.11}$$

4.1.3 Increasing the Complexity - A Distribution of Fibres?

In the partial volume model presented above, only a single fibre orientation is modelled in each voxel. In fact, there will be a *distribution*, $H(\theta,\phi)$, of fibre orientations in the voxel. In order to estimate this distribution we must build a model which, given this distribution, could predict the Diffusion Weighted MR measurements.

Such a model clearly requires some assumptions. We start by assuming that each subvoxel has only one fibre direction through it and that the MR signal is the sum of the signal from arbitrarily small subvoxels, and that the signal from each *subvoxel* behaves as described by Equation 4.8. (Note that this is a strong assumption to make, but it is *explicit* in the model. Any other model of the local diffusion characteristics of a single fibre orientation may be used as a replacement.)

$$\mu_{total} = \sum_{j \in sub-voxels} \mu_j \tag{4.12}$$

where μ_{total} is the vector of MR signal from the voxel at each gradient direction and strength, and μ_j is the same vector for each sub-voxel.

If we now consider, instead of the individual sub-voxels, the set $\Theta\Phi$ of major directions (θ,ϕ) in these subvoxels (note the discretisation of $\Theta\Phi$), then Equation 4.12 is identically equivalent to (see equation 4.8):

$$\begin{aligned}\mu_i = \sum_{(\theta,\phi)\in\Theta\Phi} \Bigg(\sum_{j\in V_{\theta\phi}} \frac{S_{0_j}}{N} \big[(1-f_j)\exp(-b_i d_j) + \\ f_j \exp(-b_i d_j \mathbf{r}_i^T \mathbf{R}_{\theta\phi} \mathbf{A} \mathbf{R}_{\theta\phi}^T \mathbf{r}_i) \big] \Bigg)\end{aligned} \tag{4.13}$$

where $V_{\theta\phi}$ is the set of all voxels whose principal fibre direction is (θ,ϕ) and N is the number of subvoxels. This equation, although fearsome at first sight, is actually very straight forward. The first part of the argument to the summation (on the top line) represents the signal due to all of the isotropic compartments, and the second part represents the signal due to all of the fibre compartments. If we now further assume that S_0 (the signal with no diffusion gradients applied) and d (the diffusivity) are constant across the voxel, then the inner summation (over subvoxels which have the same principal direction) may be replaced by a constant for the isotropic compartment, and in the anisotropic compartments, by the distribution function $H(\theta, \phi)$ defined earlier. With a little more manipulation it is easy to arrive at:

$$\begin{aligned} \frac{\mu_i}{S_0} &= (1-f)\exp\left(-b_i d\right) + \\ f \sum_{(\theta,\phi)\in\Theta\Phi} &\left(H(\theta,\phi)\exp\left(-b_i d \mathbf{r}_i^T \mathbf{R}_{\theta\phi} \mathbf{A} \mathbf{R}_{\theta\phi}^T \mathbf{r}_i\right)\right) \end{aligned} \tag{4.14}$$

If we then let the sub-voxel size tend to zero we may write:

$$\begin{aligned} \frac{\mu_i}{S_0} &= (1-f)\exp\left(-b_i d\right) + \\ f \int_0^{2\pi}\int_0^{\pi} &\mathcal{H}(\theta,\phi)\exp\left(-b_i d \mathbf{r}_i^T \mathbf{R}_{\theta\phi} \mathbf{A} \mathbf{R}_{\theta\phi}^T \mathbf{r}_i\right)\sin(\theta)d\theta d\phi. \end{aligned} \tag{4.15}$$

where $1-f$ is now the proportion of the whole voxel showing isotropic diffusion. Note that the integral is over $\sin(\theta)d\theta d\phi$ in order to maintain elemental area over the sphere. Finally, if we write the gradient direction $\mathbf{r}_i$ in spherical polar coordinates $\mathbf{r}_i = \left[\begin{array}{ccc} \sin\alpha_i \cos\beta_i & \sin\alpha_i \sin\beta_i & \cos\alpha_i \end{array}\right]$, and define γ_i as the angle between gradient direction, (α_i, β_i), and fibre direction (θ_i, ϕ_i), then the exponent inside the integral reduces dramatically. We may now write:

$$\begin{aligned} \frac{\mu_i(\alpha_i,\beta_i)}{S_0} &= (1-f)\exp\left(-b_i d\right) + \\ f \int_0^{2\pi}\int_0^{\pi} &\mathcal{H}(\theta,\phi)\exp\left[-b_i d \cos^2\gamma_i\right]\sin(\theta)d\theta d\phi. \end{aligned} \tag{4.16}$$

This equation reveals a great deal about the diffusion measurement process. The real "signal" of interest is $\mathcal{H}(\theta, \phi)$, the distribution of fibres within the voxel. When we measure the diffusion profile of this signal, we are measuring a version of this signal which is smoothed in angular space, with a kernel, predicted by this model, of $\exp\left(-bd\cos^2\gamma\right)$. We would like to deconvolve the effect of the measurement process from the signal. If we choose to infer on this model in the Bayesian framework, the resulting posterior distribution of interest will be a distribution on parameters which describe the fibre orientation distribution $\mathcal{H}(\theta, \phi)$. However, we leave the details of this estimation process, and validation thereof, as future work.

4.2 Local Parameter Estimation: Methods

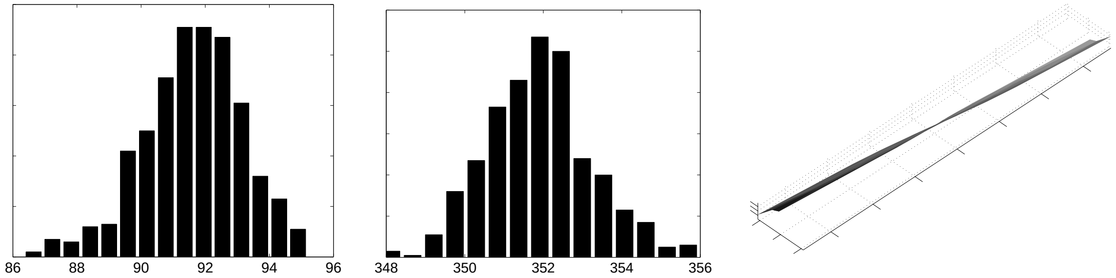
Data acquisition DT-MRI datasets were acquired on a single male healthy volunteer. The images were obtained on a 3.0 T Varian Inova scanner using a diffusion-weighted single-shot EPI sequence. To minimise eddy currents, a doubly-refocused spin-echo sequence was implemented (58) and diffusion-weighted images were subsequently registered to the same non-diffusion-weighted reference scan using affine registration (59). A birdcage radio-frequency head coil was used for both pulse transmission and signal detection. The diffusion gradients achieved a maximum gradient strength of 22 mTm^{-1}. Each data set consisted of 3 non-diffusion-weighted and 60 diffusion-weighted images acquired with a b-value of 1000 smm^{-2}. The diffusion gradients were uniformly distributed through space using the optimised scheme proposed by Jones (33). Each set of images contained 42 contiguous slices with a 2.5 mm thickness. A half k-space acquisition was performed with a matrix size set to 62×96 and a field of view of $240 \times 240 mm^2$. The images were interpolated to achieve a matrix size of 128×128 and a final resolution of $1.875 \times 1.875 \times 2.5 mm^3$. To minimise motion artifacts, peripheral gating was used such that triggering occurred on every cardiac cycle. The echo time was set to 106 ms while the effective repetition time was 14 R-R intervals. The total scan time for each dataset was approximately 15 minutes, depending on heart rate.

Estimation MCMC estimation was performed for the diffusion tensor model and for the simple partial volume model. In both cases parameters were initialised with a least squares diffusion tensor fit. The Markov Chains were then jumped 500 times without sampling as a “burn-in” (see (53)), followed by 2000 times, sampling every second jump, to give 1000 samples. A single jump of the parameter set consisted of independent jumps of each parameter. In both models samples were drawn from the precision ($\frac{1}{\sigma^2}$) with a Gibbs sampler, and from all other parameters with Metropolis Hastings samplers. Proposal distributions for Metropolis Hastings parameters were zero mean Gaussians with standard deviations tuned adaptively to give a jump acceptance rate of 0.5. The full conditional distributions for the Gibbs sampling of the precision in both models are given in the appendix at the end of this chapter. Computation time for diffusion data with 63 acquisitions is approximately 0.3 seconds per voxel on a Pentium IV 2GHz. Voxels are processed independently, so computation is easily parallelised.

4.3 Local Parameter Estimation: Results

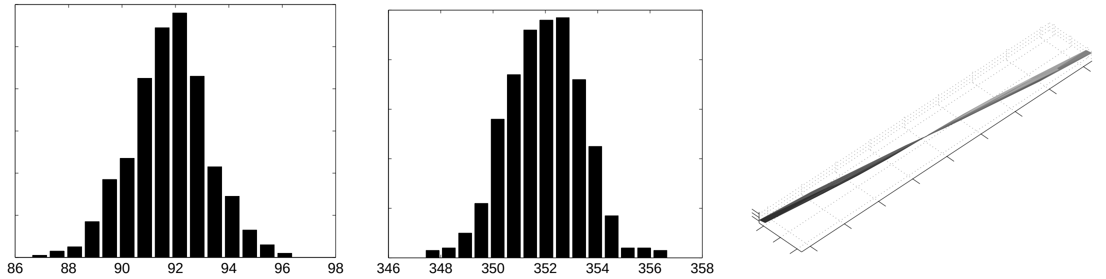
Example distributions from a single voxel. Figure 4.1 (a) and (b) show samples from the marginal posterior distributions on θ and ϕ from the *diffusion tensor* model. The voxel was chosen from the splenium of the Corpus Callosum. (c) shows (a) and (b), plotted as a joint histogram around the surface of a sphere. This is then the joint marginal posterior distribution of θ and ϕ or the marginal posterior distribution of principal diffusion direction (PDD). Note how narrow this distribution is. This represents a high confidence in our calculated PDD, which is as predicted in an area of dense white matter such as the corpus callosum. Figure 4.2 (a) and (b) show samples from the marginal posterior distributions on θ and ϕ from the simple *partial volume* model. The same voxel was chosen as in Figure 4.1. Again (c) shows (a) and (b), plotted as a joint histogram around the surface of a sphere.

Validation: Comparison with empirical measurements. The posterior *pdf*s on the parameter estimates, in either of the above models, characterise our uncertainty in these parameters. In (55), Jones proposes an empirical method for estimating this uncertainty. Following this method, we acquired 3 repeats of diffusion data with 63 gradient directions and bootstrapped, to create 1000 datasets of different combinations



(a) (b) (c)

Figure 4.1: Samples from marginal posterior distributions of the diffusion tensor model in a white matter voxel. (a) shows samples from the marginal posterior distribution on θ. (b) shows samples from the marginal posterior distribution on ϕ. (c) shows (a) and (b) plotted around a sphere, representing the marginal posterior distribution on principal diffusion direction.



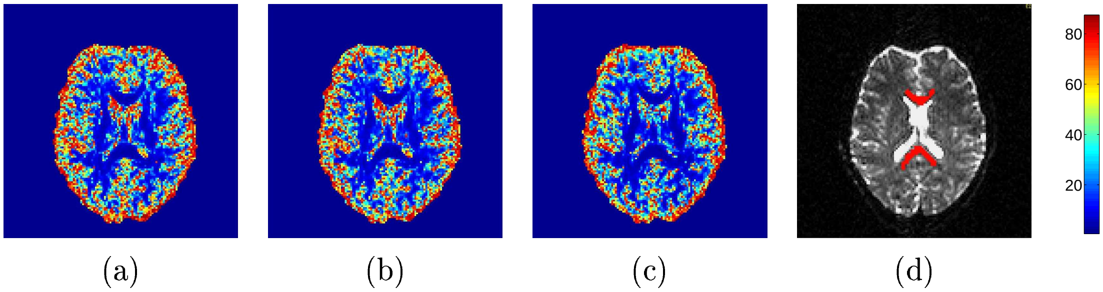
(a) (b) (c)

Figure 4.2: Samples from marginal posterior distributions of the partial volume model in a white matter voxel. (a) shows samples from the marginal posterior distribution on θ. (b) shows samples from the marginal posterior distribution on ϕ. (c) shows (a) and (b) plotted around a sphere, representing the marginal posterior distribution on principal diffusion direction.

of these repeats. We fit a diffusion tensor at each voxel in each of these new datasets, and calculated the uncertainty between the 1000 principal eigenvectors at each voxel. This uncertainty is measured as the size of the 95% confidence angle from the mean direction.

Using only *one* of these 1000 datasets we drew 1000 samples from the posterior *pdf* on principal diffusion direction at each voxel under both the diffusion tensor, and simple partial volume models. From these samples, we computed the same 95% angle from the mean direction.

Figure 4.3 shows these 95% angles for the diffusion tensor model in (a), and the partial volume model in (b). (c) shows the same angles predicted by Jones' method.



(a) (b) (c) (d)

Figure 4.3: 95% uncertainty values (in degrees), predicted from the Diffusion Tensor model (a), the partial volume model (b), and Jones' empirical method (c). (d) shows a mask of the corpus callosum used in some calculations

There are various factors to consider when comparing these results. The first is that the empirical method in (c) is not necessarily "ground truth". It has errors associated with it due to subject motion and interpolation related effects, but also, more subtly, due to the dependence within the bootstrapped datasets. This is likely to cause an *underestimate* in the measured uncertainty. The second factor is the difference in the two models. (a) and (c) predict uncertainty levels in the principal eigenvector of a diffusion tensor model. (b) predicts the same thing for the less flexible partial volume model. In areas of complex fibre structure, the partial volume model, which has only one fibre direction available to it, is forced to represent this structure as uncertainty in the single direction. (This will turn out to be extremely useful when trying to do tractography, as will be seen in chapter 5.) In contrast, the diffusion tensor model will tend to account for complex fibre structure in a voxel not only with uncertainty in the principal fibre direction, but also with a change in the diffusion profile (i.e. a change in the relative sizes of the three eigenvectors.) For this reason we would predict that, in regions of complex fibre structure, the partial volume model would show more uncertainty in principal diffusion direction than the diffusion tensor model. We would expect the two models to predict very similar uncertainties in regions of high fibre co-alignment, such as in the corpus callosum (Figure 4.3 (d)).

The mean 95% confidence angles within the brain for the three methods are: diffusion tensor model and MCMC (a) 35.4°, partial volume model and MCMC (b) 36.0° and diffusion tensor model with empirical measurements (Jones)(c) 33.9°. We further compare any two of these three methods by computing their absolute difference as a fraction of their mean value at every voxel, defining fractional deviation (Table 1):

$$FD(A,B) = FD(B,A) = 2 \times \frac{|A-B|}{A+B}$$

	A	B	C
A		0.10 (0.08)	0.15 (0.11)
B	0.10 (0.08)		0.20 (0.14)
C	0.15 (0.11)	0.20 (0.14)	

	A	B	C
A		0.04 (0.03)	0.13 (0.11)
B	0.04 (0.03)		0.12 (0.09)
C	0.13 (0.11)	0.12 (0.09)	

Table 1:Fractional Deviation values between the three methods in the whole brain (left) and within the Corpus Callosum (right). Inside each cell is the mean with the median in parentheses.

Predictions of uncertainty by MCMC on the two models are within 10% of each other throughout the brain, and within 5% in the callosal mask, showing, as predicted, very similar uncertainty where fibres are highly co-aligned, and slight differences in uncertainty in other areas. With the diffusion tensor model, uncertainties predicted by MCMC are within 15% of those predicted by the empirical method when considering the whole brain, and 13% when only considering the corpus callosum. These differences are small and may be due to errors in either or both approaches.

4.4 Discussion

In general, analysis of diffusion weighted data has involved the fitting of a model of local diffusion to the diffusion weighted data at each voxel. This model may assume that local diffusion is Gaussian in profile (the *Diffusion Tensor* model (2)) or may allow a more complex structure for local diffusion (e.g. a spherical harmonic decomposition (60; 61)). However, in all cases, the assumed model is of the *diffusion profile* and not of the underlying fibre structure, and any analysis which has occurred after the fitting of this local model has made the assumption that the parameters in this model are known absolutely.

There are two important, but separate issues here. The first, is that the parameters of real interest to the scientist are ones which relate directly to the underlying fibre structure, and not to the diffusion profile. These underlying parameters may have convincing markers within the fitted diffusion profile (for example anisotropy measures (62; 38) from the diffusion tensor fit have been shown to be a marker for collinearity of fibres within a voxel), but any attempt to recreate the fibre structure from these profiles is essentially an educated guess. There has been no model proposed to predict how a specific structure or distribution of fibre directions within a voxel will reflect itself in the measured diffusion weighted NMR signal. The second issue is that, even when fitting a model of local diffusion, the resulting parameters have uncertainty associated with them. Factors such as noise in the NMR signal (both physical and physiological) and, crucially, the inadequacy of the proposed model, lead to this uncertainty which should be incorporated in any further processing (such as tractography schemes).

In this chapter we have presented a method for the full treatment of this uncertainty. We have shown how, using Bayes' equation along with well established methods for its numerical solution, it is possible to form a complete representation of the uncertainty in the parameters in *any* generative model of diffusion, in the form of posterior probability density functions on these parameters. We have applied this Bayesian estimation technique to two simple local models of diffusion, the diffusion tensor model, and a simple partial volume model, with only a single anisotropically diffusing direction in the voxel. We have examined the results in these two cases, comparing the posterior distributions with empirical measurements of uncertainty.

However, MCMC estimation gives us a great deal of flexibility in our choice of model. In section 4.1.3, we introduce the idea of modelling a distribution of fibres, or a fibre structure within a voxel. The additive nature of the NMR signal allows us, with certain assumptions, to make predictions of the diffusion weighted signal, given knowledge of the underlying fibre structure. Effectively, we have proposed a parametrised transfer function between true fibre structure and diffusion signal. By proposing parametrised models of fibre structure it should be possible to estimate simultaneously the posterior *pdf*s on the parameters in this transfer function, and on those in the fibre model itself. This estimation procedure is left for further work, but the theory is laid out in section 4.1.3.

The issue of uncertainty in principal diffusion direction has been raised before. As we saw in section 4.1.2, in (55), Jones proposes a method for the empirical measurement of this uncertainty. This method involves acquiring a number of repeats (e.g. 3) of the complete acquisition sequence. Each repeat has a number (e.g. 60) of measurements, each acquired with different diffusion gradient directions. By bootstrapping these repeats, it is then possible to create many, (e.g 1000), new data sets, each of which is a different combination of the original repeats. By measuring the variability in principal diffusion directions between these new datasets, it is possible to estimate the uncertainty in this PDD in any one of them. Jones' method benefits from making no distributional assumptions about the noise in the image. In our method, we make the assumption that the profile of image noise *between acquisitions* is Gaussian with the *pdf* on standard deviation σ estimated at each voxel (see, for example, equation 4.10). By taking many repeats of the same measurements, Jones is able to avoid the need for such an assumption. However, there are also disadvantages with Jones' method. Firstly, and most clearly, it requires more acquisition time. In order to collect 3 repeats of the same data, 3 times as much time is needed. Secondly, subject motion and registration related interpolation effects are likely to cause small errors in the estimated uncertainty. Thirdly, and most importantly, the 1000 datasets from which the uncertainty is estimated are not independent. They are all created from a small number of original repeated acquisitions. The effect of this is likely to be an *underestimate* of the uncertainty on PDD. This issue is not fully explored in (55). In spite of these differences, in section 4.1.2 we show a pleasing correspondence between the uncertainties predicted by the two approaches.

This chapter has highlighted and distinguished three separate distributions which are of interest to the diffusion imaging experimenter. The first is the well known diffusion propagator: The probability distribution on molecular displacement in a voxel that is modelled by a Gaussian in DTI experiments. The second is the underlying distribution of fibre orientations within the voxel. This is the distribution which best describes the local fibre architecture, but it is obscured from us by the diffusion process with which we measure it. The third is the conceptual distribution describing the uncertainty that we must encounter if we choose to measure either of the physical distributions described above. This third distribution will prove to be of crucial significance in the following chapters when we use the parameters derived from diffusion imaging to trace white matter pathways through the diffusion data field. By maintaining knowledge of this distribution at each voxel, we are able to generalise diffusion tractography to the situation where the data is imperfect, increasing the robustness and interpretability of the tractography process.

4.5 Appendix

4.5.1 Gamma Distribution

x has a two-parameter gamma distribution, denoted by $\Gamma(a, b)$, with parameters a and b, if its density is given by:

$$\mathcal{P}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \tag{4.17}$$

where $\Gamma(a)$ is the Gamma function. A χ^2 distribution with ν degrees of freedom corresponds to the distribution $\Gamma(\nu/2, 1/2)$. The b parameter is a scale parameter. The one-parameter gamma distribution corresponds to $\Gamma(a, 1)$. A sample from $Ga(a, b)$ can be obtained by taking a sample from $\Gamma(a, 1)$ and dividing it by b. Note, that a gamma distribution has $mean = a/b$ and $variance = a/b^2$

4.5.2 Full conditional distibution for precision parameters

The full conditional distribution for Gibbs sampling from the precision parameters $\frac{1}{\sigma^2}$ in both models is

$$\mathcal{P}(\frac{1}{\sigma^2}|\mathbf{Y},\Omega_-) = \Gamma(a+\frac{n}{2}, b+\frac{1}{2}\sum_{i=1}^{n}(Y_i-\mu_i)^2), \tag{4.18}$$

where, $\mathbf{Y}$ is the data, Ω_- is the set of all parameters except σ, n is the number of acquisitions, Y_i is the value of the data at the i^{th} acquisition, a and b are the parameters in the Gamma prior on the precision, and μ_i is the value for the i^{th} acquisition predicted by the model. μ_i, for the diffusion tensor model is given by equation 4.6, and for the simple partial volume model, by equation 4.8.

Chapter 5

Global Connectivity Estimation

5.1 Introduction

Ever since (1) reported the tendency for diffusion to be preferred along the mean fibre orientation in a voxel, the idea of using the information available in diffusion weighted MR images to infer anatomical connectivity in the brain has caused a great deal of excitement in the neuroscience community, and an associated frenzy of activity amongst diffusion weighted methodologists. The reason for this excitement is that, currently, anatomical connectivity in the living human brain is effectively invisible. Even in the post-mortem human brain, the techniques available reveal only very limited information. Injection of fluorescent dyes post-mortem allows tracts to be traced, but only for distances of the order of tens of millimetres (e.g. (63)). Longer distance connections can be investigated either by dissection of major pathways, which reveals only the largest connections, or by histological studies of remote degeneration following a focal lesion (64), but such work is based on a relatively small number of informative patients and, even at its best, reveals no information about the route taken by the connection, only about the existence of a connection between the lesion and the remote degenerated site. By far the dominant source of information for neuroscientists interested in anatomical connectivity in the human brain has been the *non-human* primate literature. Sacrificial tracer studies in the non-human brain *can* reveal information about focal connections between distant points in the brain. Such studies have revealed a great deal but are extremely time consuming. Moreover, knowledge of connectional anatomy in non-human primates does not necessarily imply knowledge in the human brain. The human brain has many functions which are different to those of the non-human brain and the associated anatomy is likely to have different organisation. For example, consider Broca's area in the language system. In the human brain, lesions to Broca's area impair aspects of speech production (65). In the intact brain, activity in this area (BA44/45) is readily evoked by language tasks (66). Broca's area is dedicated to function which are uniquely human, and therefore although its evolutionary precursors can be studied anatomically in non-human primates (67), a true understanding of the anatomy of the area can only come from human studies.

The hope is not only that diffusion imaging will provide a unique insight into connections in the normal human brain, but also that the ability to probe fibre connectivity *non-invasively* will provide new insight into the understanding and diagnosis of many neurological disorders.

In chapter 4 we outlined techniques for estimating, at each voxel, probability distributions on every parameter in the chosen model of diffusion. In this chapter we describe a framework for using these local *pdf*s to tackle the problem of estimating anatomical connectivity from a statistical point of view. That is, we will examine the effect of uncertainty in local parameter estimates on the fibre trajectories recovered from diffusion tractography. In chapters 6 and 7 we will show examples which both validate the techniques presented here, and demonstrate how performing tractography in a probabilistic fashion allows us to address new and interesting questions which are inaccessible to traditional deterministic tractography routines.

5.2 Diffusion Tractography

5.2.1 Streamlining

The most popular method for inferring connectivity from DTI data is also the simplest. Cursory inspection of a map of principal diffusion directions (figure 5.1) clearly reveals 2 dimensional trajectories through the data. Streamlining tractography (3; 5; 4; 68) follows these trajectories in 3 dimensions to reconstruct fibre pathways.

