Cardiff University

Robust Processing of Diffusion Weighted Image Data

by

Greg D. Parker

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
School of Computer Science and Informatics

June 2014

Cardiff University

<u>ABSTRACT</u>

SCHOOL OF COMPUTER SCIENCE AND INFORMATICS

<u>Doctor of Philosophy</u>

by Greg D. Parker

The work presented in this thesis comprises a proposed robust diffusion weighted magnetic resonance imaging (DW-MRI) pipeline, each chapter detailing a step designed to ultimately transform raw DW-MRI data into segmented bundles of coherent fibre ready for more complex analysis or manipulation. In addition to this pipeline we will also demonstrate, where appropriate, ways in which each step could be optimized for the maxillofacial region, setting the groundwork for a wider maxillofacial modelling project intended to aid surgical planning.

Our contribution begins with RESDORE, an algorithm designed to automatically identify corrupt DW-MRI signal elements. While slower than the closest alternative, RESDORE is also far more robust to localised changes in SNR and pervasive image corruptions.

The second step in the pipeline concerns the retrieval of accurate fibre orientation distribution functions (fODFs) from the DW-MRI signal. Chapter 4 comprises a simulation study exploring the application of spherical deconvolution methods to 'generic' fibre; finding that the commonly used constrained spherical harmonic deconvolution (CSHD) is extremely sensitive to calibration but, if handled correctly, might be able to resolve muscle fODFs in vivo. Building upon this information, Chapter 5 conducts further simulations and in vivo image experimentation demonstrating that this is indeed the case, allowing us to demonstrate, for the first time, anatomically plausible reconstructions of several maxillofacial muscles.

To complete the proposed pipeline, Chapter 6 then introduces a method for segmenting whole volume streamline tractographies into anatomically valid bundles. In addition to providing an accurate segmentation, this shape-based method does not require computationally expensive inter-streamline comparisons employed by other approaches, allowing the algorithm to scale linearly with respect to the number of streamlines within the dataset. This is not often true for comparison based methods which in the best case scale in higher linear time but more often by $O(N^2)$ complexity.

Cardiff University

SCHOOL OF COMPUTER SCIENCE AND INFORMATICS

In accordance with university regulation I make the following statements of fact.

Firstly, I declare that, except where indicated by specific reference, the work submitted is the result of my own investigation, as are the expressed views.

Secondly, I declare that no portion of the work presented within has been submitted in substance for any other degree or award at this or any other university or place of learning, nor is it being submitted concurrently in candidature for any degree or other award.

Finally, I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be available to outside organisations.

Signed .... Greg Parker

Contents

List of Figures

List of Tables

Nomenclature

Abbreviation	Explanation
b	Diffusion weighting (context: b-value)
$b0$	Image acquired with no diffusion weighting.
$B0$	High magnitude static magnetic field produced by MRI apparatus.
BEDPOST	Bayesian Estimation of Diffusion Parameters Obtained *(*cont.) using Sampling Techniques
CBCT	Cone Beam Computed Tomography (variation on CT)
CSHD	Constrained Spherical Harmonic Deconvolution
CSF	Cerebrospinal fluid
CT	(X-Ray) Computed Tomography
dODF	diffusion Orientation Distribution Function
dRL	damped Richardson-Lucy deconvolution
DSI	Diffusion Spectrum Imaging
DTI	Diffusion Tensor Imaging
DT-MRI	Diffusion Tensor Magnetic Resonance Imaging
DW	Diffusion weighted
DW-MRI	Diffusion Weighted Magnetic resonance imaging
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
FACT	Fibre Assignment by Continuous Tracking
FAST	FMRIB's Automated Segmentation Tool
FLIRT	FMRIB's Linear Image Registration Tool
FNIRT	FMRIB's Non-linear Imaging Registration Tool
fODF	fibre Orientation Distribution Function
HARDI	High Angular Resolution Diffusion Imaging
HOMOR	Higher Order Model Outlier Rejection
MD	Mean Diffusivity
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
ODF	Orientation Distribution Function
PCA	Principal Component Analysis

PAS-MRI	Persistent Angular Structure Magnetic Resonance Imaging
PDF	Probability Distribution Function
QBI	Q-ball Imaging
RESDORE	Robust Estimation in Spherical Deconvolution by Outlier Rejection
RESTORE	Robust Estimation of Tensors through Outlier Rejection
STEAM	STimulated-Echo Acquisition Mode
SH	Spherical Harmonics
T_1	Relaxation time through transferrence of energy
	(*cont.*) into the surrounding lattice.
T_2	Relaxation time through spin-spin dephasing.
T_E	Echo time.
T_I	Inversion time.
T_R	Repetition time.

Acknowledgements

I would like to express my gratitude to my supervisors, Derek Jones, David Marshall, Paul Rosin, Stephen Richmond and Nick Drage, without your individual inputs and discussion this work would not have been possible.

On a more personal level, I would also like to the members of my family and friends who gave me the help and support I needed finish. Without you, this document would only be notes.

Sincerely, thank you all.

Chapter 1

Introduction

This thesis is a study of methods for the robust, automated reconstruction of fibrous tissue through diffusion weighted magnetic resonance imaging (LeBihan et al. (1986)) for the purpose of maxillofacial modelling and, as a later extension, analysis of large quantities of neurological image data. The search for novel methods of modelling the human maxillofacial region is inspired by a number of important applications in the orthodontic field, most noticably functional and surgical simulation (Gladilin and Roginsky (2004), Beldie et al. (2010)).

The problem of accurately and reliably capturing subject specific maxillofacial muscular architecture has proven a difficult one, largely out of reach (with respect to reliability) of currently available imaging hardware; however, the lessons learned in the attempts to do so have implications beyond the immediate field, affecting not only study of muscle fibre but all other fibrous tissue including white matter.

The challenging nature of this project has provided me with additional opportunities to examine the fields of signal processing, anatomy, general magnetic resonance imaging, neuroscience and shape/object classification as well as may other minor incursions. In keeping with the established goal and examined fields of research, this thesis will relay my work on robust processing of diffusion weighted magnetic resonance imaging signals (Chang et al. (2005), Pannek et al. (2012)), analysis on the correct use of fibre orientation retrieval methods (Basser et al. (1994a), Tournier et al. (2007), Dell'Acqua et al. (2010)), examples of these applications to in vivo maxillofacial images (accompanied by a discussion of the difficulties of acquiring these images) and methods for automated segmentation of fibre bundles – both muscle and white matter (M. Maddah (2006), O'Donnell and Westin (2007), Smith and Nayak (2010)).

1.1 Motivaton

The work described in this thesis was inspired by the need to develop novel methods, work processes and algorithms required to accurately capture and represent maxillofacial soft tissue in vivo, with particular focus on the musculature. This need is driven by a clinical need to fill a hole in current maxillofacial modelling techniques that are beginning to be used for informing surgical planning and predicting surgical outcomes.

While more detail willl be provided in Chapter 2, in recent years there has been a push within health care towards the development of patient centred treatment practices. Within surgical orthodontics, this has expressed itself as a drive towards subject-specific modelling which has the potential to both improve prediction of surgical outcomes and aid understanding subject abnormalities (Beldie et al. (2010)). State-of-the art techniques for producing such models generally combine a skeletal component, acquired through cone beam computed tomography (Haunsfield (1972)) and a skin surface scan produced by, for example, high definition laser scanning systems (Moss (1989), Aung et al. (1995)). The missing component, and one that for the moment remains out of reach, is a subject specific soft tissue model; more specifically, information on the position, attachments and orientation of muscle tissue, forcing researchers to fall back on a generic model that (while often adequate) might compromise the subject specificity of the end result (especially if that subject were to have abnormal muscular anatomy).

The reason for this omission is fairly simple, while the first two components of such a model are retrievable through off-the-shelf means, delineation of the various soft tissues is far more difficult. The standard CT scan used to retrieve bone structure, for example, cannot provide sufficient soft-tissue contrast in order to achieve the feat, nor for obvious reasons can the surface scans. What is required is a third modality which, like the other methods, remains non-invasive and non/minimally destructive.

The method most closely matching these criteria is magnetic resonance imaging; a non-invasive non-ionizing technology able to produce numerous soft tissue contrasts, of which diffusion weighted (DW) MRI appears ideally suited, being the only technique able to both identify fibrous muscle tissue in vivo and determine its principal orientation (Basser et al. (1994b), Gilbert and Napadow (2005), Heemskerk et al. (2005)). This combination of localisation and orientation could, if accurately recovered, provide the basis for a more realistic muscular component in the desired maxillofacial models. With this in mind, the technical core of this work is devoted to adapting the DW-MRI and the associated processing methods – most commonly applied to white matter which possesses different diffusive properties – to accurate and reliable application within the maxillofacial region. The corollary being that, as with the acquisition of the bone and surface components, the requisite processes should require minimal human interaction (saving the clinician time/improving accessibility of the technology), thus as an addendum to adaptation of

DW imaging and processes, effort was made wherever possible to create efficient and automated processes.

This thesis presents a modular DW-MRI processing pipeline which, despite our focus on maxillofacial imaging, can be used (in part or entirety) for the processing of most DW-MRI image data. In short, each step in the proposed pipeline provides one of key the functions necessary to transform raw, uncorrected, DW-MRI image data into a robustly labelled segmentation of the underlying fibrous anatomy. While the functions themselves are agnostic to input data (making them adaptable to many DW-MRI imaging applications), in the context of the larger maxillofacial modelling project this pipeline is intended to provide accurate segmentation of muscle fibre bundles which, when used to supplant the generic muscle structures used by current maxillofacial models, should substantially improve the subject-specificity of the resultant maxillofacial model and, by extension, the accuracy of any resultant simulation – an obvious benefit of which is increased usefulness of such models in the planning of maxillofacial surgeries.

The described pipeline is comprised of three main components: (1) A tool for the detection and elimination of corrupt DW-MRI signal elements which, if left uncorrected, would introduce bias into resultant model fittings and thus subsequently reduce the accuracy of any derived information (for example, estimated fibre orientation(s) – a key property to this project). (2) An evaluation of two spherical deconvolution methods for fibre orientation estimation with particular focus on the results of selecting an inappropriate deconvolution kernel and their applicability to low anisotropy (i.e. muscle) tissue. In the context of the proposed pipeline this study allows us to both recommend a suitable fibre model and, importantly, methods to ensure that the application of that model produces an accurate result. (3) A fast an accurate methods for automatically segmenting whole volume streamline tractographies into individual fibre bundles (each representing a particular muscle or white-matter fasciculi). While segmentation of tractography data is an integral step in most DW-MRI applications, current methods, both automated an manual, are often extremely time consuming – either requiring expert knowledge and experience in the manual case, or due to $O(N^2)$ problem formulations in the automated case (thus introducing a quadratic scaling of computation as the number of streamlines, N, increases). To solve these problems we introduce a new automated method designed to operate in low-linear time, providing accurate results in a matter of minutes in comparison to hours, thus freeing the operator (or their hardware) to perform more pressing tasks.

Chapter 2

Background

The purpose of this chapter is to place the research described within this thesis into the broader context of both a larger maxillofacial modelling project and diffusion weighted MRI in general. As a result of these multiple disciplines and the variety of topics covered in the invidual chapters, an all-encompassing literature review at this point is not feasible due to the vast amount of accumulated material that would need to be covered. Instead, this chapter will provide an overview of those topics which are common across subsequent research chapters, with more topic specific literature analysis presented at the beginning of the chapter to which it is relevant. To begin, however, I would like to start by more firmly framing the overarching goal of this project, of which the work presented here will form a component, within the scope of maxillofacial modelling and surgical planning.

2.1   Surgical Planning: A more detailed motivation

The origins of this project stem from original research conducted within the School of Dentistry at Cardiff University, specifically from the output of work lead by Prof. Stephen Richmond and his group, of which I later became a member.

As mentioned within the introduction, healthcare has taken a recent turn towards patient specific treatment, arguably in part due to the recent availability of affordable 3D imaging hardware. In the field of orthodontics, one avenue of research is to enhance 3D technologies such as (stereo-)photometry, a process first introduced in the 1940/50s (e.g. Berghagen (1951)) to produce measurements and recreate 3D structures using a combination of 2D photographs and methods more commonly found in cartography – such as contour mapping – entailing a long and tedious manual process. While such methods were subsequently advanced; for example work by Burke and Beard (1967) who introduced a grid structure to the lighting to aid the identification of common points between photographs, manual workloads were still high, leaving the process ripe for automation or replacement.

Following the advent and availability of laser scanners capable of automatically and accurately producing 3D scans of an object surface, the days of surface photometry became numbered. Numerous groups began to publish work based on this new technology (Moss (1989), Aung et al. (1995), Ismail et al. (2002), Blais (2004)). Similar flurries also occurred upon the introduction of more advanced structured light methods (Nguyen (1999), Tuncay (2001)) and 3D video capture (Caruso et al. (1989). Trotman et al. (1996)), a technique that allowed accurate dynamic modelling and analysis of 3D facial surface animations – a key by-product of underlying muscular/skeletal function, and one that must be considered when performing surgery (natural looking surface movements being a key patient outcome).

With surface modelling and dynamics well covered, a logical step is to then consider subdermal anatomy. In the field of orthodontics (bypassing the ubiquitous 2D radiographs), computed tomography (CT, Haunsfield (1972)) and its more advanced cone beam variant became the mainstay techniques for accurate reconstruction of cranial skeletal structure (Kau et al. (2005)) and, interestingly can also be used to provide reasonable 3D surface modelling (Fig. 2.1), though not to the level of accuracy required for maxillofacial modelling (Kau et al. (2007)).

In possession of accurate surface and skeletal data, a subsequent logical step is to combine the two in order to produce a more complete subject specific maxillofacial model. Current state-of-the-art attempts to do exactly this. Beldie et al. (2010), for example, were able to combine a subjects 3D surface scan (captured using a Konica-Minolta laser scanner), digital dental moulds and cone beam CT to produce an accurate static model of a subject's head (Fig. 2.2a), the remaining component to complete and drive the
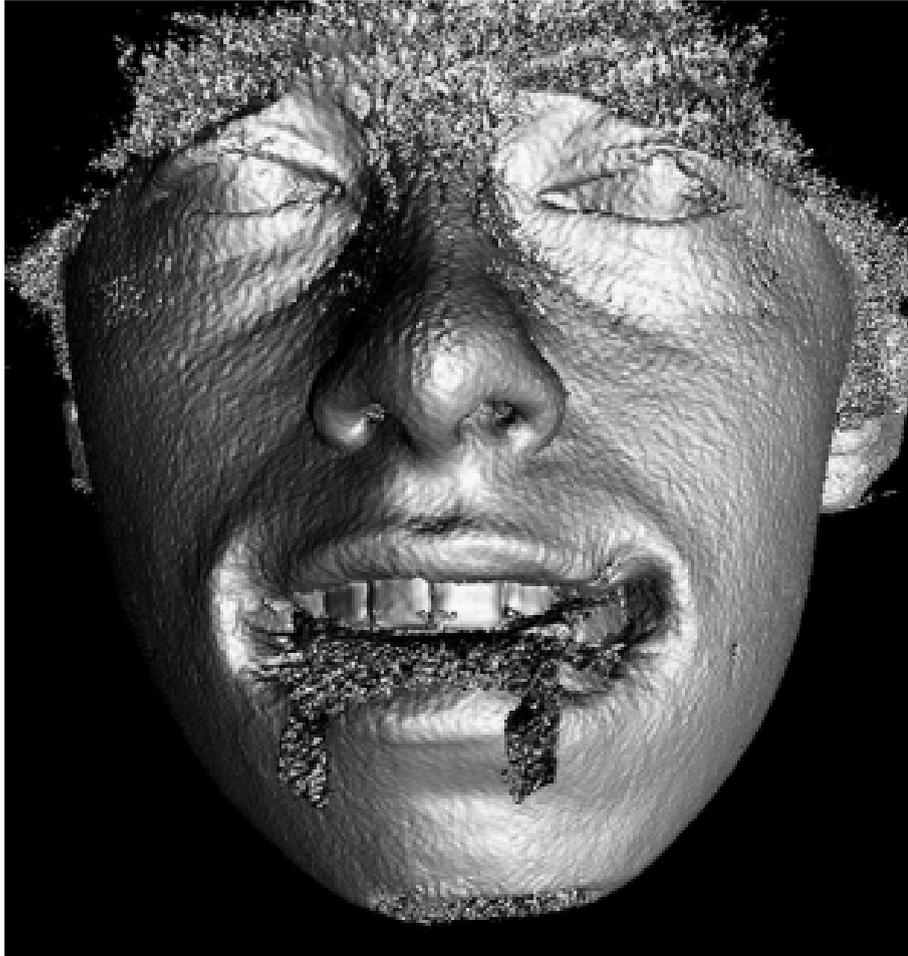
FIGURE 2.1: Example of a CBCT derived surface model. This figure is a reproduction of Figure 3 from Kau et al. (2007).

desired simulated articulation (the subject had pronounced mandibular prognathism, the surgery he was to undergo was designed to correct this and the model was intended to predict the effects of this correction with respect to normal facial movements) was a suitable model of the soft tissue architecture, specifically the muscle tissue. Unfortunately, the only internal images available were those acquired through CBCT. While able to distinguish between some forms of soft tissue, CBCT lacks the contrast required to map out the intricate muscular structures present throughout the human maxillo-facial region (Kau et al. (2007)). The effect on Beldies simulation was that the finite element (FE) model constructed to drive the skeletal articulation and surface deformation was based on a limited 20 muscle generic template, deformed to match the subjects anatomy. Individual muscles themselves were modelled taking the approximate shape of the muscle in question, but not necessarily capturing its underlying fibre geometry – thus omitting two key factors (accurate subject specific shape and architecture) required to accurately predict a muscles force output (Bovendeerd et al. (1994), a failing common with other similar methods, e.g. Gladilin and Roginsky (2004), Chabanas and Payan

(2000), Chabanas et al. (2002), Chabanas et al. (2003)). While the results of this particular simulation appeared promising, indeed the simulated expressions were qualitatively deemed successful (Fig. 2.2b), agreeing with the quantitative measurement of predicted v.s. actual skin surface outcomes, there is obvious room for improvement.
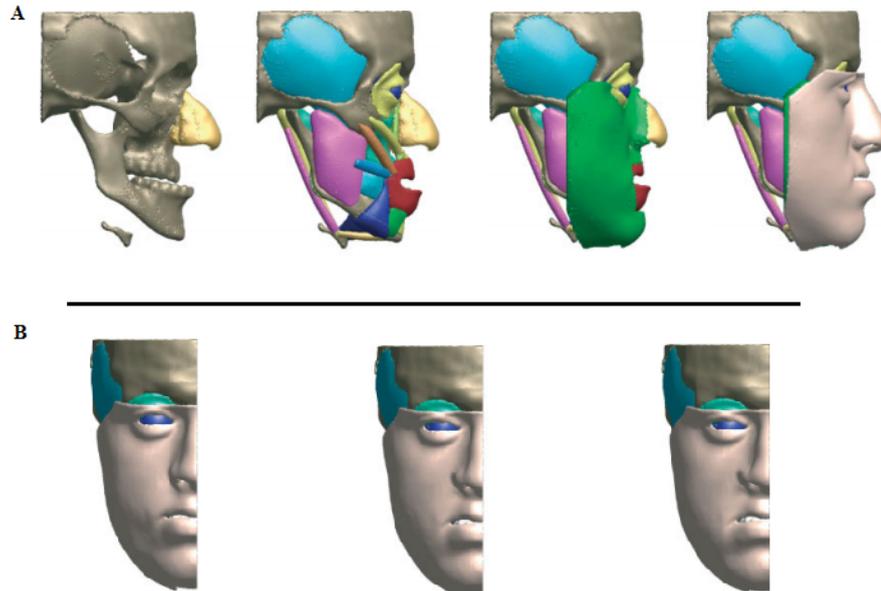
ra

FIGURE 2.2: A: Visualisation of the model construction process. Beginning with the skeletal structure, each component is built up in layers (muscle, then glanduar tissue/-fat, followed by the skin) to provide an adequate representation of the anatomy at work. Note the protruding chin that is indicative of mandibular prognathism. B: Still shots of the model transitioning from neutral to "disgust". Figure produced by combining figures 1 and 10 from Beldie et al. (2010).

Put simply, the eventual goal of my own research is to augment a Beldie type model with actual, subject-specific muscular architecture. While this was initially to take the form of producing a more accurate estimate of muscle size/location through segmentation of structural magnetic resonance images (T_1 or T_2 weighted images), perhaps modifying one of the techniques discussed by Bezdek et al. (1993) or Clarke et al. (1995) or perhaps more recent analogues (e.g. Gilles and Magnenat-Thalmann (2010)), this soon evolved into a hunt for accurate fibre architecture (due to the known association with muscle performance), leading to exploration of the diffusion-weighted MRI technologies described in later chapters and sub-sections.

2.2 Diffusion Weighted Magnetic Resonance Imaging

Magnetic resonance imaging is a relatively new technology capable of providing unprecedented power for observation of internal biological tissues without requiring the use of potentially harmful ionizing radiation or invasive procedures. At the core of MRI is the nuclear magnetic resonance phenomenon in which nuclei placed within a magnetic field will absorb and then re-emit electromagnetic (EM) radiation at a frequency dependant on the strength of the applied magnetic field and the gyromagnetic ratio of the nuclear species in question. While the NMR experiment has numerous uses, the most important variation with respect to this body of work is one producing a sensitivity to the Brownian motion of water molecules, thus allowing in vivo probing of fibrous tissues. The remainder of this section will therefore provide a briefing on the fundamentals of NMR, its adaptation to diffusion-weighted imaging, modern implementations of the DW-MRI paradigm and current DW-MRI processing techniques.

2.2.1 NMR to DW-MRI

At the centre of NMR is the concept of spin, a quantum mechanical magnetic property of nucleons. In fully paired nuclei, i.e. those possessing equal numbers of protons and neutrons in appropriate nucleon shells, each pair "cancels out" and thus there is no overall residual spin (spin quantum number $S = 0$). Where this is not the case, for example the hydrogen nucleus which is of particular interest to MRI due to its abundance within the human body (water of which ^1H is a constituent accounting for $\approx 60\%$ of human body mass), unpaired nucleons (e.g. the single ^1H proton) result in a non-zero overall spin, $S = 1/2 \times unpairednucleons$, which thus imparts a dipole magnetic moment over the nuclei with angular momentum M_s. For a given value of S, there exists $2S + 1$ possible M_s states which, in the absence of an external field, are degenerate and should thus be occupied with equal probability and oriented with random distribution.

Continuing with the hydrogen example, its spin quantum number of $1/2$ results in precisely two possible spin states, $1/2$ and $-1/2$. In the presence of a sufficiently strong external magnetic field (**B0**), interactions between **B0** and the nuclei's dipole moment have two effects. (1) Nuclei are forced into alignment with (or opposement to) **B0**, their spin vectors precessing about the fields axis with frequency $\omega = \gamma * \textbf{B0}$ – note that for ^1H the gyromagnetic ratio (γ) is approximately $42.58 MHz/T$, resulting in a $\approx$127MHz rate of precession at typical 3T MRI field strengths. (2) When spin vectors align with **B0** the energy levels of each spin state are no longer equal, spin states aligned with the field possess lower energy than those opposed and thus, in accordance with the Boltzmann distribution (Landau et al. (1996)), the distribution of spins will become slightly biased in favour of the lower energy state. With this skew in distribution it becomes possible

to excite lower energy nuclei to the higher energy level through interaction with electromagnetic radiation, the frequency of which is determined by the difference between energy states - in the case of hydrogen (with only 2 possible states) this is equal to the rate of precession, ω. (64MHz at 1.5T, 127MHz at 3T). Once excited, spins eventually return to the lower energy state, (re)emitting the absorbed energy as EM radiation at the same frequency (ω) thus forming the basis of the observed NMR signal.

Unfortunately it is not possible to observe individual nuclei undergoing such interactions in vivo, therefore with MRI we may only consider measurements on an (observable) macroscopic scale. Fortunately however, this does have the effect of simplifying the following explanations since we must only consider an ensemble result, rather than the underlying quantum mechanical formulations. Principally then, MRI can be explained in terms of an ensemble magnitization vector M, in a classical $[X \ Y \ Z]$ laboratory reference frame. Conventionally, the external field is considered to be aligned to and fixed along the positive Z axis. In the absence of **B0**, the orientational distribution of a sample of ^1H nuclei should be sufficiently random to produce no discernable ensemble magnitization, $M=[0 \ 0 \ 0]^T$. However, once **B0** is applied, the interaction of spins with **B0** will force them to align with ($M_s = 1/2$, $Z+$) or against ($M_s = -1/2$, $Z-$) the field which, remembering the thermodynamic bias towards the lower ($Z+$) energy state, results in an ensemble magnitization that may be schematically represented as $M =[0 \ 0 \ 1]^T$ – though in reality spin precession about the Z axis is initially forced into phase and will result in a small time variant $X - Y$ (transverse) component with angular frequency ω. As excitation through EM bombardment occurs, **M** will begin to rotate away from $Z+$ into the transverse and finally inverse ($Z-$) planes which, if observed from a fixed reference frame, would be observable as a downwards spiral with angular frequency (about Z) governed by the aforementioned precession. The degree of rotation of M can be governed by the length and magnitude of the EM (or radio frequency, RF) pulse, with increased magnitude or longer pulses causing a greater portion of the low energy spins to flip, increasing the rotation of M. For reasons that will be covered shortly, RF pulses are typically constructed to produce a 90° (transverse) or 180° ($Z-$) rotation. Once the RF pulse is terminated, the ensemble returns from its excited state towards thermal equilibrium with **B0**, resulting in an upwards spiralling M vector in a process known as relaxation.

There actually exist two principle forms of relaxation as described by the Bloch equations (Eq. 2.1-3, Bloch (1946)). The previously described Z axis spiral is known as T_1 relaxation and is a mechanism by which excited spins return to thermodynamic equilibrium by spontaneous transferrance of their absorbed energy into their surroundings (the "lattice") as heat, resulting in the longitudinal component of (M) (M_z) returning to **B0** alignment, the evolution of which may be formulated as:

$$\frac{dM_z}{dt} = \gamma(M_x B_x - MyBy) - \frac{M_z - M_0}{T_1} \tag{2.1}$$

where M_0 is the initial magnetization vector and $M_x/M_y/B_x/B_y$ represent the X and Y components of the magnetization vector M and total magnetic field $\mathbf{B}$ respectively – $\mathbf{B}$ being equivalent to $\mathbf{B0}$ plus a small fluctuating field, $\mathbf{B1}$, induced by the RF pulse in a direction perpendicular to the Z axis. T_1 here is a constant determined by the chemical "environment" in which the excited spins are present – interactions between the ^1H nuclei and neighbouring nuclei/molecules such as proteins or lipids may shorten or prolong the normal ^1H T_1. As an aside, it is environmentally dependant T_1 that provides a commonly used structural contrast within MRI since different organic molecules (present in different kinds of cellular structure) alter the surrounding ^1H T_1 and thus produce an anatomically linked variation in the resultant, measurable, time signature described by this equation.

The second form of relaxation, T_2, concerns a different process whereby excited spins achieve thermodynamic equilibrium amongst themselves through spin-to-spin energy transfers. Individual transfers have the effect of slightly altering the affected spins precession rate which, over a larger scale, results in a dephasing of transverse spin vector components causing an exponential decay of the transverse component of M. The time signatures of this decay may be formulated as:

$$\frac{dM_x}{dt} = \gamma(M_y B_z - M_z Bz) - \frac{M_x}{T_2} \tag{2.2}$$

and

$$\frac{dM_y}{dt} = \gamma(M_z B_x - M_x Bz) - \frac{M_y}{T_2} \tag{2.3}$$

Once again T_2 is a constant (though also dependant on proton environment) which, while not necessairly equal to T_1, can also not be larger.

In both cases, once RF excitation is terminated, the linked $\mathbf{B1}$ field no longer exists and as such $\mathbf{B}$ consists solely of (the Z aligned) $\mathbf{B0}$ field. As a result the previous equations may be simplified to:

$$\frac{dM_x}{dt} = \gamma M_y \mathbf{B0} - \frac{M_x}{T_2} \tag{2.4}$$

$$\frac{dM_y}{dt} = -\gamma M_z \mathbf{B0} - \frac{M_y}{T_2} \tag{2.5}$$

$$\frac{dM_z}{dt} = \frac{M_z - M_0}{T_1} \tag{2.6}$$

by eliminating B_x and B_y containing terms. These can then be integrated to give:

$$M_x(t) = (A cos(\gamma \mathbf{B0} t) + B sin(\gamma \mathbf{B0} t))e^{-t/T_2} \tag{2.7}$$

$$M_y(t) = (B cos(\gamma \mathbf{B0} t) - A sin(\gamma \mathbf{B0} t))e^{-t/T_2} \tag{2.8}$$

$$M_z(t) = M_0 + C e^{-t/T_1} \tag{2.9}$$

Where A, B and C are constants. Momentarily taking A equal to zero, we see that the X and Y components trace out a circle with radius decaying according to $Be^{-t/Te}$, each differing only in phase (sin vs. cos) as might be expected.

