2808915786

REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2006 Name of Author COOK, Philip Anthony

COPYRIGHT
This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION
I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOANS
Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION
University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).

B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

☑ This copy has been deposited in the Library of ______________________

☐ This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.

C:\Documents and Settings\lproctor\Local Settings\Temporary Internet Files\OLK8\Copyright - thesis (2).doc

Modelling uncertainty in brain fibre orientation from diffusion-weighted magnetic resonance imaging

Philip A Cook

Supervisors: Dr Daniel Alexander, Prof Andrew Todd-Pokropek

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of the
University of London.

Department of Computer Science
University College London

June 27, 2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Diffusion-weighted magnetic resonance imaging (DW-MRI) permits *in-vivo* measurements of water diffusion, from which we can infer the orientation of white matter fibres in the brain. We show that by ordering the measurements, we can improve the reproducibility of the fibre-orientation estimate from partially-completed DW-MRI scans, without altering the complete data set.

Tractography methods reconstruct entire fibre pathways from the local fibre-orientation estimates. Because the local fibre-orientation measurements are subject to uncertainty, the reconstructed fibre pathways are best described with a probabilistic algorithm. One way to estimate the connection probabilities is by defining a probability density function (PDF) in each voxel, and sampling from the PDF in a Monte-Carlo fashion.

We propose new models of the PDF based on standard spherical statistical methods. The models improve previous work by closely modelling the dispersion of repeated noisy estimates of the fibre orientation. We compare a simple PDF (the Watson PDF) that models circular cluster of axes to a more general PDF (the Bingham PDF) that models circular or elliptical clusters of axes. We also propose models of the PDF in regions of crossing fibres, where there are two distinct fibre populations in the voxel. We validate the PDFs by comparing them to the uncertainty in fibre orientation calculated from bootstrap resampling of a repeated brain MR acquisition. We find that the Bingham PDF produces connection probabilities that are closer to the bootstrap results than the Watson PDF.

We use the new PDF models to perform a connectivity-based segmentation of the corpus callosum in eight different subjects. The results are similar to those of previous studies on corpus callosum connectivity, despite the use of finer cortical labelling, suggesting that the dominant connections from the corpus callosum project to the superior frontal gyrus, the superior parietal gyrus and the occipital gyrus.

Contents

Acknowledgements

The first and foremost thanks must go to my supervisor Dr Danny Alexander. Before he suggested this project to me, I had never considered myself PhD material, but through all the difficulties Danny has always remained patient and helped me to continue. His guidance, encouragement and friendship over the last four years have been invaluable to the completion of this thesis.

I would also like to thank my secondary supervisor, Prof Andrew Todd-Pokropek, for his insights and guidance as I passed through the milestones on the way to the completion of this thesis. I would also like to thank Prof Simon Arridge for being my internal assessor for my preparatory reports and vivas.

Sincere thanks to Prof Sir Mike Brady and Dr Derek Jones for examining this thesis.

I am grateful to the EPSRC MIAS-IRC for funding my studies and also for providing funds for me to attend conferences in the UK and abroad. The UCL graduate school, the IEEE, and the ISMRM also generously provided funds for conference travel.

Thanks to Prof James Gee for his role in arranging the collaboration between myself and Hui Zhang that resulted in the work presented in Ch. 9. Thanks also to Hui Zhang, who visited England to work with me on that project and has been a good friend and source of interesting technical discussions before and since. Thanks to Dr Olga Ciccarelli and Dr Claudia Wheeler-Kingshott for providing the brain data used in that chapter. The brain data used in Ch. 7 and Ch. 8 was provided by Dr Geoff Parker, whom I thank for the provision of the data and also for valuable collaboration on papers and general interesting advice and discussions at conferences and in the pub. Thanks also to Dr Mark Symms and Dr Philip Boulby for their collaboration on the work in Ch. 3.

On a personal level, I would like to thank all my friends here at UCL, both past and present, who have made my time here so enjoyable. Special thanks to Jan Sikora for his unique perspective on life, which has to be experienced to be appreciated, and for being there over the many weekends I had to spend in the office to keep myself on track. I would also like to thank Kathleen Curran for her friendship and for motivating me to learn my subject more thoroughly

than I might otherwise have done. I am also grateful to Jason Riley for insisting that we visit pubs with real English beer, for the many interesting (and sometimes heated) discussions of politics, science fiction, and other things that we computer nerds care about.

Deepest thanks to Genaína Rodrigues for her encouragement and support, her friendship and understanding and for making me take time out to see the world. *Muito obrigado por tudo.*

To my family, I cannot express my gratitude for the love, support of all forms, and inspiration. Finally, I would like to thank Tessa Sundaram, who has helped me complete this thesis in just about every way imaginable, from helping find the right Matlab function to keeping me company through long nights of coding and writing.

Statement of intellectual contribution

The work presented in this thesis is my own, except for the connectivity-based segmentation work in chapter 9, which was done in collaboration with Hui Zhang from the Department of Radiology, University of Pennsylvania, who visited the UCL computer science department in January of 2005. Hui designed and implemented the automated labelling algorithm. I designed the probabilistic tractography algorithm. The work on the "seed partitioning" algorithms in Sec. 9.1 was shared equally between the two of us.

Although I have designed and carried out the experiments, and written this thesis myself, its form and substance has been guided and greatly improved by my collaboration with my supervisor, Dr Daniel Alexander. Dr Alexander also contributed a large fraction of the software in the Camino diffusion-MRI toolkit [1], which I have used extensively in my experiments, mostly to synthesise data in Ch. 3, 7 and 8. The tractography portion of the Camino code, including the implementation of the spherical statistical methods in Ch. 6 was written by me, except for the methods for estimating Bingham concentration parameters, which were provided by Dr John Kent, University of Leeds. Dr Geoff Parker, University of Manchester, also deserves a special mention for his assistance with understanding the PICo algorithm and later collaboration. My motivation to apply spherical statistical models to probabilistic tractography came about as the result of meetings between myself, Dr Alexander, and Dr Parker in 2003. This collaboration led to the publication of my first PICo-related paper at the 2004 IEEE Symposium on Biomedical Imaging [2], and continues in the ongoing Camino project [1].

The work on ordering the diffusion-weighted MR measurements in Ch. 4 was motivated by a conversation between myself, Dr Alexander, and Dr Mark Symms, Institute of Neurology, UCL. The design of the algorithm to search for an optimal ordering, and the experiments to evaluate the ordered point sets, were designed and implemented by me, but both Dr Alexander and Dr Symms provided helpful feedback and advice on my work. Dr Symms and Dr Philip Boulby (Institute of Neurology, UCL) acquired the brain data used in the experiments in Ch. 4, and they appear as co-authors on the abstracts I presented at the 13th and 14th Scientific Meetings of the ISMRM [3, 4].

Chapter 1

Introduction

Diffusion-weighted magnetic resonance imaging (DW-MRI) is a rapidly evolving field with exciting potential for understanding the anatomical connectivity of the brain. DW-MRI is sensitive to the microscopic random diffusive motion of water molecules. The random thermal motion of these molecules is hindered by the cellular structures in the brain. In white matter, the fibrous nerve tissue that connect neurons together, the water molecules are less hindered when they move along the fibres than when they move across them, because molecules moving along fibres encounter fewer cellular structures. This produces an *apparent* anisotropy in the diffusion – the diffusion process is as random in white matter as it is in free water, but the increased freedom of water molecules to move along the white matter fibres gives the appearance that the diffusion coefficient is larger along the fibres than across them.

The axon fibres that constitute white matter typically measure approximately 10^{-6} m in diameter, though some are as small as 10^{-7} m or as large as 2×10^{-5} m [5]. A typical DW-MR acquisition measures the diffusion over approximately 5×10^{-2} s, during which time the root mean-squared displacement of free water molecules (at normal body temperature of 37° C) is approximately 10^{-5} m. The axons hinder the diffusion of water molecules over the course of a DW-MR acquisition because the RMS displacement is larger than the average width of the axons.

DW-MR images are typically acquired on a grid with spacing of approximately 2×10^{-3} m, which means that a white matter voxel contains on the order of 10^5 fibres [6]. The "fibre-orientation" estimates derived from DW-MRI rely on the axons being highly organised into bundles with a common orientation at the scale of the voxel. If the axons were randomly oriented within the voxel, there would be no net diffusion anisotropy and no way of tracing the axonal paths with MRI. Histological studies suggest that white matter fibres are often highly organised into bundles with a common orientation. Fig. 1.1 shows a section from a dissected human brain, in which macroscopic bundles of white matter fibres are clearly visible. The apparent organisation of the fibre bundles is confirmed by the observation of anisotropic diffusion in both *ex-vivo* [7] and *in-vivo* [6] axonal fibres.

Figure 1.1: Dissected human brain showing microscopic white matter fibres organized into visible bundles. Image from Poupon et al [8].

The local measurements of the diffusion in each voxel are processed by “tractography” methods that attempt to reconstruct entire white matter tracts. One widely-used technique is to compute “streamlines” by following a path through the vector field of the estimated fibre orientation in each voxel [9]. Fig. 1.2 shows a schematic representation of streamlines tracked from a start point in the splenium of the corpus callosum in a human subject. Streamline tractography has been widely applied to the human brain, where several authors have tracked major fibre pathways such as the corpus callosum, cingulum, and fornix, producing results that are consistent with known anatomy [10]. Medical applications for tractography include pre-surgical planning for the surgical removal of brain tumours, where tractography can be used to visualise the path of fibre tracts close to or affected by the growth of tumour tissue [11].

A major limitation of the streamline approach to tractography is that it does not account for uncertainty in the local fibre-orientation estimates. One source of error is the random noise on the DW-MR measurements. Noise introduces an error into each fibre-orientation estimate, which in turn introduces an error into the streamline trajectory. The uncertainty in the fibre-orientation estimates depends on the scanner protocol, the mathematical model of the diffusion (most commonly the Gaussian model, as described in Ch. 3), and the microstructure of the tissue in each voxel. The cumulative effect of these errors on the streamline trajectory is difficult to

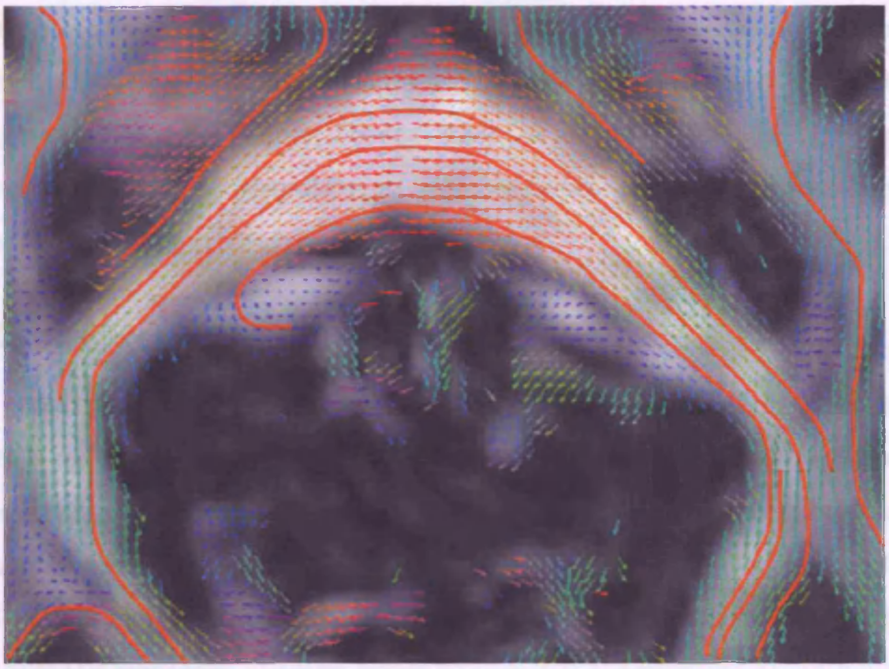


Figure 1.2: Slice from DW-MRI image of a human brain. Estimated local fibre-orientation estimate (arrows) are coloured according to orientation: red is left-right, green is anterior-posterior, blue is inferior-superior. The red lines are streamlines traced through the vector field of local fibre-orientation estimates. The image is taken from Bammer et al [10].

predict and depends on the architecture of the fibre bundle that the streamline follows.

The need to model and visualise uncertainty in tractography motivated a probabilistic approach to fibre tracking, where the uncertainty in the local fibre orientation is included in the tractography process. Monte-Carlo methods [12, 13] use a probability density function (PDF) of the local fibre orientation in each voxel. A probabilistic streamline is the same as a standard streamline, except that when a probabilistic streamline passes through a voxel, the fibre orientation is sampled randomly from the corresponding PDF (see Sec. 5.3). Many probabilistic streamlines are tracked from the same start point. The probability of connection from a starting point A to a particular position B in the brain, given the PDF in each voxel, is the probability of a probabilistic streamline started at A reaching B.

Tractography has potential for the study of neuroscience, since it allows *in-vivo* study of the anatomical connectivity in parallel with functional imaging. This provides the opportunity to study how the functional connectivity in the brain corresponds to the anatomical connections that are suggested by tractography. Johansen-Berg et al [14] find that the functional activation in the human thalamus during a motor task corresponds to the parts of the thalamus where tractography predicts a connection to the motor cortex. They reach the same conclusion with executive tasks, which activate the the prefrontal cortex and the areas of the thalamus where tractography predicts a connection to the prefrontal cortex. They use the probabilistic tractography method of Behrens et al [13] to calculate the most probable cortical connection to each point in the thalamus.

1.1 Thesis overview

Ch. 2 is a brief introduction to the principles of magnetic resonance imaging. Ch. 3 explains how MRI is applied to the study of diffusion, and introduces the methods for inferring the local fibre orientation or orientations from the DW-MRI measurements. The models that we use in subsequent chapters are the Gaussian model [15], and the multi-compartment Gaussian model [16], since these models are computationally simple, and practical to use with data from standard DW-MR sequences.

The first original contribution of the thesis is presented in Ch. 4. We present a method to optimise the MR acquisition in order to obtain the best estimate of the fibre orientations from partially completed scans. Because we measure the diffusion in anisotropic tissue at many unknown orientations, we make multiple diffusion-weighted measurements with the diffusion-weighting gradient oriented differently at each measurement. These gradient orientations are distributed as evenly as possible on the unit sphere, so that the fibre-orientation estimates have minimal dependence on the orientation of white matter fibre bundles relative to the gradient directions. A set of measurements along isotropically distributed gradient directions are typically acquired in a random order, so a subset of these directions may be significantly anisotropic. We present a method to order the gradient directions such that any subset containing the first P out of the total N directions is as isotropic as possible. We show in synthetic data (using the Gaussian model) that data from P ordered measurements gives better reproducibility of the fibre orientation and the anisotropy, than data from P unordered measurements.

Ch. 5 reviews tractography algorithms, including the Probabilistic Index of Connectivity (PICo) [12], which defines a PDF of the fibre orientation in each voxel. The following chapters evaluate new models of the fibre-orientation PDF within the PICo framework.

Ch. 6 introduces the spherical PDFs that we use as models of the fibre-orientation PDF. The Watson model assumes cylindrical symmetry about the peak of the PDF. The Bingham and Angular Central Gaussian (ACG) models are more general, and do not assume cylindrical symmetry. Ch. 7 explains how we estimate the parameters of the PDFs from statistics of the diffusion in each voxel, using the Gaussian model of diffusion. We then evaluate how well the

PDFs model the uncertainty in fibre orientation caused by image noise in synthetic data. We conclude that the Watson and Bingham PDFs are significantly better models of the uncertainty due to noise than the ACG PDF.

The second main contribution of Ch. 7 is the evaluation of the PICo PDF models in brain data. We estimate the uncertainty in the estimated fibre orientations in the brain data by bootstrap resampling [17] of eight repeated acquisitions of the DW-MRI data. The bootstrap method captures the total uncertainty in the data, not just that caused by image noise. We compare the distribution of fibre-orientation estimates predicted by the Watson, Bingham and ACG PDF models to the distributions we retrieve from the bootstrap resampling. There is some correlation between the PDF and the bootstrap uncertainty; however, the bootstrap results suggest the uncertainty is lower than that predicted by the models.

The last experiments in Ch. 7 compare PICo connection probabilities in the brain image using model-based PDFs and bootstrapping. We track the path of the fibre bundles of the corpus callosum, the structure that connects the left and right hemispheres of the brain. We initiated the tracking from each voxel in the corpus callosum along the midline between the two hemispheres. The corpus callosum contains highly anisotropic fibres; at the midline it is possible for a non-expert in neuroanatomy to identify it in a DW-MR image. This makes it a natural choice of structure to study for this thesis. We find that PICo tractography using the Bingham PDF produces connection probabilities that are most similar to those derived from bootstrap tracking.

In Ch. 8 we extend the Watson and Bingham PDFs to model two distinct fibre orientations within a voxel. Given the rather coarse resolution of DW-MR (approximately 8 mm^3 voxels) and the complex fibre architecture in the brain, many voxels contain more than one population of fibres. When we resolve two fibre orientations within a single voxel, the orientations are more sensitive to errors and the PDF is more difficult to fit to the distribution of axes. We propose a novel method to find the PDF parameters in two-fibre voxels, and compare its performance in synthetic data to a previously published method proposed by Alexander and Barker [18]. We find that Alexander and Barker's method gives more reliable estimates of the PDF parameters.

The next contribution of Ch. 8 is that we extend the calibration process that relates the parameters of the diffusion (as described by the two-compartment Gaussian model) to the parameters of the PDF. We then compare PICo results with Watson and Bingham PDFs in synthetic data. We conclude that the Bingham PDF is a better model of the uncertainty due to noise at intermediate anisotropy, where the contours of the fibre-orientation PDF are elliptical.

In the remainder of Ch. 8, we compare the PDFs to the the bootstrap results, in a similar manner to Ch. 7.

In Ch. 9, we use the one-fibre and two-fibre models of the PDF to perform PICo tracking in brain images of eight subjects. We use the PICo algorithm to segment the corpus callosum according to its connectivity to cortical grey matter, and compare the segmentation to the results of a previous study that used streamline tractography [19]. The results agree with the previous study: the majority of the corpus callosum connects to three large cortical regions. We find that attempting to resolve complex fibre architecture with the two-compartment Gaussian model does not alter the segmentation.

Chapter 2

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) [20] measures behaviour of ensembles of hydrogen nuclei (protons) in the presence of an external magnetic field, after they are deliberately disturbed from their equilibrium state. Protons in complex molecules, such as lipids, return to equilibrium at different rates to protons in much smaller water molecules, which means that the MRI signal is dependent both on the density of protons in the tissue, and on the molecular structures in which the protons are present. This gives MRI excellent soft-tissue contrast. Diffusion-weighted MRI (DW-MRI) measures the diffusion of water, which arises from the microscopic thermal motion of water molecules. Cellular structures form barriers that hinder the movement of water molecules, which makes the properties of diffusion sensitive to the local tissue microstructure.

2.1 Protons and magnetic resonance

Protons possess an intrinsic quantised angular momentum known as "spin". Spin is a quantum mechanical property so named because a classical analogy to the proton in a magnetic field is a spinning top in a gravitational field. The angular momentum of the top keeps it upright. If the top is perturbed so that its axis of rotation is no longer vertical, then gravity exerts a torque upon the top, and the axis of rotation precesses around the vertical axis. The speed of precession increases in a stronger gravitational field, or with slower rotation of the top.

The spinning proton behaves in a similar way when placed in a magnetic field. The proton aligns itself to one of two positions, which we shall call "with" or "against" the magnetic field. These positions are not exactly aligned along the axis of the field, so the proton precesses about the axis of the field. The frequency of precession is known as the Larmor frequency $L = \gamma B_0$, where B_0 is the strength of the magnetic field, and γ is a constant called the gyromagnetic ratio. Protons aligned against the field have a slightly higher energy than those aligned with it [21, Ch. 5].

The hydrogen atom is of particular interest to MRI because it is abundant in the body, and its proton is unpaired. According to the shell model of the nucleus, nuclear protons have lowest energy if they are paired with spins in the opposite direction [22, Ch. 4]. The net magnetization of a collection of paired protons in a magnetic field is therefore zero. Unpaired protons, however, are free to align with or against the field. The protons aligned with the field have slightly lower energy, which means that slightly more protons are in this state at equilibrium, resulting in a small net magnetisation M_z, which itself precesses about the B_0 field. Some other atoms (such as nitrogen) also have an unpaired proton, however these atoms are far less abundant in the body than hydrogen, and have lower gyromagnetic ratios, so they do not contribute to the conventional MRI signal [23, Ch. 3].

2.2 Measurement of magnetic resonance

The net magnetisation of hydrogen atoms is the basis of the MRI signal. The ratio of aligned protons N_- to anti-aligned protons N_+ is given by a Boltzmann distribution: $N_-/N_+ = \exp(\gamma\ B_0/kT)$, where k is Boltzmann's constant and T is the temperature of the substance being imaged. Since the human body temperature is approximately constant, we can only increase the excess of aligned protons in in-vivo imaging by increasing B_0. At clinical field strengths of 1.5 T, the thermal energy of the molecules at body temperature is much greater than the difference in energy between spins aligned with or against the field. Spins therefore constantly change between alignments, but a large ensemble of spins achieve an equilibrium such that a small excess (approximately 5 per million spins) are aligned with the field. This small excess is sufficient to produce a measurable net magnetisation because of the abundance of hydrogen atoms in the body, of which many form part of water molecules (pure water contains 5×10^{23} hydrogen atoms per gram [21]).

We define the "magnet frame" as a coordinate system where the z-axis is aligned with $\mathbf{B}_0$. The net magnetisation vector of a collection of protons is aligned with the positive z-axis in this frame. We measure magnetic resonance by disturbing the equilibrium of the protons and observing how they return to the equilibrium state. We also define a frame of reference with coordinate axes (x', y', z'), where z' aligned with z, and the x' and y' axes rotate about z at the Larmor frequency, L. If we disturb the net magnetisation such that it is no longer aligned with $\mathbf{B}_0$, then it will precess in the magnet frame with frequency L. If we apply a pulsed magnetic gradient field $\mathbf{B}_1$ along the x' axis, the magnetisation vector is rotated about the axis of $\mathbf{B}_1$ [23, Ch. 3]. The extent of this rotation is determined by the amplitude and duration of the pulse. The frequency L is therefore the resonance frequency, because the net magnetisation can be manipulated with B_1 fields much weaker than B_0.

The field $\mathbf{B}_1$ is applied for a short time, and is known as a 90° pulse, because it rotates the net magnetisation by 90° onto the y' axis. The pulse removes longitudinal magnetization along $\mathbf{B}_0$ and creates transverse magnetization along y'. This transverse magnetization is stationary in the rotating frame, but in the magnet frame it precesses about $\mathbf{B}_0$ at the resonance frequency γB_0, which produces the MRI signal by inducing a current in a radio-frequency receiver coil. After the pulse is switched off, the transverse magnetization decays and the longitudinal magnetization returns gradually to its original alignment with $\mathbf{B}_0$. This process is known as relaxation.

2.2.1 T1 relaxation

T1 relaxation is the recovery of net longitudinal magnetization after the 90° pulse. T1 relaxation primarily occurs through "spin-lattice" interactions. The protons in molecules such as water remain aligned with or against $\mathbf{B}_0$ at all times, however the molecules themselves are free to rotate and translate in space. When a molecule rotates in the x-y plane at the resonance frequency, its magnetic field oscillates at the resonance frequency, producing a molecular version of the $\mathbf{B}_1$ field produced by the 90° pulse. This field can produce changes in alignment of nearby protons. These interactions are the mechanism of T1 relaxation. The recovery of longitudinal magnetization M_z is exponential, and at time t after the 90° pulse, $M_z(t) = M_z(1 - exp[-t\mathrm{T1}])$. The T1 relaxation time is the time taken for the recovery of 65% of M_z.

The distribution of rotational frequencies in a sample is related to the thermal energy of the molecules, their size and their molecular bonds. Water molecules are small and relatively free to rotate and translate. In the body, most free water rotates at frequencies high above the resonance frequency. Water therefore has a long T1, because the rate of proton-proton transitions is low. Larger molecules, such as proteins, tend to rotate at slower frequencies, closer to the resonance frequency. Proteins in water solution can also form bonds with water molecules, further slowing their rotation. Protein solutions therefore have a shorter T1 than pure water. Lipid molecules are larger still, and their average rotation frequency is still closer to the resonance frequency of typical MR B_0 fields. Lipid molecules, such as those found in the myelin coating in brain white matter, therefore have a high rate of spin-lattice interaction and short T1 relaxation times.

2.2.2 T2 relaxation

Simultaneous with the recovery of longitudinal magnetization following a 90° pulse is the decay of transverse magnetization. This is known as T2 relaxation. Immediately after the pulse, the longitudinal magnetization is completely converted to transverse magnetization, $M_{xy}(0) = M_z$. Thereafter, $M_{xy}(t) = M_z \exp(-t/\mathrm{T2})$. For any given tissue, $T2 \leq T1$.

The molecular basis for T2 relaxation is de-phasing. When the pulse is applied, the protons are in phase, aligned with y', that is, they precess in the x-y plane at the resonance frequency. After the pulse, protons begin to dephase, because some protons experience slightly different magnetic fields (and therefore precess at a different speed). The proton-proton interactions described above cause some de-phasing, but the primary mechanism of T2 relaxation is variations in the local magnetic field [21, p.51]. Protons in an area where B_0 is slightly different precess at different rates, causing de-phasing. There are two sources of magnetic field inhomogeneity, the first is from flaws in the main magnet, which are constant. This source of de-phasing can be significantly reduced by spin-echo imaging, which is described below. The other source of local field variation is from the protons themselves. Protons that are relatively fixed in space, and rotate slowly compared to the resonance frequency, cause local differences in the B_0 field. In free water, where few molecules rotate slower than the resonance frequency, and are not fixed, there is little de-phasing due to field inhomogeneity, and T2 is long. Water molecules in protein solutions rotate more slowly and have shorter T2. Solid molecules rotate slowest of all, and have still shorter T2.

2.2.3 The spin-echo sequence

T2 images reflect the rate of de-phasing following a 90° pulse. The combined de-phasing due to T2 relaxation and the inhomogeneities in B_0 is called T2* relaxation. The spin-echo sequence [24] is designed to remove T2* effects by cancelling out the de-phasing due to field inhomogeneity. A 180° pulse is inserted between the 90° pulse and the measurement of the signal. Imagine two spins F and S, where F precesses faster than S, because of an inhomogeneity in B_0. Time starts at $t = 0$ when the 90° pulse is applied. At time $t = t_e/2$, F is ahead of S. The 180° pulse then rotates both spins by 180°. Assuming that both the inhomogeneity and the spins remain fixed, F continues to rotate faster than S, however because of the flip, F is now behind S. After a further time $t_e/2$, F and S are again in phase. If there is no T2 relaxation, the transverse magnetisation is the same at time $t = t_e$ as it was at $t = 0$. This is called the "spin-echo", and the time t_e is the echo time.

2.2.4 Image weighting

The nature of the MRI image depends strongly on the time between excitation pulses (typically a 90° pulse) and time between each pulse and the measurement of the signal. By choosing these times strategically it is possible to produce images that are "weighted" to show differences in either T1 or T2.

At time t_r after an excitation pulse, the fraction of recovered longitudinal magnetisation is $M_z(1 - \exp[-t_r/\mathrm{T1}])$. The total longitudinal magnetisation within a voxel is proportional to both N_-/N_+, the excess fraction of protons aligned with B_0, and the density of protons $N(H)$ within the voxel. Therefore the total recovered longitudinal magnetisation is proportional to $N(H)(N_-/N_+)\exp[-t_r/\mathrm{T1}]$.

The transverse magnetisation is maximum immediately after the excitation pulse, after a time t_e it decays by $\exp(-t_e/\mathrm{T2})$. The total MR signal is then:

$$N(H))(N_-/N_+)(1-\exp[-t_r/\mathrm{T1}])\exp(-t_e/\mathrm{T2})\,. \tag{2.1}$$

All MRI images contain some combination of T1, T2, and proton density contrast, however it is possible to weight the images such that a particular source of contrast is dominant. If t_r is long compared to T1, then all of the longitudinal magnetisation, M_z, is recovered, since $\exp[-t_r/\mathrm{T1}]$ tends to 1 when $t_r \gg \mathrm{T1}$. To suppress T1 contrast, t_r is made longer than the longest T1 in the tissue being imaged. If t_e is short compared to T2, then the transverse magnetisation remains, since $\exp(-t_e/\mathrm{T2})$ tends to 1 when $t_e \ll \mathrm{T2}$. To suppress T2 contrast, t_e is made shorter than the shortest T2 of the tissue being imaged.

A T1-weighted image is produced from a short t_r (to preserve T1 contrast) and a long t_e (to suppress T2 contrast). A T2-weighted image is produced from a long t_r (to suppress T1 contrast) and a long t_e (to develop T2 contrast). A proton-density-weighted image is produced from a long t_r and short t_e [21, p. 47].

2.3 Conclusions

We have given an overview of the physical processes that produce T1- and T2-weighted MR images. The behaviour of water molecules is of particular interest, since these are the subject of study in diffusion imaging. In a spin-echo sequence, the refocusing effect of the 180° pulse relies on the spins being stationary during the sequence, so that their Larmor frequency does not change between the start of the sequence and t_e. The random motion of diffusing water molecules leads to incomplete recovery of the transverse magnetisation at t_e. This relationship between the loss of transverse magnetisation and diffusion is exploited to measure diffusion with MRI, as we explain in the next chapter.

Chapter 3

Water diffusion and its measurement with MRI

Molecular diffusion arises from the random thermal motion of molecules that are in a liquid or gaseous state. Maxwell's kinetic theory of gases states that the mean kinetic energy of the molecules depends only on the temperature [25, p. 184]. Liquid water is abundant in the human body, and DW-MRI studies the statistical properties of the diffusion of water within the tissue microstructure. Water molecules in the body are sufficiently dense and energetic that they are constantly colliding. The collisions are elastic, so the total kinetic energy is the same before and after each collision. We cannot hope to predict or trace the path of single molecules, however we can make statistical inferences about the macroscopic behaviour of large collections of molecules over many collisions. If there are no barriers to the motion of the water molecules (apart from other water molecules), the probability density p of a molecule being at a particular displacement $\mathbf{r}$ from its start point after a random walk of duration t, is a zero-mean Gaussian distribution when t is much longer than the mean time between molecular collisions. Einstein [26] derived a formula for the variance of this distribution, the "diffusion coefficient", from the kinetic theory of gases.

In diffusion-MRI we estimate p separately in each image voxel, which typically occupies approximately $10^{-8}\,\mathrm{m}^3$. The diffusion time in MR imaging is approximately 50^{-3} s and the diffusion coefficient of free water at body temperature over this time is approximately $10^{-9}\,\mathrm{m}^2\,\mathrm{s}^{-1}$ [6], which means that the root mean squared displacement of water molecules over the diffusion time is approximately 3×10^{-5} m. Tissue structures that are much smaller than this therefore affect p in the voxel. In Sec. 3.1 we discuss the physiology of brain tissue and in Sec. 3.2 we discuss how this affects the observed p. We then discuss the different methods for estimating p in the brain.

3.1 Physiology of brain tissue

For historical reasons, cortical brain tissue is called "grey matter" and the connective nervous pathways are called "white matter" (the names denote the colour of the tissue in the post mortem brain) [27].

Grey matter is composed of neurons and their supporting cells. The neuronal cell bodies form only a small part of the surface area of each neuron. Fibrous dendritic spines [28] occupy almost half of the neuronal area. The dendrites reach from the neuronal nucleus and form synaptic connections with other dendrites and with axons. The axons differ from the dendrites in structure and function. Dendrites are small fibres that are part of the neuronal body and form local synaptic connections. Axons are longer, cylindrical fibres whose primary function is to transmit electrical signals between sites that do not communicate directly via dendritic synapses. Fig. 3.1 shows a diagram of a single neuron.

In the adult brain, all but the smallest axons are coated with a fatty protein called myelin, which insulates the axon and increases the efficiency of electrical communication. It is the myelin coating that gives axons their white appearance, and the "white matter" regions of the brain are composed of bundles of coherently organized axons and their supporting cells. Axons vary in diameter from less than 1×10^{-6}m to 25×10^{-6}m, with most of the fibres around 1×10^{-6}m [5]. Oligodendroglial cells form the myelin coating, each oligodendrocyte forms multiple layers of tightly wrapped myelin around the axon trunk. The total thickness of the myelin coat varies with axon diameter: the ratio of axonal diameter to total fibre diameter is almost constant at 0.6 in studies of various mammals and fibre diameters [5]. Beneath the myelin is the axonal membrane, which is 8×10^{-9}m thick and encases the axon itself. The internal space bounded by the axonal membrane is filled with axoplasm, a gel with viscosity approximately five times that of water, and small internal fibres that maintain the structure of the axon and transport nutrients along its length.

The brain is surrounded by cerebrospinal fluid, which is produced in the fluid ventricles within the cerebrum. This fluid has several functions, including the protection of the brain from impact and the transport of neural hormones.

3.2 Qualitative observations of diffusion in the human brain

The particle displacement function, p, depends on the interaction between the water molecules and the surrounding tissue structures. The behaviour of p is strongly dependent on the characteristics of the brain tissue.

The ventricles contain fluid with few barriers to the displacement of water molecules.

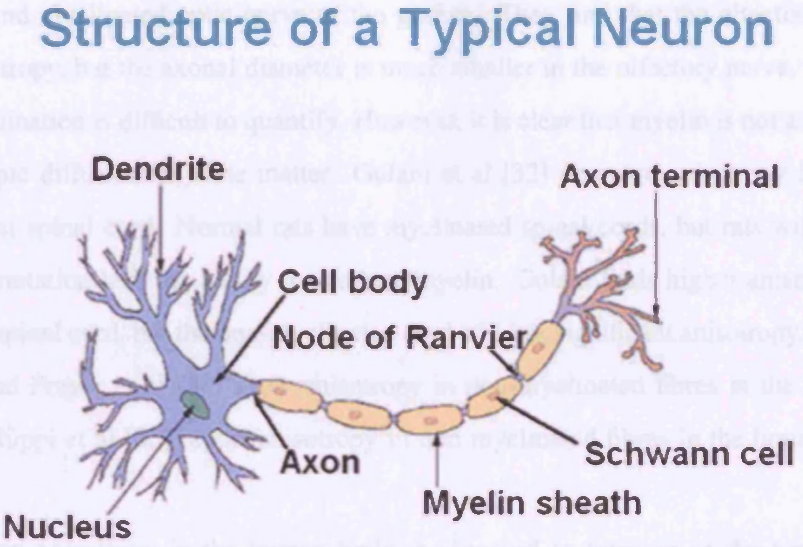


Figure 3.1: Diagram of a neuron with myelinated axon, from Young et al [29].

Pierpaoli et al [6], show that p in the cerebrospinal fluid is isotropic, with contours similar to those observed in free water.

Grey matter contains a mass of cellular structures that are dense enough to hinder the motion of water molecules significantly during DW-MRI acquisition. At the resolution of DW-MRI, the voxels contain large numbers of neurons oriented in different ways, so the cellular structures hinder diffusion equally in all directions. The root mean-squared displacement is lower than it is in cerebrospinal fluid, but p remains isotropic [6].

White matter is also densely populated with fibrous structures that hinder diffusion, resulting in a similar root mean-squared displacement to that observed in grey matter. On the scale of image voxels, the axons are highly organised, leading to an anisotropic p [6, 46]. The probability of meeting a cellular barrier is much lower for a particle that happens to move along the fibres than for a particle that moves across them, so p has a ridge along the longitudinal axis of the fibres. When more than one bundle of fibres passes through the voxel, p has ridges along the axes of the fibre bundles.

Structural organisation of cellular barriers in white matter is generally accepted to cause diffusion anisotropy, but there is no detailed explanation of how this anisotropy occurs at the cellular level. Many studies of animal and human nerve fibres have explored the contribution of different cellular features to diffusion anisotropy, these are reviewed in detail by Beaulieu [30].

The contribution of the myelin coating to anisotropy has proved difficult to quantify. Beaulieu and Allen [31] measure anisotropy in excised tissue from the non-myelinated olfactory nerve and myelinated optic nerve of the garfish. They find that the olfactory nerve has higher anisotropy, but the axonal diameter is much smaller in the olfactory nerve, so the influence of myelination is difficult to quantify. However, it is clear that myelin is not a requirement for anisotropic diffusion in white matter. Gulani et al [32] compare anisotropy in two types of excised rat spinal cord. Normal rats have myelinated spinal cords, but rats with a particular genetic mutation lack the ability to produce myelin. Gulani finds higher anisotropy in the myelinated spinal cord, but the non-myelinated cord still has significant anisotropy. Wimberger et al [33] and Prayer et al [34] show anisotropy in non-myelinated fibres in the neonatal rat brain, and Hüppi et al [35] report anisotropy in non-myelinated fibres in the brains of human neonates.

Diffusion anisotropy in the human brain is observed to increase as the brain develops and myelination increases [36]. However, the anisotropy may also increase because of greater organisation of the fibres. In the adult brain, the packing density of axonal fibres varies greatly in different fibre tracts. Pierpaoli et al [6] examine anisotropy in several tracts in the human brain where the packing density varies by an order of magnitude, and they did not find any correlation between packing density and anisotropy.

3.3 Diffusion-Weighted MRI (DW-MRI)

The DW-MRI sequence is a modified form of the spin-echo sequence. Spin-echo eliminates de-phasing caused by spatial variation in B_0 (T_2^* effects). In liquids, the re-phasing will not be complete, since diffusing spins will move during the time t_e, and hence the field they experience will not be precisely the same before and after the 180° refocusing pulse. This is a source of error in T2 imaging, but is the basis of diffusion imaging.

DW-MRI uses gradients to enhance de-phasing of protons that move during t_e. This sequence is called a "pulsed-gradient spin-echo" (PGSE) [37]. A PGSE sequence measures diffusion in one dimension. Fig. 3.2 is a schematic of the sequence. First, a 90° excitation pulse tips the spins into the transverse plane perpendicular to $\mathbf{B}_0$, where they precess in phase at the resonance frequency. Between the excitation pulse and the 180° refocusing pulse, the first diffusion-weighting gradient pulse Γ_1 is applied. During the gradient pulse, the magnetic field strength varies along the direction of the gradient, making the frequency of precession a function of position along the gradient and thereby causing spins to dephase. After the 180° refocusing pulse, the diffusion-weighting gradient Γ_2, which is identical to Γ_1, is applied. Spins that do not move, or that move perpendicular to the gradient, will experience the same field in both pulses, and thus re-phase at time t_e (labeled "TE" in Fig. 3.2), as in a standard spin-echo sequence.

Spins whose motion contains a component along the direction of the gradient during t_e will experience either a weaker or a stronger field during the second gradient and will not be in phase at t_e. Therefore the random motion of diffusing spins reduces the transverse magnetisation that we measure at t_e, and the extent of de-phasing increases with larger displacements along the gradient direction.

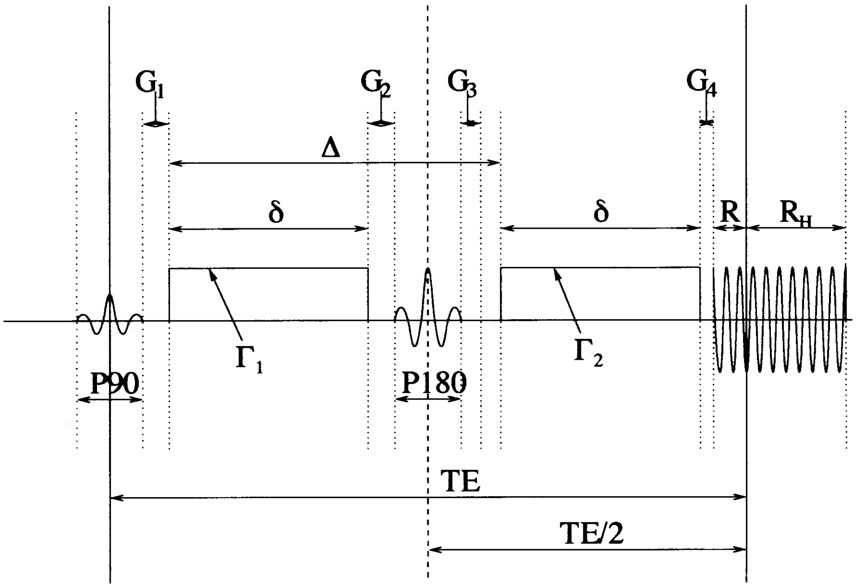


Figure 3.2: Diagram of PGSE acquisition from Alexander and Barker [38]. The gaps G_1 through G_4 reduce artefacts in the image.

The purpose of DW-MRI *in vivo* is to infer information about the tissue structure from the displacement density, p, which Stejskal and Tanner [37] show is related to the DW-MR measurement by a Fourier transform. For a measurement at the wavenumber $\mathbf{q}$, the MR measurement is

$$A^*(\mathbf{q}) = A^*(\mathbf{0}) \int p(\mathbf{r}) \exp(i\mathbf{q} \cdot \mathbf{r}) d\mathbf{r}, \tag{3.1}$$

where

$$\mathbf{q} = \gamma \delta \mathbf{g}, \tag{3.2}$$

$\mathbf{g}$ is the magnetic gradient vector, γ is the gyromagnetic ratio of the proton and δ is the duration of the gradient pulse. We use the normalised measurement $A(\mathbf{q}) = A^*(\mathbf{0})^{-1} A^*(\mathbf{q})$.

3.4 Estimating the spin displacement density

Early experiments with DW-MRI examined the "apparent diffusion coefficient" (ADC) in one dimension. The ADC is so called because the cellular structures in the imaged tissue hinder the random motion of water molecules and the resulting ADC is therefore smaller than the diffusion coefficient of free water. The ADC is calculated by assuming the relationship $A(\mathbf{q}) = \exp(-bd)$ where b (commonly called the"b-value") is $|\mathbf{q}|^2 t$.

The ADC was shown to be dependent on tissue type: Hansen [39] finds that the ADC is lower in excised brain tissue than in excised muscle tissue, and Cleveland et al [40] reports that the ADC is anisotropic in rat muscle tissue: the ADC is greater along the length of the muscle fibres than across them. Many subsequent studies have examined anisotropy in brain and nerve tissue, both *in-vivo* [41] and in excised samples [31, 15], with the aim of quantifying the anisotropy and in understanding its causes.

Measurement of the ADC in one dimension is insufficient for whole-brain imaging, because the anisotropic tissue is heterogeneously oriented. Even if the fibre orientations of all the anisotropic tissue was known, it would be impossible to acquire measurements along all the different orientations in a time that is tolerable for conscious subjects. We therefore attempt to calculate p, or the parameters of some model of p, in three dimensions.

Diffusion spectrum imaging (DSI) [42] measures $A(\mathbf{q})$ on a grid of evenly spaced $\mathbf{q}$ and computes the fast Fourier transform to obtain a discrete set of measurement of p on a grid of displacements, $\mathbf{r}$. This is a slow process because many measurements are required to provide sufficient detail in structure of p. Wedeen and Tuch [42, 43] use around 500 measurements per voxel. DSI typically uses a voxel size of approximately $64 \times 10^{-9}\mathrm{m}^{-3}$, which is necessary to achieve a sufficient signal to noise ratio.

The imaging parameters of DW-MRI are a trade off between spatial resolution, diffusion time, signal to noise ratio, and the total time spent by the subject or patient in the scanner. In current clinical scanners (field strength 1.5-3 Tesla) the spatial resolution of DW-MRI is less than for conventional MRI.

We can estimate p from fewer measurements if we assume a model of p. This reduces the minimum number of measurements p to the number of parameters in the model, which we fit to the data.

3.4.1 Gaussian model

In Diffusion-tensor MRI (DT-MRI) [15] we assume that p is a trivariate, zero-mean Gaussian distribution:

$$p(\mathbf{r}) = G(\mathbf{D}, t) = \frac{1}{\sqrt{(4\pi t)^3 \det(\mathbf{D})}} \exp\left(-\frac{\mathbf{r}^T \mathbf{D}^{-1} \mathbf{r}}{4t}\right) \tag{3.3}$$

where $\mathbf{D}$ is the diffusion tensor and t is the diffusion time. The tensor is symmetric because $p(\mathbf{r}) = p(-\mathbf{r})$ (diffusion is equally likely in both directions), and positive semi-definite because the diffusion coefficient in any direction is always greater than or equal to zero. In the literature, the diffusion tensor is sometimes called positive-definite, because in practice we do not observe zero diffusion, and hence the eigenvalues of $\mathbf{D}$ are positive. Because of symmetry $\mathbf{D}$ has six independent elements. Substituting this model into equation 3.1, we obtain

$$A(\mathbf{q}) = \exp(-t\mathbf{q}^{\mathrm{T}}\mathbf{D}\mathbf{q}), \tag{3.4}$$

and thus

$$-\log(A(\mathbf{q})) = t\mathbf{q}^{\mathrm{T}}\mathbf{D}\mathbf{q}. \tag{3.5}$$

Each measurement $A(\mathbf{q})$ provides a linear constraint on $\mathbf{D}$. Given at least six diffusion-weighted measurements with different $\mathbf{q}$, it is possible to solve for $\mathbf{D}$.

It is common to take more than six measurements to provide a more robust estimate of $\mathbf{D}$ in the presence of noise. A common approach is to acquire these measurements on a spherical shell with fixed $\mid \mathbf{q} \mid$. The directions must be spread evenly on the sphere to avoid bias. Jones [44] shows in simulation that for a total of N measurements, the best estimate of $\mathbf{D}$ comes from using N different $\mathbf{q}$ rather than NM^{-1} measurements of M different $\mathbf{q}$. The diffusion weighting is often expressed in terms of the scalar "b-value", where $b = \mid \mathbf{q} \mid^2 t$. The voxel volume in DT-MRI is typically much smaller than for DSI, approximately $8 \times 10^{-9}\mathrm{m}^3$.

The diffusion tensor may also be calculated by a direct fit of Eq. 3.4 to the measurements by numerical optimization. A direct fit can improve robustness [45] because the noise is closer to Gaussian on the measurements than on the log-measurements. We may also incorporate constraints on the diffusion tensor. For example, we can constrain the tensor to be positive definite by writing $\mathbf{D}$ in terms of its Cholesky decomposition $D = \mathbf{U}^{\mathrm{T}}\mathbf{U}$, where $\mathbf{U}$ is an upper triangular matrix. We then optimise the parameters of $\mathbf{U}$ in the equation $A(\mathbf{q}) = \exp(-t\mathbf{q}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{q})$.

The eigen system of $\mathbf{D}$ contains three eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, and their corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The eigenvectors describe the principal axes of the diffusion, and the eigenvalues are proportional to the mean squared displacements along the corresponding eigenvector. The diffusion tensor can be visualised as an ellipsoid, where the eigenvectors define the orientation and the eigenvalues define the size of the ellipsoid axes.

The introduction of DT-MRI led to a rapid growth in *in vivo* studies of diffusion in the brain, and the Gaussian model remains popular in the literature. The acquisition time is modest and the computation of $\mathbf{D}$ is fast on a standard PC. Useful scalar indices can be derived from the DT for quantitative analysis of the diffusion. The mean squared molecular displacement R^2 is related to the trace of the DT: $R^2 = 2t\mathrm{Tr}(\mathbf{D})$. The anisotropy can be quantified in different ways; a frequently used index is the fractional anisotropy [46]

$$f = \left(\frac{3}{2} \sum_{i=1}^{3} \left(\lambda_i - \frac{1}{3}\mathrm{Tr}(\mathbf{D}) \right)^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{3} \lambda_i^2 \right)^{-\frac{1}{2}} . \tag{3.6}$$

Tensor-based characterisations of the anisotropy are useful for studying heterogeneous white matter structures in the brains, where a comparison of the ADC along two or three directions as in [41] is insufficient. Bammer et al [47] report a reduction in anisotropy of normal-appearing white matter in multiple sclerosis patients, suggesting microstructural changes outside the sites of visible lesions. The physical mechanism for these changes is still being studied. Multiple sclerosis causes demyelination, axonal degeneration, and oedema (excess water at the site of inflammation), which all influence p.

The statistic $\mathrm{Tr}(\mathbf{D})$ is widely used to highlight areas of acute cerebral ischemia [48]. A reduction in anisotropy has been found in some fibres of patients with schizophrenia [49, 50] and epilepsy [51]. Multiple sclerosis lesions exhibit reduced anisotropy and increased $\mathrm{Tr}(\mathbf{D})$, [52].

The eigenvectors of the DT describe the orientation of p. If the voxel contains only fibres organised into a single bundle, the DT is highly anisotropic, the ellipsoid is prolate, like a cigar, and the primary eigenvector, $\mathbf{e}_1$, provides an estimate of the fibre orientation, which is used to track the path of white matter fibres. We discuss fibre tracking methods in Ch. 5.

A limitation of DT-MRI is that the tissue structure within a voxel is often not homogeneous, and the Gaussian model is a poor approximation of p in these cases. When a voxel contains multiple fibre bundles, p has a ridge corresponding to each bundle, but the tensor model cannot describe this situation. When two fibre bundles cross at an acute angle, $\mathbf{e}_1$ consistently points between the two fibre bundles. When the fibres are perpendicular, $\lambda_1 \approx \lambda_2 \gg \lambda_3$ and the ellipsoid is oblate, like a pancake, $\mathbf{e}_1$ may lie anywhere in the plane normal to $\mathbf{e}_3$ and the anisotropy is substantially underestimated. If three or more fibres cross, then the anisotropy will be further underestimated and the tissue in the voxel may be misclassified as isotropic grey matter.

As well as fibre crossings, the discretisation of the brain into a voxel grid often creates voxels that contain different tissue types. This also decreases anisotropy and makes $\mathbf{e}_1$ a less reliable estimate of the fibre orientation.

3.4.2 Detecting non-Gaussian diffusion

Frank [53] fits a spherical harmonic series to diffusion-weighted measurements acquired on a spherical shell. Any function on the sphere $f(\theta, \phi)$ can be expressed using spherical-harmonic basis functions:

$$f(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}. \tag{3.7}$$

A spherical harmonic series containing orders up to $l = L$ is equivalent to an order L polynomial on the sphere, and vice versa.