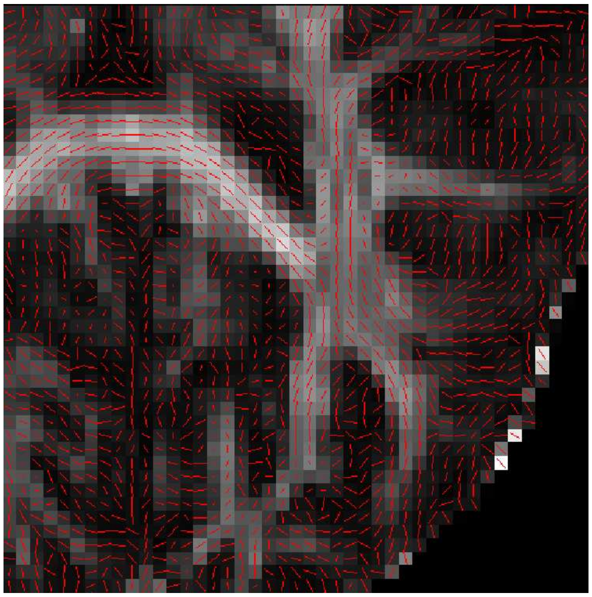


Figure 5.1: Map of projection of principal diffusion directions into slice plane, overlaid on fractional anisotropy in normal brain.

The mathematics behind the streamlining approach are laid out in Basser et al (5), where a white matter fibre tract is represented as a 3D space curve $\mathbf{r}$ parametrised by the distance s along the fibre. The evolution of $\mathbf{r}$ is described by the differential equation

$$\frac{d\mathbf{r}(s)}{ds} = \mathbf{t}(s). \tag{5.1}$$

where $\mathbf{t}(s)$ is the unit tangent vector to $\mathbf{r}(s)$ at s. If the principal diffusion direction lies parallel to the orientation of the underlying fibre tract then the tangent vector $\mathbf{t}(s)$

may be equated up to a sign with the unit eigenvector ϵ_1 evaluated at position $\mathbf{r}(s)$:

$$\mathbf{t}(s) = \epsilon_1(\mathbf{r}(s)). \tag{5.2}$$

The vector geometry can be seen in figure 5.2.

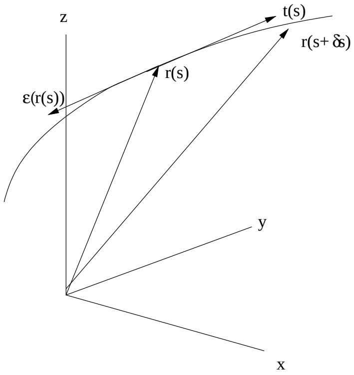


Figure 5.2: Representation of a white matter fibre trajectory as a space curve $\mathbf{r}(\mathbf{s})$. The local tangent vector $\mathbf{t}(s)$ is identified with the eigenvector $\epsilon_\mathbf{1}(\mathbf{r}(s))$, associated with the largest eigenvalue of the Diffusion tensor $\mathbf{D}$ at position $\mathbf{r}(s)$

This set of implicit differential equations is then solved by Euler's method giving:

$$\mathbf{r}(s_1) \approx \mathbf{r}(s_0) + \alpha\epsilon_\mathbf{1}(\mathbf{r}(s_0)) \tag{5.3}$$

with α small and $\mathbf{r}(0) = \mathbf{r}_0$.

The differences between the streamlining approaches proposed by Basser et al.(5), Conturo et al. (4) and Mori et al. (3; 68) centre around two key issues in solving the differential equations: How to interpolate the data field $\mathbf{D}(\mathbf{r}(s))$ and when to stop following the trajectory. Conturo et al. uses trilinear interpolation to interpolate the tensor field, calculates the principal eigenvector and continues until reaching a voxel of anisotropy lower than some threshold A_σ. Mori et al. use this same algorithm, except they choose a nearest neighbour interpoloation scheme. Basser et al. fits splines to generate a continuous approximation to the tensor field and tracks until one of the following occurs.

- 1.The tract enters a region of low anisotropy.

- 2. The radius of curvature of the tract is below 2 voxels.
- 3. The principal diffusion direction in the next voxel is not the eigenvector most collinear with the tract.

Excellent examples of the success of these algorithms in reconstructing major fibre systems can be seen in figure 5.3. A and B, taken from (69) show close correspondence between an in-vivo DTI based reconstruction (A) and a drawing from post mortem dissection (B) of fibre systems around the human brainstem and cerebellar peduncles. Figure 5.3 C, taken from (8) shows a beautiful in-vivo reconstruction of the human callosal system.

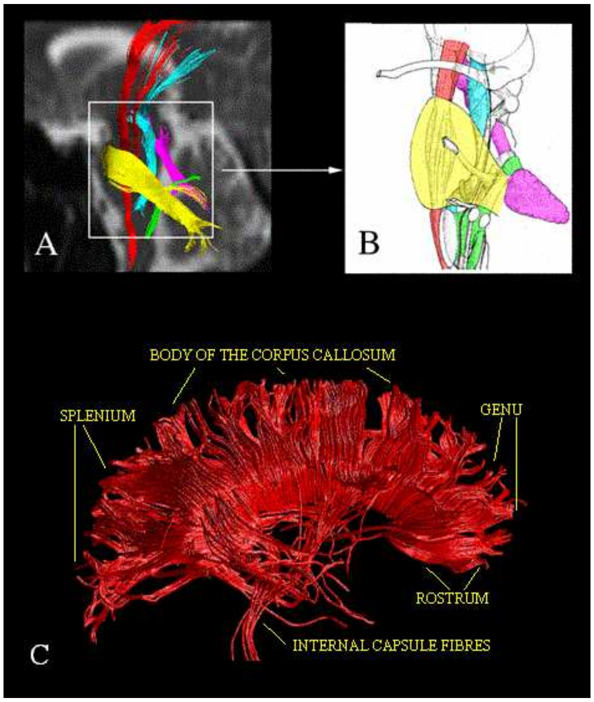


Figure 5.3: (A,B taken from (69)) (A) In-vivo DTI tractography in the fibres around the human brainstem and cerebellar peduncles. (B) drawing from post mortem dissection of the same fibre systems. In both images, the cortico-spinal tract is shown in red, the medial lemniscus in blue, the inferior cerebellar peduncle in green, and the superior cerebellar peduncle in pink. (C) (taken from (8)) In-vivo reconstruction of the human callosal system.

However, in order for these streamlining algorithms to produce a faithful representation of the underlying fibre anatomy two key assumptions must be made. First we must assume that the underlying vector field along the pathways is smoother than the resolution of the diffusion weighted image. If this were true, then a noise-free PDD field would contain all the information necessary to reconstruct the anatomical pathways from the data. Second, we must assume that the PDD field which we have measured is a faithful representation of the true PDD field. That is, we must assume that we have a noise-free measurement process, or that noise has a negligible effect on the recovered PDD field. These are strong assumptions to make, and tend to hold only in specific areas of the brain such as large homogeneous white matter tracts (See later on in this chapter, and chapter 8 for more discussion on this topic).

For these reasons such studies have been limited to pathways easily seen by post-mortem dissections, and have therefore not as yet led to any new discoveries in Neuroscience. Furthermore, between-group comparisons have proved difficult as streamlining methods have provided only a qualitative description of the recovered trajectories, with no quantitative measure of either strength of, or confidence in the connections described.

5.2.2 Distributed Methods

Streamlining methods, as described above, account only for point-to-point correspondences between the seed and end points, recovering what may be (very ambitiously) described as an “axonal pathway” between the two. However, fibre bundles of the same order of magnitude as the voxel size ($\approx 2 \times 2 \times 2mm^3$ at present) do not maintain point-to-point correspondences; they may branch (and merge) in such a way that axons passing through a single seed voxel may terminate in cortical voxels which are far apart. Furthermore, the presence of noise in the image means that even if the underlying fibre bundle *did* maintain a point-to-point correspondence, there would be uncertainty in the measurement process such that there would be some probability that the underlying white matter tract followed a path different from that predicted by the principal diffusion directions. Distributed methods attempt to account for these effects by estimating a likelihood for the connection between a seed point and every other point in the brain.

Distributed methods proposed to date include simulating a diffusion process through the data field (70), using the "fast marching" algorithm (71) to expand a volume preferentially in directions of principal diffusion (72), and using simulated annealing to find the lowest cost path through the data from *every* potential target voxel to the seed voxel (16). Each of these methods has its own specific advantage and disadvantage, but they all suffer from the same overriding problem: They all either assume some heuristic local relationship between diffusion profile and the orientational distribution of fibres in the voxel (e.g., in (70), it is assumed that the fibre distribution is $fODF(\mathbf{r}) \propto \mathbf{r}^T\mathbf{D}^7\mathbf{r}$) or some heuristic bias towards pathways with low curvature (e.g. the cost function in (72) includes a term penalising moves according to their local curvature), or (in most cases) both. This makes it extremely difficult to interpret the resulting "fibre" pathways.

5.3 A Bayesian's Perspective

In this and the following sections we will consider the streamlining problem from a probabilistic perspective. We will outline a method to use the local *pdf*s described in the previous chapter to compute global *pdf* on connection between a seed location and every voxel in the brain. The mathematical descriptions laid out here constitute the theory behind performing diffusion tractography in the presence of uncertainty in local fibre orientation and form the heart of the diffusion tractography approach used to generate the results in chapters 6 and 7.

Throughout these sections we consider that we have available to us at every voxel a probability density function describing our belief in the underlying fibre orientation such as those derived in the previous section $\mathcal{P}((\theta,\phi)_{x_i}|\mathbf{Y}_{x_i})$, where subscript x_i refers to the i^{th} voxel, and (θ,ϕ) are spherical polar coordinates describing fibre orientation. At any point in the following mathematical derivations, if these voxelwise *pdf*s are considered to be infinitely narrow, the probabilistic solution reduces to the equivalent streamlining solution. The following two sections represent a generalisation of streamlining tractography to the case where there is uncertainty in the local fibre architecture.

To illustrate this point, let us first consider this exact case where the *pdf*s are infinitely narrow, or the values of the local parameters are known with no uncertainty. What do they tell us about anatomical connectivity between voxels in the brain? In the case where our local model describes only a single fibre orientation passing through the voxel, this *global* model can only take one form:

$$\mathcal{P}(\exists A \to B|(\theta,\phi)_{\mathbf{x}}) = \begin{cases} 1 & \begin{array}{c}\text{If there is a connecting} \\ \text{path through } (\theta,\phi)_{\mathbf{x}}\end{array} \\ \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

Where $\mathcal{P}(\exists A \to B|(\theta,\phi)_{\mathbf{x}})$ is the probability of a connection existing between points A and B, given knowledge of local fibre orientation. Note that, throughout the remainder of this chapter, subscript $\mathbf{x}$ refers to "every voxel in the brain". Hence $(\theta,\phi)_{\mathbf{x}}$ refers to the complete set of principal diffusion directions.

In order to solve this equation we may simply start a connected path from a *seed point*, A, and follow local fibre orientation until a stopping criterion is met. If B lies on this path we may say that a connection exits between A and B. This procedure is at the heart of all "streamlining" algorithms (e.g. (4; 68; 5)), which choose $(\theta,\phi)_{\mathbf{x}}$ to be the principal eigen-direction of the estimated diffusion tensor at each voxel.

However, in the case where there is uncertainty associated with $(\theta,\phi)_{\mathbf{x}}$, the probability distribution describing the unknown parameters in the global model is the joint posterior distribution on the existence of a connection $(\exists A \to B)$ *and* all of the local orientations $(\theta,\phi)_{\mathbf{x}}$:

$$\mathcal{P}((\exists A \to B),(\theta,\phi)_{\mathbf{x}}|\mathbf{Y}_{\mathbf{x}}) = P(\exists A \to B|(\theta,\phi)_{x})\mathcal{P}((\theta,\phi)_{\mathbf{x}}|\mathbf{Y}_{\mathbf{x}}). \tag{5.5}$$

Expanding $\mathcal{P}((\theta,\phi)_x|\mathbf{Y}_\mathbf{x})$, the joint distribution on the field of fibre orientations, reveals:

$$\mathcal{P}((\exists A \rightarrow B),(\theta,\phi)_\mathbf{x}|\mathbf{Y}_\mathbf{x}) = P(\exists A \rightarrow B|(\theta,\phi)_x)\prod_{i=1}^{v}\mathcal{P}((\theta,\phi)_{x_i}|\mathbf{Y}_{x_i}), \tag{5.6}$$

where v is the number of voxels. That is, for any specific set of fibre orientations at every voxel, the probability of this being the true set of fibre orientations, and that this set of fibre directions implies a connection between A and B is simply the product of the probabilities of these fibre orientations at each voxel multiplied by 1 if there is a streamline connecting A to B through these orientations or 0 if there is not.

However, this joint distribution is not the distribution of real interest; we would really like to be able to answer the following question: "Accounting for all of the uncertainty in the local fibre orientations, what is the distribution on $(\exists A \rightarrow B)$ given the known data". We would like to compute the *marginal posterior distribution* on the existence of a connection $\mathcal{P}(\exists A \rightarrow B|\mathbf{Y}_\mathbf{x})$. Acquiring this distribution requires us to marginalise the joint distribution in equation 5.6 over all possible fibre orientations at every voxel. i.e. we must perform the following integrations (See chapter 3 for a detailed discussion on marginalisation):

$$\begin{aligned} &\mathcal{P}(\exists A \rightarrow B|Y) = \\ \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{2\pi}\int_0^{\pi} P(\exists A \rightarrow B|(\theta,\phi)_\mathbf{x}) &\prod_{i=1}^{v}\mathcal{P}((\theta,\phi)_{x_i}|\mathbf{Y}_{x_i}) \\ &d\theta_{x_1}d\phi_{x_1}\ldots d\theta_{x_v}d\phi_{x_v} \end{aligned} \tag{5.7}$$

That is, for each possible set of fibre orientations at every voxel $(\theta,\phi)_\mathbf{x}$, we must incorporate the probability of connection given this $(\theta,\phi)_\mathbf{x}$, and also the probability of this $(\theta,\phi)_\mathbf{x}$ given the acquired MR data. Again, note that if $\mathcal{P}((\theta,\phi)_\mathbf{x}|\mathbf{Y})$ are delta functions, equation 5.7 reduces to the streamlining solution $P(\exists A \rightarrow B|(\theta,\phi)_\mathbf{x})$. However, when local fibre orientation *is* uncertain, $\mathcal{P}(\exists A \rightarrow B|Y)$ will be non-zero for some

B not on the maximum likelihood streamlines. That is, the global connectivity pattern from A, will spread to incorporate the known uncertainty in local fibre orientation.

Even in the discrete data case, equation 5.7 represents a v dimensional (where v is the number of voxels in the brain) integral over distributions with no analytical representation (the local *pdf*s, generated with MCMC), and hence clearly cannot be solved analytically.

Fortunately, as we have seen previously, even when explicit integration is unfeasible, it is often possible to compute integrals implicitly by drawing samples from the resulting distribution. In our case, in order to draw a sample from $\mathcal{P}(\exists A \to B|Y)$ we may draw a sample from the posterior *pdf* on fibre orientation at each point in space and construct the streamline (henceforth referred to as a "probabilistic streamline") from A given these orientations. Computationally, this process is extremely cheap. Samples from the local *pdf*s at each voxel have already been generated, so to generate a single probabilistic streamline from seed point A, referring to the current "front" of the streamline as $\mathbf{z}$, it is sufficient simply to start $\mathbf{z}$ at A and:

- Select a random sample,(θ, ϕ) from $\mathcal{P}(\theta, \phi|\mathbf{Y})$ at $\mathbf{z}$.
- Move $\mathbf{z}$ a distance s along (θ, ϕ).
- Repeat until stopping criterion is met.

This probabilistic streamline is said to connect A to all points B along its path. By drawing many such samples, we may build the *spatial pdf* of $\mathcal{P}(\exists A \to B|Y)$ for all B. We may then discretise this distribution into voxels, by simply counting the number of probabilistic streamlines which pass through a voxel B, and dividing by the total number of probabilistic streamlines.

5.4 A Path Integral Equivalence

In the previous section we described an approach to the problem of estimating a *global pdf* on connectivity via the dominant streamline from the *local pdfs* on fibre orientation computed in chapter 4. This approach came from the Bayesian statistics literature and relied on a numerical marginalisation over each local *pdf*, to account for the effect of this local uncertainty on the estimated streamline.

We will now show that if this same problem was approached by a statistical physicist knowing nothing of Bayesian statistics, he might arrive at exactly the same algorithm by considering the probability of connection between points A and B as the integral over every possible path connecting A and B of the probability of that path (see figure 5.4).

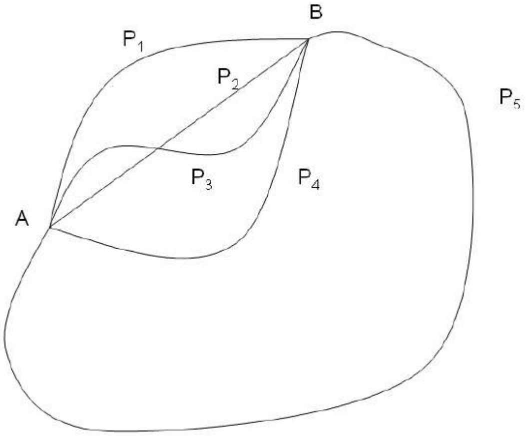


Figure 5.4: Paths connecting two points, A and B in a data field. The probability that A and B are connected is the integral over all possible paths of the probability of each path.

Let us consider a continuous path, $\mathbf{r}(s)$ (s indexes distance along the path), through a continuous *pdf* field, where at every point we have a *pdf* on fibre orientation. If we discretise this path into n steps of length l, we may write down the probability on the existence of this path of length ln given the field, F,as:

$$\mathcal{P}(\mathbf{r}(s)|ln, F) = \lim_{l \to 0, ln\ const} \mathcal{P}(\theta, \phi|\mathbf{r}(0))\mathcal{P}(\theta, \phi|\mathbf{r}(l)) \dots \mathcal{P}(\theta, \phi|\mathbf{r}((n-1)l)). \quad (5.8)$$

We may now say that a connection of length ln exists from point A to point B if any path of length ln leaving point A passes through point B.

$$\mathcal{P}((\exists A \rightarrow)_{ln} B|F, ln) = \int_{\forall (A\rightarrow B)_{ln}} \mathcal{P}(\mathbf{r}(s)|F, ln) d\mathbf{r}(s) \tag{5.9}$$

Substituting equation 5.8 into equation 5.9 gives:

$$\mathcal{P}((\exists A \rightarrow B)_{ln}|F, ln) = \int_{\forall (A\rightarrow B)_{ln}} \lim_{l\rightarrow 0, ln\ const} \left(\prod_{i=0}^{n-1} \mathcal{P}(\theta, \phi|\mathbf{r}(il)) \right) d\mathbf{r}(s), \tag{5.10}$$

which describes the probability of a connected path of length ln exists between A and B in the data field. We may now derive the probability of the existence of *any* path by integrating this expression over all path lengths, S.

$$\mathcal{P}(\exists A \rightarrow B|F) = \int_0^\infty \left[\int_{\forall (A\rightarrow B)_{ln}} \lim_{l\rightarrow 0, ln=S} \left(\prod_{i=0}^{n-1} \mathcal{P}(\theta, \phi|\mathbf{r}(il), F) \right) d\mathbf{r}(s) \right] dS \tag{5.11}$$

Now let us consider a particle, p, in the same field, F, starting from point $\mathbf{r_0}_p$ and progressing according to the algorithm described in section 5.3 above. What is the spatial probability distribution on p's location after i steps, $\mathbf{r_i}_p$? At each step, p moves a distance l according to

$$\mathcal{P}(\mathbf{r_i}_p|\mathbf{r_{i-1}}_p) = \mathcal{P}(\theta, \phi|\mathbf{r_{i-1}}_p, F), \tag{5.12}$$

where (θ, ϕ) is the direction which takes $\mathbf{r_{i-1}}_p$ to $\mathbf{r_i}_p$. The spatial density function after n independent steps is therefore

$$\mathcal{P}(\mathbf{r}_{n_p}|\mathbf{r}_{0_p}, F) = \int_{\forall \mathbf{r}_{0_p} \rightarrow \mathbf{r}_{n_p}} \prod_{i=1}^{n} \mathcal{P}(\mathbf{r}_{i_p}|\mathbf{r}_{i-1_p}) d\mathbf{r}_{1_p} \dots d\mathbf{r}_{n-1_p} \tag{5.13}$$

and we can easily see that if we let the step size tend to zero, this density is the term inside square brackets in equation 5.11 with $A = \mathbf{r}_{0_p}, B = \mathbf{r}_{n_p}$. Again, the integration over S is performed by allowing for any path length $m \leq n$. Hence, the algorithm derived in the previous section, results in pathways whos spatial *pdf* is distributed according to the path integral *pdf* of connection.

5.5 Technical Details

5.5.1 Data Interpolation

The sampling technique described in the previous section relies on the local *pdf*s existing in continuous space. Unfortunately, we only have access to MR acquisitions, and hence these local *pdf*s, on a discrete acquisition grid. We need a technique to generate samples from the local *pdf*s at a point not on the grid.

An obvious solution to this problem would be to interpolate the original data (using a standard interpolation scheme, such as sinc or trilinear interpolation), and generate the local *pdf* on fibre orientation given this new *interpolated* data. This would be extremely computationally costly, but also, on further consideration, may not conceptually be the best thing to do. In the middle of large fibre bundles, where neighbouring voxels have similar fibre orientations (each with low uncertainty), the choice of interpolation scheme will have very little effect. However, in places where neighbouring voxels may have significantly different orientations, such as at the edge of fibre bundles or where different bundles meet, such an interpolation scheme will generate a fibre orientation *in between* the orientations of the voxels on the grid. More over, the result of sinc or trilinear interpolation of data which is related to parameters in a highly nonlinear (e.g. exponent of trigonometric functions) manner is likely to produce interpolated data which does not fit well to the model, and thus the resulting most probable fibre orientation will be highly dependent on the noise in the measurements at the grid locations. An alternative to interpolating the data in this fashion is to choose an interpolation scheme which will pick a sample from *one of* the neighbouring voxels on the grid. In a probabilistic system, we also have the opportunity to use a probabilistic interpolation scheme. That is, we can choose a scheme which chooses the data from a single neighbouring point on the acquisition grid, but the probabilities of choosing each neighbour will be a function, g, of their positions relative to the interpolation site. There are many possible functions for g, but we have chosen one which is analogous to trilinear interpolation. That is, in the x-dimension, the probability of choosing data from `floor(x)` is $g(\mathtt{floor(x)}|\mathtt{x}) = \frac{\mathtt{ceil(x)-x}}{\mathtt{ceil(x)-floor(x)}}$, and from `ceil(x)` is $g(\mathtt{ceil(x)}|\mathtt{x}) = \mathtt{1} - \mathtt{g(floor(x)|x)}$, and the same in the y and z-dimensions. If a streamline, $\mathbf{z}$, were to pass through the same point twice, different nearest neighbours may be chosen, reflecting our lack of knowledge of the *true pdf* at that point.

5.5.2 Stopping criteria

Algorithms which generate streamlines based on maximum likelihood fibre orientations (principal eigenvector from a diffusion tensor fit) have tended to require harsh streamline stopping criteria based on fractional anisotropy and local curvature (angle between successive steps). Fractional anisotropy thresholds have tended to be in the range of 0.2-0.4 (e.g. (69)), and curvature thresholds have been as strict as requiring successive steps to be within 45° (e.g. (5)). These criteria are in place to reduce the sensitivity of the streamlining to noise in the image, partial volume effects, and other related problems. The aim is to reduce the possibility of seeing false positives in the results by only progressing when there is high confidence in fibre orientation, and when the orientation is anatomically plausible. The downside of these constraints is the limitations that they impose on which fibre tracts may be reconstructed and where in the brain they may occur. For example deep grey matter structures, despite displaying a high degree of order in their principal diffusion directions, tend to have low anisotropy (often below the threshold for streamlining algorithms). Streamlines will also tend to terminate well before cortex as anisotropy reduces, and uncertainty in fibre orientation increases.

In such circumstances a probabilistic algorithm has significant advantages. First, in regions where fibre orientation is uncertain (these often coincide with regions of low anisotropy), the algorithm has available to it a direct representation of this uncertainty.

Hence, even though it cannot progress along a single direction with high confidence, it *can* progress in many directions. The uncertainty in this area will be represented by voxels further along the path having lower probabilities associated with them; however a high probability of connectivity to the seed voxel may still be associated with the *region* into which the paths progress. A second useful advantage of a probabilistic algorithm is robustness to noise. It can be difficult to track beyond a noisy voxel using a non-probabilistic algorithm as it may initiate a meaningless change in path. However, with a probabilistic algorithm, paths which have taken errant routes tend to disperse quickly, so that voxels along these paths are classified with low probability. In contrast "true" paths tend to group together, giving a much higher probability of connection for voxels on these paths.

These advantages significantly reduce the need for anisotropy and curvature stopping criteria. The results presented here are generated with *no* anisotropy threshold, and with a conservative local curvature threshold of $\pm 80^\circ$ for each sample. This curvature threshold is required, as, without it, the sampled streamlines may track back along a path similar to one already visited, artificially increasing the probability along the path. In order to reduce this effect further, we check, at every step, whether the path is entering an area it has already visited, and terminate those that are.