If one then places a receiver coil in the transverse plane at a time when M is tilted away from Z and no RF excitation is occurring, the transverse component of the magnetization vector will induce an alternating current in the receiver with frequency equal to the rate of precession and magnitude decaying exponentially as (1) M rotates back towards Z through T_1 relaxation and/or (2) the transvers component of M decays through T_2 spin-spin dephasing. Depending on the pulse sequence employed, it is possible to sensitise the received signal to either form of signal decay.

2.2.2 RF Pulse Sequences

As previously stated, it is possible to design RF pulse sequences to sensitise a recovered signal to either relaxation modality. One of the simplest T_1 sensitised sequences is known as inversion recovery, principally consisting of a 180° RF pulse followed, after time T_I, by a 90° pulse. The 180° pulse inverts M ($Z-$ alignment) which is then allowed to return to equilibrium (according to Eq. 2.9) but before this is completely achieved (T_I is selected to be less than the time required to achieve full equilibrium) a 90° pulse is applied, rotating the longitudinal component of M back into the transverse plane where it can be measured, producing a signal that may be approximated as:

$$M_0(1 - 2e^{-T_I/T_1}) \tag{2.10}$$

a particular instance of E.q. 2.9 in which $C = -2M_0$. If one then measures the signal amplitude after a range of $T_I's$, the T_1 relaxation time for a sample may be easily inferred.

T_1 relaxation is, however, not of principal interest to diffusion MRI – this falls to T_2. One of the most common T_2 measurement techniques is known as the spin-echo pulse

sequence (Hahn (1950), Carr and Purcell (1954)), consisting of a 90° pulse followed, after time $T_E/2$, by a 180° pulse. The 90° pulse rotates M into the transverse plane where the spins begin to de-phase through spin-spin relaxation. After the time $T_E/2$, the 180° pulse is applied, inverting the direction of precession – i.e. turning the de-phasing spin precession differences into re-phasing differences – such that after time T_E, spins are once again in phase and able to produce a strong measurable signal (which will then begin to de-phase again in the opposite direction). Again, if multiple measurements are made at different $T_E's$, the T_2 relaxation constant of a sample can be implied.

One benefit of this RF sequence that will later be exploited in diffusion MRI is that minor inhomogeneities in the local magnetic field that would, by themselves, result in spin-dephasing by causing local variations in spin precession rates ($\omega = \gamma B_0$, B_0 being no longer constant) are cancelled out at the time T_E (when the signal is measured) by the 180° inversion applied at $T_E/2$, thus the resulting signal can be approximated as:

$$M_0(1 - e^{-T_R/T_1})e^{-T_E/T_2} \tag{2.11}$$

Remembering that while we may be concerned with spin-spin relaxation, T_1 relaxation is also occurring as the M vector returns to $Z+$ alignment following the initial 90° pulse. If insufficient time is allowed between repeat measurements (i.e. the time between repeated measurements, T_R, is too short) then magnetisation will not be correctly flipped into the tranverse plane by the 90° pulse, and thus the resulting signal will also be attenuated by T_1 effects as modelled by the bracketed component.

2.2.3 Magnetic Field Gradients: NMR to MRI

A limiting factor in the usefulness of the described NMR experiments to in vivo work is that they lack the ability to spatially localise signals generated from within a subject volume – thus it is at this point that we move from NMR to MRI. One of the principal differences between the two techniques is that MRI extends the NMR method to include deliberate inhomogeneities in the **B0** field. These take the form of superimposed magnetic gradients which, while far weaker than the main field, are able to differentially alter spin precession rates and thus, through controlled and predictable alterations to spin frequency and phase, spatially encode resultant NMR signals when combined with RF pulse stimulation. What follows constitutes a very basic (spin-warp) gradient sequence designed to achieve these aims.

Imagining once again that B_0 is applied along the $Z+$ axis in a conventional Cartesian laboratory reference frame, an imposed magnetic gradient, G, can thus be described as a vector within that space with components G_x, G_y and G_z. The effect of gradient application is such that for a position $P = (P_x, P_y, P_z)$ within the imaged volume, the

rate of precession at that point will be shifted from $\omega = \gamma B_0$ by an amount equal to $\omega_{shift}(P) = \gamma(G_x P_x + G_y P_y + G_z P_z)$.

The first manner in which this shift can be exploited is in order to achieve slice selection – a process of selectively exciting only a thin layer of spins within total volume which, for convenience, are assumed to be stacked along the Z axis. Slice selection begins with the application of a Z axis gradient ($G_x = G_y = 0, G_z \neq 0$), creating a linear longitudinal frequency shift throughout the target volume that does not vary according to position in the transverse plane ($\omega_{shift} = \gamma(0P_x + 0P_y + G_z P_z)$). Remembering that spins may only be excited through RF pulses at a frequency matching their own rate of precession, slice selective excitation is then achieved through a narrow-band $90°$ RF pulse at frequencies designed to match the gradient induced shifts between the minimum and maximum Z axis coordinates of the slice of interest. Once this excitation is achieved, magnetic gradients and RF pulses may be terminated and only those spins within the targeted slice will possess the transverse magnetisation required to produce an NMR signal.

Once a volume has been segmented in the longitudinal axis, it is then necessary to achieve localisation in the transverse plane. The first step in the process is a phase encoding step most commonly performed along the Y direction and without the need for further RF pulses. Simply put, a Y-axis gradient is applied for a short time τ, momentarily altering rates of precession according to $\omega_{shift} = \gamma(G_y P_y)$ which, given that spins were initially precessing in phase across the entire slice, has the effect of inducing a Y-axis dependant phase shift $P(y) = \gamma \tau G_y P_y$, when the gradient is terminated – thus encoding the Y position of a spin as a function of its phase (for simplicity the effects of T_2 de-phasing are omitted).

The final step required for spin localisation is X axis encoding. Principally this is achieved in the frequency domain by applying an X aligned magnetic gradient which is maintained while the NMR signal is recorded. However, noting the Figure 2.3 schematic, a separate negative gradient is applied along the frequency axis during the phase encoding step, resulting in a further de-phasing (separate from the Y axis phase encoding) that is then reversed by the positive gradient during readout, creating the measured NMR echo with maxima at a predictable temporal location – the point at which the negative and positive phase changes along the frequency encode direction cancel out (Fig. 2.3b).

The issue with such a simple sequence is that only one Y axis phase offset may be encoded per repeated experiment as only one phase encoding gradient is applied. Since the recorded MR signal represents a slice wide ensemble, production of an image requires multiple measurements over an array of phase offsets (both X and Y) in order to resolve ambiguities and retrieve individual signal intensities from each voxel – note that while this may indeed be achieved as the solution to a simultaneous equation system, Fourier methods are generally preferred due to their relative computational simplicity. Thus,
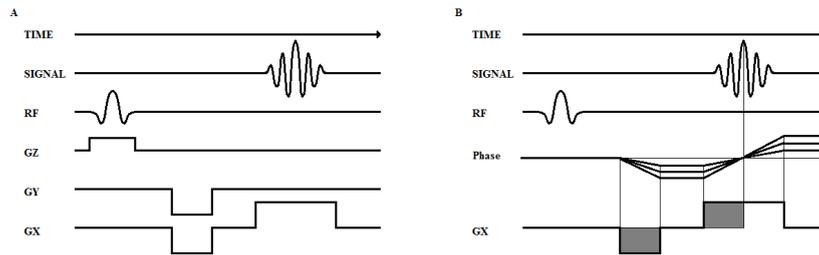
FIGURE 2.3: A: Diagram of a conventional gradient-echo pulse sequence. B: Evolution of phase along the frequency encode axis

to place ourselves in a better position to understand the particular limitations of this sequence and by extension provide a groundwork for more efficient techniques, it is prudent at this juncture to briefly examine the Fourier method and, in particular, the K-space formalism.

Conceptually, k-space is a spatial frequency domain in which the 2D Fourier transform of the targeted MR image slice is held. More practically however, k-space can be constructed as a 2D discretised matrix in which each row is populated by a discretely sampled raw slice-wide MR signal generated by a pulse/gradient sequence with known phase encode gradient magnitude, centred on the point at which the frequency encode induced phase dispersion is 0 (the point at which that phase dispersion cancels, resulting in a signal maxima as spins precess in phase). While the order in which rows are acquired may vary, their stacking within the matrix must be from largest positive phase encode gradient (up most) to largest negative (bottom most) with 0 phase encode gradient (the inflection point) at the centre. While this then explains the mechanical construction of k-space, understanding is achieved by stepping backwards slightly to consider the composition of the NMR signal. When applying a transverse gradient G, the phase shift induced after time t to spins at point P can be characterised as:

$$\phi(G, P, t) = 2\pi \int (\gamma B_0)dt + 2\pi \int (\gamma G_x P_x)dt + 2\pi \int (\gamma G_y P_y)dt \qquad (2.12)$$

and by extension, the slice-wide signal can therefore by estimated by integrating $\phi(G, P, t)$ across all P, giving:

$$S(G, t) = \int \int \rho(P)\epsilon^{i\phi(G,P,t)}dP_x dP_y \qquad (2.13)$$

where $\rho(P)$ is the spin density at point P. If we then substitute $k_x(G_x, t) = \int(\gamma G_x)dt$ and $k_y = \int(\gamma G_y)dt$, the signal (in a rotating frame) becomes:

$$S(G, t) = \int \int \rho(P)e^{2\pi i(k_x P_x + k_y P_y)}dP_x dP_y \qquad (2.14)$$

which can be recognised as a Fourier equation, implying a Fourier relationship between the MR signal and magnetic gradients (incorporated into k_x and k_y) applied in its production. If one then samples at various locations within k-space, the spin density (ρ) can be recovered using a 2D discrete Fourier transform.

In the previously described spin-warp sequence, the single phase encode gradient pulse applies a fixed phase offset with respect to Y axis position that does not change over the signal recording period. The effect with respect to k-space is that the Y axis coordinate (k_y) is fixed to a single value and thus any subsequently recorded signal would belong to a single vertical k-space level (i.e. a single row). In contrast, the paired negative/positive gradients applied along the frequency encode direction allow phase to evolve with respect to X during signal acquisition. The result of this phase evolution is that the rate of change of phase (in the spatial domain) with respect to X will be different at each discrete signal sampling point and thus, by extension, map to a different k_x coordinate – traversing k-space linearly from one extreme to the other. The combination of these phase manipulations is that for each pulse sequence repetition a single k-space row may be recorded (fixed Y, varying X) and thus, given that many rows must be recorded in order to produce an image with reasonable spatial resolution, is clearly time inefficient. The solution is fairly straight forward: attempt to recover as many k-space rows as possible per RF excitation. To this end I will now briefly cover the echo planar imaging (EPI, DeLaPaz (1994)) scheme which extends upon spin-warp to capture multiple rows per T_R. This basic sequence is also commonly used (with modifications) in diffusion MRI, the vast majority of non-simulated DW-MRI data used throughout this work, for example, was acquired through an EPI variant.

A basic EPI sequence (Fig. 2.4) begins much the same as spin-warp; initial slice selection is achieved through the same selected excitation method, combing a Z axis gradient with a targeted RF pulse. Breaking from that similarity, the first Y axis phase encode gradient is applied in the negative direction which is then flipped by a subsequent 180° RF pulse (with accompanying slice selective gradient) into the opposite k-space corner (equivalent to a positive gradient pulse without the 180° RF). Note, this combination of 90° followed by 180° RF pulses indicates a T_2 sensitization. Following the RF pulse, X axis encoding begins with the familiar short negative gradient pulse, creating the X axis phase gradient, which is subsequently reversed by a longer positive readout gradient facilitating k-space traversal. Where EPI then significantly differs from spin-warp is that the frequency encode gradient is allowed to continue to oscillate from positive to negative in a series of long gradient pulses – in essence phasing, de-phasing and then re-phasing (and so on) spins along the X axis, creating a series of RF echoes with maxima at each inflection point (0 phase). While this oscillation is occurring, the phase encode gradient is "blipped" each time the frequency encode gradient approaches 0 (as it swaps from positive to negative), altering the rate of phase change with respect to Y axis coordinate and by extension altering k_y. The net effect is that each successive echo induced by the

frequency encode oscillations corresponds to a separate k-space row, greatly increasing the area of k-space that may be traversed in a single repetition.
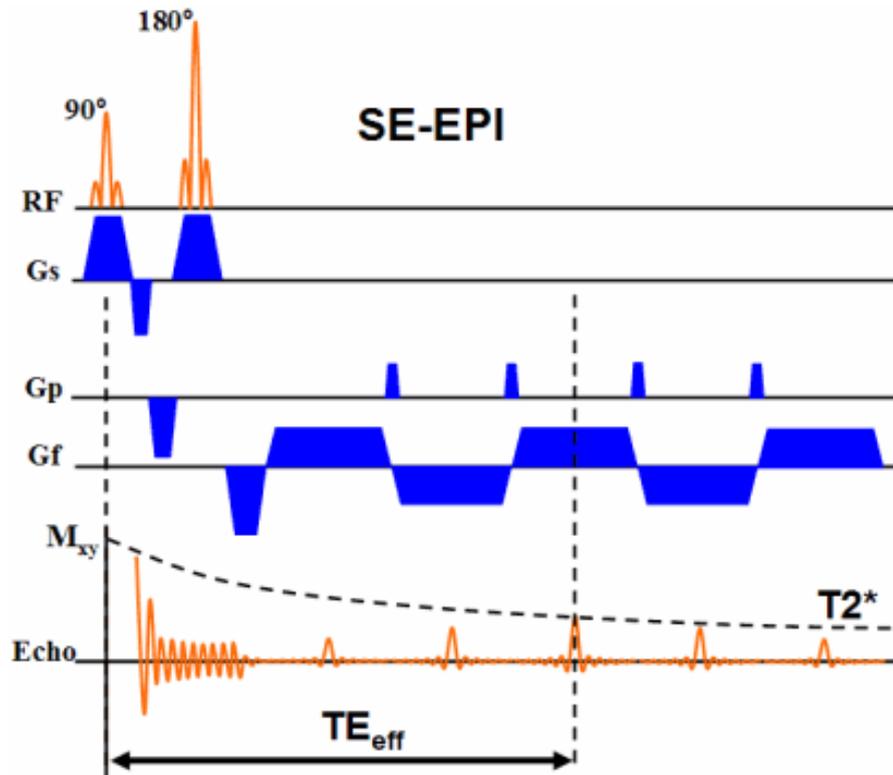
FIGURE 2.4: Schematic of a conventional EPI pulse sequence.

Setting aside the time related benefits (such as reduced imaging time and by extension reduced sensitivity to subject motion), there are drawbacks to EPI that must be considered, some of which are particularly relevant to the work presented within this thesis. Principally, EPI based acquisitions are sensitive to field inhomogeneities – be they caused by hardware, differing magnetic susceptibilities of the imaged subject (Jezzard and Balaban (1995)), or eddy currents generated by gradient slewing (Jezzard et al. (1998)) – and T_2* relaxation. Beginning with relaxation; as discussed in the pulse sequence subsection, a 90° followed by 180° excitation sensitises the received signal to T_2, or spin-spin, relaxation. As a result, from the instant the first 90° RF pulse is applied, the magnitude of the available signal is decaying according to the T_2 equation (Eq. 2.11), as such, each successive echo will produce a lower peak magnitude until, after a point, that magnitude reaches the noise floor and no further useful information may be retrieved. Practically speaking, this places a limit on the length of the echo train that may be recorded (and thus the area of k-space, directly affecting spatial resolution) while maintaining a useful SNR which varies according to the T_2 of the target spins. Unfortunately, the T_2 of muscle tissue, being much lower than that of white matter ($35 - 40ms$, Remeur et al. (1994) vs. $80ms+$, Wansapura et al. (1999)), curtails this

time period significantly, resulting in imaging difficulties that will be discussed in more detail within Chapter 5.

Secondly, on the subject of inhomogeneity, while the T_2 sensitised RF pulse sequence is reasonably insensitive to inhomogeneities (the effects of magnetic inhomogeneities on phase "cancel out" at the time of echo due to the 180° re-phasing flip), this is not the case when one also requires localisation. Briefly, the accuracy of image reconstruction relies on the ability to pinpoint a spin's location via frequency and phase encoding. A key assumption at this point is that the magnetic field remains uniform or, when gradients are applied, varies uniformly across the imaged volume such that a spin's expected frequency (and thus X axis position) may be easily determined through the $\omega(P) = \gamma B_0 + G_x P_x$ relationship. In practice, however, this is often not the case. Magnetic inhomogeneities distort the rate of precession of spins throughout the imaged volume which, depending on the magnitude of local distortions, may result in local image stretching or pile up (examples in Fig. 2.5) along the phase encode direction as expected frequency and position no longer correspond. Ignoring hardware deficiencies, one of the most common sources of inhomogeneity is magnetic susceptibility, i.e. the degree to which a material magnetizes itself in response to an external magnetic field. If one considers the human body, there are numerous materials (e.g. bone, muscle, grey matter, white matter, blood, air etc.) each possessing a different chemical composition and thus each possessing a different overall (bulk) susceptibility (Collins et al. (2002)). The consequence of these differences is that each material, when placed within a magnetic field, distorts that field in a slightly different manner. In isolation these distortions are not necessarily problematic; while the applied magnetic field would not match the expectation, the degree of distortion would be uniform (no inhomogeneity) and thus result in a uniform frequency shift that would merely translate the subject tissue within the imaged volume. This is roughly speaking the case with white/grey matter, both tissues possess similar bulk susceptibilities and thus distort B_0 in a similar manner creating a largely homogeneous region within the central brain allowing for easy imaging (while susceptibility is still an issue, this is more commonly a result of air/tissue differences affecting regions surrounding, for example, the sinuses).

In contrast with the brain, the maxillofacial region presents far more opportunities for magnetically inhomogeneous materials to exist in close proximity – consider the number of bone/muscle/air/skin/metal (fillings) boundaries within the mouth alone – creating unpredictable field inhomogeneities and thus image distortions at each intersection. As will be demonstrated and discussed in greater detail within Chapter 5, combating EPI's susceptibility to the resultant distortions will become a major challenge for accurate maxillofacial muscle modelling.
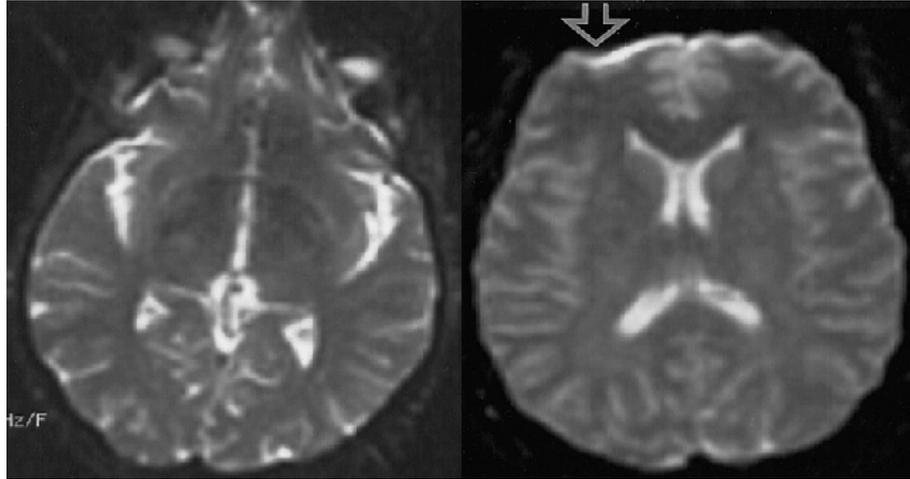
FIGURE 2.5: Examples of EPI image distortions caused by magnetic inhomogeneity. Left: Image stretching along the phase encode axis. Right: Signal pile up (hyperintensity) along the phase encode axis.

2.2.4 Diffusion

Before continuing to diffusion weighted MRI, it would first be prudent to discuss the diffusion process itself and the tissue interactions that lead to its usefulness as a biological probe. Phenomenologically, i.e. following Fick's laws (Fick (1855), Fick (1995)), diffusion can be defined as a transport phenomenon resulting in a mixing or mass transport without requiring bulk motion. Another way of viewing diffusion, as described by Einstein (1905), is that all particles undergoing diffusion all posess some level of kinetic energy determined by their thermal conditions and, as such, are constantly moving through space. What separates diffusion from other transport phenomenon, such as flow, is that the direction of movement of each particle is random and can change from time to time as the result of collisions. The result is no overall movement – i.e. a random walk/Brownian motion. While the most commonly studied forms of diffusion involve the observation of concentration gradients, it is equally possible to study the movement of particles within a homogeneous fluid (such as pure water) undergoing a process known as self diffusion; both processes being easily explainable through classical mechanics.

Starting at a phenomenological viewpoint, Fick's first law (Fick (1855)) states that the flux, J, of a diffusing substance is proportional to the spatial gradient of the substance concentration:

$$J(x,t) = -D\frac{dC(x,t)}{dx} \tag{2.15}$$

where D is the diffusion coefficient and $C(x,t)$ is the concentration at position x and time t. Taking into account the conservation of mass, the rate of change of concentration

at a given position must match the divergence of the local flux:

$$\frac{dJ(x,t)}{dx} = \frac{-dC(x,t)}{dt} \tag{2.16}$$

and thus Fick's second law states:

$$\frac{dC(x,t)}{dx} = D\frac{d^2C(x,t)}{dx^2} \tag{2.17}$$

providing a more complete picture of the evolution of concentration over time. If D may be considered constant, and the initial condition $C(x,0)$ is set to a Dirac delta function (δx) – as specified by Einstein (1905) – then Fick's second law can be solved to give:

$$C(x,t) = \frac{1}{\sqrt{4\pi Dt}}e^{\frac{-x^2}{4Dt}} \tag{2.18}$$

indicating that as time progresses, the concentration of particles spreads according to a dilating Gaussian function with zero mean and variance $2Dt$, i.e. $<x^2> = 2Dt$. Assuming isotropic diffusion, generalisation to three dimensions can be achieved with slight modification:

$$p(x(t)|x(0)) = \frac{1}{4\pi Dt}e^{-\frac{(x(t)-x(0))^2}{4Dt}} \tag{2.19}$$

where $p(x(t)|x(0))$ indicates the probability that a particle beginning at position $x(0)$, not necessarily $x = 0$, will arrive at a point $x(t)$ after time t; which is characterised by a gaussian probability function with familiar $2Dt$ variance but now $x(0)$ mean displacement. Expanding upon this, the mean squared displacement a particle will undergo over time t can be approximated as the sum of the variances along each direction which, in three dimensions, is equal to $3 \times 2Dt = 6Dt$.

In the human body, however, diffusion need not be isotropic – Chenevert et al. (1990), Beaulieu and Allen (1994) and Henkelman et al. (1994), for example, provide some of the earliest evidence of measured in vivo/vitro diffusion anisotropy within both neurological and muscular tissues (Beaulieu (2002) also goes on to provide an excellent overview of anisotropy in general). To understand the cause of this anisotropy, it is perhaps easiest to consider a simplified model of the kinds of tissue that DW-MRI routinely targets. Principally, water contributing to an observed DW-MRI signal may exist in either the intra- or inter-cellular compartments. While water within these compartments does mix (Saffman and Delebruck (1972) describes diffusion across a permiable cellular membrane), the comparatively slow rates at which this sometimes occurs (C. Meier (2003), for example, reports mixing times in excess of $700ms$, an order of magnitude greater

than typical $50 - 100ms$ DW-MRI image experiments) allows us in certain cases to make the slow diffusion simplification – i.e. model the compartments as being separated by an impenetrable cellular membrane preventing meaningful mixing (this is an assumption used by, for example, Assaf and Basser (2005) in CHARMED). Beginning with the intra-cellular compartment, the maximum displacement a water molecule can undergo is then limited by the maximum distance between boundaries of the membrane within which it is encased; thus, after a time t, measured displacements begin to plateau as more and more molecules hit the boundary until molecules may displace no further for any subsequent t (Fig. 2.6a). This leads to the label *restricted* diffusion. The extra-cellular compartment is not restricted in such a manner, molecules are free (again, over the time-course of the DW-MRI experiment) to diffuse infinitely in all directions; however, while undergoing diffusion these molecules will routinely collide with cellular membranes, *hindering* the rate at which molecules displace over a given time; such that for a given time t, the measured mean squared displacement will not match that of the $6Dt$ expected of pure water (Fig. 2.6b).
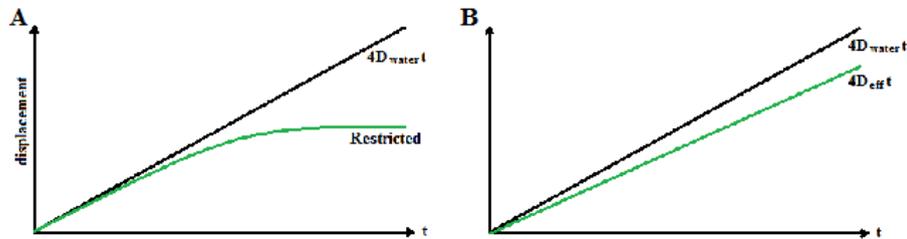
FIGURE 2.6: Schematic representation of mean squared displacement vs. time for (A) restricted and (B) hindered diffusion. Note that after some time, t, restricted diffusion prevents further displacement whereas hindered diffusion simply proceeds, though at a lower rate (D_{eff}) than that expected of pure water in isolation (D_{water}).

With this in mind, we may now consider the effects of different cellular geometries and organisation which, for a simplistic view, may be broken down into three classes – "spherical" cells, organised fibre and disorganised fibre. Within the spherical cell, the maximum distance a water molecule may displace along any direction is equal to the cellular diameter, thus while restricted, (ignoring other cellular contents such as the nuclei) intra-cellular diffusion is equally restricted in all directions and thus isotropic. Extending this externally, if spherical cells are arranged in a uniform manner, the probability of encountering a cellular membrane within a given distance along any direction is, on average, equal. Again, while the rate of diffusion in the hindered compartment is lower than pure water, it is equally hindered in all directions and thus considered isotropic – another way of thinking about this is that the presence of the cellular matrix simply lowers the *effective* diffusion coefficient equally and in all directions.

Proceeding to fibrous tissue, in the restricted compartment the maximum possible displacement now depends on the direction of travel, the elongated fibrous cell geometry means that it is possible to displace further parallel to the fibres axis than orthogonal, as such restriction is no longer isotropic (Fig 2.7a). Effects on extra-cellular water depend

on the fibrous cells organisation. If cells are strewn at random directions, e.g. within grey matter, the net effect is much like the spherical cells – the probability of encountering a cellular membrane is, on average, equal in all directions and thus so is the degree of hindrance. As a result, grey matter consisting of fibrous cells that would produce pronounced anisotropy on the microscopic scale result in near isotropic diffusion on the measurable macroscopic level. However, if these fibrous cells are aligned with each other, such as in muscle or white matter, the probability of encountering a cellular boundary when diffusing along the bundle is far lower than when diffusing across it; this directionally variant hindrance in turn creates a directionally variant anisotropy in measurable diffusion with diffusivity along the fibre pathway higher than diffusivity orthogonal to the bundle, varying in an approximately Gaussian manner that may be described in the form of a diffusion tensor by modifying the constant $2Dt$ variance in Equation 2.19 to a 3D symmetric covariance matrix (Fig. 2.7b).

$$2Dt = \begin{bmatrix} 2D_{xx}t & 2D_{xy}t & 2D_{xz}t \\ 2D_{xy}t & 2D_{yy}t & 2D_{yz}t \\ 2D_{xz}t & 2D_{yz}t & 2D_{zz}t \end{bmatrix} \tag{2.20}$$

Note that while the assumption of Gaussian diffusivity specified by these equations holds relatively well for regions containing a single fibre co-linear population, voxelisation of the imaged region often leads to individual voxels containing multiple non co-linear fibre populations that produce an ensemble signal (and thus implied diffusivities) that do not match such expectations. There are of course numerous non-Gaussian diffusion models to cover this eventuality, an overview of which will be provided within the high angular resolution diffusion imaging (HARDI) subsection (Section 2.2.7-13).
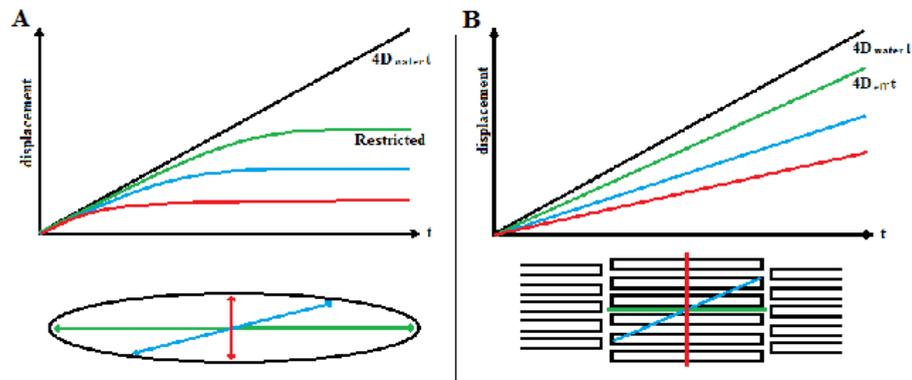
FIGURE 2.7: Schematic of orientationally variant mean squared displacement for (A) restricted and (B) hindered compartments.