A standard acquisition of N measurements with $A(\mathbf{q})$ with independent directions $\hat{\mathbf{q}}$ each give a measurement of the ADC along the direction $\hat{\mathbf{k}}$: $d(\hat{\mathbf{k}}) = -b^{-1}\log(A(\mathbf{q}))$. Each ADC is real valued, so the coefficients, a_{lm}, in Eq. 3.7 are also real. The ADC is also symmetric, $(d(\hat{\mathbf{k}}) = d(-\hat{\mathbf{k}}))$, so any ADC can be described by a spherical harmonic series of even order, $l = 0, 2, 4, \ldots, \infty$. Depending on the shape of the ADC, the coefficients of high order terms may be zero after some maximum order L. When p is perfectly isotropic, $L = 0$, since $d(\hat{\mathbf{k}})$ is constant for all $\hat{\mathbf{k}}$. If p is anisotropic but still Gaussian then $L = 2$. If p is not Gaussian then $L > 2$. Higher order terms also model the effects of noise. Alexander et al [54] use analysis of variance (ANOVA) to test whether a series with $L = M + 2$ fits significantly better than a series with $L = M$. Their algorithm fits $L = 2$ to most white matter in the brain where the DT is prolate, but fits $L = 4$ in many voxels in regions of crossing fibres such as the pons.

Özarslan and Mareci [55] use higher-order tensors in place of the second-order tensor used in most DT-MRI:

$$\log(A(\mathbf{q})) = -t\sum_{j=1}^{\infty} \mathbf{q}^{(j)}\mathbf{D}^{(2j)}\mathbf{q}^{(j)}, \tag{3.8}$$

where each term is the contraction of the order $2j$ tensor with $\mathbf{q}^{(j)}$, which is the outer product of $\mathbf{q}^{(1)} = \mathbf{q}$ and $\mathbf{q}^{(j-1)}$. Truncating the series of higher order tensors at order $j = L$, we get an order L polynomial in $\mathbf{q}$, which is equivalent to a real, symmetric, spherical harmonic series of order L [18].

3.4.3 Multi-Gaussian model

The multi-Gaussian model assumes that a voxel is a mixture of two (or more) separate tissue compartments, where p in each compartment is a Gaussian distribution. The simplest and most widely used model assumes that the exchange of water molecules between the compartments is negligible. The combined PDF is then a weighted sum of the individual Gaussian distributions:

$$p(\mathbf{r}) = \alpha_1 G(\mathbf{D}_1, t) + \cdots + \alpha_n G(\mathbf{D}_n, t), \tag{3.9}$$

where n is the number of compartments, $\alpha_i \in [0,1]$ and $\sum_{i=1}^{n} \alpha_i = 1$. Substituting into equation 3.1 gives

$$A(\mathbf{q}) = \sum_{i=1}^{n} \alpha_i \exp(-t\mathbf{q}^{\mathrm{T}}\mathbf{D}_i\mathbf{q}) \tag{3.10}$$

Because this model has more parameters than the single Gaussian model, more measurements are required to fit the parameters. Eq. 3.10 is nonlinear and is solved by optimisation of the parameters, which can easily converge on a local minimum, making the solution dependent on initial estimates. The fitting can be made more reliable by trying multiple starting points, or by imposing constraints such as fixed α or cylindrical symmetry of the tensors. The value of n is fixed and must be assigned before the optimisation. Many authors [16, 56, 57] test for a departure from Gaussian diffusion, for example by using the method in [54], and assign $n = 2$ in these voxels. When $n = 2$, Tuch [16] shows some success at resolving two fibre directions. Most studies use $n \leq 2$ in all voxels, because it is difficult to detect and fit more compartments consistently.

3.4.4 Model-free inversion

Diffusion spectrum imaging (DSI) [42] measures $A(\mathbf{q})$ on a grid of evenly spaced $\mathbf{q}$ and computes the fast fourier transform to obtain a discrete set of values of p on a grid of displacements, $\mathbf{r}$. A radial projection gives the Orientational Density Function (ODF) ϕ:

$$\phi[\hat{\mathbf{r}}] = \int_0^\infty p(\rho\hat{\mathbf{r}})d\rho, \tag{3.11}$$

which is the integral of p in the radial direction along a certain orientation $\hat{\mathbf{r}}$. Therefore $\phi[\hat{\mathbf{r}}]$ gives the probability of a particle displacement into a differential solid angle about the direction $\hat{\mathbf{r}}$, so ϕ peaks in directions of the ridges of p. This projection discards the radial information in the MRI measurements.

If we are only interested in the angular structure of p, it is more efficient to make measurements spread over a spherical shell at fixed $\mid \mathbf{q} \mid$ and approximate ϕ. Methods such as q-ball imaging and the persistent angular structure (PAS-MRI) can resolve multiple fibre orientations in clinical scanners with far fewer measurements per voxel than DSI.

Jansons and Alexander [58] use the persistent angular structure (PAS) $\widetilde{p}$, where:

$$p(\mathbf{r}) = \widetilde{p}(\hat{\mathbf{r}})\rho^{-2}\delta(\mid \mathbf{r} \mid -\rho), \tag{3.12}$$

ρ is a constant and δ is the delta function ($\delta = 1$ if $\mid \mathbf{r} \mid = \rho$, 0 otherwise). According to this formulation p is zero everywhere except the surface of the sphere of radius ρ. This results in a form of p that contains only angular information. Jansons derives a general parametric form of $\widetilde{p}$ that maximises the entropy with respect to the data. The parameters must be found by numerical optimisation, which is very time consuming. The peaks of the PAS can describe multiple fibre directions, and can be computed with far fewer measurements than DSI.

Tuch's q-ball algorithm [43] uses the Funk transform [59], to generate a function that closely resembles ϕ, from measurements at fixed $\mid \mathbf{q} \mid$. Lin et al [60] propose a similar method to q-ball, which they demonstrate with 253 measurements per voxel. Q-ball is much faster than PAS-MRI, since it does not involve an optimisation process. Tuch demonstrates q-ball using 492 measurements per voxel, however, Alexander [61] applies it to data with 60 measurements per voxel. Alexander shows that q-ball and PAS-MRI perform well with 60 measurements, though q-ball requires a slightly higher signal to noise ratio (SNR) to resolve two fibre orientations as consistently as PAS-MRI. Fig. 3.3 shows a slice of a q-ball reconstruction in brain data with 60 measurements per voxel, overlaid on a map of the tensor fractional anisotropy. The peaks of the ODF appear to correspond to the orientation of the white matter fibre bundles. The zoomed region shows a fibre-crossing region, where two peaks are resolved by the q-ball method.

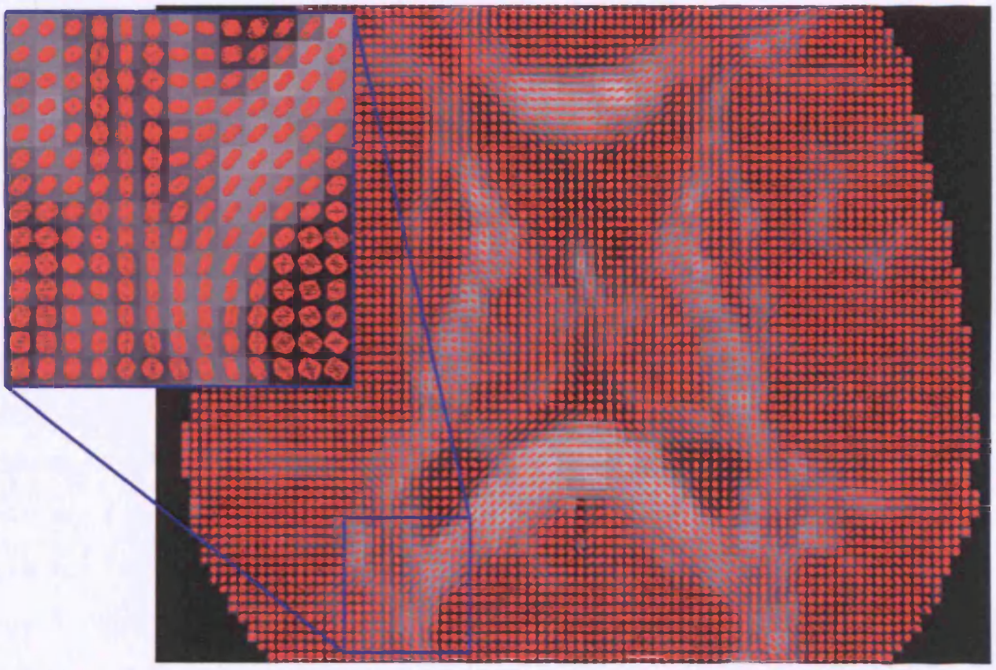


Figure 3.3: Q-ball reconstruction of the ODF in the brain, overlaid on a fractional anisotropy map. The peaks of the ODF align with the fibre orientation in white matter. Image provided by Kiran Seunarine, generated using the Camino toolkit [1].

3.5 Conclusions

The experiments in this thesis use both the Gaussian model and the multi-compartment Gaussian model. We focus on DT-MRI first in Ch. 4. The Gaussian model is computationally inexpensive compared to more advanced algorithms, which facilitates methods and experiments that require large samples of synthetic data. Also, the diffusion tensor provides easily computable statistics of anisotropy, such as the fractional anisotropy, which has been shown to correlate with the uncertainty in the associated fibre-orientation estimate [56]. The multi-compartment Gaussian model is a natural extension of this model, which is still reasonably computationally efficient and allows us to use the same statistics of anisotropy to model the uncertainty as a function of the tensors fitted in multi-fibre voxels.

Chapter 4

Optimal orderings of diffusion-weighted MRI measurements on the sphere

Diffusion-tensor MRI [62] (DT-MRI) assumes a Gaussian model of the spin displacement density. Each measurement A at wavenumber $\mathbf{q}$ provides a linear constraint on the diffusion tensor D:

$$\log[A(\mathbf{q})/A(\mathbf{0})] = -t\mathbf{q}^{\mathrm{T}}D\mathbf{q}\,, \tag{4.1}$$

where t is the diffusion time. Since D has six independent elements, DT-MRI requires at least seven measurements with independent $\mathbf{q}$. A standard approach is to acquire N measurements $A(\mathbf{q}_1), \ldots, A(\mathbf{q}_N)$ on a spherical shell in $\mathbf{q}$-space, with $\mid \mathbf{q}_1 \mid = \mid \mathbf{q}_2 \mid \cdots = \mid \mathbf{q}_N \mid$, and M measurements with $\mathbf{q} = \mathbf{0}$. The gradient directions $\hat{\mathbf{q}}_1, \ldots, \hat{\mathbf{q}}_N$ are spread isotropically on the sphere to minimise dependence of the error in the measured diffusion tensor on the orientation of imaged tissue. Several algorithms in the literature determine isotropically spread sets of N directions. Tuch et al [16] use the vertices of a tessellated icosahedron as gradient directions. Jones et al [63] model the directions as antipodal pairs of charged particles (at $\hat{\mathbf{q}}_i$ and $-\hat{\mathbf{q}}_i$), joined by a light, rigid rod that passes through the centre of the unit sphere. Each pair is free to rotate on the surface of the sphere. Jones et al search for the point set that minimises the sum of the forces between the particles. Papadakis et al [64] maximise the minimum distance between any two charge pairs. Hasan et al [65] review and compare several methods and evaluate their suitability for fitting the diffusion tensor. They conclude that for estimating diffusion anisotropy, using N unique directions provides no advantage over acquiring $N/6$ repeated measurements in each of 6 directions. However, Jones [44] shows that for highly anisotropic tissue, N unique directions provide more stable estimates of the principal direction and anisotropy of the diffusion tensor. Unique gradient directions also allow resolution of multiple fibre-orientations within a single voxel [18].

In this chapter, we define the electrostatic point set as the directions found by minimising the electrostatic energy of equal and opposite pairs of charged particles on the sphere. We obtain electrostatic point sets using Jansons and Alexander's procedure [58]. They choose the lowest energy configuration from 500 local minima computed from different random starting configurations. Previous approaches [63, 64] use one starting position, so their results may be suboptimal, particularly for large N. The electrostatic energy of the two pairs of charges i and j is

$$E_{ij} = 2\sqrt{2}[(1+\hat{\mathbf{q}}_i \cdot \hat{\mathbf{q}}_j)^{-\frac{1}{2}} + (1-\hat{\mathbf{q}}_i \cdot \hat{\mathbf{q}}_j)^{-\frac{1}{2}}] \,. \tag{4.2}$$

The total energy of the system of N pairs is

$$E_N = \sum_{i=2}^{N}\sum_{j=1}^{i-1} E_{ij} \,. \tag{4.3}$$

If a scan is interrupted, because of excessive patient motion for example, the set of directions acquired may be anisotropically distributed, because subsets of an isotropic point set are not in general isotropic. Several neurological conditions currently studied with diffusion-MRI, such as epilepsy, dementia, and schizophrenia, can cause patients to become non-compliant or to have seizures during the acquisition. Partial scans would generally be excluded from a research study, but a partial scan may still be useful for clinical diagnosis if the subset of gradient directions used before the interruption is close to isotropic.

Dubois et al [66], propose a method to improve results from partial scans by separating gradient directions that are positioned close together in the acquisition sequence. They introduce weights to the energy in Eq. 4.3 and they use point sets that minimise

$$E_N = \sum_{i=2}^{N}\sum_{j=1}^{i-1} E_{ij} w_{ij} \,. \tag{4.4}$$

They propose three weighting schemes. In the first, pairs are divided into subsets of six directions, $w_{ij} = 1$ if pairs i and j are in the same subset, and $w_{ij} = 0.6$ if the pairs are not in the same subset. In the second scheme, pairs in adjacent subsets have an intermediate weight $w_{ij} = 0.8$. In the third scheme, there are no subsets and the weights decrease gradually according to the order in the acquisition sequence: $w_{ij} = 1$ if $\mid i-j \mid \leq 10$, $w_{ij} = \mid i-j \mid^{-\alpha}$ if $\mid i-j \mid > 10$, and α is a positive scalar. They calculate the point sets with a simulated annealing algorithm.

Dubois et al demonstrate the method with a point set of 18 directions. They scan the brain of a healthy adult volunteer using both their point sets of 18 directions and with electrostatic point sets of 6, 12, and 18 directions. They show that maps calculated using the first $P = 6, 12, 18$ directions of their point sets appear similar to those produced from the corresponding electrostatic point set of P directions. A drawback of this approach is that the point set that minimises Eq. 4.4 does not minimise Eq. 4.3, and is less isotropic than the electrostatic point set. Empirically, we find also that the weights in Eq. 4.4 increase the number of local minima, making the global minimum of Eq. 4.4 harder to locate.

In a previous work [3] we divide the electrostatic point set into equally-sized subsets, and optimise the ordering of the pairs to make each subset as isotropic as possible. Specifically, we minimise

$$E_s = \sum_{i=2}^{N} \sum_{j=1}^{i-1} E_{ij} \delta_{ij} , \tag{4.5}$$

where $\delta_{ij} = 1$ if pairs i and j are in the same subset, and zero otherwise. The purpose of the method is to create isotropic subsets, so that the diffusion tensor can be fitted to the measurements in each subset. One potential application for this is motion correction using tensor information derived from the subsets at different points during the scan. It also improves the results of partial scans compared to an unordered point set, however the method treats subsets independently, so it does not minimise the energy of the first P pairs. Therefore, a partial scan consisting of, for example, the first two subsets, or the first subset and half of the second, may be significantly anisotropic.

We propose a new method to optimise the ordering of acquisition and thus improve the quality of partial scans without compromising the quality of the complete scan. We test the ordered point sets and compare them with the Dubois point sets in simulation and we also demonstrate the ordered point set using brain data from a healthy volunteer.

4.1 Methods

This section presents the method to find an ordering of an isotropic point set of size N gradient directions that gives the most well separated subsets $\{\hat{\mathrm{q}}_1, \ldots, \hat{\mathrm{q}}_P\}$, $6 \leq P \leq N$. In contrast to Dubois' approach, the method does not alter the quality of the complete scan, since none of the directions are modified.

We cannot know in advance how many measurements the scanner will complete before interruption, and so we require that each subset of P directions is well separated. To achieve this, the method optimises $(N - 6)$ nested subsets of the first P directions, in contrast to the earlier work by Cook et al [3], which divides directions into non-overlapping subsets. We aim to minimise the electrostatic energy of all subsets simultaneously, so we search for the ordering that minimises

$$f = \sum_{P=6}^{N} E_P P^{-2} , \tag{4.6}$$

where E_P is the electrostatic energy of the subset $\{\hat{\mathrm{q}}_1, \ldots, \hat{\mathrm{q}}_P\}$. The electrostatic energy of P isotropically distributed pairs is approximately proportional to P^2; the normalisation factor P^2 in Eq. 4.6 ensures that each subset $\{\hat{\mathrm{q}}_1, \ldots, \hat{\mathrm{q}}_P\}$ contributes similarly to the objective function.

An exhaustive search of all $N!$ orderings is feasible for small N but rapidly becomes intractable as N grows. We search for the ordering that minimises f using the simulated annealing [67, p. 444] method in Algorithm 1. The parameters T, ϵ, R and S control the annealing schedule. We decrease T to $(1-\epsilon)T$ after trying S random changes to the configuration, so positive changes in f are progressively less likely to be accepted. The counter R records how many times T has been lowered since we improved the best configuration. Simulated annealing is a heuristic algorithm and we have no guarantee that a particular minimum is the global minimum. We therefore test a range of annealing schedules to provide an estimate of the optimum algorithm parameters. We compute orderings over a range of N. Specifically, we compute orderings for $N = 12, 18, 30, 61$. We find for each N that $f_{\min}$ never reduces when T is less than approximately 10^{-5}. The optimisation ends when T is below $T_{\min}$ or when T has been lowered 10^6 times without a change in $f_{\min}$. We set $T_{\min} = 10^{-8}$ for each N and fix $S = 10^4$.

```
1. Read unordered electrostatic point set.
2. $T = T_{\max}; R = 0$
3. $f_{\min} = f$
4. while $R < 10^6$ and $T > T_{\min}$
   (a) for $x = 1 \ldots S$
       i. Choose two points $i$ and $j$ and swap their position.
       ii. Calculate the change $\delta_f$ in $f$.
       iii. If $\delta_f \leq 0$ accept the change, else accept the change with probability $\exp[-\delta_f/T]$.
       iv. if $f < f_{\min}$, set $f_{\min} = f$ and store the current configuration.
   (b) If $f_{\min}$ changed during 4a, then $R = 0$. Else $R = R + 1$.
   (c) $T = (1 - \epsilon)T$.
5. Return the best configuration.
```

Algorithm 1: The simulated annealing algorithm.

We choose the initial temperature $T_{\max}$ for the simulated annealing using a simple calibration algorithm. Starting from a random ordering, we run a single annealing iteration (statement 4a in Algorithm 1) and determine the temperatures T_{10}, T_{30}, and T_{50} at which 10%, 30%, and 50% respectively of the upward steps ($\delta_f > 0$) are accepted.

We vary the cooling parameter ϵ between 10^{-3} and 10^{-6}. With $T_{\text{max}} = T_{50}$, the algorithm terminates in approximately 5 minutes with $\epsilon = 10^{-3}$ and 48 hours with $\epsilon = 10^{-6}$ on a 2.4 GHz PC workstation.

For $N = 12$, we perform an exhaustive search of all possible orderings to find the global minimum configuration. The annealing algorithm finds this minimum whenever $\epsilon \leq 10^{-4}$, with $T_{\text{max}} = T_{10}, T_{30}, T_{50}$. For $N = 18$ and 30, we cannot search all possible orderings, however the annealing algorithm finds the same minimum when $\epsilon \leq 10^{-4}$, with $T_{\text{max}} = T_{10}, T_{30}, T_{50}$. With $N = 61$, the annealing finds the same minimum with $T_{\text{max}} \geq T_{30}$ and $\epsilon = 10^{-6}$. Higher values of N will likely require slower annealing (smaller ϵ and higher T_{max}).

4.2 Experiments and results

This section shows some experiments to compare different point sets with scan interruption. We show quantitative results from simulations and some qualitative results from human-brain data. We compare the first P directions from different point sets:

1. Ordered electrostatic.
2. Dubois [66].
3. Subset electrostatic [3].
4. Unordered electrostatic [58]

For comparison, we compute the same statistics from electrostatic point sets with $N = P$. We compare two values of N: 18, as used by Dubois et al [66], and 61, typical of high angular resolution acquisitions and used in [3].

We compute point set 2 for each N using the third weighting scheme from [66], which Dubois et al propose for use when scan interruption may occur at any point in the acquisition. In this scheme, $w_{ij} = 1$ if $\mid i - j \mid \leq 10$, $w_{ij} = \mid i - j \mid^{-\alpha}$ if $\mid i - j \mid > 10$. We determine the value of α empirically by finding the α (to the closest 0.1) that gives a point set with the minimum f in Eq. 4.6. For $N = 18$, $\alpha = 0.4$, for $N = 61$, $\alpha = 0.3$. To find the optimal point set for each α, we follow [58] and use a Levenberg-Marquardt optimisation to minimise Eq. 4.4 from each of 50,000 random starting points and pick the lowest minimum.

We compute point set 3 for $N = 18$ by dividing the pairs into three groups of 6 pairs. We optimise Eq. 4.5 using simulated annealing as described in Sec. 4.1, with $T_{\text{max}} = T_{50}$ and

$\epsilon = 10^{-5}$. For $N = 61$, we divide the pairs into one group of 16 pairs, and three groups of 15 pairs. We run the annealing with $T_{\text{max}} = T_{50}$ and $\epsilon = 10^{-6}$. After annealing for $N = 61$, we choose directions from the first subset that give the minimum E_P with $P = 6$ and make these the first six directions of the subset. This ensures that the first six directions are as isotropic as possible without changing the subsets.

4.2.1 Evaluation of ordered points

We evaluate the point sets using synthetic data. Given a Gaussian test function with diffusion tensor D we synthesise the first P measurements. We then add complex Gaussian noise and take the modulus. The diffusion time, $\mid \mathbf{q} \mid$ and signal to noise ratio are the same as in the human brain data (described below). For $N = 61$, the number of unweighted measurements, M, varies according to P. For $N = 18$, $M = 1$. The diffusion tensor is calculated by a linear least-squares fit to the log of the noisy measurements.

For each D we obtain 5000 independent noisy estimates of the fractional anisotropy (FA) [46] and the principal direction at each of 1000 random rotations of D, giving a total of 5×10^6 trials.

We use three different diffusion tensors typical of those found in the brain. All three are cylindrically symmetric with a trace of $2.1 \times 10^{-9}\,\text{m}^2\,\text{s}^{-1}$. Tensor D_1 has FA $= 0.8$, which is typical of major white-matter fibre tracts. Tensor D_2 has FA $= 0.4$, and D_3 has FA $= 0.1$, which is often used as a threshold in tractography to detect grey matter.

The concentration of the estimated principal axes $\mathbf{x}_1 \ldots \mathbf{x}_{5000}$ about their mean axis $\mathbf{m}$ is found by fitting the parameter κ of the Watson distribution: $p(\mathbf{x}) = \exp[\kappa(\mathbf{m} \cdot \mathbf{x})^2]$ [68]. A larger value of κ means that the principal directions are more concentrated about $\mathbf{m}$. A uniform distribution of directions on a sphere has zero concentration, while identically aligned axes have infinite concentration.

4.2.2 Comparing the point sets

Figures 4.1-4.6 show the results for D_1 only. The trends are similar for D_2 and D_3 over all values of P, so we omit these to avoid cluttering the plots. We define the "angle bias" as the angle between the mean of the estimated principal directions, $\mathbf{m}$, and the true principal direction for each orientation of D.

We plot $\min(\kappa)$ over the 1000 orientations of D for each point set and value of P. An anisotropic point set has a high angle bias, and produces to high concentration at some orientations of D and low concentration at others. Aside from point set 4, all point sets have a mean angle bias of less than 3 degrees with $P > 6$, and point sets with the highest $\min(\kappa)$ have the lowest angle bias. At $P = 6$, point set 2 has a higher angle bias but $\min(\kappa) \approx 0$.

Fig. 4.1 plots the minimum concentration of the estimated principal directions against P with $N = 18$. The minimum concentration for point set 1 is lower than for point set 2 for most P, but higher when $P = 6$ and $P = 18$. Point set 3 performs worse than point sets 1 and 2 when P is not a multiple of six, but is still better than point set 4.

Fig. 4.2 plots the minimum concentration of the principal directions against P with $N = 61$. Point sets 1 and point set 2 produce similar results, although point set 2 has lower $\min(\kappa)$ when $P = 6$. Point set 3 has similar performance to point sets 2 and 1; at some P point set 3 is slightly worse and at others slightly better. As with $N = 18$, point sets 1, 2 and 3 are all better than point set 4.

Fig. 4.3 shows the mean FA over all trials for $N = 18$. Point sets 1, 2 and 3 give similar results to the optimal $N = P$ point set with $P > 6$. Point set 3 is slightly less accurate than point sets 1 and 2, but still better than point set 4. Fig. 4.4 shows the standard deviation of FA with $N = 18$. Point set 2 has higher standard deviation than point set 1 with $P = 6$, but slightly lower standard deviation for $P = 7$ and $P = 8$. Point set 3 has higher standard deviation than point sets 1 and 2 for $P < 12$.

Fig. 4.5 shows the mean FA for $N = 61$. Fig. 4.6 shows the standard deviation of FA for $N = 61$. In both plots, point sets 1, 2 and 3 are similar for $P > 6$, and give more accurate and consistent measurements of FA than point set 4.

4.2.3 Human brain data acquisition

We acquire human brain data on a 3T scanner (GE Excite II platform, running G3M4 software), equipped with $0.04\,\mathrm{T\,m^{-1}}$ gradients, using an 8 channel phased array coil. Imaging parameters are $\delta = 21\,\mathrm{ms}$, $\Delta = 0.029\,\mathrm{s}$ (δ and Δ are defined in Fig. 3.2), $b = 1200\,\mathrm{s\,mm^{-2}}$ and $t_e = 0.073\,\mathrm{s}$. The image consists of fifty one 2.7 mm thick slices, with a 96×96 matrix, reconstructed to 256×256, with a 22 cm field of view. Cardiac gating was employed with a TR of 30RR and triggering occurring on every QRS complex.

The sequence acquires 7 unweighted ($b = 0$) images and 61 diffusion-weighted images. The gradient directions were calculated using the electrostatic energy minimisation algorithm in [58] and ordered using the procedure in the methods section. The $b = 0$ images are acquired at regular intervals throughout the scan, whenever the ratio of weighted to unweighted images is greater than $9 : 1$. This ratio of weighted to unweighted image acquisitions is a compromise between the optimal ratio for measuring anisotropy and that for measuring fibre orientations [38]. The signal to noise ratio in white matter in the $b = 0$ images is approximately 20.

Figure 4.7 compares the colour-coded principal direction maps with $P = 10$ from the

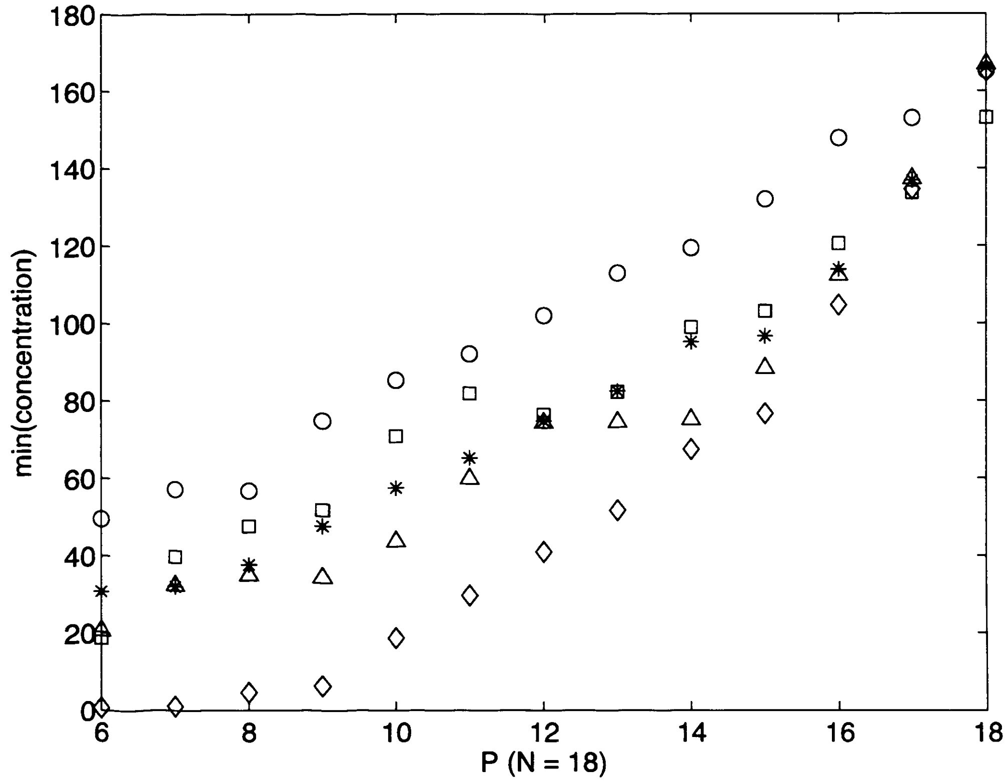


Figure 4.1: Plot of min(κ) against P with $N = 18$, for test function D_1 (FA $= 0.8$). The key is: electrostatic $N = P$ (circles), point set 1 (ordered, asterisks), point set 2 (Dubois, squares), point set 3 (electrostatic subsets, triangles) and point set 4 (unordered electrostatic, diamonds).

ordered point set, with $P = 10$ from the unordered point set, and with the full scan ($P = 61$). The map from the 10 ordered directions is closer to the results of the complete scan than the map from 10 unordered directions.

4.3 Discussion

We conclude that ordering the points significantly improves the results from a partially completed scan compared to a unordered point set. The Dubois point sets can give slightly more precise estimates of the diffusion tensor principal direction with small N, however the optimisation required to compute these point sets becomes difficult at high N, and ordering performs as well as the Dubois point sets at large N. The Dubois point sets can perform poorly with $P = 6$, because they do not explicitly optimise the first six directions, since the weights in Eq. 4.4 are equal for the first ten points. The subset electrostatic point sets [3] are also susceptible to this problem when the subsets contain more than 6 directions, though the problem can be alleviated to some extent by reordering the directions within the first subset.

We expect ordered point sets to be most useful in a clinical setting, where diffusion-

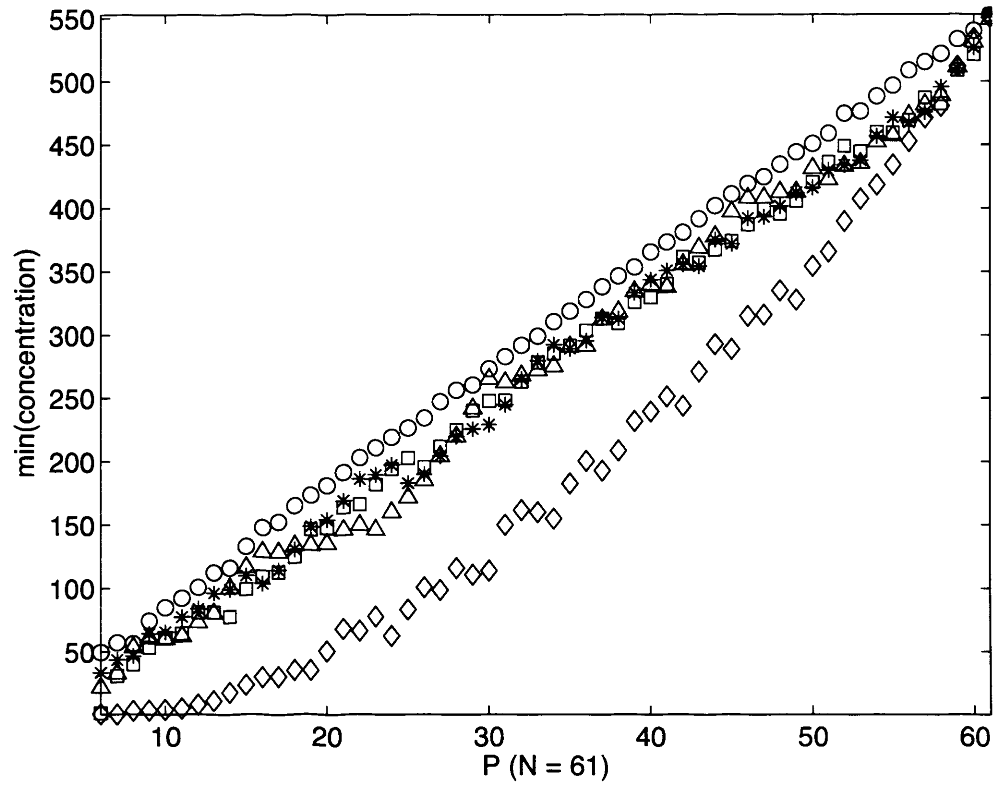


Figure 4.2: Plot of $\min(\kappa)$ against P with $N = 61$, for test function D_1 (FA $= 0.8$). The key follows that of Fig. 4.1.

weighted imaging is becoming increasingly common. Time constraints may prevent a restless patients from being rescanned, or patients can become non compliant or distressed (e.g. , from an epileptic seizure), after part of a high resolution diffusion-weighted image has been acquired. An ordered point set improves the measurements of the fibre orientation and fractional anisotropy from a partial scan and since ordering does not alter the data in a complete scan, it is simple to adapt existing acquisition sequences.

In this chapter, we have presented a method to optimise the acquisition scheme in order to improve the voxel-based measurements of the fibre orientation with DT-MRI. The next chapter introduces tractography, where these local measurements are used to trace the path of white matter fibre bundles in the brain.

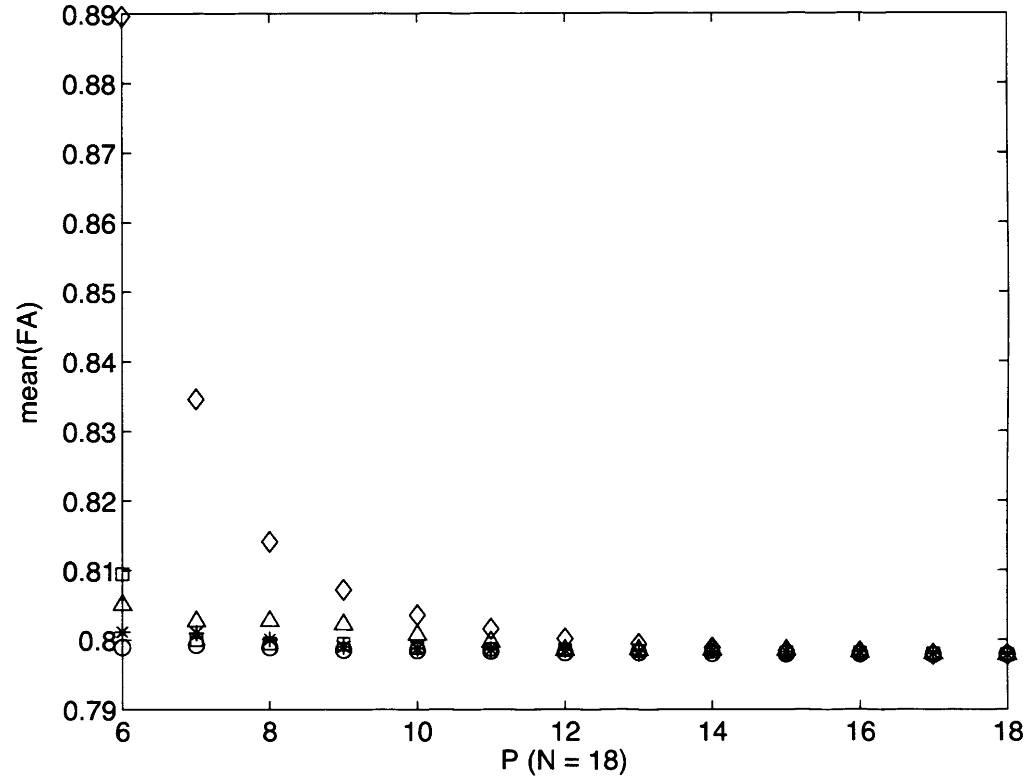


Figure 4.3: Plot of mean(FA) against P with $N = 18$ for test function D_1 (FA $= 0.8$). The key follows that of Fig. 4.1.

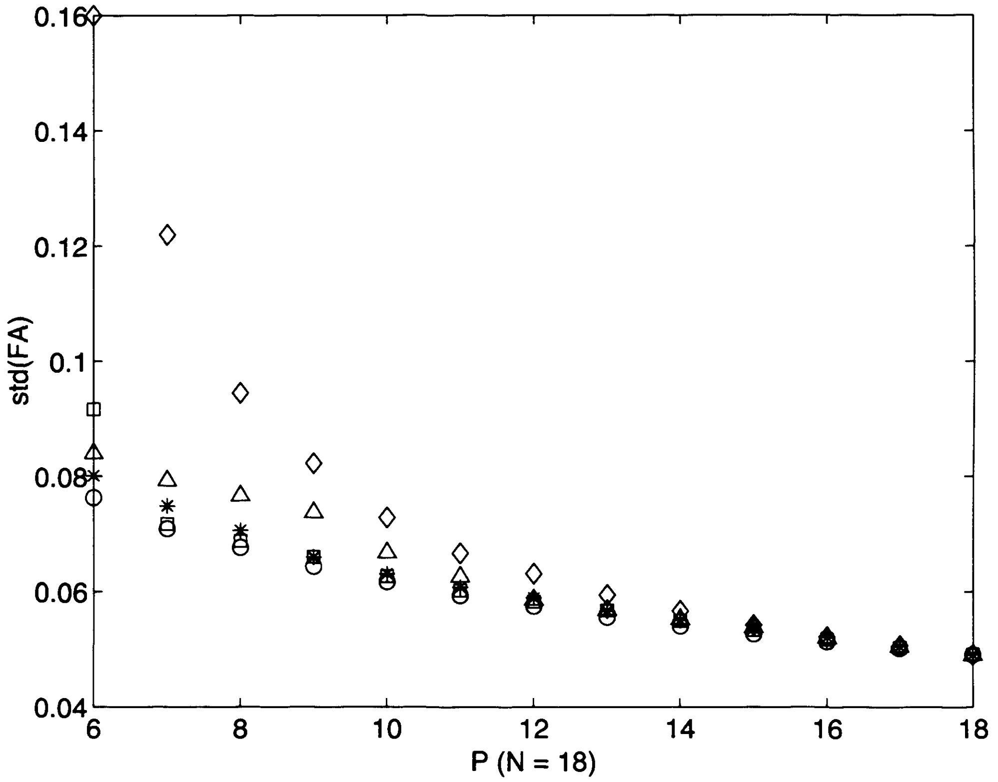


Figure 4.4: Plot of std(FA) against P with $N = 18$, for test function D_1 (FA $= 0.8$). The key follows that of Fig. 4.1.

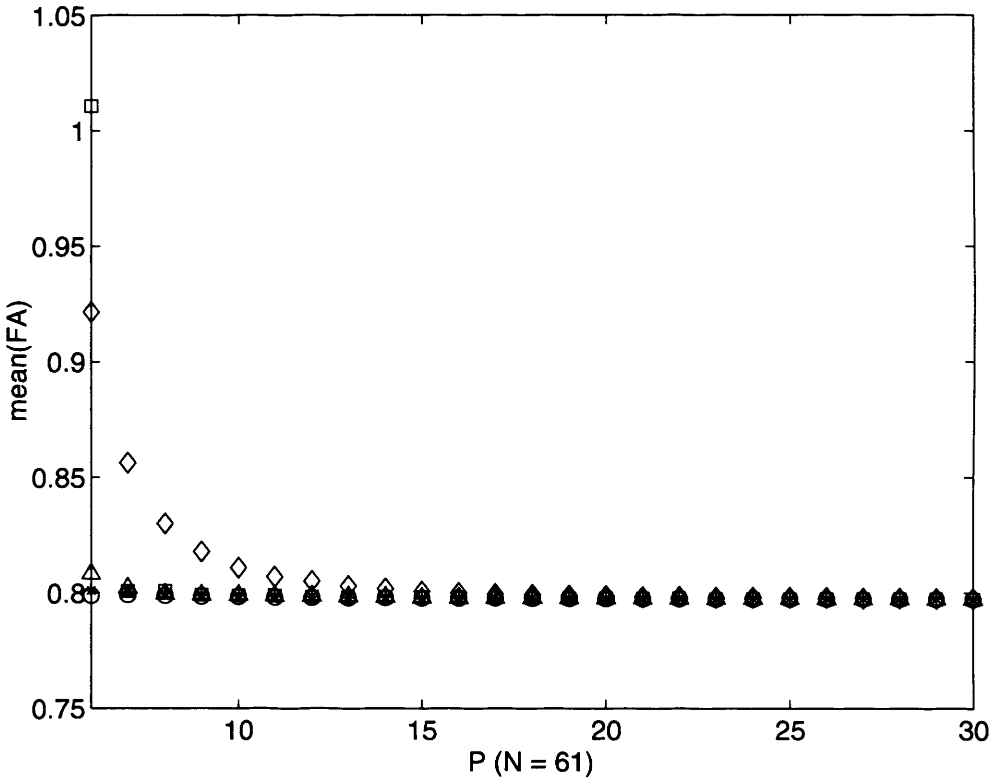


Figure 4.5: Plot of mean(FA) against P with $N = 61$ for test function D_1 (FA $= 0.8$). The key follows that of Fig. 4.1.

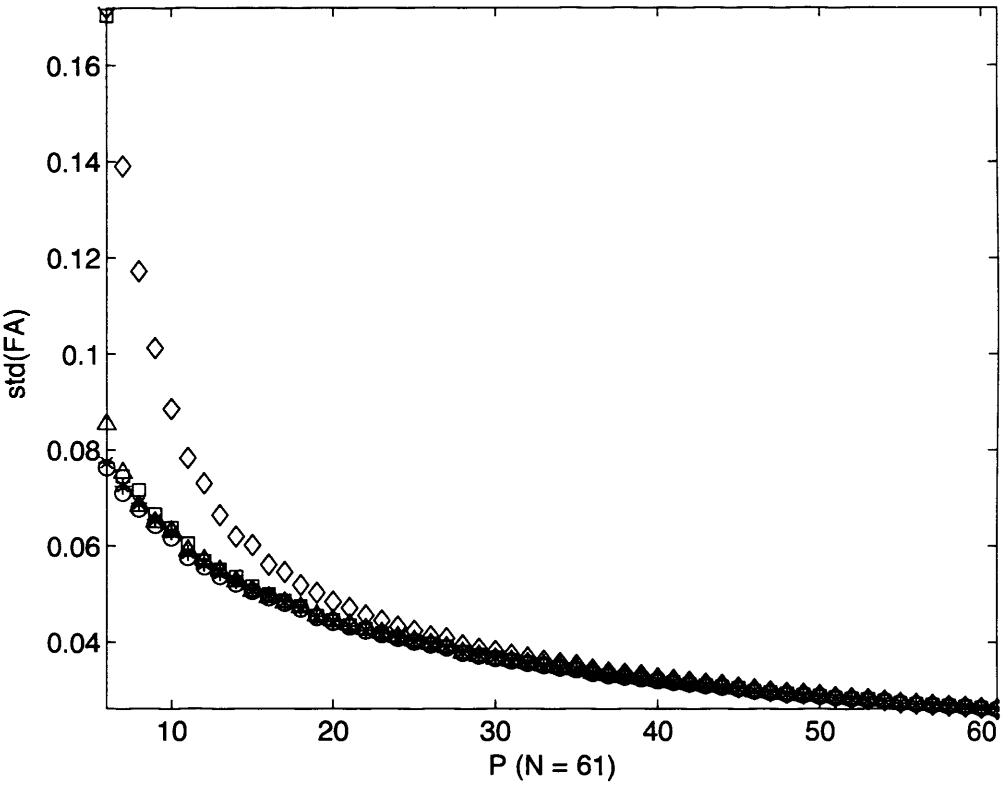


Figure 4.6: Plot of std(FA) against P with $N = 61$, for test function D_1 (FA $= 0.8$). The key follows that of Fig. 4.1.

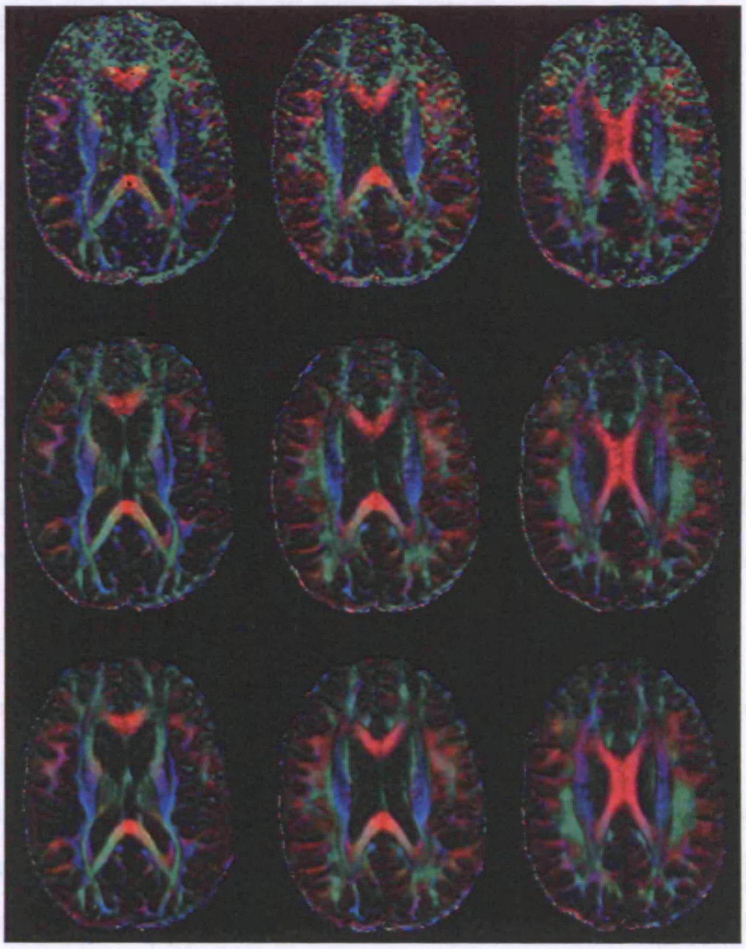


Figure 4.7: Colour-coded principal direction map from the human brain data, using ten random directions (top row), ten ordered directions (middle row), and all sixty-one directions (bottom row).

Chapter 5

Tractography with DW-MRI

Many recent studies have proposed methods to examine the nervous connectivity of the brain by tracing paths between sites that are connected by white matter fibres. This area of MRI research is commonly called tractography. Tractography methods can be split into two broad classes, deterministic and probabilistic. Deterministic methods estimate a single maximum-likelihood fibre-path. Probabilistic methods determine the probability of connection between any two points in the brain. Both methods have been used widely with DT data, though the principles extend naturally to other models of p. The work presented in this thesis relates to probabilistic streamline-based tractography, and we examine these methods in detail here.

5.1 Streamline tractography

Streamline tractography algorithms estimate the path of a fibre bundle from a given seed point. The algorithms work by propagating a particle along the local fibre orientation from the seed, until the end of the fibre bundle is reached. The path of the particle is a *streamline.* Fibre tracts are often visualised by tracking many streamlines seeded within the tract. This can give the appearance that streamlines represent individual fibres, in fact they are estimates of the path of a large bundle of axons.

5.1.1 Fibre assignment by continuous tracking

Fibre Assignment by Continuous Tracking (FACT) [9] was the first method to trace fibre paths using streamlines in DT data. The FACT method follows $\mathbf{e}_1$ in each voxel, so the stream particle changes direction only when the path crosses a voxel boundary. The streamline is continuous in that it may enter and exit the voxel at any point, and is not forced to pass through voxel centres.

To ensure that streamlines are constrained to white matter, FACT tests the angular dispersion of $\mathbf{e}_1$ in the neighbourhood surrounding a voxel. If the angular dispersion is high, then the tract is judged to have left white matter, and the streamline stops. FACT was first demonstrated by Mori et al in the rat brain [9], but the method has since been applied to the human brain (for example, by Stieltjes et al [69]).

5.1.2 Interpolated tracking

Conturo et al [70] track streamlines in DT data using a fixed, sub-voxel step size s. The streamline is propagated using Euler's method [67, p. 710]. Starting from a position $\mathbf{r}_0$, the next position of the streamline $\mathbf{r}_1$ is $\mathbf{r}_0 + s\mathbf{e}_1(\mathbf{r}_0)$. This requires a continuous approximation of the data to provide $\mathbf{e}_1(\mathbf{r})$ at sub-voxel intervals. They use trilinear interpolation of each measurement and recalculate the tensor at each step in the tracking process. Conturo finds that bundles of streamlines seeded in white matter trace paths that are visually consistent with known anatomy. Conturo also uses two regions of interest to reduce false positive connections. Seed points are placed in both regions and any streamline that does not pass through both regions is discarded. When regions of interest are placed at the ends of known fibre tracts, the number of spurious fibre trajectories is reduced because an erroneous streamline is unlikely to trace back onto the known path.

Basser [71] et al generate a continuous approximation of the diffusion tensor field, instead of interpolating the DW-MR data directly. They interpolate the tensors using spline-based interpolation of the diffusion tensors fitted to the data in each voxel [72]. They use the Runge-Kutta method [67, p. 710] to track through the interpolated tensor field, which is more precise than the Euler method. In the fourth-order variant of the Runge-Kutta method, the trajectory of the streamline along a step from $\mathbf{r}_0$ to $\mathbf{r}_1$ is a weighted average of four smaller steps:

$$\mathbf{v}_a = s\mathbf{e}_1(\mathbf{r}_0) \tag{5.1}$$

$$\mathbf{v}_b = s\mathbf{e}_1(\mathbf{r}_0 + \frac{s}{2}\mathbf{v}_a) \tag{5.2}$$

$$\mathbf{v}_c = s\mathbf{e}_1(\mathbf{r}_0 + \frac{s}{2}\mathbf{v}_b) \tag{5.3}$$

$$\mathbf{v}_d = s\mathbf{e}_1(\mathbf{r}_0 + \frac{s}{2}\mathbf{v}_c) \tag{5.4}$$

$$\mathbf{r}_1 = \frac{1}{6}(\mathbf{v}_a + 2\mathbf{v}_b + 2\mathbf{v}_c + \mathbf{v}_d). \tag{5.5}$$

Tench [73] et al also use the Runge-Kutta method, but they interpolate the raw data, rather than the tensor field.