5.6 Results

5.6.1 Simulations

Data was simulated at each voxel from a diffusion tensor with eigenvalues $\{1, 0.4, 0.2\} \times 10^{-3} smm^{-2}$, giving a Fractional Anisotropy of 0.65. Real and imaginary noise was added in Fourier space with standard deviation of $\sqrt{2} \times \frac{S_0}{SNR}$. Data was reconstructed via the Fast Fourier Transform, and the absolute value was computed. Local *pdf*s were then calculated at each voxel as in chapter 4, and tractography was carried out as described in the previous sections of this chapter.

Figures 5.6,5.7,5.8 show samples from the connectivity *pdf* when tracing through the linear field in figure 5.5 under different Signal to Noise ratios. In each case (A) did not perform loop checking as described in section 5.5, and (B) did. SNRs were 4 in figure 5.6, 8 in figure 5.7 and 15 in figure 5.8. Tracking was carried out over 6 voxels (≈ 12 steps), to examine the effect of SNR on the resulting distribution.

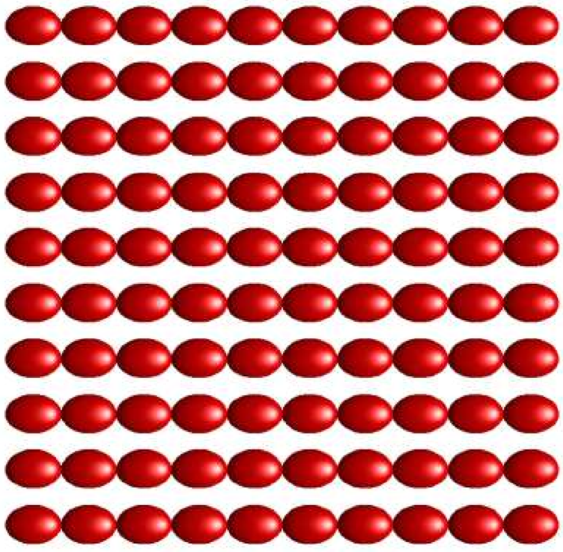


Figure 5.5: Noise free tensor field for the SNR simulations in the following figures. Tensors have eigenvalues $\{1, 0.4, 0.2\} \times 10^{-3} smm^{-2}$. Tractogrpahy is seeded in the central voxel.

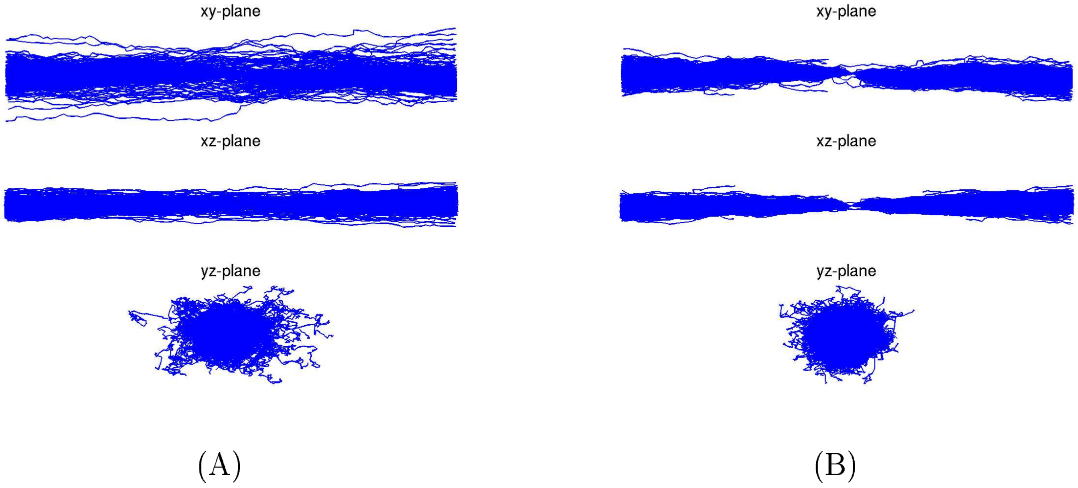
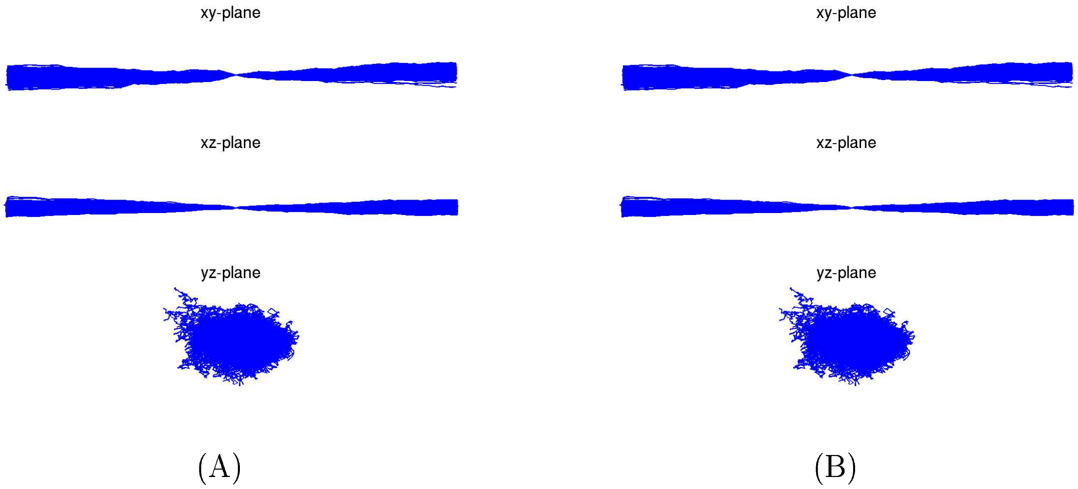
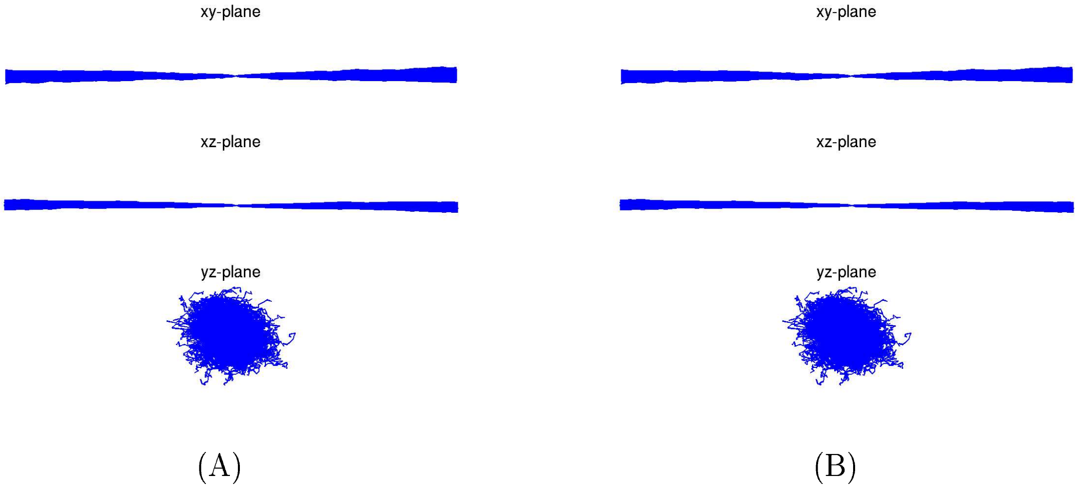


Figure 5.6: Simulated tracking through tensor field in Figure 5.5 with SNR of 4. (A) Without loopchecking (B) with loopchecking.



(A) (B)

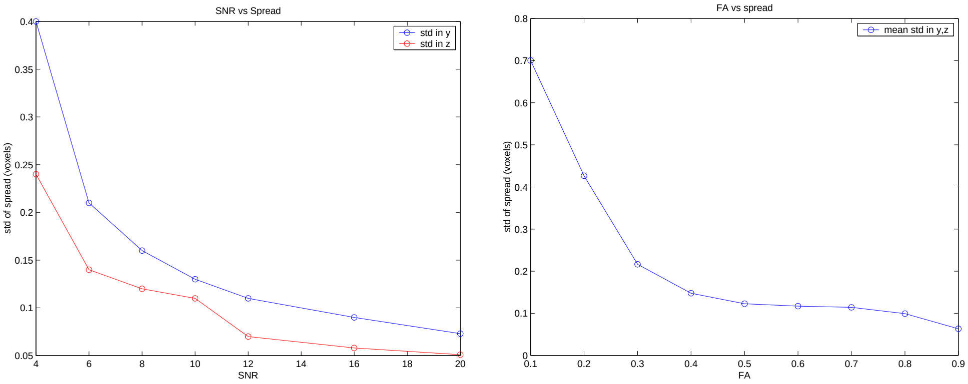
Figure 5.7: Simulated tracking through tensor field in Figure 5.5 with SNR of 8. (A) Without loopchecking (B) with loopchecking.



(A) (B)

Figure 5.8: Simulated tracking through tensor field in Figure 5.5 with SNR of 15. (A) Without loopchecking (B) with loopchecking.

These results are quantitated in figure 5.9 (A) and (B). (A) shows the dependence of the distribution spread on signal to noise ratio (i.e the effect that image noise has on the recovered distributions). The standard deviations of position in y (blue) and z (red) are plotted against SNR after the streamlines have travelled 6 voxels in x. Note that the spread is greater in y than in z as the simulated diffusion tensors had an eigenvalue of $0.4 \times 10^{-}3$ in y and $0.2 \times 10^{-}3$ in z. Typical SNR levels in diffusion weighted images might be around 10-15. (B) shows the dependence of the distribution on the underlying diffusion anisotropy of the system - determined by the biophysical properties of the anatomy and the true orientational distribution of fibres. A SNR of 10 was assumed for these simulations.



(A) (B)

Figure 5.9: Effect of SNR (A) and anisotropy (B) on the spread of the connectivity distribution.

5.6.2 Visual System

Figure 5.10 shows probabilistic tractography in the human visual system. Red to yellow encodes probability of connection.(**a**) From a voxel in putative lateral geniculate nucleus of thalamus (LGN), the connectivity distribution was traced anteriorly along the optic tract, and posteriorly to the visual cortex, consistent with the well-established anatomy of the visual system. (**b**) Seeding a voxel in the optic tract generated a connectivity distribution that followed two distinct paths (meaning that some of the multiple pathways generated followed one path, whereas others followed the other path). One path went through the LGN and on to visual cortex, corresponding to the optic radiations (as in **a**). A second path corresponded to the brachium of the superior colliculus (see axial slice in **b**). The separation of these two paths can be seen in the coronal sections in **b**. The more inferior and medial path follows the brachium and the more superior path will follow the optic radiation (as in **a**). Seeding a voxel in the white matter just lateral to the LGN generated a distribution that travelled indirectly to the visual cortex, via the temporal lobe, forming Meyer's loop. The path

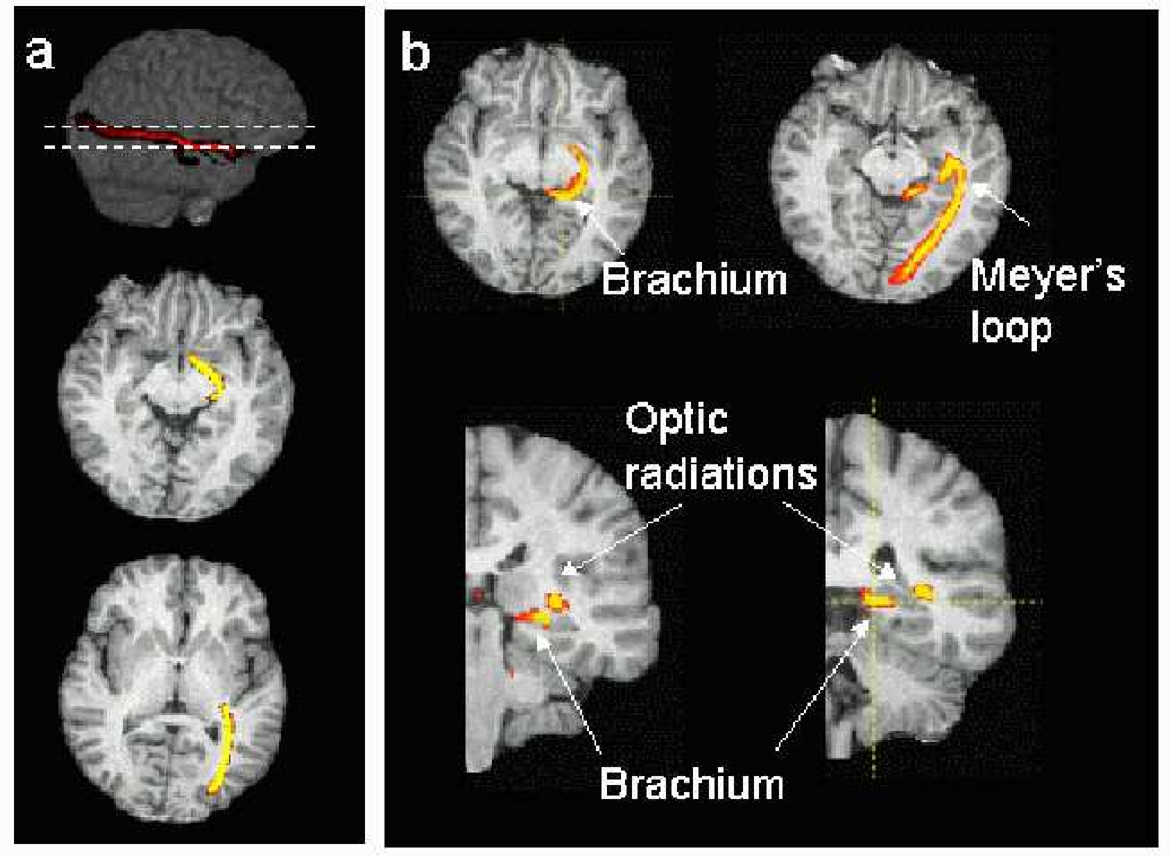


Figure 5.10: Connectivity distributions in the human visual system. (a) shows a distribution seeded in lateral geniculate nucleus of thalamus (LGN). (b) shows a distribution seeded in the optic tract. See text for details.

displayed here is a maximum intensity projection over 4 axial slices, overlaid on a single slice.

5.6.3 A note on interpretation

The implication of accounting for the uncertainty in local fibre orientations, and hence estimating a spatial probability distribution of connectivity from the seed point, is that the recovered connectivity distribution is spread in space (see, for example thalamic pathways in chapter 6). It is tempting to think of this distribution as a distribution of connections from the seed point. This is not the case. According to the model used earlier in this section, this spatial *pdf* represents *confidence bounds* on the location of the most probable single connection. It is certainly true that some of the uncertainty estimated locally is likely to be due to partial volume effects, such as a spread of fibre orientations in the voxel, and therefore the presence in the brain of multiple connection sites from the seed *may* result in a diffuse spatial *pdf*. However, while the model of diffusion at each voxel includes only a single fibre orientation, the global inference is clearly on a single pathway.

5.7 Discussion

In this chapter we have presented an algorithm for tracing pathways through a vector field in the presence of uncertainty. We have shown that the *pdf* on connectivity from a seed point may be obtained either by considering the integral over all possible fibre orientations at each voxel (i.e. by marginalising the joint posterior distribution on fibre orientation at *every* voxel) or by considering the integral of the probability of a path over all possible paths. We have shown that these two integrals are equivalent, and have developed an algorithm for performing these very high dimensional integrals in a computationally efficient manner. This algorithm turns out to be similar in nature to others presented, along with this method, at ISMRM 2002 (73; 74; 75), effectively repeatedly sampling local *pdf*s to create streamlines, and regarding these streamlines as samples from a global *pdf*. A crucial difference between these algorithms and the one presented here is that the local *pdf*s used here are meaningful quantities derived rigorously from the data and model. Hence we are able to define the global distribution of interest and show that the algorithm presented converges on this distribution. The global *pdf*s recovered are both meaningful and easily interpretable. The methods referenced above all use heuristic experience-based relationships between the shape of the fitted diffusion tensor and the assumed local *pdf*s, making it very difficult to interpret the resulting distributions.

Section 5.6 shows resulting distributions from both simulated and real data; the simulations in 5.6.1 explore the relationship between the uncertainty in the fibre trajectories (i.e. the spread of the distribution) and factors including MR noise and biophysical complexity. The results in figure 5.9 show these effects quantitatively. In highly coaligned white matter ($FA > 0.7$), with MR sequences typically used for tractography ($SNR > 10$, ≈ 60 diffusion encoding directions), the uncertainty caused by MR noise is very low (the distributions have cone angles of only 1 or 2 degrees). However, even in good signal to noise conditions, biophysical complexity (and therefore low underlying diffusion anisotropy) will cause high uncertainty. This can be seen clearly in figure 5.10, which shows tractography results in the human visual system. The majority of the connectivity distributions lie in highly aligned white matter, resulting in minimal uncertainty in principal diffusion direction. However in (b) which is seeded in the optic tract, the distribution encounters a fibre split, where the optic tract splits into fibres progressing to LGN and then the optic radiations, and fibres progressing to the superior colliculus via the superior collicular brachium. This biological complexity is reflected as uncertainty in the local pdfs and the connectivity distribution spreads into both pathways.

The work in this chapter has addressed an important problem in tractography, namely how to quantify our belief in the resulting fibre trajectories. However, there are other important problems to solve before diffusion tractography can come close to the specificity achieved by invasive techniques. The discussion on interpretation in subsection 5.6.3 touches on some of these problems; belief in the fibre trajectories is not the quality which scientists would most like to quantify. A quick sample of local neuropsychologists reveals “The likelihood of neuronal activity in one area directly influencing neurons in another”, or (only slightly) more realistically, “The number of fibres connecting two areas” as quantities of real interest. In order to get closer to these (in some cases) anatomically meaningful quantifications the most obvious requirement is that, instead of local pdfs on dominant fibre orientation, we have an estimate of (at worst) the actual distribution of fibres within the voxel (the fibre orientation distribution function, or $fODF$, denoted $\mathcal{H}(\theta, \phi)$ in 4.1.3). Even if this information were available there would be non-trivial problems to be solved. See chapter 8 for a detailed discussion on this topic, and on how the framework presented in this chapter will apply even when the inference is on objects far more complicated than the location of the dominant streamline.

Despite these limitations, a probabilistic approach to tractography has significant advantages over deterministic approaches. The internal representation of uncertainty within the algorithm allows a probabilistic tractography scheme to continue tracing robustly through areas of low diffusion anisotropy where conventional methods might fail. This will prove to be crucial in chapter 6 where connections are traced from thalamic grey matter all the way to cortical grey matter. The ability to quantify belief in a meaningful manner significantly increases the interpretability of the tractography output ("Is this really a pathway or is it just noise - how can we be sure? What other paths might have been taken under different noise conditions"). Finally, the existence of a distributed representation of our belief in a pathway allows us to answer questions which are unavailable to conventional techniques. For example, in chapter 6, we are able to parcellate thalamic grey matter based on the cortical zone with the *highest probability* of connection and in chapter 7 we use similarities between connectivity *distributions* to cluster cortical grey matter into functionally discrete subunits.

Chapter 6

Results from The Human Thalamus

Evidence concerning anatomical connectivities in the human brain is sparse and based largely on limited post-mortem observations. Diffusion tensor imaging has previously been used to define large white matter tracts in the living human brain, but has had limited success at tracing pathways into grey matter. Here we identify specific connections between human thalamus and cortex using the probabilistic tractography algorithm described in the previous chapter. Classification of thalamic grey matter based on cortical connectivity patterns reveals distinct subregions whose locations correspond to nuclei described previously from histological studies. The connections that we find between thalamus and cortex are similar to those reported for non-human primates and are reproducible between individuals. Our results provide the first quantitative demonstration of reliable inference of anatomical connectivity between human grey matter structures using diffusion data and the first connectivity-based segmentation of grey matter.[1] [2]

---

[1] A preliminary version of the material in this chapter was published in *Nature Neuroscience*(13)

[2] The research in this chapter was carried out in close collaboration with Heidi Johansen-Berg who appeared as joint first author on the resulting paper.

6.1 Introduction

The anatomical connectivity pattern of a brain region determines its function (76). However, although invasive tracer studies have produced a large body of evidence concerning connectivity patterns in non-human animals (77; 78; 79), direct information concerning brain connections in humans is very limited. Injection of fluorescent dyes post-mortem allows tracing of tracts, but only for distances of tens of millimetres (63). Longer distance connections can be investigated by dissection of major tracts or histological studies of remote degeneration following a focal lesion (64), but such work is based on a relatively small number of informative patients.

A specific, important focus for investigation is the thalamus because nearly all incoming information to the cortex is routed through this deep grey matter structure. The thalamus is divided into cytoarchitectonically distinct nuclei which have different patterns of anatomical connectivity that are well characterised for non-human animals (80; 81). The ability to identify these cytoarchitectonically distinct nuclei would have many uses; for example the ability to make specific inferences about thalamic activations in functional imaging studies and the ability to localise nuclei which are the targets for functional neurosurgery in movement disorders (82). However, boundaries between thalamic nuclei that can be visualised histologically (83) are not clear in-vivo even using contrast-optimised magnetic resonance (MR) imaging (84). If connectivity information were available locally, it could be used as an alternative method for defining boundaries between thalamic sub-regions and would provide information directly relevant to function.

Diffusion imaging characterises the apparent diffusion properties of water (85; 2). In tissue with a high degree of directional organisation, the self-diffusion of water protons is different in different directions. In brain white matter the principal diffusion direction corresponds well with orientation of major fibres in the voxel (36). Recent developments in DTI techniques have enabled tracing of large fibre tracts in the living human brain (86; 87; 3; 88; 89; 90; 9). However, as noted in chapter 5, conventional 'streamlining' tract tracing algorithms typically can only progress when there is high certainty of fibre direction (i.e., when local diffusion anisotropy exceeds a specified threshold) (3; 90; 9). This has limited their usefulness in defining pathways near grey matter and thus to their cortical or sub-cortical targets.

Here, using the probabilistic tractography algorithm described in chapter6, we demonstrate how anatomical connectivity progressing fully into grey matter can be inferred. We have used this approach to provide a first comprehensive description of the connections between thalamus and cortex in the human brain *in vivo*. An additional novel result of this approach is the discrimination of human thalamic subregions on the basis of their connections with the cortex.

6.2 Results

6.2.1 Connectivity Distributions

We used a fully automated probabilistic tractography algorithm (see chapters 4 and 5) to form connectivity distributions from individual voxels within the thalamus of a single subject. From these distributions, we traced pathways all the way to the cortex (Figure 6.1 a-d). For example, seeding a voxel in the lateral geniculate nucleus (LGN) generated pathways to the visual cortex and optic tract (Figure 6.1 **a**). The course of generated pathways depends on seed point locations (Figure 6.1 b).