2.2.5 Measuring Diffusion Through MRI: DW-MRI

In order to convey the principles behind diffusion measurement through MRI it is easiest to begin by examination of a DW-MRI pulse sequence, one of the simplest and earliest of which is the Pulse Gradient Spin Echo (PGSE) method described by Stejskal and Tanner (1965), a schematic of which is presented in Figure 2.8.
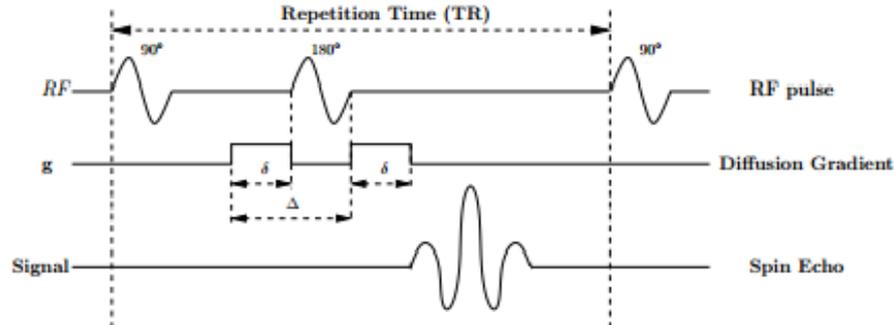
FIGURE 2.8: Schematic representation of the Stejskal-Tanner pulse gradient sequence.

PGSE ostensibly consists of two matching magnetic field gradients with magnitude G applied for duration δ at spacing Δ, positioned either side of the 180° RF pulse of a standard T_2 sensitising 90/180° spin-echo RF sequence. The first gradient imparts a positionally dependant phase shift on the excited spins which, if they do not move, will be eliminated by the second pulse, thus returning spins to precession in phase, producing a strong RF echo across the volume. In reality however, spins will (as part of water molecules) diffuse over the time Δ such that when the second gradient pulse is applied, the phase changes accrued under each gradient pulse will not cancel out (different position $\rightarrow$ different magnitude), resulting in a residual phase dispersion when compared to non-displaced spins at the current position. In the ensemble, the further and more frequently spins displace from their original position relative to the gradient axis, the larger the residual phase dispersions be within a given volume which by extension leads to greater T_2 signal attenuation.

According to Stejskal and Tanner (1965), the relationship between diffusivity and signal attenuation can then be modelled as:

$$S = S_0 e^{-bD} \tag{2.21}$$

where S is the expected signal magnitude, S_0 is the signal magnitude without diffusion sensitising gradients, D is the effective diffusivity and $b = -\gamma^2 G^2 \delta^2 (\Delta - \delta/3)$, more commonly referred to as the b-value or diffusion weighting (units, s/mm^2). As an aside, S_0 is usually captured when G and thus b is zero and is often referred to as the $b0$

image. By trivial rearrangement, diffusivity (D) along the magnetic gradient can then be approximated as:

$$D = \frac{-ln(S/S_0)}{b} \tag{2.22}$$

where S and S_0 are recorded during MR imaging and b is a known, user specified, parameter.

2.2.6 DW-MRI Signal Processing Models and Metrics: Diffusion Tensor MRI

One of the simplest DW-MRI processing techniques is to fit the self diffusion tensor described in Equations 2.19 and 2.20 to observed DW-MRI signals (Diffusion tensor imaging, DTI or DT-MRI, Basser et al. (1994a)). Considering that the gradient G may be applied along any 3D orientation as a combination of orthonormal gradients with magnitudes G_x, G_y and G_z, Equation 2.21 can be adapted to show that the normalised attenuation expected along any G can be related to the diffusion tensor through:

$$ln(\frac{S}{S_0}) = \sum_{i=1}^{3} \sum_{j=1}^{3} b_{ij} D_{ij} \tag{2.23}$$

which, for clarification, becomes:

$$ln(\frac{S}{S_0}) = -(b_{xx}D_{xx} + 2b_{xy}D_{xy} + 2b_{xz}D_{xz} + b_{yy}D_{yy} + 2b_{yz}D_{yz} + b_{zz}D_{zz}) \tag{2.24}$$

when expanded, where b_{xy} (for example) is calculated according to $\gamma^2 G_x G_y \delta^2 (\Delta - \delta/3)$ and D_{ij} is the $i'th$ row $j'th$ column entry in the 3×3 diffusion tensor covariance matrix. The multiples of two for the off diagonal components, e.g. xy, account for the D matrix symmetry, indicating that only 6 unique parameters must be estimated in order to estimate the diffusion tensor. If one rearranges Equation 2.24 in a similar manner to 2.22, signal measurements can be acquired over a minimum of 6 unique gradient directions (resulting in 6 unique b-matricies and 6 unique DW-signal intensities) plus one non-diffusion weighted image (S_0), leaving the components of diffusion tensor as the only unknown parameters and thus allowing them to be estimated through regression of the resultant equation system (for which there are numerous popular methods ranging from multivariate linear regression as originally proposed by Basser et al. (1994a), to more complex robust techniques, e.g. Chang et al. (2005) or Mangin et al. (2002)).
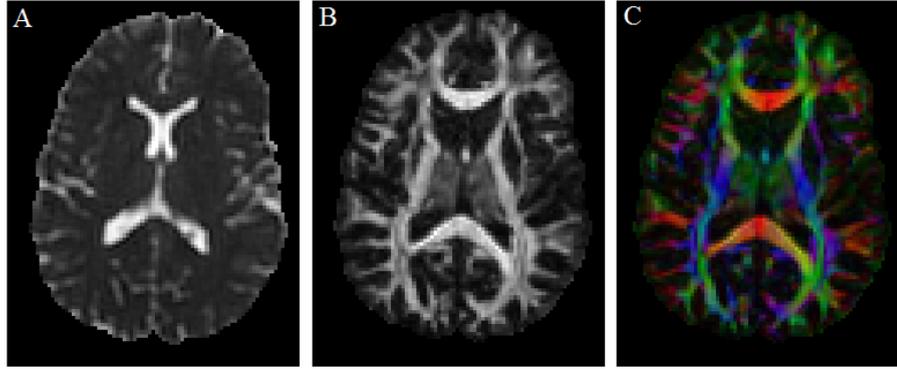
FIGURE 2.9: DTI derived quantitative metrics and visualisations. A: Mean diffusivity.
B: Fractional Anisotropy. C: DEC Map.

Following estimation and eigenanalysis, the diffusion tensor can provide a range of useful information. For signals resulting from a single co-linear fibre population, the eigenvectors of a diffusion tensor can be thought of as describing an ellipsoid with its long axis, i.e. the direction of greatest variance described by the 1st eigenvector/value, aligned to the direction of greatest diffusivity and by proxy the orientation of that fibre bundle – a necessary property for fibre tractography. Proceeding to eigenvalues, each eigenvalue then provides an estimation of the effective diffusivity along the corresponding eigenvector, allowing calculation and mapping of statistics such as the trace (used clinically to, for example, diagnose stroke by Gonzales et al. (1999)):

$$Trace = \lambda_1 + \lambda_2 + \lambda_3 \tag{2.25}$$

which may alternatively be calculated (without eigenanalysis) as $Trace = d_{xx} + d_{yy} + d_{zz}$, i.e. the sum of the diagonal elements of the diffusion tensor, and from which we may then derive Mean Diffusivity (MD, Fig. 2.9a):

$$MD = \frac{Trace}{3} \tag{2.26}$$

and Fractional Anisotropy (FA, Fig. 2.9b):

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{2.27}$$

a measure introduced by Pierpaoli and Basser (1996a) providing an 0-1 scaled index of anisotropy, within which healthy white matter tends to produce measurements at the upper end of the scale (0.5-0.9) owing to its tightly packed narrow axon diameter arrangement providing strong hindrance perpendicular to the fibre orientation; whereas muscle, with less densely packed and larger fibrous cells with arguably more porous cellular membranes (axons possess a myelin sheath not present in muscle tissue) hinder

transverse diffusion far less severely, resulting in a significantly lower FA measure ($\approx$ $0.2 - 0.3$).

Using the first eigenvector as an index into the RGB colour space and FA as a brightness scaling factor, it is then possible to create directionally encoded colour (DEC, Pajevic and Pierpaoli (1999), Fig 2.9c) maps commonly used to quickly impart both the presence of fibrous tissue (high brightness implying high anisotropy which, by extension, might imply a coordinated fibrous structure) and its orientation (Y axis red, X axis green, Z axis blue, combinations of which depict non axially aligned fibre orientations).

As previously alluded to, DT-MRI does however have a serious limitation. By construction, the diffusion tensor formulation expects either isotropic diffusion ($\lambda_1 = \lambda_2 = \lambda_3$) or anisotropic diffusion that may easily be described in an ellipsoidal Gaussian manner such as that produced by a single co-linear fibre population ($\lambda_1 > \lambda_2 \approx \lambda_3$).

The problem, as shown by Jeurissen et al. (2013), is that at common spatial resolutions these assumptions frequently do not hold. Specifically, Jeurissen shows that in the human brain a large portion (upto $\approx 90\%$ according to CSHD, Tournier et al. (2007), or $\approx 63\%$ with BEDPOSTx, Behrens et al. (2007)) of white matter containing voxels contain signal contributions from non-colinear sources – crossing, kissing or fanning fibre populations – creating the Partial Volume Effect (PVE, Alexander et al. (2005)). Put simply, the presence of two or more fibre populations within a voxel creates a non-Gaussian diffusion probability density function (PDF), and thus DW-MRI signal profile, that while potentially describable under certain circumstances (crossing, but not fanning) through a combination of two or more overlaid tensors (Tuch (2002)), can never be accurately described using one tensor. A common result, for example, is that if a single diffusion tensor is fit to a two-fibre signal, the resultant description is generally oblate ($\lambda_1 \approx \lambda_2 > \lambda_3$) and, while it may lie in the plane described by the orientation of each fibre, cannot accurately describe the orientations of either due to the inherent limitation of attempting to describe two principle directions with a single eigenvector. Thus, in order to surpass this Gaussian diffusion limitation, other models must be employed.

2.2.7   Methods for Resolution of Non-Gaussian Diffusion

The non-Gaussian label currently encompasses a range of methods intended to overcome the limitations of the diffusion-tensor model or, more precisely, provide disambiguation of crossing fibre orientations – fibre fanning presents additional challenges that will be discussed briefly in Section 2.2.14. For the purposes of this chapter we will attempt to provide an overview of some of the main technologies; first focusing on q-space methods (for their relevance to the field) and spherical deconvolution techniques (for relevance to this work) but then al so briefly covering other interesting alternatives.

2.2.8   Q-Space

Returning to the topic of gradient induced phase shifts; combining the assumptions of a narrow gradient (whereby $\delta << \Delta$, ensuring the effects of diffusion during gradient application are negligible) and instantaneous gradient ramping (i.e. gradient application follows the square waveform, ensuring the full magnitude is applied across all δ), then the phase (ϕ) accrued by a spin starting a position $x(0$ during the first pulse of a PGSE experiment can be estimated as:

$$\phi_0 = \gamma \delta G \cdot x(0) \tag{2.28}$$

where γ and G are, respectively, the gyromagnetic ratio and gradient vector. Using the same logic, we can also assume that the phase accrued during the second pulse, with spin at position $x(\Delta)$ can be estimated as:

$$\phi_\Delta = \gamma \delta G \cdot x(\Delta) \tag{2.29}$$

such that, given the second gradient vector is designed to negate the first (equal in magnitude, opposite in orientation), zero net phase is accrued by a stationary spin over the course of the experiment. Should that spin not remain stationary, however, the resultant signal after a net phase shift of ϕ_{net} can be approximated as:

$$e^{i\phi_{net}} = e^{i\gamma \delta G \cdot (x(0) - x(\Delta))} \tag{2.30}$$

However, given that it is not possible to record the signal produced by a single spin, we must once again consider the ensemble result. It is therefore helpful to begin to think of the expected signal produced by a particular spin as the product of two probabilities; the first ($P(x(\Delta)|x(0), \Delta)$) describing the probability of a spin starting at position $x(0)$ and then diffusing to position $x(\Delta)$ over time Δ, and the second ($f(x(0), 0)$) describing the possibility of actually finding a spin at position $x(0)$ at time 0 (logically, if no spin was present, no signal would be expected). Expanding upon this line of thought, the signal (S)) produced by any given spin ensemble can then be modelled as the summation of all signal produced by spins starting (with known probability) at all possible $x(0)$ and then, over time Δ, diffusing to all possible $x(\Delta)$ (which includes the possibility of the spin remaining stationary). This can be expressed in the form:

$$S(G, \Delta) = \int e^{i\gamma \delta G \cdot (x(0) - x(\Delta))} P(x(\Delta)|x(0), \Delta) \int f(x(0), 0) \ dx(0) dx(\Delta) \tag{2.31}$$

which, if one then defines:

$$q = \frac{\gamma \delta G}{2\pi} \tag{2.32}$$

as specified by Callaghan et al. (1988), can be rewritten as:

$$S(q, \Delta) = \int e^{i2\pi q \cdot (x(0) - x(\Delta))} P(x(\Delta)|x(0), \Delta) \int f(x(0), 0) \ dx(\Delta) dx(0) \tag{2.33}$$

thus linking the expected signal to the q-space gradient wave vector. Furthermore, if we then recall that according to Brownian motion, each spin is essentially following its own independent random walk, the probability $P(x(\Delta)|x(0), \Delta)$ can actually be thought of as equivalent to the to the diffusion propagator and, since on a single voxel scale the signal only depends on spin displacement, can be rewritten in this instance as $P(\Delta x, \Delta)$ where $\Delta x = x(0) - x(\Delta)$, i.e. a displacement vector. Building upon this, the ensemble average propagator for any given voxel, $\overline{P}$, describing the probability of a spin in that ensemble achieving displacement Δx over the period Δ can be written as:

$$\overline{P}(\Delta x, \Delta) = \int P(\Delta x, \Delta) f(x(0), 0) \ dx(0) \tag{2.34}$$

which when combined with Equation 6 produces:

$$S(q, \Delta) = \int \overline{P}(\Delta x, \Delta) e^{i2\pi q \cdot \Delta x} \ d\Delta x \tag{2.35}$$

establishing an inverse fourier relationship between the ensemble average propagator and the observed signal. It was this relationship that was then seized upon by Wedeen et al. when establishing Diffusion Spectrum Imaging (DSI, Wedeen et al. (2000), Wedeen et al. (2005)); a technique that attempts to reconstruct $\overline{P}$ through a Fourier transform of the normalised (with respect to S_0, the non diffusion weighted intensity) signal across a densely sampled q-space:

$$\overline{P}(\Delta x) = \int S(q) e^{-2\pi i q \cdot \Delta x} \ dq \tag{2.36}$$

which, to account for physiological noise creating unwanted phase shifts (corrupting the recovered image), was eventually modified to the modulus Fourier transform:

$$\overline{P}(\Delta x) = \int |S(q)| e^{-2\pi i q \cdot \Delta x} \ dq \tag{2.37}$$

in which S now implies the normalised signal. With the ensemble average propagator recovered, it was then shown to be possible to approximate underlying fibre orientation through interrorgation of its angular structure:

$$ODF(u) = \int\limits_{\mathbb{R}+} \overline{p}(xu)x^2 \ dx \tag{2.38}$$

which stipulates that the magnitude of the (diffusion) orientation distribution function (ODF) along a direction defined by the unit vector u can be approximated by the summation of all $\overline{P}$ along the associated radial projection (as indicated by the scaling factor x). If we were to then repeat this process across many possible u a wider picture of the diffusion ODF would emerge in which peaks in magnitude would correspond to underlying fibre orientation(s).

There are, however, drawbacks to DSI; principal amongst which is that in order to properly estimate $\overline{P}$, q-space must be densely sampled. In practice, owing to the nature of most 3D Fourier transform implementations, this takes the form of sampling q-space across the vertices of a super-imposed grid structure which, depending on the particular application, might require hundreds if not thousands of individual pulse gradient experiments in order to complete. Given the time required to acquire these samples and the extremely high diffusion weightings often required to reach the outer q-space shells (resulting in high signal attenuation and thus reduced SNR), clinical DSI is generally seen as infeasible from both a scan-time and hardware requirements perspective.

This is not to say that the q-space formalism itself is not useful. As an example, Q-ball imaging (QBI), introduced by Tuch (2004), is one method designed to address the time deficiencies of DSI (other approaches attempt, as an example, compressed sensing Lee and Singh (2010) to reduce the number of sampling points required while still maintaining full q-space coverage). Rather than extensively sampling q-space, Tuch was able to show that computing the Funk Radon Transform (FRT) of data acquired on a single spherical q-space shell (sampling with fixed gradient magnitude, duration and mixing time but varying gradient orientation, analogous to a DT-MRI acquisition with a large number of sampling points) could retrieve an ODF equivalent to DTI, though with a blurring matching that of a 0th order Bessel function with width inversely proportional to that of the q-space shell (i.e. a higher diffusion weighting results in a sharper ODF which results in higher angular resolution).

In purest form, if $f(w)$ is a function on the sphere then the FRT of $f(w)$ is defined as the sum over the corresponding equator, i.e. the set of points perpendicular to the vector of interest, such that the FRT of $f(w)$ with respect to the vector u may be written as:

$$FRT[f(w)](u) = \int f(w)\delta(w^T u) \ dw \tag{2.39}$$

where, in this instance, δ is a Dirac delta function. In practice, the function of interest is provided through a q-space sampling in a 3D Cartesian space, depicted as $f(x)$ where x is a 3D vector and this will require mapping to a spherical form. This is achieved by modifying the FRT to incorporate a radius r', such that the FRT_m (modified) at a particular r' may be written as:

$$FRT_m[f(x)](u, r') = \int f(x)\delta(x^T u)\delta(|x| - r') \ dx \tag{2.40}$$

which, by extension, allows the ODF to be estimated as:

$$ODF(u) = \frac{1}{Z}FRT_m[S(q, \Delta)](u, q') \tag{2.41}$$

where q' is the radius of the q-space sampling shell and Z is a normalising constant intended to compensate for the non-consideration of solid angle factors during radial integration; something that may be more easily appreciated by altering Tuch's ODF formulation to format similar to Wedeens (Eq. 10):

$$ODF(u) = \int p(xu) \ du \tag{2.42}$$

where we may note that the x^2 term present in Wedeen's formulation is eliminated in Tuch's, altering the effects of the q-space radius and, as such, creating the need for the normalisation. Numerous authors have since been picked up on this discrepancy, creating constant solid angle QBI variants to rectify the issue (e.g. **?**).

Finally, it is worth noting that in practice that, due to hardware limitations, the narrow gradient assumption cannot. As a result, any PDF or ODF derived from these or similar q-space methods are, at best, a good approximation of the relationships described here. Returning to Equation 2.21, as an example, the diffusion weighting b includes the $(\Delta - \delta/3)$ term as one possible method for compensating for more practical wide gradients.

2.2.9 Spherical Deconvolution

The previous DSI and QBI methods can both be described as model free. This is to say that they retrieve the ODF (or other diffusion related statistics) without making assumptions about the nature of the diffusivity that created the measured signals. Spherical deconvolution (SD) methods depart from this model free approach and instead make an explicit assumption that the observed DW-signal results from some combination of a known response function, $R(\theta)$, such that:

$$S(\theta, \phi) = \sum_i f_i A_i R(\theta) \tag{2.43}$$

where $S(\theta, \phi)$ is the observed signal, f_i is the fraction of that signal contributed by the i'th fibre and A_i is a rotation operation altering the response function $R(\theta)$ such that its axis lies along (θ_i, ϕ_i) – the i'th fibre orientation. More succinctly, this may be rewritten as:

$$S(\theta, \phi) = fODF(\theta, \phi) \otimes R(\theta) \tag{2.44}$$

where the fibre ODF (fODF) gives the fraction of fibres within the sample that are aligned along the (θ, ϕ) direction. This Equation (2.43) then serves to illuminate the core assumption of spherical deconvolution techniques – that the observed signal is the result of the spherical convolution of an underlying fODF with a known axially symmetric fibre response function (hence its expression as a function of only θ) thus, given that the response function and signal are known or can be measured, the fODF (parameter of interest) may then be retrieved by deconvolving the response from the signal. Note that additionally, spherical deconvolution does not attempt to imply the ODF from the diffusion PDF like the previous q-space methods (a relationship that has not yet been conclusively proven (Tournier et al. (2004))), but rather attempts to estimate the fibre ODF directly. While this has advantages, there is also the potential for serious issues should the estimated response function $R(\theta)$ not match reality.

While there are numerous implementations of spherical deconvolution, I will focus on two – Tournier's Constrained Spherical Harmonic Deconvolution (CSHD, Tournier et al. (2007)) and Dell'Acqua's damped Richardson-Lucy Deconvolution (dRL, Dell'Acqua et al. (2010)) for the reason that these are the methods most commonly used within this work and thus are most relevant to the thesis.

2.2.10 Constrained Spherical Harmonic Deconvolution

CSHD uses a combination of spherical and rotational harmonics to describe the components of Equation 2.44 and complete the deconvolution process. In purest form, spherical harmonics describe an orthonormal basis set in spherical space such that any spherical function $F : S_2 \rightarrow C$ can be rewritten as a linear combination of spherical harmonics:

$$F(\theta, \phi) = \sum_{l=0}^{\infty} C_{l,m} Y_{l,m} \tag{2.45}$$

where $Y_{l,m}$ represents a spherical harmonic of the l'th order ($l \geq 0$) and m'th phase factor ($-l \leq m \leq l$) with complex coefficient:

$$C_{l,m} = \int\limits_{0}^{2\pi} \int\limits_{0}^{\pi} F(\theta, \phi) Y_{l,m}^{*}(\theta, \phi) \qquad (2.46)$$

By bringing in rotational harmonics (a complete orthonormal basis set in rotational space), Healy et al. (1998) showed that a spherical harmonic convolution of order l may be written as the matrix operation:

$$S_l = Rh_l O_l \qquad (2.47)$$

where S_l and O_l are $2l+1$ element vectors describing a spherical harmonic decomposition of the signal/fODF and Rh_l is a $2l+1 \times 2l+1$ matrix describing a rotational harmonic decomposition of the convolution kernel. In this very simple formulation, deconvolution may then be achieved by inverting Rh_l.

Attempting to model the diffusion process and cope with noise in the measured signal, Tournier's implementation (**?**) differs from this template. The DW-MRI signal (and associated fODF) is both an axially symmetric and real valued function on the sphere – this has several benefits, allowing simplification of the process. Again beginning with the spherical harmonics, axial symmetry allows one to omit odd order harmonics in any decomposition since their axial asymmetry means that they do not provide sufficient improvements in description relative to the cost of fitting their parameters (extra DW-MRI signals would need to be acquired, wasting time). When constrained to real values, spherical harmonics then exhibit conjugate symmetry, halving the number of parameters to be estimated. The axial symmetry also has an effect on the rotational harmonics. Tournier discovered that if representing the response function as Z axis aligned, only the $n = m = 0$ elements possess non-zero magnitude and as such, the $l'th$ order rotational harmonics (Rh_l in Eq. 2.47) may be represented by a single real scalar constant, rh_l, whose value may be computed by scalar division of the lth order, $m = 0$ SH coefficient of an ideal single fibre response function (Z axis aligned) by the corresponding coefficient of a spherical harmonic fit to the Z axis delta function. Such that in the case of diffusion:

$$S_l = Rh_l O_l = rh_l O_l \qquad (2.48)$$

However, deconvolution rarely uses only a single order of harmonics, a point that will be addressed shortly.

Moving onwards, the deconvolution itself is an ill-posed problem, recovery through simple inversion is very susceptible to noise in the DW-MRI signal and thus not a desirable method. One of the common effects of such noise, as shown through the original (non-constrained) spherical harmonic deconvolution method (Tournier et al. (2004)) is large negative lobes appearing within the fODF, something that is obviously impossible in reality (no negative diffusion). To address this problem a constrained non-negative spherical harmonic deconvolution was proposed, formulated around a modified Tikhonov regularisation driving small/negative fODF components to 0, leaving only large positive lobes (i.e. the fibres), thus constrains the fODF to biologically and physically plausible options. The iterative formulation Tournier adopts can be written:

$$f_{i+1} = \arg\min\{||Af_i - S||^2 + \lambda||Lf_i||^2\} \tag{2.49}$$

Breaking this down into two pieces: The first component $||Af_i - b||^2$ provides the sum of squared errors between the DW-MRI signal, S, and the fODF approximation which is produced by multiplying the SH representation of the fODF (f_i) by a matrix A which maps f_i into intensities along the directions in which S was measured.

The second component, $||Lf_i||^2$ penalises the fODF approximation for having non-zero intensities along directions that are deemed to be unimportant (i.e. should be zero). Since these directions cannot be known a priori they are estimated from the previous fODF approximation in the following manner:

Given an fODF approximation f, its intensities u along the set of directions N can be calculated by multiplication with a mapping matrix P, such that:

$$u = Pf \tag{2.50}$$

We may then choose some threshold τ, often $\approx 10\%$ of mean(u), allowing construction of a matrix L such that:

$$L_{m,n} = \begin{cases} P_{m,n} & u_m < \tau \\ 0 & u_m \geq \tau \end{cases} \tag{2.51}$$

where P is defined as:

$$P_{i,j} = \begin{cases} Real\{Y_{l,m}(\theta,\phi)\} & m \geq 0 \\ Imag\{T_{l,m}(\theta,\phi)\} & m < 0 \end{cases} \qquad \begin{aligned} m &\in [-l, l] \\ j &= \frac{1}{2}(l+1) + m \end{aligned}$$

i.e. a set of SH modified to account for the real valued nature of diffusion.

The effect of multiplying f_i by L is such that along directions where u was greater than the threshold (i.e. "important") the intensities produced Lf_i will be 0, hence not punished, whereas directions along which u was less than τ (i.e. should be zero) will retain their full intensity and hence be punished appropriately.

Returning to a point made earlier, it is not common to perform (de)convolution using a single order of spherical harmonics (as it is in Eq. 2.47 and 2.48). The fODF is more commonly specified in terms of the set of even order harmonics (including the 0th) up to some cut-off point, L_max, with the addition of each increasing angular frequency and thus angular resolution of the resultant fODF. Thus for a given L_max, the number of parameters required to describe an fODF is equal to $(L_{max}+1)(L_{max}+2)/2$ accounting a weight on each of the $2l + 1$ phase factors for any given order and omitting the odd ordered harmonics. As an example, this means that a common L_{max} of 8 requires weights on 45 separate harmonics to describe the fODF, and thus a minimum of 45 unique DW-MRI signal directions to calculate a deconvolution. If we briefly examine the A term in Equation 2.48, which provides a mapping between the SH fODF decomposition and the resultant intensities along the set of directions of the signal S, this matrix is constructed through the multiplication:

$$A = QR \tag{2.53}$$

where Q maps the spherical harmonic coefficients to corresponding DW signal directions (constructed in an identical manner to P in Eq. 2.52) and R performs the spherical convolution with the kernel $Rh(\theta)$. Remembering Eq. 2.48, this kernel is described in the form of rotational harmonics, though due to the axially symmetric nature of the diffusion response function, only $m = n = 0$ harmonics possess non-zero magnitude. Thus the R matrix is constructed as:

$$R_{i,j} = rh_l \tag{2.54}$$

where:

$$j = \frac{1}{2}2l(l+1) + m \tag{2.55}$$

creating a diagonal matrix with one rotational harmonic value per spherical harmonic order. This, along with the construction of P/Q to consider all harmonics up to and including L_{max} allows spherical (de)convolution across the entire range of harmonics to be conducted in a single operation. With the formulation set out, fODF optimisation is allowed to iterate from an initial estimation produced using a (filtered) least-squares

method until the L matrix, encoding the set of directions which should be zero, converges – indicating that an reasonable solution has been found. Fibre orientations may then be retrieved by differentiating the SH representation fODF of (or through a variety of other methods) to find local maxima.