In the "tensorline" scheme [74, 75] the velocity of the stream particle is an adaptable function of the path of approach, the local diffusion tensor information, and the path of approach deflected by the local diffusion tensor, the tensor deflection (TEND) term. Given a previous tracking step $\mathbf{v}_{in}$, the next step direction is

$$\mathbf{v}_{out} = f\mathbf{e}_1 + (1-f)((1-g)\mathbf{v}_{in} + g\mathbf{D}\cdot\mathbf{v}_{in}), \tag{5.6}$$

where f and g are weights between 0 and 1. When **D** is prolate, the TEND term will deflect $\mathbf{v}_{in}$ towards $\mathbf{e}_1$. If **D** is oblate the TEND term will deflect $\mathbf{v}_{in}$ toward the plane of the oblate DT, and if **D** is isotropic then the TEND term will be close to $\mathbf{v}_{in}$. The TEND algorithm can be adapted to choose f and g intelligently. For example, if the DT is highly prolate, then f should be close to 1. When the DT is oblate, $\mathbf{e}_1$ is unreliable and it is sensible to decrease f. A low value of g gives more weight to $\mathbf{v}_{in}$, which discourages sharply bending tracts without terminating the streamlines.

5.1.3 Stopping criteria for streamlines

Streamline algorithms are designed to terminate when the streamline has left white matter, or when the streamline trajectory is no longer trustworthy, for example if the streamline intersects itself or changes direction too rapidly. An anisotropy threshold terminates streamlines that enter an isotropic region [70]. A curvature threshold [71] terminates streamlines that curve too much over over a short path (e.g., one voxel). White matter fibre bundles generally curve slowly, and sharp curvature is often associated with erroneous paths. However, some pathways, such as the u-fibres, have naturally high curvature [76], and a restrictive curvature threshold makes tracking in these regions impossible.

5.2 Reliability of streamline tractography

Many investigations into the reproducibility of DT-MRI fibre orientation measurements have not explicitly dealt with tractography. This includes theoretical studies by authors such as Anderson [77], and empirical studies. Lin et al [78] trace fibre paths in the rat brain with a chemical tracer. They use a contrast-enhancing agent to highlight the optic tracts, and fitting a polynomial curve to the path of the optic tract. Lin compares the gradient of this curve to $\mathbf{e}_1$ in voxels along the path of the curve. The mean angular deviation of $\mathbf{e}_1$ from the polynomial curve was 1.1°, with standard deviation $\sigma = 13.27°$. Tuch et al [16] test the reliability of their multi-Gaussian model by adding noise to simulated tensors and calculating the mean error in the orientation of the primary eigenvectors.

Most evaluation of the reliability of tractography has used synthetic data, where the ground truth is explicitly defined. Both Tournier et al [79] and Lazar et al [76] test streamline tractography in synthetic tensor fields. Lazar finds that interpolation decreased the divergence of streamlines from an ideal path, except where the tensor field is divergent (where tensors neighbouring the ideal path pointed away from the path), in which case FACT tracked closer to the ideal path. Tensorline tracking generally underestimates tract curvature, but this bias towards low curvature (which is stronger when the tensor is oblate) stabilises tensorline trajectories in regions of crossing fibres. Other studies have used animal models such as the macaque monkey [80, 81], and found agreement in the gross anatomy between histological studies and the tractography.

Stieltjes et al [69], tracked streamlines in the human brainstem of six healthy subjects, using Conturo's two-region idea, an anisotropy threshold, and a curvature threshold. A group of raters defined the ROIs and assessed the similarity of the resulting tracts. Stieltjes concluded that the anisotropy threshold was the most effective at producing reproducible tracts. The recommended parameters were an anisotropy threshold of $f = 0.25$, and an a threshold allowing a maximum of 40 degrees curvature between adjacent voxels on the streamline path.

The need for restrictive thresholds to achieve reproducibility show the limitations of deterministic tractography. One of the biggest problems is that no estimate of the confidence in the streamline path is given, so a spurious streamline carries as much weight as any other.

5.3 Probabilistic streamline tractography

Probabilistic methods estimate the connection probability between two points A and B, given the set of fibre orientation measurements in the image. Koch et al [82] simulate the diffusion of a water molecule, which takes a random walk with probabilities of each walk step defined by the local molecular displacement density, p, for which they use the Gaussian model. A particle takes a random walk from a manually defined start point, with each step sampled from the local p. This process is repeated many times and the the probability of connection to a voxel is the fraction of particles that pass through the voxel during the random walk. While this method produces higher probabilities of connection along the path of white matter, since particles are more likely to move along fibres than across them, it assumes that the probability density of directions of molecular displacements along $\mathbf{x}$ is equivalent to the probability density of the local fibre orientation, which in general is not the case. In bundles of fibres with high anisotropy, the probability of a random walk moving perpendicular to $\mathbf{e}_1$ is much higher than the probability that the fibre bundle is oriented perpendicular to $\mathbf{e}_1$. Only at very low anisotropy do the PDF of directions of molecular displacement and the fibre-orientation PDF become comparable [12].

The Probabilistic Index of Connectivity (PICo), which was introduced independently by Parker et al [83] and Lazar and Alexander [84], and developed in more detail by Parker et al [12], defines a probability density function (PDF) P of the fibre orientation $\mathbf{x}$ in each voxel, in contrast to the PDF on molecular displacements used by Koch. The probability of connection between A and B is the integral over all possible paths that connect A and B. This integral is too complex to solve analytically, so we find the connection probabilities by generating "probabilistic streamlines". Starting from a seed point at A, we draw a sample orientation $\mathbf{x}$ according to the local PDF, and move a small step along this orientation. The tracking continues until a stopping criteria is met, and the streamline connects A to all points along its length. Behrens [85, p.69-70] shows that as the step size tends to zero, the probabilistic streamlines are samples from the integral over all possible paths. Note that the integral is over all possible paths of a single tract, and the "connection probability" represents the belief that a single coherently-organized bundle of white matter fibres traces a path from A to B. The connection probability cannot be interpreted as a distribution of fibres: it is wrong to suggest that a connection probability of 0.5 means that half of the fibres in the voxel A connect to B.

In the limit of no uncertainty, $P(\mathbf{x})$ is a delta function, which is 1 if $\mathbf{x}$ is aligned with $\mathbf{e}_1$ and zero otherwise, and the probabilistic streamlines follow the maximum likelihood estimate of the fibre orientation in each voxel. When there is uncertainty in the data, probabilistic streamlines from the same seed disperse according to the uncertainty in the local fibre orientations along the path of the streamline. We generate a large set of N probabilistic streamlines, and the probability of connection between the seed point and a voxel v is the fraction of probabilistic streamlines that pass through voxel v. A similar method was proposed by Behrens et al [13]. The methods differ in the definition of the voxel PDF: PICo and related algorithms fit a model of the PDF, while Behrens uses a Bayesian approach to sample directly from the fibre-orientation PDF in each voxel. We discuss both methods separately in the following subsections.

The stopping criteria for probabilistic streamlines are less restrictive than for deterministic approaches, since streamlines that enter regions with high uncertainty will disperse widely and give low connection probabilities. Behrens et al use no anisotropy threshold and a curvature threshold of 80 degrees within a voxel, to prevent streamlines from looping back into fibre paths that they had already tracked. Parker et al likewise use no anisotropy threshold but a slightly more restrictive curvature threshold, limiting curvature to 70 degrees.

5.3.1 Model-based PDFs

The advantage of the PICo model-based approach over the Bayesian method proposed by Behrens et al [13] is that the PDF parameters are pre-computed once per imaging scheme, which requires approximately five hours of computation time on a standard PC for single-fibre calibration, while the calibration method for the Bayesian method must be carried out once per subject, and requires approximate 24 hours of computation time. The advantage of the Bayesian method is that it does not assume any model of the fibre-orientation PDF, while PICo must both assume a model of the PDF and correlate its variance to some parameters of the diffusion.

In PICo, the peak orientation of the PDF is estimated from the data. We assume that the measured fibre orientation μ is an unbiased estimate of the true fibre orientation $\mathbf{x}$ from the

PDF. It remains to determine the variance of the PDF from the data.

Parker et al propose models that relate the fractional anisotropy f of the diffusion tensor to the concentration of the PDF. The simplest model assumes that the fibre orientation has rotational symmetry about the peak. This is true when the diffusion has rotational symmetry about the principal direction. Pierpaoli et al [6] suggest that this is the case in many parts of the brain.

Fig. 5.1 shows the coordinate system that Parker et al use to define the PDFs. The tensor principal direction, $\mathbf{e}_1$, has polar coordinates (θ, ϕ) (a). In the coordinate frame of the tensor, $\mathbf{e}_1 = z'$, and the probability of the true fibre orientation being $\mathbf{x}$ is a function of $\delta\theta'$ and $\delta\phi'$.

With rotational symmetry, the PDF depends only on the angle $\delta\theta'$. Any axis found by rotating $\mathbf{x}$ about $\mathbf{e}_1$ by an angle $\delta\phi'$ has the same probability density.

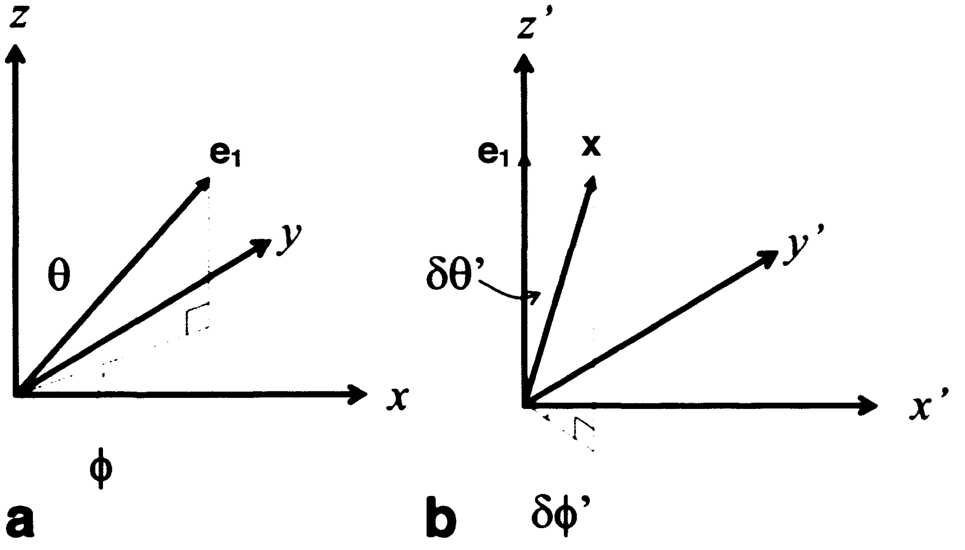


Figure 5.1: The coordinate system for PICo from Parker et al [12].

In the initial PICo paper, Parker et al [12] model the uncertainty in symmetric tensors with a one-dimensional Gaussian distribution $N(0, \sigma)$ of $\delta\theta'$, subject to the constraint that $\delta\theta' \leq \frac{\pi}{2}$. To obtain a fibre-orientation estimate from this PDF, we sample $\delta\theta'$ from the Gaussian distribution and $\delta\phi'$ from a uniform distribution $U[0, 2\pi]$. Parker relates the standard deviation σ of the distribution to f by a sigmoid function. While the model has the desired result of decreasing uncertainty as f increases, the sigmoid relationship between f and uncertainty is a heuristic choice.

Pierpaoli et al [6] found that the diffusion is not symmetric about the principal direction of the tensor in some areas of white matter, including the pyramidal tract and the internal capsule.

In these regions, the uncertainty in fibre orientation depends on $\delta\theta'$ and $\delta\phi'$. Parker et al use a sigmoid model where the standard deviation of the distribution of $\delta\theta'$, $N(0,\sigma)$, is dependent on the sampled $\delta\phi'$. The standard deviation σ is proportional to $\mathbf{e}^T\mathbf{D}\mathbf{e}$, where $\mathbf{e}$ is the vector with $\delta\theta' = \frac{\pi}{2}$ and $\delta\phi' = U[0, 2\pi]$. The axes sampled from this distribution are less concentrated along $\mathbf{e}_2$ than along $\mathbf{e}_3$.

In later work, Parker and Alexander [56] use a modified form of the Gaussian PDF as a model of the uncertainty in $\mathbf{e}_1$ caused by image noise. The diffusion is assumed to have circular symmetry about $\mathbf{e}_1$. Parker and Alexander find empirically the relationship between f and the concentration of the PDF by emulating the scanner sequence used to acquire the data. Given a Gaussian p with a diffusion tensor of anisotropy f, they synthesise many sets of measurements according to Eq. 3.4. Each trial gives an estimate of $\mathbf{e}_1$ and therefore a sample of $\delta\theta'$. They fit the value of σ from $N(0,\sigma)$ to the samples of $\delta\theta'$. This procedure is repeated over a range of f, to produce an estimate of σ as a function of f. Lu et al [86] propose a calibrated model using perturbation theory to describe the error in the diffusion tensor. This model also allows for elliptical contours of the PDF.

Models for multiple fibres

Parker and Alexander [56] extend the single-fibre PICo PDFs to the Gaussian compartment model with $n = 2$. They synthesise noisy data from the test function $p = 0.5[G(\mathbf{D}_1, t) + G(\mathbf{D}_2, t)]$, where $\mathbf{D}_1$ and $\mathbf{D}_2$ are cylindrically symmetric diffusion tensors with identical anisotropy, and the principal directions of $\mathbf{D}_1$ and $\mathbf{D}_2$ are orthogonal. They then fit a fourth-order spherical harmonic series to the data, and re-sample from the series to obtain a larger number of measurements, with inherent smoothing from the spherical harmonic model. They fit the parameters of the two diffusion tensors to this resampled data, and extract the principal directions. This process yields a set of fibre orientations, one from each tensor at each trial, which they separate into two even groups, one for each fibre population. They then calculate $\delta\theta'$ for the axes in each group, and fit the parameter σ of the Gaussian distribution to the samples of $\delta\theta'$. This procedure is repeated over a range of f. Parker and Alexander plot σ against f and fit a biexponential curve to this plot to give σ as a function of f.

They report that the value of σ in each compartment depends only on f of the diffusion tensor. The anisotropy of the other compartment, and the angle between the fibres, does not affect the uncertainty. This is surprising because the diffusion-weighted measurements do depend on both compartments. The spherical harmonic resampling may smooth out these differences, but in doing so they may introduce errors to the fibre-orientation estimates.

The general two-Gaussian model is $p = \alpha G(\mathbf{D}_1, t) + (1-\alpha)G(\mathbf{D}_2, t)$. Parker and Alexander do not claim that the value of σ is independent of α. We expect that α does affect the uncertainty, with larger compartments having lower uncertainty than smaller ones.

5.3.2 Limitations of the model-based PDFs

The model-based calibration approach to PICo has some inherent limitations. Firstly, the approach considers only sources of uncertainty that can be modelled during the calibration process. Often, only imaging noise is modelled. Other sources of uncertainty, such as patient motion or eddy-current distortions, are difficult or impossible to model and are not accounted for.

The calibration process gives an estimate of the uncertainty in fibre orientation assuming p is Gaussian, with a known diffusion tensor. In the brain, p is not necessarily Gaussian, and the diffusion tensor is not known but estimated from noisy measurements. Noise generally causes a positive bias in anisotropy [87], and since anisotropy is inversely related to uncertainty in the fibre orientation, the concentration of calibrated PDFs is generally overestimated. Furthermore, the model itself may not be a good approximation of the true uncertainty in the fibre orientation, for example it may assume circular contours of the PDF, even if the true contours are elliptical.

5.3.3 Bayesian methods

Behrens et al [13] use a Markov Chain Monte Carlo (MCMC) [88] method to sample the fibre-orientation PDF. Markov Chain Monte Carlo provides samples from the PDF $P(\omega \mid Y)$ of possible parameters ω of some model of p given the observed data Y. The fibre orientation PDF is a function of all the parameters in ω. According to Bayes' theorem:

$$P(\omega \mid Y) = \frac{P(Y \mid \omega)P(\omega)}{P(Y)} \tag{5.7}$$

where the denominator in Eq. 5.7 is the integral over all parameter vectors in the parameter space Ω:

$$P(Y) = \int_{\Omega} P(Y \mid \omega)P(\omega)\mathrm{d}\omega \tag{5.8}$$

$P(Y|\omega)$ is the probability of the data given the model and $P(\omega)$ is the prior probability of ω. If we do not have any prior information about ω, i.e. all information that we have about ω comes from Y, then $P(\omega)$ is uniform. Behrens et al use non-informative prior distributions but ensure positivity where sensible, for example $A(\mathbf{0})$ must not be negative.

The MCMC method provides samples from $P(\omega \mid Y)$ without computing $P(Y)$, which cannot be done analytically. The probability of the data given the parameters is the joint probability of the diffusion-weighted measurements y_i,

$$P(Y \mid \omega) = \prod_i^n P(y_i \mid \omega) \tag{5.9}$$

$$P(y_i \mid \omega) = N(\mu_i, \sigma) \tag{5.10}$$

where N is the normal distribution with variance σ^2, and mean μ_i, and μ_i is the predicted value of y_i given ω. Behrens et al propose two models of p: a partial volume model and the standard Gaussian model.

For the Gaussian model, the predicted measurement is $\mu_i = A(\mathbf{0}) \exp(-b_i \mathbf{r}_i^{\mathrm{T}} \mathbf{D} \mathbf{r}_i)$, where b_i is the b-value for measurement i. There are eight parameters in ω for this model, six for the diffusion tensor, $A(\mathbf{0})$ and σ.

The partial volume model contains one anisotropic compartment with diffusion only in one direction, and one isotropic Gaussian compartment. The predicted value of each measurement is

$$\mu_i = A(\mathbf{0})[(1-\nu)\exp(-b_i d) + \nu \exp(-b_i d \mathbf{r}_i^{\mathrm{T}} \mathbf{R} \mathbf{A} \mathbf{R}^{\mathrm{T}} \mathbf{r}_i)]\,, \tag{5.11}$$

where d is the diffusion coefficient, ν is the fraction of the voxel occupied by the isotropic compartment, $\mathbf{A}$ = diag[100] and $\mathbf{R}$ rotates $\mathbf{A}$ to the orientation (θ, ϕ). This model has six parameters: θ, ϕ, ν, d, $A(\mathbf{0})$, and σ. This model is forced to represent crossing fibres as uncertainty in a single fibre orientation, while the tensor model can account for the data by varying the shape of the tensor as well as its orientation. The partial volume model therefore gives greater uncertainty in the fibre orientation in regions of crossing fibres, which is desirable for probabilistic tractography.

Friman and Westin [89] present a modified version of the Bayesian framework. In the diffusion tensor model, they constrain the tensor to be cylindrically symmetric and use a prior distribution for the tensor orientation that gives higher likelihood to low curvature of the streamlines. They also fix the eigenvalues of the tensor and the parameters $A(\mathbf{0})$ and σ at their maximum likelihood estimate. This simplifies the calculation of $P(\omega \mid Y)$ and allows them to sample from the fibre-orientation PDF without using a MCMC method, however it is unclear what effect the restrictive priors have on the fibre-orientation PDF.

Multiple fibres

Hosey et al [90] present a two-fibre PDF that combines ideas from the diffusion tensor model and the partial volume model. For a single fibre:

$$\mu_i = A(\mathbf{0})[(1-\nu)\exp(-b_i d) + \nu \exp(-\mathbf{r}_i^{\mathrm{T}} \mathbf{D}_1 \mathbf{r}_i) \tag{5.12}$$

where $\mathbf{D}$ is a cylindrically-symmetric tensor. This extends naturally to two fibres:

$$\mu_i = A(\mathbf{0})[(1 - \nu_1 - \nu_2)\exp(-b_i d) + \nu_1 \exp(-\mathbf{r}_i^{\mathrm{T}}\mathbf{D}_1\mathbf{r}_i) + \nu_2 \exp(-\mathbf{r}_i^{\mathrm{T}}\mathbf{D}_2\mathbf{r}_i)]\,, \qquad (5.13)$$

Hosey et al sample the parameters of the one-fibre model M_1 and the two-fibre model M_2 by MCMC. In addition to using these samples for tractography, they calculate the likelihood of the data $P(Y \mid M_1)$ and $P(Y \mid M_2)$ for voxel classification. They use the two-fibre model in a voxel when the statistic $P(Y \mid M_2)/P(Y \mid M_1)$ is above a threshold [88], otherwise they use the one-fibre model.

5.4 Tractography without streamlines

Other tractography algorithms, such as those based on fast marching methods, evolve a reconstructed pathway from a seed point but do not define streamlines. Fast Marching Tractography [91] advances a front along the voxel boundaries in all directions. The front advances fastest where there is good coherence in the orientation of primary eigenvectors (as in white matter). Tournier et al [92] present a similar method that does not advance the front in whole voxel steps. The front expands from a seed point by defining "child points" spread in a cone about the local tensor $\mathbf{e}_1$. Each child point is assigned an "index of connectivity" from an orientation density function, which is a function of the local diffusion tensor's orientation and anisotropy, such that directions close to $\mathbf{e}_1$ of a highly anisotropic tensor have a high index of connectivity. Zhang et al [93] compute the evolution of a front by simulating the diffusion of water. The process begins with a unit concentration C of water in a seed voxel, with zero concentration elsewhere. The front evolves according to the diffusion equation:

$$\frac{\partial C}{\partial t} = \nabla \cdot (\mathbf{D}\nabla C)\,, \qquad (5.14)$$

where $\mathbf{D}$ is the diffusion tensor. The front expands to contain all voxels with a nonzero concentration of water at time t.

Other techniques treat tractography as an inverse problem, such as the "spaghetti energy" model of Poupon et al [94]. This method defines an energy related to curvature, in analogy to the thermal energy that must be transferred to spaghetti to make it flexible. Highly curved paths have the highest energy. The authors explain the method in terms of Bayes' theorem: the optimal streamline path maximises $P(D \mid T)$, where $P(D)$ is the prior probability of a set of directions D that define a fibre path, and $P(T)$ is the prior probability of the tensor data T. By Bayes' rule,

$$P(D \mid T) = \frac{P(T \mid D)P(D)}{P(T)}\,; \qquad (5.15)$$

the spaghetti energy alters $P(D)$ to favour low curvature paths, and $p(T \mid D)$ is proportional to the diffusivity along the directions D, i.e. the method maximises the alignment of D with directions of maximum diffusivity, subject to the regularization term that favours low curvature.

5.5 Bootstrap methods in tractography

The bootstrap method [17] is a technique to improve estimates of uncertainty from small data sets. The source data in each voxel is the set of $N \times R$ measurements $A(\mathbf{q}_i)_r$, which contains R repeated measurements at each of N wavenumbers. This data is sampled with replacement, where each sample consists of N measurements, one at each $\mathbf{q}_i$, and the measurement at each wavenumber is chosen at random from one of the R possibilities. Therefore, we first construct a vector $\boldsymbol{\tau}$ of N elements, with each element drawn at random from the uniform distribution $U[1, R]$. Then for each wavenumber $\mathbf{q}_i$, we select the measurement from acquisition $\boldsymbol{\tau}(i)$, such that the complete bootstrap sample is $[A(\mathbf{q}_1)_{\boldsymbol{\tau}(1)}, A(\mathbf{q}_2)_{\boldsymbol{\tau}(2)} \ldots A(\mathbf{q}_N)_{\boldsymbol{\tau}(N)}]$. The result of bootstrap resampling is a set of different combinations of the measurements at each wavenumber. We can therefore assess the uncertainty in derived measures (such as the diffusion tensor) by calculating them for all of the bootstrap samples and applying standard statistical methods.

The bootstrap sample value at each wavenumber $\mathbf{q}_i$ is drawn from a small set of actual measurements $A(\mathbf{q}_i)_1 \ldots A(\mathbf{q}_i)_R$. These measurements are themselves random samples from an underlying distribution of possible values. This is both a strength and potential weakness of the technique. It is a strength because the uncertainty is estimated directly from the data, without making any assumptions about how the data should be distributed. However, any inferences derived from bootstrapping depend on how well the set of R measurements represents the underlying distribution from which they are sampled, which depends strongly on the size of R. Bootstrapping must therefore never be thought of as a substitute for acquiring more data, since additional acquisitions would yield additional samples of $A(\mathbf{q}_i)$ from the *true* distribution, while bootstrap resampling draws from the restricted set of R measurements.

Jones [95] uses bootstrapping to estimate the uncertainty in fibre-orientation estimates of DT-MRI in the brain. Jones uses two separate DW-MR images acquired from a volunteer in a single scanner session. The volunteer was immobilised, so the resulting images were in near perfect spatial alignment. Jones creates 1000 bootstrap samples of the data in each voxel. The signal along each diffusion-weighting gradient is chosen randomly from either the first or second DW-MR image. Jones then fits the diffusion tensor to the bootstrap samples, and calculates the dispersion of the fibre-orientation estimates. Jones visualises these results using a “cone of uncertainty” showing the 95% confidence interval on $\mathbf{e}_1$. The uncertainty is large in regions of fibre crossings, and fairly small in highly anisotropic white matter regions (the 95% confidence interval is 3.5° in the body of the corpus callosum). This quantification of uncertainty is important for probabilistic tractography, however Jones' results are specific to his imaging parameters. Also, Jones uses only two images to form the bootstrap samples, which tends to underestimate the uncertainty [96]. Behrens et al [13] use the same procedure to validate their MCMC tractography algorithm with the partial volume model. The cone of uncertainty is similar, with a mean difference over the whole brain of 15%. It is not clear whether the difference is due to the underestimation of uncertainty in the bootstrap, errors in the partial volume model, or both.

The bootstrap method has also been applied to tractography. Lazar and Alexander [97] study the concentration of fibre orientations derived from bootstrap resampling of eight repeated acquisitions. They conclude that the concentration of fibre orientations from bootstrap resampling is similar to that predicted by their noise-based calibration approach [76] in homogeneous white matter fibres such as the corpus callosum, and present images comparing probabilistic streamlines from bootstrap resampling and from the PDF model. They find that the bootstrap results are generally consistent with a model-based approach, except where there was evidence of branching in the fibre bundles. Jones and Pierpaoli [98] also use the bootstrap method to track probabilistic streamlines. The fibre orientations at each iteration of the tracking process come from bootstrap resampling of the data, rather than samples from a PDF. They conclude that the uncertainty in a fibre pathway depends not only on the local uncertainty, which determines the probability of the streamline following an incorrect path, but also on the architecture of tissue surrounding the fibre being tracked, because that determines what happens to streamlines that do follow the wrong path. Streamlines in fibre bundles that adjoin grey matter are less likely to produce false positive connections than streamlines in a bundle of fibres close to other white matter tracts, since streamlines that reach grey matter disperse quickly and are terminated by most tracking algorithms, whereas streamlines that jump into neighbouring tracts do not disperse and continue tracking.

5.5.1 Limitations of bootstrap methods

Bootstrap resampling systematically underestimates the true uncertainty in the data [99], and this underestimation is a function of the number of measurements. Increasing the number of repeats reduces the correlation between bootstrap samples and gives a better estimate of the true uncertainty. O'Gorman and Jones [96] investigate in synthetic data how the number of repeated measurements affect the estimated uncertainty in the principal direction of diffusion tensors in DT-MRI. They find that with eight repeats, the size of the 95% cone of uncertainty is underestimated by approximately 0.5 degrees. Their results suggest that the bootstrap with eight repeats is a good approximation of the uncertainty in diffusion tensor orientations. They also find that approximately 600 bootstrap samples are sufficient to estimate the uncertainty; increasing the number of bootstrap samples (in contrast to increasing the number of actual data measurements) does not improve the results.

5.6 Conclusions

We have reviewed streamline tractography and its use in a probabilistic framework. The PICo method uses the anisotropy of diffusion, calculated from the diffusion tensor or tensors in each voxel, to predict the parameters of a fibre-orientation PDF in each voxel given a specific MR acquisition scheme. The calibration step, which maps tensor anisotropy to uncertainty, can be performed once per imaging scheme, assuming that the SNR in white matter is constant, which makes PICo less computationally expensive than Bayesian methods, where the PDF is calculated directly for each subject.

PICo assumes that the parameters of the diffusion tensor or tensors in a voxel correlate strongly with the uncertainty, and that the models of the fibre-orientation PDF are a good approximation of the distribution of fibre-orientation estimates observed in the calibration process. Bootstrap estimates of the uncertainty provide one way of testing these assumptions, but bootstrapping tends to underestimate the uncertainty, so we must also test the effects of bootstrapping on the parameters of the fibre-orientation PDFs.

The fibre-orientation PDFs used by Parker and Alexander [56] assume cylindrically symmetric diffusion tensors, where the second and third eigenvalues of the diffusion tensor are equal. The PDF that models uncertainty in asymmetric tensors in [12] is calibrated heuristically, while the symmetric PDF is calibrated directly from experiments in synthetic data [56, 100]. In the following chapters, we propose new models of the fibre-orientation PDF that model uncertainty in asymmetric tensors, and we implement a calibration method to give the PDF parameters as a function of tensor shape. We then extend the models to deal with voxels containing two distinct fibre populations. We compare the uncertainty predicted by the new models to that predicted by bootstrap resampling of a human brain data set.

Chapter 6

Statistical distributions for axial data

The models of the fibre-orientation PDF that we evaluate in this thesis make use of standard statistical distributions specifically designed for spherical data. Existing applications of these models include astronomy, for the analysis of the source orientation of cosmic rays [101] and geology, for analysis of paleomagnetic data such as the orientation of ferromagnetic domains within rocks [102]. This chapter introduces the distributions that we use as models for the fibre orientation PDF in later chapters. These distributions model axial data, so $p(\mathbf{x}) = p(-\mathbf{x})$ for any vector $\mathbf{x}$ on the unit sphere. They are therefore preferable to directional distributions, for which $p(\mathbf{x}) \neq p(-\mathbf{x})$, for modelling the fibre-orientation PDF. The PDFs we present here generalise to higher dimensional spaces as detailed by Mardia and Jupp [68] but we concentrate on models for three-dimensional spheres.

6.1 Watson Distribution

The Watson distribution has the PDF

$$p(\mathbf{x}) = W(\mathbf{x}; \boldsymbol{\mu}, \kappa) = M\left(\frac{1}{2}, \frac{3}{2}, \kappa\right)^{-1} \exp\left[\kappa(\boldsymbol{\mu} \cdot \mathbf{x})^2\right], \tag{6.1}$$

where the unit vector $\boldsymbol{\mu}$ is the mean orientation, the scalar parameter κ is the concentration about $\boldsymbol{\mu}$, and M is the confluent hypergeometric function of the first kind [68, p. 181] (see Appendix B). The Watson distribution is rotationally symmetric about $\boldsymbol{\mu}$, and is a good candidate model of the DT-MRI fibre-orientation estimate when the tensor has rotational symmetry.

For $\kappa > 0$, the distribution is "bipolar" and has maxima at $\pm\boldsymbol{\mu}$. As κ increases, the distribution becomes more concentrated about $\boldsymbol{\mu}$. For $\kappa < 0$, the distribution is a "girdle distribution" concentrated around the great circle orthogonal to $\boldsymbol{\mu}$. Concentration of the girdle distribution increases as $\kappa \to -\infty$.

For prolate, symmetric tensors, where $\lambda_1 \gg \lambda_2 \approx \lambda_3$, the best estimate of the fibre orientation is $\mathbf{e}_1$, and the probability of an error $(\delta\theta', \delta\phi')$ (in the notation of Parker et al [12]) depends only on $\delta\theta'$. Repeated measurements of $\mathbf{e}_1$ would produce two equal and opposite clusters of axes centred on the true axis of $\mathbf{e}_1$. A bipolar Watson distribution with $\boldsymbol{\mu} = \mathbf{e}_1$ gives a probability $p(\delta\theta') \propto \exp[\kappa \cos^2(\delta\theta')]$, with no dependence on $\delta\phi'$.

When two identical bundles of fibres cross within a single voxel, the tensor becomes increasingly oblate with increasing angular separation of the fibre bundles. When the bundles cross at right angles, the tensor is completely disk-shaped, with $\lambda_1 \approx \lambda_2 \gg \lambda_3$. Further measurements of $\mathbf{e}_1$ would be equally likely to lie anywhere in the plane of the fibre crossing, but they would be unlikely to be close to $\mathbf{e}_3$, because hindrance of particle mobility is much greater along this axis. A girdle Watson distribution with $\boldsymbol{\mu} = \mathbf{e}_3$ is maximum in the plane perpendicular to $\mathbf{e}_3$, with $p(\mathbf{x}) \propto \exp[\kappa(\mathbf{e}_3 \cdot \mathbf{x})^2]$.

6.1.1 Maximum likelihood estimation of parameters

Mardia [68] derives the maximum likelihood estimates of the Watson parameters given a set of samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$. We shall use these methods later to fit the Watson distribution to sets of axes derived from simulated DW-MRI acquisitions.

The log-likelihood function is

$$\begin{aligned} l(\boldsymbol{\mu}, \kappa; \mathbf{x}_1, \ldots, \mathbf{x}_n) &= \kappa \textstyle\sum_{i=1}^{n} (\mathbf{x}_i^T \boldsymbol{\mu})^2 - n \log M\left(\tfrac{1}{2}, \tfrac{3}{2}, \kappa\right) \\ &= n\left\{\kappa \boldsymbol{\mu}^T \bar{\mathbf{T}} \boldsymbol{\mu} - \log M\left(\tfrac{1}{2}, \tfrac{3}{2}, \kappa\right)\right\}, \end{aligned} \tag{6.2}$$

where $\bar{\mathbf{T}}$ is the scatter matrix of n samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$:

$$\bar{\mathbf{T}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T . \tag{6.3}$$

The maximum likelihood estimate $\hat{\boldsymbol{\mu}}$ of the mean axis comes from the eigenvectors $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ of $\bar{\mathbf{T}}$. If $\kappa > 0$, then l is maximum when $\boldsymbol{\mu} = \mathbf{t}_1$, where $\mathbf{t}_1$ is the eigenvector with the largest eigenvalue. If $\kappa < 0$, then $l(\boldsymbol{\mu}, \kappa; \mathbf{x}_1, \ldots, \mathbf{x}_n)$ is maximised when $\boldsymbol{\mu} = \mathbf{t}_3$, where $\mathbf{t}_3$ corresponds to the smallest eigenvalue.

To find $\hat{\kappa}$, we differentiate equation 6.2 with respect to κ and set $\frac{dl}{d\kappa} = 0$, giving

$$D_3(\hat{\kappa}) = \hat{\boldsymbol{\mu}}^T \bar{\mathbf{T}} \hat{\boldsymbol{\mu}}, \tag{6.4}$$

with

$$D_3(\hat{\kappa}) = \frac{M\left(\frac{3}{2}, \frac{5}{2}, \kappa\right)}{3M\left(\frac{1}{2}, \frac{3}{2}, \kappa\right)} \tag{6.5}$$

since

$$\frac{d}{d\kappa} M(a, b, \kappa) = \frac{a}{b} M(a+1, b+1, \kappa) . \tag{6.6}$$

We solve equation 6.4 for $\hat{\kappa}$ numerically, using the Newton-Raphson method [67, p. 362]. To decide whether to fit a bipolar or girdle distribution we maximise $l(\mathbf{t}_1, \hat{\kappa}_b; \mathbf{x}_1, \ldots, \mathbf{x}_n)$ and

$l(\mathbf{t}_3, \hat{\kappa}_g; \mathbf{x}_1, \ldots, \mathbf{x}_n)$ and choose the mean axis and concentration parameter that produce the greatest l.

6.2 Bingham Distribution

The Bingham distribution provides a model for axial data with or without cylindrical symmetry. The PDF is

$$p(\mathbf{x}) = M(\frac{1}{2}, \frac{3}{2}, \mathbf{A})^{-1} \exp\left(\mathbf{x}^T \mathbf{A} \mathbf{x}\right) , \tag{6.7}$$

where

$$\mathbf{A} = (\boldsymbol{\mu}_3, \boldsymbol{\mu}_2, \boldsymbol{\mu}_1) \begin{pmatrix} \kappa_3 & 0 & 0 \\ 0 & \kappa_2 & 0 \\ 0 & 0 & \kappa_1 \end{pmatrix} (\boldsymbol{\mu}_3, \boldsymbol{\mu}_2, \boldsymbol{\mu}_1)^T , \tag{6.8}$$

and $\kappa_1 \leq \kappa_2 \leq \kappa_3 \leq 0$. The concentration parameters are defined only up to an additive constant, since $\mathbf{A}$ and $\mathbf{A} + c\mathbf{I}$ are parameters of the same distribution for any real c. Conventionally, κ_3 is set to 0, so equation 6.7 simplifies to

$$p(\mathbf{x}) = M(\frac{1}{2}, \frac{3}{2}, \mathbf{A})^{-1} \exp\left(\kappa_1(\boldsymbol{\mu}_1 \cdot \mathbf{x})^2 + \kappa_2(\boldsymbol{\mu}_2 \cdot \mathbf{x})^2\right) . \tag{6.9}$$

When $\kappa_1 = \kappa_2$, the Bingham distribution is equivalent to a bipolar Watson distribution with mean axis $\boldsymbol{\mu}_3$ [68, p. 234]. When $\kappa_2 = 0$, the Bingham distribution is the girdle Watson with polar axis $\boldsymbol{\mu}_1$.

6.2.1 Maximum likelihood estimation of parameters

Mardia gives the following derivation for the parameters of a Bingham distribution. Let $\bar{\mathbf{T}}$ again be the scatter matrix of n samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$, then the log-likelihood of the samples is

$$l(\mathbf{A}, \mathbf{x}_1, \ldots, \mathbf{x}_n) = n \left[\log \operatorname{Tr}(\mathbf{A}\bar{\mathbf{T}}) - \log M(\frac{1}{2}, \frac{3}{2}, \mathbf{A})\right] . \tag{6.10}$$

We express $\mathbf{A}$ and $\bar{\mathbf{T}}$ in terms of their eigen systems, $\mathbf{A} = \mathbf{U}\mathbf{K}\mathbf{U}^{\mathrm{T}}$ and $\bar{\mathbf{T}} = \mathbf{V}\mathbf{L}\mathbf{V}^{\mathrm{T}}$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, $\mathbf{K} = \operatorname{diag}(\kappa_1, \kappa_2, \kappa_3)$, and $\mathbf{L} = \operatorname{diag}(t_1, t_2, t_3)$. The maximum likelihood estimates of $\mathbf{U}$ and $\mathbf{K}$ are $\hat{\mathbf{U}}$ and $\hat{\mathbf{K}}$. The term $\operatorname{Tr}(\mathbf{A}\bar{\mathbf{T}})$ is maximised when $\hat{\mathbf{U}} = \mathbf{V}$ [68, p. 201]. Mardia shows that the eigenvalues of $\bar{\mathbf{T}}$ are related to the concentration parameters κ_i:

$$t_i = \frac{\partial \log M(\frac{1}{2}, \frac{3}{2}, \hat{\mathbf{K}})}{\partial \kappa_i} . \tag{6.11}$$

We use the method presented by Kent [103] to estimate $\hat{\mathbf{K}}$. Kent presents methods to compute the right hand side of equation 6.11, which allows us to solve equation 6.11 for $\hat{\mathbf{K}}$ using the Newton-Raphson method. Kent states that his methods produce a relative error of no more than 0.1% in the fitted parameters, though he does not give any numerical results to validate this claim.

6.3 Angular Central Gaussian Distribution

The Angular Central Gaussian (ACG) distribution was introduced by Tyler [104] as an alternative model to the Bingham distribution. The ACG density function is

$$p(\mathbf{x}) = \mid \Lambda \mid^{-\frac{1}{2}} (\mathbf{x}^T \Lambda^{-1} \mathbf{x})^{-\frac{3}{2}}, \tag{6.12}$$

where Λ is a positive-definite, (3×3) matrix. The ACG distribution is the distribution of directions from a tri-variate Gaussian distribution N with mean $\mathbf{0}$ and covariance matrix Λ: if $\mathbf{y}$ are distributed according to $N(\mathbf{0}, \Lambda)$, then $\mid \mathbf{y} \mid^{-1} \mathbf{y}$ has ACG distribution with matrix parameter Λ. This means that random samples from the ACG distribution may be generated without the use of a rejection method, using the same routines that generate multivariate Gaussian vectors. Another computational advantage of the ACG is that the normalising constant, $\mid A \mid^{\frac{1}{2}}$, is much easier to compute than the hypergeometric constant of the Bingham distribution.

The parameter matrix is defined up to a multiplicative constant, because Λ and $c\Lambda$ describe the same distribution for any nonzero c.

6.3.1 Maximum likelihood estimation of parameters

Tyler shows that the maximum likelihood estimate $\hat{\Lambda}$ is a solution to the equation

$$\hat{\Lambda} = \frac{3}{n} \sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{T}} \hat{\Lambda}^{-1} \mathbf{x}_i . \tag{6.13}$$

Tyler presents a simple iterative algorithm to find $\hat{\Lambda}$. Let $\hat{\Lambda}_0 = \mathbf{I}$, then:

$$\hat{\Lambda}_{k+1} = 3 \left[\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} (\mathbf{x}_i^{\mathrm{T}} \hat{\Lambda}^{-1} \mathbf{x}_i)^{-1} \right] \left[\sum_{i=1}^{n} (\mathbf{x}_i^{\mathrm{T}} \hat{\Lambda}^{-1} \mathbf{x}_i)^{-1} \right]^{-1} . \tag{6.14}$$

As with the Bingham distribution, the principal directions of Λ correspond to those of $\bar{\mathbf{T}}$.

6.4 Conclusions

We have introduced three distributions to use as models of the fibre-orientation PDF. The Watson distribution, which has circular contours on the sphere, is analogous to the Gaussian PDF used by Parker et al [12], except that it can take the girdle form with negative κ in order to model the distribution of $\mathbf{e}_1$ from oblate tensors. The Bingham and ACG distribution both allow elliptical contours on the sphere, which we use to model cases where the distribution of fibre-orientation estimates is not circular about the mean, such as when the diffusion tensor does not have cylindrical symmetry. We present a method to calibrate these distributions for PICo tractography in Ch. 7, and in Ch. 8, we use the distributions to model uncertainty in voxels where we resolve two fibre bundles.

Chapter 7

Models of the single-fibre PDF

The experiments in this chapter evaluate the Watson, Bingham and Angular Central Gaussian models of the fibre-orientation PDF in synthetic data and in the brain. We set out the framework for the comparison of the models and evaluate them in synthetic data. The experiments in synthetic data test which models most closely reproduce the uncertainty in fibre orientations caused by noise. The synthetic data is produced by emulating the scanner sequence used in the acquisition of the brain data. The experiments in brain data test which models most closely reproduce the total uncertainty in fibre orientations, recovered from bootstrap resampling of repeated images of the same subject.

7.1 Calibration of models

The PICo algorithm requires calibration for each imaging scheme. We assume in this calibration that the Gaussian model is a good approximation of the particle displacement density, p. Given a diffusion tensor, the acquisition scheme and an estimate of the variance of the noise, we synthesise a large set of noisy MR acquisitions from a Gaussian p with the specified tensor. We then fit the diffusion tensor to each acquisition, extract the fibre orientations, and estimate the concentration. To calibrate PICo we perform this procedure over a range of tensor shapes and construct a lookup table (LUT) that gives an estimate of the concentration in each voxel of a DT-MR image, given the diffusion tensor in each voxel.

The calibration is specific to the MR acquisition sequence and the type of diffusion tensor fitting routine used to extract the fibre orientations. Chapters 7 and 8 include experiments using synthetic data and a series of eight brain images acquired from a healthy volunteer. We therefore generate synthetic data by emulating the scanner sequence used in the acquisition of the brain data and the PICo PDFs are calibrated for the brain MR images, with tensors fitted using a nonlinear least-squares approach. The fitting routine optimises the parameters of the Cholesky decomposition of the diffusion tensor, as described in Sec. 3.4.1, with the Levenberg-Marquardt algorithm. We explain the details of the brain MR acquisition in Sec. 7.1.1.

7.1.1 The MR acquisition sequence

A series of eight complete acquisitions were made in a single visit to the scanner. We number these images 1 through 8, and the "average brain" is the mean of all 8 images. The subject remained still between the acquisitions, so there is a good spatial correspondence between the eight acquisitions. The data was acquired on a Philips Achieva 3T scanner. The imaging parameters are $\Delta = 0.0284$ s, $\delta = 13.52 \times 10^{-3}$ s (δ and Δ are defined in Fig. 3.2), $| \mathbf{g} | = 0.062 \times 10^{-3}\,\mathrm{T\,m^{-1}}$. These values translate to a "b-value" ($b = | \mathbf{q} |^2\, t$) of $1200\,\mathrm{s\,mm^{-2}}$.

The data was acquired on a 112×128 matrix, reconstructed to 128×128, resulting in an in-plane voxel size of 1.84×10^{-3} m. The slice thickness is 2.1×10^{-3} m. To reduce scanner time, 34 contiguous slices were acquired, which do not cover the entire brain. The acquisition used $t_r = 6$ s and $t_e = 56 \times 10^{-3}$ s, with SENSE [105] parallel imaging (factor 2). A single image with no diffusion weighting, which we call the unweighted image, was acquired, followed by 61 diffusion weighted images with $b = 1200\,\mathrm{s\,mm^{-2}}$. The 61 gradient directions are distributed isotropically on the hemisphere using an electrostatic minimisation algorithm [58].

We use the following procedure to estimate the signal to noise ratio (SNR) in white matter in the unweighted image. We define a region of interest (ROI) containing N voxels in a white matter region. We calculate the mean unweighted signal $M(A^*(\mathbf{0}))$, averaged over all images $i = 1 \ldots 8$, for each voxel in the ROI. The estimated noise in image i is $\epsilon(i) = A_i^*(\mathbf{0}) - M(A^*(\mathbf{0}))$. We calculate the mean signal $M(A^*(\mathbf{0}))$, the mean noise $M(\epsilon)$ and the variance of the noise $\sigma^2(\epsilon)$ over the ROI:

$$M(A) = \sum_{i=1}^{8} \sum_{v=1}^{N} \frac{A_{i,v}(\mathbf{0})}{8N} \tag{7.1}$$

$$M(\epsilon) = \sum_{i=1}^{8} \sum_{v=1}^{N} \frac{\epsilon_v(i)}{8N} \tag{7.2}$$

$$\sigma^2(\epsilon) = \sum_{i=1}^{8} \sum_{v=1}^{N} \frac{[\epsilon_v(i) - M(\epsilon)]^2}{8N - 1}. \tag{7.3}$$

The estimated SNR is then $M(A)/\sigma(\epsilon)$. We estimate the SNR in the corpus callosum of the subject to be 17.

7.1.2 Voxel classification

We segment the brain from the background within the image volume using the Brain Extraction Tool [106]. Within the brain, we fit the diffusion tensor in voxels with Gaussian diffusion (using the same nonlinear fitting routine we use for the calibration of PICo), and a two-Gaussian compartment model in voxels with non-Gaussian diffusion. The methods for fitting the two-Gaussian compartment model to the data are explained in Sec. 8.4. We classify the diffusion in each voxel as Gaussian or non-Gaussian with the spherical harmonic algorithm of Alexander et al [54]. Each voxel is classified as Gaussian (order 0 or 2 spherical harmonic) or non-Gaussian (order 4 spherical harmonic). We average the measurements from the eight acquisitions to enhance signal to noise, and run the spherical harmonic classification algorithm on the average measurements. We choose the threshold for the F-test interactively using a graphical tool included in the Camino package [1]. The voxel classification is the same in all experiments, so that the fibre orientation estimates either come from a single-fibre population, or a two-fibre population. We use this voxel classification for the experiments in the brain in chapters 7 and 8. In synthetic data, we use the ground truth voxel classification.

7.1.3 Data synthesis

We generate synthetic data by emulating the scanner sequence used to acquire the brain data. For each wavenumber, q, the ideal signal $A^{\dagger}(\mathbf{q})$ is the sample of the Fourier transform of a test function p at $\mathbf{q}$. The test function is either Gaussian (Eq. 3.3) or a two-Gaussian compartment model (3.9). The trace of the diffusion tensors in each test function is $2.1 \times 10^{-9}\mathrm{m}^2\,\mathrm{s}^{-1}$, which is the typical trace in brain white matter. The trace in white matter is approximately constant in healthy subjects [6].

We add complex Gaussian noise to the ideal signal and take the modulus:

$$A(\mathbf{q}) = \mid A^{\dagger}(\mathbf{q}) + c \mid , \tag{7.4}$$

where the real and imaginary components of c are taken from the normal distribution $N(0, \sigma)$, and σ is calculated to give a signal to noise ratio of 17 in the unweighted images, which is the noise level we estimate in the brain data.

7.1.4 LUT generation procedure

We construct a LUT for each model that returns the concentration parameters of the PDF given the shape of the diffusion tensor. Given the tensor eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$, we use two parameters to index the LUT: $x = \lambda_1 \lambda_3^{-1}$ and $y = \lambda_2 \lambda_3^{-1}$. The calibration algorithm generates synthetic data and fits the PDF parameters:

```
1. for $x = 1; x \leq x_{\max}; x+ = \delta$
    (a) for $y = 1; y \leq x; y+ = \delta$
        i. Calculate tensor $\mathbf{D}$
        ii. For each trial $i = 1 : T$
```

A. Synthesise noisy data from Gaussian p with tensor $\mathbf{D}$: add noise to ideal signal $A^{\dagger}(\mathbf{q}) = \exp(-t\mathbf{q}^{\mathrm{T}}\mathbf{D}\mathbf{q})$ (Sec. 7.1.3).

B. Fit tensor to noisy data and extract fibre-orientation estimate $\mathbf{x}_i = \mathbf{e}_1$.

(b) Fit the model concentration parameters to $\mathbf{x}_i \ldots \mathbf{x}_T$.

(c) Record the concentration parameters in the LUT.

The computational expense of calibration scales quadratically with the extent of the LUT, defined by $x_{\mathrm{max}} - x_{\mathrm{min}}$, and the resolution, defined by δ. The computation scales linearly with the number of trials, T, used to calculate each set of concentration parameters. In the remainder of this section we examine the range of tensor shapes observed in the brain data and perform some experiments to determine the range, resolution, and number of trials that we require to calibrate PICo for this imaging scheme.

7.1.5 LUT range

The useful extent of the LUT is determined by the range of diffusion anisotropy in the brain. Fig. 7.1 shows a histogram of x in one of the brain volumes. We compute LUTs with $x_{\mathrm{min}} = 1$ and $x_{\mathrm{max}} = 15$. With $x = 15$ and $y = 1$, the tensors have fractional anisotropy of 0.93.