6.2.2 Commonly connected thalamic sub-regions

We segmented the cortex into large anatomically-defined regions (see methods) corresponding to known connection areas of the major thalamic nuclear groups in non-human primates (Figure 6.2 a). For every thalamic voxel probabilistic tractography was used to define probability levels for connectivity to each ipsilateral cortical zone. Each thalamic voxel then was classified according to the ipsilateral cortical zone with which it had the highest probability of connectivity. This resulted in clusters of commonly con-

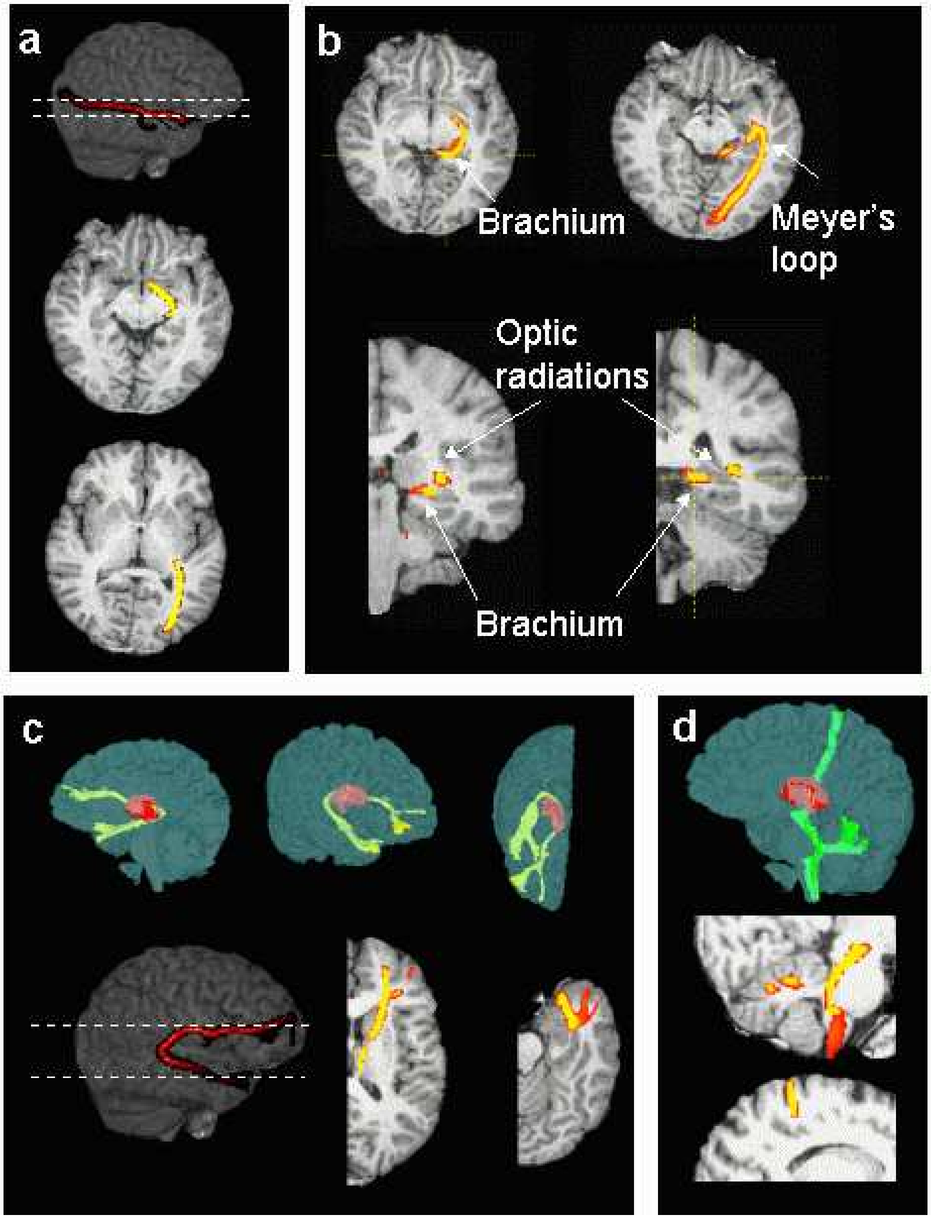


Figure 6.1: Tracing connectivity distributions from individual seed voxels. Voxels are colour coded according to whether the probability of pathways travelling through that voxel is high (yellow) or low (red). (**a**) From a voxel in putative LGN, the connectivity distribution was traced anteriorly along the optic tract, and posteriorly to the visual cortex, consistent with the well-established anatomy of the visual system. (**b**) Seeding a voxel in the optic tract generated a connectivity distribution that followed two distinct paths (meaning that some of the multiple pathways generated followed one path, whereas others followed the other path). One path went through the LGN and on to visual cortex, corresponding to the optic radiations (as in **a**). A second path corresponded to the brachium of the superior colliculus (see axial slice in **b**). The separation of these two paths can be seen in the coronal sections in **b**. The more inferior and medial path follows the brachium and the more superior path will follow the optic radiation (as in **a**). Seeding a voxel in the white matter just lateral to the LGN generated a distribution that travelled indirectly to the visual cortex, via the temporal lobe, forming Meyer's loop. The path displayed here is a maximum intensity projection over 4 axial slices, overlaid on a single slice. (**c**) From a voxel in putative Medio-Dorsal necleus (MD), the connectivity distribution progressed anteriorly to the lateral prefrontal cortex and also, at first posteriorly, around the posterior edge of the thalamus, and then anteriorly to the anterior temporal cortex. (**d**) From a voxel in putative VL, the distribution both ascended to the anterior bank of the central sulcus (M1) and descended. The descending distribution followed two distinct paths, one entered the cerebellum and branched, terminating in the cerebellar cortex, the other continued further down the brainstem.

nected voxels within the thalamus. On the basis of the strong correspondence between our connectivity-based clusters in the human thalamus, known locations of thalamic nuclei in humans (83), and connections in non-human primates (80), we propose that these correspond to different thalamic nuclei or nuclear groups (Figure 6.2). Note, that the description cannot distinguish between thalamo-cortical and cortico-thalamic connections as DWI is sensitive only to the orientation of fibre tracts and not to their anatomical polarity.

In the monkey, the mediodorsal nucleus (MD) is reciprocally connected to the prefrontal cortex (PFC) (91; 92) via the anterior thalamic peduncle and to temporal regions including the temporal pole (93), olfactory cortex (94) and amygdala (95) via the inferior thalamic peduncle. Our results suggest that a similar organisation is found in the human brain as a large medial, dorsal region of the thalamus had highly probable prefrontal and temporal connections. We suggest that this region includes MD. Studies in non-human primates have shown that the ventral posterior nucleus (VP) projects to primary and secondary somatosensory areas (SI,SII) (96). We found a similar ventral posterior region with a strong probability of somatosensory connections (Figure 6.2 b,c) and suggest this corresponds to the human VP. In monkeys, the ventral lateral (VL) and ventral anterior (VA) nuclei project to primary (M1) and premotor cortex (PMC) (97). A lateral region, anterior to putative VP, was shown here to have a high probability of motor cortical connectivity (Figure 6.2 b,c). We suggest that this region includes VL and VA. In the non-human primate, the lateral posterior nucleus (LP) and parts of the pulvinar (Pu) project to posterior parietal cortex (PPC) and extrastriate cortices (80), and medial and inferior parts of the pulvinar are connected to the temporal lobe (98; 99). In the human brain we found a posterior region that is connected to these areas, and propose that it corresponds to the LP/Pu complex (Figure 6.2 b,c). Confidence in the connectivity parcellations was increased by the demonstration that the pattern of connections was comparable between left and right thalami (Figure 6.2 b,c).

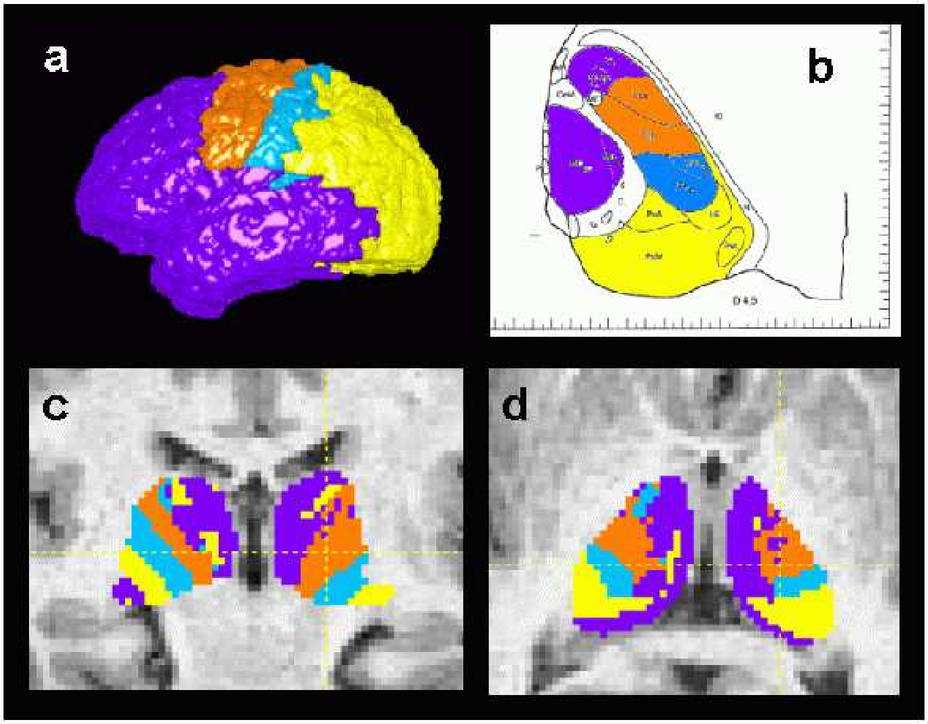


Figure 6.2: Connectivity-based segmentation of the thalamus in a single subject. (**a**) Division of the cerebral cortex. (**b**) An axial section based on a histological atlas of the human thalamus with nuclei outlined by black lines (83). Nuclei have been colour coded according to the cortical zone to which we predict they would show the strongest connections, on the basis of literature from non-human primates (80; 91; 92; 93; 94; 95; 96; 97). (**c**)(**d**) Classifying thalamic voxels based on the zone with the highest probability of connection resulted in clusters of commonly-connected voxels. The clusters correspond to histologically-defined locations of major nuclei (**b**). The medial, anterior purple area in **c** and **d** is thought to include the mediodorsal nucleus and nuclei within the anterior complex, which are connected to prefrontal cortex and the temporal lobe. The more posterior purple area is thought to include parts of the lateral and inferior pulvinar which connect to the temporal lobe. The yellow area is thought to include the anterior pulvinar and the lateral posterior nucleus which project mainly to posterior parietal and extrastriate areas. The blue area is thought to include the ventral posterior lateral nucleus, which projects to somatosensory cortices. The orange area is thought to include the ventral lateral and ventral anterior nuclei, which project to motor and premotor cortices.

6.2.3 Nuclear subdivisions

We further divided the cortical surface in the left hemisphere (for example, M1 was separated from PMC, Figure 6.3 a) and re-ran the automated tractography-based segmentation for the left thalamus to test for finer thalamic subdivisions. Within the large lateral clusters, smaller, distinct regions connecting to the somatosensory cortices, M1, and PMC were distributed along a posterior-anterior axis (Figure 6.3 b). From the known anatomy in the non-human primate, in which VP projects to SI/SII (96), VLp to area 4 (100), and VLa and VA to lateral and medial area 6 (101), we suggest that these correspond to VP, VLp, and VLa/VA. On the cortical surface we also separated PPC and the occipital cortex. We propose that the inferior/lateral area (coloured red in Figure 6.3 c) with most highly probable connections to the occipital lobe corresponds to the lateral geniculate nucleus (LGN). The posterior region that connects to PPC (coloured yellow in Figure 6.3) may include LP and anterior and lateral Pu. Medial to this, there are areas that connect to the occipital lobe and to the temporal cortex (coloured red and purple, respectively, in Figure 6.3). These may include parts of medial and inferior Pu, which have been shown to connect to ventral and anterior parts of the temporal lobe in non-human primates (98; 99).

6.2.4 Distribution of thalamic pathways within the internal capsule

Thalamo-cortical pathways are highly organised within the internal capsule (102). To test the degree to which we could detect this organisation we used the same tractography approach to classify white matter voxels within the internal capsule according to the thalamic region with which they were most strongly connected. The distribution of thalamic pathways within the internal capsule was as expected (102) (Figure 6.4).

6.2.5 Probabilistic representation of connections

In figures 6.2 and 6.3, classification of thalamic voxels is generated from the highest probability of connectivity to cortical targets. However, information is available on the probability of connectivity to every cortical region for each voxel. Representation of these probabilities demonstrates that some voxels appear connected to more than one cortical region (Figure 6.5). For example, in the cluster that we propose may correspond to VA, some pathways reached PMC and others reached PFC consistent with the known distribution of cortical connections of this nucleus in the monkey (91; 103) (Figure 6.5). We also detected PFC connections from the medial pulvinar

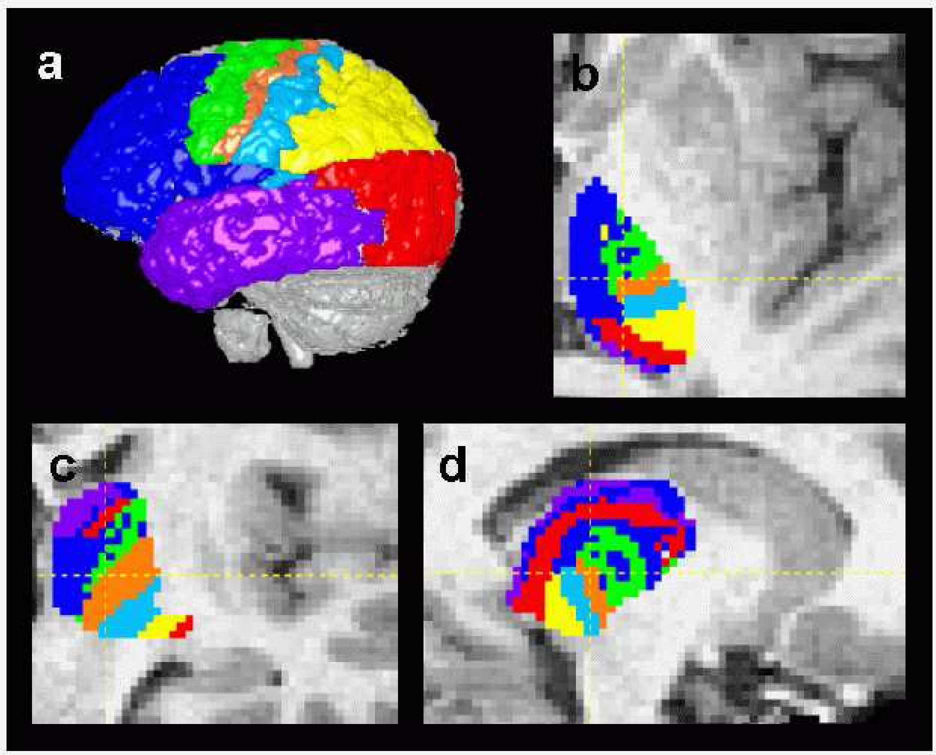


Figure 6.3: Connectivity-based segmentation of the thalamus. (**a**) Division of the cortex. (**b**)(**c**)(**d**) Classification of thalamic voxels based on probable connections to these cortical areas. We propose the following: the anterior, superior and medial purple regions (visible in **c** and **d**) include some of MD, which receives inputs from the temporal lobe and parts of the anterior complex (AV) that project to limbic areas in the medial temporal region; the more posterior and inferior purple regions (posterior to the red area, visible in **b** and **d**) includes parts of the medial and inferior pulvinar that project to the temporal lobe; the dark blue area includes some of MD, VA and parts of the anterior complex (AM, AD) that project to the prefrontal cortex; the yellow area includes the anterior parts of the pulvinar that project to PPC; the pale blue area includes LP and VPL which project to somatosensory cortices; the orange area includes VLp which projects to M1, the green area includes VLa and VA which project to premotor areas; the red areas include LGN (visible in coronal section, **c**), parts of the inferior pulvinar (most inferior red areas visible in sagittal section, **d**) and some intralaminar nuclei.

(Figure 6.5), again consistent with monkey studies (103).

6.2.6 Paths between thalamus and cortex

Each of the large cortical masks used here includes functionally and anatomically distinct subregions. Information is also available on pathway targets within each cortical mask, and the path by which they travel to the cortex. We explored this with connections from MD to temporal lobe. A similar approach could be taken with other thalamic subregions and cortical areas.

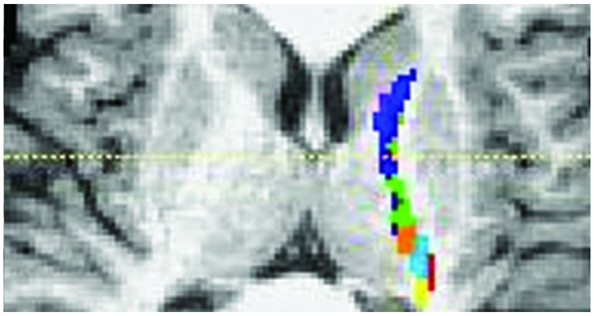


Figure 6.4: Classification of internal capsule white matter based on connections to putative thalamic nuclei. Voxels are colour-coded according to the thalamic region in Figure 6.3 with which they show the strongest connection. The anterior limb of the internal capsule contained probable pathways from putative MD that terminated in the prefrontal cortex. The posterior limb of the internal capsule contained, listed anteriorly to posteriorly, pathways from putative VA/VL to premotor cortex, from VL to motor cortex, from VP to sensory cortices, and from the pulvinar to temporal and visual cortices.

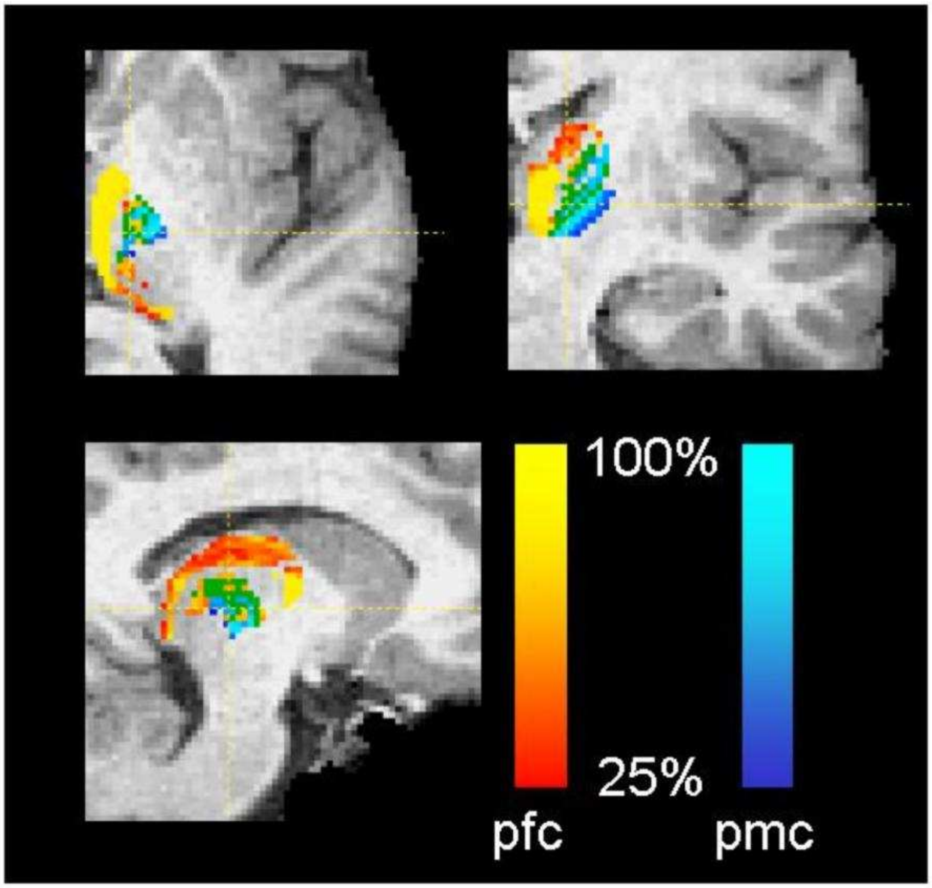


Figure 6.5: Probabilistic mapping of cortical connections. Regions of overlapping connections are illustrated in green. Probabilistic mapping of connections to prefrontal cortex (pfc) and premotor cortex (pmc) demonstrates the presence of a region, which we propose corresponds to the ventral anterior nucleus, with a suprathreshold probability of connections to both regions.

We generated pathways from all voxels within MD with predominantly temporal lobe connections (Figure 6.6 a) and found that pathways between MD and the temporal lobe take at least two distinct paths (Figure 6.6 b) and terminate in different regions (Figure 6.6 b,c,d).

6.2.7 Internal Medullary Lamina

The classification of thalamic voxels described thus far has been based only on probable connections to cerebral cortex. However, in monkey thalamus there are thalamic regions with predominantly subcortical connections and only weak or diffuse cortical connections (80). Similarly, in our data there were regions within the thalamus for which the probability of connection to cortical grey matter was very small. To better define these regions that show lower probabilities of cortical connections, we re-ran analyses with thresholding at various levels (Figure 6.7). As the threshold was increased, a region between the proposed lateral and medial nuclear groups was defined that did not show suprathreshold connectivity probability to any cortical region (Figure 6.7). Pathways from this region were generated mainly to the ipsilateral basal ganglia or to contralateral subcortical targets via the corpus callosum. We propose that the thalamic region defined in this way includes parts of the internal medullary lamina and the intralaminar nuclei. These nuclei project predominantly to the striatum in the monkey (105), although they also have diffuse cortical connections (80).

6.2.8 Reproducibility between subjects

To explore the reproducibility of the connectivity-based thalamic clusters between subjects we analysed data from an additional seven subjects. In five out of seven subjects the resulting clusters were very similar in location, ordering and size to those in the original subject (Figure 6.2.8). In two out of seven subjects the diffusion data was of poorer quality and most paths generated from some thalamic regions failed to reach the cortical masks. However, even in those two subjects, pathways from the thalamus

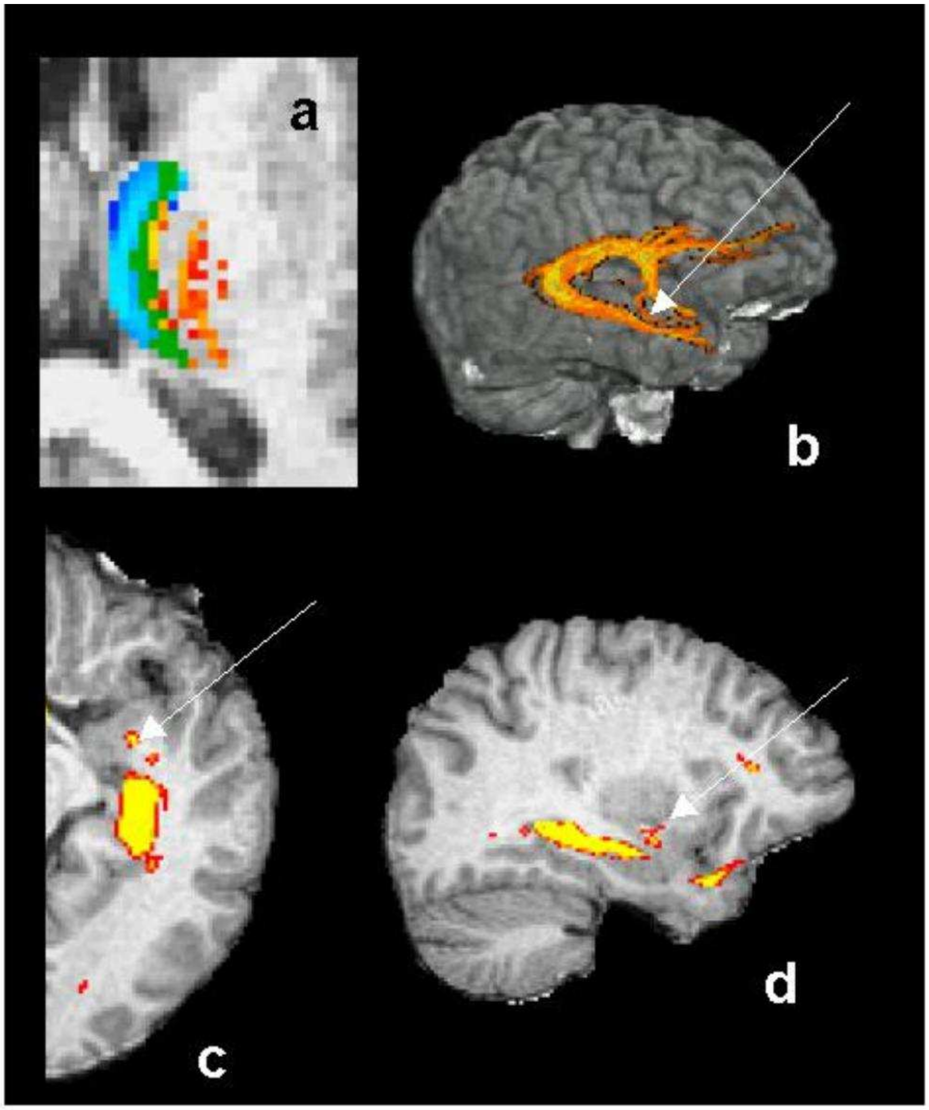


Figure 6.6: Paths from thalamus to temporal lobe. (**a**) Probabilistic mapping of connections to the temporal (turquoise to blue) and prefrontal (red to yellow) regions or both (green). To investigate the paths and termination sites of connectivity distributions to the temporal cortex we generated a representation of pathways from all medial voxels with suprathreshold connections to the temporal (and not prefrontal) region (i.e. the blue/turquoise area in **a**). (**b**) Distributions took two distinct paths to the temporal lobe. The most probable pathway from this region travels around the posterior edge of the thalamus, extends anteriorly through the hippocampal formation, then spreads before terminating in the superior temporal gyrus (temporal pole), the anterior middle temporal gyrus and piriform lobe (as in Figure 6.1 c). A less probable pathway (indicated by arrows in **b, c** and **d**) extended anteriorly and inferiorly along the medial wall of the thalamus, then turned laterally into the amygdala (**c,d**). A similar path, which travels from the amygdala via the inferior thalamic peduncle and enters the anterior head of the thalamus, has been described for the non-human primate (104)

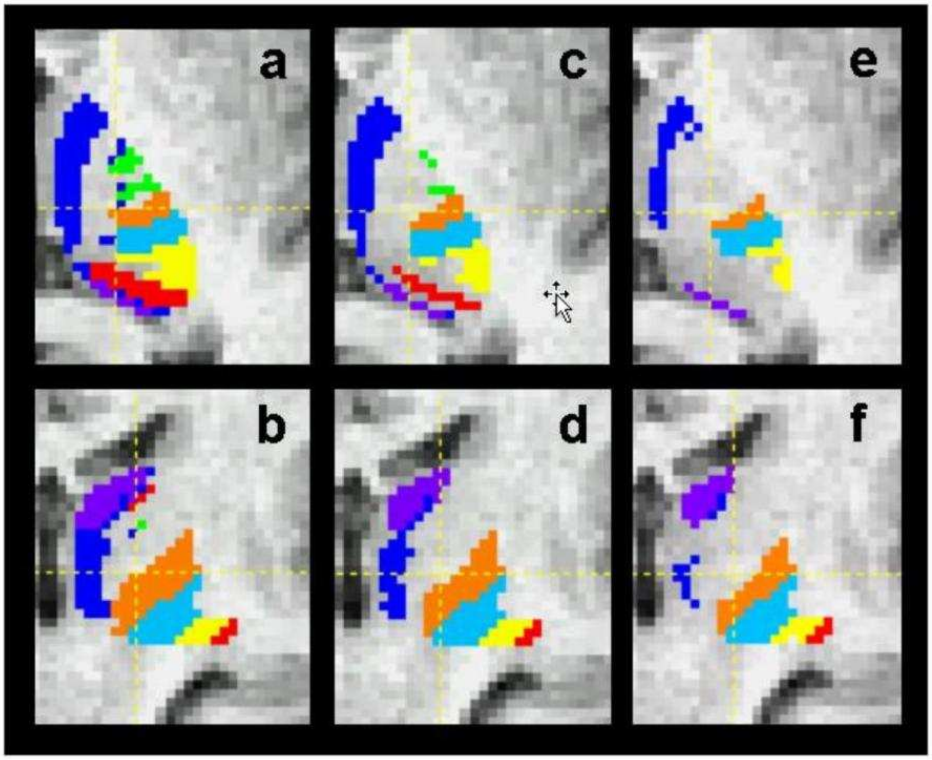


Figure 6.7: Effects of thresholding thalamic clusters based on the probability of connection to cortex. Thresholding at probabilities of 10% (**a**,**b**), 40% (**c**,**d**) and 80% (**e**,**f**).

that did reach cortex had a very similar distribution to that observed in the original subject. Explicit representation of pathway uncertainty provides a good criterion to limit false positive results using our approach.