2.2.11 Damped Richardson-Lucy Deconvolution

The damped Richardson-Lucy technique takes a different approach based on the assumption that the signal S, along direction u, produced by a single fibre with diffusion-tensor derived axial and radial diffusivities of λ and β can be adequately estimated as the function:

$$S(u) = e^{-b(\lambda \cos^2(\theta(r,u)) + \beta \sin^2(\theta(r,u)))} \tag{2.56}$$

where b is the b-value and $\theta(r, u)$ is the angle between the fibre orientation, r, and direction of measurement u. Extending this, it can be assumed that the DW-MRI signal produced by an arbitrary number of fibres can be estimated as:

$$S(u) = \sum_{i=1}^{p} f_i e^{-b(\lambda_i \cos^2(\theta(r,u)) + \beta_i \sin^2(\theta(r,u)))} \sum_{j=1}^{q} f_j e^{-b\beta_j} \tag{2.57}$$

where f_i provides a weight on each fibre component and the second term

$$\sum_{j=1}^{q} f_j e^{-b\beta_j} \tag{2.58}$$

described a series of q isotropic components, with corresponding diffusiviy β_j, providing a constant signal regardless of measurement orientation. This model can be rewritten:

$$S(u) = \sum_{i=1}^{p} f_i e^{-b\alpha_i \cos^2(\theta(r,u))} + \sum_{j=1}^{q} c_j \tag{2.59}$$

where $\alpha_i = \lambda_i - \beta_i$ and $c_j = f_j e^{-b\beta_j}$, a scalar term that is constant along all directions (the isotropic component).This shows that, assuming all fibres contributing to the signal have the same diffusion profile, the single fibre response function in Equation 2.43 oriented along direction r, can be approximated as:

$$R(u, r) = e^{-b\alpha cos^2(\theta(r,u))} \tag{2.60}$$

where α defines the "shape" of the common diffusion profile.

In order to achieve deconvolution, Dell'Acqua turns to the Richardson-Lucy (RL) method (Dell'Acqua et al. (2007)), a well known iterative procedure for a latent image that has been blurred by a known point spread function. In the context of deconvolution, the observed signal is the blurred image, the fODF is the latent image and the response function matches the point spread function which convolves with the fODF to "blur" it into the observed signal. A non-damped version of the RL algorithm may be written as follows:

$$f_{i+1} = [f_i] \frac{[H^T S]}{[H^T H f_i]} \tag{2.61}$$

where i refers to the $i'th$ iteration step, f is a column vector of length N containing fODF values along N uniformly distributed directions, S is the signal of length M and H is an $M \times N$ matrix mapping a response function oriented along the N possible orientations to the signal intensities that it would produce along the M DW-MRI imaging directions. The square bracket operators depict element wise operations. While not strictly following the original Richardson-Lucy deconvolution for Poisson noise, this adaptation is known to converge to a non-negative linear least-squares solution in the presence of Gaussian noise which was deemed a more desirable property in spherical deconvolution of DW-MRI signal data.

The issue with this formulation however, is that it does not compensate for isotropic diffusion components in the DW-MRI signal (for this matter, neither does CSHD). The result is that, where isotropic partial volume is present, fODFs are known to suffer noise susceptibility (since a lower portion of the DW-MRI signal is devoted to signal produced by the isotropic compartments) and ringing due to a partial de-activation of the non-negative constraints caused by a non-zero background intensity (isotropic components could be modelled as a directionally invariant constant fODF magnitude).

An example of the effects of isotropic partial volume can be observed in Figure 2.10. Note that without the isotropic component (black plot) the fODF exhibits the desired 0 magnitude along non-relevant orientations, while inclusion of the isotropic component (black dotted line, blue plot) allows the fODF to oscillate freely, causing a spurious fODF peak, due to partial deactivation of the non-negative constrained caused by a non-zero background.

Dell'Acqua's solution was to introduce a damping component to the Richardson-Lucy deconvolution, taking note of the dynamic range of the recovered fODF to suppress artefacts caused by isotropic partial volume. This new dRL algorithm may be written as follows:
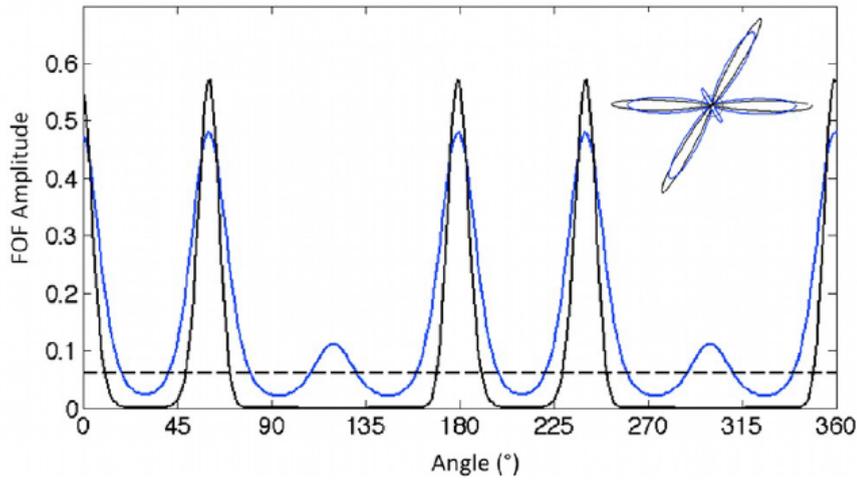
FIGURE 2.10: Example given by Dell'Acqua et al. (2010), Figure 1, demonstrating changes in an fODF resulting from the introduction of an isotropic component.

$$f_{i+1} = [f_i](1 - [u_i]\frac{[H^T S - H^T H f_i]}{[H^T H f_i]})$$ (2.62)

where u is an N element damping vector:

$$u_i = 1 - \mu r_i$$ (2.63)

with:

$$r = 1 - \frac{[f_i^v]}{[f_i^v + \eta^v]}$$ (2.64)

and:

$$\mu = \max(0, 1 - 4\sigma(S))$$ (2.65)

where $\sigma(S)$ is the standard deviation of S. Beginning with the construction of the damping vector u_i. The μ component is an operation designed to modulate the damping between voxels and obtain a classification of tissue based on signal anisotropy, tied to the standard deviation (larger standard deviations implying larger anisotropies). According to Dell'Acqua, the resultant value tends to vary between 0.2 for highly anisotropic single-fibre signals and 1 for totally isotropic values, the 0 simply provides a bare minimum should highly isotropic signals drive the $1 - 4\sigma(S)$ value negative. The desirability of this feature should be fairly obvious, where signals are highly anisotropic they are less likely to contain an isotropic component and thus more likely to be adequately described

by the original RL formulation, damping in these circumstances may undesirably serve to eliminate small though genuine fODF peaks and thus its importance should be down weighted in such cases.

Proceeding to the r vector, this selectively damps individual fODF components depending on their magnitude, where the fODF values are low, the threshold term η dominates and the result of the division tends to 0. This causes the corresponding r vector component to approach 1, sending corresponding elements in the u vector towards $1 - \mu$, resulting in a low valued fODF components damping out in the next $i + 1'th$ iteration at a rate governed by the isotropy of the given individual signal (μ). The v element in r is simply a geometric thresholding term, governing the rate at which r transitions from 0 to 1 (turning the damping on/off), the smaller the v the broader the transition.

Giving an overview of the algorithm; by subtracting the projection of the predicted signal (fODF convolved with the circulant H matrix) from the projection of the actual signal, then normalising by the simulated projection, small differences between the simulated and actual signals will cause corresponding elements in the division to tend to 0, resulting in a 0 change update, while larger differences will, modulated by u, receive larger updates in the appropriate direction (negative or positive). A side effect of this formulation is that the rate of convergence of the fODF is linked to the isotropy of the signal in question (the value of μ is governed by $\sigma(S)$) which, given that this algorithm is semi-convergent – i.e. will converge to a good solution and then, if allowed to continue iterating, diverge onto poorer solutions (**?**) - raising questions about the efficacy of selecting a fixed number of iterations across the entire brain. Empirically speaking however, this does not seem to be a significant issue – slow converging high anisotropy single fibre signals easily to produce a (directionally) appropriate fODF within the limited number of iterations afforded by lower anisotropy single fibre/crossing fibre (see chapter 4); though I have not explicitly checked the effects on fODF peak magnitudes (a property used in Hindrance Modulated Orientational Anisotropy, HMOA, Dell'Acqua 2012), perhaps this may be worthy of future investigation.

Finally, once a fODF is recovered, spherical harmonics are often fit to the N discrete samples and peak directions/magnitudes recovered in a similar manner as CSHD.

2.2.12   Spherical Deconvolution: Closing Comments

A factor common to a number (though not all, e.g. PAS-MRI which can be treated as a special case of spherical deconvolution, (Seunarine and Alexander (2006))) spherical deconvolution methods is the assumption that one single fibre response function may adequately reflect actual fibre responses throughout an entire volume. This is a large assumption to make, and so while it may hold true for a significant portion of the fibrous voxels within an image, there are consequences (unique to each deconvolution

method and the degree of miss-estimation) for inappropriate response function selection. Chapter 4 covers these consequences and their likely causes, in detail, for the above two spherical deconvolution algorithms.

2.2.13 Overview of Other Novel HARDI Methods

While discussing HARDI techniques in general it is perhaps easiest to divide them into model free and model dependant reconstructions – DSI being an example of the former, CSHD of the latter.

One of the earliest examples of a model-free method was apparent diffusion coefficient (ADC) modelling (Frank (2002), Alexander et al. (2002)), the basic idea was that if one fit even order spherical harmonics to the apparent diffusion coefficient profile ($S = S_0 e^{-bADC}$), non-Gaussian diffusion could be detected by examining the relative power assigned to each order – higher order harmonics have higher angular frequencies and are thus more likely to be useful in describing complex ADC profiles arising from crossing fibres. While effective with respect to its intended purpose, it was noted by Tuch (2002) amongst others that maxima in the ADC do not necessarily correspond with fibre orientation and thus results cannot be used to infer orientations.

Later model-free approaches were more successful in recovering fibre orientation, while I have already covered DSI and QBI, PAS-MRI (Jansons and Alexander (2003)) presents another interesting example. PAS-MRI operates on the assumption that angular and radial structure in diffusion MRI can be considered as independent, allowing acquisitions to focus on a single q-space shell – largely ignoring radial structure while (hopefully) retaining sufficient angular information to reconstruct a PDF of the form:

$$p(x) = \frac{\tilde{p}(\hat{x})}{r^2}\delta(|x| - r) \tag{2.66}$$

where δ is a Dirac delta function, r is a constant, $\hat{x}$ is a unit vector along direction x and $\tilde{p}(\hat{x})$ represents a persistent angular structure function, i.e. the object of interest. While the maximum entropy formulation used to recover $\tilde{p}(\hat{x})$ produces accurate fibre orientation estimations, the proposed non-linear iterative algorithm is extremely computationally challenging (often requiring multi-hour reconstruction times) which has hampered the uptake of this method (though more recent works have attempted to correct this shortcoming, e.g. Seunarine and Alexander (2006)).

Continuing to model dependant reconstructions, some of the earliest HARDI methods were simply extensions of the previous DT-MRI technique to describe a DW-MRI signal as a mixture of Gaussians. Tuch (2002), for example, proposed that a signal may be modelled as:

$$S(q_k) = \sum_i f_i \exp[-q_k D_i q_k \tau] \tag{2.67}$$

i.e. a weighted (f_i) combination of compartments corresponding to individual Diffusion Tensors (D_i). The problem with such techniques is that the fitting of individual diffusion tensors is an inverse problem and since there is no a priori knowledge of the number of tensors to fit, leading to instability if one chooses incorrectly (Alexander et al. (2002)) – something that is easy to do when one considers how frequently the number of components per voxel changes over the brain.

Finishing with an example of a more complex technique, CHARMED (Assaf and Basser (2005)) combines hindered and restricted models of diffusion into a single integrated system such that:

$$S(q, \Delta) = f_h S_h(q, \Delta) + \sum_{i=1}^{n} f_i S_r^i(q, \Delta) \tag{2.68}$$

where $f_h S_h(q, \Delta)$ represents a hindered compartment contributing signal fraction f_h (later modelled as a single diffusion tensor) and $f_i S_r^i(q, \Delta)$ represents one of n restricted compartments modelled as Neuman restricted cylinders (Neuman (1974)). In a departure from the previously discussed methods, CHARMED acquisitions are multi-shelled, collecting spherically distributed measurements over a range of b-values (from 0 to 8571 s/mm^2 in Assaf and Basser (2005)), with the highest b-value shells eliminating signals from all but perpendicular diffusion in restricted compartments (remembering $S = S_0 e^{-bD}$, if b is very large, D must be very low, i.e. transverse diffusion in a restricted compartment, in order for a measurable signal to survive the attenuation), allowing separation of diffusion resulting from the hindered and restricted compartments, as well as between individual restricted compartments.

The problem with the CHARMED model is that, along with the longer acquisition time required to measure multiple shells, fitting suffers the combined drawbacks of both the multi-Gaussian and PAS-MRI models. Specifically, you must both choose the appropriate number of restricted compartments a-priori (same drawback as the multi-Gaussian model) and, in fitting the CHARMED model, use computationally expensive non-linear methods (same drawback as PAS-MRI). While there are methods to address the first drawback (a priori selection), they are themselves computationally expensive and beyond the scope of this work.

2.2.14 Tractography

Until this point I have discussed methods leading to voxel wise estimates of fibre orientation. While useful, it is often preferable to build upon these discrete estimates in order to reconstruct the large-scale fibre architecture present within an imaged volume. One class of techniques for doing this, and the one that we will focus on, is tractography. As before, it is probably easiest to explain tractography by examining a simple implementation – Fibre Assignment by Continuous Tracking (FACT, Mori et al. (1999)) being one of the earliest examples. The FACT algorithm proceeds as follows:

1. Place a seed at the centre of the target voxel

2. Generate a streamline from the seed, proceed along the principal diffusion vector of the host voxel.

3. When the streamline intersects a voxel boundary:
   3.1. IF $R \geq Threshold$
   3.1.1. Continue along the trajectory specified by the principal diffusion vector of the intersected voxel.
   3.2. ELSE
   3.2.1. Terminate streamline.
   3.2.2. GOTO 5.

4. GOTO 3.

5. IF streamline length $\geq$ length threshold
   5.1 Retain streamline.

6. ELSE
   6.1 Reject Streamline.

repeated for each voxel in the image or targeted sub-region. As this algorithm shows, tractography is the process of building a set of representative streamlines whose path through the image is informed by orientations specified by the diffusion weighted data in such a way that, when viewed as an ensemble, they create a visual representation of the underlying fibre. In the case of FACT, which is based on the DT-MRI, the principal diffusion vector, v_c for the current voxel, v_n for the incident voxel, is simply the first eigenvector of the diffusion tensor.

A key factor in any tractography algorithm is the choice of termination criteria, for which FACT uses the value R:

$$R = \sum_i^s \sum_j^s ||\frac{V_{\lambda_i} \cdot V_{\lambda_j}}{s(s-1)}|| \tag{2.69}$$

which provides a measure of coherence between the principal eigenvectors (V_{lambda_i}) of the s nearest neighbours to the incident voxel. Where R is low, the voxel and its surroundings are judged to be incoherent, suggesting an absence of organised fibre and thus an undesirable region through which to propagate streamlines.

While a reasonable start, FACT has numerous deficiencies, one obvious choice being that due to the method through which streamlines propagate, the difference between integration steps is by definition unequal – depending case by case on the point of intersection with a voxel and the length one must travel along that voxels orientation to reach the next boundary. Combined with the assumption that orientation remains constant throughout the voxel, the resulting streamlines can be rather jagged (especially at low resolutions, resulting in larger voxels) with the uneven curvature leading to misinterpretation of the underlying geometry.

Fortunately, this particular concern was subsequently addressed by Basser et al. (2000) through the introduction of interpolation to tractography. By fixing the step size and interpolating the tensor field at each integration point (obtaining a unique estimate), the resulting streamlines have the potential to be far smoother (or at least more uniform) than those produced by FACT, generally improving the quality of the resultant reconstruction.

Furthermore, FACT (as with Basser's method amongst others, e.g. Lazar et al. (2003)) is by definition deterministic, it makes no attempt to account for any uncertainty in the orientational estimates upon which the results are based. Briefly, anything from image mis-registration or signal corruption to normal signal noise may cause bias in the orientational estimates that can produce marked effects in the resultant tractography; common artefacts may include tract "jump", whereby a streamline artefactually passes from one fibre tract to another similarly oriented fibre tract in close proximity due to error induced perturbation, or premature termination where bias results in artefactual triggering of some termination criterion, most commonly an angular threshold preventing the difference in orientation between consecutive steps from exceeding some prescribed value (a common replacement for the R value employed by FACT). If the tractography method does not acknowledge the possibility of such artefacts, or at least indicate where they may be present, it is difficult to trust the accuracy of any result – especially longer streamlines since errors are known to accumulate as one moves further from the seed (Basser et al. (2000)).

One way to deal with this uncertainty is to use a probabilistic tractography method (of which there are numerous examples, e.g. Behrens et al. (2002) or Parker et al. (2003)) which, in one way or another, attempt to construct and then sample from probability distributions covering the range of likely fibre orientations predicted by a given set of noisy signals – allowing reconstruction of (hundreds or thousands of) potential streamline paths which may then be aggregated to provide the most likely result(s) and/or

provide confidence intervals at each streamline step. Looking at an early example, the Probabilistic Index of Connectivity (PICo) method by Parker et al. (2003) attempts to link the uncertainty in an estimate to the shape of the derived diffusion tensor, while this is not directly measuring uncertainty (and thus may not observe the expected link between uncertainty and signal SNR), it is not an entirely unreasonable approximation given that lower anisotropy signals are known to be more susceptible to noise (e.g. via sorting bias). With a shape derived uncertainty calculated at each voxel, PICo then outlines a Monte Carlo framework, sampling from the inferred cones of uncertainty in order to produce an array of streamlines which are then aggregated in the form of voxel intersection maps (Fig 2.11, adapted from Parker et al. (2003)). When appropriately thresholded, these intersection maps can then be used to provide an indication of likely connectivity for streamlines seeded at a specific point/region. Though as hinted before, the concatenation of errors as a streamline propagates means that longer or curving paths, even if perfectly visible through dissection, are more difficult to reconstruct – leaving the most likely connectivity indicated by the amalgamated result unduly biased towards shorter path lengths, even if this may not be true in reality (an observation that is true for a large number of similar methods).
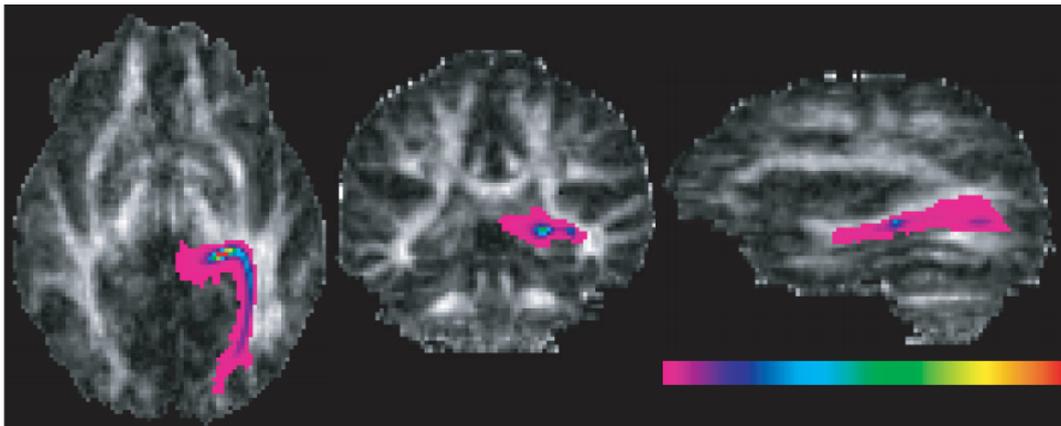
FIGURE 2.11: Example aggregated voxel intersection map for probabalistic tractography through the Optic Radiation.

Moving forwards; Jones (2003) shows that bootstrapping can be a powerful tool for provision of more directly data-driven uncertainty estimates. The process initially involved capturing repeat measurements of a given signal and then recombining individual measurements in different combinations to give particular instances of the signal in question (the traditional bootstrap). By repeating this numerous times and deriving orientational estimates from the resultant signals, Jones showed that useful properties such as the cone of uncertainty (Fig. 2.12) could be estimated. The issue of course is that recording the signal multiple times correspondingly increased the acquisition times. This was in Jones (2006) where Jones proposed the "Wild" bootstrap approach, essentially manipulating the residuals between a single recorded signal and its DT-MRI fit (introducing a level of model dependence) to produce new signals and thus orientational estimates from within

the estimated noise distribution. By combining this wild bootstrap with a Monte-Carlo framework (similar to that proposed by Parker et al. (2003)), Jones was then able to perform probabilistic tractography using data derived uncertainty estimates without increasing acquisition costs, but still closely matching traditionally bootstrapped results (Fig. 2.13).
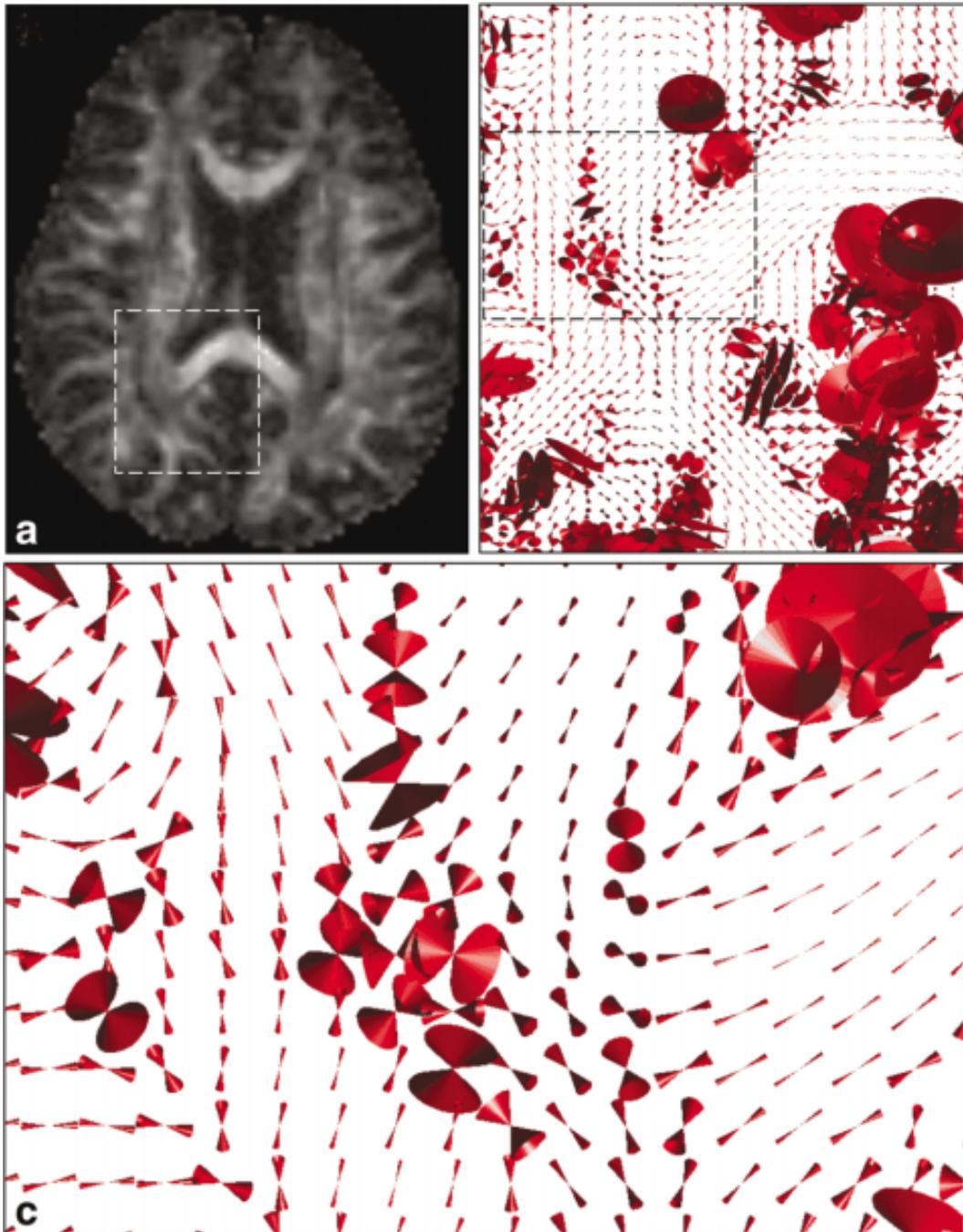
FIGURE 2.12: Visualisation of bootstrap-derived cones of uncertainty (showing 95% confidence angle) at the level of the splenium in the Corpus Callosum. Reproduction of Figure 2 Jones (2003).
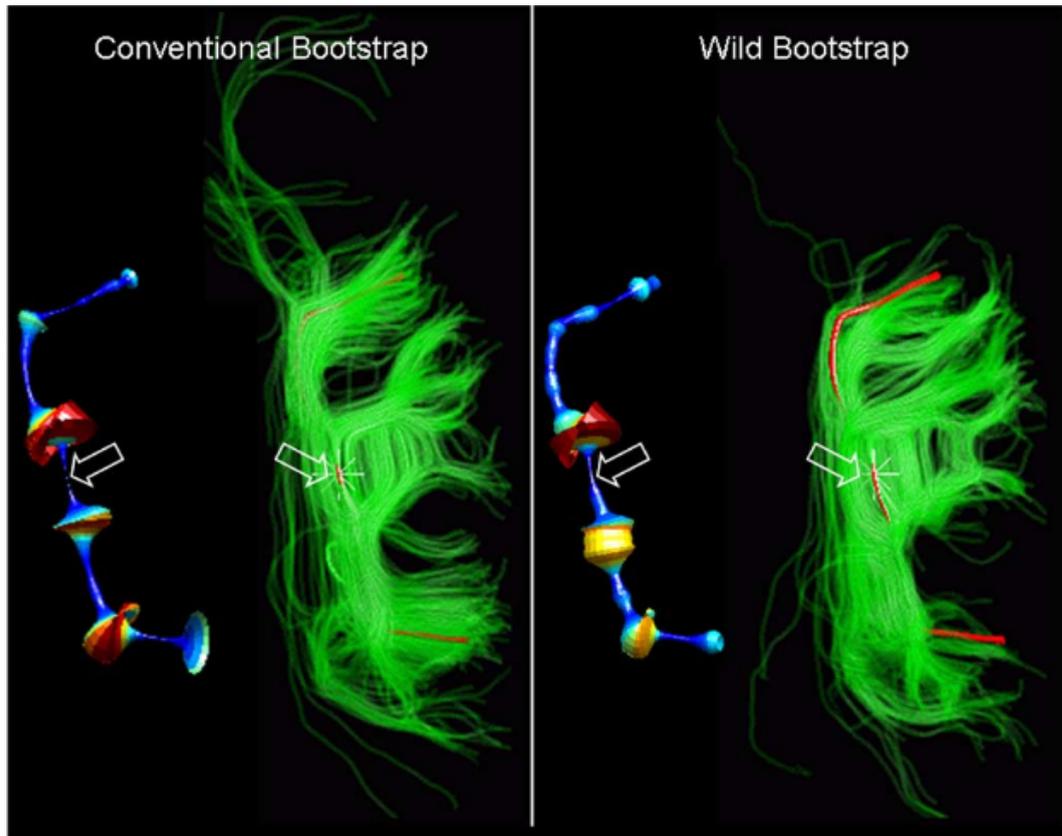
FIGURE 2.13: Visualisation of conventional vs. "wild" boostrap results. Reproduction of Figure 2 Jones (2006).

An issue in common with all tractography algorithms discussed until this point is that they take their orientational information from the DT-MRI model, one that we know from previous sections is deficient in the area of non-Gaussian diffusion. Fortunately, for the most part (the effects on probabilistic tractography will be discussed shortly) adaptation of DT-MRI tracking to HARDI data is fairly trivial. An example of a deterministic HARDI tractography algorithm (matching that used by ExploreDTI toolbox, Leemans et al. (2009)); is as follows:

1. Calculate the fODF at the centre of each voxel.

2. Place seed points at the vertices of a regular grid superimposed over the fODF image.

3. FOR each seed:
   3.1. Interpolate an fODF estimate for the current position.
   3.2. Use peak finding methods to determine the likely fibre orientations originating from the current position.
   3.3. FOR each peak of significant magnitude:

   3.3.1. Proceed [Step size] along the indicated orientation
   3.3.2. Interpolate a new fODF at this location and find the minimally subtending peak.
   3.3.3. IF this peak subtends at a sub-threshold angle and has super-threshold magnitude:

   3.3.3.1. GOTO 3.3.1
   3.3.4. ELSE
   3.3.4.1. Terminate the streamline.

4. FINALLY: Reject all sub-length-threshold streamlines.

In comparison with the FACT algorithm the only differences are that Basser-like interpolation has been included, the principal diffusion vector has been replaced by the minimally subtending fODF peak and combined peak magnitude and angular deviation check has replaced the originally prescribed R value for determining the point at which streamline propagation should cease – the assumption being that low magnitude peaks are more likely to represent unimportant/artefactual fODF lobes and that streamlines should follow a smooth path through the associated white matter.