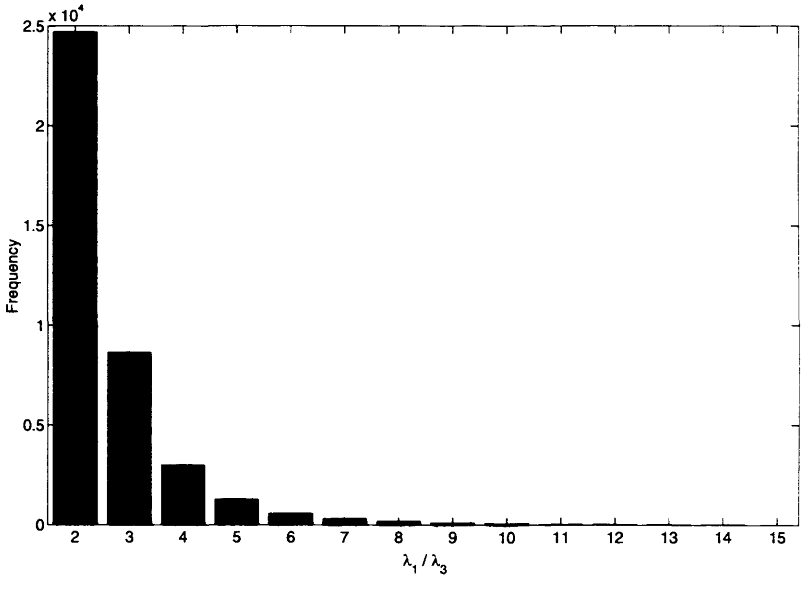


Figure 7.1: Histogram of diffusion tensors in the brain data, binned according to λ_1/λ_3.

The LUT range is also dependent upon the noise level in the image, since the noise conditions limit the maximum diffusion coefficient that may be reliably measured. As the diffusion coefficient along a particular direction increases, the signal (according to Eq. 3.4) decreases exponentially, and once it has decreased sufficiently for the measured signal to be dominated by noise, the diffusion coefficient will be systematically underestimated. This underestimation is more pronounced along fibre bundles, where the diffusion coefficient is high, than across them, where the diffusion coefficient is low, causing anisotropy to be underestimated (assuming that the trace remains constant). Jones and Basser [107] study this effect and conclude that it is unlikely to be significant at the b-value of $1200\,\mathrm{s\,mm^{-2}}$ that is used in the brain data in this thesis.

7.1.6 Experiment (i): number of trials required at each LUT entry

Experiment (i) tests how many fibre-orientation estimates are required for each entry in the LUT to obtain acceptable reproducibility of the concentration parameters. As a baseline we construct a LUT for the Bingham PDF with $T = 10000$ trials per entry, $x_{\max} = 10$, $y_{\max} = 10$, $\delta = 0.2$. We compare the baseline values to those produced from LUTs with $500, 1000, 2000$ and 5000 trials. We also produce another LUT with 10000 trials, which shows the reproducibility of the baseline. Table 7.1 shows the mean absolute and percentage error in the parameters κ_1 and κ_2 over all entries in the LUT.

Trials	mean percentage error		mean absolute error	
	κ_1	κ_2	κ_1	κ_2
500	5.98	25.54	47.18	13.94
1000	3.61	20.72	30.50	9.73
2000	2.79	12.20	23.32	7.26
5000	1.95	6.48	16.30	5.40
10000	1.65	5.53	13.71	4.33

Table 7.1: Errors in the Bingham parameters with increasing numbers of trials in the calibration.

The parameter κ_2 has the largest percentage error but the smallest absolute error. In oblate tensors $|\ \kappa_2\ |$ is much smaller than $|\ \kappa_1\ |$, and in the event that the diffusion tensor is completely oblate, with $\lambda_1 = \lambda_2 \gg \lambda_3$, κ_2 is close to zero. Fig. 7.2 shows the percentage error in κ_2 with 5000 trials at each entry. The percentage error is above 10% only when $x = y$. We conclude from this experiment that 5000 trials are sufficient.

7.1.7 Experiment (ii): LUT resolution

We use bilinear interpolation to estimate the concentration from the LUT given any x and y. Experiment (ii) examines the error introduced by interpolation at various values of the step size δ. We construct a Bingham LUT where $\delta = 0.05$, $x_{\max} = 10$, $y_{\max} = 10$, $T = 5000$, and

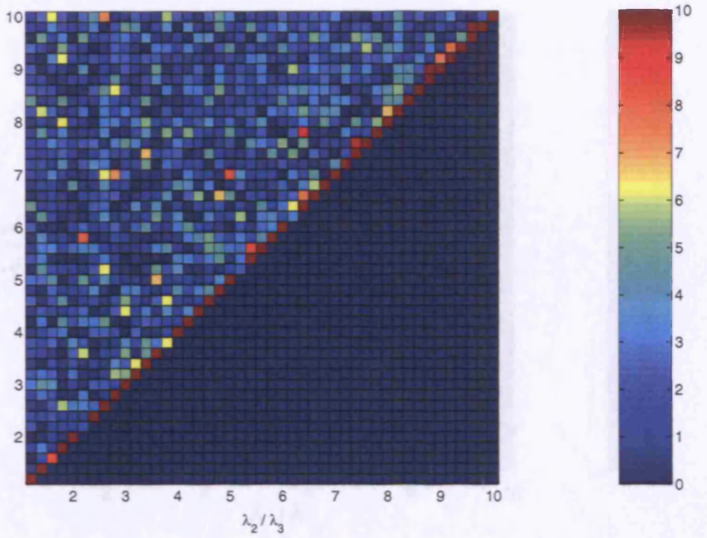


Figure 7.2: Percentage error in κ_2 in experiment (i), with 5000 trials per LUT entry.

sub-sample the table in both dimensions at intervals of $0.1, 0.2, 0.5$, and 1.0. Table 7.2 shows the mean error (over all entries in the subsampled LUT) between the actual LUT values and the predicted values from interpolation of the subsampled tables.

δ	mean percentage error		mean absolute error	
	κ_1	κ_2	κ_1	κ_2
0.1	1.56	4.57	12.73	3.71
0.2	2.04	11.39	16.33	4.67
0.5	2.47	49.93	18.37	6.424
1.0	3.18	203.2	20.31	14.50
2.0	4.50	808.7	23.95	45.47

Table 7.2: Errors in the Bingham parameters in an interpolated LUT with increasing step size.

The mean errors in the LUT with $\delta = 0.1$ are smaller than the errors between successive runs of the LUT generation algorithm with $T = 10000$ in experiment (i). Fig. 7.3 shows the percentage error in κ_1 with a step size of 0.1. Most values are correct within 5% and the error exceeds 10% only when $x = y$. We conclude that a step size of 0.1 and $T = 5000$ trials are reasonable values for the LUT generation. Using these values, the calibration process takes approximately eight hours on a standard 2.4 GHz PC workstation.

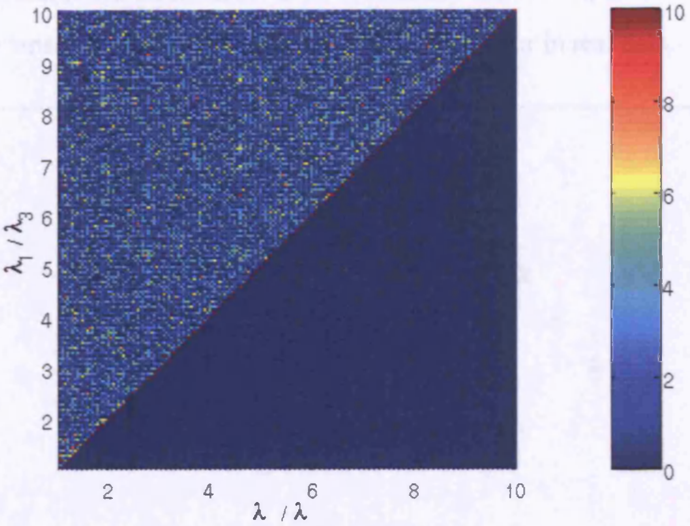


Figure 7.3: Percentage error in κ_1 in experiment (ii), with $\delta = 0.1$.

7.1.8 Experiment (iii): Effect of noisy LUT indices

The LUT gives concentration parameters from a Gaussian test function given the shape parameters of the diffusion tensor x and y. In practice, we have only a noisy estimate of x and y. This causes an error in the concentration assigned to each PDF, even if the LUT has zero errors. Experiment (iii) shows the effect of this error in synthetic data.

We synthesise 1000 sets of noisy measurements from a Gaussian test function with a cylindrically symmetric diffusion tensor $\mathbf{D}_f$ with anisotropy $f = 0.1$. For each set of measurements we fit the noisy diffusion tensor $\mathbf{D}_{fi}$, $1 \leq i \leq 1000$, and calculate the parameters x and y. We then obtain the Bingham parameters κ_1 and κ_2 from the LUT for each noisy tensor $\mathbf{D}_{fi}$, where the LUT has resolution $\delta = 0.1$ and sample size $T = 5000$. We repeat this procedure over a range of anisotropy up to $f = 0.9$.

Fig. 7.4 plots κ_1 against f. The blue circles are κ_1 predicted for the tensor $\mathbf{D}_f$ from each test function. The red squares and error bars show the mean and standard deviation of κ_1 returned from the LUT when the LUT indices x and y are calculated from the noisy $\mathbf{D}_{fi}$. There is little systematic error, but the noise introduces a random error in the concentration. Fig. 7.5 plots the standard deviation of κ_1 and κ_2 from $\mathbf{D}_{fi}$ as a percentage of the value predicted by the LUT for the corresponding $\mathbf{D}_f$. At low anisotropy, the percentage error is high, but it levels off to around 20% at moderate to high anisotropy. This source of error has obvious implications for PICo tractography, as we shall examine in experiments (v) and (xi). We conclude that increasing the PICo calibration time by using $\delta < 0.1$ or $T > 5000$ would not improve tractography for this imaging scheme, because the errors in the LUT are already small compared to those introduced by the error in the tensor eigenvalues that we expect to encounter in real data.

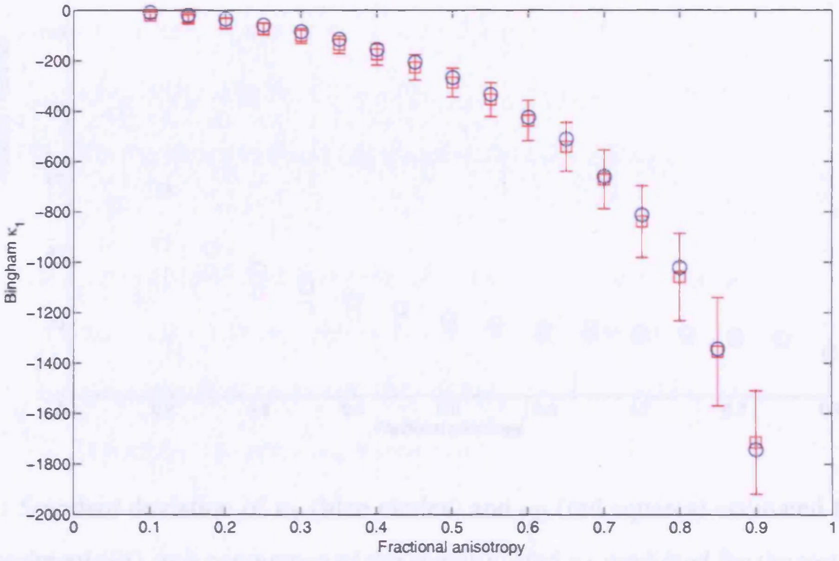


Figure 7.4: Bingham κ_1 against f in experiment (iii). The blue circles are κ_1 predicted for the tensor $\mathbf{D}_f$ from each test function. The red squares and error bars show the mean and standard deviation of κ_1 returned from the LUT when the LUT indices x and y are calculated from the noisy $\mathbf{D}_{fi}$.

7.2 The PICo tracking algorithm

We define the algorithm here as it is used in the tractography experiments. The tracking method is similar to FACT, because streamlines may enter and exit voxels at any point, and follow the voxel fibre orientation without interpolation. In a similar way to Behrens et al [13], we apply no anisotropy threshold, and a curvature threshold that terminates streamlines only if they curve by more than 80 degrees over the length of the longest side of a voxel.

1. Initialise the fibre orientation PDF in each voxel v.

2. connection probability in each voxel is $p_v = 0$.

3. for all iterations $i = 1 : N$ and v in $[-1, 1]$, do

   (a) Start a new streamline at the centre of the seed voxel.

   (b) Generate a new sample $\mathbf{x}_v$ from the fibre orientation PDF in voxel v.

   (c) $\mathbf{x}_v = v\mathbf{x}_v$

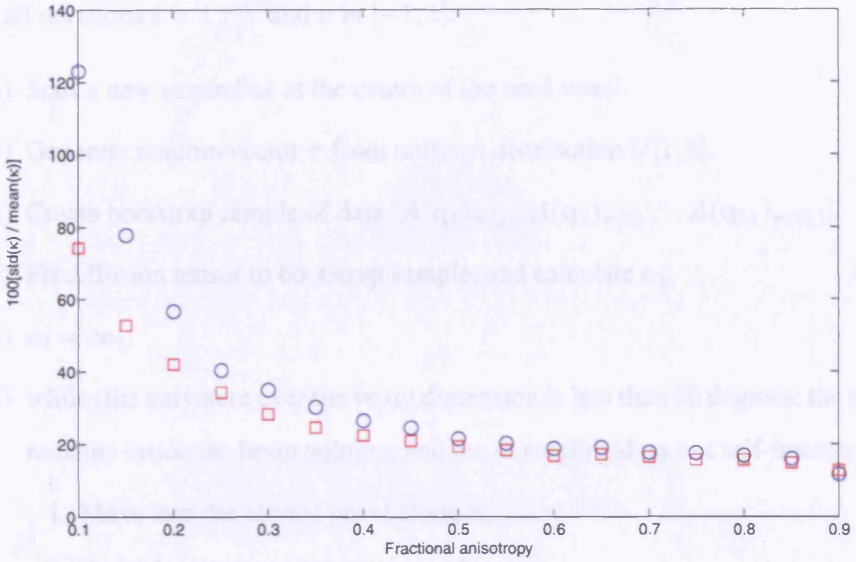


Figure 7.5: Standard deviation of κ_1 (blue circles) and κ_2 (red squares) estimated from noisy data in experiment (iii), as a percentage of the actual κ_1 and κ_2 predicted for the test function.

(d) while (the curvature over the voxel dimension is less than 80 degrees; the streamline remains inside the brain volume; and the streamline does not self-intersect):

    i. Move into the closest voxel along $\mathbf{x}_v$.

    ii. Record current position in streamline.

    iii. Generate a new sample $\mathbf{x}_v$ from the fibre orientation PDF in v.

    iv.

(e) for all voxels v on streamline path

    i. Connection probability $p_v = p_v + \frac{1}{N}$.

4. Output PICo map, which is a 3D image of identical dimensions to the data, where the intensity in each voxel v is p_v.

7.2.1 Bootstrap tracking

The algorithm for producing bootstrap PICo maps is identical, except that there are no models of the fibre orientation PDF. At each iteration, the fibre orientations in each voxel are estimated by fitting the diffusion tensor or two-tensor compartment model to a bootstrap sample (see Sec. 5.5) of the diffusion-weighted measurement in the voxel.

1. Initialise the fibre orientation PDF in each voxel v.
2. connection probability in each voxel is $p_v = 0$.

```
3. for all iterations $i = 1 : N$ and $v$ in $[-1, 1]$.
    (a) Start a new streamline at the centre of the seed voxel.
    (b) Generate random vector $\boldsymbol{\tau}$ from uniform distribution $U[1, 8]$.
    (c) Create bootstrap sample of data $[A(\mathsf{q}_1)_{\boldsymbol{\tau}(1)}, A(\mathsf{q}_2)_{\boldsymbol{\tau}(2)} \ldots A(\mathsf{q}_{61})_{\boldsymbol{\tau}(61)}]$.
    (d) Fit diffusion tensor to bootstrap sample, and calculate $\mathbf{e}_1$.
    (e) $\mathbf{e}_1 = v\mathbf{e}_1$.
    (f) while (the curvature over the voxel dimension is less than 80 degrees; the streamline remains inside the brain volume; and the streamline does not self-intersect):
        i. Move into the closest voxel along $\mathbf{e}_1$.
        ii. Record current position in streamline.
        iii. Generate random vector $\boldsymbol{\tau}$.
        iv. Create bootstrap sample of data in voxel $v$.
        v. Fit diffusion tensor and calculate $\mathbf{e}_1$.
    (g) for all voxels $v$ on streamline path
        i. Connection probability $p_v = p_v + \frac{1}{N}$.
4. Output PICo map, which is a 3D image of identical dimensions to the data, where the intensity in each voxel $v$ is $p_v$.
```

7.3 PICo tractography in synthetic data

The experiments in this section test how well the PDFs model the uncertainty caused by noise in the principal direction of a Gaussian test function. The experiments use the models to calculate PICo probability maps in the synthetic fibre pathway shown in Fig. 7.6. The image dimensions are 80 × 32 × 33 and the voxel dimensions are $2 \times 2 \times 2$ mm^3. We seed streamlines at a single point in the image, and they follow a helical pathway. A streamline seeded in the noise-free image traces out a path of length 145 mm, with a radius of curvature of 46 mm. We carry out each experiment in this section in eight tori, with the fractional anisotropy f of each torus set to a different value between 0.2 and 0.9.

We construct a gold standard synthetic PICo map by tracking streamlines in 5000 identical images. In each image, we add noise to the measurements as described in Sec. 7.1.3, fit the tensors, and then track a streamline from the seed point. The collection of streamlines define a PICo connection probability map, except that the fibre orientations at each iteration are estimated directly from the noisy data and not sampled from a PDF. We measure sum of squared

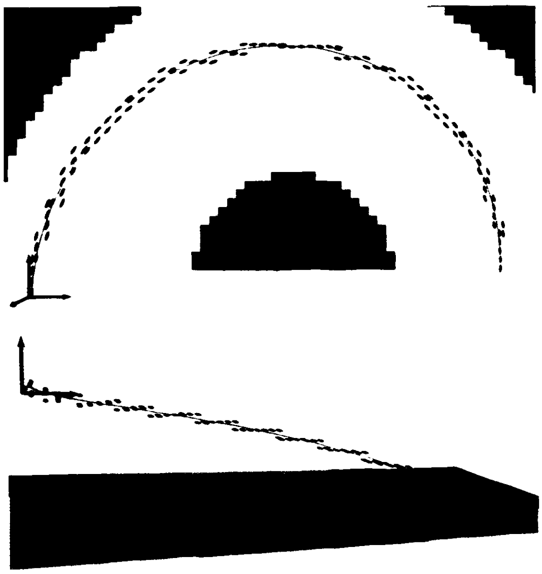


Figure 7.6: Path of a streamline in the synthetic fibre pathway, in the absence of noise. The diffusion tensors in voxels intersected by the streamline are also shown. The streamline is shown from an axial (top) and coronal perspective (bottom).

differences (SSD) S between PICo maps generated using our models (also with $N = 5000$ iterations) and the gold standard. The best model of the fibre-orientation PDF will produce maps with the lowest SSD to the gold standard.

7.3.1 Experiment (iv): PICo maps in noise free data

In experiment (iv), we compute 50 PICo maps with each PDF in the noise-free data, where the orientation and shape of each tensor is not affected by noise. This experiment tests the ability of the PICo PDFs to reproduce the uncertainty in fibre orientations, when the tensor shape and orientation is known.

Figure 7.7(a) shows the sum of squared differences to the gold standard in experiment (iv). The Watson and Bingham models both have smaller errors than the ACG model. Figure 7.7(b) shows the normalised cross correlation between the Watson and Bingham and the gold standard. The correlation is greater than 0.99 over the range of anisotropy, but decreases with increasing anisotropy. We conclude from this result that the Watson and Bingham distributions are accurate models of the actual PDF, though they fit the data less well at high anisotropy. We also conclude that the PICo process is stable after 5000 iterations. The ACG model consistently overestimates the uncertainty. Fig. 7.8 shows an example PICo map from the gold standard, the Watson model and the ACG model. The fibre orientations sampled from the ACG PDF are more dispersed than those from the Watson PDF. The pathway has been rotated to ease visualisation,

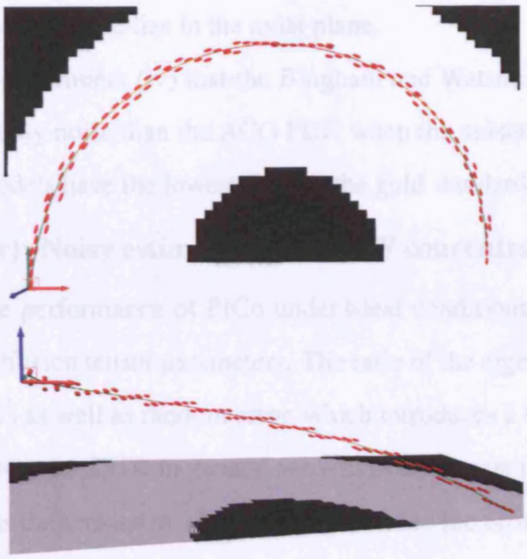


Figure 7.6: Path of a streamline in the synthetic fibre pathway, in the absence of noise. The diffusion tensors in voxels intersected by the streamline are also shown. The streamline is shown from an axial (top) and coronal perspective (bottom).

differences (SSD) S between PICo maps generated using our models (also with $N = 5000$ iterations) and the gold standard. The best model of the fibre-orientation PDF will produce maps with the lowest SSD to the gold standard.

7.3.1 Experiment (iv): PICo maps in noise free data

In experiment (iv), we compute 50 PICo maps with each PDF in the noise-free data, where the orientation and shape of each tensor is not affected by noise. This experiment tests the ability of the PICo PDFs to reproduce the uncertainty in fibre orientations, when the tensor shape and orientation is known.

Figure 7.7(a) shows the sum of squared differences to the gold standard in experiment (iv). The Watson and Bingham models both have smaller errors than the ACG model. Figure 7.7(b) shows the normalised cross correlation between the Watson and Bingham and the gold standard. The correlation is greater than 0.99 over the range of anisotropy, but decreases with increasing anisotropy. We conclude from this result that the Watson and Bingham distributions are accurate models of the actual PDF, though they fit the data less well at high anisotropy. We also conclude that the PICo process is stable after 5000 iterations. The ACG model consistently overestimates the uncertainty. Fig. 7.8 shows an example PICo map from the gold standard, the Watson model and the ACG model. The fibre orientations sampled from the ACG PDF are more dispersed than those from the Watson PDF. The pathway has been rotated to ease visualisation, so that the true path of a streamline lies in the axial plane.

We conclude from experiment (iv) that the Bingham and Watson PDFs are better models of the uncertainty caused by noise than the ACG PDF, when the anisotropy of the test function is known, since these models have the lowest SSD to the gold standard.

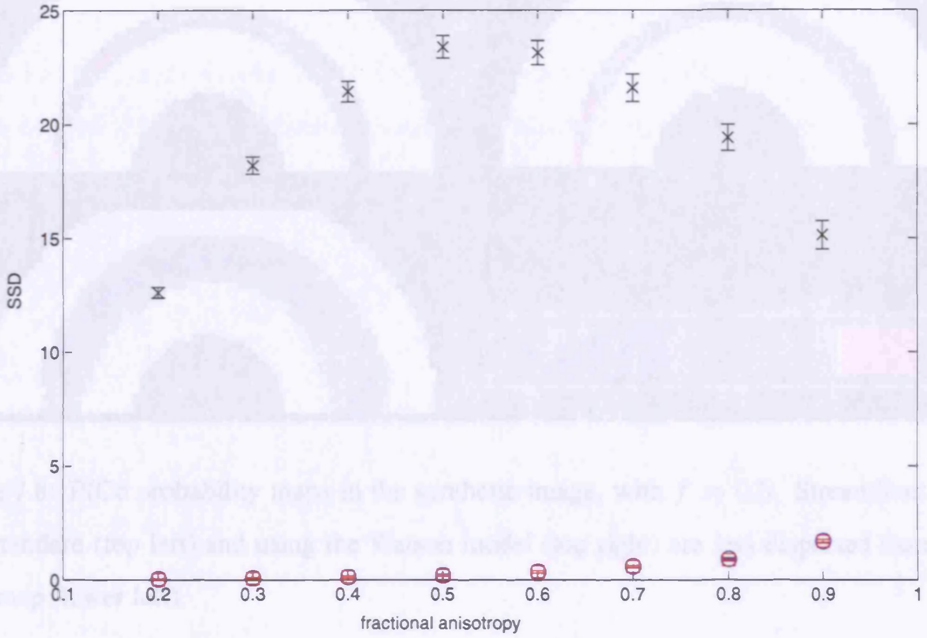
7.3.2 Experiment (v): Noisy estimates of the PDF concentration

Experiment (iv) tests the performance of PICo under ideal conditions. In practice PICo uses noisy estimates of the diffusion tensor parameters. The ratio of the eigenvalues x and y are subject to a positive bias [87] as well as random error, which introduces a bias in the concentration parameters we sample from the LUT: in general we will overestimate the concentration.

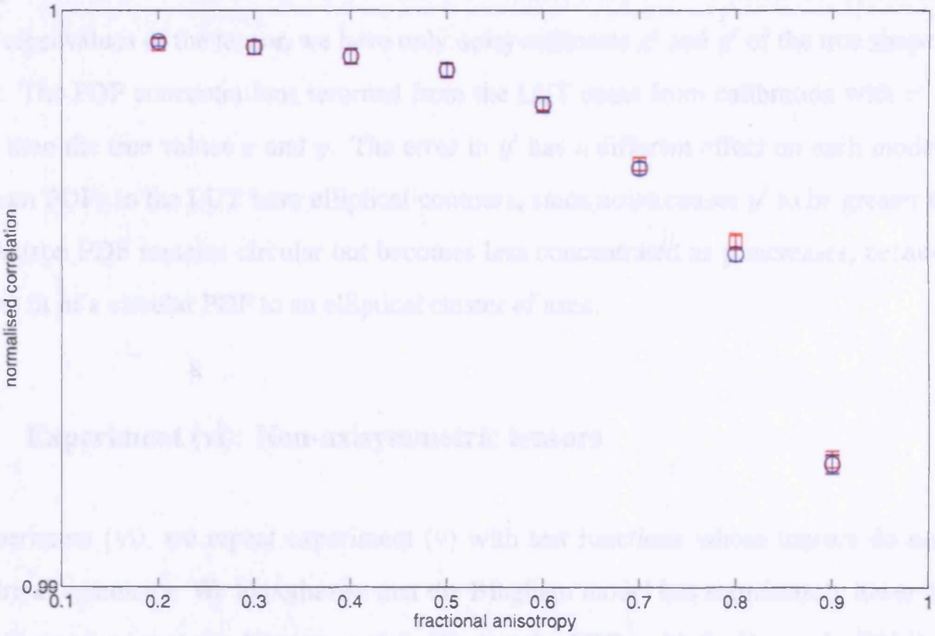
Experiment (v) tests the sensitivity of each PDF model to the error in tensor shape caused by noise. We again compute 50 PICo maps with each PDF. For each PICo map, we add noise to the ideal image and fit the diffusion tensors. The orientation of the PDFs are determined from the noise-free image, as in experiment (iv), but the tensor eigenvalues are determined from the noisy images. Fig. 7.9 shows the SSD to the gold standard. The SSD is similar to the noise-free result (Fig. 7.8), except at $f = 0.2$ and $f = 0.9$. At $f = 0.2$ the SSD of the Watson model is slightly higher, which suggests that it is more sensitive to the error in the LUT index than the Bingham model. At $f = 0.9$, all three models have larger errors. At $f = 0.9$ the fibre orientations in the gold standard are highly concentrated, and a small error in the concentration produces relatively large changes in the connection probabilities.

When we use noisy PDF concentration for PICo in experiment (v) and plot the mean SSD, S, over the 50 PICo maps for each PDF, we can only detect differences between the PDF performances if the differences are much larger than the variance of each PDF's performance across different images. This is the case for the ACG PDF, which has higher SSD than the other PDFs in all 50 PICo maps. However, the differences between the Watson and Bingham PDFs are less clear, because they are small compared to the differences in S over the different images. We can emphasise differences in the performance of each PDF by measuring the relative performance of two PDFs A and B in the same image. We measure $d = S(I_n, A) - S(I_n, B)$ for each image I_n, $1 \leq n \leq 50$. We detect a significant difference between A and B by computing the mean $\bar{d}(A, B)$ and standard error $\sigma_{\bar{d}}(A, B)$; there is a significant difference at the 95% confidence level if $\mid \bar{d}(A, B) \mid \; > \; 1.96\sigma_{\bar{d}}(A, B)$.

Fig. 7.10 shows $\bar{d}(\text{Watson}, \text{Bingham})$ for experiment (v). When this statistic is greater than zero, the SSD for the Bingham model is on average lower than for the Watson model, over the 50 PICo maps. The Bingham model has lower SSD at $f = 0.2, 0.8, 0.9$ than the Watson model, which is interesting because the tensors in the test function are axisymmetric and hence



(a) Sum of squared differences to the gold standard for axisymmetric tensors in experiment (iv).



(b) Correlation to the gold standard for axisymmetric tensors in experiment (iv).

Figure 7.7: Sum of squared differences to the gold standard in experiment (iv). Data points are the mean SSD calculated from 50 different runs of the PICo algorithm. Error bars show the standard deviation. Markers are blue circles for the Watson model, red squares for the Bingham model, and black crosses for the ACG model.

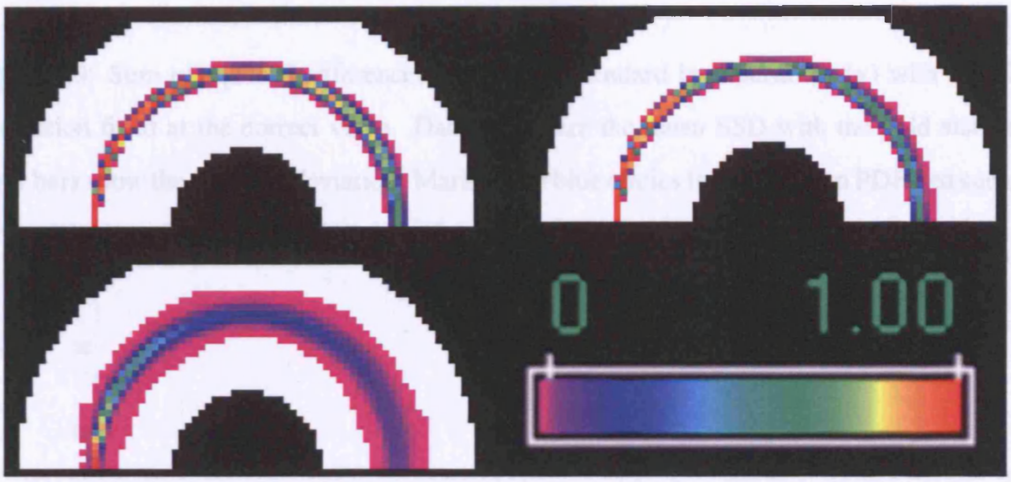


Figure 7.8: PICo probability maps in the synthetic image, with $f = 0.5$. Streamlines in the gold standard (top left) and using the Watson model (top right) are less dispersed than in the ACG map (lower left).

we expect that the true PDF has circular contours. However, since the noise introduces an error in the eigenvalues of the tensor, we have only noisy estimates x' and y' of the true shape of the tensor. The PDF concentrations returned from the LUT come from calibration with x' and y' rather than the true values x and y. The error in y' has a different effect on each model. The Bingham PDFs in the LUT have elliptical contours, since noise causes y' to be greater than 1. The Watson PDF remains circular but becomes less concentrated as y increases, because it is the best fit of a circular PDF to an elliptical cluster of axes.

7.3.3 Experiment (vi): Non-axisymmetric tensors

In experiment (vi), we repeat experiment (v) with test functions whose tensors do not have cylindrical symmetry. We hypothesise that the Bingham model has significantly lower SSD to the gold standard than the Watson model. We test the PDFs with $\lambda_2/\lambda_3 = \lambda_1/(2\lambda_3)$, $\lambda_1 = 3\lambda_3, 5\lambda_3, 7\lambda_3, 9\lambda_3$, with $Tr(\mathbf{D})$ the same as the symmetric tensors. The principal direction of the tensors is unchanged. Fig. 7.11 shows the SSD for each PDF. The ACG model still has the largest SSD, even though the ACG PDF can have elliptical contours. The difference between the Watson and Bingham PDFs are less obvious, so we compute the statistic $\bar{d}$ as we did for experiment (v). Fig. 7.12 shows the results. The Bingham PDF is significantly better than the Watson PDF at the 95% confidence level, and the difference is larger at high anisotropy.

Figure 7.9: Sum of squared differences to the gold standard in experiment (v) with the PDF orientation fixed at the correct value. Data points are the mean SSD with the gold standard. Error bars show the standard deviation. Markers are blue circles for the Watson PDF, red squares for the Bingham PDF, and black crosses for the ACG PDF.

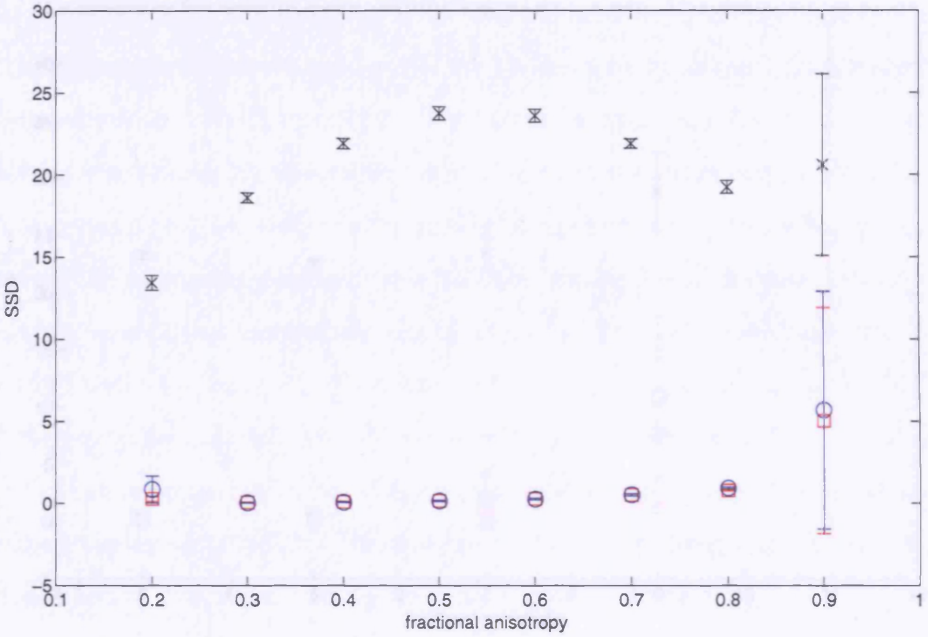


Figure 7.10: The statistic $\bar{d}$(Watson, Bingham) in experiment (v). The error bars are the 95% confidence region for the estimate of $\bar{d}$.

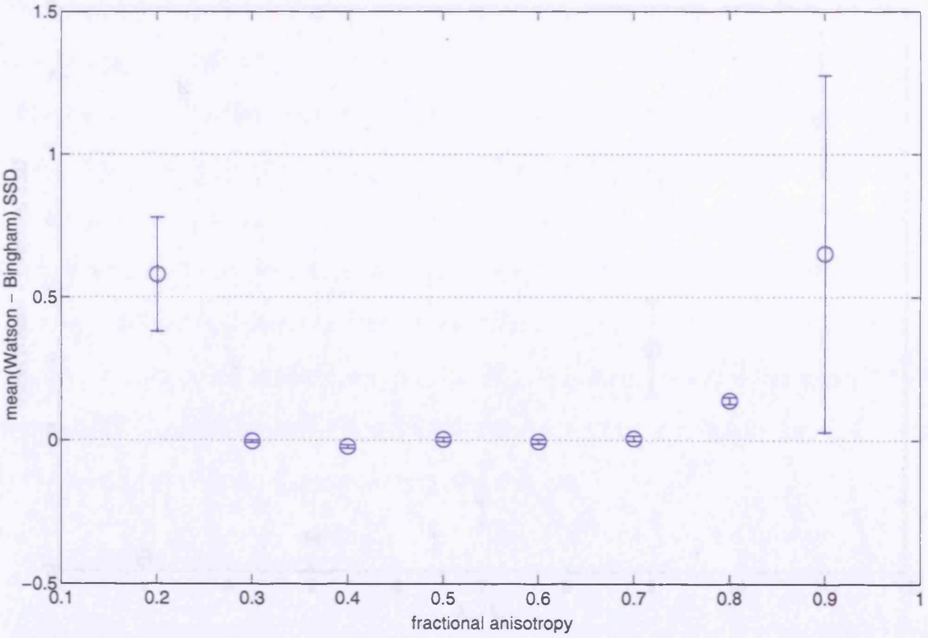


Figure 7.11: Sum of squared differences to the gold standard in experiment (vi) with the PDF orientation fixed at the correct value. Data points are the mean SSD with the gold standard. Error bars show the standard deviation. Markers are blue circles for the Watson PDF, red squares for the Bingham PDF, and black crosses for the ACG PDF.

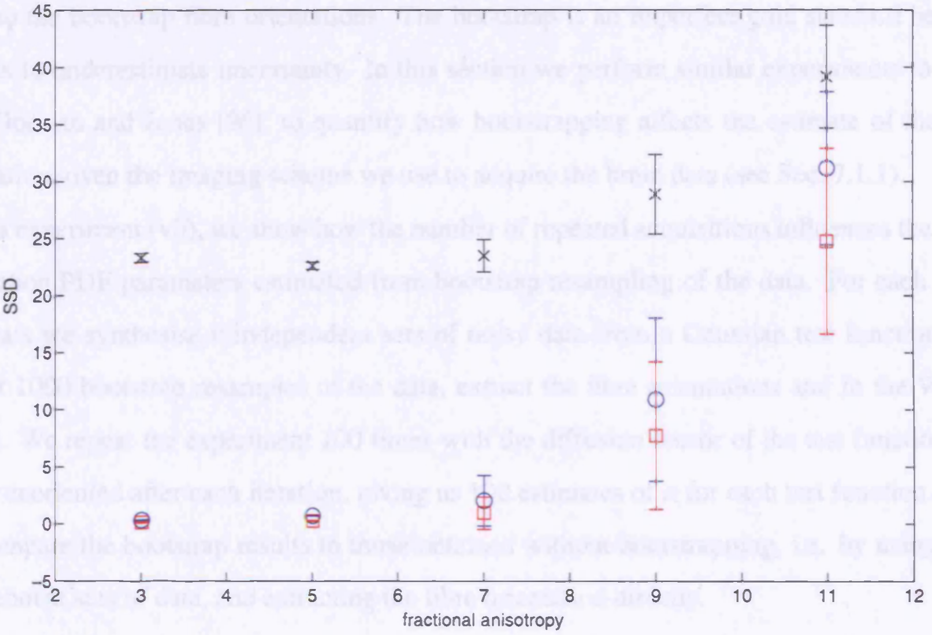
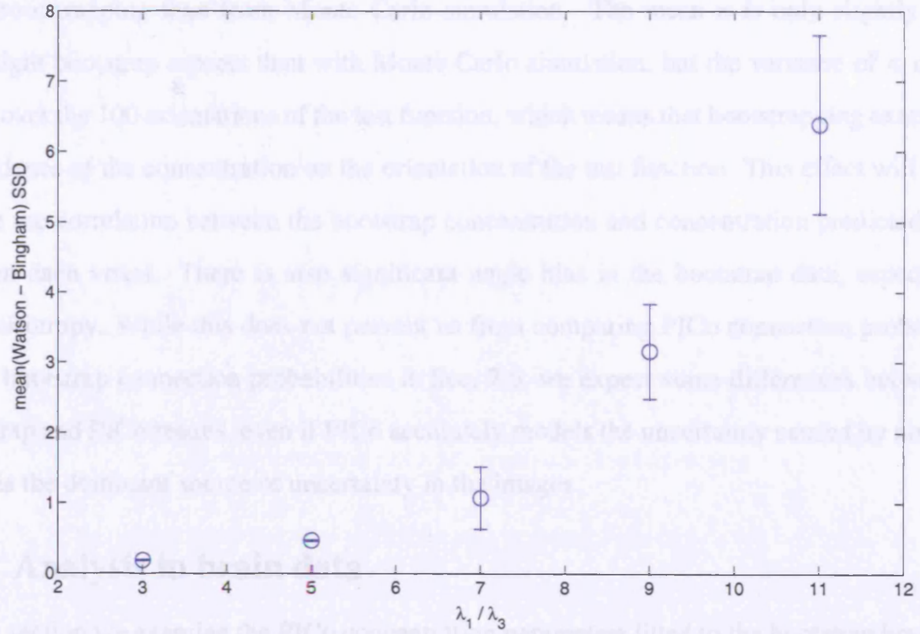


Figure 7.12: The statistic d(Watson, Bingham) in experiment (vi). The error bars are the 95% confidence region for the estimate of d.



7.4 Experiment (vii): bootstrap simulation of one-fibre measurements

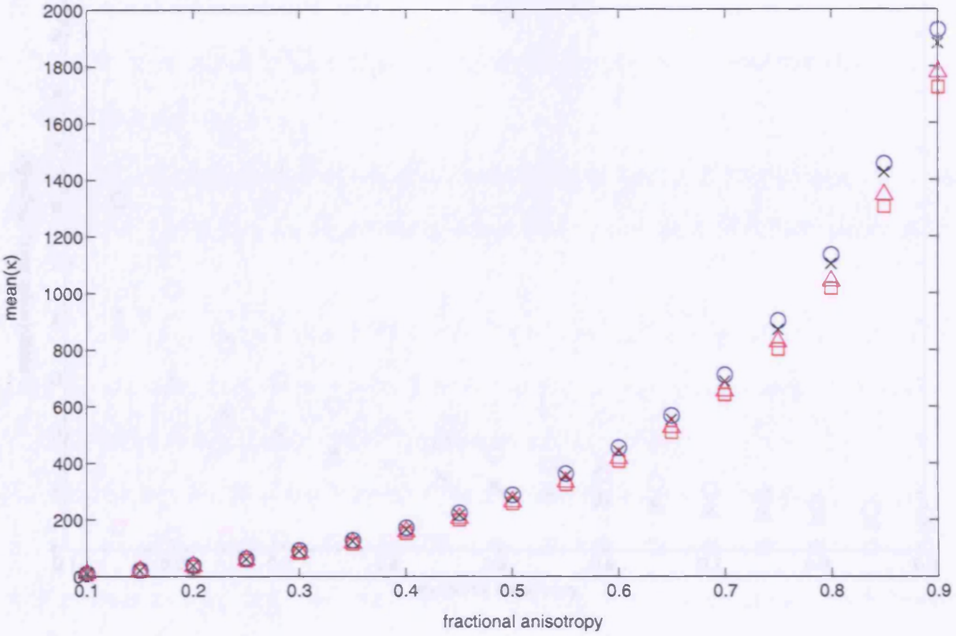
Bootstrap resampling of the brain data gives us an estimate of the concentration of fibre orientations in the brain. In Sec. 7.5 we compare the concentration predicted from the LUTs to that fitted to the bootstrap fibre orientations. The bootstrap is an imperfect gold standard because it tends to underestimate uncertainty. In this section we perform similar experiments to those of O'Gorman and Jones [96], to quantify how bootstrapping affects the estimate of the fibre orientation given the imaging scheme we use to acquire the brain data (see Sec. 7.1.1).

In experiment (vii), we show how the number of repeated acquisitions influences the fibre-orientation PDF parameters estimated from bootstrap resampling of the data. For each set of r repeats we synthesise r independent sets of noisy data from a Gaussian test function. We extract 1000 bootstrap resamples of the data, extract the fibre orientations and fit the Watson model. We repeat the experiment 100 times with the diffusion tensor of the test function randomly reoriented after each iteration, giving us 100 estimates of κ for each test function and r. We compare the bootstrap results to those obtained without bootstrapping, i.e. by using 1000 independent sets of data, and extracting the fibre orientation directly.

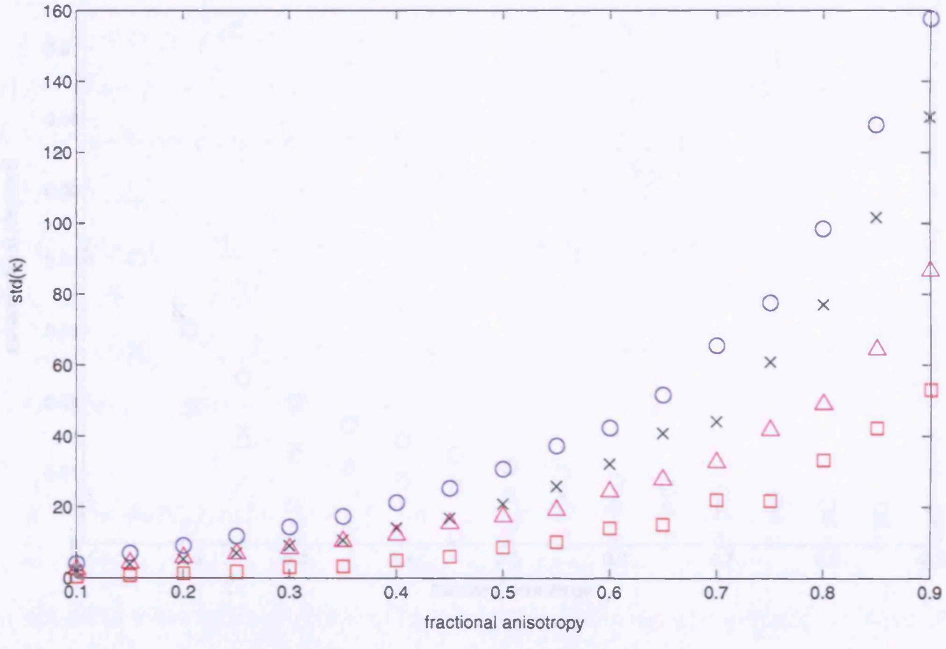
Fig. 7.13 shows the mean and standard deviation of κ for $r = 4, 8$ and 20. Fig. 7.14 shows the "angle bias", the angle between the mean of the bootstrap fibre orientations and the actual orientation of the test function. Bootstrap estimates of the fibre orientation with eight repeats underestimate the uncertainty in the fibre orientation, because the mean κ is higher from bootstrapping than from Monte Carlo simulation. The mean κ is only slightly higher with eight bootstrap repeats than with Monte Carlo simulation, but the variance of κ is much larger over the 100 orientations of the test function, which means that bootstrapping exacerbates dependence of the concentration on the orientation of the test function. This effect will tend to reduce the correlation between the bootstrap concentration and concentration predicted by the LUT in each voxel. There is also significant angle bias in the bootstrap data, especially at low anisotropy. While this does not prevent us from comparing PICo connection probabilities to the bootstrap connection probabilities in Sec. 7.6, we expect some differences between the bootstrap and PICo results, even if PICo accurately models the uncertainty caused by noise and noise is the dominant source of uncertainty in the images.

7.5 Analysis in brain data

In this section we examine the PICo concentration parameters fitted to the bootstrap brain data. For every voxel in the brain, we generate 1000 bootstrap samples of the data, extract 1000 fibre-

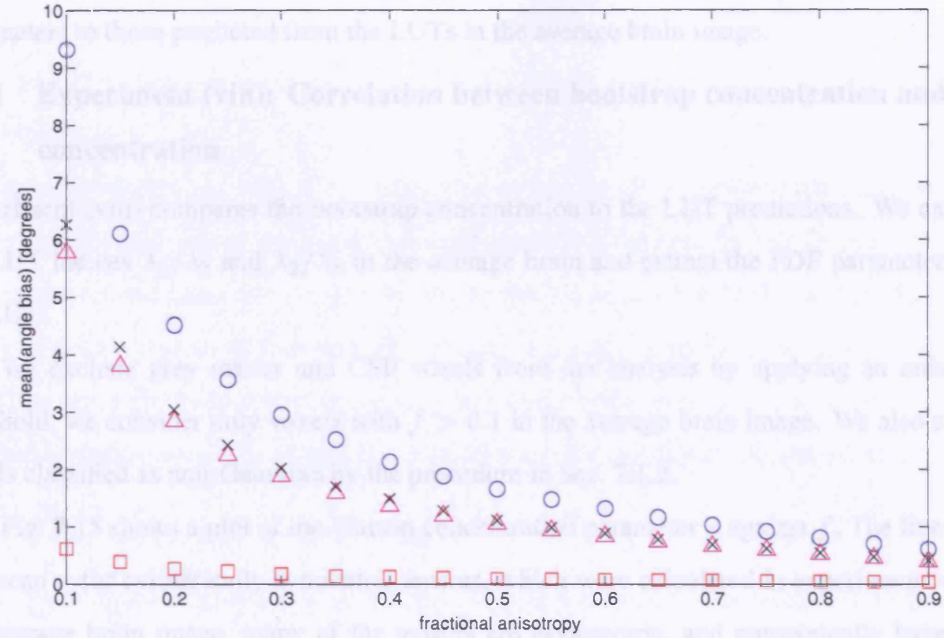


(a) Mean κ as a function of tensor anisotropy.

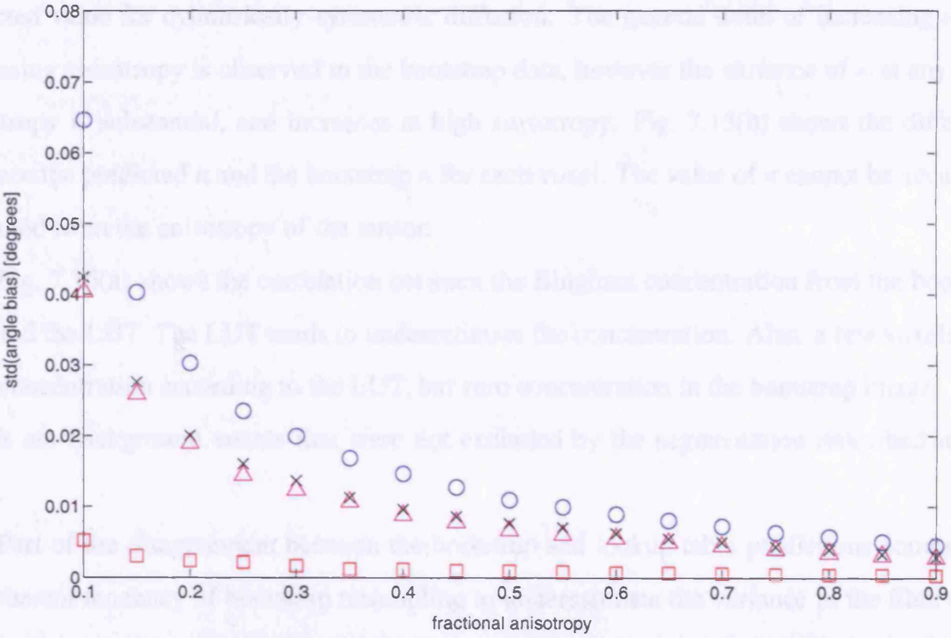


(b) Standard deviation of κ as a function of tensor anisotropy.