6.3 Group Results

6.3.1 Population Connectivity Maps

Connection probabilities to each of 7 cortical masks (as in Figure 6.3) were calculated in 11 subjects. To characterise voxel-wise correspondence in thalamic connections across subjects quantitatively, we co-registered binarised masks of thalamic regions showing greater than 25% probability of connection to each of the defined cortical areas. The resulting group probability maps are centred on localised regions of high probability (across the group) of connection to each cortical region (Figure 6.9). Note that even

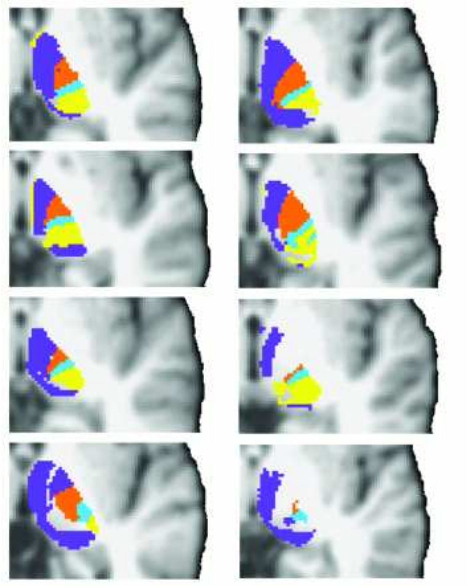


Figure 6.8: Comparison of connectivity-based segmentation of the thalamus between eight subjects (results from subject 1 are described in greater detail elsewhere in the manuscript). Voxels are colour coded as in Figure 6.2 if at least 1% of generated pathways reached the cortex. Brains have been realigned to the same standard brain template and a single axial slice (at Talairach Z = 0) is presented. The connectivity-based clusters found in subjects 1-6 were very similar. In subjects 7 and 8 there are some 'gaps' in the thalamus which correspond to regions where fewer than 1% of pathways reached any cortical mask. The patterns in these two subjects are similar to those seen in subject 1 when much higher thresholds were used (see Figure 6.7).

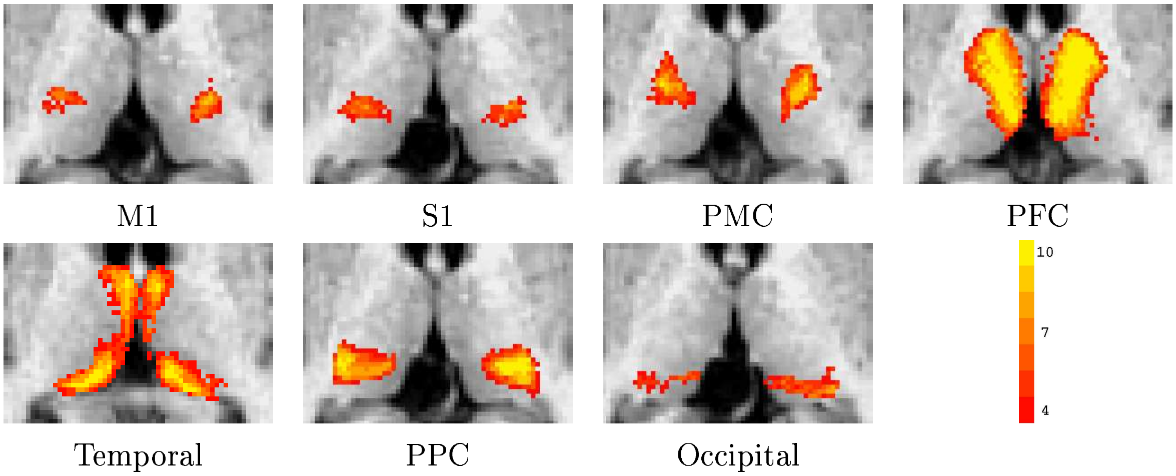


Figure 6.9: Group probability maps. Axial images showing overlap of thalamic sub-regions across subjects in voxels showing greater than 25% probability of connection to selected cortical mask (indicated in top left of each image) using a colour scale running from red (4/11 subjects) to yellow (11/11 subjects). Slices are taken at the average Z co-ordinate (given in bottom right of each image) across the left and right hemispheres for the centre of gravity of that cluster.

the smallest connectivity-defined regions (e.g. projections to M1) show areas of high probability, revealing a high degree of consistency across the population. Figure 6.10 shows a complete Talairach atlas of a binarised version of Figure 6.9.

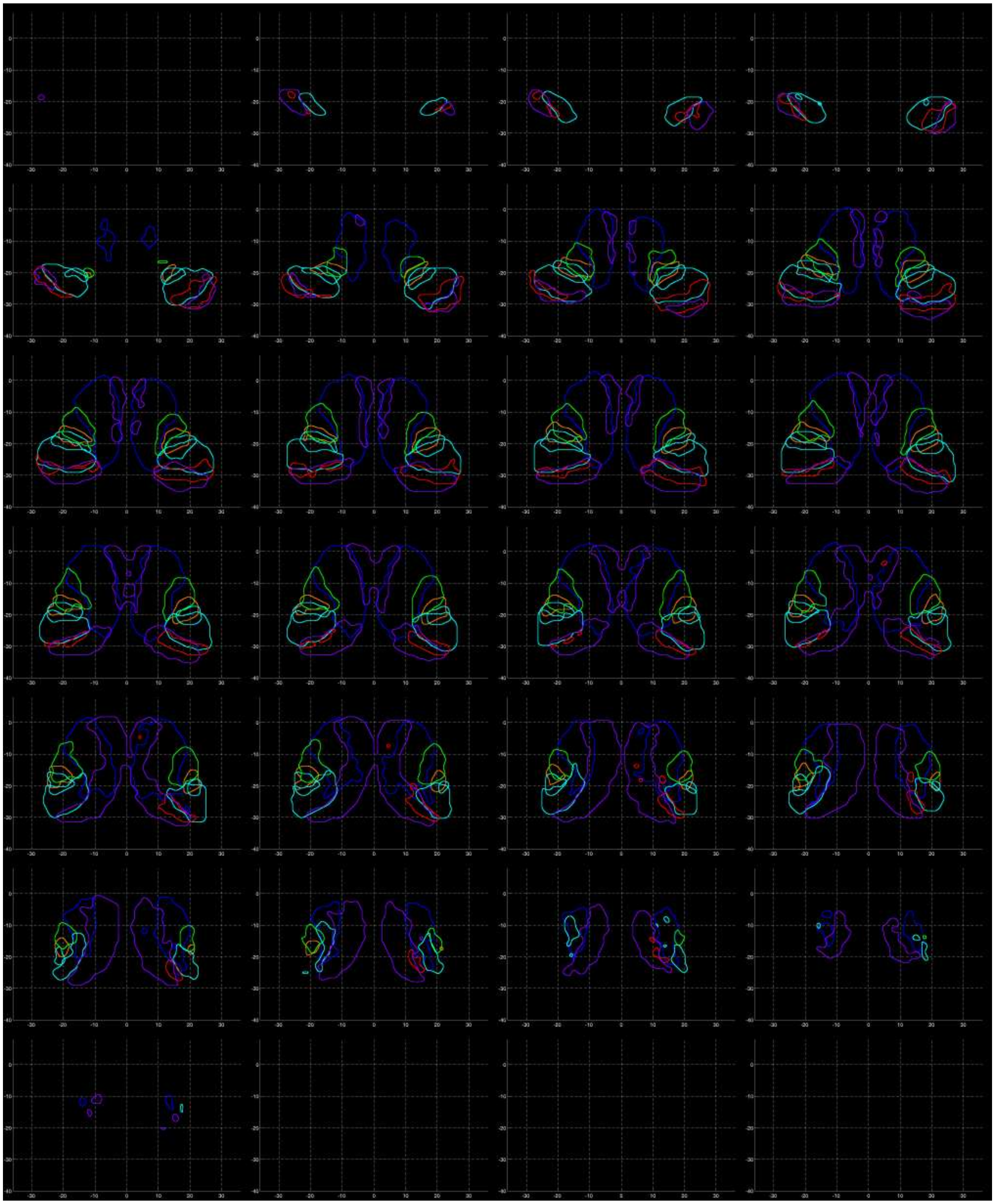


Figure 6.10: Axial slices through the whole thalamus showing edges of thresholded group probability maps for connection to each cortical region.

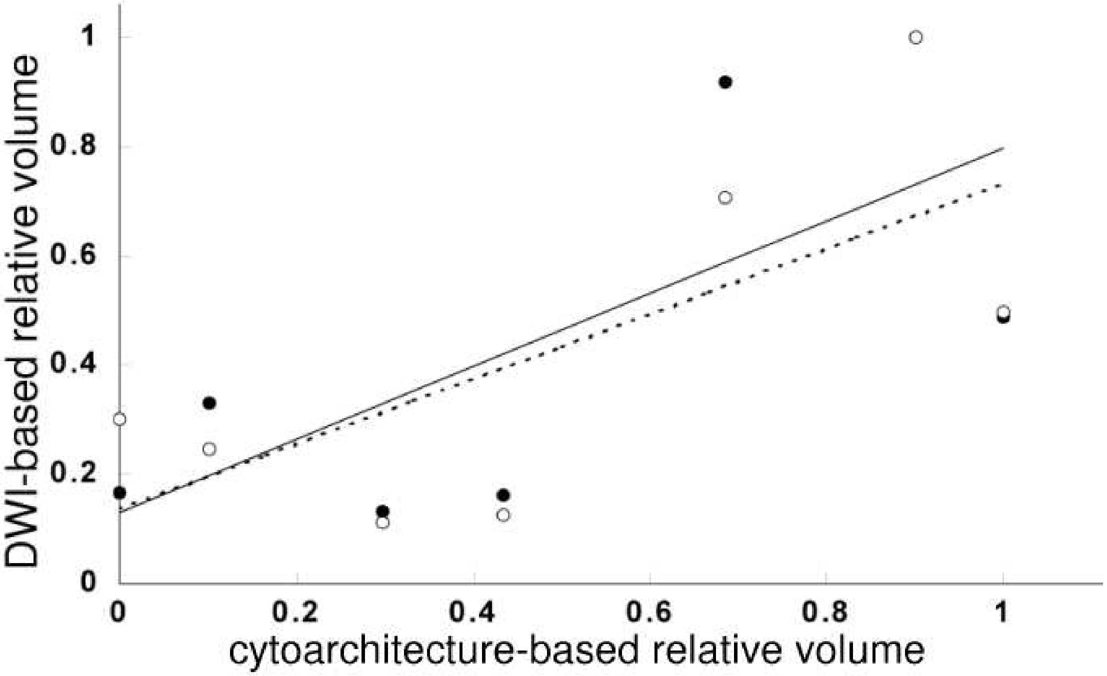


Figure 6.11: Correspondence between relative volumes based on cytoarchitectonic data and on DWI. Data is shown for the left (filled circles) and right (open circles) hemispheres. Regression lines are shown for the left (solid line) and right (dashed line) hemisphere data separately.

6.3.2 Volumetric Measurements

Having established evidence for between-subject reproducibility of our thalamic parcellation, we further tested the hypothesis that the sub-regions found in this way (i.e., the individual subject CDRs) correspond to thalamic nuclei or nuclear groups. The close correspondence between the relative locations of our CDRs and cytoarchitectonically-defined nuclei in the thalamus post-mortem (Figure 6.3) supports this claim. We also compared the relative volumes of individual connectivity-defined regions and cytoarchitectonically -defined nuclear groups in (83). We found a strong correlation between relative volumes defined using DWI and those based on cytoarchitecture (Figure 6.11, left hemisphere: r=0.71, p=0.038; right hemisphere: r=0.70, p=0.04).

6.3.3 Co-localisation with centres of functional activation

We have proposed that the atlas based on our group probability maps could be used to assign likely anatomical labels to sites of brain activation or lesions. To test directly the functional-anatomical validity of the boundaries defined by our group maps, we assessed the correspondence between centres of group maps and previously reported centres of gravity of functional activations localised to the thalamus. We found centres of thalamic activation during simple upper limb motor tasks (e.g. (106; 107; 108; 109)). Activation centres for the motor tasks co-localised well with thalamic regions connecting to motor and premotor cortices (Figure 6.3.3(a),(b)). Nine out of ten previously reported centres of motor activation fell within the M1/PMC volume ($p = 8.4 \times 10^{-5}$). Similarly, activation centres for frontal tasks co-localised with the thalamic region connecting to the prefrontal cortex, with fourteen out of sixteen previously reported centres of activation falling within the PFC volume (p=0.0025)(Figure 6.3.3(c)).

6.4 Discussion

We have shown that it is possible to trace connections from the thalamus to the grey matter of the human cerebral cortex *in vivo* using diffusion imaging. To our knowledge this is the first demonstration of anatomically specific inferences of connectivity between grey matter structures using diffusion data. This approach therefore can provide novel data on human brain connectivity. Here, we have shown that connections between the thalamus and cortex in the human are largely similar to those previously found from tracer studies in non-human primates (80; 91; 92; 93; 94; 95; 96; 97; 98; 99; 100; 101) and are reproducible between individuals. This similarity to data from non-human primates is apparent not only in the distribution of connections to different cortical sites, but also in the paths that are followed between the thalamus and the cortex. The probability of connection to different cortical zones was also used as an anatomical classifier for individual thalamic voxels. This generated clusters of voxels showing common patterns of anatomical connectivity to the neocortex. The relative locations

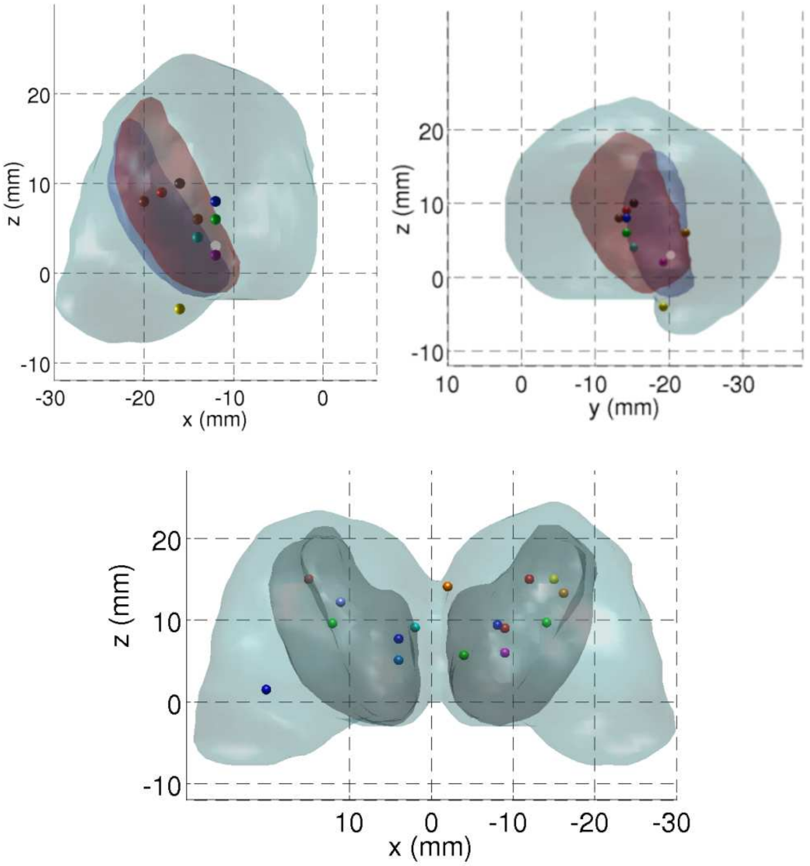


Figure 6.12: Correspondence between connectivity-defined and functionally-defined thalamic subregions. In each figure the light grey surface represents the whole thalamus, within which spheres represent centres of functional activations and volumes showing where at least 4/11 subjects had more than 25% probability of connection to particular cortical targets are represented as semi-transparent surfaces. Top: Centres of thalamic activation for right hand movement tasks and volumes with high probability of connection to primary motor cortex (dark grey) and premotor cortex (red). Bottom: Centres of activation for executive and memory tasks and the volume with high probability of connection to prefrontal cortex (dark grey). Note that some studies provided more than one activation location.

and sizes of these clusters corresponded well to previously histologically-defined human thalamic nuclei. The approach used here is generalisable and therefore may be used to map connectivities and to anatomically parcellate other grey matter structures.

This type of result should complement functional imaging studies and has the potential to provide new insights into understanding disorders associated with variations in brain structure. In section 6.3 we demonstrated these uses with functional and anatomical validation of the connectivity-based segmentations and characterisation of their population variability. We have defined the variation of the thalamic parcellation between individuals quantitatively by generating a probabilistic thalamic atlas and used this to test more rigorously the hypothesis that the parcellation corresponds to functional anatomical divisions within the thalamus. First, we found good agreement between volumes of the thalamic subregions obtained using our method and comparable sub-regions from previous cytoarchitectonic data. Second, the locations of the individual regions correspond well with data from prior functional imaging experiments. We showed, for example, that thalamic activations with motor paradigms map into a region corresponding to the ventral lateral nucleus, while activations associated with tasks involving executive control and working memory co-localised with the connectivity-defined sub-region including the mediodorsal nucleus.

Tracing fibres to cortex has been notoriously difficult when using the maximum likelihood approaches generally employed for DTI tractography. As fibres approach cortex, diffusion anisotropy reduces and calculated principal diffusion directions become increasingly uncertain due to image noise (55). This problem is so pronounced that streamlining algorithms used to date have had to apply an arbitrary anisotropy threshold which forces the early termination of reconstructed pathways (3; 90; 9). This has limited attempts to trace pathways directly from deep grey matter, which typically has low anisotropy. In these circumstances a fully probabilistic algorithm has significant advantages. First, because an explicit representation of uncertainty in path direction is generated, the relative probabilities of directions can be estimated, and the pathway can continue even if the probability is low for any single direction. A second useful advantage of a probabilistic algorithm is robustness to noise. It can be difficult to track beyond a noisy voxel using a non-probabilistic algorithm as it may initiate a meaningless change in path. However, with a probabilistic algorithm, paths which have taken errant routes tend to disperse quickly, so that voxels along these paths are classified with low probability. In contrast "true" paths tend to group together, giving a much higher probability of connection for voxels on these paths.

We have used patterns of thalamo-cortical connectivity inferred from diffusion images to parcellate anatomically distinct regions of the thalamus. While related by use of a common imaging technique, this approach is distinct from the differentiation of thalamic regions on the basis of local diffusion properties (110) and offers significant advantages as we consider long-distance connectivities. For example, the lateral sensorimotor nuclear group (VLa, VLp, VP) contains voxels that have very similar principal eigenvectors from a diffusion tensor fit, but pathways from the nuclear subdivisions terminate in distinct anatomical regions, allowing their discrimination if connections are followed up to the cortex. Also, by considering not only diffusion measurements within a voxel, but also information available at a more global scale, the thalamic segmentation achieved has a finer spatial resolution than that of the original diffusion images. Together, these characteristics of our approach have allowed us to make strong tests of predictions concerning homologies between human and non-human primate thalamic organisation.

There are limitations to this approach. Firstly, it is impossible to define fibre tract polarity (e.g., distinguishing thalamo-cortical from cortico-thalamic connections) using diffusion tractography. Secondly, the method is sensitive primarily to major pathways, and therefore smaller pathways, pathways with sharp path inflections or pathways that cross other tracts are not always detected by our method. For example, from seeds placed within the grey matter of the LGN we were able to detect the dorsal portion of the optic radiation that travels directly to the visual cortex (Figure 6.1 a). However, detection of the more ventral part of the optic radiation that travels to visual cortex via the temporal cortex (forming ‘Meyer’s loop’) was possible only from seeds placed in the white matter just lateral to the LGN. Use of higher resolution imaging and of analysis approaches that explicitly model complex fibre structures within a voxel (111) may increase sensitivity to such paths.

In interpreting the connectivity distributions presented here it is important to note that our current approach does not allow us to infer the true distribution of fibre structure within a voxel, but to estimate the uncertainty associated with the principal direction in order that we may quantify our confidence in the tractography results. Therefore the distributions that we present are not estimates of true fibre spread from a voxel, but of the confidence bound on the location of the most probable single connection. The width of these confidence bounds will depend both on experimental noise and on any mismatch between our model (a single population of parallel fibres) and the diffusion data (which may reflect complex structure, such as crossing fibres).

It has previously been difficult to directly validate diffusion tractography due to a lack of alternative methods providing similar data in humans, as well as the relative paucity of animal imaging data. Here we took two approaches to validation of the connectivity-based parcellation: first by comparison to cytoarchitectonic atlases and second by comparison to functional activations.

The demonstration that relative locations and volumes defined using a connectivity-based parcellation of the thalamus generally correspond well with volumetric measurements from cytoarchitectonic maps is a powerful test of the ability of the method to define correct targets and, by implication, to identify appropriate paths. We would not expect the absolute volumes to correspond between these two methods. While cytoarchitectonic measurements were made up only of volumes from major nuclei, the measurements based on diffusion imaging would have included contributions from smaller nuclei and from white matter regions within the thalamic volume. Therefore, not unexpectedly, the absolute volumes of sub-regions are greater for the connectivity-based measures.

Relating the thalamic parcellation to locations of centres of functional activations offers an alternative approach to its validation (and also illustrates an important potential application of the parcellation method). Although thalamic activation is frequently reported in imaging studies (112; 109; 106; 107), authors rarely assign activations to a specific nucleus, due to the problems of inferring nucleic architecture with current approaches. Here we have chosen to test the functional validity of connectivity-defined volumes with data from well-characterised tasks to minimise confounds in interpretation. Centres of activation associated with two distinct types of tasks showed good functional-anatomical correspondences: motor activations co-localised to sub-region corresponding to the ventrolateral nucleus and activations during executive tasks were mainly located within the subregion corresponding to the mediodorsal nucleus. The atlas based on our group maps presented here can be used to assign a probabilistic anatomical label to activation foci within the thalamus more generally. This could prove particularly useful in cases where functional connectivity between the thalamus and neocortex is otherwise ambiguous. For example, although thalamic activation is frequently reported in studies of pain (112; 113), the nature of the thalamic processing remains unclear and would be illuminated substantially by a clarification of precisely which thalamic nuclei are involved in processing nociceptive versus non-nociceptive stimulation in the human brain.

While the patterns that we found correspond well with those defined by direct anatomical studies of non-human primates and are consistent with previous functional and anatomical studies in human, further validation of this approach is desirable. For example, cytoarchitectonic boundaries could be defined on the same population imaged and parallel diffusion and manganese-tracer imaging (114) could test the correspondence between probabilistic tracts and real fibre pathways in non-human primates.

The current study was limited to the investigation of thalamic connections to large cortical regions. However, the non-human primate literature has provided strong evidence for finer grained topographic mapping, e.g., between subregions of MD and smaller regions in the prefrontal cortex (103; 115). Such investigations in the living human brain could proceed as a straightforward extension of the current investigation. There are obvious applications of our method to clinically important problems. Probabilistic tractography could help characterise developmental and acquired brain disorders, e.g., testing for alterations in fronto-thalamic circuitry in Schizophrenia (116) for the relations between quantitative differences in cortico-cortical connectivity and learning abilities (117). The ability to define regional boundaries reliably in grey matter could be useful, e.g., in improved targeting of specific sub-cortical structures in the treatment of movement disorders by functional neurosurgery (82).