Adaptation of previous probabilistic techniques can also, depending on the precise implementation, be equally trivial. Jeurissen et al. (2011), for example, showed that the bootstrapping methods used by Jones may be adapted to spherical harmonic deconvolutions, this time performing a bootstrap using residuals to the spherical harmonic fit to the DW-MRI signal; providing a better representation of uncertainty in crossing-fibre region - the DT-MRI model, along with being unable to resolve multiple fibre orientation, would also find uncharacteristically large residuals caused by fitting a Gaussian to inherently non-Gaussian diffusion.

As one possible alternative to bootstrapping, Behrens et al. (2007) presents a novel method that builds upon earlier work (Behrens et al. (2002)) in which DW-MRI signal is assumed to be described by a mixture of Gaussian components (similarly to multi-Tensor model) which are then fit using Bayesian estimation; updated (in the 2007 work) to

include shrinkage priors that ensure that multiple components are fit only when necessary – solving the previously discussed multi-tensor instability problem. Tractography is then performed using (a variation on) the familiar Monte Carlo framework, where instead of picking an orientation from the distribution determined through a bootstrap, one is sampled from the posterior distribution provided by the Bayesian fit - again resulting in one of a number possible streamline variations, which when aggregated may provide information on the most likely paths from a point.

As a point on which to end, one of the largest challenges still affecting tractography algorithms of all types is signal ambiguities. This is to say that with respect to the observed DW-MRI signal, crossing, kissing and fanning structures produce very similar measurements, though the implications of each on the resultant tractography and subsequent interpretation are profound. Disambiguation of these different configurations is by no means a solved problem, though some headway has been made. As an example, imagine two fibres bundle intersecting in an ambiguous crossing/kissing configuration. From a HARDI fODF point of view, there is very little information to inform the correct outwards trajectory for an incoming streamline – it will simply follow the minimally subtending outbound fODF lobe with a nominally 50/50 chance of selecting the appropriate solution. It is at this point that concepts such as Tractometry (Bells et al. (2011)) may provide a promising avenue for investigation. By incorporating additional information into the streamline representation, i.e. axon diameter – a parameter that may be retrieved through CHARMED, even from partial volumed environments – additional constraints could be woven into the tractography algorithm (Sherbondy et al. (2010)) to ensure a continuity of diffusive properties and not just trajectories. In cases, and admittedly these are not all cases, where the diffusive properties of the ambiguous tissues are dissimilar, the additional information could substantially improve the probability of selecting the correct fibre trajectory – substantially reducing convincing, though false positive results.

Fanning, however, is trickier still. The problem with fanning is that one of the core assumptions in a large number of diffusion processing algorithms is broken – the diffusion is not necessairly symmetric about an axis. This is not to say that the resultant signal is not symmetric – it is, and thus the diffusion appears to be – but if one imagines placing a voxel over the exact point at which a fibre begins to fan, the true trajectory on one side of that voxel would be continuous and uniform, but on the other side there would be an array of equally valid exit trajectories. This presents a one in, many out scenario which current methods cannot represent on a voxel-wise basis (forced axial symmetry). In the absence of new measurement techniques, the current avenue of pursuit is to examine fibre orientations on a multi-voxel scale. Recent work by Rowe et al. (2013), for example, is showing promising progress with particle filtering methods that disambiguate macro-voxel orientational divergence and convergence patterns, informing the choices made at

a local level while conducting streamline propagation. While seemingly not yet mature, it would appear that such methods will be an important object of study in future years.

2.2.15   Closing Comments

Over the course of this background chapter we have attempted to convey the key concepts and realisations that underpin diffusion weighted MRI. From the base mechanisms of NMR, we have described a path through echo manipulation to diffusion weighting to signal processing and finally tractography which bringing us to the last major point of commonality between the subsequent work. From this point onwards, each individual method chapter will build upon some aspect of the methods described within this chapter, incorporating addtional topic specific background where necessary, to convey a novel contribution to this discipline.

Chapter 3

Methods for Rejection of Corrupted Diffusion-Weighted Signal Elements

Prologue

This chapter discusses methods for ensuring the reliability of measures derived from corrupted diffusion weighted image data, providing the first step in a proposed robust signal processing pipeline.

A diffusion weighted signal may be described as corrupted if one or more of its elements deviates from its expected value by a larger magnitude than can be explained through thermal noise alone. Where this is the case, these artefactual deviations in signal intensity can introduce significant bias into the estimates of derived diffusion metrics and, by extension, have the potential to significantly alter the result of subsequent analysis.

One way to compensate for this corruption is to search the DW-MRI signals for these corrupted elements and then reject/eliminate them from any subsequent model fit (from which metrics will be derived), in effect ensuring that only the 'good' data are used in further analysis. While a limited of number techniques have been designed to achieve exactly this and (as will soon be demonstrated) function well against moderately corrupted data, they appear to lose robustness as the degree of corruption (particularly the number of corrupted elements) rises, resulting in erroneous increased rejection of good data (i.e. false positives) that may in itself introduce bias by reducing the number of points available for model fitting. For reasons covered in Chapter 5, this is of particular concern for maxillofacial imaging because currently achievable data quality in this region is particularly poor, thus it is paramount that we retain as many of the uncorrupted signal elements as possible.

After discussing the background and relevant works, we introduce our own contribution: a novel, scalable and automated algorithm for identification of fODF fitting outliers (corruption) with improved performance against highly corrupted signal data. The novelty of this technique stems from the amalgamation of a (modified) Richardson-Lucy deconvolution previously used for fODF estimation (Dell'Acqua et al. (2007)) with elements of a non-negative sparse coder (Hoyer (2002)) training phase, allowing us to selectively update the Richardson-Lucy dictionary to account for (and later detect) outliers in the fODF fitting process. By varying the number of voxels that are simultaneously processed within this framework (and thus contributing to the dictionary update), it is possible to efficiently detect common artefacts over an arbitrary number of voxels (i.e. from a single voxel to an entire slice, or any subregion within). The chapter concludes with an evaluation of the proposed and competing methods against a variety of simulated and real (scanner acquired) signal data, providing a vehicle for the discussion of the relative performance of each technique under a variety of situations.

Extracts from this work were presented at the 2013 ISMRM conference, Parker et al. (2013c).

3.1 Introduction

Diffusion weighted magnetic resonance imaging (DW-MRI) is a valuable tool for clinical and experimental neuroscience, providing the ability to both estimate fibre tract orientations and obtain related quantitative indices. As may be expected, however, the reliability of these derived metrics is strongly dependent on the quality of the recovered DW-MRI images (Jones and Basser (2004a), Walker et al. (2011)). As a simple example, the magnitude of typical thermal noise in complex MR imaging is known to be Rician distributed (Edelstein et al. (1984), Gudbjartsson and Patz (1995)); where diffusion weightings are high (or alternatively, SNR low) this can lead to an artificial overestimation of the DW-MRI signal as intensities hit the rectified noise floor (a minimum measurable intensity determined by background noise) which, by extension, causes an underestimation of the apparent diffusion coefficient (Dietrich et al. (2001)) leading to a further underestimation of fractional anisotropy – an important metric for many DW-MRI studies.

While numerous methods have been proposed to compensate for thermal noise (e.g. Basu et al. (2006), Dietrich et al. (2001)), this is not the sole source of DW-MRI signal corruption (Pusey (1986), Bihan et al. (2006)). There are, in fact, a host of other potential pitfalls that might affect image quality; some vary spatially – e.g. local susceptibility artefacts (Andersson et al. (2003), Jezzard and Balaban (1995)), temporally – e.g. bulk subject rotation/translation (Anderson and Gore (1994), Trouard et al. (1996)) or cardiac pulsation Pierpaoli et al. (2003) and even as the result of imaging

hardware instability (e.g. timing errors or RF inhomogeneity, Smith and Nayak (2010)). Again, there are numerous methods that have been proposed to deal with these pitfalls. Single-shot EPI (Edelman et al. (1994)), for example, provides an extremely fast image capture method which, at $\approx 100ms$ (Turner and Bihan (1990)), essentially "freezes" the subject in place for the duration of the pulse sequence, eliminating slower movements as a source of (intensity) artefacts. Likewise, cardiac gating (e.g. Lanzer et al. (1984), Lanzer et al. (1985)) can ensure that images are captured at low points within the cardiac cycle, eliminating periodic localised signal losses caused by the imaging sequence corresponding with an arterial pressure wave. There are, however, still problems. Sudden large motions (relative to the ≈ 10 micron sensitivity of the typical DW-MRI pulse sequence) can overcome even the SS-EPI acquisition, leading to irrecoverable signal corruption. This is particularly problematic when dealing with difficult subject groups – e.g. the very young, disabled or those otherwise prone to sudden involuntary movement – to the extent that it is not uncommon for individual volumes (corresponding to one or more unique DW-MRI gradient image) to suffer from irrecoverable signal drop out (coincidentally the most common outcome of all corruption mechanisms, Chang et al. (2012)). A useful (if slightly outdated) summary of bulk motion related artefacts is presented in a review article by Norris (2001).

A common practice for dealing with these kinds of artefact is to attempt to identify the corrupted or outlier signal elements first and then eliminate them from any subsequent model fitting – the idea being that regardless of the robustness of the model fitting technique, it is still more likely to converge to a plausible result if the corrupted data is not present to begin with. Methods designed to achieve this task may crudely be broken into two groups, those that work slice or volume wide (Jiang et al. (2009), Zhou et al. (2011)) and those that operate on individual voxels (e.g. Mangin et al. (2002), Chang et al. (2005), Chang et al. (2012), Pannek et al. (2012)).

Beginning with volume or slice wide methods, these techniques share the common assumption that artefacts along a given gradient direction affect all, or at least the majority, of the voxels within that slice/gradient volume. The power of this assumption is that where it holds true, e.g. scanner induced pattern artefacts, the statistical power afforded by examining multiple similarly affected signals allows for accurate and computationally simple (compared to voxel wise techniques) detection. The point at which such methods invariably fail is where corruption affects only a small subset of signals within the volume/slice, a classic example being cardiac pulsation artefacts that can lead to signal drop out but only severely affect a small area (Pannek et al. (2012)). Depending on the sensitivity of the particular technique such errors will either go undetected (with obvious consequences to derived metrics in affected areas) or cause a large number of non-corrupted signal elements to be incorrectly eliminated as the algorithm highlights corruptions affecting only a small subset of voxels belonging to the slice/volume.

At the opposite end of the spectrum, single voxel methods attempt to identify outliers on a per-voxel basis. This makes them more sensitive to localised corruption, but does so at the expense of both increased computational costs and the need to implement a more constrained model for what constitutes an "acceptable" diffusion profile (note that while slice/volume wide methods may also implement a specific diffusion model, e.g. Jiang et al. (2009), this is not a necessity, e.g. Zhou et al. (2011)). Of the available voxel wise methods, two principal variations exist. Beginning with the earliest, RESTORE (Chang et al. (2005)) and its later variant *informed* RESTORE (Chang et al. (2012)) operate under the diffusion tensor model (Basser et al. (1994a)). The algorithm itself can be broken down into two components, firstly the labelling of corrupted voxels and then an iterative process designed to detect and eliminate individual corrupted signal elements. Labelling is relatively simple, firstly an estimation of background noise (henceforth SD_{noise}) is made by either measuring the standard deviation of noise in a background region (using the method prescribed by Henkelman (1985)) or, in the later iRESTORE, by measuring residuals to the diffusion tensor fit, Walker et al. (2011)). Diffusion tensors are then fit (using constant-weighted non-linear least squares regression) to each voxel and those with one or more residual $\geq 3 \times SD_{noise}$ - i.e. with variations outside those explainable by random thermal noise - are labelled as corrupted. Each labelled voxel is then processed in turn by iterating the following until convergence:

1. For the current diffusion tensor fit, determine optimal signal weights using the Geman-McClure M-estimator (GMM, Geman and McClure (1987)).

2. Re-estimate the diffusion tensor using a weighted non-linear least squares regression with the GMM supplied weights.

3. Check whether the current tensor estimation is sufficiently similar to the previous iteration as to imply convergence, if not repeat from (1).

Once convergence has occurred, the identification of corrupted signals is then achieved by eliminating those signal elements whose residuals to the final iteration's diffusion tensor fit lie outside $3 \times SD_{noise}$. A final diffusion tensor is then approximated by a non-linear least squares fitting to the remaining signal data set. The accompanying flow diagram for this algorithm is given in Figure 3.1.

While demonstrably better than fitting tensors to signals including the corrupted elements, RESTORE is not without flaws. Situations could arise where, either due to underestimation of SD_{noise} or there being more corrupted data than "good", a significant number of signal elements would be rejected, leading to an ill-conditioned matrix with adverse effects on the quality of derived diffusion tensors (Chang et al. (2009)). While this was largely solved in the later iRESTORE modification through the implementation of condition number (Skare et al. (2000)) constraints, the RESTORE family of methods are still fundamentally limited by the diffusion tensor model. Given that a
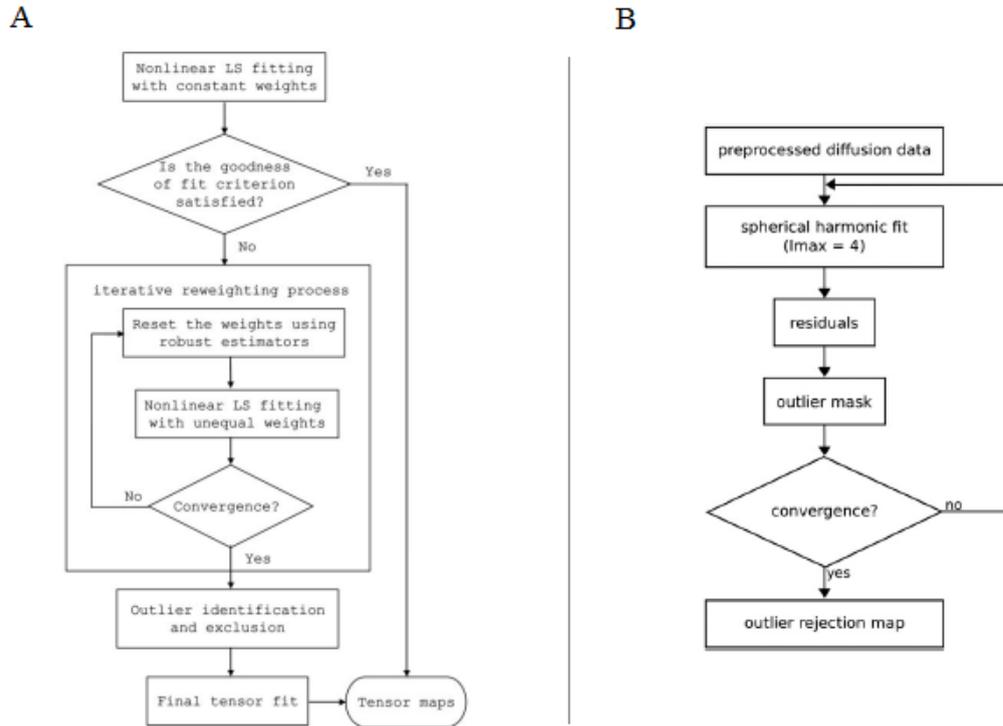
FIGURE 3.1: Process flow diagrams for (A) RESTORE, a reproduction of Figure 1, Chang et al. (2005) and (B) HOMOR, a reproduction of Figure 1, Pannek et al. (2012).

majority of white-matter voxels contain two or more fibre populations (Jeurissen et al. (2013)), the subsequent non-Gaussian diffusion profiles cannot be well described using a single tensor and thus distinguishing between residuals caused by genuine corruption and those caused by incorrect model selection is not easily achieved. Put simply, for non single fibre population signals the limitations of the diffusion tensor mean that either RESTORE algorithm is still likely to be discarding perfectly valid signal data in an effort to fit an ill equipped model – a drawback that is likely to become increasingly apparent as diffusion weightings are increased (e.g. for HARDI imaging) and the differentiations between single and multi-fibre population signals become more pronounced.

To address this limitation Pannek et al. (2012) proposed HOMOR. Operationally, HOMOR remains similar to the earlier RESTORE methods. There is a voxel labelling phase consisting of a $3 \times SD_{noise}$ goodness of fit test, followed by an iterative process designed to eliminate outliers. The difference between HOMOR and RESTORE lies in the assumed diffusion models; moving away from the diffusion tensor, HOMOR opts for a fourth order spherical harmonic representation. Recalling the previous chapter, one advantage of a spherical harmonic representation (with respect to the diffusion tensor) is the ability to accurately represent axially symmetric non-Gaussian diffusion; however, truncation to the fourth order severely limits the ability to represent higher angular frequencies. The assumption made by HOMOR is that, unlike fODF's whose sharp impulses require high angular frequencies for accurate representation, fourth order harmonics should be

sufficient to represent most normal diffusion weighted signals within the $1000s/mm^2$ to $3000s/mm^2$ b-value range, given that they are assumed to vary smoothly for the majority of fibre configurations. Should any high frequency information be present within the DW signal – i.e. potential corruption – the fourth order model will not prove sufficiently descriptive and thus the corresponding signal elements are theorised to produce a large residual that should allow for their identification and elimination. Algorithmically, HOMOR consists of an initial fourth order harmonic fit, allowing an initial labelling of corrupted voxels by highlighting those with $3 \times SD_{noise}$ residuals (with SD_{noise} calculated according to Walker et al. (2011)). Following labelling, each highlighted voxel is then subjected to a further fourth order harmonic fit and those elements with supra threshold residuals eliminated. This fit and eliminate process is then iterated until all remaining signal elements lie under the $3 \times SD_{noise}$ threshold. The accompanying flow diagram for this algorithm is shown in Figure 3.1.

Results presented in the HOMOR publication seem to demonstrate the expected improvements. Compared with RESTORE, HOMOR tends to both detect fewer voxels containing outliers and, where detections correspond, fewer corrupted signal elements – particularly at higher diffusion weightings. A reasonable assumption here is that the higher order model does indeed provide a better description of non-Gaussian diffusion and thus fewer erroneous detections/eliminations are occurring as a function of fibre geometry rather than genuine signal corruption. With that said, while the HOMOR publication does make a reasonable case that this is true, there are no real "gold standard" examinations. This is to say that while the frequency of detection and element elimination are examined (as well as the effects on derived parameters), little attempt is made to determine whether the *right* signals are being eliminated which, for methods such as these, is paramount.

To address these concerns, part of this chapter will be dedicated to providing a more thorough examination of HOMOR within a controlled simulated environment. However, this is not the main focus. In conjunction with the examination of HOMOR, we present a comparison with our own technique, RESDORE: Robust Estimation in Spherical Deconvolution by Outlier Rejection. While primarily a voxel wise method following the established detection and iterative elimination template, RESDORE possesses a scalable formulation that allows easy adaptation to simultaneous processing of an arbitrary number of voxels detecting, with increasing sensitivity as more voxels are included, common artefacts at a region, slice or even volume wide scale – thus providing a single novel technique for both localised and global error detection. While this will shortly be covered in more detail, the core RESDORE algorithm resembles an amalgamation of Richardson-Lucy deconvolution (Dell'Acqua et al. (2007)) and a non-negative sparse coder training phase (Hoyer (2002)). In combination, these two methods allow case specific manipulation of an (initially) isotropic dictionary element to account for

common outliers (across the provided signal samples) in the spherical deconvolution and by extension their detection and elimination.

The work presented in this chapter therefore consists of a performance comparison (where applicable) between HOMOR, RESTORE and our own novel RESDORE technique. Beginning with a series of simulations, we will attempt to highlight the strengths and weaknesses of each algorithm – paying particular attention to the accuracy and reliability of rejections. Following these simulations we will then proceed to more subjective analysis of real subject data. For reasons that will be explained shortly, the frequency and location of corrupted signal detection in vivo is not of particular concern (the images and indeed detection technique/parameters are the same as those used by Pannek et al. (2012), thus any strenuous examination would be a duplication of previous work), however there will be examination of comparative differences in derived metrics. Finally, a heavily corrupted case study will be examined to demonstrate algorithm performance on both a local and slice-wide scale.

3.2   RESDORE

Algorithm initialisation depends on the intended target (for example, a goodness of fit criterion must be checked if proceeding voxel-wise – a point that will be discussed later) however, at the core, RESDORE consists of a modified sparse coding algorithm (Olshausen and Field (1996)). Sparse coding can be thought of as the attempt to describe an observed signal using a linear combination of the smallest possible number of dictionary elements which, with various modifications, has applications in compressive sensing (Lustig et al. (2007)), image de-noising (Hyvarinen et al. (1999)) and simulation of neuronal activities (particularly the optical systems, Vinje and Gallant (2000)). Of particular interest to this work, however, is the manner in which non-negative sparse coder (NNSC) variants can learn dictionary elements from training data – the non-negative constraint is especially important given the properties of diffusion discussed at length in the previous chapter. The particular algorithm of interest was proposed by Hoyer (2002) and proceeds as follows:

Algorithm for NNSC

1. Initialise A_0 and S_0 to random positive matrices and rescale each column of A_0 to the unit norm. Set $t = 0$

2. Iterate until convergence:
   1. $A' = A^t - \mu(A^t S^t - -X)(S^t)^T$
   2. Any negative values in A' are set to zero.
   3. Rescale each column of A' to unit norm, then set $A^{t+1} = A'$.
   4. $S^{t+1} = S^t . * ((A^{t+1})^T X)./((A^{t+1})^T (A^{t+1}) S^t + \delta)$
   5. Increment t.

such that the output would be $X \approx AS$, where X is a set of exemplar signals, A is a dictionary (with each column encoding a separate element or basis vector) and S provides a matrix of sparse weighting vectors. For reference, the operators $.*$ and $./$ represent, respectively, the Hadamard product and Hadamard division of the related matrices.

If we disassemble this algorithm, steps 2.1 to 2.3 comprise a gradient decent designed to multiplicatively (at a rate limited by the tuning parameter μ) optimise a set of basis vectors (A) to best describe X given a constant S. Step 2.4 comprises a separate optimisation designed to select optimal non-negative S values (with sparseness modulated by the tuning parameter δ) to best describe X given a fixed A. As the algorithm is allowed to iterate, the two phases essentially trade against each other until the dictionary and weights reach a local optimum.

The problem with sparse coding in this manner is that since the A and S matrices are initialised randomly, there is no guarantee that repetitions of the experiment will converge to the same result and no way to include *a priori* information in the formulation. In principal though, if we were to arrive at a dictionary using noise free training data, residuals to the sparse fit to new data (achieved simply by iterating step 2.4 and beginning with a uniform S) might give indications of corrupted signals since they would not be easily represented by the learned dictionary elements – much in the same manner as HOMOR employs the fourth order spherical harmonic series.

Fortunately however, recalling Section 2.2.11 the equation in step 2.4 should be recognisable (minus the tuning parameter δ) as the Richardson-Lucy deconvolution with modifications for Gaussian noise – an optimisation that has already seen use in the recovery of fODFs from DW-MRI images (Dell'Acqua et al. (2007)). This tells us that the dictionary A can be easily modified to incorporate DW-MRI assumptions; more specifically, we can modify A such that each element represents the DW-MRI signal (assuming Eq.

3.1) produced by fibrous tissue oriented along one of a given set of directions. In doing so this ensures that any S fit to an observed signal will, without modification of the dictionary through step 2.1, correspond the output of a Richardon-Lucy fODF recovery.

$$Sig(v,r) = exp^{-b(\lambda cos^2(\theta) + \beta sin^2(\theta))}$$

$$
\begin{aligned}
&\text{v} = \text{gradient direction} \\
&\text{r} = \text{direction of principal diffusion} \\
&\lambda = \text{axial diffusivity} \\
&\beta = \text{radial diffusivity} \\
&\theta = \text{angle subtended by v and r} \\
&\text{b} = \text{diffusion weighting}
\end{aligned}
$$

(3.1)

$$Sig(v,r) = exp^{-b\alpha cos^2(\theta)}$$

$$\text{Where } \alpha = \lambda - \beta$$

(3.2)

Returning to the NNSC dictionary update step; we modify the original Richardson-Lucy dictionary such that each element represents an "alpha" response (Eq. 3.2, described by Dell'Acqua et al. (2013)) – essentially eliminating representation of radial diffusivity – and then expand the dictionary to include a single isotropic component with diffusivity corresponding to the eliminated radial response (Figure 3.2a). With the directed elements unable to represent radial diffusivity and a sufficiently tuned sparseness modifier (δ), any subsequent spherical deconvolutions will be forced to include some multiple of the isotropic component in order to provide an accurate description of the observed signal. Using this to our advantage, the dictionary update step can be modified (Equation 3.3, written in MATLAB notation) such that errors to the fODF fit are corrected by manipulation of the isotropic component only.

$$A(:,1) = A(:,1) - \mu * ((A * S - X) * S(1,:)')$$

where the first dictionary element, A(:,1), is initially isotropic

(3.3)

The element elimination procedure then progresses as follows. To initialise, the algorithm is first allowed to run a small number of times ($\approx 20 - 50$, or $\approx 1/4$ of intended iterations) without dictionary update, providing an initial fODF estimate from which to move forwards (the Richardson-Lucy algorithm is known to be robust to corrupted data and thus should provide a reasonable first approximation, Dell'Acqua et al. (2007)). In the remaining iterations, where the observed signal is adequately described by some combination of axially symmetric response functions (i.e. not corrupted), errors with respect to the *current* iteration's fODF estimate (S) can be corrected by improved weights on those symmetric elements in later iterations – the isotropic component will only be included to account for missing radial diffusivity, any significant errors should have been
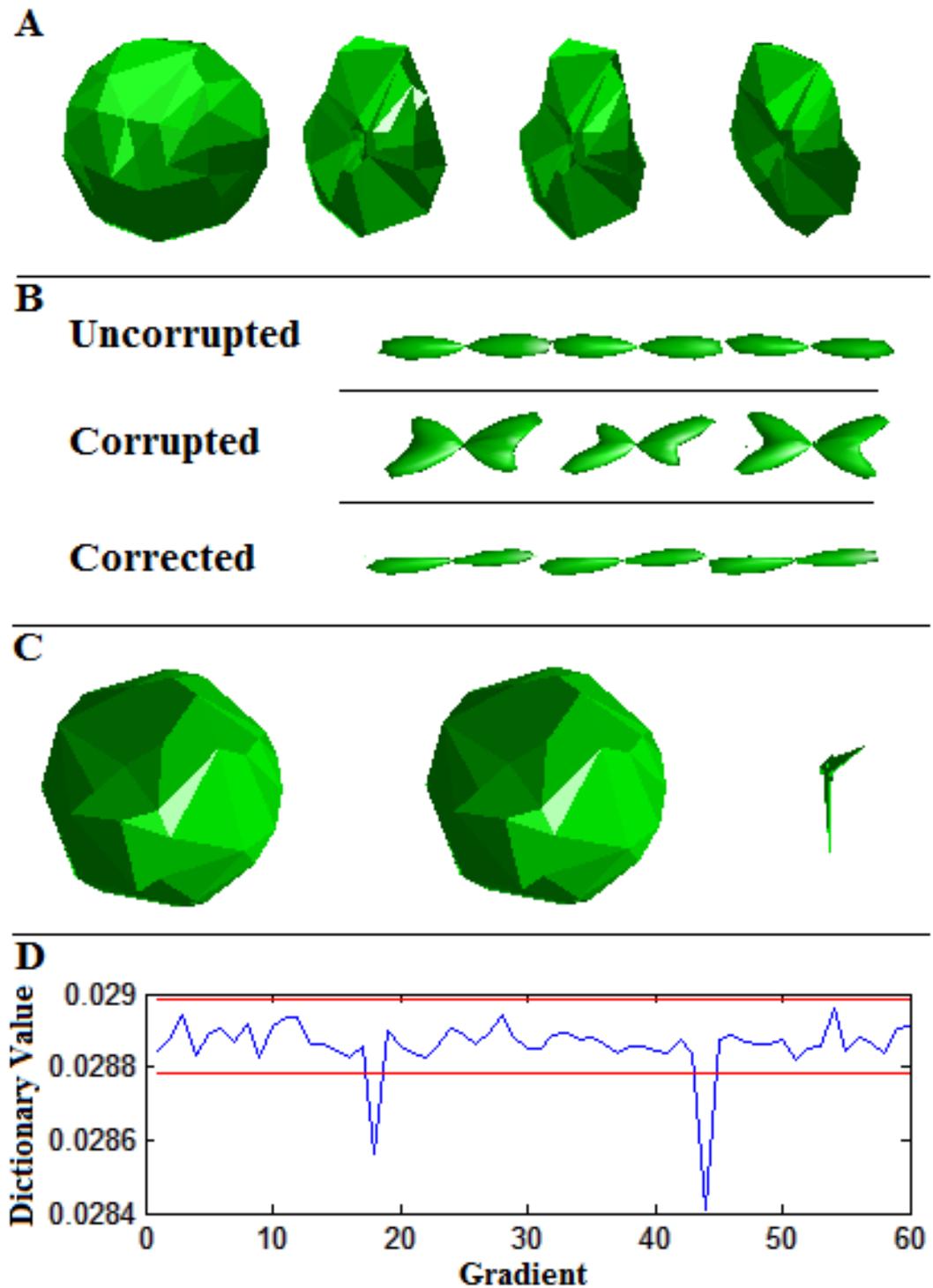
FIGURE 3.2: (A) Example of the modified RESDORE dictionary consisting of one isotropic element (left) followed by the first three of a series of alpha response functions. (B) Richardson-Lucy fODFs resulting from application to (1) an uncorrupted signal (2) corrupted signal (gradients 19 and 44 multiplied by 0.3) (3) signal following RESDORE correction (C) Resultant deformation to the isotropic component. Left: Original isotropic element. Center: Distorted following correction process. Right: Difference. (D) Blue: Plot of modified dictionary values (equivalent to center image in (C)). Red: Outlier thresholds.

eliminated by the preliminary fit and thus iterative updates will not significantly alter the isotropic profile. Where the signal cannot be described by a combination of axially symmetric responses, the only way in which the fODF fit may be significantly improved is by compensating for the residuals through update of the isotropic element (Fig. 3.2b-d). In a sense, this isotropic component therefore acts as a 'sink' for errors outside the assumed diffusion model, both improving the fODF fit (though we still would not recommend its use in any subsequent tractography) and importantly providing insight into the distribution of error through examination of deformations to its original isotropic structure (Figure 3.2).