Figure 7.13: Bootstrap estimation of diffusion tensor orientation in synthetic data. Bootstrap resampling underestimates the true uncertainty (red squares). The bootstrap results are shown for four repeats (blue circles), eight repeats (black crosses), and twenty repeats (magenta triangles).



(a) Mean angle bias as a function of tensor anisotropy.



(b) Standard deviation of the angle bias as a function of tensor anisotropy.

Figure 7.14: Bootstrap estimation of diffusion tensor orientation in synthetic data. The angle bias is consistently larger in bootstrap samples than in Monte-Carlo simulation (red squares). The bootstrap results are shown for four repeats (blue circles), eight repeats (black crosses), and twenty repeats (magenta triangles).

orientation estimates and fit the parameters of the Watson and Bingham PDF. We compare these parameters to those predicted from the LUTs in the average brain image.

7.5.1 Experiment (viii): Correlation between bootstrap concentration and LUT concentration

Experiment (viii) compares the bootstrap concentration to the LUT predictions. We calculate the LUT indices λ_1/λ_3 and λ_2/λ_3 in the average brain and extract the PDF parameters from the LUT.

We exclude grey matter and CSF voxels from the analysis by applying an anisotropy threshold, we consider only voxels with $f > 0.1$ in the average brain image. We also exclude voxels classified as non-Gaussian by the procedure in Sec. 7.1.2.

Fig. 7.15 shows a plot of the Watson concentration parameter κ against f. The lines show the mean κ for cylindrically symmetric tensors, which were calculated in experiment (vii). In the average brain image, many of the tensors are asymmetric, and consequently have lower κ. Some tensors are sufficiently oblate that they are better modelled by the girdle Watson distribution, with $\kappa < 0$. At higher anisotropy, the κ are more tightly clustered around the expected value for cylindrically-symmetric diffusion. The general trend of increasing κ with increasing anisotropy is observed in the bootstrap data, however the variance of κ at any given anisotropy is substantial, and increases at high anisotropy. Fig. 7.15(b) shows the difference between the predicted κ and the bootstrap κ for each voxel. The value of κ cannot be accurately predicted from the anisotropy of the tensor.

Fig. 7.16(a) shows the correlation between the Bingham concentration from the bootstrap data and the LUT. The LUT tends to underestimate the concentration. Also, a few voxels have high concentration according to the LUT, but zero concentration in the bootstrap image. These voxels are background voxels that were not excluded by the segmentation described in Sec. 7.1.2.

Part of the disagreement between the bootstrap and lookup table predictions comes from the inherent tendency of bootstrap resampling to underestimate the variance in the fibre orientations, as we saw in experiment (vii). This causes the bootstrap concentration to be artificially high, with an increased dependence on the orientation of the fibres relative to the gradient directions.

Another possible cause of disagreement between the LUT and the bootstrap is the error in the LUT indices x and y. If the diffusion has cylindrical symmetry, then $y = 1$ and $\kappa_1 = \kappa_2$, i.e. the Bingham PDF has circular contours. Noise causes an increase in y, which could cause the Bingham concentration $\mid \kappa_1 + \kappa_2 \mid$ to be underestimated. Fig. 7.16(b) shows the difference between κ_2/κ_1 in the bootstrap and LUT concentration. The mean value of κ_2/κ_1 in the LUT concentration is 0.36, in the bootstrap data it is 0.37. This suggests that cylindrical asymmetry caused by noise is not a major cause of the differences between the LUT and bootstrap concentration.

Another limitation of the PICo calibration is that it uses a Gaussian test function, while the value of x and y from the brain data only represents the best fit of a Gaussian model to the data, which may be poorly approximated by the Gaussian model. We have excluded non-Gaussian voxels that we detect with the spherical harmonic algorithm, but some voxels with non-Gaussian diffusion undoubtedly remain. It is possible that these non-Gaussian effects break the relationship between anisotropy and concentration that we expect from the LUTs.

7.6 Tractography in brain data

In this section we compare PICo probability maps generated from bootstrap data to those produced by using the PDFs and LUTs. The tractography algorithm for the bootstrap PICo maps is described in Sec. 7.2.1. We define a region of interest (ROI) across the corpus callosum of the subject. Fig. 7.17 shows a sagittal slice of the fractional anisotropy image with the ROI highlighted.

7.6.1 Experiment (ix): Stability of PICo connection probabilities

The bootstrap tracking process is much more computationally expensive than the PDF sampling. Experiment (ix) tests the stability of connection probabilities in brain image 1, to find the minimum number of iterations that give acceptable reproducibility of the results. We compute four PICo maps for each seed in the ROI using the Bingham model, one with 1000 iterations, one with 2000 iterations and two with 5000 iterations. Fig. 7.18 shows the SSD between each pair of images. Increasing the iterations above 1000 does reduce the errors, but the SSD at 1000 iterations is below 0.1 for all seed points in the ROI (mean 0.04) so we use 1000 iterations to generate the bootstrap PICo maps with the limitation that we cannot detect differences between models that are less than the difference between subsequent PICo runs in the same image.

7.6.2 Experiment (x): PICo calibrated directly from bootstrap fibre orientations

Experiment (x), tests how well each PDF models the distribution of fibre orientations from the bootstrap data, given the actual bootstrap fibre orientations computed in experiment (viii). We compute connection probability maps from every voxel in the ROI, first using the bootstrap data, and then using the Watson, Bingham and ACG PDF fitted to a sample of bootstrap fibre orientations in experiment (viii).

This experiment tests the ability of each PDF to model the distribution of fibre orientations extracted from the bootstrap data when the PDF orientation and concentration are known. Large differences between the bootstrap tractography and PICo results suggest intrinsic deficiencies in the PDFs as models of the fibre-orientation PDF. We conclude from experiment (iv) that the Watson and Bingham PDFs are good models of the uncertainty caused by image noise, when the test function that produces the data is Gaussian. Experiment (x) tests whether the PDFs accurately model bootstrap fibre orientations from the brain, when the models are calibrated directly from the bootstrap fibre orientations. Fig 7.19 shows the SSD between the bootstrap PICo maps and the models fitted to the bootstrap fibre orientations in experiment (viii). We conclude from this that the Bingham PDF is the best model of the bootstrap uncertainty, since it consistently has the lowest SSD of all models: the mean and standard deviation of the SSD over all seed points is 0.07 ± 0.03, the mean and standard deviation of the SSD between repeated runs of the PICo algorithm in experiment (ix) is 0.04 ± 0.02. At the 95% confidence level, $\bar{d}(\text{Watson} - \text{Bingham}) = 0.6 \pm 0.17$. Fig. 7.20 shows the difference between the Watson and Bingham PICo maps for the seed point where the Watson model has the highest SSD to the bootstrap connection probability map. The largest difference in connection probability is 0.09. For the other seed points, the differences are typically smaller, in one seed point where $d(\text{Watson} - \text{Bingham}) = 0.6$, the mean difference between the non-zero Bingham and Watson connection probabilities is 0.006.

7.6.3 Experiment (xi): PICo calibrated by tensor shape

Experiment (xi) compares the bootstrap results to the PICo maps, with PDF concentration defined from LUTs. This experiment is similar to experiment (v). The orientation of the PDFs is the same as in experiment (x), but the concentration is determined from the LUTs, with the estimated tensor eigenvalues subject to the effects of noise.

Fig. 7.21 shows the SSD when the tensor eigenvalues are estimated in the average brain. The trends are similar to those in Fig. 7.22, the mean and standard deviation of the SSD for the Bingham model is 0.57 ± 0.4.

Fig. 7.22 shows the SSD when the PDF concentration is determined from the LUTs, with the tensor eigenvalues estimated from brain image 1. The pattern is similar to the result of experiment (x), and the Bingham PDF remains the best model. The mean and standard deviation of the SSD for the Bingham model is 1.00 ± 0.95, an order of magnitude larger than the SSD when the Bingham PDF is calibrated directly from the bootstrap data. The Watson SSD with the tensor eigenvalues taken from brain image 1 is 2.97 ± 2.2.

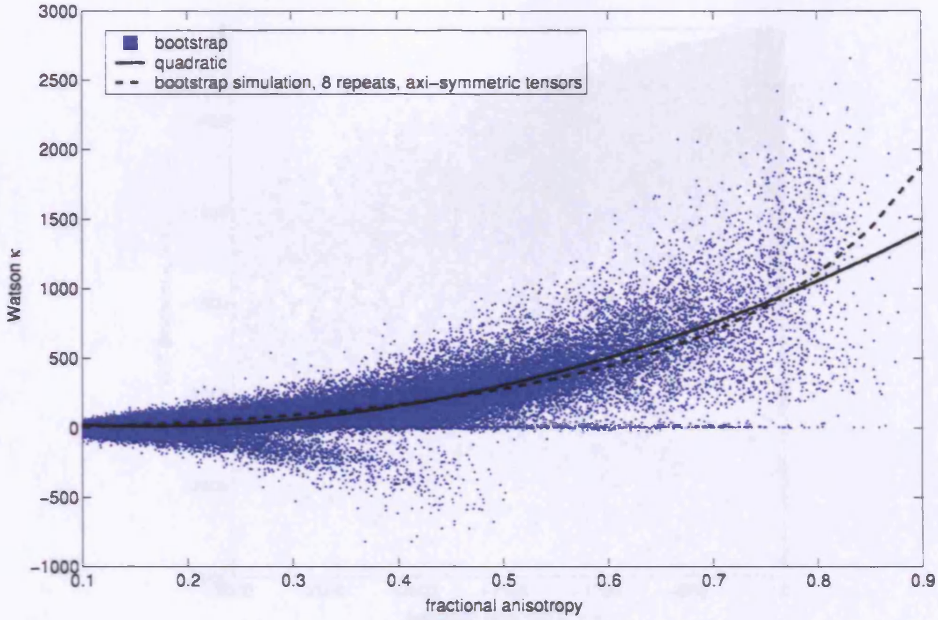
Fig. 7.23(a) shows the difference in connection probabilities between the bootstrap tracking and PICo using the Bingham PDF with concentration calculated from the bootstrap fibre orientations. Fig. 7.23(b) shows the differences between the bootstrap tracking and PICo using the Bingham PDF with concentration predicted from the LUT, where the indices x and y of the LUT were calculated from the tensors in brain image 1. The difference in connection probabilities is much larger in the latter case. In both images, the differences become larger with increasing distance from the seed point in the centre of the corpus callosum, but this does not necessarily mean that the differences in concentration are also larger further along the streamline path, because the differences in streamline trajectories at the scale of the voxel accumulate as the streamlines disperse from the seed point.

Since the increase in SNR resulting from averaging significantly lowers the SSD, we conclude that the error in the tensor eigenvalues is a significant source of error in the PDF concentration for PICo in the brain.

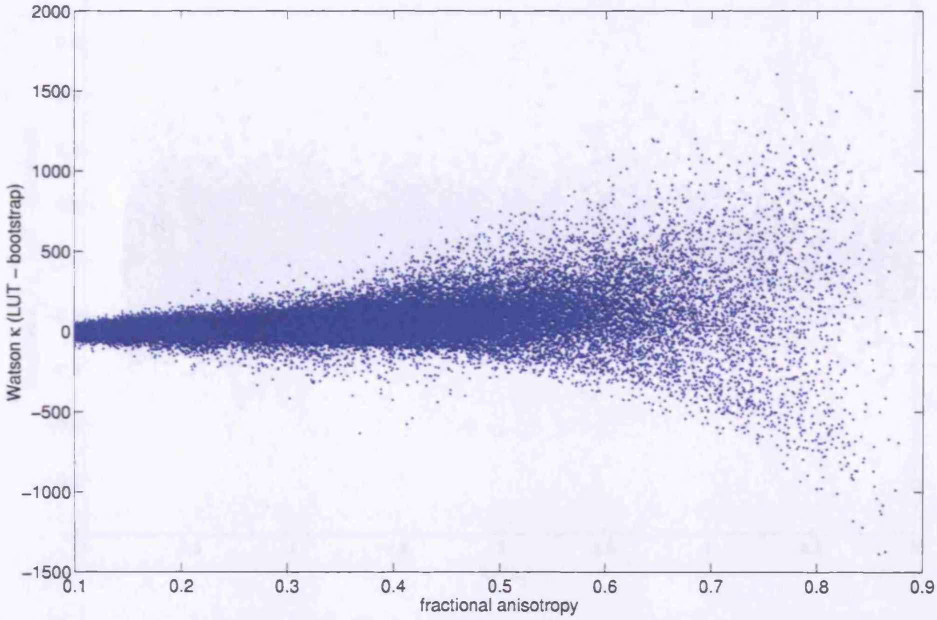
7.7 Conclusions

We have evaluated the Watson, Bingham and ACG models of the fibre-orientation PDF. The experiments in synthetic data suggest that the Watson and Bingham PDFs are significantly better models of the uncertainty due to noise than than the ACG PDF. The bootstrap experiments show some correlation between the tensor anisotropy and the concentration. There is still significant disagreement between the bootstrap and LUT concentration, however these errors do not in most cases cause large differences in the connection probabilities from PICo (with PDF concentration defined in the average brain) compared to those obtained from bootstrap fibre tracking. The largest source of error comes from using noisy estimates of the tensor eigenvalues to obtain the concentration parameters from the LUT.

The Bingham PDF produces consistently lower SSD to the bootstrap connection probabilities than the Watson PDF. The difference between the two models is apparent whether the PICo concentration parameters are fitted to the bootstrap fibre orientations or taken from the LUT.

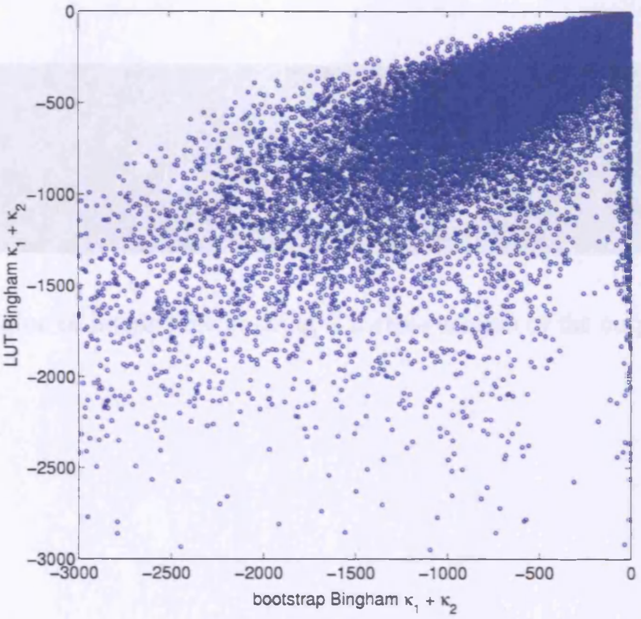


(a) Watson κ against f in experiment (viii). The points show κ calculated from the bootstrap data, plotted against f of the corresponding voxel in the average brain. The solid line is a quadratic curve fitted to the data points. The dashed line is the predicted κ for cylindrically-symmetric tensors, based on a bootstrap experiment with eight repeats, from experiment (vii).

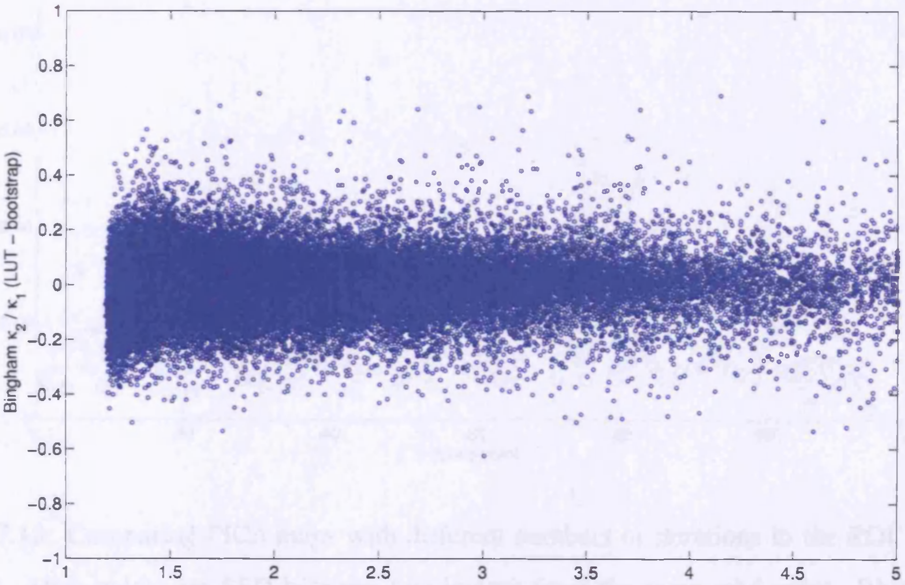


(b) Watson κ predicted from the LUT, subtracted from κ found in the bootstrap data in experiment (viii).

Figure 7.15: Watson κ plotted against the value of f in the corresponding voxel in the average brain data.



(a) Bingham concentration from the bootstrap data plotted against that predicted from the LUT using the values of x and y in the average brain.



(b) Difference in the statistic κ_2/κ_1 calculated from the bootstrap data and from the LUT, plotted against x of the corresponding voxel in the average brain.

Figure 7.16: Bingham parameters in the bootstrap data and predicted from the LUT in experiment (viii).

Figure 7.17: The region of interest for tracking is a cross section of the corpus callosum of the subject.

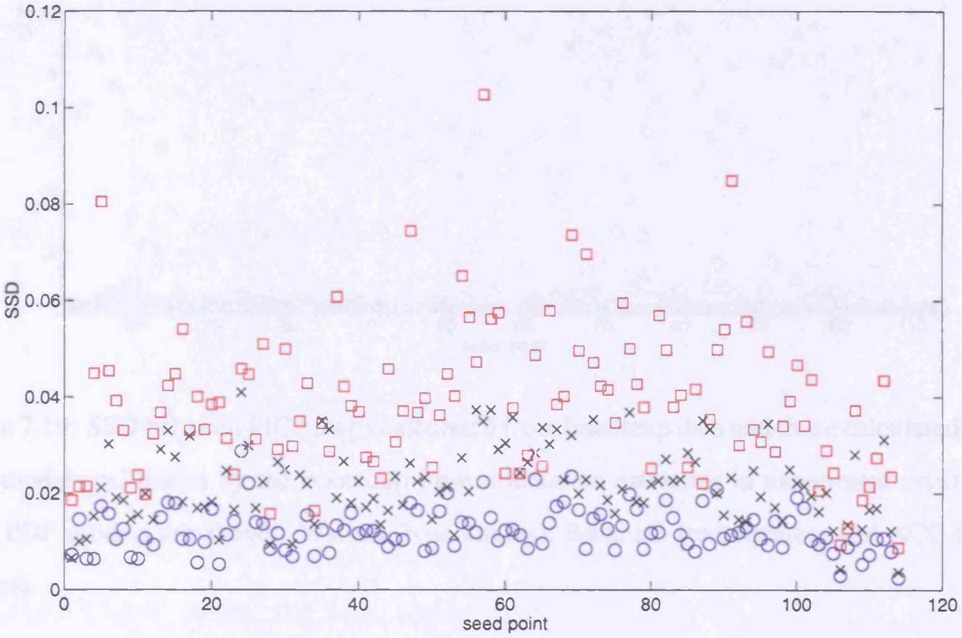


Figure 7.18: Comparing PICo maps with different numbers of iterations in the ROI in brain image 1. Data points are SSD between two images from the same seed point. Blue circles are the SSD between two different maps each with 5000 iterations. Red squares are the SSD between 1000 and 5000 iterations, black crosses are the difference between 2000 and 5000 iterations.

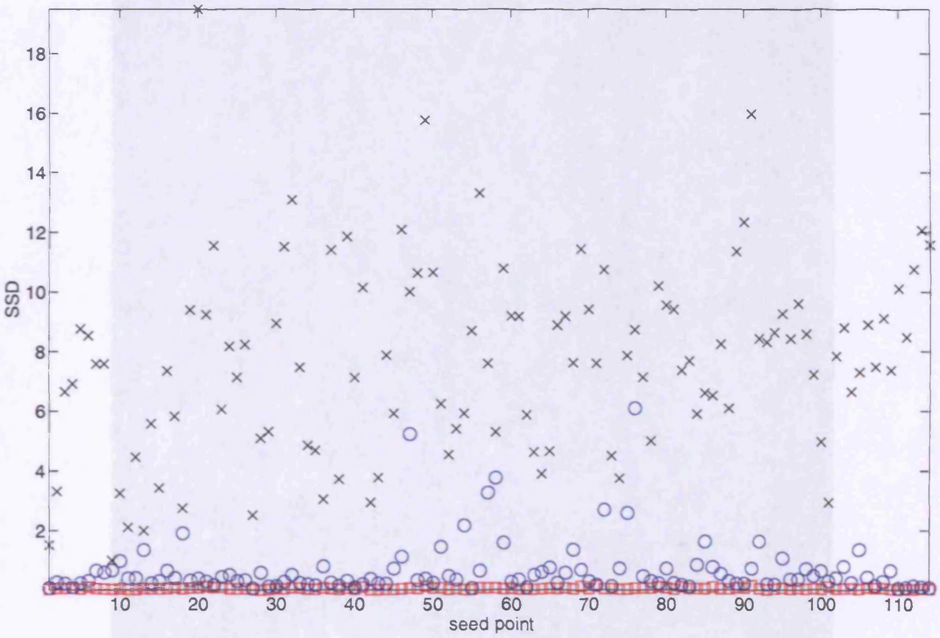


Figure 7.19: SSD between PICo maps calculated from bootstrap data and those calculated using PDF models calibrated by the bootstrap fibre orientation estimates in experiment (viii). The three PDF models are shown: Watson (blue circles), Bingham (red squares) and ACG (black crosses).

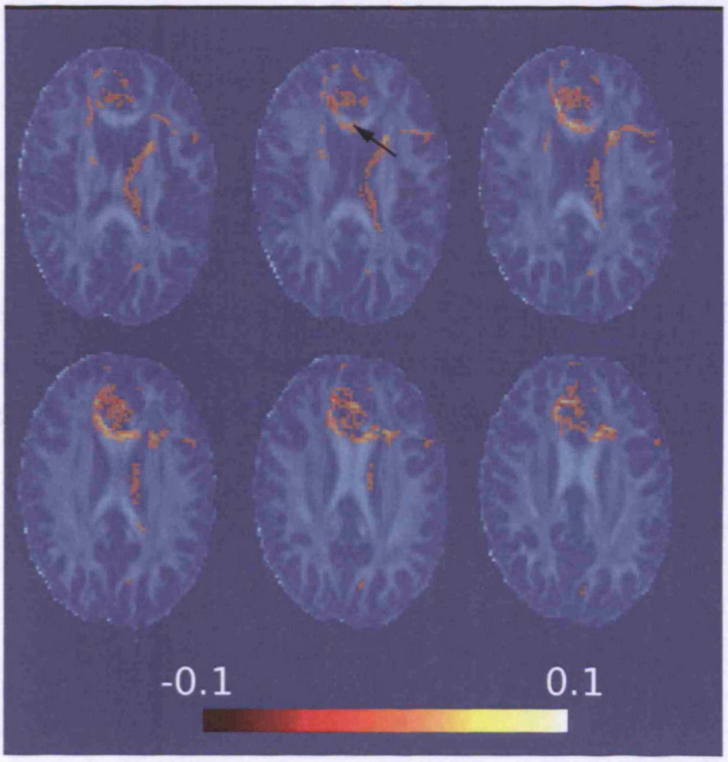


Figure 7.20: Differences in connection probability between the Bingham and Watson model in experiment (x). The SSD to the bootstrap image for this seed is 6.1 for the Watson model and 0.1 for the Bingham model. The arrow highlights the seed point.

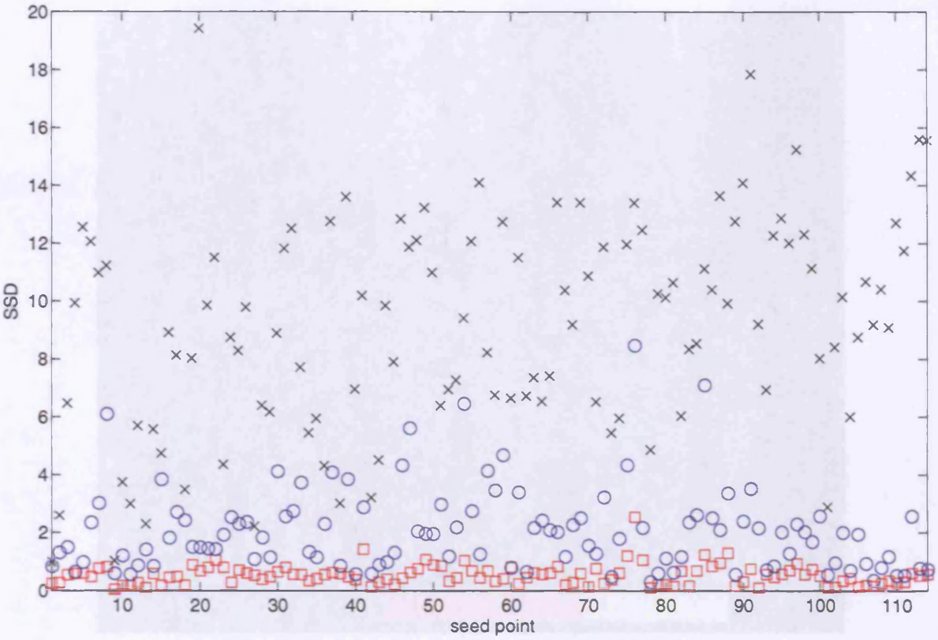


Figure 7.21: SSD between PICo maps calculated from bootstrap data and PDF models with tensor anisotropy estimated from the average brain. The three PDF models are shown: Watson (blue circles), Bingham (red squares) and ACG (black crosses).

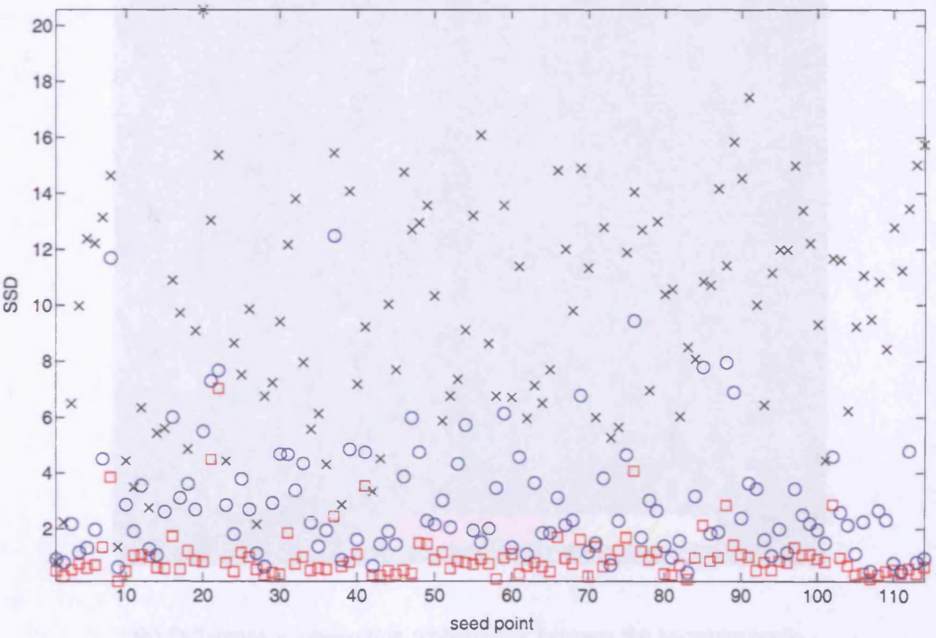
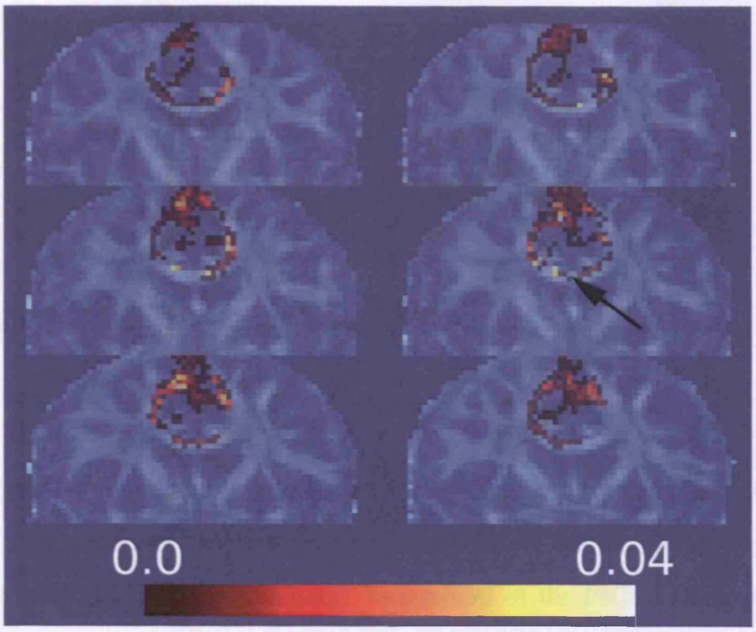
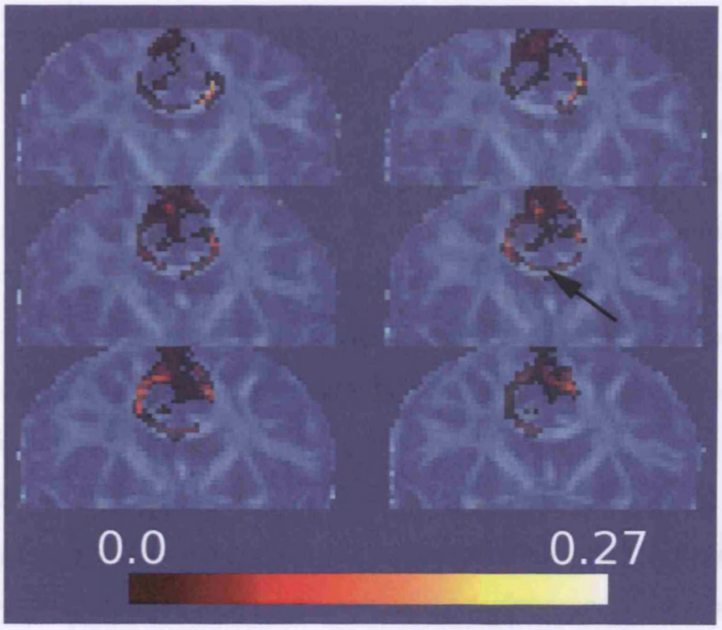


Figure 7.22: SSD between PICo maps calculated from bootstrap data and PDF models with tensor anisotropy estimated from brain image 1. The three PDF models are shown: Watson (blue circles), Bingham (red squares) and ACG (black crosses).



(a) Difference in connection probability between the bootstrap tracking and PICo with the Bingham concentration parameters calculated from the bootstrap fibre orientations in experiment (viii).



(b) Difference in connection probabilities between the bootstrap tracking and PICo with the Bingham concentration parameters taken from the LUT, with the tensor shape parameters from brain image 1.

Figure 7.23: Difference in connection between PICo (using the Bingham PDF) and the bootstrap tracking from a single seed point, which is highlighted by the arrow.

Chapter 8

Models of the two-fibre PDF

Multi-fibre reconstruction algorithms resolve two or more distinct fibre orientations within a voxel. Parker and Alexander [56] extend PICo to model the uncertainty in fibre orientations in voxels containing two fibre orientations. The model of the PICo PDF in these voxels is calibrated by synthesising a set of fibre-orientation estimates $[\mathbf{x}_{11}, \mathbf{x}_{21}; \mathbf{x}_{12}, \mathbf{x}_{22}; \ldots; \mathbf{x}_{T1}, \mathbf{x}_{T2}]$ from T trials of a two-fibre reconstruction algorithm, and fitting the PDF parameters to these axes. Parker and Alexander calculate the fibre orientations from a two-compartment Gaussian model in [56] and in later work [100] calibrate PICo with fibre orientations calculated using PAS-MRI.

In this chapter, we extend the Watson and Bingham PDFs to model the uncertainty in fibre orientation in two-fibre voxels. The test function for the calibration is a two-compartment Gaussian model. Parker and Alexander [56] divide the fibre-orientation estimates from a fitted two-compartment Gaussian model into two sets, one for each principal direction, and fit the single-fibre PDF to each group of axes. For each trial, the axis most closely aligned to one of the principal directions of the test function is assigned to that set. However, this division introduces a potential for misclassification. Alexander and Barker [38] later proposed an iterative method, which we examine in the next section.

We propose a single mixture model for the PDF, which fits the whole set of axes at once, without dividing the axes into two groups. We compare this approach to the iterative method proposed by Alexander and Barker. We then evaluate the performance of the models in synthetic data and in bootstrap brain data.

8.1 Iterative sorting

The iterative method proposed by Alexander and Barker [38] assigns the fibre orientations from each trial to two equally-sized sets, s_1 and s_2, where each set contains the axes from one fibre population. The axes are initially assigned to the sets in the order that they returned from the two-tensor fitting routine. The algorithm then proceeds as follows:

```
1. counter = 0;
2. swap = true;
3. while (counter < 100 & swap)
4. compute mean orientation $\mu_1$ of $s_1$ and $\mu_2$ of $s_2$
5. swap = false;
    (a) for all axes $\mathbf{x}_{t1}$ in $s_1$ from trial $t$: if ($| \mathbf{x}_{t1} \cdot \mu_1 | < | \mathbf{x}_{t1} \cdot \mu_2 |$)
        i. assign $\mathbf{x}_{t1}$ to $s_2$ and $\mathbf{x}_{t2}$ to $s_1$.
        ii. swap = true
        iii. counter = counter + 1
```

Once the algorithm terminates, we fit a spherical PDF to each set of fibre orientations using the standard methods described in Ch. 6.

Alexander and Barker [38] find that the sorting process usually converges after three iterations. This is the case when the clusters of fibre orientations are reasonably well separated, but is not always true when the clusters are not well separated. Occasionally the algorithm oscillates, switching a few axes back and forth between the sets, so we terminate the algorithm after 100 iterations.

8.1.1 Initialisation step

In this thesis, we initialise the algorithm differently. When the principal directions of the test function $\mathbf{e}_{11}$ and $\mathbf{e}_{12}$ are known (for example, in PICo LUT calibration) we use these to make the initial division. For each pair of fibre orientations from trial t, we assign $\mathbf{x}_1$ to s_1 and $\mathbf{x}_2$ to s_2 if $| \mathbf{e}_{11} \cdot \mathbf{x}_1 | > | \mathbf{e}_{12} \cdot \mathbf{x}_1 |$, otherwise we assign $\mathbf{x}_1$ to s_2 and $\mathbf{x}_2$ to s_1. When the true fibre orientations are not known, we substitute a pair of axes from one of the trials for $\mathbf{e}_1$ and $\mathbf{e}_2$.

8.2 Two-Watson model

The two-Watson model is an equally-weighted mixture of two Watson distributions. One distribution models the axes from each fibre population, and the mixture is equal because half of the axes in the sample come from each compartment. We optimise the parameters of both distributions within a single objective function, to maximise the likelihood of the entire set of $2T$

axes. The PDF of the model is:

$$p(\mathbf{x}) = \frac{1}{2}M\left(\frac{1}{2},\frac{3}{2},\kappa_1\right)^{-1}\exp\left[\kappa_1(\boldsymbol{\mu}_1\cdot\mathbf{x})^2\right]\frac{1}{2} + M\left(\frac{1}{2},\frac{3}{2},\kappa_2\right)^{-1}\exp\left[\kappa_2(\boldsymbol{\mu}_2\cdot\mathbf{x})^2\right]. \tag{8.1}$$

We optimise the parameters of the two Watson distributions with a Levenberg-Marquardt minimisation [67] of the negative log-likelihood ψ of the fibre orientation estimates:

$$\psi = -\sum_{i=1}^{T}\log[p(\mathbf{x}_{i1})] + \log[p(\mathbf{x}_{i2})]. \tag{8.2}$$

We constrain the optimisation to return bipolar distributions, by optimising $\omega_n^2 = \sqrt{\kappa_n}$, $n = 1, 2$. We define the modal axes $\boldsymbol{\mu}_n$ in terms of spherical polar coordinates θ_n and ϕ_n. The Levenberg-Marquardt algorithm fits the parameters $[\theta_1, \phi_1, \omega_1, \theta_2, \phi_2, \omega_2]$ by minimising ψ in Eq. 8.2.

We divide the axes from each trial into two sets, s_1 and s_2, as in Sec. 8.1.1 to provide an initial estimate of the parameters of each Watson distribution. Without this initialisation step the optimisation often converges on a local minimum.

8.3 Two-Bingham model

In a similar way to the two-Watson model, we optimise the parameters of a mixture of two Bingham distributions to maximise the likelihood of the entire set of $2T$ axes. The two-Bingham model has PDF:

$$p(\mathbf{x}) = \frac{1}{2}M(\frac{1}{2},\frac{3}{2},\mathbf{A}_1)^{-1}\exp(\mathbf{x}^T\mathbf{A}_1\mathbf{x}) + \frac{1}{2}M(\frac{1}{2},\frac{3}{2},\mathbf{A}_2)^{-1}\exp(\mathbf{x}^T\mathbf{A}_2\mathbf{x}), \tag{8.3}$$

where the parameter matrix $\mathbf{A}_n$, $n = 1, 2$, may be decomposed into its eigen system:

$$(\boldsymbol{\mu}_{3n}, \boldsymbol{\mu}_{2n}, \boldsymbol{\mu}_{1n})\begin{pmatrix}\kappa_{3n} & 0 & 0\\ 0 & \kappa_{2n} & 0\\ 0 & 0 & \kappa_{1n}\end{pmatrix}(\boldsymbol{\mu}_{3n}, \boldsymbol{\mu}_{2n}, \boldsymbol{\mu}_{1n})^T. \tag{8.4}$$

We set $\kappa_{3n} = 0$ as explained in Sec. 6.2. Substituting into Eq. 8.3 gives

$$\begin{aligned}p(\mathbf{x}) =& \frac{1}{2}M(\frac{1}{2},\frac{3}{2},\mathbf{A}_1)^{-1}\exp\left(\kappa_{11}(\boldsymbol{\mu}_{11}\cdot\mathbf{x})^2 + \kappa_{21}(\boldsymbol{\mu}_{21}\cdot\mathbf{x})^2\right)\\ &+ \frac{1}{2}M(\frac{1}{2},\frac{3}{2},\mathbf{A}_2)^{-1}\exp\left(\kappa_{12}(\boldsymbol{\mu}_{12}\cdot\mathbf{x})^2 + \kappa_{22}(\boldsymbol{\mu}_{22}\cdot\mathbf{x})^2\right).\end{aligned} \tag{8.5}$$

We use a conjugate gradient method [67] to minimise the negative log likelihood ψ of the fibre-orientation estimates:

$$\psi = -\sum_{i=1}^{T}\log[p(\mathbf{x}_{i1})] + \log[p(\mathbf{x}_{i2})]. \tag{8.6}$$

We define two rotation matrices, $\mathbf{R}_{\theta,\phi}$ rotates the vector $[0,0,1]^{\mathrm{T}}$ to the orientation (θ,ϕ) and $\mathbf{R}_{\gamma}$ is a rotation of γ degrees about the axis (θ,ϕ). The principal axes of each distribution, $\boldsymbol{\mu}_{1n}$ and $\boldsymbol{\mu}_{2n}$ are perpendicular, so we substitute $\boldsymbol{\mu}_{1n} = \mathbf{R}_{\theta,\phi}[0,0,1]^{\mathrm{T}}$ and $\boldsymbol{\mu}_{2n} = \mathbf{R}_{\psi_n}\mathbf{R}_{\theta_n,\phi_n}[1,0,0]^{\mathrm{T}}$. As with the Watson model, we wish to restrict the concentration parameters $\kappa_{1n} < \kappa_{2n} \leq 0$. We therefore substitute $\kappa_{1n} = -\omega_n^2$, $\kappa_{2n} = \omega_n^2 \cos^2(\theta_n)$.

The conjugate gradient method minimises fits the parameters $[\theta_1$, ϕ_1, γ_1, κ_1, ω_1, θ_2, ϕ_2, γ_2, κ_2, ω_2, $\theta_2]$ that minimise ψ in Eq. 8.6. We use conjugate-gradient method here, because it does not require us to compute the second derivative of the objective function, which makes the method easier to implement.

We divide the axes from each trial into two sets, s_1 and s_2, as in Sec. 8.1.1, to provide an initial estimate of the parameters of each Bingham distribution. Without this initialisation step the optimisation often converges on a local minimum.

8.4 Experiment (xii): multi-Gaussian compartment fitting routines

The multi-Gaussian compartment model (Sec. 3.4.3) with $n = 2$ has a maximum of 13 parameters: six for each tensor and one mixing parameter. However, it may be desirable to fit a model with fewer parameters to improve robustness and to simplify the PICo calibration process. One possible constraint is to fix the mixing parameter $\alpha = 0.5$, so there are 12 parameters. We can also restrict the diffusion tensors to be cylindrically symmetric, so the second and third eigenvalues are the same. We do this as explained by Alexander [18], by substituting $\mathbf{D} = \alpha \mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} + \beta \mathbf{I}$, where $\mathbf{e}_1$ is the principal direction of the tensor, I is the identity tensor, and $\mathbf{D}$ has eigenvalues $\lambda_1 = \alpha + \beta$, $\lambda_{2,3} = \beta$.

This reduces the number of free parameters from 13 to 9. If we enforce cylindrical symmetry and fix the mixing parameter $\alpha = 0.5$, there are 8 free parameters.

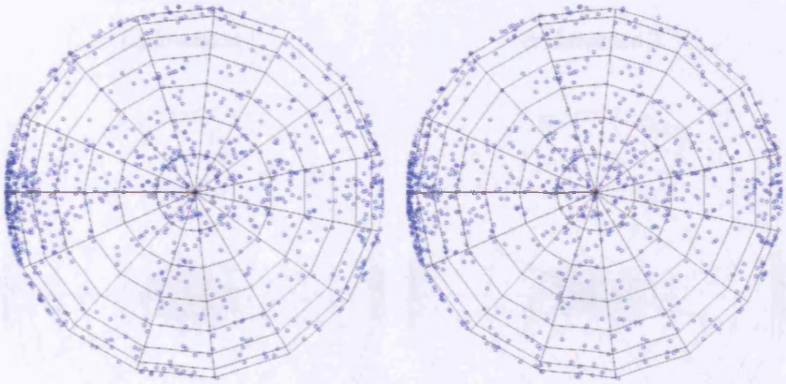
Experiment (xii) makes a qualitative comparison between the performance of each inversion in synthetic data. We generate 1000 noisy sets of diffusion-weighted data using a two-Gaussian test function $\frac{1}{2}G(\mathbf{D}_1, t) + \frac{1}{2}G(\mathbf{D}_2, t)$, where $\mathbf{D}_1 = \mathrm{diag}[\lambda_1, \lambda_2, \lambda_2]$ and $\mathbf{D}_2 = \mathrm{diag}[\lambda_2, \lambda_2, \lambda_1]$. We estimate the orientation and anisotropy of the tensors using four inversion routines:

1. full inversion (13 parameters)
2. equal mixing inversion (12 parameters).
3. cylindrical inversion (9 parameters)
4. restricted inversion (8 parameters)

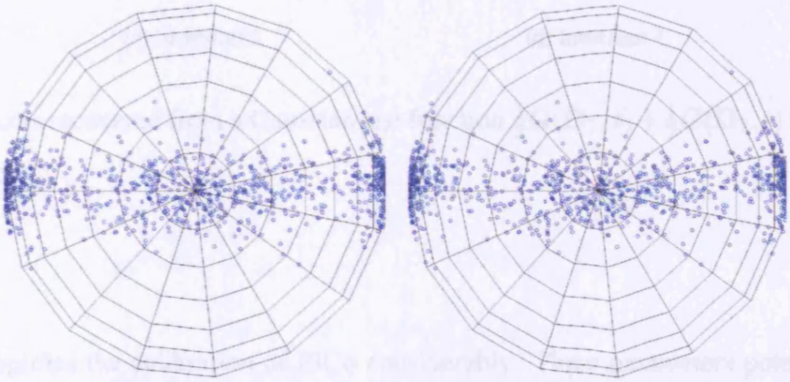
Both tensors have the same fractional anisotropy, f, which we vary between $f = 0.4$ and $f = 0.8$.

8.4.1 Concentration of fibre orientations

Inversions 3 and 4 have more concentrated fibre orientations, which are grouped in elliptical clusters. At $f = 0.4$ (Fig. 8.1) and $f = 0.6$ (Fig. 8.2), these inversions produce significantly more concentrated fibre-orientation estimates than the other inversions. The inversions are more similar at $f = 0.8$ (Fig. 8.3).



(a) Inversion 1. (b) Inversion 2.

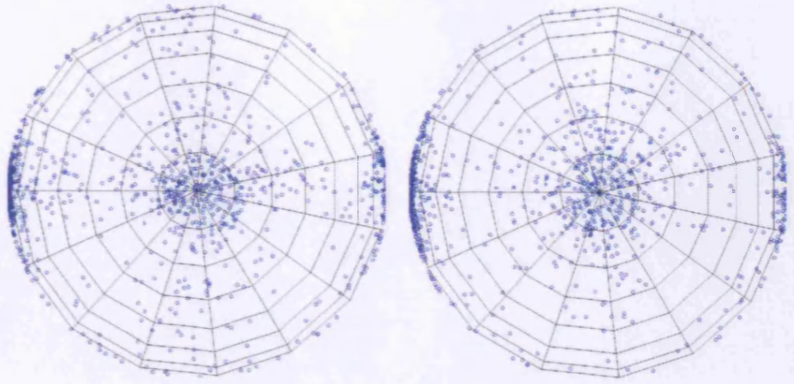


(c) Inversion 3. (d) Inversion 4.

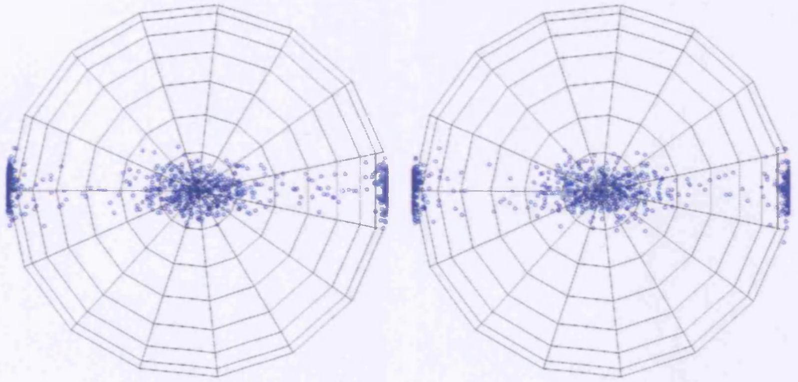
Figure 8.1: Axes recovered from a Gaussian test function $\frac{1}{2}G(\mathbf{D}_1, t) + \frac{1}{2}G(\mathbf{D}_2, t)$ with $f = 0.4$ for both tensors.

8.4.2 Estimation of fractional anisotropy

Fig. 8.4 histograms the anisotropy of the fitted tensors at $f = 0.4$, Fig. 8.5 shows the estimated anisotropy at $f = 0.8$. The histograms from inversion 1 are the least biased from the true values. We conclude that fixing the mixing parameter improves the estimate of f. The restricted



(a) Inversion 1. (b) Inversion 2.

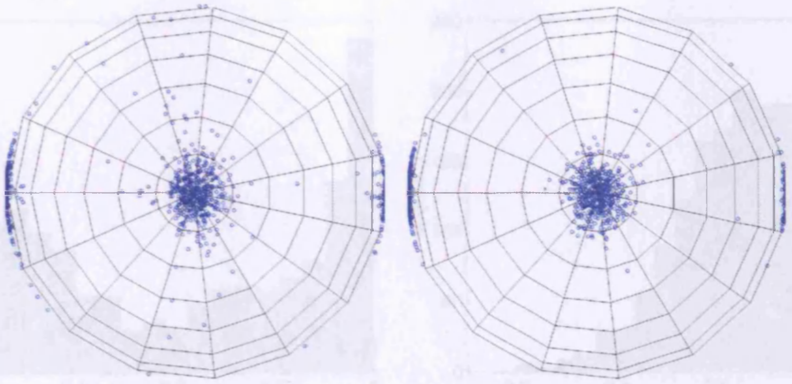


(c) Inversion 3. (d) Inversion 4.

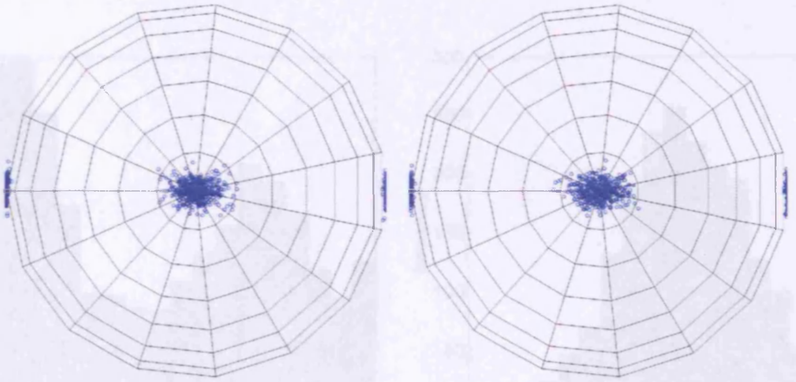
Figure 8.2: Axes recovered from a Gaussian test function $\frac{1}{2}G(\mathbf{D}_1, t) + \frac{1}{2}G(\mathbf{D}_2, t)$ with $f = 0.6$ for both tensors.

inversion simplifies the calibration of PICo considerably. Three parameters potentially affect the concentration of fibre orientations: the anisotropy of the two tensors, and the angle between their principal directions. Calibrating PICo with an unequal mixing parameter would add an extra dimension to the LUT, while calibration of non-cylindrically symmetric tensors would require an even more lengthy calibration, since the concentration would be dependent on λ_1/λ_3 and λ_2/λ_3 and the angle between the second eigenvectors of each tensor.

We therefore assume that each compartment contains cylindrically symmetric tensors in the calibration of the PICo PDFS, which is a reasonable assumption because such symmetry is often observed in anisotropic white matter containing a single bundle of white matter fibres [6].



(a) Inversion 1. (b) Inversion 2.



(c) Inversion 3. (d) Inversion 4.