6.5 Methods

6.5.1 Data Acquisition

Diffusion-weighted (DW) data were acquired in 11 healthy subjects (6 male, 2 female, aged 26 to 33) using diffusion-weighted echo planar imaging (60×2.3mm thick slices, field of view = $220 \times 220\text{mm}^2$, matrix = 96×96; images were reconstructed on a 128×128 matrix giving a final resolution of $1.7 \times 1.7 \times 2.3\ \text{mm}^3$) implemented on a General Electric 1.5 T Signa Horizon scanner with a standard quadrature head-coil and maximum gradient strength of 22 mT m^{-1}. Informed written consent was obtained from all subjects in accordance with ethical approval from the Institute of Neurology. The diffusion weighting was isotropically distributed along 54 directions (δ = 34ms, Δ = 40ms, b-value = 1150 smm^{-2}, the optimum for white matter diffusion-weighted imaging (DWI) measurements)(33). 6 diffusion-weighted volumes (b-value = 300 smm^{-2}) and 6 volumes with no diffusion weighting were acquired. The high angular resolution of the diffusion-weighting directions (33) improves the robustness of *pdf* estimation by increasing the signal-to-noise ratio per unit time and reducing directional bias. Cardiac gating (118) was used to minimise artefacts from pulsatile flow of the cerebrospinal fluid. The total scan time for the DWI protocol was approximately (depending on heart rate) 20 minutes.

The high resolution T1-weighted scan was obtained with a 3D inversion recovery prepared spoiled gradient echo (IR-SPGR) (FOV = 310×155; matrix = 256×128; in-plane resolution = $1.2 \times 1.2\ \text{mm}^2$; 156×1.2mm thick slices; TI = 450ms; TR = 2s; TE = 53ms).

6.5.2 Thalamic segmentation.

We manually outlined the whole thalamus and a number of cortical zones (see below) on each subject's T1-weighted image. For tissue-type segmentation, skull stripping and registration, tools from FSL (www.fmrib.ox.ac.uk/fsl) were used. We performed probabilistic tissue type segmentation and partial volume estimation on the T1-weighted image (119). We thresholded these results to include only voxels estimated at greater than 35% grey matter occupancy and used this to mask our cortical regions. We skull-stripped the diffusion-weighted and T1-weighted images (120) and performed affine registration (59) between the first non diffusion-weighted volume and the T1-weighted image to derive the transformation matrix between the two spaces. From each voxel in the thalamus mask, we drew samples from the connectivity distribution (as described in the previous chapter), maintaining knowledge of location in structural and DWI spaces, and recorded the proportion of these samples that passed through each of the cortical masks as the probability of connection to that zone. Hard segmentation was performed by classifying the seed voxel as connecting to the cortical mask with the highest connection probability (e.g., Figures 6.2 and 6.3). In order to determine whether voxels showed strong or weak cortical connectivity, thresholding was carried out based on the probability of reaching any cortical mask (e.g., Figure 6.7). Some voxels showed a high probability of connection to more than one cortical mask; in order to investigate overlapping or branching connections from individual voxels, a probability was assigned for each mask at each voxel, based on the percentage of cortically-connected pathways that reach the mask (e.g., Figure 6.6).

6.5.3 Cortical masks.

We defined cortical zones which correspond to known thalamic connection sites (80). First, we defined four exclusive cortical zones using the following landmarks on the T1-weighted image:

Prefrontal/temporal zone: The prefrontal cortex was defined as the lateral and orbital cortex anterior to the inferior and superior precentral sulci, from the dorsal to the orbital brain surface. Medially, this region included the cortex superior to the cingulate sulcus, anterior to the vertical line from the anterior commissure, perpendicular to the plane between the anterior and posterior commissures (VAC) and also the cingulate gyrus, from its anterior edge to the level of the central sulcus. The temporal lobe was defined as the lateral cortex inferior to the lateral fissure, bordered posteriorly by the temporo-occipital incisure where visible, or the anterior occipital sulcus. Medially, the temporal lobe was defined as the cortex anterior to a line from the temporal incisure to the posterior commissure.

Motor zone: This region included M1, SMA and PMC. Laterally, this included cortex anterior to the central sulcus up to and including both banks of the inferior and superior precentral sulci. Medially, this region included cortex anterior to the central sulcus up to the VAC line. Both lateral and medial parts of this region extended from the dorsal surface of the brain to the level of cingulate sulcus.

Somatosensory zone: This region included SI and SII. SI was defined as the posterior bank of central sulcus plus postcentral gyrus, from the dorsal surface of the brain to the lateral fissure. SII was defined as the superior bank of lateral fissure posterior to the posterior edge of postcentral gyrus.

Parieto-occipital zone: The posterior parietal cortex was defined as the lateral cortex posterior to postcentral gyrus, bordered posteriorly and inferiorly by a line from the posterior-occipital fissure to the lateral fissure. Medially, posterior parietal cortex included regions posterior to the postcentral gyrus, bordered posteriorly by the parietal-occipital fissure and inferiorly by the cingulate gyrus. The occipital lobe was defined as the lateral cortex posterior to anterior occipital sulcus and the temporo-occipital incisure, bordered superiorly by a line from the parietal-occipital fissure to the lateral fissure and the medial cortex inferior to the parietal-occipital fissure and posterior to a line from the temporal incisure to the posterior commissure. The optic tract, where visible, was also included with the occipital lobe in this target mask.

Further analyses were carried out with seven cortical masks. We separated primary motor cortex (anterior bank of central sulcus and posterior half of precentral gyrus) from the premotor and supplementary motor areas (the remainder of the motor zone). We also separated the posterior parietal cortex from the occipital lobe and separated the temporal lobe from the prefrontal cortex.

The insular cortex was not considered in this analysis.

Chapter 7

Recent Developments

7.1 New Findings with Diffusion Tractography?

In chapter 6, we showed examples in which we were able to confirm a similarity between anatomical connections in human derived from non-invasive diffusion tractography and those in non-human primate derived from invasive tracer methods. This type of study should lend itself well to many other structures in the brain where the anatomy is known in primate, and expected to be similar in human. The major advantage of using DTI in these studies is its non-invasive nature, potentially allowing the kind of connectivity maps shown in chapter 6 (e.g. figure 6.3) to be used to search for differences between control and patient groups. Taking the thalamus as a specific example, the connectivity between medio-dorsal nucleus of thalamus and prefrontal cortex has been highlighted as a potential substrate for Schizophrenia (116). The kind of analysis presented here seems ideally suited to the investigation of this kind of clinical question. The potential to identify subcortical nuclei for the first time in a non-invasive fashion also has potential clinical applications. Functional neurosurgery in patients with movement disorders (such as Parkinson's disease) relies on the targeting of specific subcortical nuclei for (for example) electrode implantation. A non-invasive technique for locating such nuclei with increased accuracy could only benefit the outcome of the surgery.

However, an interesting question which has not yet been addressed by the diffusion tractography literature is whether, given the caveats with the methodology discussed in chapters 5, 6 and, in some depth, in 8, we are able to use diffusion tractography to confirm or deny *new* hypotheses about the connectional architecture of the human brain. Specifically, can we, with any confidence, describe anatomical connectivity patterns which are different from those derived from non-human data?

The ideal scenario for answering this question would be that diffusion tractography would advance to the stage where a direct validation with invasive tracer studies in the same non-human primate brain would reveal no "false-positives" in the results derived from diffusion imaging. In this case, new results in the *human* brain could be interpreted with confidence as connections unique to humans, and not as potential artefacts of the methodology. However, we are not yet at the stage to perform such a study. Firstly, diffusion tractography *can* lead to false positives. In areas of complex fibre structure, where the chosen models of diffusion break down, simple fibre tractography may end up following entirely wrong directions. Secondly, sacrificial tracer methodologies only trace connections from a small number (1-3) of very specific seed locations in each experiment. A test on the scale required to inspire confidence in diffusion tractography would require many invasive experiments at a large cost, both financially and in terms of animal life. Such a study is not appropriate while the methodologies used for diffusion tractography are changing at their current pace.

However, even in the absence of such a study - in fact, even in the expectation that in some cases such a study would fail - it still may be possible to find evidence for previously unknown connections from diffusion tractography. One study, currently in progress, is using precisely the methodology described in chapters 4 and 5 to investigate cortico-cerebellar circuitry. Projections to cerebellum from cortex pass through cerebral peduncle (see figure 7.1 and (121)). An examination of cerebral peduncle similar to that performed in thalamus in chapter 6 should provide some insight into the areas of cortex sending signals to cerebellum, and hence into the type of information which cerebellum is processing. The traditional view of cerebellum is that it processes mainly low-level information such as signals from cortical motor areas (122). In accordance

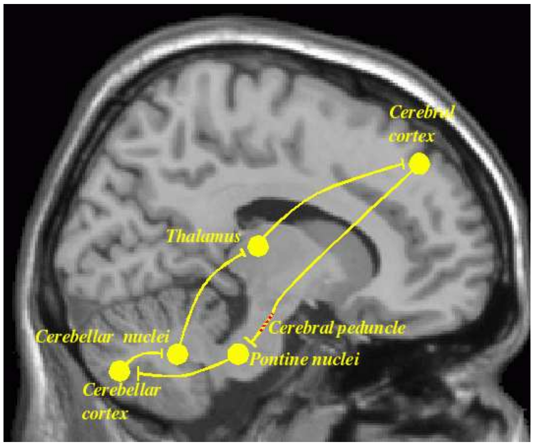
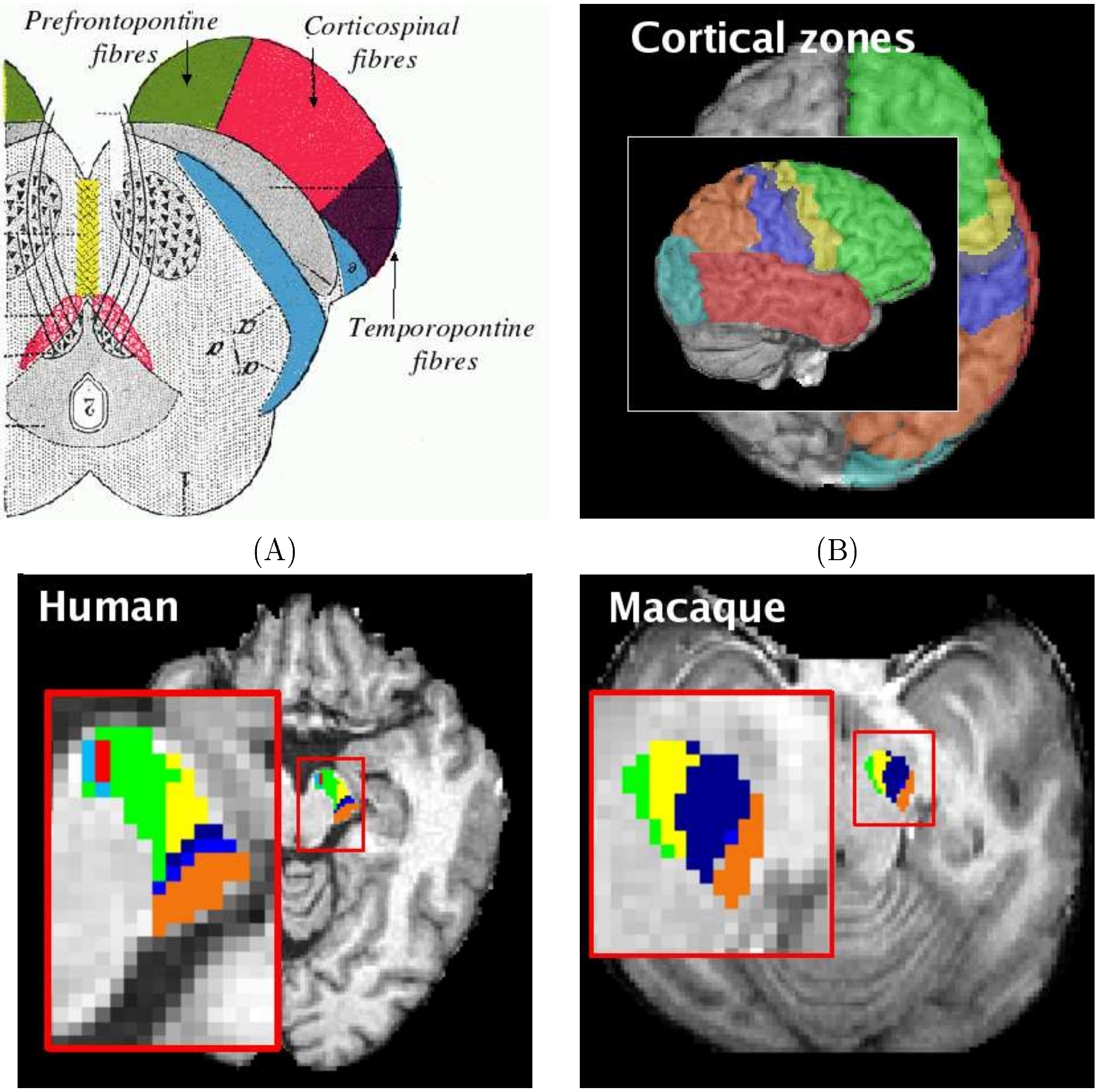


Figure 7.1: Schematic of cerebello-cortical circuitry. Projections from cerebellum terminate on thalamic neurons that project to cerebral cortex. These cortical neurons send return projections to cerebellum via cerebral peduncle, and relay neurons in the pontine nuclei.

with this view, non-invasive tract tracing studies in macaque monkey have revealed that connections from cortex to pons (and thence to cerebellum) are organised topologically as they pass through cerebral peduncle, with anterior to posterior areas in cortex being represented medially to laterally in peduncle (123; 124), but that the overwhelming majority of fibres which pass through peduncle have their source in motor/sensory areas in cortex (red area in figure 7.2 (A)) (121; 125). In human, however, the situation may be slightly different. Recent functional imaging studies have revealed cerebellar activations in much higher level cognitive tasks (10; 11), suggesting a different role for cerebellar circuitry in the human brain.

A connectivity-based segmentation, such as was carried out in chapter 6, would seem to be an ideal tool to investigate this question. If, as in macaque, the majority of cerebral peduncle is found to have a high probability of connecting to motor areas, this would offer supporting evidence for the conventional view of cerebellar processing. However, if a large part of peduncle is found to have a high probability of connection to frontal lobe, and specifically, prefrontal cortex, this would favour the alternative view, that cerebellar circuitry has a role to play in cognitive processing.

Probabilistic tractography was carried out in 11 subjects exactly as in chapter 6 but seeded from voxels in the cerebral peduncle. Again, exactly as in chapter 6, each



(A) (B)

(C) (D)

Figure 7.2: Cortical connectivity in the cerebral peduncle (A) derived from invasive tracer studies in macaque, (b) derived from diffusion tractography in human. (C) Colour code for cortical zones in (B).

voxel was assigned to the cortical zone with which it had the highest probability of connection (cortical regions are shown in figure 7.2 (B)). Results were consistent across eleven subjects.

A typical result of the connectivity based segmentation may be seen in figure 7.2 (C). It can be seen that the results from diffusion tractography in human exhibit exactly the *topology* predicted from tracer studies in macaque, but that the relative *volumes* associated with the different cortical zones are strikingly different. Most noticeably, the relative volumes associated with premotor and primary motor/sensory cortical regions (yellow, navy blue and royal blue) are much smaller than in the predictions from macaque, and the relative volume associated with prefrontal cortex (green) is much larger.

This result is very appealing. If the relative volumes seen in the connectivity based segmentation reflect the ratios in which different cortical areas send projections to the pontine nuclei and on to cerebellum, then this would provide evidence in favour of a role for cerebellum in cognitive processing. However, we must first consider an alternative. We are attempting to make a scientific statement based on a comparison of data from macaque with data from human, but the two sets of data come from very different sources. Is it not possible that we are seeing some kind of measurement bias? How confident can we be that the tractography results reflect the underlying truth in human? Again, is it possible to make new findings with diffusion tractography?

One possible answer to this question lies in the topology found in both (A) and and (C) in figure 7.2. The exact same topology is present in both sets of data, giving extra weight to the differences seen in the relative volumes. That said, far stronger evidence is shown in figure 7.2 (D). Here diffusion tractography and connectivity-based segmentation are carried out in the cerebral peduncles of a macaque monkey. The result shows the same topology as the human data and the predictions from macaque tracer studies but this time, as in figure 7.2 (A), the majority of the volume of peduncle has a high probability of connection to motor regions, with prefrontal cortex having only a small representation. Comparing (D) with (A) allows us to compare results from two different methodologies without the confounding effects of two different species. Comparing (D) with (C) allows us to compare the two different species without the confounding effects of two different methodologies[1].

---

[1] This work was carried out in collaboration with Narender Ramnani, FMRIB Centre, Oxford

7.2 Connectivity Matrices - Defining Anatomical/Functional Subunits

We saw in chapter 6 that, using probabilistic tractography, we were able to gather connectivity information between thalamus, which includes many discrete functional subunits, and manually labelled cortical zones which each share function with at least one of those subunits. We were then able to use this connectivity information to parcellate thalamus into commonly connected clusters, which correspond well with histologically defined thalamic nuclei - the functional subunits of interest. However, in order to achieve this parcellation, we required a strong hypothesis for the cortical connectivity of the individual thalamic nuclei. In the case of thalamus, this information was readily available, as thalamo-cortical connections in human are expected to be similar to those in monkey, which have been studied extensively. However, there are many areas in the brain where we are much worse informed, but where we would still hope to be able to learn from connectivity based parcellations.

7.2.1 Intra-connected systems

The brain is split into systems which are densely intra-connected, but only sparsely inter-connected. Often, these systems are themselves split in a similar fashion, so that regions of the brain may be classified in a hierarchical fashion. A schematic of such a system is shown in figure 7.3.

A defining feature of such a system is that, if the nodes in the system are arranged as the rows and columns of a matrix with each matrix element being the connectivity strength between the two nodes in question, then this “connectivity” matrix may be permuted (by a simple permutation of the ordering of the nodes) such that the elements in the matrix appear in clusters around the leading diagonal. An example of a connectivity matrix from a hierarchically connected system with binary connections may be seen in figure 7.4.

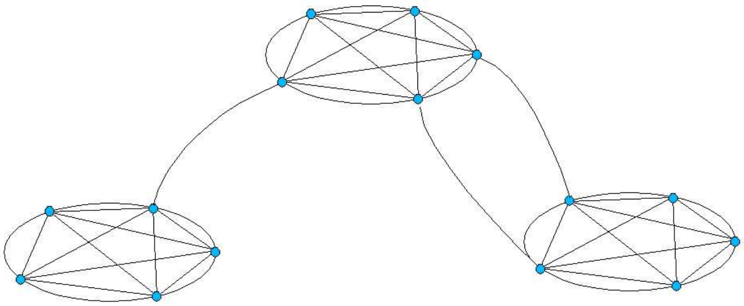


Figure 7.3: Schematic of a hierarchically organised network. Functional subunits are densely intra-connected but sparsely inter-connected.

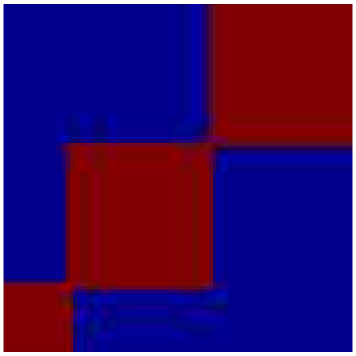


Figure 7.4: Example connectivity matrix from an “ideal” hierarchically connected system. Connections strengths are binary, and appear in clusters around the leading diagonal to form densely intraconnected subunits

As the underlying connectivity structure becomes less and less hierarchical, the structure in the connectivity matrix will become less and less identifiable. For example, if links were made between the two lower subunits of the system, in figure 7.3, then we would see clusters of connectivity in the off diagonal corners in figure 7.4. However, the existence of discrete densely intra-connected subunits in the system will still reveal itself with clusters of high connectivity around the leading diagonal, and in their absence it should prove impossible to organise the matrix in such a fashion.

If diffusion tractography data, when organised in matrix form, could be permuted to reveal such a structure, then we would be able to parcellate the brain into anatomical subsystems without the need for any prior hypothesis for the connectivity of any particular regions.

7.2.2 Matrix Permutation

The search for organised structure in matrices forms a large research area in its own right. Algorithms range from the very simple (for example Hierarchical Clustering simply repeatedly groups into a single "node" the two "nodes" in the current matrix with the highest connectivity) to the very complex (for example "Independent Component Analysis (ICA)" searches for mixtures of statistically independent processes which may represent the data (126)). An important topic for further work in this area will be to develop algorithms which efficiently search the connectivity matrix for clusters in the processing hierarchy of the kind expected to represent functional subunits in the brain. However, the early results shown in the next section were generated by an algorithm with, at least explicitly, a different aim. The "NCut" algorithm, described below, aims to find the matrix node reordering which minimises the sum of each element value multiplied by its square distance from the leading diagonal.

$$p = arg \min_{\Omega_p} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{A}(i,j) \times (i-j)^2 \right] \tag{7.1}$$

where $\mathbf{A}$ is the $N \times N$ connectivity matrix, Ω_p is the space of all possible node permutation vectors, and p is the "optimal" permutation vector searched for by NCut.

Effectively, Ncut is searching for a node reordering which forces large connectivity values close to the leading diagonal. In the case of a perfectly hierarchically connected system, such as the matrix in figure 7.4, this reordering will organise the matrix such that connected clusters appear contiguously in the reordered matrix. The major advantage of the NCut algorithm is its computational efficiency. The reordering process merely requires the computation of a single eigenvector of an $N \times N$ matrix. This may be expected to prove of great importance for reordering the types of connectivity matrices we can generate with diffusion tractography. Even at present voxel resolution, grey matter in a diffusion weighted image comprises around 100000 voxels, giving a potential $10^5 \times 10^5$ grey matter connectivity matrix. Data mining on this scale is simply unfeasible with the majority of available algorithms.

NCut The Ncut algorithm (127) provides a computationally efficient approximate solution to the problem stated in equation 7.1. This problem may be restated (see (127),(128)) as one of finding the reordering vector p which minimises the following quadratic matrix equation.

$$p = arg \min_{\Omega_p} p^T \mathbf{Q} p \tag{7.2}$$

where

$$q_{ij} = \begin{cases} -a_{ij} & \text{for } i \neq j \\ -\sum_{j=1, j\neq i}^{N} q_{ij} & \text{for } i = j \end{cases} \tag{7.3}$$

Now a heuristic is introduced which makes the problem tractable at the expense of computing a guaranteed optimal solution. Instead of limiting the search to the discrete set of permutation vectors Ω_p, we relax the problem to find the minimum over real vectors, $\mathbf{x}$, of length N, under the caveat that the following trivial solutions are ignored: $\mathbf{x} = \mathbf{0}$ and $\mathbf{x} = \mathbf{e}$, where $\mathbf{0} = [0, 0, \ldots, 0]$ and $\mathbf{e} = [1, 1, \ldots, 1]$.

$$x = arg \min_{x \in \Re : x^T \mathbf{e} = 0, ||x||_2 = 1} x^T \mathbf{Q} x \tag{7.4}$$

which is solved by taking x to be the eigenvector corresponding to the second smallest eigenvalue of $\mathbf{Q}$ (see (127)), equivalently referred to as the Fiedler vector of $\mathbf{A}$. This solution is a real vector, not the desired permutation vector; however we can use the ordering of the elements in $\mathbf{x}$ to induce a permutation vector $\mathbf{p} \in \Omega_p$. We choose $\mathbf{p} \in \Omega_p$ such that $p_i \leq p_j$ if and only if $x_i \leq x_j$. Spectral reordering via the Ncut algorithm is the application of this ordering $\mathbf{p}$ to matrix $\mathbf{A}$.

Simulated Data Figure 7.5 shows the results of applying NCut spectral reordering to two types of simulated connectivity matrix. The first is a random matrix with no underlying structure, (A). Neither the Fiedler vector, (B) and sorted in (C), nor the reordered matrix, (D), show any structure. The second is the "ideal" hierarchical connectivity matrix from figure 7.4 but with the nodes of the matrix permuted in a random order (E). The Fiedler vector, (F) and sorted in (G), shows 3 distinct clusters, and the reordered matrix, (H), shows the recovered structure from figure 7.4.