The process for identification of outliers (and by extension corrupted signal elements) is by comparison fairly simple. Taking the components of the formerly isotropic dictionary element, we first calculate the interquartile range (IQR) as well as the 25% and 75% quartiles (denoted q_1 and q_3 respectively) then apply thresholds labelling any element outside $[q_1 - 1.5IQR, q_3 + 1.5IQR]$ as containing outliers for elimination (Figure. 3.2d, red lines indicate thresholds), effectively applying a three standard deviation check (Fig 3.3, a standard technique for univariate data) using the variance of the dictionary component deformation, rather than an externally calculated (i.e. SD_{noise}) threshold.
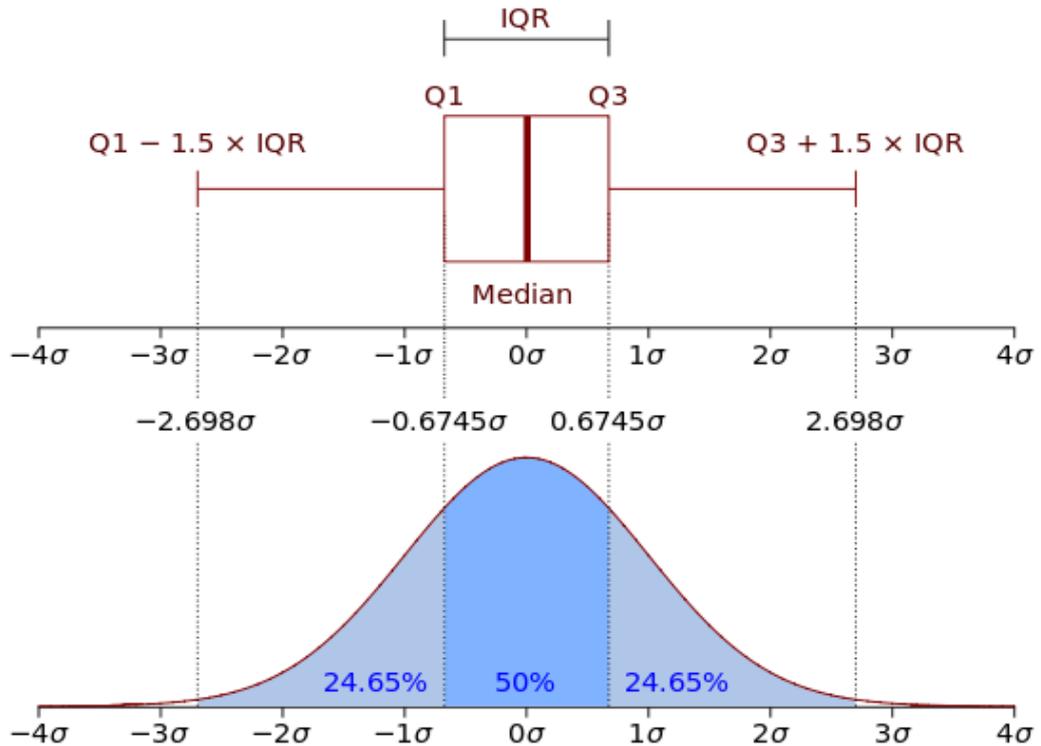
FIGURE 3.3: Relationship between the interquartile range test and the standard deviation of a univariate sample.

Given the adapted NNSC/Richardson-Lucy formulation, there is no limit (bar computational hardware) to the number of signals that may be simultaneously processed. Equation 3.3 merely stipulates that any dictionary updates are not in this case optimized for the individual signals, but for the signal corpus as a whole. In practice this means that when given a corpus containing only randomly distributed thermal noise where there is no overall bias to the resultant residuals (either in direction or magnitude), the best 'correction' is for the isotropic component to go unchanged. If, however, there is a consistent bias along given set of directions the effects of individual voxels on the overall correction will not cancel (as they would in thermal noise) and thus the overall response would be to modify the corresponding dictionary component(s). In this way we will demonstrate that RESDORE can be applied to an arbitrary number of voxels with minimal sensitivity to underlying thermal noise but with an increasing sensitivity to underlying corruption (owing to the increase in statistical power provided by additional examples).

3.3 Simulation

3.3.1 Data Simulation

Crossing fibre DW-MRI signal data were simulated assuming a 60 direction sampling scheme (Jones et al. (1999)) with 6 b0, component anisotropies ranging from 0.1 to 0.9 (0.2 interval), crossing angles ranging from $0°$ (essentially single fibre) to $90°$ ($10°$ interval), fixed (component) mean diffusivity $0.7 \times 10^{-3} mm^2/s$ and b-values ranging from $1000s/mm^2$ to $3000s/mm^2$ ($1000s/mm^2$ interval). Gaussian noise was then added in quadrature (500 repetitions per signal tuple), generating signal to noise ratios ranging from 10:1 to 70:1 (20:1 interval).

3.3.2 Voxel Labelling

The first component of all voxel-wise methods is the initial detection and labelling phase. Although it was earlier stated that the patterns and spatial distribution of in vivo detection are not a primary concern for this work (recalling that the image data and voxel labelling methods will conform with those acquired/employed by Pannek et al. (2012) – only a tertiary examination will be required to ensure that the methods are correctly implemented), the following subsection will show that owing to different fitting procedures (RESTORE) or dictionary manipulation (RESDORE), it is possible for the rejection component of an algorithm to outperform its labelling component, reliably rejecting signal components with lower degrees of corruption than would be detectable during the labelling phase. Thus, in order to conduct a thorough comparison between rejection techniques, one must first account for what they are likely to detect (i.e. a

method may perform flawless rejection, but if the detection component is unreliable the overall result might be poor in comparison). With this in mind, we will begin by briefly examining the detection performance of the four following methods (two of which are based on existing literature, two on the Richardson-Lucy deconvolution employed by RESDORE):

1. Detection by residuals outside $3 \times SD_{noise}$ to a non-linear least squares diffusion tensor fit (i.e. the RESTORE method).

2. Detection by residuals outside $3 \times SD_{noise}$ to a fourth order spherical harmonic fit (using fitting procedures of the MRTrix toolbox (Tournier et al. (2012) as specified by HOMOR).

3. Detection by residuals outside $3 \times SD_{noise}$ to a Richardson-Lucy deconvolution (Dell'Acqua et al. (2007)).

4. Detection by residuals outside $3 \times SD_{noise}$ to a damped Richardson-Lucy deconvolution (Dell'Acqua et al. (2010)).

To provide detectable corruption, signals were modified by multiplying, at random, signal elements from combinations of one, five and ten gradient directions by values ranging from 0 to 2 (i.e. total drop out to hyperintensity) at a 0.1 interval; essentially expanding upon the corruption model employed by Chang et al. (2005) (multiplying random combinations of elements by 0.5/1.5), thus facilitating easier comparison between that work and our own. For the purposes of this experiment an exact value for SD_{noise} is already known (derived from parameters of the simulation – dependent of course on the specific SNR), therefore estimation from data is unnecessary. Finally, both the Richardson-Lucy and damped Richardson-Lucy algorithms were configured for a 752 direction dictionary with individual responses calibrated to an anisotropy of 0.9 with the damped variant using geometric and regularisation thresholds of 8 and 0.04 respectively (no isotropic dictionary element is included at this stage).

Beginning with the simplest possible case, Figure 3.4 displays detection results for a single fibre configuration at b-values ranging from $1000s/mm^2$ to $3000s/mm^2$ at an SNR of 30:1 and with component anisotropy of 0.7. Confirming suspicions raised by the HOMOR publication, detection performance using the tensor method (red) is severely hampered at high diffusion weightings, recording significant increases in the number of false positive detections as the DW-MRI signal increasingly deviates from the Gaussian assumption – observable by comparing Figure 3.4a with Figure 3.4c at the 1× (i.e. no corruption) interval. An important characteristic demonstrated by all detection techniques is a reduction in overall sensitivity as diffusion weightings increase. Ignoring for a moment the diffusion tensor based detections (whose superior performance against high degrees of corruption is likely as much to do with the inapplicability of the model as the
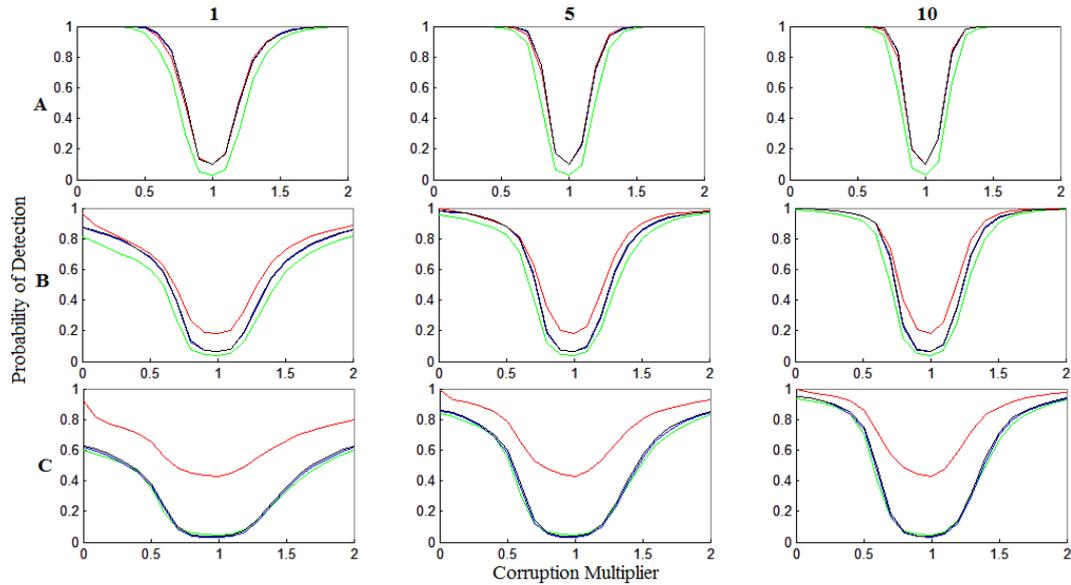
FIGURE 3.4: Single voxel single fibre detection results at (A) $1000s/mm^2$ (B) $2000s/mm^2$ and (C) $3000s/mm^2$ diffusion weightings against (column-wise) combinations of 1, 5 and 10 corrupted signal elements. Colour codes: Red – diffusion tensor (RESTORE). Green – spherical harmonics (HOMOR), Blue – Richardson-Lucy and Black – damped Richardson-Lucy

accuracy of those detections), there is a significant drop in detection rates as diffusion weightings increase. Examining Figure 3.5, which displays the likelihood of detection for a single corrupted gradient with respect to the angle subtended between that gradient and the fibres orientation, there is a readily apparent decline in sensitivity parallel to the fibre orientation (for both signal drop out – Figure 3.5a $0.5\times$ corruption multiple – and signal hyper-intensity – Figure 3.5b $1.5\times$ corruption multiple) which worsens with increased diffusion weighting.

The most probable cause for this deterioration is diminishing absolute differences between the corrupted and non-corrupted signal data. For a given diffusivity, increasing the diffusion weighting will increase the corresponding signal attenuation. As a result, if we then multiply each signal by a fixed percentage (i.e. our corruption values ranging from 0 to 2), the absolute difference between corrupted and non-corrupted signals will naturally be smaller at higher diffusion weightings owing to the lower base signal intensity. Examining the diffusion profile as a whole, in a single fibre system the highest diffusivity (i.e. lowest uncorrupted signal intensity) corresponds with the fibre axis. By extension, corruption via multiplication will produce smaller absolute deviations along the fibres axis than orthogonal, allowing model fitting regimes to better adjust to axial corruption (it is naturally less costly to adjust a model fit to account for small error than it is to account for a large error), thus increasing the probability that such errors will fall under the $3 \times SD_{noise}$ threshold avoiding detection.

In practical terms, the previous observation has two implications: (1) Given that diffusion weighted signals cannot have an intensity less than zero, signals undergoing drop
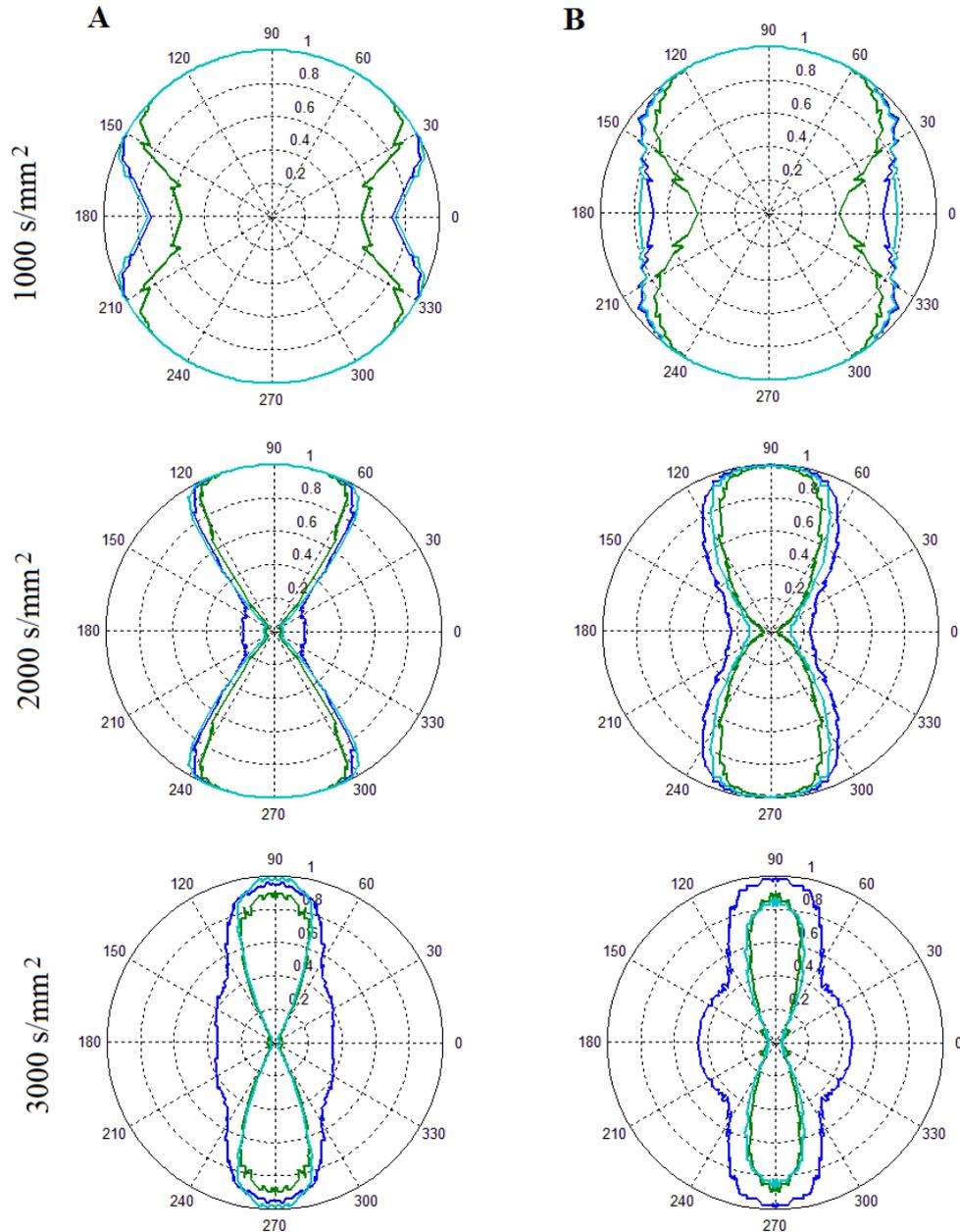
FIGURE 3.5: Plot of the probability of detection against the angle between corrupted gradient and fibre orientation. (A) One gradient corrupted by a 0.5 multiple. (B) One gradient corrupted by a 1.5 multiple. Dark blue – diffusion tensor. Cyan – spherical harmonics. Green – Richardson-Lucy. Black – Damped Richardson-Lucy (results near identical to standard Richardson-Lucy).

out can, at worst, only deviate from the expected intensity **by** the expected intensity (i.e. the $0\times$ multiple, though practically speaking the rectified noise floor, Jones and Basser (2004b), insures the signal never reaches 0) thus, as our simulations show, increases in diffusion weighting will lead to increasing difficulties in detecting corruption along the fibre axis – for example, note that in Figure 3.5, reliable detection of a 50% drop out is only possible at near orthogonal angles to the fibre at $3000s/mm^2$ but may be detected

easily at most angles given a $1000s/mm^2$ weighting. (2) On the opposite end of the spectrum, hyper-intensity is not bound to any particular multiple of the true signal intensity (limited instead by the dynamic range of the scanner). This means that while small (relative to expected) overestimations of the signal intensity will be increasingly difficult to detect as b-values increase, the maximum potential deviations are decoupled from the direction of measurement and, as such, for a given absolute deviation it may actually be easier to detect additive corruption along the fibres orientation than orthogonal since, in terms of proportion, the differences will be larger.
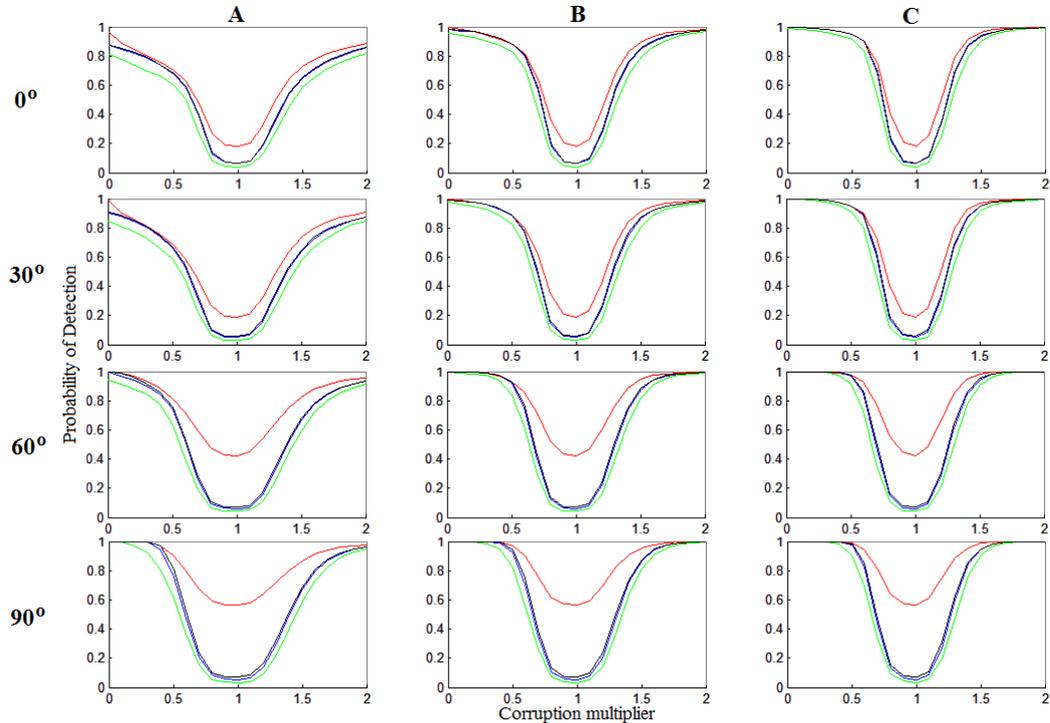
FIGURE 3.6: Single voxel crossing fibre detection results ranging from 0° to 90°. (A) Single corrupted gradient. (B) Two corrupted gradients. (C) Three corrupted gradients. Colour codes: Red – diffusion tensor (RESTORE). Green – spherical harmonics (HOMOR), Blue – Richardson-Lucy and Black – damped Richardson-Lucy

Adding a second fibrous component, Figure 3.6 demonstrates the effect of crossing angle on detection reliability (component $FA = 0.7$, $2000s/mm^2$, 30:1 SNR angles from 0° to 90°, 30° interval). As the angle of intersection is widened the expected deficiencies of the tensor model become apparent, demonstrating a large number of false positive detections at angles exceeding $\approx 60°$. In contrast to this, the higher order methods demonstrate substantial improvements in detection reliability as crossing angles increase, particularly in cases of signal drop out. The reason for this is likely twofold. Firstly, in a crossing fibre system the maximum potential apparent diffusivity is by definition lower than that of a comparable single fibre system. This difference results in a higher signal minima and, by extension, larger and more easily detectable absolute differences between corrupted and non-corrupted signal intensities – particularly as the crossing angle widens and the individual components signals increasingly dis-entwine. Secondly;

owing to a larger number of free parameters, higher order models are naturally able to provide a better description of an observed signal than the simpler diffusion tensor alternative. However, where that signal is relatively non-complex, i.e. a single fibre that may be adequately described by the tensor model, then the additional complexity of the higher order models essentially goes towards providing a 'better' description of the inherent noise/corruption, essentially making outliers harder to detect (i.e. over fitting). By including a second fibre component, the previously free parameters are now called upon to provide descriptions of a genuinely complex signal, naturally decreasing the degree of over fitting to noise/corruption and by extension making such artefacts more readily apparent in the form of larger residuals.

For reference, Figure 3.7 displays a radial plot of the in-plane detection probabilities at each angular separation. Note however that, for ease of visualisation, the signals used in production of this Figure were two dimensional (no Z axis component).Thus, while not accurately describing real 3D detection profiles (detection at a given angle becomes easier out of the fibres plane due to lower signal attenuations, these Figures do not capture that reality), the link between maxima in the apparent diffusivity (see Figure 3.7c depicting normalised apparent diffusivity for each fibre configuration) and minima in the detectability of signal corruption (Figure 3.7a – 0.5× multiple. Figure 3.7b, 1.5× multiple).

Completing this preliminary investigation of detectability, Figure 3.8 displays the effects of (A) b-value, (B) FA and (C) SNR on the 90° crossing fibre configuration (one corrupted gradient only). Starting with diffusion weighting, as in the single fibre case, increasing diffusion weighting leads to reduced sensitivity (for reasons already covered at length). Proceeding to fractional anisotropy, trends in detectability depend on the approach used. At lower anisotropies, for example, signals are sufficiently similar to a single fibre response for diffusion tensor to provide reasonable detection, though this quickly drops of as anisotropy is increased and the differences between single and crossing-fibre responses become more apparent. Similarly, the spherical harmonic method also appears to function best at low anisotropies (which makes sense, given that these are the most likely to provide the expected smoothly varying profile), dropping off slightly as anisotropy increases and the diffusion profiles increase in sharpness. Finally, the Richardson-Lucy methods (both providing similar results) evolve in the opposite direction, providing slight improvements in detection as one progresses from $FA = 0.3$ to $FA = 0.7$, though faltering (with large numbers of false positive detections) at the final $FA = 0.9$ step – given the $FA = 0.9$ alpha calibration (see Chapter 4 for the reasoning behind this selection), it makes sense that detection would improve as anisotropy increases (the signals will increasingly correspond to the model), though at $FA = 0.9$ it is probable that the elimination of the radial component in the alpha calibration causes radial elements of the simulated signal to trigger false positive detections.
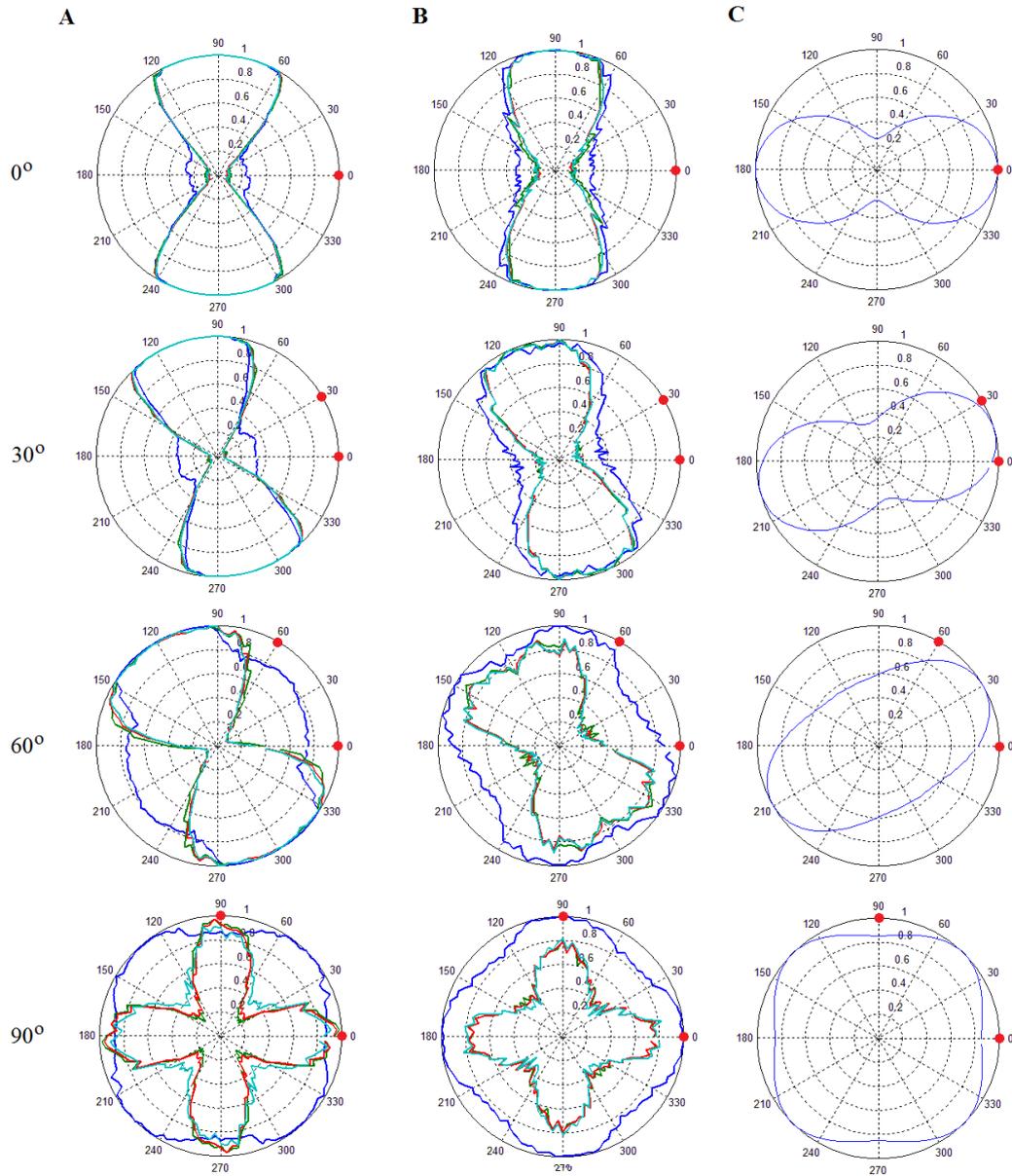
FIGURE 3.7: Plot of the probability of detection against the angle between corrupted gradient and the axis aligned fibre component orientation. (A) One gradient corrupted by a 0.5 multiple. (B) One gradient corrupted by a 1.5 multiple. (C) 2D ADC profile of the described fibre configuration, included for reference. Dark blue – diffusion tensor. Cyan – spherical harmonics. Green – Richardson-Lucy. Red – Damped Richardson-Lucy. The large red dots indicate individual component orientations.