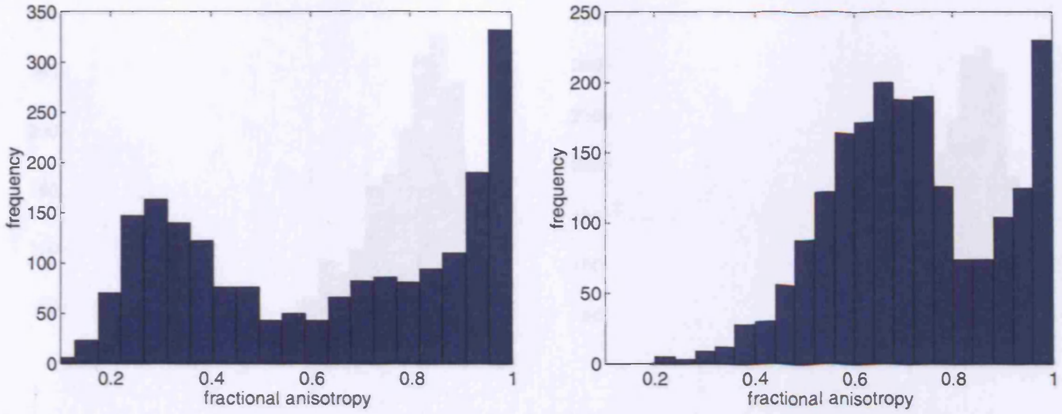
Figure 8.3: Axes recovered from a Gaussian test function $\frac{1}{2}G(\mathbf{D}_1, t) + \frac{1}{2}G(\mathbf{D}_2, t)$ with $f = 0.8$ for both tensors.

8.4.3 Mixing parameter

We further assume in the two-fibre model that the mixing parameter is 0.5, which we may expect to be the average value over all the voxels containing two fibres. Fig. 8.6 shows the histogram of the mixing parameter of the largest compartment, in the noisy synthetic data with $f = 0.8$, using inversion 1 and 3. Inversion 3 appears to be unbiased while 1 is biased away from the true value of 0.5.

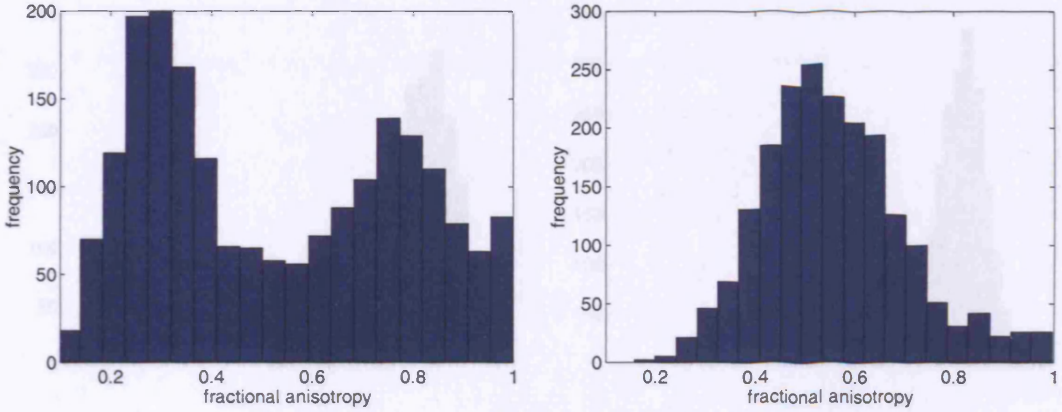
8.4.4 Experiment (xiii): concentration of fibre orientations from unequally mixed compartments

Experiment (xiii) shows a qualitative analysis of the effects of fixing the mixing parameter at 0.5 in the model fitting when the mixing parameter of the test function is not 0.5. We generate 1000 noisy trials of the diffusion-weighted data using a two-Gaussian test function $\alpha G(\mathbf{D}_1, t)+$



(a) Inversion 1, $f = 0.4$.

(b) Inversion 2, $f = 0.4$.

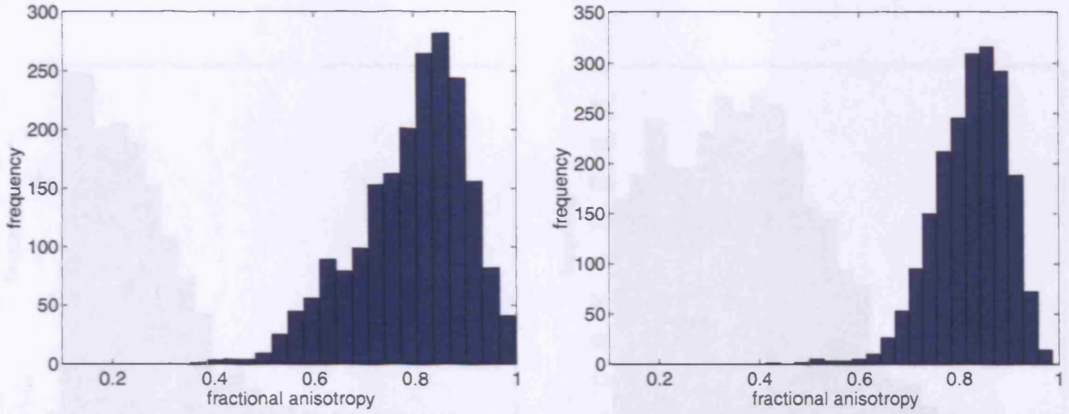


(c) Inversion 3, $f = 0.4$.

(d) Inversion 4, $f = 0.4$.

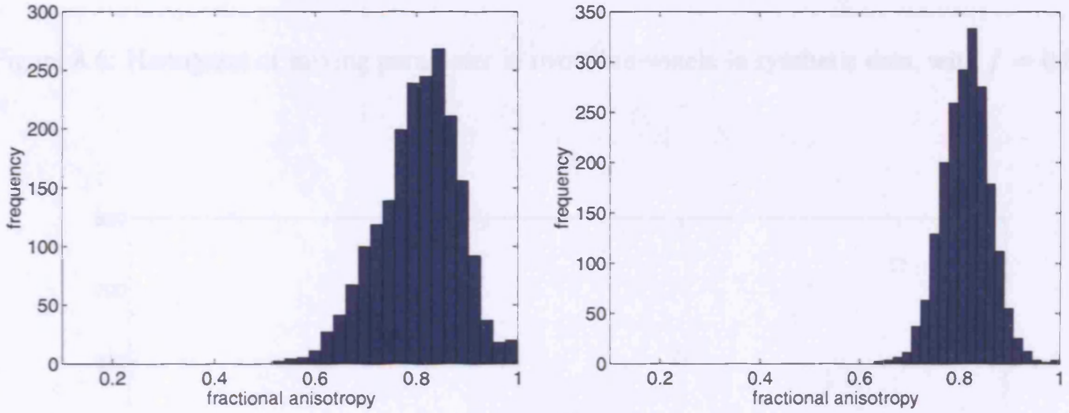
Figure 8.4: Histogram of f estimated from a Gaussian test function $\frac{1}{2}G(\mathbf{D}_1, t) + \frac{1}{2}G(\mathbf{D}_2, t)$ with $f = 0.4$ for both tensors.

$(1 - \alpha)G(\mathbf{D}_2, t)$, $f = 0.9$ for both tensors, $\mathbf{D}_1 = \text{diag}[\lambda_1, \lambda_2, \lambda_2]$ and $\mathbf{D}_2 = \text{diag}[\lambda_2, \lambda_2, \lambda_1]$. We vary α between 0.5 and 0.8. We recover the fibre orientations from the data using inversions 3 and 4. We fit the two-Watson PDF to the fibre orientations estimated from inversion 3. We use the estimated anisotropy of the tensors (as calculated using inversion 4) at each trial to retrieve the concentration from the Watson PDF calibrated with $\alpha = 0.5$ (the calibration procedure is described in Sec. 8.7). Fig. 8.7 shows the concentration of each compartment plotted against the mixing parameter. As the mixing parameter departs from 0.5, the calibrated concentration of fibre orientation estimates from the larger compartment is an underestimation. However, the anisotropy of the smaller compartment is underestimated, which means that the concentration is approximately correct. In practice, the error due to uneven mixing is limited, because as α



(a) Inversion 1, $f = 0.8$.

(b) Inversion 2, $f = 0.8$

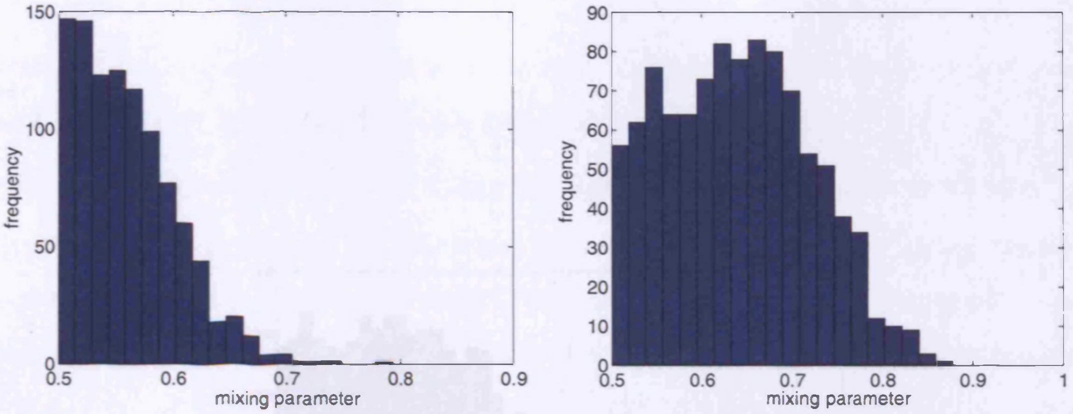


(c) Inversion 3, $f = 0.8$.

(d) Inversion 4, $f = 0.8$

Figure 8.5: Histogram of f estimated from a Gaussian test function $\frac{1}{2}G(\mathbf{D}_1, t) + \frac{1}{2}G(\mathbf{D}_2, t)$ with $f = 0.8$ for both tensors.

approaches 1, the smaller compartment becomes increasingly difficult to detect. Fig. 8.8 shows the histogram of mixing parameters found in the average brain (the average of eight repeated acquisitions) using inversions 1 and 3. Using inversion 3, approximately 50% of all two-fibre voxels have a mixing parameter of greater than 0.6, in these voxels the concentration of the larger compartment will be underestimated on average by the LUT.



(a) Inversion 3.

(b) Inversion 1.

Figure 8.6: Histogram of mixing parameter in two-fibre voxels in synthetic data, with $f = 0.8$.

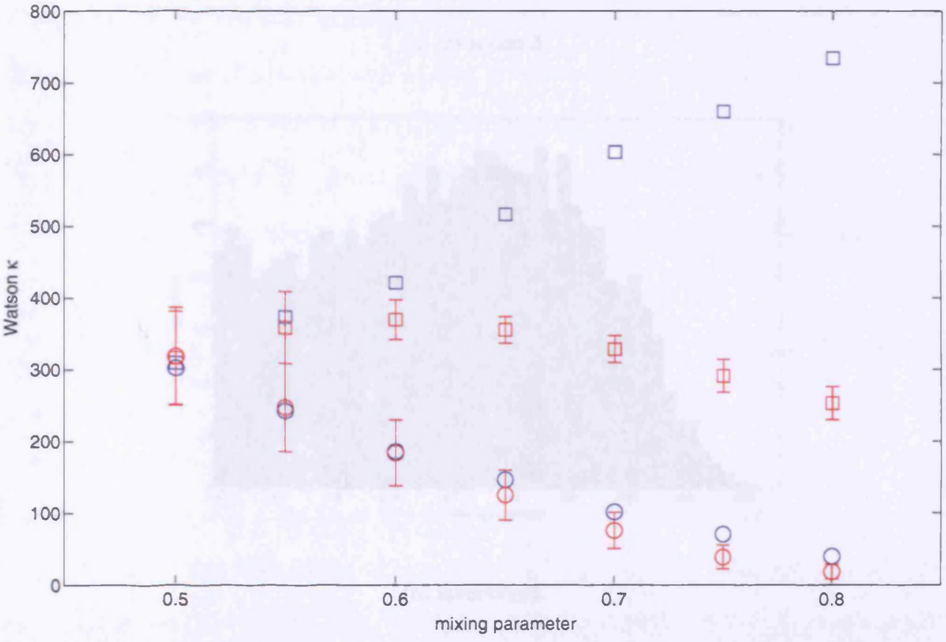
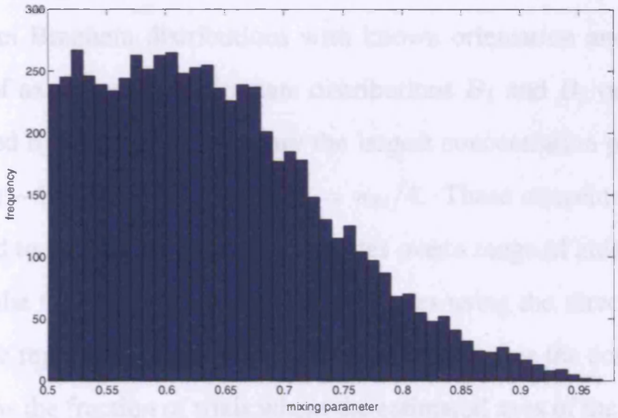
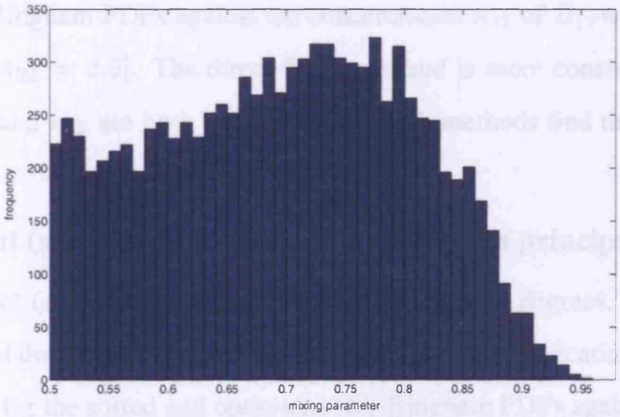


Figure 8.7: Concentration of fibre orientations as a function of the the mixing parameter in experiment (xiii). The larger compartment (squares) has higher concentration than the smaller compartment. The blue series is the concentration of fibre-orientations fitted directly to the data. The red series is the predicted concentration from a calibration procedure where the mixing parameter is assumed to be 0.5.



(a) Inversion 3.



(b) Inversion 1.

Figure 8.8: Histogram of mixing parameter in two-fibre voxels in the average brain.

8.5 Robustness of direct fitting and iterative sorting with Bingham-distributed axes

In this section we compare the direct fitting of the models to the iterative sorting method using axes sampled from Bingham distributions with known concentration.

8.5.1 Experiment (xiv): direct fitting and sorting with unequal concentration

Experiment (xiv) investigates the differences between the direct fitting and sorting methods in data sampled from Bingham distributions with known orientation and concentration. We sample 1000 pairs of axes from two Bingham distributions B_1 and B_2 oriented such that $\boldsymbol{\mu}_{11}$ and $\boldsymbol{\mu}_{12}$ are separated by 90 degrees. We vary the largest concentration parameter κ_{1i} of each distribution between -10 and -400, with $\kappa_{2i} = \kappa_{1i}/4$. These concentration parameters are similar to those fitted to the fibre-orientation estimates over a range of anisotropy from $f = 0.4$ to $f = 0.9$. We fit the two-Bingham model to these axes using the direct fitting and iterative sorting approach. We repeat the entire process 50 times, and define the consistency fraction for each fitting method as the fraction of trials where the estimated axes of the distributions are less than 3 degrees from their true orientation and the estimated concentration of each distribution differs from the true value by less than 5%. Fig. 8.9 plots the consistency fraction for the sorted and optimised two-Bingham PDFs against the concentration κ_{11} of B_1, with the concentration of B_2 $[\kappa_{12} = -10, \kappa_{22} = 2.5]$. The direct fitting method is more consistent than the sorting method. When κ_{11} and κ_{12} are both less than -20, both methods find the correct parameters for all 50 trials.

8.5.2 Experiment (xv): Acute crossing angles between principal directions

We repeat experiment (xiv) with $\boldsymbol{\mu}_{11}$ and $\boldsymbol{\mu}_{12}$ separated by 60 degrees. Reducing the angle between the principal directions increases the potential for misclassification. Fig. 8.10 plots the consistency fraction for the sorted and optimised two-Bingham PDFs against the concentration κ_{11} of B_1, with the concentration of B_2 $[\kappa_{12} = -10, \kappa_{22} = 2.5]$. Direct fitting is still highly consistent. The sorting method is less consistent as the concentration of B_1 increases, because the misclassifications cause a relatively larger error in highly concentrated distributions. When κ_{11} and κ_{12} are less than -20, both methods find the correct parameters for all 50 trials.

8.6 Direct fitting and sorting of fibre orientations from synthetic MRI data

According to experiments (xiv) and (xv), the direct fitting method is better than the sorting method at estimating the concentration of Bingham-distributed data at low concentration. The

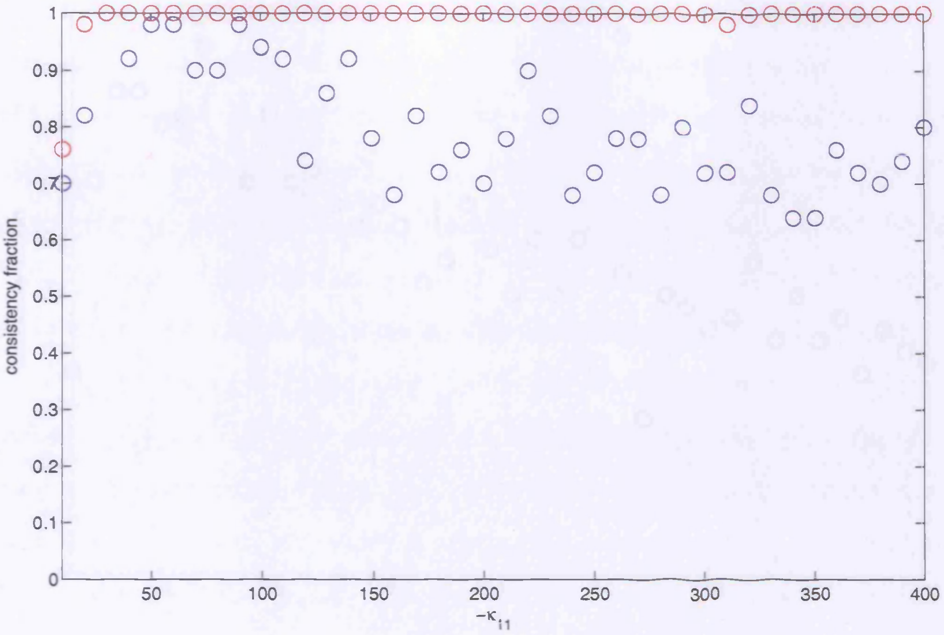


Figure 8.9: Consistency fraction in experiment (xiv). The optimised model fitting (red) finds the correct PDF parameters more often than the sorting algorithm (blue) at low concentration.

experiments in this section test whether the direct fitting is better when fitted to axes from synthetic MRI data.

8.6.1 Experiment (xvi): Reproducibility of concentration parameters

This experiment tests the stability of the direct fitting and sorting techniques among different sets of fibre orientations generated from the same test function. The test function is $p = 0.5[G(\mathbf{D}_1, t) + G(\mathbf{D}_2, t)]$ with $\mathbf{D}_1 = \mathrm{diag}[\lambda_1, \lambda_2, \lambda_2]$ and $\mathbf{D}_2 = \mathrm{diag}[\lambda_2, \lambda_2, \lambda_1]$. We generate 2000 noisy data sets and extract the fibre orientation estimates. We fit the PDF parameters by optimising the two-Watson and two-Bingham models, and fitting the Watson and Bingham PDFs to each set of fibre orientations after iterative sorting. We repeat data synthesis and fitting 30 times. At each iteration we reorient the test function with one random rotation of both diffusion tensors. Fig. 8.11 shows the mean and standard deviation of the concentration of the first compartment, as a function of the anisotropy of the tensors. Both the optimised and sorted PDF parameters are similar, even at $f = 0.4$ where the anisotropy is low. The reproducibility of the concentration is high for the Bingham and also for the Watson PDF, even though the axes are grouped in elliptical clusters. We conclude that both methods are reliable for finding the concentration of fibre orientations from this test function.

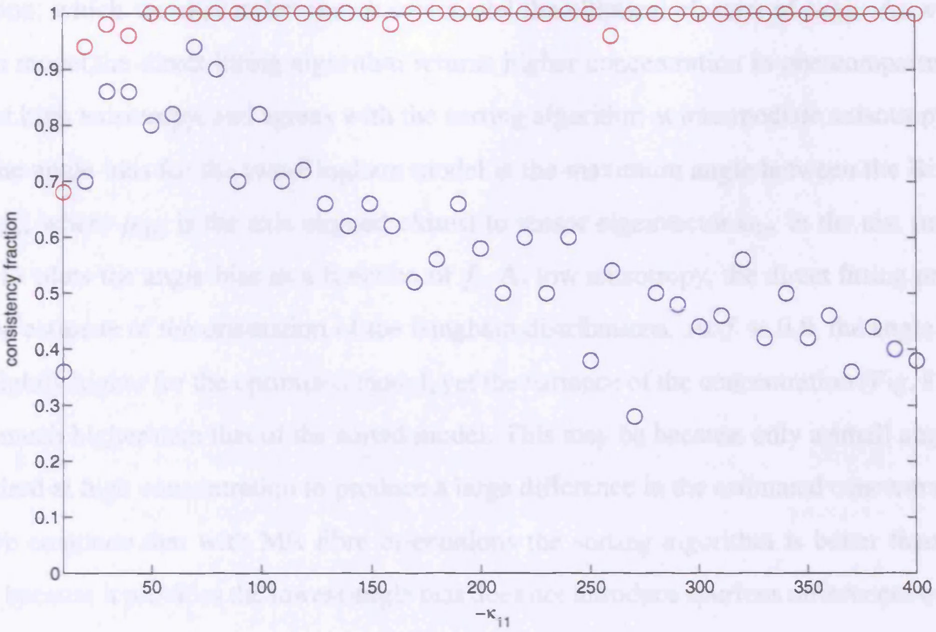


Figure 8.10: Consistency fraction in experiment (xv). The optimised model fitting (red) finds the correct PDF parameters more often than the sorting algorithm (blue) at low concentration.

8.6.2 Experiment (xvii): Reproducibility of concentration at an acute crossing angle

We repeat experiment (xvi) with D_2 rotated such that the angle between the principal directions e_{11} and e_{12} is 60 degrees. The clusters of fibre orientations are less well separated, which may cause the PDF fitting to fail. Fig. 8.12 shows the concentration for the Watson and Bingham models found by direct fitting and iterative sorting. The sorted Watson PDF (Fig. 8.12(a)) has low concentration for all values of f, and at $f = 0.9$ the sorting algorithm returns a negative κ. The optimised Watson PDF has similar concentration except at $f = 0.4$ and $f = 0.9$. It appears that the variance of κ_1 is high, but in fact, the direct fitting systematically assigns a high concentration to one compartment and a very low concentration to the other. This happens even if the orientation of the PDFs is fixed during the optimisation. The sorting algorithm does not suffer from this problem at low anisotropy, but at $f = 0.9$ it fits a negative κ. At intermediate anisotropy the Watson concentration is very low. We conclude from this result that the Watson model is a poor approximation of the fibre orientation PDF at acute crossing angles. At high anisotropy with orthogonal tensors, the fibre orientation estimates resolve into approximately circular clusters, as we see in experiment (xii). With a crossing angle of 60 degrees the clusters of axes maintain an elliptical shape when the tensors are highly anisotropic.

Fig. 8.12(b) shows the concentration of the two-Bingham model, which has higher concentration, which we expect, because it can model the elliptical clusters of axes. As with the Watson model,the direct fitting algorithm returns higher concentration in one compartment at low and high anisotropy, and agrees with the sorting algorithm at intermediate anisotropy.

The angle bias for the two-Bingham model is the maximum angle between the Bingham axis μ_{3n}, where μ_{3n} is the axis aligned closest to tensor eigenvector $\mathbf{e}_{1n}$ in the test function. Fig. 8.13 plots the angle bias as a function of f. At low anisotropy, the direct fitting provides a biased estimate of the orientation of the Bingham distributions. At $f = 0.9$, the angle bias is only slightly higher for the optimised model, yet the variance of the concentration (Fig. 8.12(b)) is still much higher than that of the sorted model. This may be because only a small angle bias is required at high concentration to produce a large difference in the estimated concentration.

We conclude that with MR fibre orientations the sorting algorithm is better than direct fitting, because it provides the lowest angle bias does not introduce spurious differences between the concentrations of identical compartments at acute crossing angles.

8.7 Calibration of models

The two-tensor PDFs are calibrated by constructing LUTs that give the concentration of the PDF as a function of tensor anisotropy and of the angle between the principal directions, given the imaging scheme and the signal to noise ratio in the data. The calibration algorithm is:

```
for ($f_1 = f_{\min}; f_1 \leq f_{\max}; f_1 = f_1 + \delta_f$) for ($\theta = \theta_{\min}; \theta \leq \theta_{\max}, \theta = \theta + \delta_\theta$)
    1. for ($f_2 = f_{\min}; f_2 \leq f_1; f_2 = f_2 + \delta_f$)
        (a) Calculate tensor $\mathbf{D}_{s1}$ with anisotropy $f_1$, and $\mathbf{D}_{s2}$ with anisotropy $f_2$.
        (b) Rotate $\mathbf{D}_{s2}$ by $\theta$ degrees about the second eigenvector.
        (c) For each trial $i = 1 : T$
            i. Synthesise noisy data from Gaussian compartment model.
            ii. extract fibre-orientation estimates $\mathbf{x}_1, \mathbf{x}_2$.
    2. Fit the model of the fibre orientation PDF to $\mathbf{x}_i \ldots \mathbf{x}_{2T}$.
    3. Record the parameters of the PDF in the LUT.
```

Since we use the restricted inversion, we do not consider the effects of the mixing parameter, or of non cylindrically-symmetric tensors.

The experiments in this section evaluate the range of anisotropy and crossing angles where we can resolve two fibre populations given the imaging parameters of the brain data. Parker and Alexander [56] find that the concentration of fibre orientations in one compartment is independent of the orientation and anisotropy of the other compartment. They fit a fourth-order spherical harmonic series to the data in two-fibre voxels, and re-sample from the series to obtain a larger number of measurements, with inherent smoothing from the spherical harmonic model. They fit the full (13 parameter) two-tensor model to the resampled data. We use the restricted two-tensor model and do not use spherical harmonic resampling. We evaluate the effects of the crossing angle and the anisotropy on the concentration in detail in experiments (xviii) and (xix).

8.7.1 Experiment (xviii): crossing angle

Experiment (xviii) tests how the concentration depends on the crossing angle for data sampled from a two-Gaussian test function. We generate 1000 noisy trials of the diffusion-weighted data using a two-Gaussian test function with $f = 0.9$ for both tensors, with crossing angles between 50 and 90 degrees. Fig 8.14 shows the concentration of fibre orientation estimates using the restricted inversion, and Fig. 8.15 shows the concentration of fibre orientation estimates from the full inversion. Both figures show the concentration found by iterative sorting. The concentration is dependent on the crossing angle when the test function is Gaussian, for both the restricted and the full two-tensor models. At crossing angles of 55 degrees or less, the sorting method assigns different concentrations to each compartment, suggesting that the fibre orientation estimates cannot be reliably resolved into two distinct clusters when the tensors cross at 55 degrees or less.

8.7.2 Experiment (xix): anisotropy of compartments

Parker and Alexander [56] model the concentration of fibre orientations from each compartment independently, assuming that the anisotropy of the other compartment has no effect on the concentration. Experiment (xix) tests whether this assumption is valid for data from a Gaussian test function. We generate 1000 noisy trials of the diffusion-weighted data using a two-Gaussian test function $p = 0.5[G(\mathbf{D}_1, t) + \mathbf{D}_2, t)]$, where $f(\mathbf{D}_1) = 0.9$, and $f(\mathbf{D}_2)$ varies between 0.3 and 0.9, and the two tensors cross at 90 degrees. We also generate 1000 trials from the test function $p = 0.5[G(\mathbf{D}_2, t) + \mathbf{D}_3, t)]$, where $f(\mathbf{D}_2) = f(\mathbf{D}_3)$ and f varies between 0.4 and 0.9. and the two tensors also cross at 90 degrees. Fig. 8.16 shows the concentration of each compartment plotted against $f(\mathbf{D}_2)$. We conclude that the concentration of orientations is dependent on the anisotropy of the other compartment.

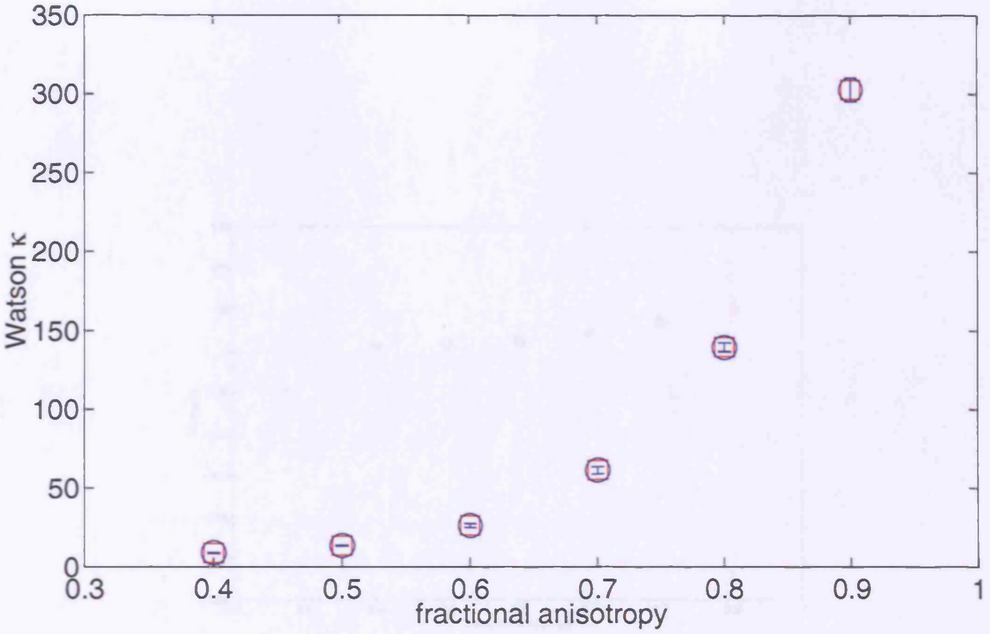
8.7.3 LUT Generation procedure

We examine the range of LUT parameters f and θ in the average brain (the mean of all 8 repeated acquisitions of the subject) to guide the choice of LUT range. Experiments (xx) and

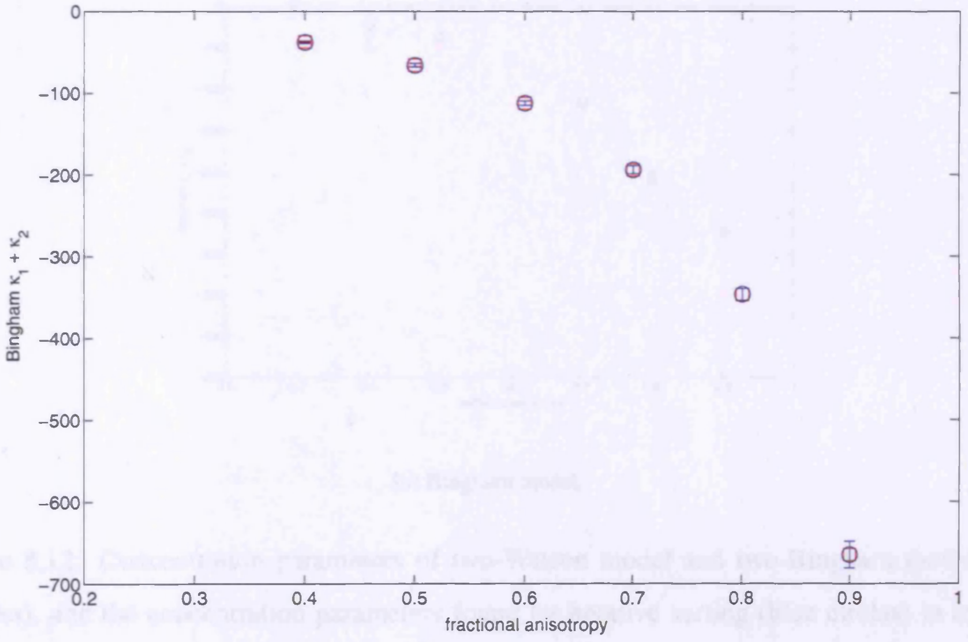
(xxi) test different resolutions and sample sizes for the calibration process. Fig. 8.17 shows a histogram of estimated anisotropy in two-fibre voxels in the average brain.

We set $f_{\mathrm{min}} = 0.3$; according to the LUTs, the concentration of two tensors of $f = 0.3$ crossing at 90 degrees is low, $\kappa = 6$ for the Watson PDF and $\kappa_1 = -16$, $\kappa_2 = -3$ for the Bingham PDF. According to the histogram of anisotropy in Fig. 8.17, 90% of the tensors have $f \geq 0.3$. We set $f_{\mathrm{max}} = 0.94$, which corresponds to a ratio of $\lambda_1/\lambda_3 = 18$. This is slightly more anisotropic than the maximum anisotropy observed in single-tensor voxels (Fig. 7.1). Approximately 10% of the tensors in two-fibre voxels have $f > 0.94$, but this is almost certainly an artefact of the two-tensor fitting procedure and not a physical effect, since no single-tensor voxels have anisotropy this high.

We set the angular range $\theta_{\mathrm{min}} = 50$ and $\theta_{\mathrm{max}} = 90$ degrees. Fig. 8.18 shows the histogram of crossing angles in the mean brain, 82% of two-fibre voxels have a crossing angle greater than 50 degrees and experiment xviii suggests that the concentration of fibre orientations at smaller angles is close to zero. There is probably a bias in the histogram towards larger crossing angles, because as the crossing angle tends to zero, the compartments become harder to distinguish.

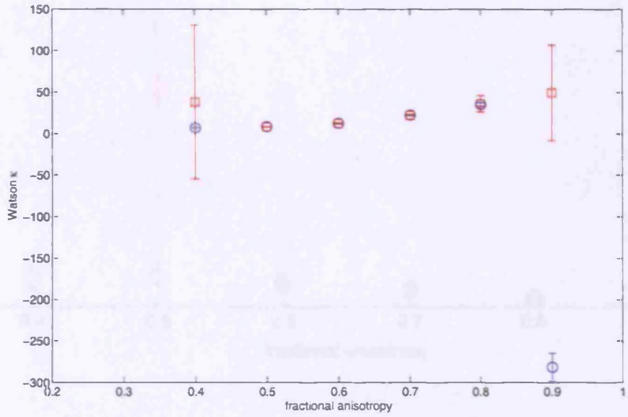


(a) Watson model.

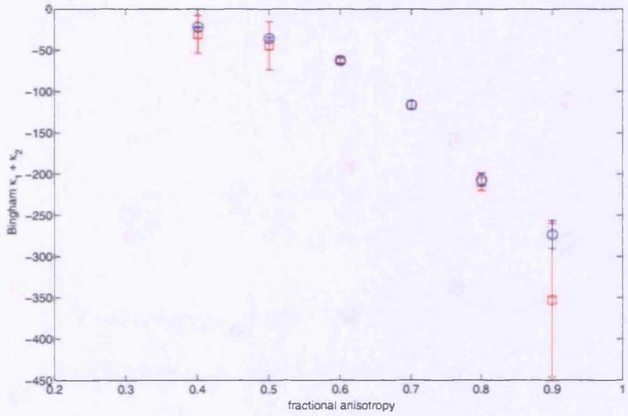


(b) Bingham model.

Figure 8.11: Concentration parameters of two-Watson model and two-Bingham model (red squares), and the concentration parameters found by iterative sorting (blue circles) in experiment (xvi).



(a) Watson model.



(b) Bingham model.

Figure 8.12: Concentration parameters of two-Watson model and two-Bingham model (red squares), and the concentration parameters found by iterative sorting (blue circles) in experiment (xvii).

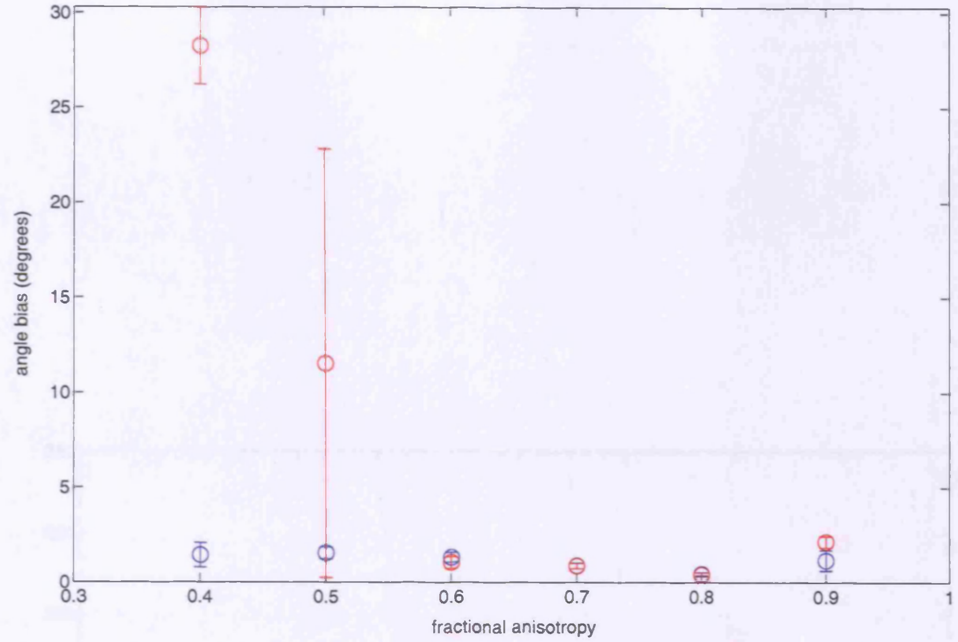


Figure 8.13: Angle bias for the optimised (red) and sorted (blue) two-Bingham models in experiment (xvii).

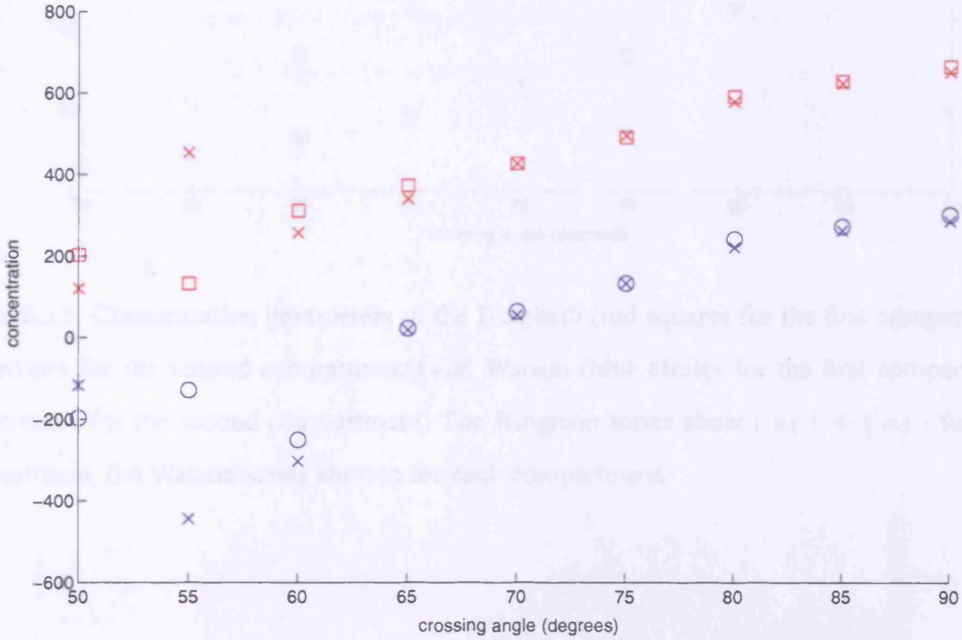


Figure 8.14: Concentration parameters of the Bingham (red squares for the first compartment, red crosses for the second compartment) and Watson (blue circles for the first compartment, blue crosses for the second compartment) PDFs for two compartments, using the restricted inversion. The Bingham series show $|\kappa_1| + |\kappa_2|$ for each compartment, the Watson series show κ for each compartment.

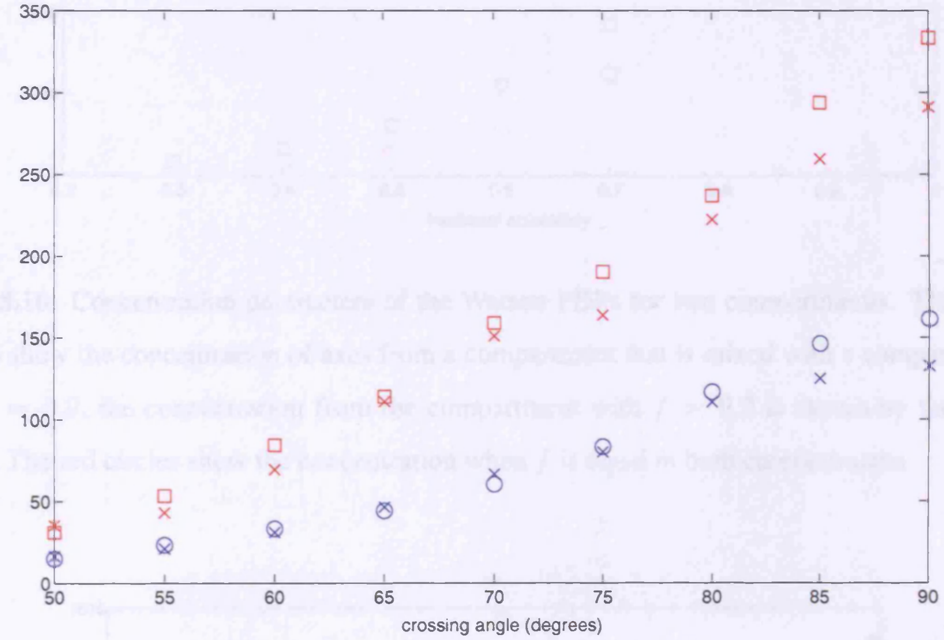


Figure 8.15: Concentration parameters of the Bingham (red squares for the first compartment, red crosses for the second compartment) and Watson (blue circles for the first compartment, blue crosses for the second compartment) The Bingham series show $\mid \kappa_1 \mid + \mid \kappa_2 \mid$ for each compartment, the Watson series show κ for each compartment.

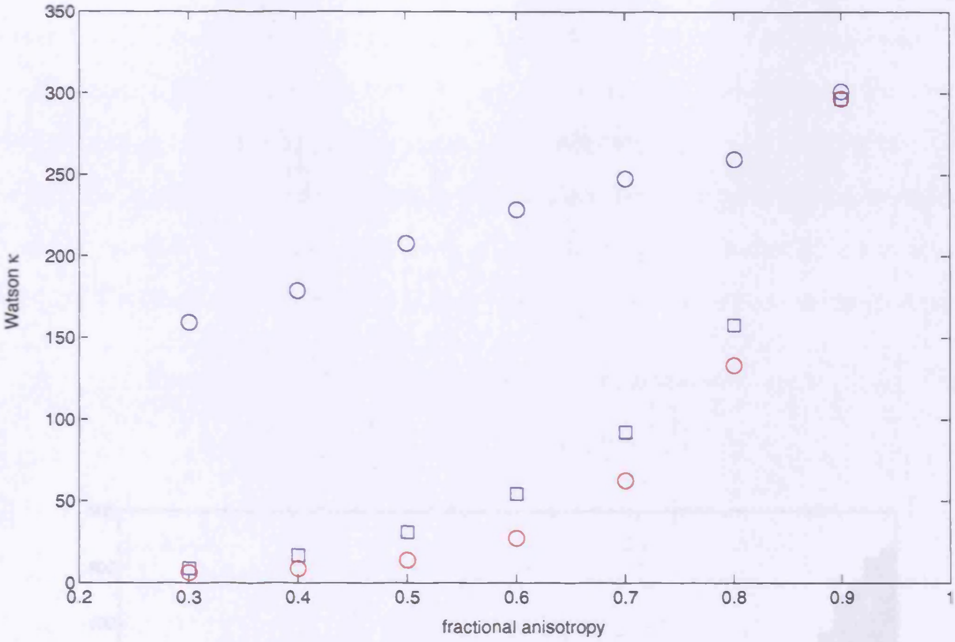


Figure 8.16: Concentration parameters of the Watson PDFs for two compartments. The blue squares show the concentration of axes from a compartment that is mixed with a compartment with $f = 0.9$, the concentration from the compartment with $f = 0.9$ is shown by the blue circles. The red circles show the concentration when f is equal in both compartments.

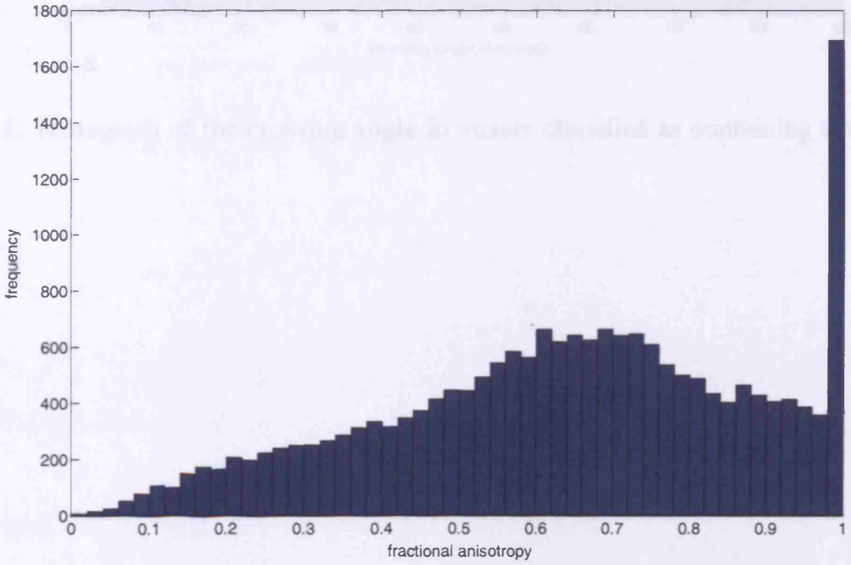


Figure 8.17: Histogram of tensor anisotropy in voxels classified as containing two compartments.

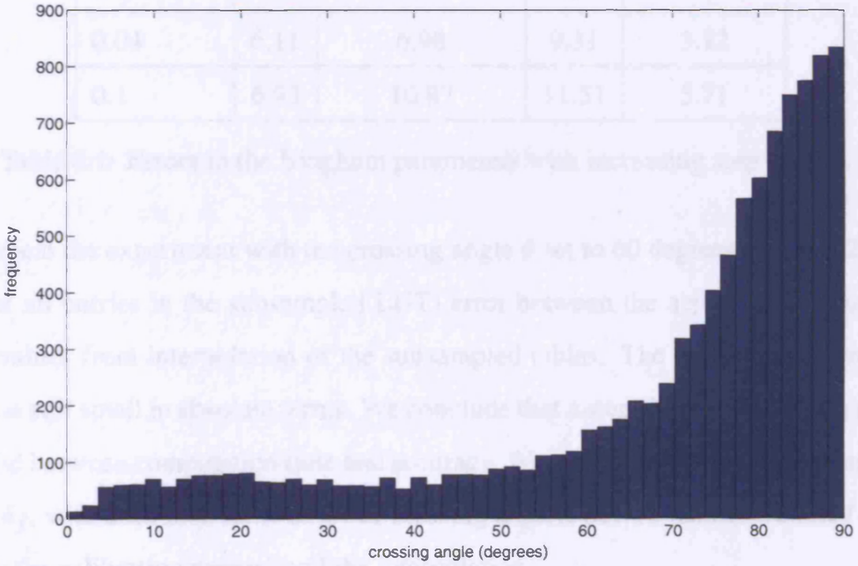


Figure 8.18: Histogram of the crossing angle in voxels classified as containing two compartments.

8.7.4 Experiment (xx): two tensor LUT resolution

We generate a LUT using the algorithm in Sec. 8.7 with $f_{\min} = 0.3$, $f_{\max} = 0.94$ and $\theta_{\min} = \theta_{\max} = 90$ degrees. We use bilinear interpolation to estimate the concentration for any f_1 and f_2 within the range. We construct a Bingham LUT with step size $\delta_f = 0.01$, with $T = 2000$ trials at each point in the table, and sub-sample the table in both dimensions at intervals of 0.02, 0.04, and 0.1. Table 8.1 shows the mean (over all entries in the subsampled LUT) error between the actual LUT values and the predicted values from interpolation of the subsampled tables.

Resolution	mean percentage error		mean absolute error	
	κ_1	κ_2	κ_1	κ_2
0.02	4.80	5.39	7.21	2.40
0.04	6.11	6.98	9.31	3.12
0.1	6.93	10.87	11.51	5.71

Table 8.1: Errors in the Bingham parameters with increasing step size δ_f.

We repeat the experiment with the crossing angle θ set to 60 degrees. Table 8.2 shows the mean (over all entries in the subsampled LUT) error between the actual LUT values and the predicted values from interpolation of the subsampled tables. The error from interpolation is larger, but is still small in absolute terms. We conclude that a step size $\delta_f = 0.02$ is a reasonable compromise between computation time and accuracy. We could save computation time by using a variable δ_f, with a smaller value at lower crossing angles, but we maintain a fixed $\delta_f = 0.02$ to simplify the calibration process and the interpolation.

Resolution	mean percentage error		mean absolute error	
	κ_1	κ_2	κ_1	κ_2
0.02	6.73	5.31	6.69	1.06
0.04	8.56	7.18	8.56	1.47
0.1 10.07	11.06	10.52	12.39	2.19

Table 8.2: Errors in the Bingham parameters with increasing step size δ_f.

8.7.5 Experiment (xxi): two tensor LUT sample size

Experiment (xxi) tests the variance of the estimated concentration as a function of the number of trials T used to fit the two-Watson and two-Bingham models. We generate LUTs according to the algorithm in Sec. (8.7) with $f_{\min} = 0.3$, $f_{\max} = 0.9$, $\delta_f = 0.05$ and $\theta_{\min} = \theta_{\max} = 90$ degrees. As a baseline we compute a LUT with $T = 5000$ trials. Table 8.3 shows the mean (over all entries in the subsampled LUT) error between the baseline concentration and those calculated with smaller T. We conclude that $T = 2000$ is a reasonable sample size.