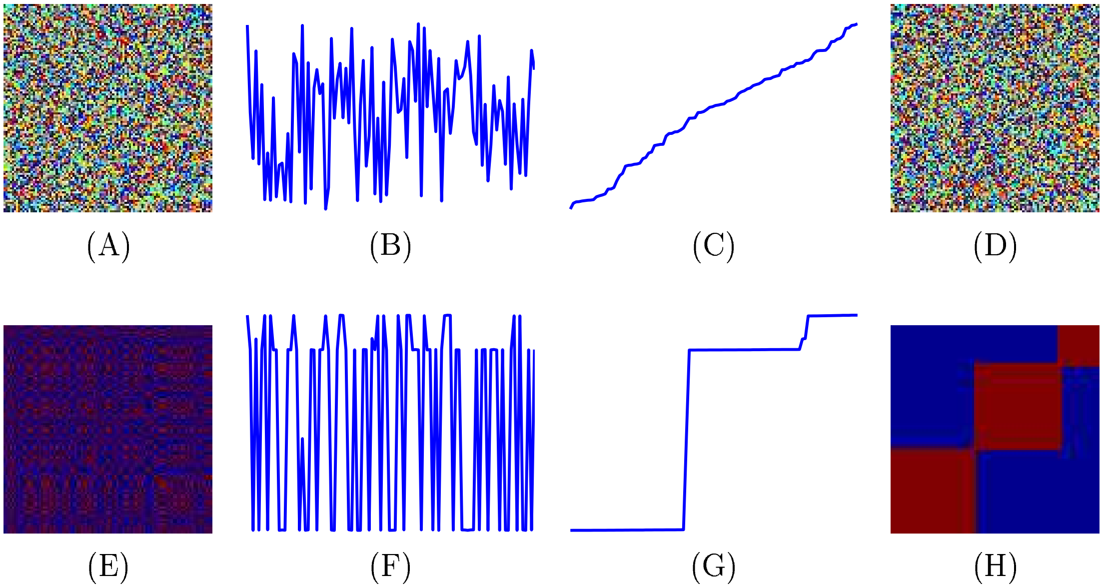


Figure 7.5: Results of applying spectral reordering to simulated data. Top row - reordering of a random matrix (drawn from $\mathcal{U}(0,1)$). (A) original matrix. (B) Fiedler vector. (C) Sorted Fiedler vector. (D) reordered matrix. Bottom Row - reordering of a matrix with an underlying hierarchical structure such as in figure 7.4. (E) original matrix. (F) Fiedler vector. (G) sorted Fiedler vector. (H) reordered matrix.

7.2.3 Connectivity Data

A good starting point for this type of unsupervised clustering approach is the visual system, which, in primate brain, exhibits exactly the kind of hierarchical connection structure which the matrix permutation approach relies on. In simple terms, visual information is sent from primary visual cortex (V1), along two broadly parallel processing streams (129) (See figure 7.6). The ventral pathway processes object form information and includes projections to inferior temporal lobe. The dorsal pathway processes information on object motion and includes projections to posterior parietal cortex. These two processing streams are densely intra-connected but sparsely inter-connected, hence, at the coarsest scale, we would expect to see structure in our connectivity matrix much like that in figure 7.4

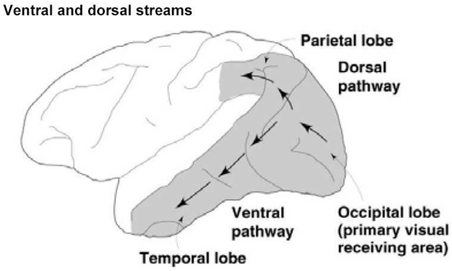


Figure 7.6: Schematic of the organisation of primate visual system. Visual information is sent from primary visual cortex along two broadly parallel processing streams. See text for details. Taken from (129).

Methods Diffusion weighted data were acquired as in chapter 6. Masks were drawn manually of parietal, temporal and occipital lobes using the same landmarks as in chapter 6. Probabilistic tractography (chapter 5) was carried out from each voxel on a $10 \times 10 \times 10mm^3$ grid within these masks. Connectivity values were stored from each seed voxel, to every other voxel in the masks. Connectivity information was arranged in a matrix, $\mathbf{A}$. The natural logarithm of this matrix was taken to decrease the effective range of the matrix, and hence render it more suitable for spectral reordering, $\mathbf{B} = \log \mathbf{A}$. Symmetry was imposed on the matrix $\mathbf{C} = (\mathbf{B} + \mathbf{B}^T)/2$. Connectivity matrix $\mathbf{C}$ was submitted to the NCut spectral reordering routine described above.

Results Figure 7.7 shows the connectivity matrix in its original form (left) and after spectral reorganisation (right). The reordered matrix was examined by eye, and divided, at the coarsest scale, into three clusters (horizontal coloured bars in the right hand panel of figure 7.7). These clusters were then mapped back onto their anatomical location, as can be seen in figure 7.8. The first cluster (blue) included lateral parietal lobe, corresponding to putative dorsal visual stream. The second cluster (yellow) included inferior temporal lobe corresponding to putative ventral visual stream. The third cluster (red) included medial occipital lobe, corresponding to putative primary visual cortex.

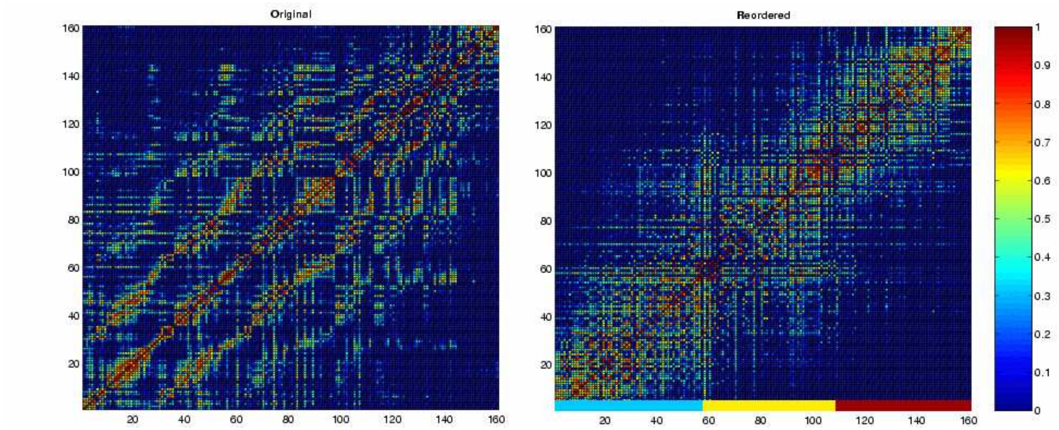


Figure 7.7: Connectivity matrices from the human visual system. (A) Before spectral reordering. (B) After spectral reordering. Clusters were chosen according to the coloured horizontal bars.

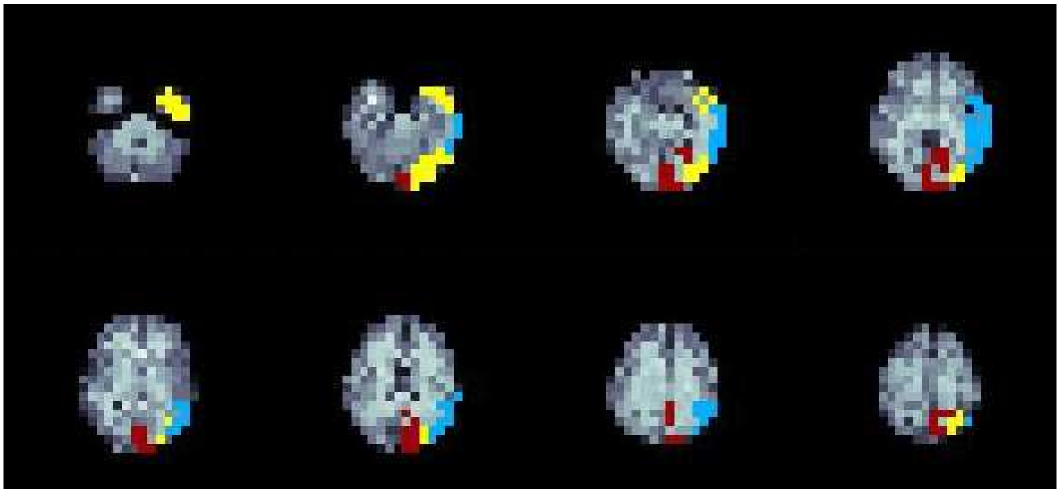


Figure 7.8: Connected clusters mapped into anatomical space. Clusters were defined according to the coloured horizontal bars in figure 7.7(B)

7.2.4 Discussion

There are evidently many open problems with this type of exploratory diffusion connectivity analysis. The most obvious of these is the development of an algorithm which detects and labels clusters in the reordered matrix in an unbiased fashion. In the analysis presented above, the three clusters detected were labelled by eye. Not only is this unsatisfactory due to its subjective nature and associated bias, but it also potentially ignores much of the available information. If the matrix were examined at different scales, it is likely that many other "clusters" might emerge. An unbiased measure for judging when a "cluster" is a cluster, and where it sits in the cluster hierarchy would dramatically increase not only the credibility of such studies, but also their power. A second crucial question to ask is what is the best data to search for clusters in. Again, the approach taken in the above analysis is clearly far from satisfactory. We were forced to take the log of the probability values derived in chapter 5 in order to reduce the effective range of the data to make it suitable for clustering. The reason for this is the very nature for the probability values themselves. As discussed in chapters 5 and 6, the probabilities in each voxel represent the probability of the *dominant* streamline from the seed voxel passing through this voxel. Not only does this mean that probability values away from the dominant streamline will be lower than if we had access to true "connectivity" data, but also they will depend on, for example, the voxel dimensions of the grid in which you store them (there is a greater probability that the dominant streamline will pass through a large voxel than a small one!). Access to data which more closely represents the true levels of "flow of information" between cortical areas would clearly be of benefit to this style of connectivity analysis.

7.3 Discontinuities in the "Connectivity Field"

Even in the absence of a complete representation of connectivity, diffusion tractography may still form a rich source of information. Network studies, such as the one presented in the previous section, will not prove as easy in areas of the brain where, for example, diffusion tractography is less reliable due to complex fibre architecture or areas where the most interesting information about the network does not lie in the dominant fibre pathways. A less ambitious goal might be, instead of trying to infer on the entire network by tracing the exact white matter routes, to attempt to cluster seed voxels which exhibit very similar connectivity patterns or, equivalently, to draw boundaries between adjacent brain areas with markedly different patterns. By simply searching for discontinuities in the "Field of Connectivity", we no longer rely on tracing the true fibre structure through the white matter. We only require that a sharp change in connectivity between adjacent brain regions will be reflected as a sharp change in the connectivity pattern generated by tractography.

7.3.1 Definition of the SMA-preSMA boundary by changes in connectivity.

Introduction [2]Medial area 6 consists of two cytoarchitectonically distinct regions in monkey(130) and two(131) or three(132) in human. These cytoarchitectonically distinct regions are not visible using conventional imaging techniques. However, in non-human primates there is a change in connectivity along medial frontal cortex: preSMA connects to prefrontal/anterior cingulate cortex whereas SMA proper connects to sensorimotor regions(131). If this change is visible in the data generated by probabilistic tractography then it should be possible to use it to identify the corresponding cytoarchitectonic boundary. A simple validation of this boundary may be performed by comparing its anterior-posterior position with that of the anterior commisure, a structure which is easily visible on a T1-weighted scan, and the current heuristic for the location of the SMA-preSMA boundary.

Methods Diffusion-weighted data and a T1-weighted image were acquired in 11 subjects. Each re-aligned T1-weighted image was used to define a medial frontal mask on a single axial slice (MNI 152 Z=56), which included both SMA and preSMA. Probabilistic tractography (as in chapters 4 and 5) was run from all voxels in this seed mask. Probabilities of connection from each seed voxel (at $1.2mm^3$ resolution) to every other voxel in the brain (re-sampled to $5mm^3$) were binarised and stored in a matrix, $\mathbf{A}$. The cross-correlation matrix of $\mathbf{A}$ was computed and stored in $\mathbf{B}$. (Note, the matrix $\mathbf{B}$ is of dimension (number of seeds × number of seeds), and element B_{ij} contains the correlation between the connectivity distribution from voxel i and the connectivity distribution from voxel j. This matrix contains information reflecting the similarity of the connectivity distributions of every pair of seed voxels). The nodes in $\mathbf{B}$ were permuted

---

[2] This work was carried out in collaboration with Heidi Johansen-Berg

according to the spectral NCut algorithm described in the previous section. Clusters around the diagonal of this reordered matrix represent clusters of similar connectivity in the brain. Discontinuities between clusters in the reordered matrix represent sharp changes in connectivity, and putative cytoarchitectonic boundaries between brain regions.

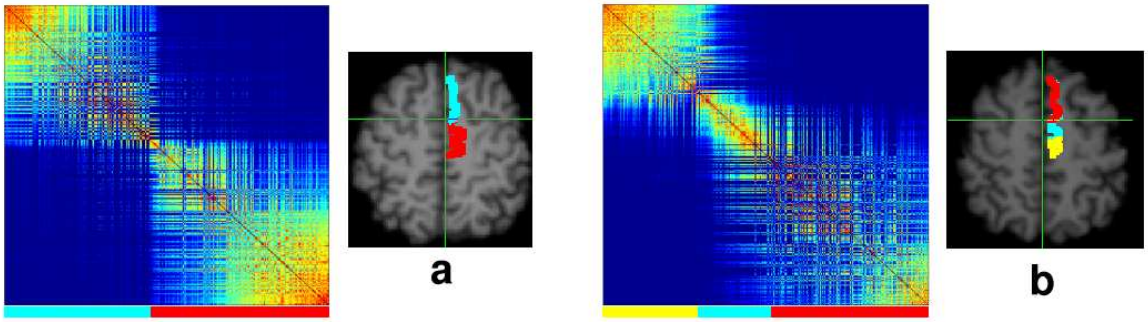


Figure 7.9: Reordering connectivity correlation matrices in SMA/preSMA. In 8/11 subjects two clusters were apparent in the reordered correlation matrices (a typical example is shown in (a)). In 3/11 subjects three clusters were apparent (a typical example is shown in (b)).

Results In eight out of eleven subjects the re-ordered connectivity matrix could be easily divided (by eye) into two clusters whereas in three out of eleven subjects there were three clusters apparent (typical examples are shown in figure 7.9). In all subjects, when clusters were mapped back onto the brain, they appeared as distinct regions along the anterior-posterior axis (figure 7.9). Clusters from the 8 subjects in whom two clusters were identified were overlaid to define a probabilistic atlas putative pre-SMA/SMA-proper (figure 7.10). The putative SMA/Pre-SMA boundary lay directly superior to the anterior commisure (figure 7.10).

Discussion A change in connectivity along medial frontal cortex was detected by identifying clusters in reordered connectivity correlation matrices. Mapping these clusters back onto the brain defined a boundary that we hypothesise corresponds to a functionally relevant cytoarchitectonic border. Future imaging studies will test how well this border co-localises with a functional boundary.

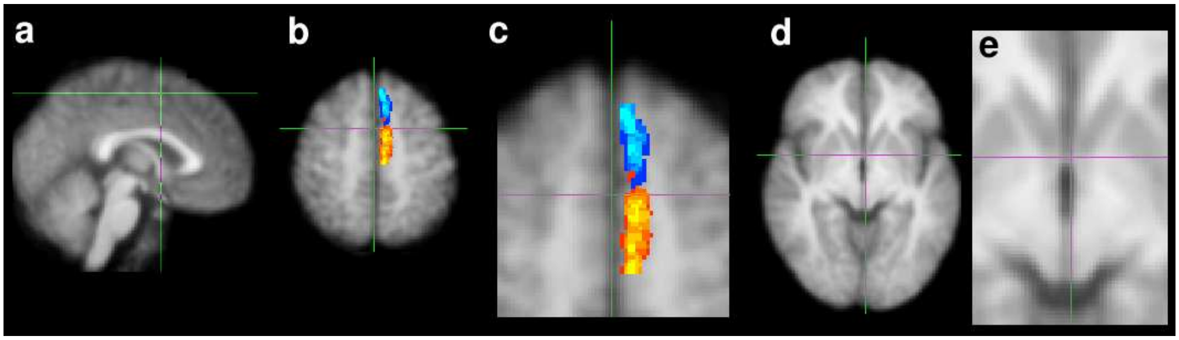


Figure 7.10: Probabilistic atlas of connectivity defined SMA/Pre-SMA. (a,d,e) Show the anterior-posterior position of the putative boundary. (a) in coronal section and (d,e) on the axial slice containing the anterior commisure (AC). The boundary lies directly superior to the AC. (b,c) show the probabilistic atlas. Pre-SMA is coded with dark-blue to light blue representing 1-8 subjects, and SMA is coded with red to yellow representing 1-8 subjects.

This approach is based solely on detecting changes in connectivity and is therefore able to define grey matter boundaries even when connectivity targets cannot be accurately specified or connectivity information is incomplete. This provides a novel means for parcellating cortical grey matter in vivo using information that is directly relevant to function.

Chapter 8

Conclusions

In this thesis we have developed a statistical framework in which to infer on anatomical white matter connectivity from diffusion weighted MR data, which provides a generalisation of previous "streamlining" techniques to the case when there is uncertainty in local fibre orientations. We have compared probabilistic predictions derived from Bayesian statistics with those derived from repeated acquisitions of the same data, and have shown that, by inferring in a probabilistic fashion, we improve the robustness and interpretability of diffusion based tractography to an extent that connectivity may be established between grey matter regions in the human brain.

We have used this "probabilistic diffusion tractography" to infer on cortico-thalamic circuitry in the human brain, providing a topographic map of the the thalamus coded by cortical projection/reception sites. The similarity of this map to predictions from invasive tracer studies in non-human primate provides compelling evidence in favour of the concept of diffusion tractography as a useful probe of anatomical connectivity. In fact, the co-localisation of the connectivity defined regions with the functional activations of the same nuclei, along with the correspondence between connectivity defined *volumes* and thalamic nuclear volumes measured post-mortem, provide the first attempt at quantitative validation of diffusion-based tractography.

In recent developments, we have applied the same methodology to infer on cortico-cerebellar circuitry, revealing new findings about the prefrontal projections to cerebellum. We have also suggested new methods for using tractography data to identify functionally related and functionally distinct grey matter regions.

There are still many open questions in the field of diffusion tractography; many of them relate to new and interesting potential applications and two of the most exciting are mentioned in sections 8.1.2 and 8.1.3. However, perhaps the most important question for the future of diffusion tractography relates to the diffusion measurement process itself; specifically, to the question of how we may interpret our diffusion-weighted measurements in terms of the underlying fibre architecture: How can we relate fibre structure to the diffusion we measure?

8.1 Some Open Questions

8.1.1 Relating Fibre Structure to Diffusion

Although the probabilistic framework is generic, the specific application of diffusion tractography throughout this thesis has been the inference on the location of the *dominant* streamline from a seed location. As was discussed in chapters 5 and 6, this dominant streamline, although certainly of interest, is far from the complete picture in terms of the anatomical connectivity of the seed location. The reason that we are limited to inferring on a single dominant streamline is that the model we originally fit to the diffusion signal in chapter 4 allowed for only a single fibre direction within a voxel. In fact, each voxel will contain a continuous distribution of fibre orientations (from here on referred to as the fibre orientation distribution function (fODF)), resulting from fibre divergence, convergence and crossing. A key area for future work in the field of diffusion tractography is to develop models or relationships which allow us to relate the measured diffusion weighted NMR signal to (at worst) this fODF or (at best) to the actual morphology of the fibre system in the voxel. Recent work on the model-free recovery of the diffusion function (16; 133) has brought this possibility one stage closer. Q-space, and Q-ball imaging allow us to recover, respectively, the (sampled) ensemble average diffusion propagator (the *pdf* on diffusion within a voxel (see chapter 2)) and the diffusion spin ODF (the orientational *pdf* on diffusion within a voxel) without the requiring us to make any assumptions about the forms of these functions. Crucially, this work has resulted in observations of areas in the brain where the recovered diffusion functions are far from what we would predict with a Gaussian model of diffusion. For example, areas in the brain where fibre pathways cross are revealed by diffusion functions with multiple modes, where each mode is along the direction of one of the crossing pathways. However, unfortunately these techniques do not yet allow us to recover the distributions which relate directly to the underlying fibre architecture (the fODF or the fibre morphology in the voxel).

The most obvious approach to achieving this goal yields a problem of such enormous complexity that it seems unlikely ever to provide us with an adequate solution. The number of biophysical unknowns which influence diffusion in and around white matter is so great that as yet, we cannot even state with any confidence which of them contribute to the diffusion anisotropy we rely on for tractography. It seems then that building a low-level biophysical model of white matter with a view to predicting the diffusion we witness in the NMR signal is a not a problem to be tackled in the near future.

All hope is not yet lost. It would certainly be possible to tackle the problem experimentally at a much higher level. Diffusion measurements from phantoms where we knew the ground truth fibre architecture could help to understand the measurements we take in the living brain. However, perhaps the most promising approach to tackling this essentially biophysical problem, is to ignore completely the biophysics. It may be possible to take an extremely simple model of the relationship between fibre architecture and diffusion signal (such as the one in section 4.1.3) but infer on it in such a way that we recover the same distributions as we would if we knew the true model. We can potentially achieve this by using sensible anatomical constraints on the distributions we estimate. For example, we know that every fibre that leaves one voxel enters the relevant neighbour. We know that white matter fibres connect regions of grey matter. It may be that such general anatomical "prior" information could provide us with enough constraints to infer reliably on the fibre distributions of interest without the need for a complete understanding of the biophysics.

8.1.2 Functionally-relevant brain atlases.

There has been a growing trend in recent years to build images which are in some way representative of an "average" human brain (e.g. (134)). Such representative brains serve an important function to the neuroscience community; they provide a common frame of reference in which neuroscientists can examine and compare their experimental results. Currently, the representative brain most widely used by neuroscientists is the MNI 152 brain, which is an arithmetic average of MRI scans of 152 brains at each imaging voxel. However, before experimental results may be examined in the light of the MNI brain and, indeed, before the 152 brains could be averaged originally, they must undergo the process of registration. That is, they must be transformed in some way such that the same coordinates in each image refer to the same anatomical location in the brain (for a detailed explanation of registration and atlases, see (21)). At present, this transformation process aims to match either visible features in the structural image (feature based registration) or image intensities at each voxel (intensity based registration). However, the brain location of real interest to the neuroscientist is not the physical location of the active neurons defined by neighbouring sulci and gyri, but their conceptual location in the axonal network which defines neuronal function. The relationship between these structural and functional locations is not wholly consistent across individuals (e.g. (135)); hence, the explanation of neuronal activation in the context of their visible anatomical location (termed "Brain Mapping"), while clearly useful, is far from presenting a complete picture.

In this thesis (chapters 6 and 7) we have introduced methods for defining and parcellating brain regions *according* to their location in the axonal network; that is, according to their connections. An elegant extension to this work would be to devise registration strategies to match images based on these parcellations such that we may build a functionally relevant brain atlas which is parcellated and labelled according to local connectivity. Not only would this provide a more logical frame of reference in which to examine experimental results, but it would also allow us, at least on a gross scale, to build a picture of the axonal network.

8.1.3 Combination with FMRI

Functional MRI (FMRI) provides a means for measuring the decrease in deoxyhaemoglobin content resulting from the increase in demand for oxygenated blood when neurons fire or "activate" (21). Neuroscientists use FMRI in conjunction with psychological experiments to build a picture of the areas of the brain involved with processing information in the task of interest. These "activation maps" provide us with information on the *functional localisation* of human brain regions. As discussed in chapters 6 and 7, we would expect to see a strong relationship between the *function* of a brain region and the *structure* of its anatomical wiring or connectivity. The function is constrained by the anatomy.

Figure 8.1: Combined FMRI-diffusion experiment. FMRI result (displayed in red to yellow): Group activation from 8 subjects for a working memory task. Thalamic and prefrontal cortical activations were apparent. Probabilistic tractography was seeded from the peak of the FMRI activation. The resulting connectivity distribution terminated in cortex at the location of the prefrontal activation.

With this in mind, tractography data can clearly be useful in informing the interpretation of results of FMRI experiments. For example, Figure 8.1 shows (in red to yellow8uiui) a projection through 7 slices of an activation map from a working memory experiment in 8 subjects. Activation clusters are shown in medial thalamus, insular, and lateral prefrontal cortex. In more superior slices there was also a parietal lobe activation. Probabilistic tractography was carried out as in chapter 5 seeded at the peak of the thalamic activation cluster. The resulting connectivity distribution terminated in prefrontal cortex at exactly the location of the prefrontal activation. No connections were found between the thalamic and parietal activating clusters. However seeding in the external capsule generated pathways connecting the frontal and parietal activations. This kind of result is clearly useful in helping us to understand the interaction between spatially discrete brain regions during the FMRI experiment.

However, this type of "visual-inspection" analysis, while potentially useful, could certainly be improved upon. A major goal of systems level neuroscience research is to highlight and describe the neuronal *networks* involved in processing the task at hand. The diffusion and BOLD datasets provide us with related but complementary information about these networks during the course of the FMRI experiment and, as such, are prime candidates for data fusion - the merging of more than one type of data derived from the same source, but in different fashions, to infer on that source.

In the Bayesian framework, the process of data fusion is a natural extension of inference with only a single type of data. Different generative likelihoods predict the effect of the source as seen in each type of data, and are combined with each other and the priors to generate the posterior distribution on the source. Here, in the BOLD data the underlying network is witnessed in the grey matter as correlations between the BOLD signal and the experimental design. In the diffusion data, it may be seen as fast diffusion along a white matter route between connected grey matter regions. Generative likelihoods for the entire network in the diffusion data may be built as a simple extension of the likelihood for a pathway (e.g. chapter 5) and a local fibre direction (e.g. chapter 4). In the BOLD data, a network likelihood should include parameters for the spatial positions of different activating units of the networks, and parameters for the temporal relationship between the BOLD signals in these units.