Comparing the four methods in general; as expected, the diffusion tensor model employed by RESTORE is largely inadequate at higher diffusion weightings (Fig 3.4c) or in the presence of high anisotropy crossing fibre (Fig 3.8b). Using such a method these conditions will undoubtedly lead to a large number of false positive detections and, by extension, elimination of large quantities of perfectly acceptable DW-MRI signal data (following the rejection phase) for sake of force-fitting an unsuitable model. Proceeding to the higher order detection models, generally speaking all methods would perform
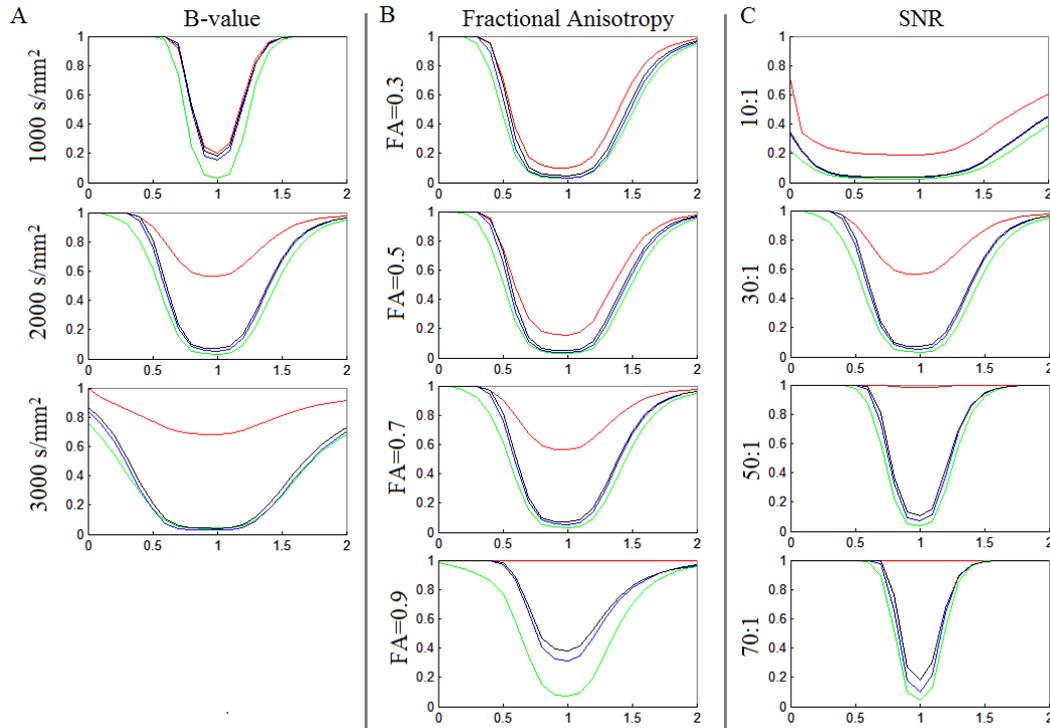
FIGURE 3.8: Single voxel crossing fibre detection results, one corrupted element only. (A)$FA = 0.7$, 30:1 SNR, 90° intersection, b-values ranging from $1000s/mm^2$ to $3000s/mm^2$. (B) $2000s/mm^2$, 30:1 SNR, 90°, component anisotropies ranging from $FA = 0.3$ to $FA = 0.9$. (C) $2000s/mm^2$, $FA = 0.7$, 90°, SNR ranging from 10:1 to 70:1. Red – Tensor detection, Green – Spherical Harmonic detection. Blue – Richardson-Lucy detection. Black – Damped Richardson-Lucy detection.

similarly under a majority of conditions. However, if one wishes to achieve optimal detection performance then different algorithms could be preferential depending on the precise aims. At lower anisotropies and diffusion weightings (Figure 3.4a, 3.8), for example, the spherical harmonic method generally demonstrates a noticeable reduction in false positives. On the other hand, the spherical harmonic method appears to be slightly less sensitive to corruption across the wider spectrum of multipliers – notice that in Figures 3.4, 3.6 and 3.8, the green line (spherical harmonics) is almost invariably below that of the Richardson-Lucy based methods. While the difference is relatively minor in most cases, it does become pronounced with signal drop-out in the $FA = 0.5$ to $FA = 0.7$ region (e.g. Figure 4.8b) – though this is generally not an issue as the number of corrupted gradients increases (e.g. Figure 3.6 compare sub-Figure A with C).

For general purposes then, any of the higher order methods are likely to be suitable, though given that a spherical harmonic fit is far easier to calculate than a full Richardson-Lucy deconvolution, for reasons of computational efficiency the spherical harmonics are the option of choice. With this in mind, as a general rule of thumb we can expect, on average, reliable detection of corruption with multiplies in the regions of $\leq 0.5\pm0.1$ and $\geq 1.5\pm0.1$ (See Figures 3.4, 3.6 and 3.8) with improved chances of detection (as would seem

obvious) is the corrupted gradient orthogonally subtends the local fibre orientation or multiple gradients are corrupted (again, increasing the chance of orthogonal corruption).

3.3.3 Signal Element Elimination

The second component of a voxel-wise method is the elimination step, in which voxels highlighted by the previously explored detection step are subjected to further processing in order to determine which signal elements triggered the detection and should thus be eliminated. For this step we will examine the performance of three methods, RESTORE, HOMOR and RESDORE, each implemented as follows:

1. RESTORE: Implementation extracted from the ExploreDTI processing toolbox (Leemans and Jones (2009a)). The source code was deemed consistent with the algorithm described by Chang et al. (2005).

2. HOMOR: Implemented as described in Pannek et al. (2012) using calls to the MRTrix toolbox in order to perform spherical harmonic fitting.

3. RESDORE: Implemented (in MATLAB) according to the description provided in Section 3.2. The sparseness constraint (δ) was set to 0 (rendering the fODF fit algorithm identical to Dell'Acqua et al. (2007)) and the dictionary update damper (μ) set to 0.05. The algorithm was allowed to iterate 200 times.

Figure 3.9 provides elimination results for a single fibre configuration with $FA = 0.7$, $SNR = 30 : 1$ and diffusion weightings varying from $1000s/mm^2$ to $3000s/mm^2$ - table 3.1 provides an explanation of the colour coding.

TABLE 3.1: Colour coding legend for Figure 3.9

Colour	Meaning
Red	Perfect result.
Blue	All corrupted elements detected but false positives also included.
Black	Incomplete detection of corrupted elements, but also no false positives.
Green	Sum of red and black lines (≥ 1 corrupted element detected, no false positives).

To briefly explain Table 1, and by extension the information relayed by associated figures, each category represents one of a larger set of possible outcomes to the signal element elimination experiment (note that not all outcomes are examined, the four possibilities selected are simply those which, following preliminary examination, best highlight differences between the three algorithms). To begin with the simplest, red lines convey the probability of achieving a perfect result, i.e. the probability that all of the deliberately corrupted signal elements will be correctly eliminated with no false positives. Expanding upon this, blue lines convey the probability that all corrupted elements will be identified

while also allowing for the possibility of additional false-positive eliminations – a common result for both the HOMOR and RESTORE algorithms. In reality, however, it is not always possible to eliminate the full set of corrupted elements; for reasons covered in later within this Section it is, for example, more difficult to detect corruption parallel to the fibre orientation than perpendicular, leading to possibility of only a subset of the corrupted elements will be successfully eliminated. To account for this, black lines indicate the probability of eliminating a smaller subset of the corrupted elements while not allowing for additional false positive eliminations. Finally, the green line (essentially the sum of the red and black lines) represents a kind of summary term, indicating the probability of eliminating at least one corrupted element (up to and including the whole set) while, again, not allowing for any false positive eliminations. While there are of course additional scenarios, e.g. partial recovery with false positives, these were found to convey relatively little additional information that could not be extracted from the mean true and false positive statistics also included within this Section (presented as separate figures).
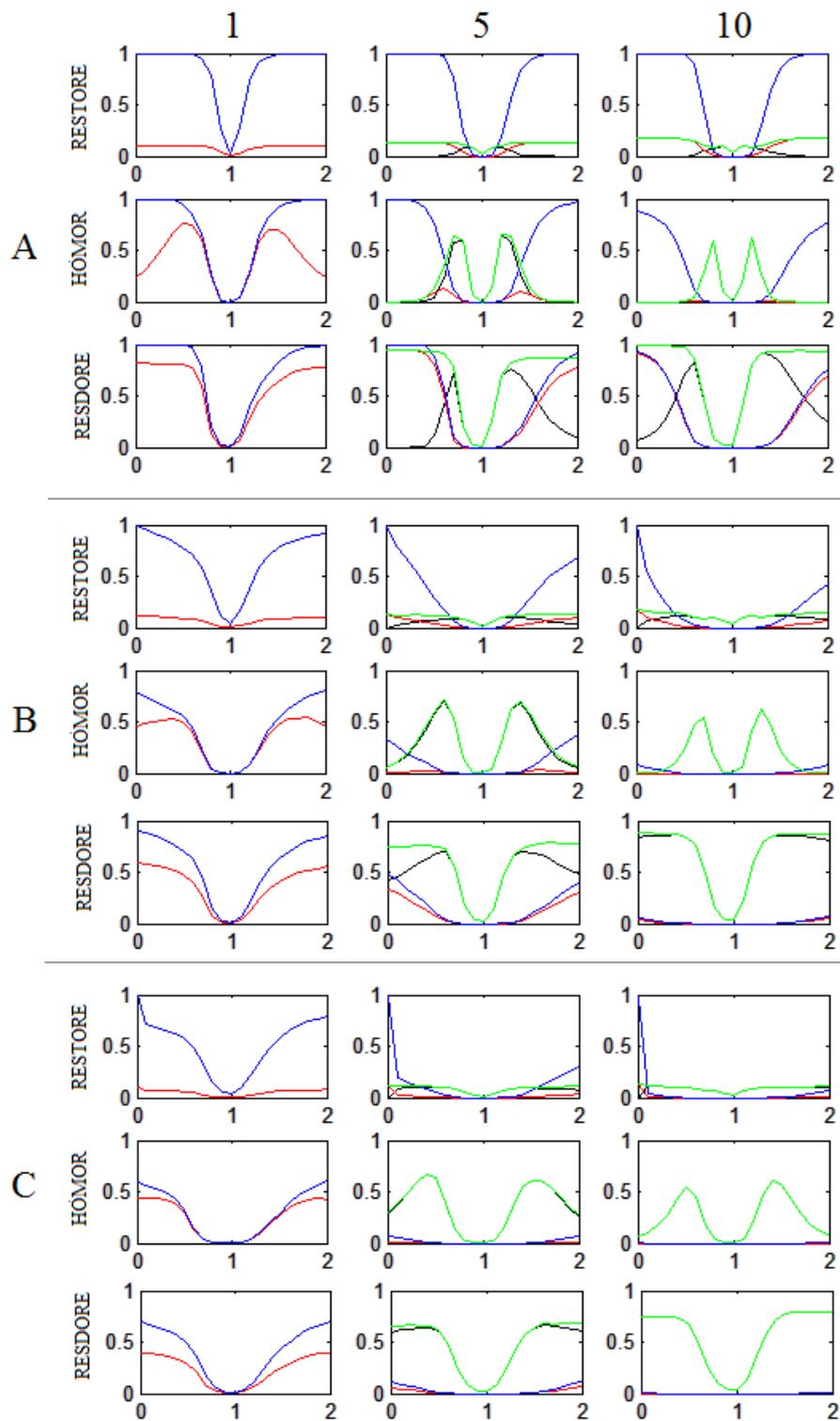
FIGURE 3.9: Single voxel single fibre elimination results at (A) $1000s/mm^2$. (B) $2000s/mm^2$. (C) $3000s/mm^2$ against (column-wise) signals containing 1, 5 and 10 corrupted elements. Colour coded according to table 3.1. Column convention will apply to all following figures up to and including Figure 3.18.
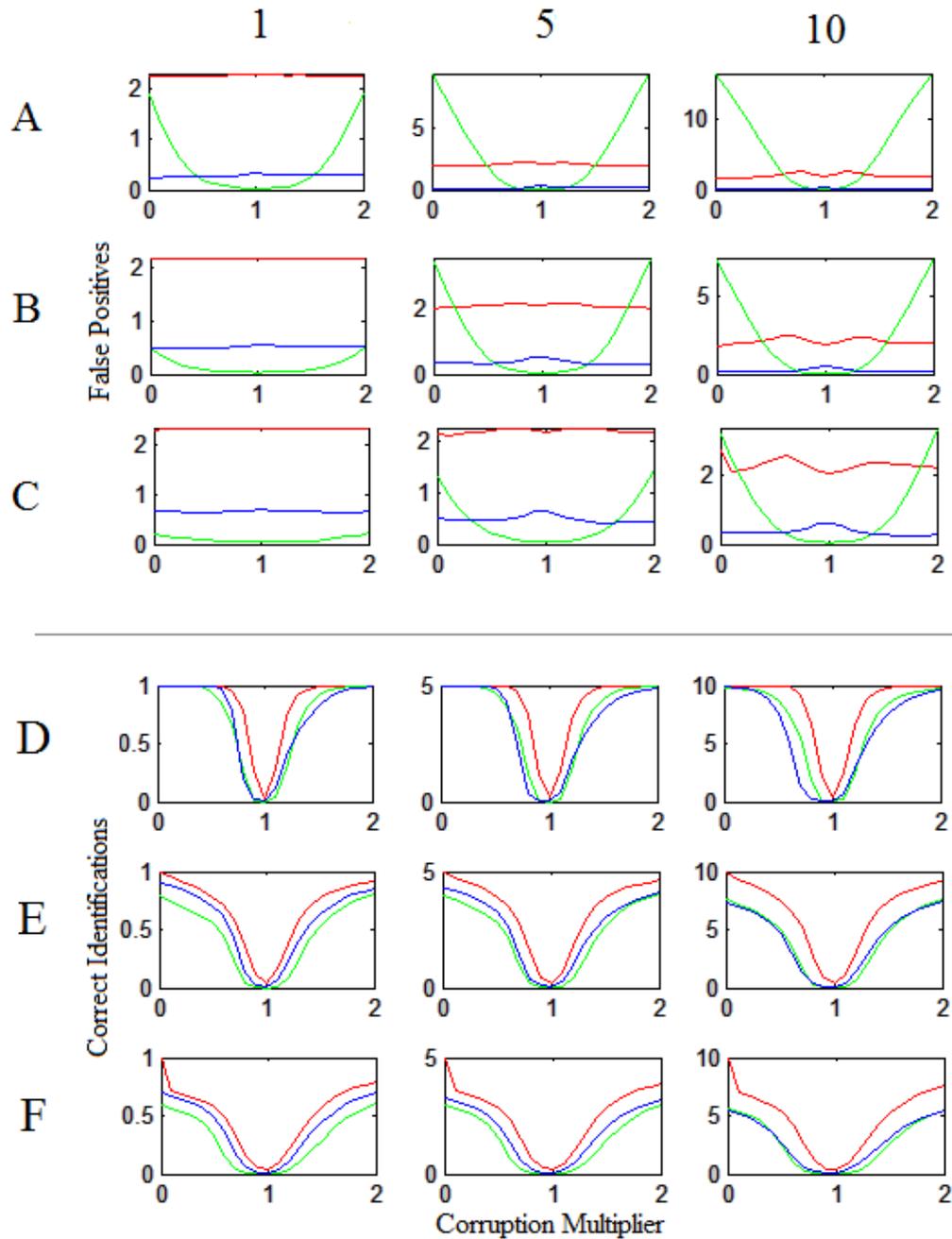
FIGURE 3.10: Single voxel single fibre false positives. (A) $1000s/mm^2$. (B) $2000s/mm^2$. (C) $3000s/mm^2$. Single voxel single fibre set recovery. (D) $1000s/mm^2$. (E) $2000s/mm^2$. (F) $3000s/mm^2$. Red - Restore, Green - HOMOR, Blue - RESDORE.

Analysis of relative performance depends on the criterion of interest. Beginning at low diffusion weightings, if one simply wishes to eliminate all corrupted elements regardless of additional false positives then examination of the blue lines indicates that RESTORE is a suitable choice; appearing to be the more sensitive than the alternatives to low magnitude corruption (note the tighter 'V' shaped profile) and most likely to identify all corrupted elements. Considering that the tensor model is designed for exactly this application (single fibre, low diffusion weighting), these findings are not entirely unexpected. Comparing HOMOR with RESDORE under the same conditions, both have comparable performance, however, if we then observe false positives as a criterion of interest, Figure 3.10a (separate colour scheme, please see caption) indicates a substantial increase in the number of false positives generated by HOMOR as the extent of the corruption (either by magnitude or number of corrupted elements) increases, while RESTORE and RESDORE remain comparatively consistent (with RESDORE generating fewer false positives). The differences in apparent sensitivity between RESTORE and the two higher order models can be then explained by examining Figure 3.10d. Here we see that RESTORE is in almost all cases more likely to recover more of the intentionally corrupted set (true positives) at a given multiplication, while HOMOR and RESDORE are more evenly matched.

An interesting observation is the opposite manner in which HOMOR and RESDORE react to increased corruption – specifically with respect to the production of false positives. As discussed in greater detail later, HOMOR tends to increase the number of false positive detections as the corruption worsens while RESDORE achieves the opposite. The reason stems from differences in the manner in which each algorithm defines corruption and will be key to future comparisons. As covered in Section 3.2, the RESDORE elimination phase does not rely on any estimation of SD_{noise} for its elimination criteria, instead, corrupted elements are highlighted by examining elements of the (formerly) isotropic dictionary component for deformations outside approximately three standard deviations of variation within **that** component. The result of this formulation is that the 'definition' of corruption varies on a per voxel basis and, importantly, all processed voxels are capable of producing a detectable outlier. When a signal with no/very minimal corruption is analysed, updates to the isotropic component will be limited to compensating for standard thermal noise with relatively low variance thus, expecting corruption, the algorithm may select the element(s) most affected by noise for elimination, rejecting on average ≈ 0.34 elements per trial at the $1\times$ multiple which, according to the Empirical rule (Ross (1987)) that suggests that our ≈ 2.8 standard deviation check would capture $\approx 99.5\%$ of normal variance, appears reasonably consistent with statistical expectations ($60 \times 0.005 = 0.3 \approx 0.34$, where 60 is the number of DW signal elements). Once true corruption is present, however, this has the effect of widening the variance of the deformations to the isotropic element (since large compensations will be required to deal with the corrupted elements) which, by extension, will reduce the significance of noise driven deformations leading to the observed reduction in false positives as the degree of

corruption rises. HOMOR, on the other hand, does not possess a situationally adaptive model (remembering that RESDORE deforms the isotropic component to ensure a more accurate fODF fit); thus, as the degree of corruption increases so will the cost of not altering the spherical harmonic fit to account for the corrupted elements – especially where the number of corrupted elements is large. It appears, therefore, that below a certain threshold of corruption the fit will remain true to the non-corrupted signal (explaining the low number of false positive surrounding the $1\times$ multiple), but once that threshold is passed the model fit must be altered to compensate. Note that this does not mean the fit is altered to perfectly match the corrupted elements, indeed, since such elements are still detected by the $3 \times SD_{noise}$ threshold this cannot be the case. Instead we are simply suggesting that the fit is increasingly biased by the corrupted elements, resulting in a 'drift' away from the ideal solution. Once the model fit is compromised, a third source of residuals is introduced – corruption, noise and now fitting error – which could easily push borderline (due to noise) elements over a detection threshold that is held rigid regardless of individual voxel conditions, resulting in the observed increase in false positive detections as corruption (and thus drift) worsens.

At higher diffusion weightings (Figures 3.10e and 3.10f, $2000s/mm^2$ and $3000s/mm^2$ respectively), results indicate the same reduction in sensitivity observed in the previous subsection, with each algorithm finding it increasingly difficult to detect the full set of corrupted signal element (most likely due to a difficulty in detecting fibre-axis aligned corruption). This observation echoes in Figure 3.9b-c in which the black lines (partial recovery, no false positives) essentially replace the full recovery line (red) in Figure 3.9a. Comparing individual methods, where the number of corrupted elements is relatively low, both HOMOR and RESTORE appear to be competitive choices (Fig. 3.9c, column 1), producing on average fewer false positive detections than the more sensitive RESTORE method. As the degree of corruption increases, HOMOR rapidly loses its competitiveness and the choice falls down to either RESTORE (sensitivity, Figure 3.10e-f, columns 5 and 10) or RESDORE (accuracy, Figure 3.10b-c, columns 5 and 10) depending on individual criteria.

An interesting point, however, is that when we examine the false positive rates at these higher diffusion weightings (Fig. 3.10b-c), any increase in b-value sharply reduces HOMORs worst case performance. Going back to Section 3.3.2, these changes can likely be explained by considering the same absolute deviation effects. The increased signal attenuation at higher diffusion weighting will lead to a smaller absolute deviation between corrupted and non-corrupted versions of a given signal which, by extension, reduces the fitting error caused by the corrupted signal and thus the resultant drift. Combined with the increased effect of noise (as represented by the non diffusion weighted SNR) at higher attenuation, residuals resulting from fitting drift will be comparatively negligible and thus far harder to incorrectly identify as the result of corruption.

Figures 3.11 and 3.12 introduce a second fibre component to the simulation, allowing performance to be tested across differing crossing fibre situations as the angle of intersection widens from 30° to 90° (other parameters remain constant, $2000s/mm^2$, $SNR = 30 : 1$, component $FA = 0.7$). Beginning with RESTORE, the introduction of crossing fibres begins to demonstrate the deficiencies of the tensor model. While still able to identify corrupted elements with high sensitivity (Figure 3.12d-f), the accuracy of the eliminations becomes increasingly poor as the inter-fibre angle widens. Examining Figure 3.13a-c, we observe that, with a small concession, RESTORE's false positive rates are largely insensitive to the number of corrupted elements or indeed the magnitude of their corruption. Instead, the primary force driving an increase in false positives for RESTORE appears to be the angle of separation. Put simply, this is strong evidence that a significant majority of the RESTORE eliminations are as a result of systematic deficiencies in the tensor model. This is further backed up by Figure 3.10d-f in which we note that even where the correct elements are eliminated, there is no significant reduction in the false positive count.

Comparing HOMOR and RESDORE, for low degrees of corruption (i.e. a single element) it does appear that HOMOR will produce the more accurate eliminations (note the red line, Figure. 3.11 column 1). This is not to say that RESDORE is incapable of detecting the corrupted elements, indeed performance in that respect is roughly equivalent to HOMOR (Fig. 3.12d-f), but rather that due to the previously discussed voxel-specific testing, RESDORE is slightly more likely to introduce a false positive elimination under low corruption conditions (Fig. 3.12a-c, column 1). This situation tends to reverse, however, when the number of corrupted elements is increased. As with the single fibre case, this has the effect of increasing the number of false positives generated by HOMOR while simultaneously decreasing those generated by RESDORE. As a result, the probability of accurate detection through RESDORE appears (relatively) to improve the worse the signal quality becomes – an observation that bodes well for the method's intended purpose.
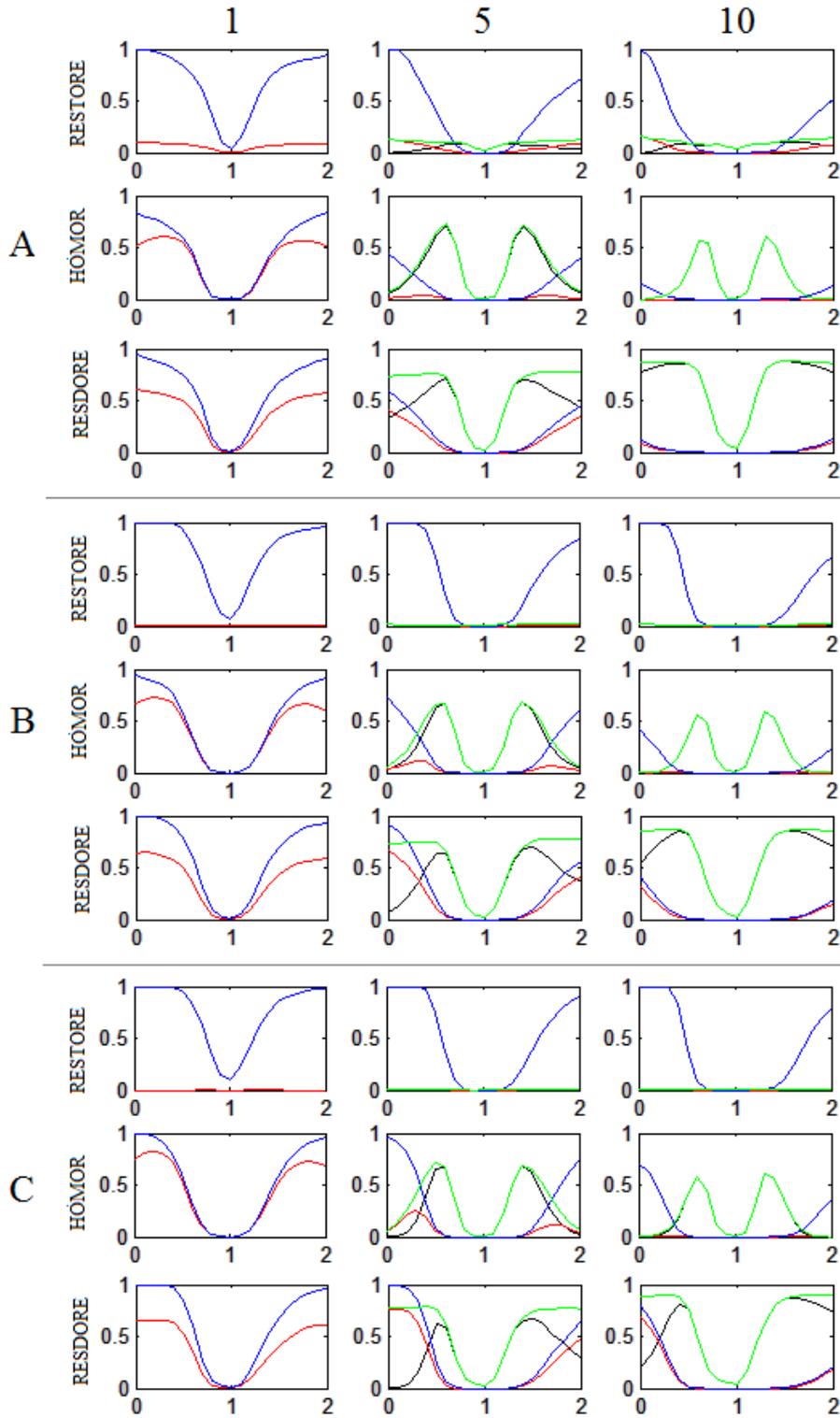
FIGURE 3.11: Single voxel crossing fibre detection results. (A) 30° of separation. (B) 60°. (C) 90°. Colour coded according to table 3.1.
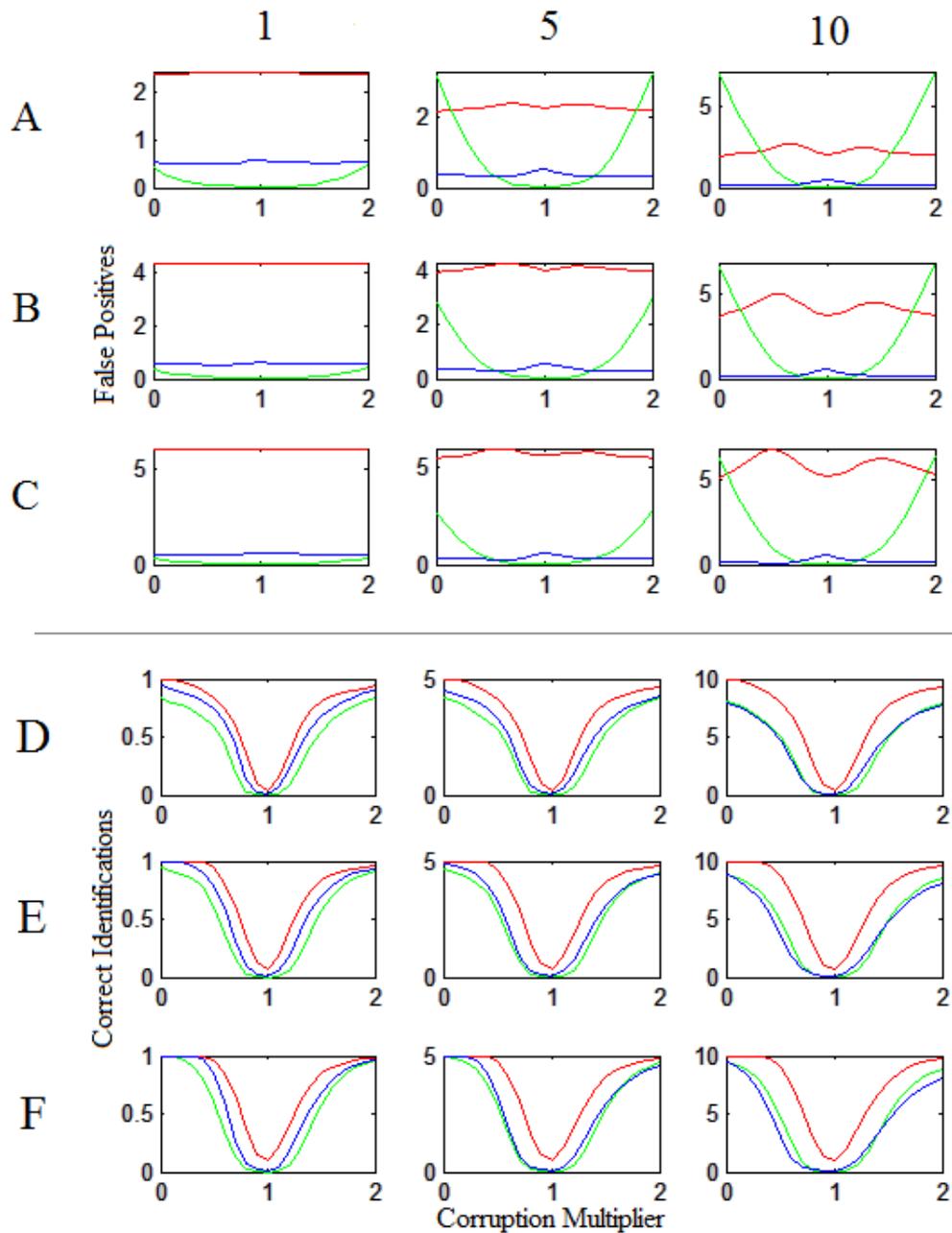
FIGURE 3.12: Single voxel crossing fibre false positives. (A) 30° of separation. (B) 60°. (C) 90°. Single voxel single fibre set recovery. (D) 30°. (E) 60°. (F) 90°. Red - Restore, Green - HOMOR, Blue - RESDORE.