T	mean percentage error		mean absolute error	
	κ_1	κ_2	κ_1	κ_2
500	9.98	12.98	13.39	4.59
1000	7.36	8.88	15.07	3.00
2000	6.27	6.38	7.99	2.35
4000	4.87	5.95	7.70	2.28
5000	4.22	4.73	5.81	1.61

Table 8.3: Errors in the Bingham parameters with increasing numbers of trials in the calibration.

Given the parameters $f_{\text{min}} = 0.3$, $f_{\text{max}} = 0.94$ $\delta_f = 0.02$, $\theta_{\text{min}} = 50$, $\theta_{\text{max}} = 90$ and $T = 2000$, it remains to choose the angular step size δ_θ for the generation of the LUTs. Experiment (xviii) suggests that the concentration varies linearly as a function of the crossing angle for highly anisotropic tensors, but this is unlikely to be the case for all combinations of f_1 and f_2 we may find in a two-fibre voxel. We set $\delta_\theta = 5$ degrees, which is the smallest step that still allows the full LUT to be computed in less than 24 hours of computation time on a standard $2.8GHz$ PC workstation.

8.8 Analysis in synthetic data

We test the two-Bingham and two-Watson models in a synthetic fibre crossing, by comparing them to a gold standard PICo map as in Ch. 7. The test functions for the data are Gaussian, with cylindrically-symmetric tensors. A streamline seeded in the noise-free image traces out a path of length 80mm, from left to right. The test function for these voxels is $G(\mathbf{D}_1, t)$, where $\mathbf{D}_1$ is oriented along the x-axis. A perpendicular fibre pathway of width 14mm intersects the main path, the test function in these voxels is $G(\mathbf{D}_2, t)$, where $\mathbf{D}_2$ is oriented along the y-axis. The voxels in the intersection between the two paths are two-fibre voxels, and the test function for these voxels is $0.5[G(\mathbf{D}_1, t) + G(\mathbf{D}_2, t)]$. The anisotropy of both tensors is equal, and varies between $f = 0.4$ and $f = 0.9$.

The tractography algorithm is as described in Sec. 7.2 with one modification: in voxels where there are two tensors, we draw samples from both PDFs and the streamline tracks along the axis most closely aligned to the previous streamline trajectory.

8.8.1 Experiment (xxii): PICo maps in noise free data

Experiment (xxii) tests the models in the noise free data, as we did for the single-fibre PDFs in experiment (iv). We define a gold standard synthetic PICo map by tracking streamlines in 5,000 identical copies of the fibre crossing image. In each image, we add independent Gaussian noise to the DW measurements, fit the tensors, and then track a streamline from the designated seed point. The collection of streamlines define a PICo connection probability map, except that the fibre orientations at each iteration are estimated directly from the noisy data and not sampled from a PDF. We measure sum of squared differences S between PICo maps generated using the models (also with 5,000 iterations) and the gold standard.

Fig. 8.19 shows the SSD for the Bingham and Watson models as a function of the anisotropy of the tensors. The Bingham PDF has lower SSD to the gold standard when $f < 0.8$, but at high anisotropy, the clusters of fibre orientations are well modelled by both distributions.

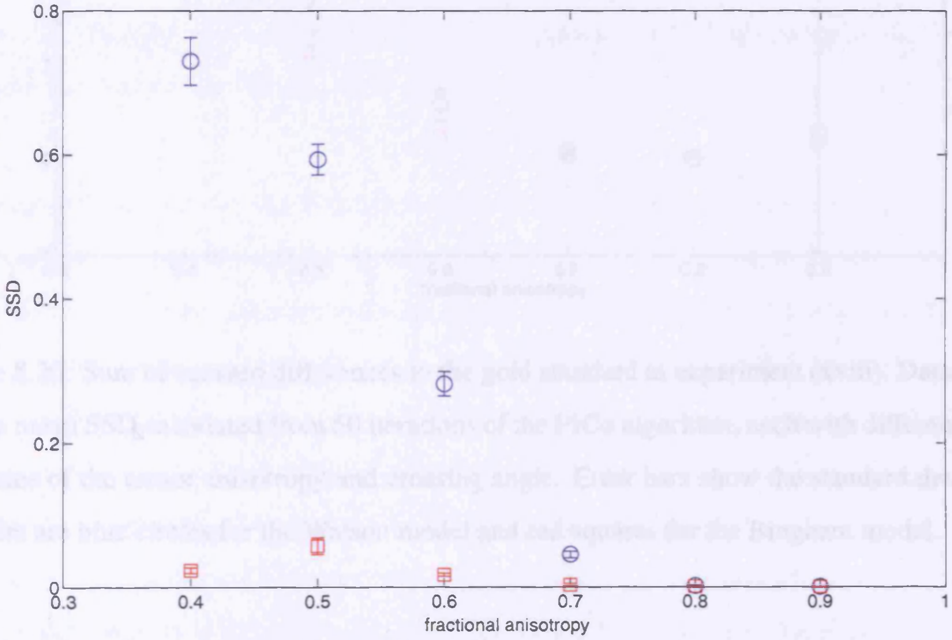


Figure 8.19: Sum of squared differences to the gold standard in experiment (xxii). Data points are the mean SSD calculated from 50 iterations of the PICo algorithm. Error bars show the standard deviation. Markers are blue circles for the Watson model and red squares for the Bingham model.

8.8.2 Experiment (xxiii): Noisy estimates of the PDF concentration

Experiment (xxiii) compares the two models with noisy estimates of the anisotropy and crossing angle. The estimate of the anisotropy and crossing angle is taken from noisy data, however the peaks of the PDFs remain fixed at the correct value. The experiment tests how the error in PDF concentration (which depends on the estimated crossing angle) affects each PDF, given the noise conditions of the data. Fig. 8.20 shows that the error to the gold standard is much larger for both models, but the Bingham model has lower errors. Fig 8.21 shows the statistic $\bar{d}(\text{Watson}, \text{Bingham})$ (defined in Sec. 7.3.2); the Bingham PDF has lower SSD at $f < 0.7$, but there is no significant difference at higher anisotropy.

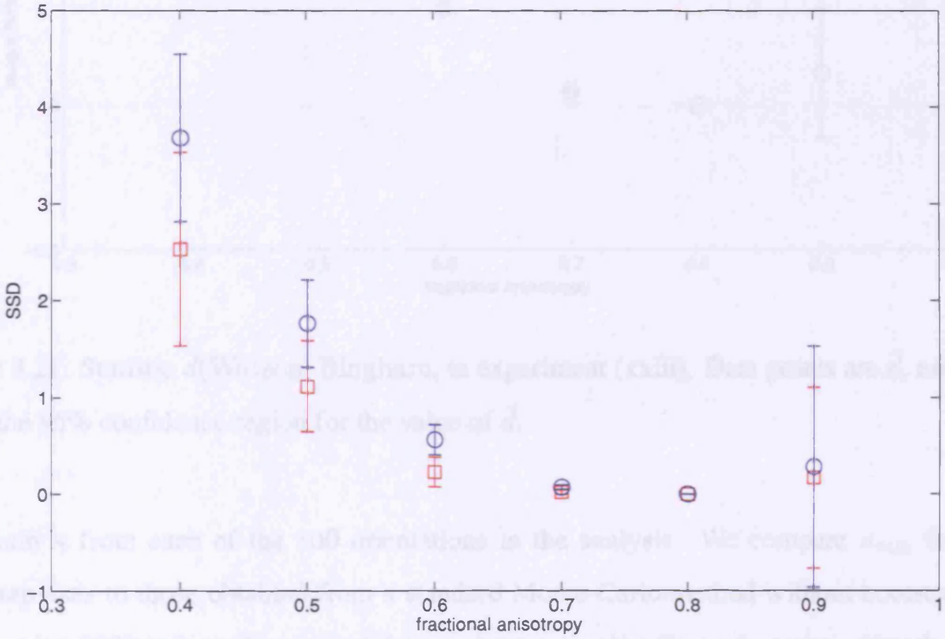


Figure 8.20: Sum of squared differences to the gold standard in experiment (xxiii). Data points are the mean SSD calculated from 50 iterations of the PICo algorithm, each with different noisy estimates of the tensor anisotropy and crossing angle. Error bars show the standard deviation. Markers are blue circles for the Watson model and red squares for the Bingham model.

8.9 Bootstrap simulation of two-fibre measurements

In experiment (xxiv), we show how the number of repeated acquisitions influences the two-fibre PDF parameters estimated from bootstrap resampling of the data.

For each of r repeats, $r = 2, 4, 6, \ldots, 16$, we synthesise r independent sets of noisy data from a Gaussian compartment test function, $p = 0.5[G(\mathbf{D}_1, t), G(\mathbf{D}_2, t)]$. The anisotropy of both tensors is $f = 0.8$, and $\mathbf{D}_2$ is oriented perpendicular to $\mathbf{D}_1$. We extract 1000 bootstrap resamples of the data, extract the fibre orientations and fit the Watson PDF. We repeat the experiment 100 times with the test function randomly reoriented after each iteration, giving us 100 estimates of κ_1 and κ_2 and the corresponding axes μ_1 and μ_2 for each r. We use the

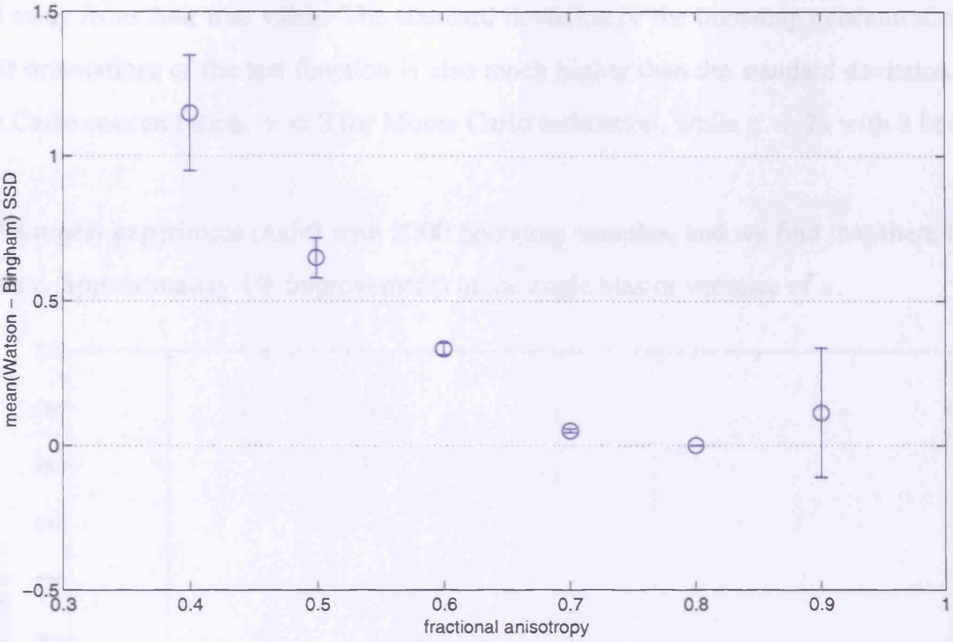


Figure 8.21: Statistic $\bar{d}$(Watson, Bingham, in experiment (xxiii). Data points are $\bar{d}$, error bars show the 95% confidence region for the value of $\bar{d}$.

minimum κ from each of the 100 orientations in the analysis. We compare $\kappa_{\min}$ from the bootstrap axes to those obtained from a standard Monte-Carlo method without bootstrapping, i.e. by using 1000 independent sets of data, and extracting the fibre orientations directly.

Fig. 8.22 shows the mean and standard deviation of $\min(\kappa)$ as a function of the number of repeats. Fig. 8.23 shows the "angle bias", which is the angle between the mean of the bootstrap fibre orientations and the known orientation of the tensors. Since there are two principal directions, we define the angle bias as the largest of the two angles between a mean axis μ and the closest tensor principal direction. The concentration is not consistently overestimated with bootstrapping, but the variance of κ over the 100 orientations of the test function is higher than for the Monte-Carlo simulation (the standard deviation of the Monte-Carlo concentration is 5.4 over the 100 orientations). This suggests that bootstrap resampling exacerbates the bias that arises from the sparse sampling of diffusion along 61 gradient directions. Although the gradient directions are spread as evenly on the sphere as possible, 61 directions cover the sphere sparsely and there is inevitably better concentration at some orientations of the test function than at others.

The bootstrap results also have much larger angle bias, the mean angle bias is 0.07 degrees for the Monte-Carlo estimate and 1.8 degrees for 8 bootstrap repeats. The bootstrap estimate of κ is not an unbiased estimate of the true concentration, because the peaks of the PDFs are biased away from their true value. The standard deviation of the bootstrap concentration over the 100 orientations of the test function is also much higher than the standard deviation of the Monte Carlo concentration: $\sigma = 3$ for Monte Carlo estimation, while $\sigma = 24$ with 8 bootstrap repeats.

We repeat experiment (xxiv) with 2000 bootstrap samples, and we find that there is little difference (approximately 1% improvement) in the angle bias or variance of κ.

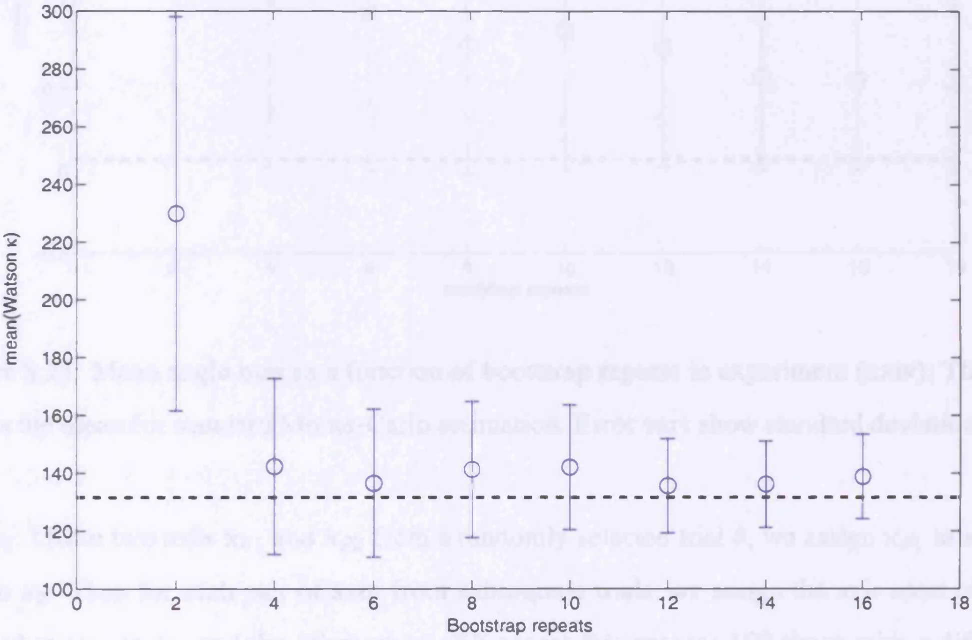


Figure 8.22: Mean Watson concentration as a function of bootstrap repeats in experiment (xxiv). The line shows the mean for standard Monte-Carlo estimation. Error bars show standard deviation.

8.10 Experiment (xxv): PDF fitting with an unknown test function

The PICo calibration algorithm uses a test function with known principal directions, these directions can be used to drive the fitting of the PDF models. When fitting the two-fibre models to fibre-orientations extracted from the bootstrap brain data, we have two clusters of axes, but the mean axis of each cluster is unknown. Experiment (xxv) tests the ability of iterative sorting algorithm to fit the PDF parameters when the true fibre orientation is unknown. We generate 1000 noisy trials of the diffusion-weighted data using a two-Gaussian test function, $p = 0.5[G(\mathbf{D}_1, t) + \mathbf{D}_2, t)]$, where $f(\mathbf{D}_1)$ varies between 0.4 and 0.9, $\mathbf{D}_1 = \mathrm{diag}[\lambda_1, \lambda_2, \lambda_2]$ and $\mathbf{D}_2 = \mathrm{diag}[\lambda_2, \lambda_2, \lambda_1]$.

We initialise the sorting algorithms by splitting the set of fibre orientations into two sets s_1

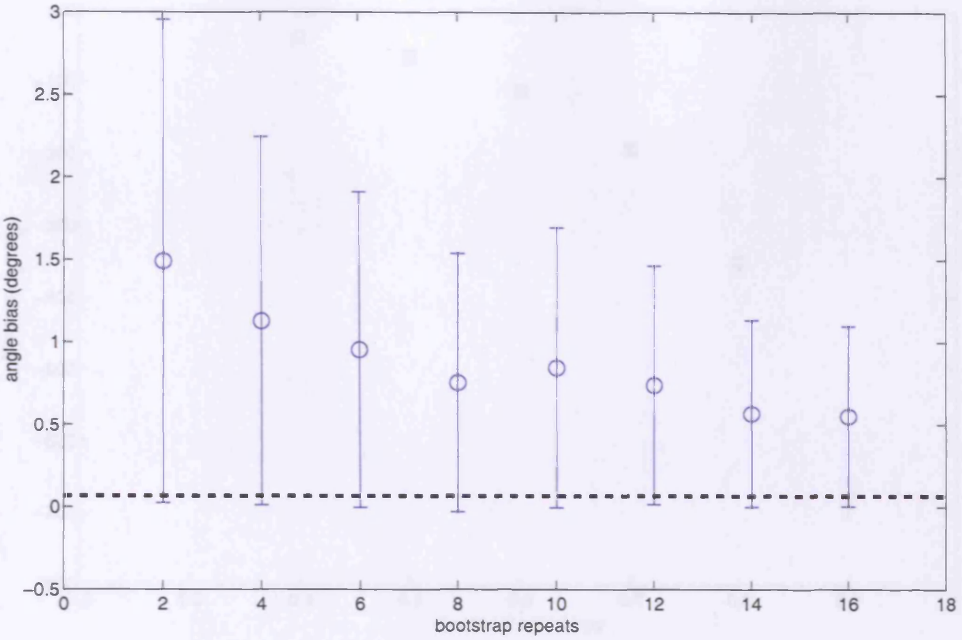


Figure 8.23: Mean angle bias as a function of bootstrap repeats in experiment (xxiv). The line shows the mean for standard Monte-Carlo estimation. Error bars show standard deviation.

and s_2. Given two axes $\mathbf{x}_{\theta 1}$ and $\mathbf{x}_{\theta 2}$ from a randomly selected trial θ, we assign $\mathbf{x}_{\theta 1}$ to s_1 and $\mathbf{x}_{\theta 2}$ to s_2. Then for each pair of axes from subsequent trials, we assign the axis most closely aligned to $\mathbf{x}_{\theta 1}$ to s_1, and the other to s_2. We repeat this process 100 times with a different starting trial θ.

Fig. 8.24 shows the mean and standard deviation of the concentration over the 100 iterations, compared to the concentration parameters we calculated using the same test function in experiment xvi), where the iterative algorithm is initialised using the principal directions of the test function rather than a pair of axes from one trial. The concentrations differ only slightly from experiment (xvi).

We repeat this experiment with $\mathbf{D}_2$ rotated such that the principal directions of the tensors cross at 60 degrees. Fig. 8.25 shows the mean and standard deviation of the concentration over the 100 iterations. The concentration is similar to that from experiment (xvii) except at $f = 0.8$ and $f = 0.9$, but even then the difference is small. Fig. 8.26 shows the angle bias. The bias is slightly higher than when the fibre orientations are known, for all the sampled points, but is much higher at $f = 0.4$, though it is still lower than the angle bias of the optimisation method (Fig. 8.13).

We conclude from experiment (xxv) that the iterative sorting algorithm can accurately find the concentration and orientation of the PDFs except in voxels with low anisotropy and an

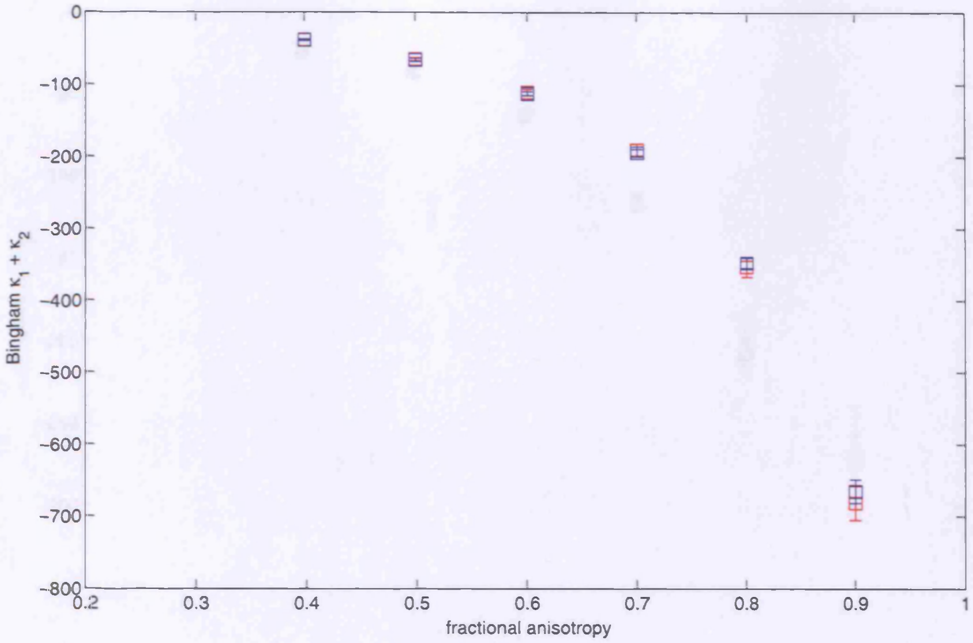


Figure 8.24: Concentration of two compartments calculated with the iterative sorting method, with different starting points in experiment (xxv) (red), compared to the concentration calculated from the same test function in experiment (xvi) (blue). The crossing angle between the two tensors is 90 degrees.

acute crossing angle, but these voxels have low concentration, so the effect of the angle bias on connection probabilities is probably small.

8.11 Experiment (xxvi): Two-fibre PDFs in brain data: LUTs and bootstrap

This experiment examines the correlation between the PDF concentration parameters fitted to the bootstrap data, against those predicted from the LUTs. A strong correlation between bootstrap and LUT concentration means that the anisotropy and crossing angle of the tensors are a good indication of the PDF concentration.

For every voxel classified as non-Gaussian in the average brain image, we generate 1000 bootstrap samples of the data, extract the 2000 fibre-orientation estimates and fit the parameters of the two-Watson and two-Bingham PDFs using the optimisation and the sorting method. We also compute two-fibre LUTs using the sorting method. We compare the Bingham concentration parameters to those taken from the LUTs using the anisotropy and crossing angle of the two tensors in the average brain image.

Fig. 8.28 shows the correlation between the bootstrap and LUT two-Bingham concentra-

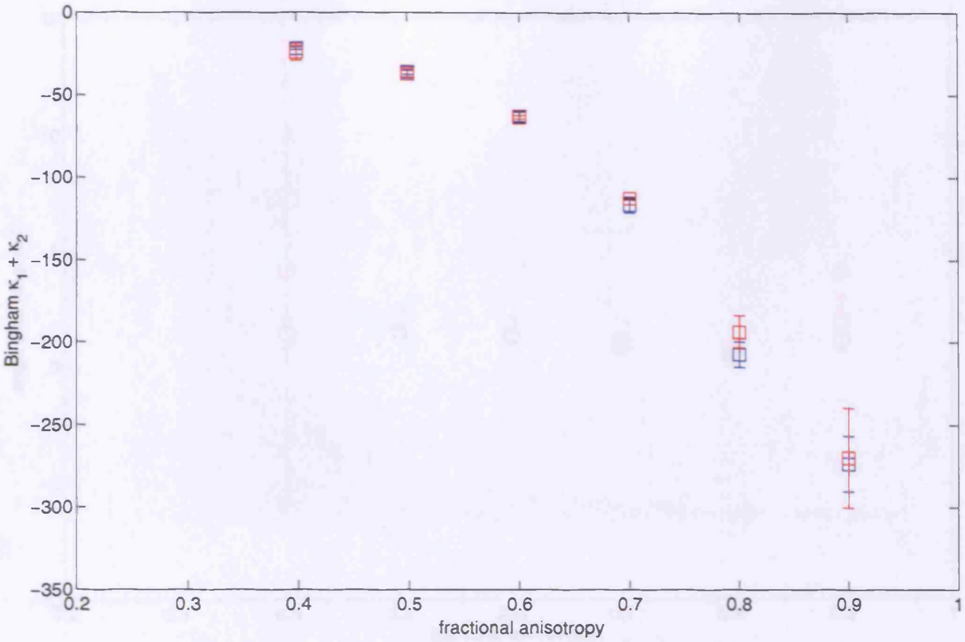


Figure 8.25: Concentration of two compartments calculated with the iterative sorting method, with different starting points in experiment (xxv) (red), compared to the concentration calculated from the same test function in experiment (xvii) (blue). The crossing angle between the two tensors is 60 degrees.

tion using the sorting method. There appears to be some correlation between the bootstrap and the LUT concentration, though there are some voxels where the LUT either overestimates or underestimates the concentration compared to the bootstrap result.

The differences between the two are probably caused by a combination of factors besides the underestimation of uncertainty in bootstrap resampling. As in the single-fibre case, there is an error in estimating the LUT indices (anisotropy and crossing angle) in each voxel, Also, some voxels may have crossing angles and anisotropy outside the range modelled by the LUT. Unequal mixing of the compartments is also not included in the calibration, experiment (xiii) suggests that the most noticeable effect of this is an underestimate of concentration in the larger compartments. Also, the assumption of cylindrical symmetry may not be justified. Fig. 8.27 shows the histogram of λ_2/λ_3 in the mean brain, computed with the full two-tensor model. Noise causes the measured λ_2/λ_3 to be greater than 1 even if the underlying test function is cylindrical, but in the brain there may be genuine cylindrical asymmetry.

More fundamentally, the LUT is calibrated with two-Gaussian compartment test functions, and this might not be a good approximation of the diffusion in the brain data. Also, the voxel classification may contain false-positives due to noise, and fit a fourth-order spherical harmonic

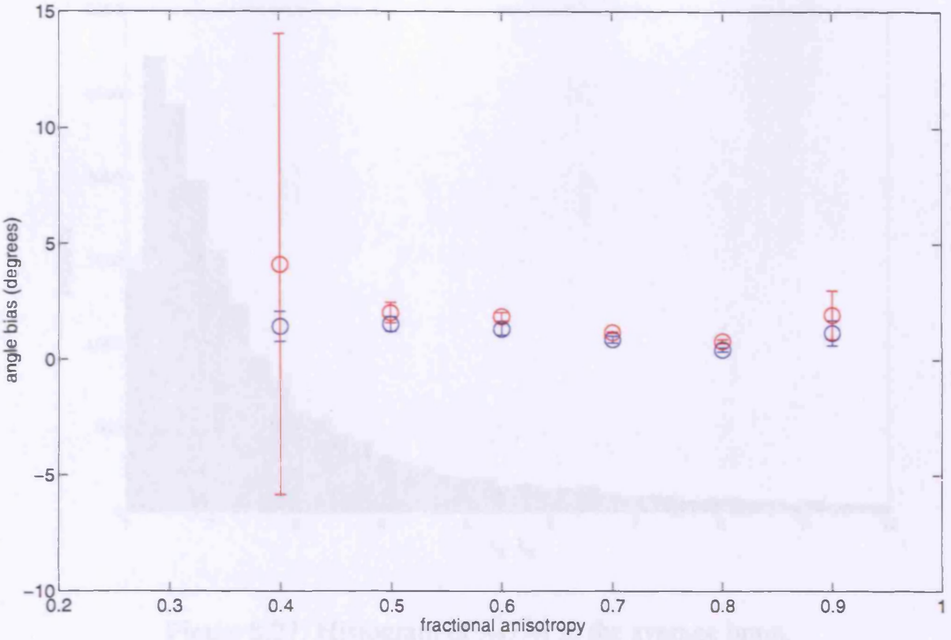


Figure 8.26: Angle bias in the iterative sorting method with different starting points in experiment (xxv), with a crossing angle of 60 degrees between the two tensors. The angle bias is higher when the true fibre orientations are unknown (red) than when they are known (blue).

in some voxels where there are not two distinct fibre orientations.

8.12 Tractography in brain data

In this section we compare PICo probability maps generated from bootstrap data to those produced by using the models and LUTs. We track in the same region of interest (ROI), across the corpus callosum of the subject, as we use in Ch. 7. The bootstrap tractography algorithm is the same as in Sec. 7.2.1, except that in voxels classified as containing two fibre orientations, we fit a restricted two-Gaussian compartment model and extract $\mathbf{e}_1$ from both tensors. The streamline then tracks along the tensor $\mathbf{e}_1$ that is most closely aligned to the previous streamline trajectory. We track 2000 streamlines from each seed point, which is twice as many as we use for the single-fibre case, since the fibre-orientation PDF is more complex than in the single-fibre case.

8.12.1 Experiment (xxvii): PICo calibrated directly from bootstrap fibre orientations

We compare the bootstrap tractography results to PICo tractography with the models calibrated directly from the bootstrap fibre orientations in each voxel. This experiment tests the goodness of fit of the models to the bootstrap fibre orientations. We track 2000 streamlines from each seed point in the ROI using the optimised two-Watson and two-Bingham models. Fig. 8.29

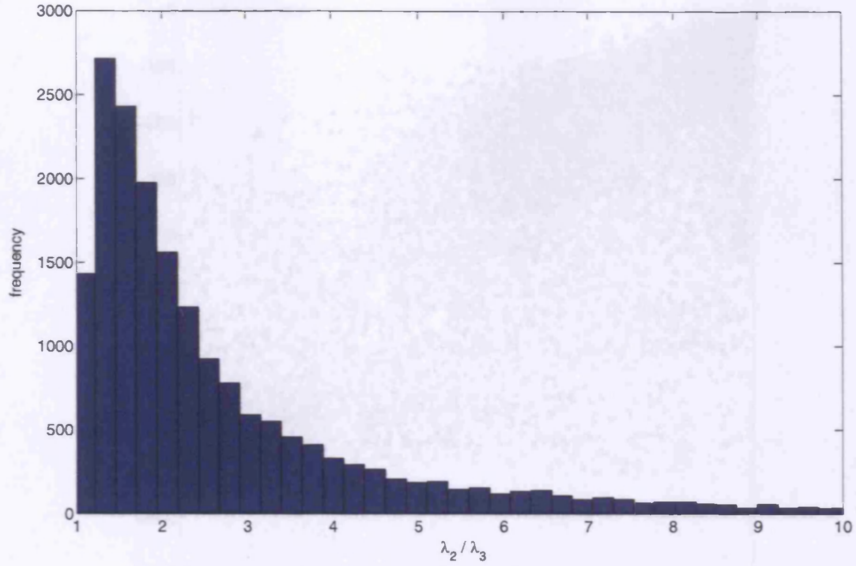


Figure 8.27: Histogram of λ_2/λ_3 in the average brain.

shows the SSD between the connection probability maps from each model and those from the bootstrap tracking. As in the single-fibre experiments, the Bingham results are more similar to the bootstrap tractography. It is not possible to quantify how much of the difference in the SSD is due the two-fibre PDFs, because the Watson and Bingham models are also different in the single-fibre voxels (experiment (x)).

The mean SSD for the Bingham connection probability images is 0.18 ± 0.2, which is larger than in the single-fibre case, (0.07 ± 0.03).

8.12.2 PICo calibrated by anisotropy

Experiment (xxviii) quantifies the effects of the error in PDF concentration introduced by a noisy estimate of the anisotropy. In a similar way to experiment (xi), we take the orientation of the PDFs fitted to the bootstrap data, but take the concentration from the LUT, indexed by the anisotropy and the crossing angle of the tensors fitted in the mean brain. Fig. 8.30 shows the SSD between the PICo PDFs and the bootstrap connection probabilities, when the concentration parameters are calculated from the tensors in the average brain. As in the single-fibre case, the difference between the Bingham and Watson models remains when the estimate of the anisotropy is noisy. The SSD of the Bingham model to the bootstrap data is 1.12 ± 0.92, compared to 0.57 ± 0.4 in the single-fibre case.

When the PICo concentration parameters are calculated from brain image 1, the SSD for the Bingham model is 1.48 ± 1.2, compared to 1.00 ± 0.95 in the single fibre case. Fig. 8.31 shows the difference between the bootstrap and Bingham connection probabilities for a

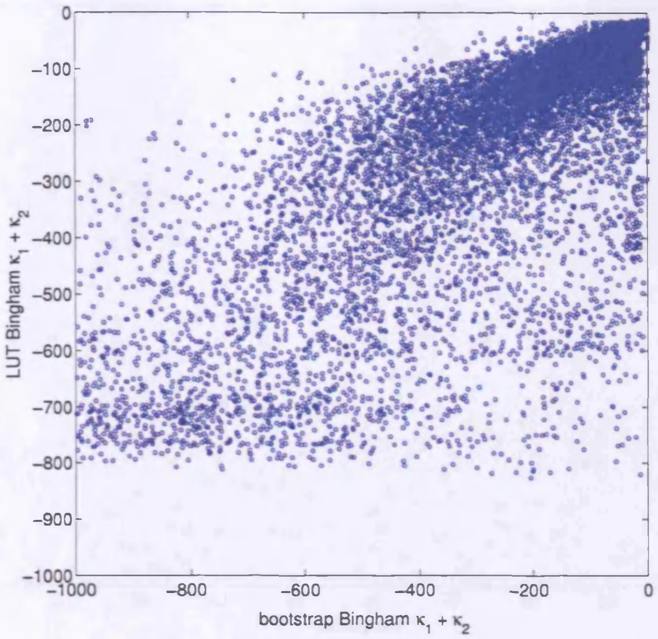


Figure 8.28: Bootstrap PDF concentration plotted against LUT concentration using the sorting method.

seed where the SSD is 1.45, close the mean value. The mean SSD for the Watson model is 2.93 ± 2.0 almost twice as large as for the Bingham model, but actually slightly smaller than the mean SSD for the Watson model in the single-fibre image (2.97).

8.13 Summary and conclusions

The calibration of PICo PDFs for the two-Gaussian compartment model is challenging because of the large number of parameters in the model and the difficulty of modelling two clusters of fibre-orientation estimates with a spherical PDF. Experiment (xii) compared four two-tensor fitting routines. The experiment tested the fitting routines on a two-Gaussian test function with an equal mixing parameter, with cylindrically symmetric tensors crossing at 90 degrees. Bearing in mind that we only tested one of many possible realistic test functions, we made the following observations. Restricting the mixing parameter in the fitting improves the estimate of the fractional anisotropy. This is important for PICo because the LUT is indexed by this statistic. Restricting fitting to cylindrically-symmetric tensors increases the concentration of fibre orientation estimates, but causes the clusters from each fibre orientation to become elliptical rather than circular, which requires a more complex model of the fibre-orientation PDF.

The choice of two-tensor fitting routine is also guided by the computational burden of the PICo calibration. The full 13-parameter two-tensor model is clearly unsuitable for PICo because

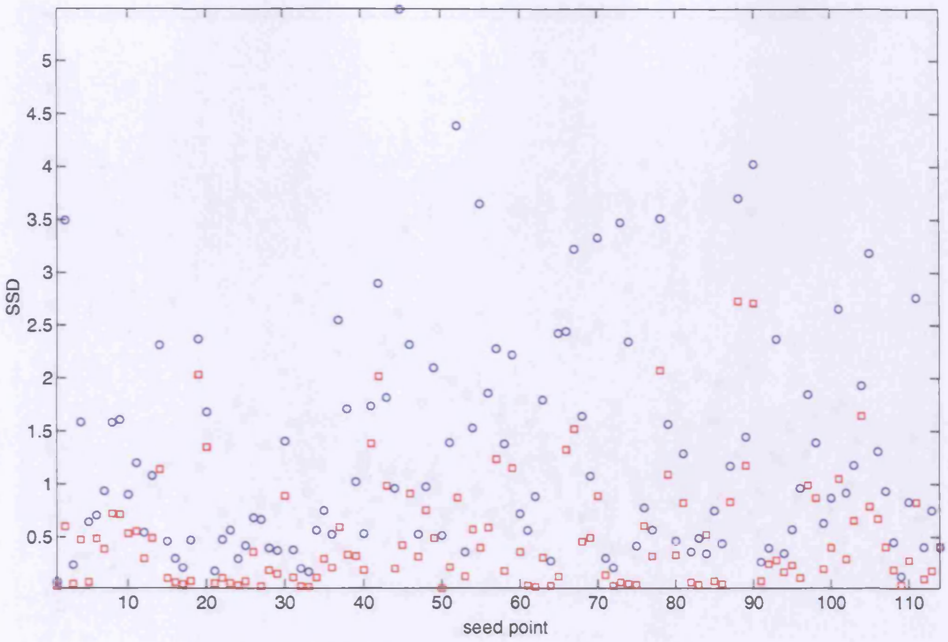


Figure 8.29: SSD between bootstrap tractography results and PICo using models calibrated from the bootstrap fibre orientations in experiment xxvii. The red squares are the SSD for the Bingham model, the blue circles are the SSD for the Watson model.

the estimated anisotropy and mixing parameter are too unstable to be reliable indices for the LUT and because it would require too many parameters in the calibration. With a fixed mixing parameter and cylindrically-symmetric tensors, the LUT has three parameters: the anisotropy of each tensor, and the angle between the principal directions. Experiments (xx) and (xxi) suggest that the calibration process requires approximately 24 hours on a standard PC workstation. We could potentially add the mixing parameter to the calibration, but we would have to reduce the resolution of the LUT in at least one dimension, in order to keep the computational burden acceptable. Experiment (xviii) suggests that the crossing-angle step size, δ_θ, in the LUT could be larger than 5 degrees, since the relationship between the crossing angle and the concentration is approximately linear. However, experiment (xviii) used a highly anisotropic test function ($f = 0.8$) and further experiments would be required to test the behaviour of less anisotropic test functions at varying crossing angles.

We propose two models of the fibre-orientation PDF: the two-Watson model and the two-Bingham model, both mixtures of the PDFs we used as single-fibre distributions in Ch. 7. The elliptical clusters fibre orientations from the two-tensor fitting routine are better modelled with Bingham distributions, and experiment (xxviii) shows that the two models produce significantly different connection probabilities in the brain.

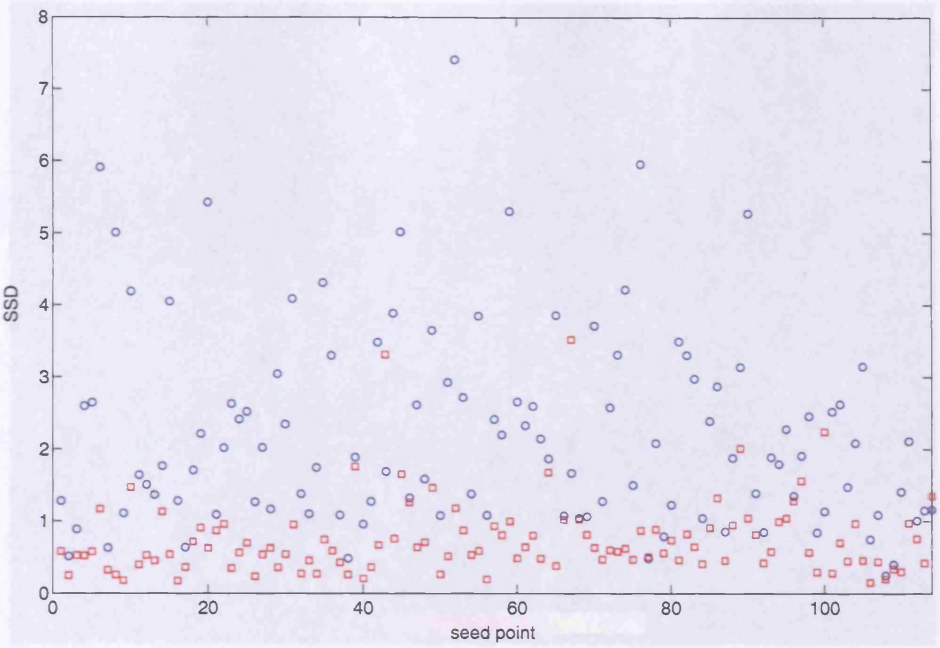


Figure 8.30: SSD for the two-Bingham model (red squares) and two-Watson (blue circles) between the bootstrap and the PICo connection probabilities, with PICo concentration estimated from the LUT in the average brain.

In the calibration of the PICo PDFs, we obtain many noisy estimates of the fibre orientations from a two-tensor fitting routine. The ordering of these axes is unspecified, so it is often not clear how to divide the axes into separate clusters from each fibre population. We propose fitting the parameters of a two-Watson and two-Bingham model directly to the data, and compare this to an iterative sorting approach [18] where the axes are divided into two groups, and a single-mode PDF is fitted separately to each group of axes. Experiment (xiv) shows that the direct fitting is better than the sorting at finding the parameters of Bingham distributions, when the axes are sampled from Bingham distributions and one or both of the clusters of axes have very low concentration. However, the direct fitting method appears to be more sensitive to the outliers that are found in clusters of axes extracted from noisy MR data. Experiment (xvii) shows that the sorting method gives better estimates of the PDF orientation, and more plausible estimates of the concentration of fibre-orientation estimates from synthesised MR data.

We compare the PDF concentration from bootstrap fibre orientations to the concentration estimated from the LUT calibration approach. The bootstrap fibre orientations are an imperfect gold standard, as we see in experiment (xxiv). Bootstrapping increases the dependence of the concentration on the orientation of the diffusion relative to the gradient directions. There is some correlation between the bootstrap concentrations and those predicted by the LUT, but

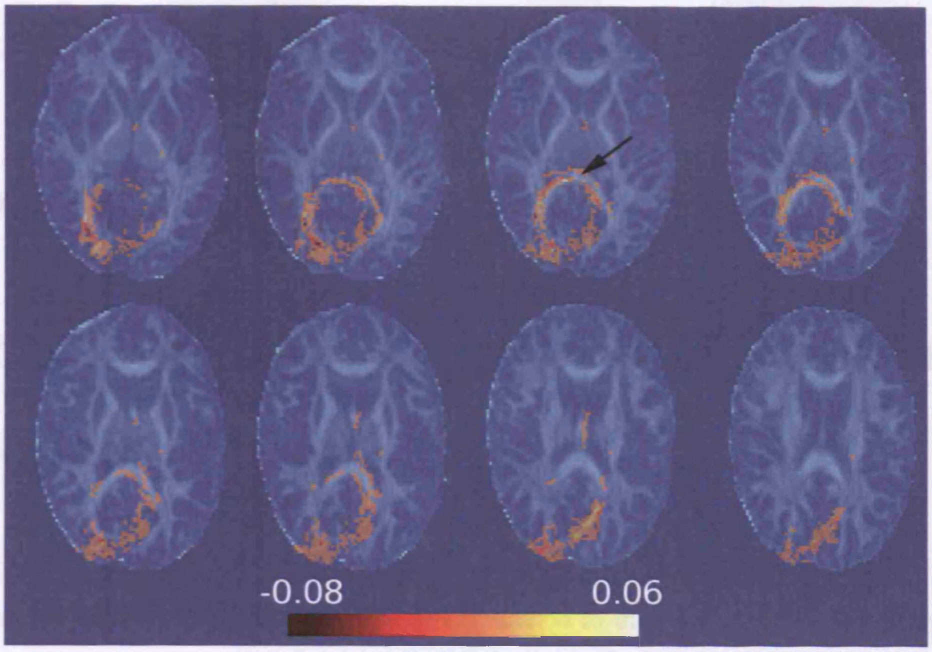


Figure 8.31: Difference in connection probability between the bootstrap tracking and the Bingham model, with concentration estimated from the LUT in brain image 1. The arrow highlights the seed point.

the LUT tends to underestimate the concentration compared to the bootstrap method. The differences between the two are probably caused in part by the deficiencies of the bootstrap method, and in part by deficiencies in the PICo calibration. PICo models only the uncertainty due to noise; other sources of error, such as artefacts due to magnetic susceptibility effects, may also be present in the brain data, however we would expect those to reduce concentration of the bootstrap fibre orientations, rather than increase it.

We perform probabilistic tractography in the brain using the bootstrap fibre orientations, the Bingham and Watson PDFs fitted to the bootstrap data, and the Bingham and Watson PDFs calibrated from the LUTs. The connection probability maps calculated using the Bingham PDF models fitted to the bootstrap data are similar to those calculated from bootstrap tracking (experiment (xxvii)). This suggests that the two-Bingham PDF is a good model of the fibre-orientation PDF. However, the PICo calibration process introduces errors in the concentration that make the differences between PICo and bootstrap connection probabilities much larger, but the Bingham model still has smaller errors than the Watson model (experiment (xxviii)).

Chapter 9

Connectivity based partitioning of the corpus callosum

This chapter details an application of the PICo method to the problem of connectivity-based partitioning of the corpus callosum. Connectivity-based partitioning is a method for segmenting a region of interest in the brain based on anatomical connectivity to other brain regions. Given a segmentation of cortical grey matter into distinct labelled regions, we use tractography to determine which labelled region is most likely to be connected to each voxel in the region of interest.

Part of this work is published in [108], where we applied the Watson PDFs. In this thesis we use the two-Bingham PDFs. Compared to earlier efforts [19, 109, 110], we make two main innovations. Firstly, we leverage advances in atlas-based image segmentation to produce cortex partitioning of subjects automatically. Secondly, we resolve crossing fibres and use a novel model of uncertainty as described in Ch. 8.

9.1 Method

We segment the corpus callosum of 8 healthy subjects (six male, two female). The diffusion-weighted images contain 66 diffusion weighted measurements in each voxel. The first 54 measurements are at a fixed $\mid \mathbf{q} \mid = q$, $\delta = 0.034$s, $\Delta = 0.04$s (δ and Δ are defined in Fig. 3.2), $\mid \mathbf{g} \mid = 0.022\mathrm{T\,m^{-1}}$, giving a b-value of $1050\mathrm{s\,mm^{-2}}$. The gradient directions are spread evenly on the hemisphere to minimise then dependence of the tensor on the orientation of the tissue relative to the gradient directions. Six images have no diffusion-weighting, and six of the 54 diffusion-weighted measurements are repeated with $\mid \mathbf{q} \mid = q/2$ as part of an experimental acquisition procedure, which is unrelated to this work.

We refer to the brain region of interest as the seed region and each voxel in the seed region as a seed point. We define the "connectivity map" for each seed point as the collection of probabilistic streamlines emanating from the seed.

The algorithm to partition the seed region is as follows:

1. Label the grey matter regions of the subject in the space of the diffusion data.
2. Generate the connectivity map for each seed point from the diffusion data.
3. For each seed point, determine the labelled grey matter region to which the seed has highest likelihood of being connected according to the connectivity maps.

We describe the details of each of these three steps in order.

9.1.1 Cortical region labeling

We require cortical regions to be defined in the same space as the connectivity maps. However the diffusion-weighted images lack the resolution for accurate delineation of cortex. Thus a high resolution T1-weighted image of the subject is acquired, on which the cortical regions of interest are defined. These regions are then transferred onto the diffusion data after co-registration of the T1-weighted image to a non-diffusion weighted image acquired as part of the diffusion MRI sequence hereafter referred to as the unweighted image.

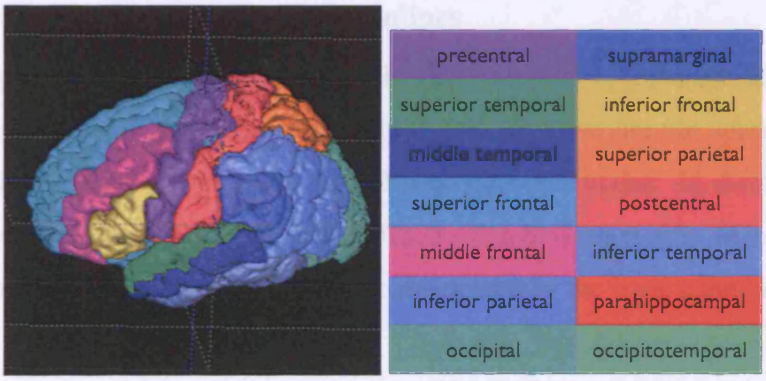


Figure 9.1: A surface rendering of the labeled atlas used in this work with cortical labels.

One feature of our method is the automation of the cortical region labeling step by warping an atlas [111] into alignment with the T1-weighted image. Figure 9.1 shows the brain atlas and the cortical labels. Fig. 9.1.1 shows the warped atlas after registration to the T1-weighted image of one subject. To improve the quality of the co-registration between the T1-weighted and unweighted images, we also acquire a high-resolution EPI image, which is used as an intermediate representation of the brain configuration in the diffusion data space. Specifically, the T1-weighted image is non-rigidly aligned to the EPI image, which in turn is non-rigidly registered to the unweighted image. The corresponding transformations are computed to enable

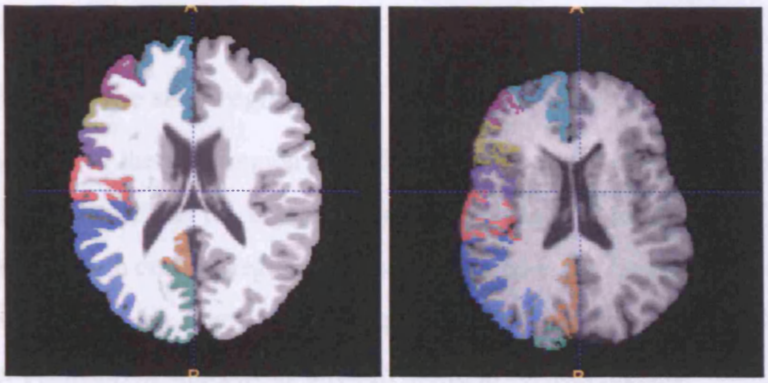


Figure 9.2: (Left) The grey matter labels for one hemisphere, shown superimposed on the underlying structural image of the atlas. (Right) The atlas is registered to the T1-weighted image of one subject, and the warped grey matter labels for one hemisphere are shown superimposed on the subject's structural image.

atlas labels defined on the T1-weighted image to be transferred to the unweighted image. The non-rigid registration algorithm used optimizes an intensity cross-correlation metric under the constraints of a diffeomorphic transformation model in multiresolution fashion [112].

9.1.2 Generation of probabilistic streamlines

The tensor fitting routine is the same we use in Ch. 7 and Ch. 8. In single-fibre voxels, we fit the elements of the diffusion tensor directly to the data. We use the spherical-harmonic classification [54] to find voxels where the diffusion is non-Gaussian. In these voxels we fit the restricted two-tensor inversion defined in Ch. 8, which assumes that the two tensors are cylindrically-symmetric and that each of the compartments in the voxel contributes equally to the signal.