If analysed separately, but interpreted together, as in figure 8.1, BOLD- and diffusion-weighted data may be able to give us some conceptual insight into neuronal processes contributing to the FMRI data. If analysed together, such that each dataset may inform our inference about the other, then their combination may enable us to infer on structural/functional networks in the living human brain.

Bibliography

[1] M.E. Moseley, Y. Cohen, M.F. Kucharaczyj, J. Mintorovitch, H.S. Asgari, M.F. Wendland, J. Tsuruda, and D. Norman. Diffusion-weighted mr imaging of anisotropic water diffusion in cat central nervous system. *Radiology*, **176**, 439–445. (1990).

[2] P.J. Basser, J. Matiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. B*, **103**, 247–254. (1994).

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

[4] T. Conturo, N. Lori, and T. Cull. Tracking neuronal fiber pathways in the living human brain. *Proc Natl Acad Sci USA*, **96**, 10422–10427. (1999).

[5] P. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magn Reson. Med.*, **44**, 625–632. (2000).

[6] How Stuff Works. http://health.howstuffworks.com/brain1.htm.

[7] C. Poupon, C. Clark, V. Froulin, D LeBihan, I. Bloch, and J-F Mangin. Inferring the brain connectivity from MR diffusion tensor data. In *Annual Meeting of MICCAI*, pages 452–462, (1999).

[8] M. Catani, R. J. Howard, S. Pajevic, and D. K. Jones. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage*, **17**, 77–94. Sep 2002.

[9] P.J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.*, **44**, 625–632. (2000).

[10] N. Ramnani and R. C. Miall. A system in the human brain for predicting the actions of others. *Nat Neurosci*, **7**, 85–90. Jan 2004.

[11] J. A. Fiez, M. E. Raichle, D. A. Balota, P. Tallal, and S. E. Petersen. PET activation of posterior temporal regions during auditory word presentation and verb generation. *Cereb Cortex*, **6**, 1–10. Jan 1996.

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

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

[14] 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. *Philosoph. Mag.*, **4**, 161. (1828).

[15] A. Fick. On liquid diffusion. *Philos. Mag.*, **10**, 30–39. (1855).

[16] David Tuch. *Diffusion MRI of Complex Neural Architecture.* PhD thesis, Harvard-MIT Division of Health Sciences and Technology, (2002).

[17] P.T. Callaghan. *Principles of nuclear magnetic resonance microscopy.* Oxford University Press, (1993).

[18] E.L. Hahn. Spin echoes. *Phys. Rev.*, **80**, 580–594. (1950).

[19] H.C. Torrey. Bloch equations with diffusion terms. *Phys. Rev.*, **104**, 563–565. (1956).

[20] E.O. Stejskal and J.E Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.*, **42**, 288–292. (1965).

[21] P Jezzard, P.M. Matthews, and S.M. Smith. *Functional MRI. An introduction to methods.* Oxford University Press, (2001).

[22] E.O. Stejskal. Use of spin echoesd in a pulsed magnetic-field gradient to study anisotropic, restricted diffusion and flow. *J. Chem. Phys.*, **43**, 3597 3603. (1965).

[23] A. Einstein. Investigations on the theory of Brownian motion. In *R. Furthe and A.D. Cowper, editors.* Dover, New York, (1956). Collection of papers translated from German.

[24] R.L. Cooper, D.B. Chang, A.C. Young, C.J. Martin, and B. Ancker-Johnson. Restricted diffusion in biophysical systems. *Biophys. J.*, **14**, 161–1771. (1974).

[25] J.E. Tanner. Self diffusion of water in frog muscle. *Biophys. J.*, **28**, 107–116. (1979).

[26] P.C. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, **242**, 190–191. (1973).

[27] P. Mansfield and P. K. Grannel. NMR diffraction in solids? *J. Phys. Chem.:Solid State Phys*, **6**, L422. (1973).

[28] A. N. Garroway, P. K. Grannell, and Mansfield P. Image formation in NMR by a selective irradiative process. *J. Phys. Chem.:Solid State Phys*, **7**, L457. (1974).

[29] D.G. Taylor and M.C. Bushell. The spatial mapping of translational diffusion coefficients by the NMR imaging technique. *Phys. Med. Biol.*, **30**, 345–349. (1985).

[30] D. LeBihan, E. Breton, D. Lallemaud, P. Grenier, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion neurological disorders. *Radiology*, **161**, 401–407. (1986).

[31] M.E. Moseley, Y. Cohen, J. Mintorovitch, L. Chilenuittand, H. Shimizu, M.F. Kucharaczyj, M.F. Wendland, and P.R. Weinstein. Early detection of regional cerebral ichemia in cats: Comparisons of diffusion- and T2-weighted MRI and spectroscpoy. *Magn. Reson. Med*, **14**, 330–346. (1990).

[32] R. Bammer S.L. Keeling M. Augustin K.P. Pruessman C.Leussler F. Ebner F. Fazekas. Improved diffusion-weighted single-shot EPI in stroke using sense. In *Annual Meeting of the ISMRM*, page 161, (2001).

[33] D. K. Jones, M. A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn Reson Med*, **42**, 515–525. Sep 1999.

[34] Westin C-F. Maier S. Khidir B. Everett P. Jolesz F. Kikinis R. Image processing for diffusion tensor magnetic resonance imaging. In *Annual Meeting of Medical Image Computing and Computer Assisted Interveneion*, pages 441–452, (1999).

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

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

[37] Neil J.J. Shiran S.I. McKinstry R.C. Schefft G.L. Snyder A.Z. Almli C.R. Akbudak E. Aronovitz J.A. Miller J.P. Lee B.C. Conturo T.E. Normal brain in human newborns: apparent diffusion coefficient and diffusion anisotropy measured using diffusion tensor MR imaging. *Radiology*, **209**, 57–66. (1998).

[38] Christian. Beaulieu. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed*, **15**, 435–455. Nov 2002.

[39] A. C. Guo, J. R. MacFall, and J. M. Provenzale. Multiple sclerosis: diffusion tensor MR imaging for evaluation of normal-appearing white matter. *Radiology*, **222**, 729–736. Mar 2002.

[40] A. T. Toosy, D. J. Werring, R. W. Orrell, R. S. Howard, M. D. King, G. J. Barker, D. H. Miller, and A. J. Thompson. Diffusion tensor imaging detects corticospinal tract involvement at multiple levels in amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*, **74**, 1250–1257. Sep 2003.

[41] M.S. Buschbaum, C.Y. Tang, S. Peled, H. Gudbjartson, D. Lu, E.A. Haslett, J. Downhill, M. Haznedar, J.H. Fallon, and S.W. Atlas. MRI white matter diffusion anisotropy and PET metabolic rate in schizophrenia. *Neuroreport*, **16**, 425–430. (1998).

[42] J. Foong, M. Maier, C.A. Clark, G.J. Barker, D.H. Miller, and M.A. Ron. Neuropathalogical abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study. *J. Neurol. Neurosurg. Psychiatry*, **68**, 242–244. (2000).

[43] R.M. Steel, M.E. Bastin, S. McConnell, I. Marshall, D.G. Cunningham-Owens, S.M Lawrie, E.C. Johnstone, and J.J.K. Best. Diffusion tensor imaging (DTI) and proton magnetic resonance spectroscopy (^{1}H MRS) in schizophrenic subjects and controls. *Psych Res: Neuroimag*, **106**, 161–170. (2001).

[44] T. Klingberg, M. Hedehus, E. Temple, T. Salz, J. Gabrieli, M. Moseley, and Poldrack R. Microstructure of temporo-parietal white matter as a basis for reading ability: Evidence from diffusion tensor magnetic resonance imaging. *Neuron*, **25**, 493–500. (2000).

[45] Qian. Dong, Robert. C. Welsh, Thomas. L. Chenevert, Ruth. C. Carlos, Pia. Maly-Sundgren, Diana. M. Gomez-Hassan, and Suresh. K. Mukherji. Clinical applications of diffusion tensor imaging. *J Magn Reson Imaging*, **19**, 6–18. Jan 2004.

[46] Peter M Lee. *Bayesian Statistics: An Introduction*. Wiley-New York, (1997).

[47] D. Gamerman. *Markov Chain Monte Carlo.* Chapman and Hall, (1997).

[48] R. Choudrey. *Variational Methods for Bayesian Independent Component Analysis.* PhD thesis, University Of Oxford, (2002).

[49] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society,* **53**, 370–418. (1763).

[50] S.M. Stigler. Laplace's 1774 memoir on inverse probability. *Statistical Science,* **1(3)**, 359–378. (1986).

[51] J.M. Bernardo and A.F.M. Smith. *Bayesian Theory.* Wiley, (2000).

[52] D.J.C Mackay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, Cambridge, (1995).

[53] W.R. Gilks, S. Richardson, and G.J. Spiegelhalter. *Markov Chain Monte Carlo in Practice.* Chapman and Hall, (1996).

[54] M.W. Woolrich. *Model-based Approaches to FMRI Analysis.* PhD thesis, University of Oxford, (2001).

[55] Derek. K. Jones. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magn Reson Med,* **49**, 7–12. Jan 2003.

[56] Pierpaoli C. Jezzard P. Diffusion tensor imaging of the human brain. *Radiology,* **201**, 637–648. (1996).

[57] Basser P.J. New histological and physiological stains derived from diffusion tensor MR images. *Ann. NY Acad Sci,* **820**, 123–138. (1997).

[58] T. G. Reese, O. Heid, R. M. Weisskoff, and V. J. Wedeen. Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. *Magn Reson Med,* **49**, 177–182. Jan 2003.

[59] M. Jenkinson and S. Smith. A global optimisation method for robust affine registration of brain images. *Med Image Anal,* **5**, 143–156. Jun 2001.

[60] D. C. Alexander, G. J. Barker, and S. R. Arridge. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magn Reson Med*, **48**. (2002).

[61] Lawrence. R. Frank. Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magn Reson Med*, **47**, 1083–1099. Jun 2002.

[62] Pierpaoli C. Basser P.J. Toward a quantitive assessment of diffusion anisotropy. *Magn. Reson. Med.*, **36**, 893–906. (1996).

[63] E.J. Mufson, D.R. Brady, and J.H. Kordower. Tracing neuronal connections in postmortem human hippocampal complex with the carbocyanine dye dii. *Neurobiol. Aging*, **11**, 649–653. (1990).

[64] J.M. Van Buren and R.C. Burke. *Variations and connections of the human thalamus. 1. The nuclei and cerebral connections of the human thalamus.* Springer-Verlag, New York, (1972).

[65] A. R. Damasio and N. Geschwind. The neural basis of language. *Annu Rev Neurosci*, **7**, 127–147. (1984).

[66] S. Bookheimer. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci*, **25**, 151–188. (2002).

[67] M. Petrides and D. N. Pandya. Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur J Neurosci*, **16**, 291–310. Jul 2002.

[68] Mori S. KAuffman W. et al. In vivo visualisation of human neural pathways by magnetic resonance imaging. *Ann Neurol*, **47**, 412–414. (2000).

[69] B. Stieltjes, W. E. Kaufmann, P. C. van Zijl., K. Fredericksen, G. D. Pearlson, M. Solaiyappan, and S. Mori. Diffusion tensor imaging and axonal tracking in the human brainstem. *Neuroimage*, **14**, 723–735. Sep 2001.

[70] M. Koch. *Measurement of the Self-Diffusion Tensor of Water in the Human Brain.* PhD thesis, Max Planck Institute of Cognitive Neuroscience- Liepzig, (2001).

[71] J. Sethian. A fast marching level set method for monotonically advancing fronts. *Proc. Natl. Acad. Sci. USA*, **93**, 1591–1595. (1996).

[72] Parker G. Wheeler-Kingshott C. Barker G. Distributed anatomical brain connectivity derived from diffusion tensor imaging. Accepted for Proceedings of $XVIII^{th}$ International Conference on Information Processing in Medical Imaging, (2001).

[73] Brun, A. and Bjornemo, M. and Kikinis, R. and Westin, C-F. White matter tractography using sequential importance sampling. In *Annual Meeting of the ISMRM*, page 1131, (2002).

[74] Lazar, M. and Alexander, A. L. White matter tractography error analysis in a brain diffusion tensor field. In *Annual Meeting of the ISMRM*, page 1125, (2002).

[75] Parker, G.J.M. amd Barker,G.J. and Thacker, N. and Jackson, A. A framework for a streamline-based probabilistic index of connectivity (pico) using a structural interpretation of anisotropic diffusion. In *Annual Meeting of the ISMRM*, page 1165, (2002).

[76] R.E. Passingham, K.E. Stephan, and R. Kotter. The anatomical basis of functional localization in the cortex. *Nat. Rev. Neurosci.*, **3**, 606–616. (2002).

[77] H. Barbas and D.N. Pandya. Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *J. Comp. Neurol.*, **256**, 211–228. (1987).

[78] D.C. Van Essen, W.T. Newsome, J.H. Maunsell, and J.L. Bixby. The projections from striate cortex (v1) to areas v2 and v3 in the macaque monkey: asymmetries, areal boundaries, and patchy connections. *J. Comp. Neurol.*, **244**, 451–480. (1986).

[79] J.W. Scannell, G.A. Burns, C.C. Hilgetag, M.A. O'Neil, and M.P. Young. The connectional organization of the cortico-thalamic system of the cat. *Cereb. Cortex*, **9**, 277–299. (1999).

[80] E.G. Jones. *The Thalamus.* Plenum Press, New York, (1985).

[81] R.W. Guillery and S.M. Sherman. Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system. *Neuron*, **33**, 163–175. (2002).

[82] J.D. Speelman, R. Schuurman, R.M. De Bie, R.A. Esselink, and D.A. Bosch. Stereotactic neurosurgery for tremor. *Mov. Disord.*, **17 Suppl 3**, S84–S88. (2002).

[83] A. Morel, M. Magnin, and D. Jeanmonod. Multiarchitectonic and stereotactic atlas of the human thalamus. *J. Comp. Neurol*, **387**, 588–630. (1997).

[84] V.A. Magnotta, S. Gold, N.C. Andreasen, J.C. Ehrhardt, and W.T. Yuh. Visualization of subthalamic nuclei with cortex attenuated inversion recovery mr imaging. *Neuroimage*, **11**, 341–346. (2000).

[85] Basser P.J. Matiello J. Le Bihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J*, **66**, 259–267. (1994).

[86] G.J. Parker, C.A. Wheeler-Kingshott, and G.J. Barker. Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging. *Neuroimage*, **21**, 505–512. (2002).

[87] O. Ciccarelli, G.J. Parker, A.T. Toosy, C.A. Wheeler-Kingshott, G.J. Barker, P.A. Boulby, D.H. Miller, and A.J. Thompson. From diffusion tractography to quantitative white matter tract measures: a reproducibility study. *Neuroimage*, **18**, 348–359. (2003).

[88] C. Poupon, C.A. Clark, V. Frouin, J. Regis, I. Bloch, D. Le Bihan, and J. Mangin. Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *Neuroimage*, **12**, 184–195. (2000).

[89] D.K. Jones, A. Simmons, S.C. Williams, and M.A. Horsfield. Non-invasive assessment of axonal fiber connectivity in the human brain via diffusion tensor mri. *Neuroimage*, **42**, 37–41. (1999).

[90] T.E. Conturo, N.F. Lori, T.S. Cull, E. Akbudak, A.Z. Snyder, J.S. Shimony, R.C. McKinstry, H. Burton, and M.E. Raichle. Tracking neuronal fiber pathways in the living human brain. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 10422–10427. (1999).

[91] Jr. Tanaka, D. Thalamic projections of the dorsomedial prefrontal cortex in the rhesus monkey (macaca mulatta). *Brain Res.*, **110**, 21–38. (1976).

[92] T.J. Tobias. Afferents to prefrontal cortex from the thalamic mediodorsal nucleus in the rhesus monkey. *Brain Res.*, **83**, 191–212. (1975).

[93] H.J. Markowitsch, D. Emmans, E. Irle, M. Streicher, and B. Preilowski. Cortical and subcortical afferent connections of the primate's temporal pole: a study of rhesus monkeys, squirrel monkeys, and marmosets. *J. Comp. Neurol.*, **242**, 425–458. (1985).

[94] H. Yarita, M. Iino, T. Tanabe, S. Kogure, and S.F. Takagi. A transthalamic olfactory pathway to orbitofrontal cortex in the monkey. *J. Neurophysiol.*, **43**, 69–85. (1980).

[95] F.T. Russchen, D.G. Amaral, and J.L. Price. The afferent input to the magnocellular division of the mediodorsal thalamic nucleus in the monkey, macaca fascicularis. *J. Comp. Neurol.*, **256**, 175–210. (1987).

[96] E.G. Jones. Lack of collateral thalamocortical projections to fields of the first somatic sensory cortex in monkeys. *Brain*, **52**, 375–384. (1983).

[97] E.G. Jones, S.P. Wise, and J.D. Coulter. Differential thalamic relationships of sensory-motor and parietal cortical fields in monkeys. *J. Comp. Neurol.*, **183**, 833–881. (1979).

[98] D. Raczkowski and I.T. Diamond. Cortical connections of the pulvinar nucleus in galago. *J. Comp. Neurol.*, **193**, 1–40. (1980).

[99] M.J. Webster, J. Bachevalier, and L.G. Ungerleider. Subcortical connections of inferior temporal areas te and teo in macaque monkeys. *J. Comp. Neurol.*, **335**, 73–91. (1993).

[100] C. Asanuma, W.T. Thach, and E.G. Jones. Distribution of cerebellar terminations and their relation to other afferent terminations in the ventral lateral thalamic region of the monkey. *Brain Res*, **286**, 237–265. (1983).

[101] C. Asanuma, W.T. Thach, and E.G. Jones. Distribution of cerebellar terminations and their relation to other afferent terminations in the ventral lateral thalamic region of the monkey. *Brain Res.*, **286**, 237–265. (1983).

[102] MB Carpenter. *Neuroanatomy.* Williams and Wilkins, London, (1978).

[103] P.S. Goldman-Rakic and L.J. Porrino. The primate mediodorsal (md) nucleus and its projection to the frontal lobe. *J. Comp. Neurol.*, **242**, 535–560. (1985).

[104] J.P. Aggleton and M. Mishkin. Projections of the amygdala to the thalamus in the cynomolgus monkey. *J. Comp. Neurol.*, **222**, 56–68. (1984).

[105] A. Parent, A. Mackey, and L. De Bellefeuille. The subcortical afferents to caudate nucleus and putamen in primate: a fluorescence retrograde double labeling study. *Neuroscience*, **10**, 1137–1150. (1983).

[106] E. Gerardin, A. Sirigu, S. Lehericy, J. B. Poline, B. Gaymard, C. Marsault, Y. Agid, and D. Le Bihan. Partially overlapping neural networks for real and imagined hand movements. *Cereb Cortex*, **10**, 1093–1104. (2000).

[107] E. Gerardin, S. Lehericy, J. B. Pochon, M. S. du Tezenas, J. F. Mangin, Y. Agid, D. Le Bihan, and C. Mersault. Foot, hand, face and eye representation in the human striatum. *Cereb Cortex*, **13**, 162–169. (2003).

[108] B. Haslinger, P. Erhard, F. Weilke, A. O. Ceballos-Baumann, P. Bartenstein, V. Grafin, M. Schwaiger, B. Conrad, and H. Boecker. The role of lateral premotor-cerebellar-parietal circuits in motor sequence control: a parametric fmri study. *Brain Res Cogn Brain Res*, **13**, 159–168. (2002).

[109] D.J. McGonigle, A.M. Howseman, B.S. Athwal, K.J. Friston, R.S. Frackowiak, and A.P. Holmes. Variability in fmri: an examination of intersession differences. *NeuroImage*, **11**, 708–734. (2000).

[110] Mette. R. Wiegell, David. S. Tuch, Henrik. B. W. Larsson, and Van. J. Wedeen. Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *Neuroimage*, **19**, 391–401. Jun 2003.

[111] D.S. Tuch, T.G. Reese, M.R. Wiegell, N. Makris, J.W. Belliveau, and V.J. Wedeen. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magn. Reson. Med*, **48**, 577–582. (2002).

[112] K.D. Davis, C.L. Kwan, A.P. Crawley, and D.J. Mikulis. Functional mri study of thalamic and cortical activations evoked by cutaneous heat, cold, and tactile stimuli. *Neurophysiol*, **80**, 1533–1546. (1998).

[113] L. R. Becerra, H. C. Breiter, M. Stojanovic, S. Fishman, A. Edwards, A. R. Comite, R. G. Gonzalez, and D. Borsook. Human brain activation under controlled thermal stimulation and habituation to noxious heat: an fMRI study. *Magn Reson Med*, **41**, 1044–1057. May 1999.

[114] R.G. Pautler, A.C. Silva, and A.P. Koretsky. In vivo neuronal tract tracing using manganese-enhanced magnetic resonance imaging. *Magn. Reson. Med.*, **40**, 740–748. (1998).

[115] J. Kievit and H.G. Kuypers. Organization of the thalamo-cortical connexions to the frontal lobe in the rhesus monkey. *Exp. Brain. Res.*, **29**, 299–322. (1977).

[116] N.C. Andreasen, D.S. O'Leary, T. Cizadlo, S. Arndt, K. Rezai, L.L. Ponto, G.L. Watkins, and R.D. Hichwa. Schizophrenia and cognitive dysmetria: a positron-emission tomography study of dysfunctional prefrontal-thalamic-cerebellar circuitry. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 9985–9990. (1996).

[117] T. Klingberg, M. Hedehus, E. Temple, T. Salz, J.D. Gabrieli, M.E. Moseley, and R.A. Poldrack. Microstructure of temporo-parietal white matter as a basis

for reading ability: evidence from diffusion tensor magnetic resonance imaging. *Neuron*, **25**, 493–500. (2000).

[118] C.A.M Wheeler-Kingshott, P.A. Boulby, M. Symms, and G.J. Barker. Optimised cardiac gating for high-resolution whole brain DTI on a standard scanner. In *Annual Meeting of the ISMRM*, page 1118, (2002).

[119] A.W. Anderson. Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magn. Reson. Med.*, **46**, 1174–1188. (2001).

[120] S. M. Smith, Y. Zhang, M. Jenkinson, J. Chen, P. M. Matthews, A. Federico, and N. De Stefano. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage*, **17**, 479–489. Sep 2002.

[121] P. Brodal. The corticopontine projection in the rhesus monkey. Origin and principles of organization. *Brain*, **101**, 251–283. Jun 1978.

[122] P. Brodal and J. G. Bjaalie. Salient anatomic features of the cortico-ponto-cerebellar pathway. *Prog Brain Res*, **114**, 227–249. (1997).

[123] P.M. Levin. The efferent fibres of the frontal lobe of the monkey, macaca mulatta. *J Comp Neurol*, **63**, 369–419. (1936).

[124] O. Nyby and J. Jansen. An experimental investigation of the corticopontine projection in Macaca mulatta. *Skr. Norske Vidensk.-akad., I. Mat.-Nat. Kl.*, **3**, 1–47. (1952).

[125] M. Glickstein, J. G. 3rd. May, and B. E. Mercier. Corticopontine projection in the macaque: the distribution of labelled cortical cells after large injections of horseradish peroxidase in the pontine nuclei. *J Comp Neurol*, **235**, 343–359. May 1985.

[126] S. Roberts and R Everson. *ICA: Principles and practice.* Cambridge University Press, (2001).

[127] S. T. Barnard, A. Pothen, and H. D. Simon. A spectral algorithm for envelope reduction of sparse matrices. *Numerical Linear Algebra with Applications*, **2**, 317–334. (1995).

[128] D. J. Higham. Unravelling small world networks. *J. Comp. and Applied Math.*, **158**, 61–74. (2003).

[129] L. Ungerleider and M. Mishkin. Two streams of visual processing. *Analysis of Visual Behaviour*, pages 549–586. (1982).

[130] O. Vogt and C. Vogt. Ergebnisse unserer Hirnforshung (Results of our Neuroscience). *J. Psychol Neurol*, **25**, 277–462. (1919).

[131] K. Zilles, G. Schlaug, S. Geyer, G. Luppino, M. Matelli, M. Qu, A. Schleicher, and T. Schormann. Anatomy and transmitter receptors of the supplementary motor areas in the human and nonhuman primate brain. *Adv Neurol*, **70**, 29–43. (1996).

[132] V. Vorobiev, P. Govoni, G. Rizzolatti, M. Matelli, and G. Luppino. Parcellation of human mesial area 6: cytoarchitectonic evidence for three separate areas. *Eur J Neurosci*, **10**, 2199–2203. Jun 1998.

[133] D.S. Tuch, T.G. Reese, M.R. Wiegell, and V.J. Wedeen. Diffusion MRI of complex neural architecture. *Neuron*, **in press**. (2003).

[134] A.C. Evans, S. Marrett, P. Neelin, and MORE? Anatomical mapping of functional activation in stereotactic coordinate space. *Neuroimage*, **1**, 43–53. (1992).

[135] K. Amunts, A. Schleicher, U. Burgel, H. Mohlberg, H. B. Uylings, and K. Zilles. Broca's region revisited: cytoarchitecture and intersubject variability. *J Comp Neurol*, **412**, 319–341. Sep 1999.