We generate streamlines using the Bingham PICo PDFs. We generate LUTs following the procedures in Ch. 7 and Ch. 8. We estimate the signal to noise ratio in this data to be approximately approximately 14 in white matter in the unweighted images. We use the direct fitting method to find the parameters of the two-Bingham PDFs. This is consistent with the earlier published version of this work [108], but in light of experiment (xvii), we expect that the sorting method would produce better results.

The tractography method is the same as in Ch. 8. We define the region of interest as a cross section of the corpus callosum along the mid-sagittal line (between the cerebral hemispheres). We use a simple region-growing algorithm to segment all connected voxels with anisotropy $f > 0.4$. This anisotropy threshold is higher than the usual grey-white matter threshold of $f \approx 0.1$. We use this threshold to eliminate voxels that contain a mixture of white matter and

CSF or of white and grey matter.

9.1.3 Labelling of the seed region

We label each voxel in the seed region according to the cortical region (as defined in [111] and shown in Fig. 9.1) that has the highest connection probability to the seed. The connection probability between each cortical region and the seed voxel is the fraction of streamlines from the seed that enter the region. Because tracking terminates upon entry to a labelled region, we do not consider connectivity that passes through multiple cortical regions.

9.2 Results

The fibre bundles passing through the corpus callosum extend to both the left and right cortex, so we construct two sets of connectivity maps. Each map represents the connections to one hemisphere of the brain. Fig. 9.3 shows the labelled corpus callosum of all eight subjects.

The three largest partitions in the corpus callosum we find are consistent with the earlier study [19]. The voxels within these partitions are most likely to connect to the superior frontal gyrus, the superior parietal gyrus and the occipital gyrus respectively. Fig. 9.4 shows the size of each partition as a percentage of the size of the corpus callosum for the Bingham connectivity map. On average over all the controls, the sizes of the three largest partitions are 54%, 21% and 9% for the left hemisphere and 57%, 23% and 10% for the right hemisphere. The fourth largest partition to the left hemisphere is the postcentral gyrus, with an average size of 6%. The fourth largest partition to the right hemisphere is the superior temporal gyrus, with an average size of 3%. Postcentral connections to the right hemisphere 0.8% on average.

We repeat the tractography without using the two-Bingham PDFs, to see what influence resolving crossing fibres has on the results. We fit a single diffusion tensor in each image and perform PICo tractography as before, using the Bingham PDFs. Fig. 9.5 shows the results, which are very similar to the results when we use the voxel classification and two-Bingham PDFs. The failure of the two-fibre methods to significantly alter the results has several possible explanations. The voxels containing crossing fibres may not be consistently identified. Fig. 9.6 shows a slice of a fractional anisotropy image, with a known fibre-crossing region (studied in, for example, [54] and [113]) highlighted, where fibres of the cortico-spinal tract are thought to cross those of the corpus callosum. Note that the anisotropy appears to be lower in the crossing region than in the surrounding white matter, because the Gaussian model is a poor approximation of the DW-MR measurements. The corresponding voxel classification from the spherical harmonic algorithm is also shown overlaid on the anisotropy. Only some of the voxels in the fibre-crossing region are classified as non-Gaussian. Using two-fibre PDFs will only make a difference to the labelling if the streamlines are diverted into a different cortical zone as a result of resolving the fibre-crossings. The architecture of the fibre pathways and cortical zones may make this unlikely. In addition to the difficulties associated with identifying crossing fibres, the tracking algorithm is biased towards following a low-curvature path, since we follow the direction most closely aligned to the previous streamline trajectory.

In [108], we used the voxel PDFs to make a random choice of which principal direction to follow in voxels with two fibre populations. If the (Watson) PDFs in a voxel are $W_1(\mathbf{x}; \boldsymbol{\mu}_1, \kappa)$ and $W_2(\mathbf{x}; \boldsymbol{\mu}_2, \kappa)$, and the orientation of the streamline trajectory on entry to the voxel is $\mathbf{x}$, then the probability of tracking along a sample from W_1 is $p_1 = W_1(\mathbf{x})/[W_1(\mathbf{x}) + W_2(\mathbf{x})$ and the probability of tracking along a sample from W_2 is $1 - p_1$. If two equally-concentrated PDFs have peaks almost equally aligned with the incoming streamline trajectory, then the streamline is almost equally likely to follow a sample from either one. The tracking algorithm in this thesis would always choose whichever PDF was most closely aligned to the incoming direction, even if the difference between the two was small. However, further work is required to explore more fully whether making a probabilistic choice of fibre orientation is theoretically sound, and how such a method could be compared to bootstrap tractography, where no PDF is defined.

The partitioning results in Fig. 9.4 are consistent with those in [108] for the three largest partitions, which constitute over 90% of the partitioned area of the corpus callosum, but the postcentral gyrus partition is not consistently larger in the left than in the right connectivity maps.

9.3 Conclusions

Compared to the work by Huang et al. [19], we have chosen a significantly finer cortical labeling. Despite that, it appears majority of the connections through the corpus callosum are restricted to several large cortical regions. Additionally, several regions, notably the middle and inferior temporal gyrus, the occipitotemporal gyrus and the supramarginal gyrus, are not found in the partitions. It is possible that the connections to these regions are mediated through other regions, either by ipsilateral or U-fibre connections. Alternatively, such connections may exist in the corpus callosum but are not resolved by the partitioning method. The tractography estimates the probability that a single fibre pathway connects to the labelled region. Multiple cortical regions may have a nonzero connection probability to the seed but this cannot be interpreted as a distribution of distinct fibre pathways. The PDF that we sample from in PICo is the probability of a single fibre pathway existing along the path of each streamline. We cannot resolve multiple connections to the seed voxel.

We were unable to find any difference in the segmentation when using the two-Bingham PDFs in conjunction with the spherical-harmonic voxel classification. A more reliable voxel classification algorithm, or an alternative method of resolving fibre crossings (such as q-ball [43] or PAS-MRI [58]) may alter the results considerably.

It has been shown in postmortem studies that there are wide-spread heterotopic commissure connections across the corpus callosum [114]. We consistently find more connections to the left postcentral gyrus than to the right. Partitioning studies such as this have the potential to offer new evidence of the existence of the heterotopic connections and, furthermore, possibly locate where such connections traverse across the corpus callosum. Other potential applications include the study of connectivity in diseases such as schizophrenia, where previous research has suggested a reduction in the size of the corpus callosum [115].

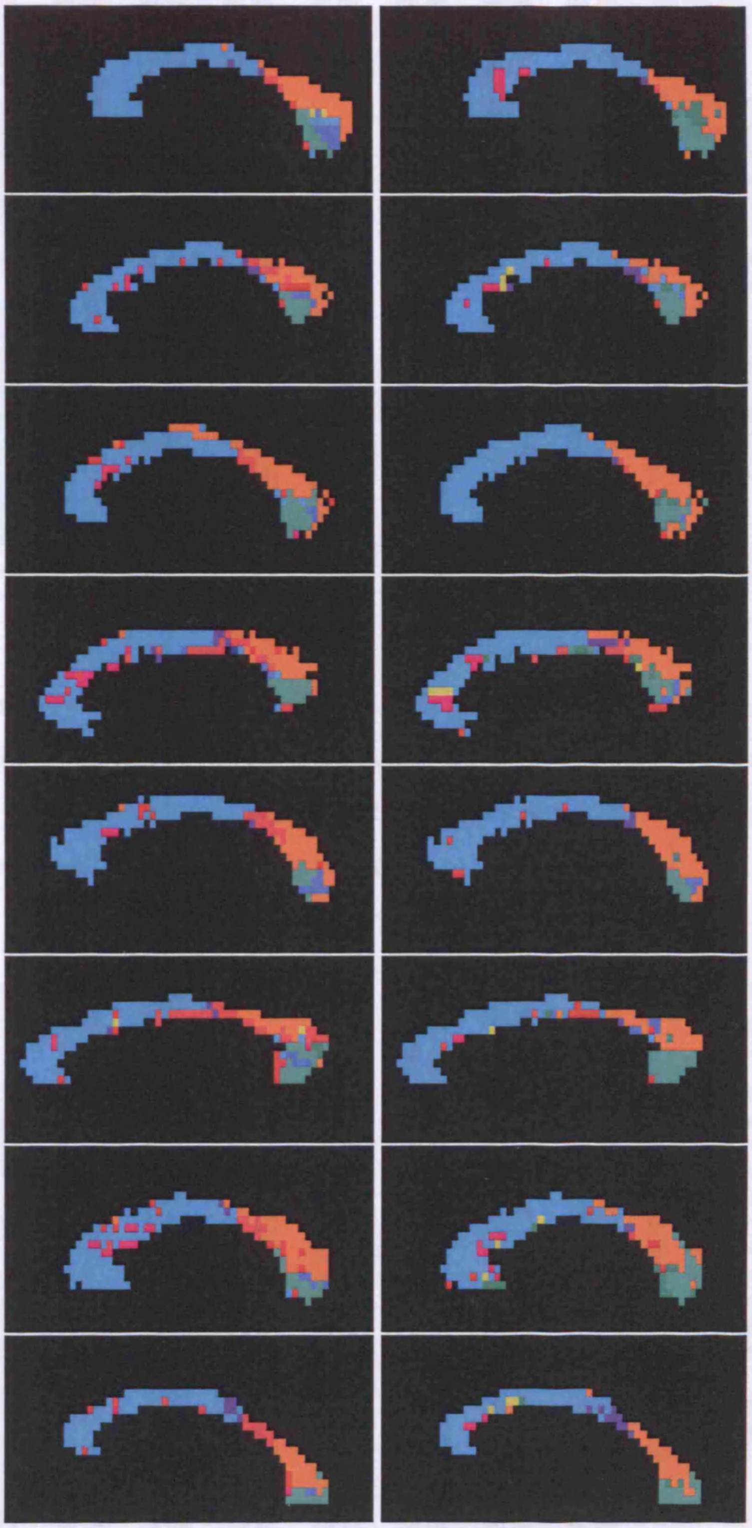


Figure 9.3: Partitions for the left and right side connectivity from the Bingham connectivity map, coloured according to the labels in Fig. 9.1.

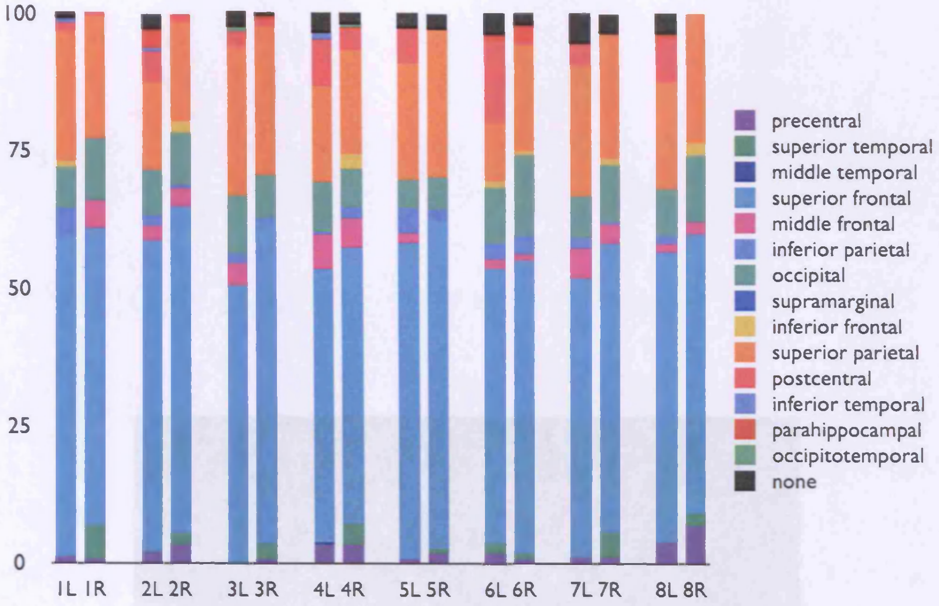


Figure 9.4: Results of tracking with Bingham PICo PDFs, using the two-fibre PDFs in non-Gaussian voxels. The number of voxels assigned each label is shown as a percentage of the total voxels in the corpus callosum segmentation of each subject. The black region are voxels from which no streamlines reached a labelled region.

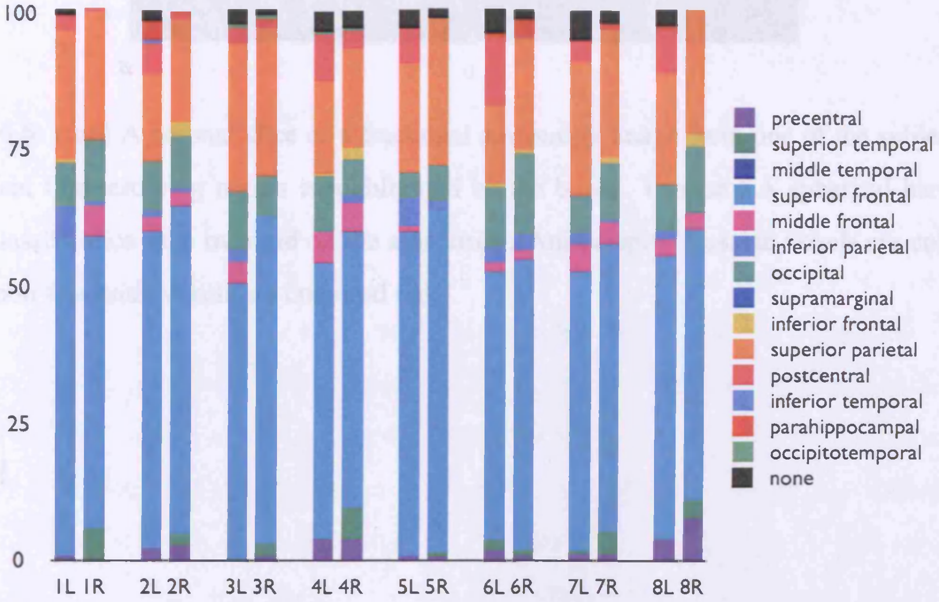


Figure 9.5: Results of tracking with Bingham PICo PDFs using only the single-fibre PDFs.

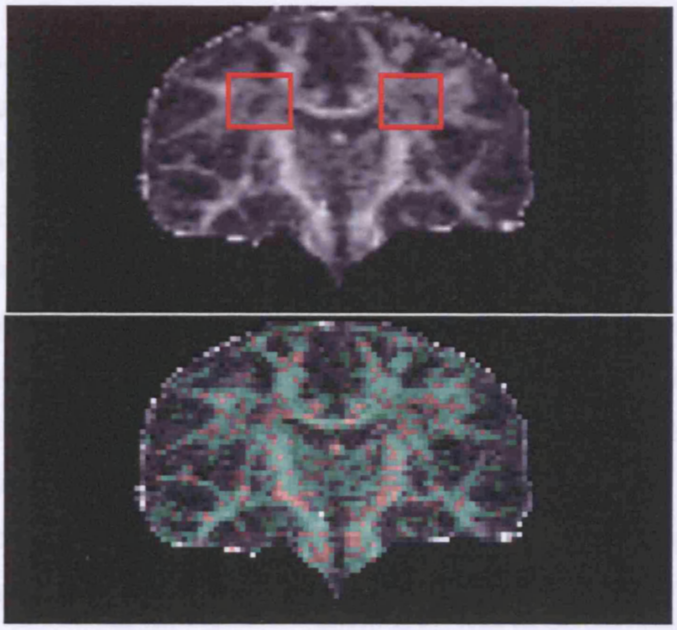


Figure 9.6: (top) A coronal slice of a fractional anisotropy image from one of the subjects. A prominent fibre crossing region is highlighted by the boxes. (bottom) A spherical harmonic voxel classification map overlaid on the anisotropy. Anisotropic Gaussian voxels are coloured green, non-Gaussian voxels are coloured red.

Chapter 10

Conclusions and future directions

We have presented new models of the fibre-orientation PDF for voxels containing a single fibre population. The Watson PDF is similar to the Gaussian model used by Parker and Alexander [56], with circular contours on the sphere. Unlike Parker's Gaussian model, the Watson PDF can also model a distribution of fibre orientations concentrated in a plane, such a distribution is seen when there are crossing fibres in a voxel and the diffusion tensor is oblate.

The Bingham PDF generalises the Watson PDF to distributions that do not have rotational symmetry about the peak. It can therefore model the elliptical clusters of fibre orientations observed in voxels where the diffusion is anisotropic but does not have cylindrical symmetry about the fibre orientation.

The ACG PDF can also model elliptical clusters of fibre orientations, but tends to overestimate the uncertainty. The ACG PDF is less sensitive to outliers than the Bingham or Watson models, however this means that when sampling from the ACG PDF we are more likely to obtain an outlying axis. This means that the ACG PDF disperses streamlines more widely than the other models (see Fig. 7.8). There are relatively few outliers in the single-tensor fitting at the SNR of the brain data, which makes the ACG PDF least suitable of the three candidates.

We show in synthetic data that the Bingham PDF provides the best model of the uncertainty due to noise, since it provides lower errors to the gold standard of the uncertainty. In the bootstrap brain data, we also find that the Bingham model provides the best model of the fibre-orientation PDF as defined by the bootstrap data. In practice, our ability to model the fibre-orientation PDF is limited, because we must rely on a LUT, which is indexed according to tensor parameters which are subject to errors. We show that even when the parameters are taken from the average brain, the difference in connection probabilities between PICo and the bootstrap tractography grows from a mean SSD of 0.07 to a mean of 0.57. When the parameters are taken from a single brain image, the mean SSD is 1.0.

For the two-fibre case, we test a mixture model of the Watson and Bingham models. We show that if we impose constraints to improve the stability of the two-tensor fitting, the clusters of fibre orientations from the fitting routine are elliptical, and best modelled by a Bingham distribution. Experiment (xiv) suggested that direct fitting of the two-Bingham parameters via optimisation may improve the accuracy of the estimated concentration. However, in axes obtained from MR data, the iterative sorting method of Alexander and Barker [38] proved to be better, at low crossing angles. The differences between the bootstrap fibre tracking and two-fibre PICo are larger than in the single-fibre image, the mean SSD with tensor parameters taken from brain image 1 is 1.48. Still, the two-Bingham model has lower SSD than the two-Watson model.

The distributions of fibre orientations in two-fibre voxels are less concentrated and have more outliers than the single-fibre distributions. The ACG model could be better in these voxels, however it us not such a good model in the single fibre case. A future work could perform PICo using the Bingham PDF in one-fibre voxels and the ACG PDF in two-fibre voxels. We could also use the Watson PDF in the two-fibre voxels, to quantify how much difference the two-Bingham PDF specifically makes to the connection probabilities. Another interesting idea is to model the effects of a variable mixing parameter in the two-fibre fitting. It is possible that a variable mixing parameter will destabilise the estimated anisotropy and cause a larger error in the concentration, but ignoring the mixing parameter also causes an error in the estimated concentration.

We use the Bingham PDFs to perform a connectivity based segmentation of the corpus callosum. We use direct fitting in Ch. 9 to find the parameters of the two-Bingham PDF. Based on experiment (xvii), we expect that the sorting method would be better. The experiments in Ch. 8 and Ch. 9 depend on spherical harmonic voxel classification to distinguish two-fibre voxels. The classification thresholds are set manually and there is no correct threshold that is known to produce the best classification. Using the two-fibre models did not alter the results of the corpus callosum segmentation, however it would be interesting to see how perturbing the thresholds affects the results of experiments (xxvi) and (xxvii), where we compare PDF concentration at the voxel level.

Another interesting future work would be a full comparison between the model-based approaches and the Bayesian methods proposed by Behrens et al, [13], Friman and Westin [89] and Hosey et al [90].

Appendix A

Generation of pseudo-random samples from axial distributions

A.1 Watson distribution

The routines for simulating the bipolar and girdle Watson distribution give the spherical polar coordinates of the sample axis $\mathbf{x}$. The values R_x refer to pseudo-random variates from a uniform distribution ($R \in [0, 1]$). These should be set to new values each time a line is executed.

A.1.1 Bipolar distribution

The mean axis is assumed to be the Cartesian z-axis, where $\theta = \phi = 0$. In the implementation, the mean axis can be specified, and the samples generated from the algorithm below are rotated appropriately. We use the rejection method [67, p.290], shown by Best and Fisher [116] to generate the samples.

1. Set $C = \frac{1}{\exp(\kappa)-1}$
2. $U = R_1, V = R_2$
3. Set $S = \frac{1}{\kappa}\log(\frac{U}{C} + 1)$
4. If $V > \exp(\kappa S^2 - \kappa S)$, go to 2
5. Sample direction = $(\cos^{-1} S, 2\pi R_3)$

A.1.2 Girdle distribution

The polar axis is assumed to be $\theta = \phi = 0$. As in the bipolar case, the polar axis can be specified at run time.

1. Set $C_1 = \sqrt{|\kappa|}, C_2 = \tan^{-1}(C_1)$
2. $U = R_1, V = R_2$

3. Set $S = \frac{1}{C_1} \tan(C_2 U)$
4. If $V > (1 - \kappa S^2) \exp(\kappa S^2)$, go to 2
5. Sample direction = $(\cos^{-1} S, 2\pi R_3)$

A.2 Bingham distribution

The Bingham is simulated by the rejection method, using a modified Watson distribution as a wrapper function. We require the Watson distribution that wraps the Bingham distribution as closely as possible, as shown in figure A.1. For most distributions the best choice of wrapper is a bipolar distribution with $\boldsymbol{\mu} = \mathbf{e}_1$. The wrapper must always be greater than or equal to the underlying Bingham distribution, so

$$\frac{1}{C_{bw}} \exp\left[\kappa_3(\boldsymbol{\mu}_3 \cdot \mathbf{x})^2\right] \geq \frac{1}{C_B} \exp\left[\kappa_1(\boldsymbol{\mu}_1 \cdot \mathbf{x})^2 + \kappa_2(\boldsymbol{\mu}_2 \cdot \mathbf{x})^2\right] \tag{A.1}$$

The normalisation constant C_{bw} is different from the constant in equation 6.1. The minimum value of C_W is found by considering the case where $\mathbf{x}$ is orthogonal to $\boldsymbol{\mu}_3$.

$$\frac{1}{C_{bw}} \geq \frac{1}{C_B} \exp\left[\kappa_1(\boldsymbol{\mu}_1 \cdot \mathbf{x})^2 + \kappa_2(\boldsymbol{\mu}_2 \cdot \mathbf{x})^2\right]. \tag{A.2}$$

Because $\kappa_1 \leq \kappa_2$, the right hand side of A.2 is maximum when $\mathbf{x}$ is equal to $\pm\boldsymbol{\mu}_1$, yielding

$$\begin{aligned} \frac{1}{C_{bw}} &\geq \frac{1}{C_B} \exp\left[\kappa_2\right] \\ C_{bw} &\leq C_B \exp\left[-\kappa_2\right]. \end{aligned} \tag{A.3}$$

Substituting the maximum value of C_{bw} into A.1 and taking logs of both sides gives

$$\kappa_3(\boldsymbol{\mu}_3 \cdot \mathbf{x})^2 + \kappa_2 \geq \kappa_1(\boldsymbol{\mu}_1 \cdot \mathbf{x})^2 + \kappa_2(\boldsymbol{\mu}_2 \cdot \mathbf{x})^2. \tag{A.4}$$

The right hand side is maximised when $\mathbf{x}$ equals $\boldsymbol{\mu}_3$, so

$$\kappa_3 \geq -\kappa_2. \tag{A.5}$$

Substituting this into equation A.4, we get

$$\kappa_1(\boldsymbol{\mu}_1 \cdot \mathbf{x})^2 + \kappa_2[(\boldsymbol{\mu}_2 \cdot \mathbf{x})^2 + (\boldsymbol{\mu}_3 \cdot \mathbf{x})^2)] - \kappa_2 \leq 0. \tag{A.6}$$

Given that $\kappa_1 < \kappa_2$ and $(\boldsymbol{\mu}_1 \cdot \mathbf{x})^2 + (\boldsymbol{\mu}_2 \cdot \mathbf{x})^2 + (\boldsymbol{\mu}_3 \cdot \mathbf{x})^2 = 1$ for all $\mathbf{x}$, it is easy to see that equation A.6 is satisfied for all $\mathbf{x}$.

When $\kappa_1 \ll \kappa_2$,the Bingham distribution is similar to a girdle Watson distribution with $\boldsymbol{\mu}_3 = \mathbf{e}_3$. In this instance it is more efficient to use a girdle Watson as a wrapper. Equation A.2 becomes

$$\frac{1}{C_{bw}} \exp\left[\kappa_3(\boldsymbol{\mu}_1 \cdot \mathbf{x})^2\right] \geq \frac{1}{C_B} \exp\left[\kappa_1(\boldsymbol{\mu}_1 \cdot \mathbf{x})^2 + \kappa_2(\boldsymbol{\mu}_2 \cdot \mathbf{x})^2\right] \tag{A.7}$$

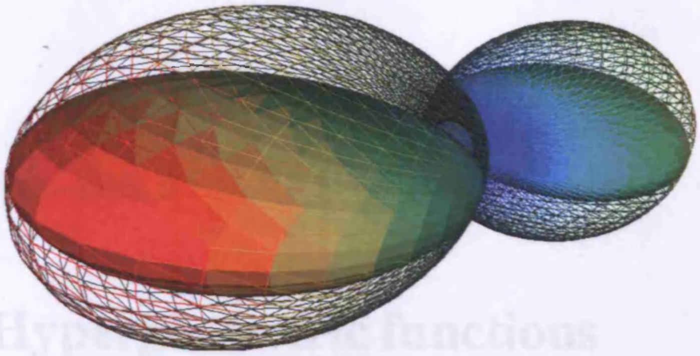


Figure A.1: Contours of a bipolar Watson distribution wrapping a Bingham distribution.

which is maximum when $\mathbf{x}$ is orthogonal to $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, giving

$$C_{bw} \leq C_B. \tag{A.8}$$

Substituting this into A.1, it follows that $\kappa_3 \geq \kappa_1$.

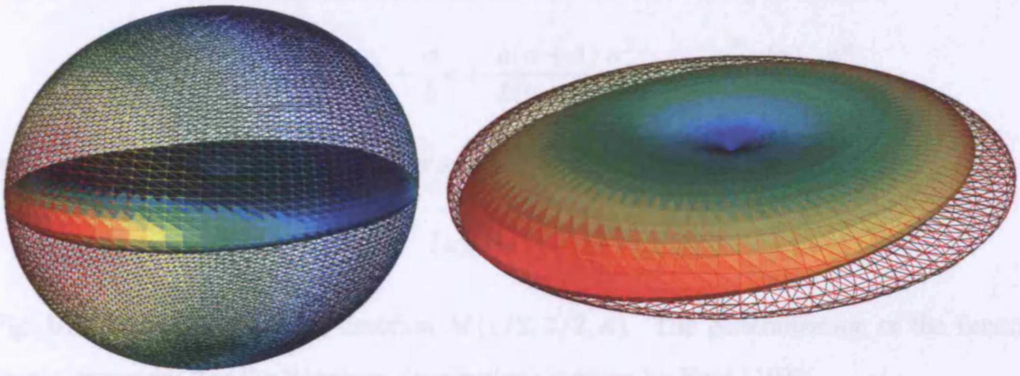


Figure A.2: The girdle distribution (right) is a more efficient wrapper than the bipolar (left) in some cases.

A.3 Angular Central Gaussian Distribution

The ACG distribution is closely related to the trivariate Gaussian distribution (see Ch. 6). We therefore use a standard method to generate trivariate Gaussian samples with zero mean and covariance matrix $\mathbf{\Lambda}$. Let N_x be samples from the standard normal distribution $N(0, \mathbf{\Lambda})$. We decompose $\mathbf{\Lambda}$ into its eigen system, $\mathbf{\Lambda} = \mathbf{U}\mathbf{L}\mathbf{U}^{\mathrm{T}}$, with $\mathbf{L} = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. This process is done only once. The algorithm to generate a sample is then:

1. $\mathbf{z} = (N_1 * \sigma_1, N_2 * \sigma_2, N_3 * \sigma_3)^{\mathrm{T}}$
2. return $\mathbf{U}\mathbf{z} \mid \mathbf{z} \mid^{-1}$.

Appendix B

Hypergeometric functions

The normalizing constant for the Watson distribution in three dimensions is the confluent hypergeometric function of the first kind [68]:

$$M(a,b,\kappa) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 \exp(2t) t^{a-1}(1-t)^{b-a-1} dt \,, \tag{B.1}$$

where Γ is the Gamma function. The function also has a series representation:

$$M(a,b,\kappa) = 1 + \frac{a}{b}\kappa + \frac{a(a+1)}{b(b+1)}\frac{\kappa^2}{2!} \cdots = \sum_{n=1}^{\infty} \frac{(a)_n}{(b)_n}\frac{\kappa^n}{n!} \,, \tag{B.2}$$

where $(a)_n$ and $(b)_n$ are Pochhammer symbols [117]

$$(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)} \,. \tag{B.3}$$

Fig. B.1 shows a plot of the function $M(1/2, 3/2, \kappa)$. The generalisation of the function to matrix argument (for the Bingham distribution) is given by Kent [103]:

$$M(\frac{1}{2}, \frac{3}{2}, \mathbf{A}) = 2\sqrt{\pi} \exp[\kappa_1] \sum_{i,j=0}^{\infty} \frac{\Gamma(i+\frac{1}{2})\Gamma(j+\frac{1}{2})\alpha_1^i \alpha_2^j}{\Gamma(i+j+\frac{3}{2}) i! j!} \,, \tag{B.4}$$

where $\alpha_1 = -\kappa_1$ and $\alpha_2 = \kappa_2 - \kappa_1$.

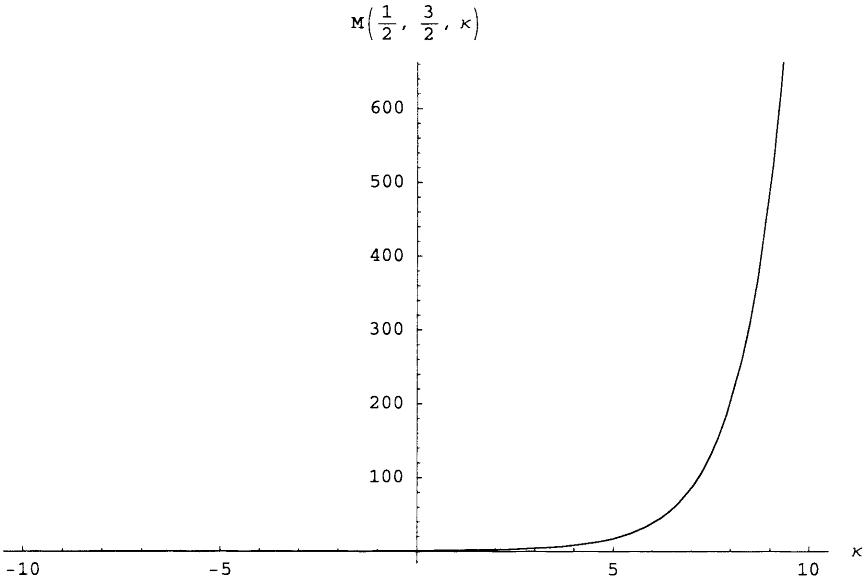


Figure B.1: Plot of the confluent hypergeometric function of the first kind against κ.

Bibliography

[1] P A Cook, Y Bai, M G Hall, S Nedjati-Gilani, K K Seunarine, G J M Parker, and D C Alexander. Camino: Open-source diffusion-MRI reconstruction and processing. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 2759, 2006.

[2] P A Cook, D C Alexander, and G J M Parker. Modelling noise-induced fibre-orientation error in diffusion-tensor MRI. In *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pages 332–336, 2004.

[3] P A Cook, P A Boulby, M R Symms, and D C Alexander. Optimal acquisition order of diffusion-weighted measurements on a sphere. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 1305, 2005.

[4] P A Cook, P A Boulby, M R Symms, and D C Alexander. Ordering diffusion-weighted MRI measurements improves results from partially completed scans. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 1035, 2006.

[5] S G Waxman (Editor). *Physiology and Pathobiology of Axons*. Raven Press, 1978.

[6] C Pierpaoli, Peter Jezzard, P J Basser, Alan Barnett, and Giovanni Chiro. Diffusion tensor MR imaging of the human brain. *Radiology*, 201:637–648, 1996.

[7] C Beualieu and P S Allen. Water diffusion in the giant axon of the squid: implications for diffusion-weighted MRI of the nervous system. *Magnetic Resonance in Medicine*, 32:579–583, 1994.

[8] C Poupon, C A Clark, V Frouin, J Régis, I Bloch, D Le Bihan, and J F Mangin. Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *NeuroImage*, 12:184–195, 2000.

[9] S Mori, Barbara J Crain, VP Chacko, , and PCM van Zijl. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45:265–269, 1999.

[10] R Bammer, B Acar, and M E Moseley. In vivo MR tractography using diffusion imaging. *European Journal of Radiology*, 45:223–234, 2003.

[11] S Mori and Peter C. M. van Zijl. Fiber tracking: principles and strategies: a technical review. *NMR In Biomedicine*, 15:468–480, 2002.

[12] G J M Parker, C A M Wheeler-Kingshott, and H A Haroon. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *Journal of Magnetic Resonance Imaging*, 18:242–254, 2003.

[13] 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. *Magnetic Resonance in Medicine*, 50:1077–1088, 2003.

[14] H Johansen-Berg, T E J Behrens, E Sillery, O Ciccarelli, A J Thompson, S M Smith, and P M Matthews. Functionalanatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. *Cerebral Cortex*, 15:31:39, 2005.

[15] P J Basser, J Matiello, and D Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.

[16] 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. *Magnetic Resonance in Medicine*, 48:577–582, 2002.

[17] B Efron. Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7:1–26, 1979.

[18] D C Alexander. An introduction to computational diffusion MRI: the diffusion tensor and beyond. In J Weichert and H Hagen, editors, *Visualization and Image Processing of Tensor Fields*. Springer, 2005.

[19] H Huang, J Zhang, H Jiang, S Wakanaa, L Poetschera, M I. Miller, P C M van Zijl, A E Hillis, R Wytik, and S Mori. DTI tractography based parcellation of white matter: Application to the mid-sagittal morphology of corpus callosum. *NeuroImage*, 26:195–205, 2005.

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

[21] R C Smith and R C Lange. *Understanding Magnetic Resonance Imaging.* CRC Press, 1998.

[22] A Das and T Ferbel. *Introduction to Nuclear and Particle Physics.* Wiley, 1994.

[23] J P Hornak. *The Basics of MRI.* J P Hornak, 1996. `http://www.cis.rit.edu/htbooks/mri/index.html`.

[24] H Y Carr and E M Purcell. Effects of free precession in nuclear magnetic resonance experiments. *Physics Review*, 94, 1954.

[25] F Mandl. *Statistical Physics (2nd edition).* Wiley, 1988.

[26] A Einstein. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Annalen Der Physik*, 17:549–560, 1905.

[27] H Gray. *Gray's Anatomy.* Magpie, 1993.

[28] A Peters, S L Palay, and H D Webster (editors). *The Fine Structure Of The Nervous System: Neurons And Their Supporting Cells, 3rd Edition.* Oxford University Press, New York, 1991.

[29] J Young, A Fritz, G Liu, K Thoburn, J Kres, and S Roffers. Anatomy and physiology: US National Cancer Institute online training module. `http://training.seer.cancer.gov/ss_module00_bbt/unit02_sec04_b_cells.html`.

[30] C Beualieu. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR In Biomedicine*, 15:435–455, 2002.

[31] C Beaulieu and P S Allen. Determinants of anisotropic water diffusion in nerves. *Magnetic Resonance in Medicine*, 31:394–400, 1994.

[32] V Gulani, A G Webb, I D Duncan, and P C Lauterbur. Apparent diffusion tensor measurements in myelin-deficient rat spinal cords. *NeuroImage*, 16:378–388, 2002.

[33] D M Wimberger, T P Roberts, A J Barkovich, L M Prayer, M E Moseley, and J Kucharczyk. Identification of "premyelination" by diffusion-weighted MRI. *Journal of Computer Assisted Tomography*, 19:28–33, 1995.

[34] D Prayer, T Roberts, A J Barkovich, L Prayer, J Kucharczyk, M Moseley, and A Arieff. Diffusion-weighted MRI of myelination in the rat brain following treatment with gonadal hormones. *Neuroradiology*, 29:320 – 325, 1997.

[35] P S Hüppi, S E Maier, S Peled, G P Zientara, P D Barnes, F A Jolesz, A Ferenc, and J J Volpe. Microstructural development of human newborn cerebral white matter assessed in vivo by diffusion tensor magnetic resonance imaging. *Pediatric Research*, 44:584–590, 1998.

[36] H Sakuma, Y Nomura, K Takeda, T Tagami, T Nakagawa, Y Tamagawa, Y Ishii, and T Tsukamoto. Adult and neonatal human brain: diffusional anisotropy and myelination with diffusion weighted mr imaging. *Radiology*, 180:229233, 1991.

[37] E O Stejskal J E Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *Journal of Chemical Physics*, 42:288–292, 1965.

[38] D C Alexander and G J Barker. Optimal imaging parameters for fiber-orientation estimation in diffusion MRI. *NeuroImage*, 27:357–367, 2005.

[39] J R Hansen. Pulsed NMR study of water mobility in muscle and brain tissue. *Biochimica et Biophysica Acta*, 230:482–486, 1971.

[40] Cleveland; GG, Chang; DC, Hazlewood; CF, and Rorschach; HE. Nuclear magnetic resonance measurement of skeletal muscle: anisotropy of the diffusion coefficient of the intracellular water. *Biophysical Journal*, 16:1043–1053, 1976.

[41] ME Moseley, YC Cohen, J Kucharczyk, HS Asgari, MF Wendland, J Tsuruda, and D Norman. Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology*, 176:439–445, 1990.

[42] V J Wedeen, T G Reese, D S Tuch, J G Dou, R M Weiskoff, and D Chessler. Mapping fiber orientation spectra in cerebral white matter with Fourier-transform diffusion MRI. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, 2000.

[43] D S Tuch. *Diffusion MRI of complex tissue structure.* Doctoral dissertation, Harvard-MIT Division of Health Sciences and Technology, 2002.

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

[45] Lin-Ching Chang, D K Jones, and C Pierpaoli. RESTORE: Robust estimation of tensors by outlier rejection. *Magnetic Resonance in Medicine*, 52:1088–1095, 2005.

[46] P J Basser and C Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of Magnetic Resonance B*, 111:209–219, 1996.

[47] R Bammer, M Augustin, S Strasser-Fuchs, T Seifert, P Kapeller, R Stollberger, F Ebner, H Hartung, and F Fazekas. Magnetic resonance diffusion tensor imaging for characterizing diffuse and focal white matter abnormalities in multiple sclerosis. *Magnetic Resonance in Medicine*, 44:583–591, 2000.

[48] Q Dong, R C Welsh, T L Chenevert, R C Carlos, P Maly-Sundgren, D M Gomez-Hassan, and S K Mukherji. Clinical applications of diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 19:6–18, 2004.

[49] I Agartz, JL Anderson, and S Skare. Abnormal brain white matter in schizophrenia: A DT-MRI imaging study. *NeuroReport*, 12:2251–2254, 2001.

[50] J Foong, M R Symms, G J Barker, M Maier, D H Miller, and M A Ron. Investigating regional white matter in schizophrenia using diffusion tensor imaging. *NeuroReport*, 13:333–336, 2002.

[51] K Arfanakis, B P Hermann, B P Rogers, J D Carew, M Seidenberg, and M E Meyerand. Diffusion tensor MRI in temporal lobe epilepsy. *Magnetic Resonance Imaging*, 20:511–519, 2002.

[52] D J Werring, C A Clark, G J Barker, A J Thompson, and D H Miller. Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology*, 52:1626–32, 1999.

[53] Lawrence R Frank. Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine*, 47:1083–1099, 2002.

[54] D C Alexander, G J Barker, and S R Arridge. Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data. *Magnetic Resonance in Medicine*, 48:331–340, 2001.

[55] E Özarslan and T H Mareci. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magnetic Resonance in Medicine*, 50:955–965, 2003.

[56] G J M Parker and D C Alexander. Probabilistic monte-carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. In C. J. Taylor and J. A. Noble, editors, *Information Processing in Medical Imaging*, volume 2732 of *Lecture Notes in Computer Science*, pages 684–695. Springer-Verlag, 2003.

[57] R Blyth, P A Cook, and D C Alexander. Tractography with multiple fibre directions. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 240, 2003.

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

[59] P Funk. Uber eine geometrische anwendung der abelschen integral- gleichnung. *Annals of Mathematics*, 77:129–135, 1916.

[60] C P Lin, W Y I Tseng, L Kuo, V J Wedeen, and J H Chen. Mapping orientation distribution function with spherical encoding. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 2120, 2003.

[61] D C Alexander. A comparison of q-ball and PASMRI on sparse diffusion MRI data. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 90, 2004.

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

[63] D K Jones, M A Horsefield, and A Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine*, 42:515–525, 1999.

[64] N G Papadakis, C D Murrills, L D Hall, C L H Huang, and T A Carpenter. Minimal gradient encoding for robust estimation of diffusion anisotropy. *Magnetic Resonance Imaging*, 18:671–679, 2000.

[65] K M Hasan, D L Parker, and A L Alexander. Comparison of gradient encoding schemes for diffusion-tensor MRI. *Journal of Magnetic Resonance Imaging*, 13:769–780, 2001.

[66] J Dubois, C Poupon, Y Cointepas, F Lethimonnier, and D Le Bihan. Diffusion gradient orientation schemes for dti acquisitions with unquiet subjects. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 90, 2004.

[67] W H Press, S A Teukolsky, W T Vetterling, and B P Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.

[68] K V Mardia and P E Jupp. *Directional Statistics*. Wiley, 2000.

[69] B Stieltjes, W E Kaufmann, P C M 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, 2001.

[70] 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. In *Proceedings of the National Academy of Sciences, USA*, volume 96, pages 10422–10427, 1999.

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

[72] S Pajevic, A Aldroubi, , and P J Basser. A continuous tensor field approximation of discrete DT-MRI data for extracting microstructural and architectural features of tissue. *Journal of Magnetic Resonance*, 154:85–100, 2002.

[73] C R Tench, P S Morgan, M. Wilson, and L D Blumhardt. White matter mapping using diffusion tensor MRI. *Magnetic Resonance in Medicine*, 47:967–972, 2002.

[74] D Weinstein, G Kindlmann, and E Lundberg. Tensorlines: Advection-diffusion based propagation through diffusion tensor fields. In *IEEE Visualization*, pages 249–253, 1999.

[75] M Lazar, D M Weinstein, J S Tsuruda, K M Hasan, K Arfanakis, M E Meyer, B Badie, H A Rowley, V Haughton, A Field, and A L Alexander. White matter tractography using diffusion tensor deflection. *Human Brain Mapping*, 18:306 321, 2003.

[76] M Lazar and A L Alexander. An error analysis of white matter tractography methods: synthetic diffusion tensor field simulations. *NeuroImage*, 20:1140–1153, 2003.

[77] A W Anderson. Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magnetic Resonance in Medicine*, 46:1174–1188, 2001.

[78] C P Lin, W Y I Tseng, H C Cheng, and J H Chen. Validation of diffusion tensor magnetic resonance axonal fiber imaging with registered manganese-enhanced optic tracts. *NeuroImage*, 14:1035–1047, 2001.

[79] J.D. Tournier, F. Calamante, M.D. King, D.G. Gadian, , and A. Connelly. Limitations and requirements of diffusion tensor fiber tracking: An assessment using simulations. *Magnetic Resonance in Medicine*, 47:701–708, 2002.

[80] G J M Parker, K E Stephan, G J Barker, J B Rowe, D G MacManus, C A M Wheeler-Kingshott, O Ciccarelli, R E Passingham, R L Spinks, R N Lemon, and R Turner. Initial demonstration of in vivo tracing of axonal projections in the macaque brain and comparison with the human brain using diffusion tensor imaging and fast marching tractography. *NeuroImage*, 15:797–809, 2002.

[81] D S Tuch, J J Wisco, M H Khachaturian, L B Ekstrom, R Kotter, and W Vanduffel. Q-ball imaging of macaque white matter architecture. *Philosophical transactions of the Royal Society of London Series B*, 360:869–879, 2005.

[82] M A Koch, D G Norris, , and M Hund-Georgiadis. An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage*, 16:241–250, 2002.

[83] G J M Parker, G J Barker, N A Thacker, and A Jackson. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of anisotropic diffusion. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 1165, 2002.

[84] M Lazar and A Alexander. White matter tractography using random vector (RAVE) perturbation. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 523, 2002.

[85] T Behrens. *MR diffusion tractography: methods and applications.* Dphil thesis, Magdalen College, University of Oxford, 2004.

[86] Y Lu, Z Ding, J Qi, J C Gore, and A Ws Anderson. Probabilistic connectivity mapping using the asymmetric uncertainty of diffusion tensor imaging. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 1312, 2005.

[87] C Pierpaoli and P J Basser. Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine*, 36:893–906, 1996.

[88] W R Gilks, S Richardson, and G J Spiegelhalter. *Markov chain Monte Carlo in practice.* Chapman and Hall, 1996.

[89] O Friman and C-F Westin. Uncertainty in fiber tractography. In James S Duncan and Guido Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention*, volume 3749 of *Lecture Notes in Computer Science*, pages 107–114. Springer-Verlag, 2005.

[90] T Hosey, G Williams, and R Ansorge. Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magnetic Resonance in Medicine*, 54:1480–1489, 2005.

[91] G J M Parker, C A M Wheeler-Kingshott, and G J Barker. Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging. *IEEE Transactions in Medical Imaging*, 21:595–512, 2002.

[92] J D Tournier, F Calamante, D G Gadian, and A Connelly. A novel fibre-tracking technique: front evolution using a fibre orientation probability density function. In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 538, 2002.

[93] J Zhang, N Kang, and S E Rose. Approximating anatomical brain connectivity with diffusion tensor MRI using kernel-based diffusion simulations. In G E Christenssen and M Sonka, editors, *Information Processing in Medical Imaging*, volume 3565 of *Lecture Notes in Computer Science*, pages 64–75. Springer-Verlag, 2005.

[94] C Poupon, J F Mangin, CA Clark, V Frouin, J Régis, D Le Bihan, and I Bloch. Towards inference of human brain connectivity from MR diffusion tensor data. *Medical Image Analysis*, 5:1–15, 2001.

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

[96] R L O'Gorman and D K Jones. How many bootstraps make a buckle? In *Proceedings of the Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 225, 2005.

[97] M Lazar and A L Alexander. Bootstrap white matter tractography (BOOT-TRAC). *NeuroImage*, 24:524–532, 2005.

[98] D K Jones and C Pierpaoli. Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. *Magnetic Resonance in Medicine*, 53:1143–1149, 2005.

[99] M R Chernick. *Bootstrap Methods: A Practitioner Guide*. Wiley, 1999.

[100] G J M Parker and D C Alexander. Probabilistic anatomical connectivity using persistent angular structure obtained from diffusion weighted imaging. *Philosophical transactions of the Royal Society of London Series B*, 360:893–902, 2005.

[101] S Razzaque and J P Ralston. On the global anisotropy of cosmic ray data above 4×10^{19} eV. `http://dx.doi.org/10.1088/1475-7516/2003/07/007`, 2003.

[102] T C Onstott. Application of the Bingham distribution function in paleomagnetic studies. *Journal of Geophysical Research*, 85:1500–1510, 1980.

[103] J T Kent. Asymptotic expansions for the Bingham distribution. *Applied Statistics*, 36:139–144, 1987.

[104] D E Tyler. Statistical analysis for the angular central Gaussian distribution. *Biometrika*, 74:579–590, 1984.

[105] J S van den Brink, Y Watanabe, C K Kuhl, T Chung, R Muthupillai, M V Cauteren, K Yamada, S Dymarkowski, J Bogaert, J H Maki, C Matos, J W Casselman, and R M Hoogeveen. Implications of SENSE MR in routine clinical practice. *European Journal of Radiology*, 46:3–27, 2003.

[106] S M Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17:143–155, 2002.

[107] D K Jones and P J Basser. “squashing peanuts and smashing pumpkins”: How noise distorts diffusion-weighted MR data. *Magnetic Resonance in Medicine*, 52:979993, 2004.

[108] P A Cook, H Zhang, B Avants, P Yushkevich, D C Alexander, J C Gee, O Ciccarelli, and A J Thompson. An automated approach to connectivity-based partitioning of brain structures. In James S Duncan and Guido Gerig, editors, *Medical Image Computing and Computer Assisted Intervention*, volume 3749 of *Lecture Notes in Computer Science*, pages 164–171. Springer-Verlag, 2005.

[109] J C Gee, H Zhang, A Dubb, B A Avants, P A Yushkevich, and J T Duda. Anatomy-based visualizations of diffusion tensor images of brain white matter. In J Welckert and H Hagen, editors, *Visualization and Image Processing of Tensor Fields*. Springer, 2005.

[110] T E Behrens, H Johansen-Berg, M W Woolrich, S M Smith, C A 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. *Nature Neuroscience*, 6:750–757, 2003.

[111] A Dubb, P A Yushkevich, Z Xie, R C Gur, R E Gur, and J C Gee. Regional structural characterization of the brain of schizophrenia patients. In C Barillot, D R Haynor, and P Hellier, editors, *Medical Image Computing and Computer-Assisted Intervention*, volume 3217 of *Lecture Notes in Computer Science*, pages 688–695, 2004.

[112] B A Avants, P T Schoenemann, and J C Gee. Lagrangian frame diffeomorphic image registration: morphometric comparison of human and chimpanzee cortex. *Medical Image Analysis*, 10:397–412, 2006.

[113] M R Wiegell, H B W Larsson, and V J Wedeen. Fiber crossing in human brain depicted with diffusion tensor MR imaging. *Radiology*, 217:897–903, 2000.

[114] G Di Virgilio and S Clarke. Direct interhemispheric visual input to human speech areas. *Human Brain Mapping*, 5:347–354, 1997.

[115] A S David P W Woodruff, I C McManus. Meta-analysis of corpus callosum size in schizophrenia. *Journal of Neurology, Neurosurgery and Psychiatry*, 58:457–61, 1995.

[116] D J Best and N I Fisher. Goodness-of-fit and discordancy tests for samples from the watson distribution on the sphere. *Australian Journal of Statistics*, 28:13–21, 1986.

[117] E W Weisstein. Fibonacci numbers, from MathWorld - a Wolfram web resource. `http://mathworld.wolfram.com/PochhammerSymbol.html`.