Figures 3.13 and 3.14 repeat the experiment while this time altering the fractional anisotropy of the fibre components from $FA = 0.3$ to $FA = 0.9$. Other parameters remain fixed at $2000s/mm^2$, 30:1 SNR and a 90° angle of intersection. Beginning with Figure 3.13; as might be expected, increasing component anisotropy leads to a decrease in RESTORE accuracy as the signal increasingly deviates from it's expected single fibre response – note that as with the previous crossing angle experiment, RESTORE false positives here are primarily driven by the anisotropy and not the imposed corruption (Fig. 3.14). Comparing HOMOR and RESDORE, as a general trend HOMOR is once again more likely to produce an accurate result given a single corrupted element, though this quickly reverses as the number of elements increases. Examining FA specific traits, while more accurate at lower anisotropies, HOMOR also appears less sensitive – note the profile produced by each line moves from a 'u' shaped profile to a 'v' as anisotropy increases, especially in the single corrupted element case. In comparison, the opposite appears true for RESDORE. While more sensitive than HOMOR at lower anisotropies (Fig. 3.14d), there is a relative drop in true positive eliminations at extremely high anisotropies (Fig 3.14f, component $FA = 0.9$, most apparent in the ten corrupted element simulation) leading to a decline in solution quality (Fig. 3.13c).

Examining the false positives for both techniques; Figure 3.14a-c indicates no significant increase in RESDORE false positives at any anisotropy. HOMOR, on the other hand, appears more likely to produce false positives as the anisotropy increases. If we consider that the angular resolution required to describe a $FA = 0.9$ crossing-fibres signal at $2000s/mm^2$ is likely to be far higher than that of a $FA = 0.3$ signal; and also consider that HOMOR intentionally limits its angular frequencies (to enforce a smoothness constraint) by truncating at the fourth order harmonic, it is plausible that the model is on the limit of its descriptive abilities. Factoring in a combination of residuals caused by descriptive limitations and the previously described fitting drift, it is not unreasonable to expect thresholds to be more frequently exceeded as anisotropy rises, particularly at high b-values that serve to emphasise the difference between a low and high anisotropy signal.
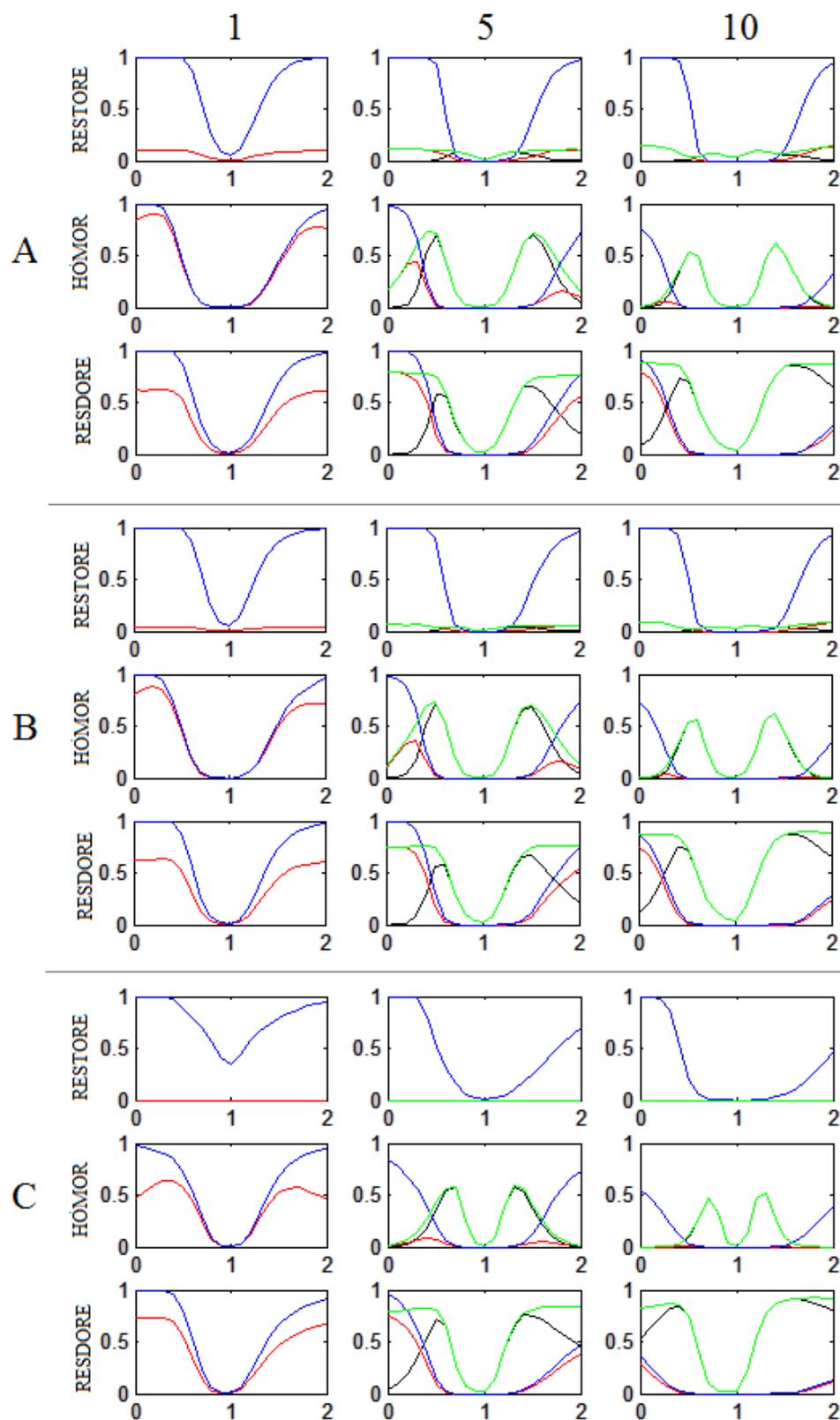
FIGURE 3.13: Single voxel crossing fibre detection results. (A) $FA = 0.3$. (B) $FA = 0.5$. (C) $FA = 0.9$. Colour coded according to table 3.1.
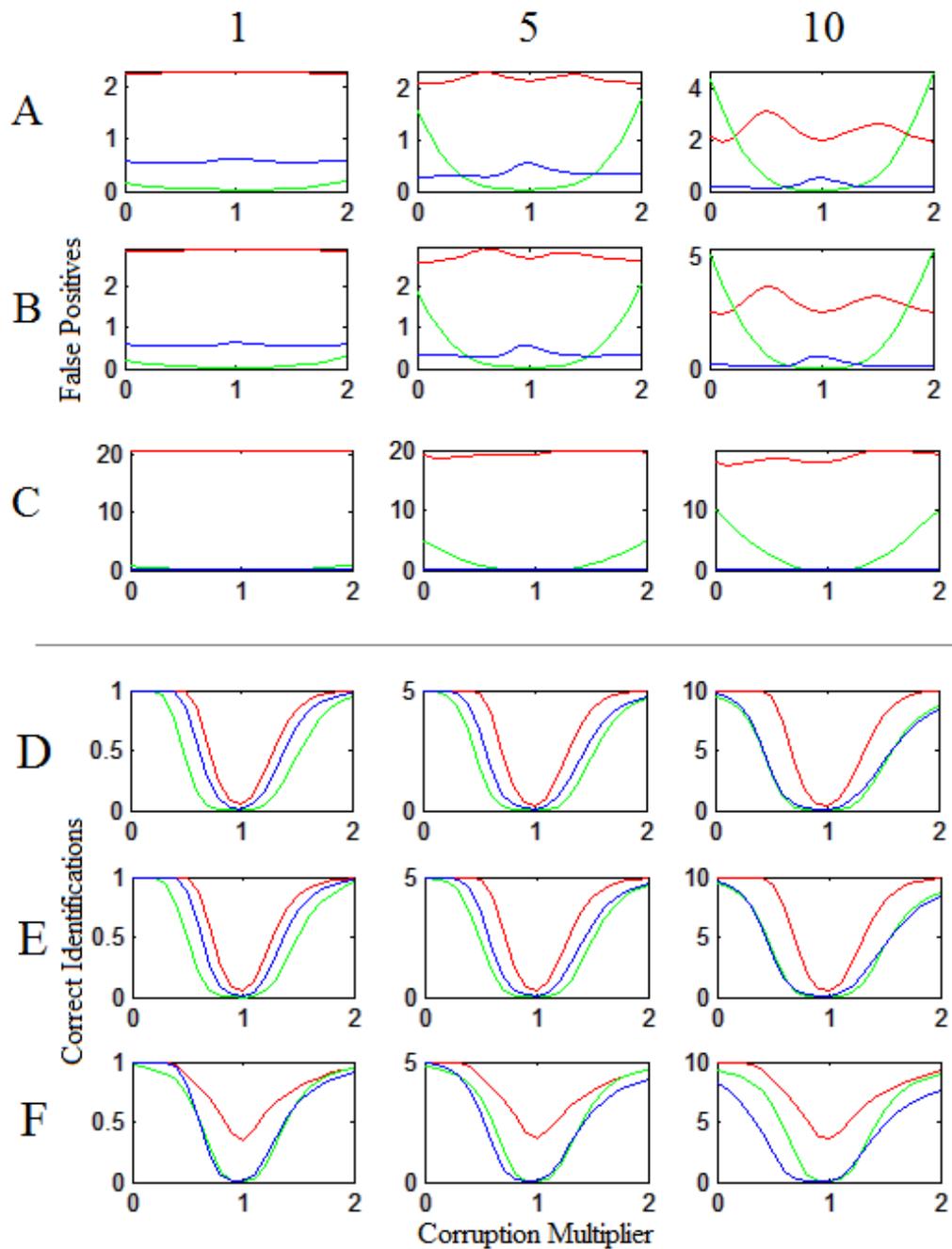
FIGURE 3.14: Single voxel crossing fibre false positives. (A) $FA = 0.3$. (B) $FA = 0.5$. (C) $FA = 0.9$. Single voxel single fibre set recovery. (D) $FA = 0.3$ (E) $FA = 0.5$ (F) $FA = 0.9$. Red - Restore, Green - HOMOR, Blue - RESDORE.

Finally, Figures 3.15 and 3.16 repeat the crossing fibre experiment, this time altering (separately) diffusion weightings ($1000s/mm^2$ and $3000s/mm^2$ and (for later analysis) signal to noise ratios ($SNR = 10 : 1$ and 50:1) with remaining parameters fixed as before ($2000s/mm^2$, $SNR = 30 : 1$, $FA = 0.7$, 90° intersection unless otherwise stated). In order to save figure space, RESTORE results are omitted from Figure 3.15 (though still present in Figure 3.16, allowing for comparison).

As with the single fibre case, a low diffusion weighting (Fig. 3.16a) leads HOMOR to produce large numbers of false positives as the severity of corruption (with respect to magnitude or number of elements) increases. While still able to recover the majority of the corrupted elements (Fig. 3.15a – blue, 3.16e), the prevalent false positives lead to a poor overall result (Fig. 3.15a, red/green). When the diffusion weightings increase the number of false positive detections are significantly reduced (Fig 3.16b), however it then also becomes more difficult for HOMOR to identify the full set of corrupted elements (Fig. 3.16f). Overall, HOMOR is less likely to identify a complete set at higher diffusion weightings but, when doing so, is also less likely to falsely identify other signal elements as corrupted.

In comparison, RESDORE once again produces a low number of false positives across the b-value range (Fig. 3.16a-b). Combined with a corrupted set recovery that retains parity with HOMOR (Fig. 3.16e-f, suffering the same loss at higher diffusion weightings), overall results are generally superior to HOMOR in the high magnitude/high frequency corruption areas (Fig. 3.15a-b), with HOMOR possessing a slight edge in other circumstances - a pattern that appears consistent across the majority of tests thus far.
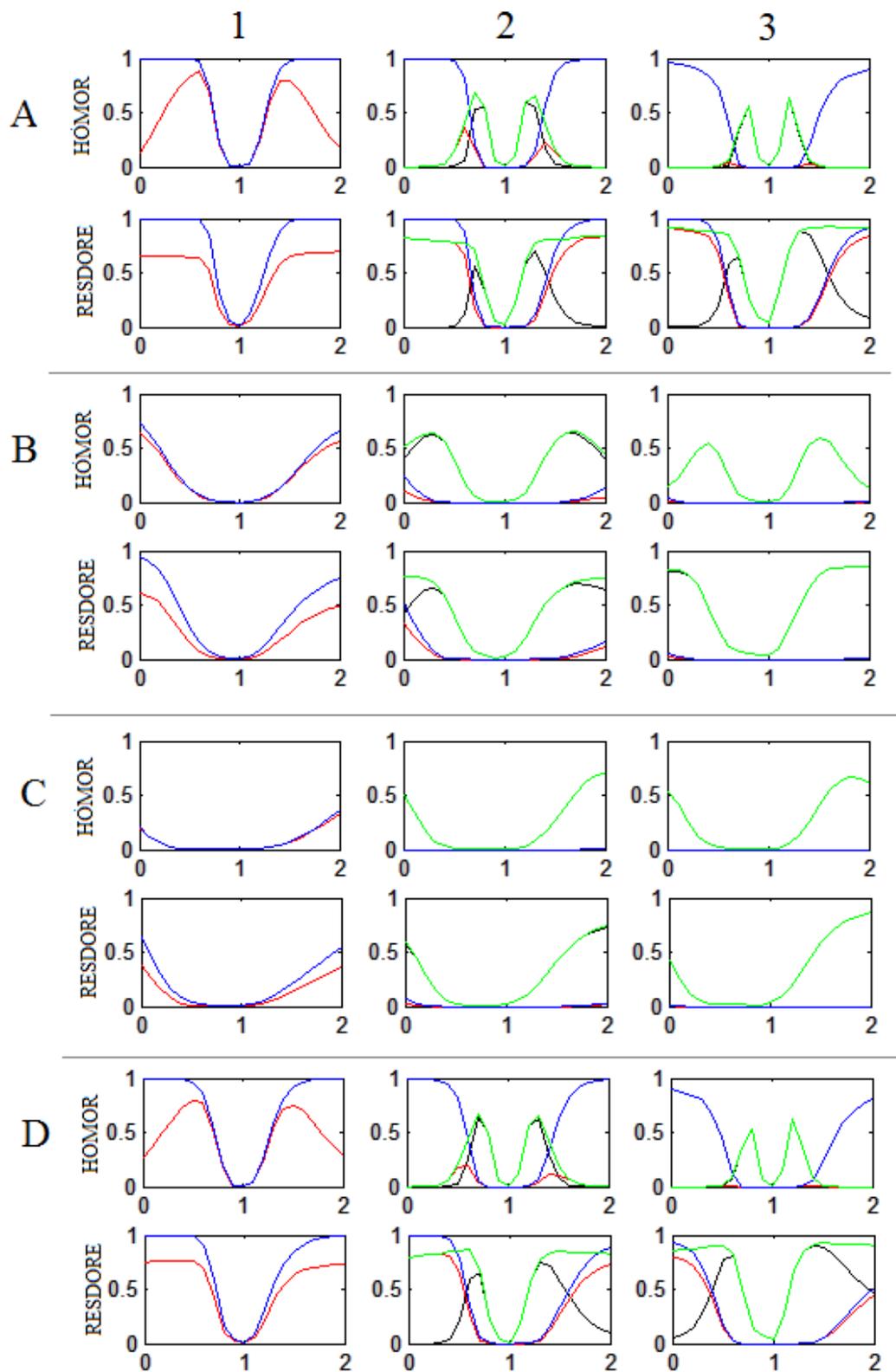
FIGURE 3.15: Single voxel crossing fibre detection results. (A) $1000s/mm^2$ (B) $3000s/mm^2$ (C) $10:1$ SNR (D) $50:1$ SNR. Colour coded according to table 3.1.
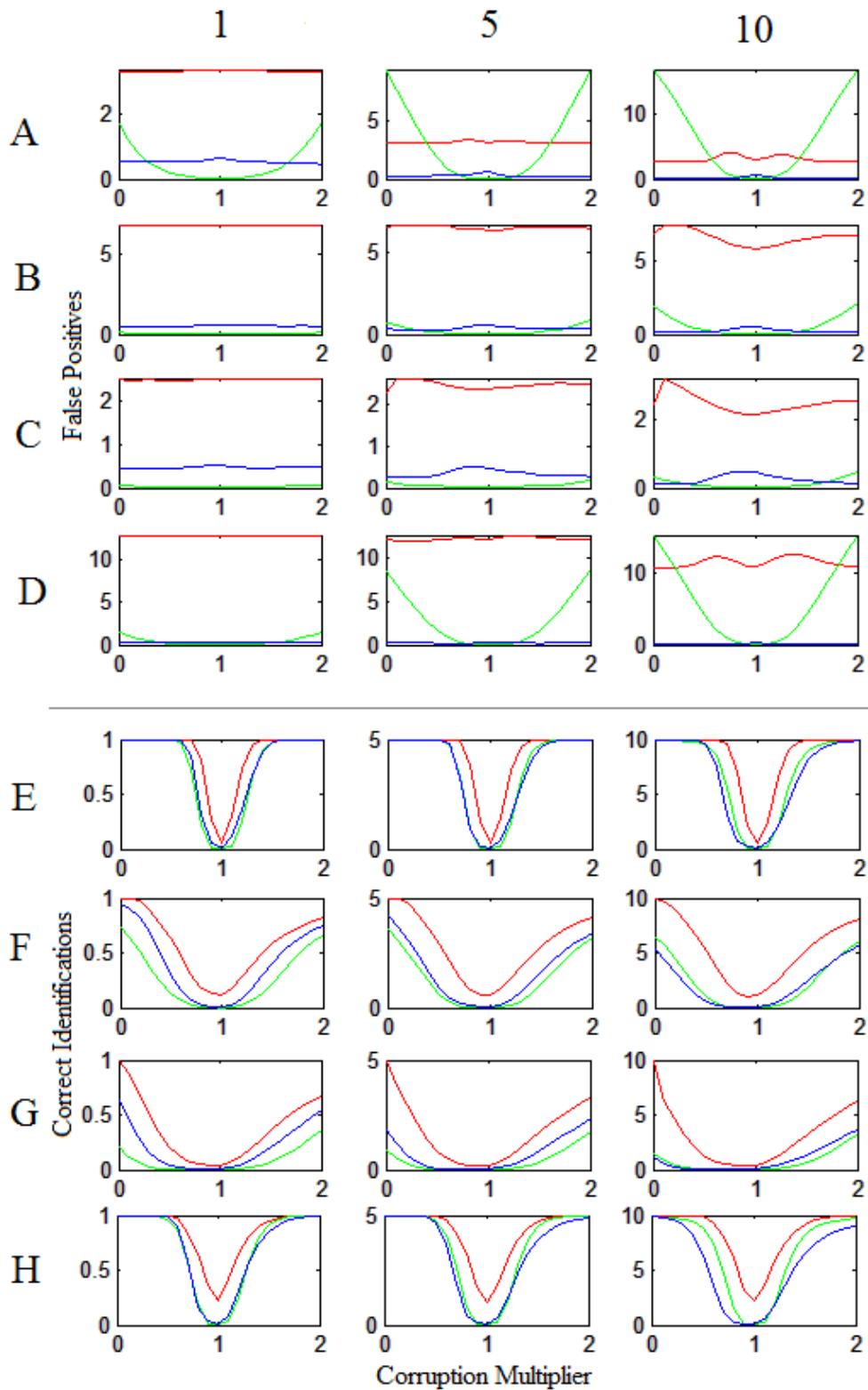
FIGURE 3.16: Single voxel crossing fibre false positives. (A) $1000s/mm^2$ (B) $3000s/mm^2$ (C) $10:1$ SNR (D) $50:1$. Single voxel crossing fibre set recovery (E) $1000s/mm^2$ (F) $3000s/mm^2$ (G) $10:1$ SNR (H) $50:1$ SNR. Green - HOMOR, Blue - RESDORE.

Comparing the three SNR points (a comparable 30:1 SNR result is available in Figure 3.11c and 3.12c/f), all methods observe the (expected) general trend towards increased sensitivity as noise decreases. HOMOR, however, also demonstrates a counter-intuitive increase in false positive production as the SNR increases. The reason for this can likely be explained as an extension of previous arguments. Consider the $3 \times SD_{noise}$ threshold; this is designed to account for noise induced residuals to a spherical harmonic fit that reasonably approximates the fit expected for an equivalent noise/corruption free signal. The problem, however, is that when a signal is corrupted the resultant fit 'drifts' away from the ideal, creating a second source of residuals caused by a poor model fit unrelated to standard signal noise. Where SNR is low, SD_{noise} is naturally high and thus small drift residuals are likely to be trivial in comparison and therefore go unnoticed amongst the background noise (low false positives). However, where SNR is high, SD_{noise} shrinks while the independent drift residuals remain approximately constant. Thus, proportionately, drift residuals play a larger effect at high signal to noise ratios and, since the $3 \times SD_{noise}$ threshold is only designed to account for background noise, elements with large drift could more easily be mis-identified as a corrupted signal element – hence the counter-intuitive increase.

Examining RESDORE, however, increasing SNR actually reduces apparent sensitivity relative to HOMOR for large corruptions (Fig. 3.16h, column 10, note though that in columns 1 and 5 both methods are equally sensitive). One possible explanation for this decrease is that, while producing a large number of false positives at higher degrees of corruption, the HOMOR drift mechanism may also conceivably act to increase sensitivity to true corruption, particularly along the fibre axis which, on empirical visual inspection, is more likely to be recovered through HOMOR than RESDORE (thus creating the relative performance increase).

3.3.4   Combining Simulations

Sections 3.3.2 and 3.3.3 provide, respectively, useful insights into the labelling and elim-
ination characteristics of each of the examined algorithms along with valuable insight
into the mechanisms by which they can fail. To complete the analysis, the logical final
step is to test the performance of combined labelling and elimination algorithms – that
is to say, label test data according to the method specific labelling algorithm and then
process the positively labelled (i.e. suspected of corruption) data through corresponding
eliminations processes. Due to the nature of the labelling response curves produced by
Section 3.3.3, we can already make some assumptions about the way in which the results
of the complete (two stage) implementations will differ from results in Section 3.3.3. In
the majority of cases, signals undergoing severe corruption (i.e. multiples below -0.5 or
above 0.5) are labelled with near 100% likelihood and thus, across any reasonable sample
size, the vast majority would be forwarded to elimination steps. In this case, the results
of those eliminations would be little/no different from Section 3.3.3 in which we already
process all data as if it had been positively labelled. The difference, if any, will come
in the low magnitude corruption areas (i.e. [-0.5 0.5] multiples) where the likelihood of
detection is not 100%.

To explore this possibility, we will briefly examine the performance of the combined
labelling and elimination algorithms against two basic fibre configurations. Figure 3.17
presents results (averages across positively labelled voxels *only*) against a single fibre
configuration with $b = 2000 s/mm^2$, $SNR = 30 : 1$ and a component anisotropy of
0.7, while Figure 3.18 presents similar information for a comparable 90° crossing fibre
configuration (all other parameters remain identical). Beginning with overall elimination
results (Figures 3.17 and 3.18a), when compared to equivalent Section 3.3.3 results
(Figures 3.9b and 3.13c respectively), the addition of the labelling process results in
an apparent increase in sensitivity to outliers across most corruption multipliers. The
reason for this is fairly simple; as the results of Section 3.3.2 show, labelling methods
are less sensitive to corruption along gradients aligned with, or near to, minima in the
apparent diffusion coefficient. As a result of this orientationally variant sensitivity, the
set of voxels that are positively labelled will be biased towards corruption aligned with
ADC maxima which, as covered in Section 3.3.3, is easier for the elimination algorithms
to identify – resulting in the apparent increases in sensitivity (since we average results
across those voxels passes by the labelling phase). The problem, however, is that the
less readily detectable minima-aligned corruption goes untested by the elimination step
and, as such, will remain in the data as a source of bias.
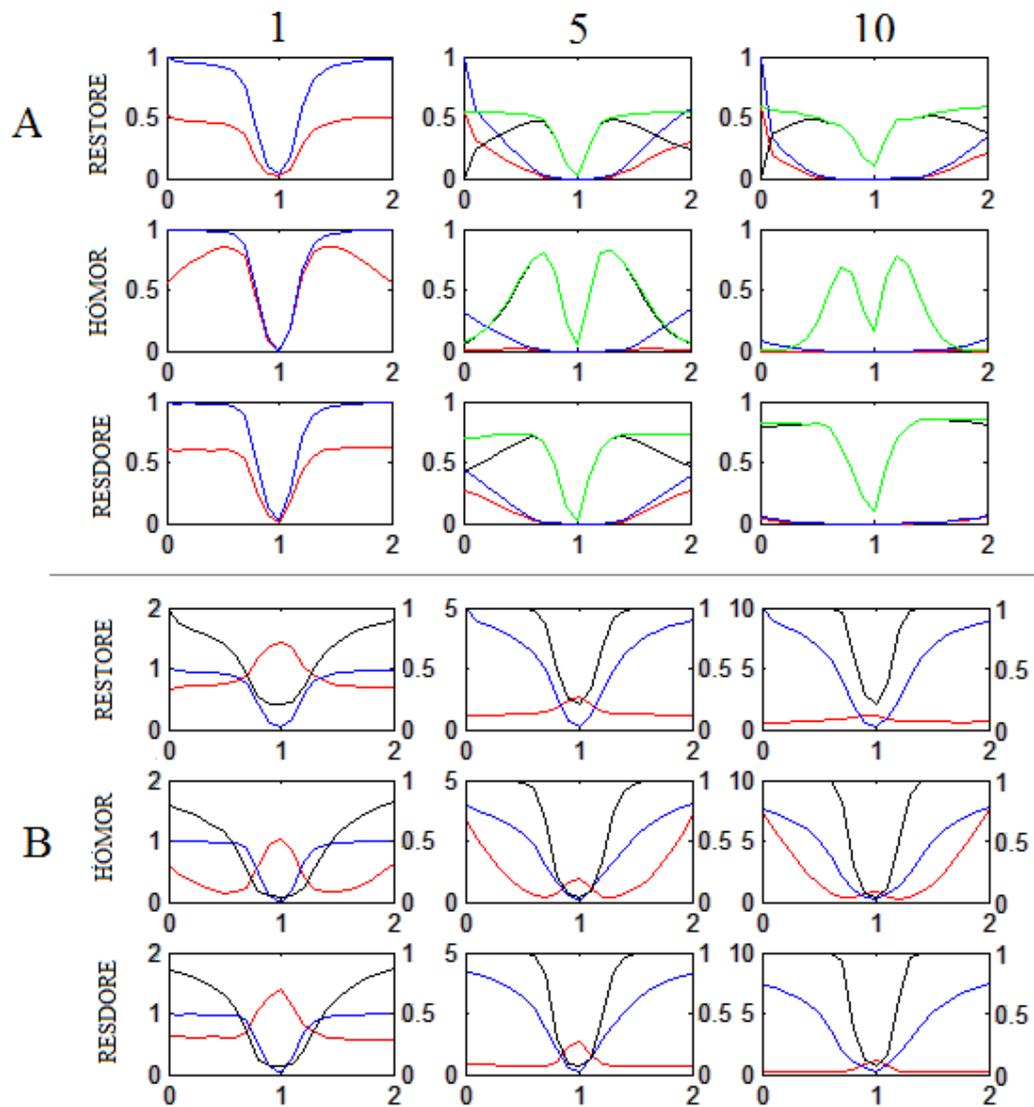
FIGURE 3.17: (A) Single voxel combined single fibre detection results, colour coded according to co Colour coded according to Table 3.1. (B) Red: false positive detections, scaled to left axis. Blue: Number of corrupted elements correctly identified, scaled to left axis. Black: signal labelling profile, scaled to right axis.
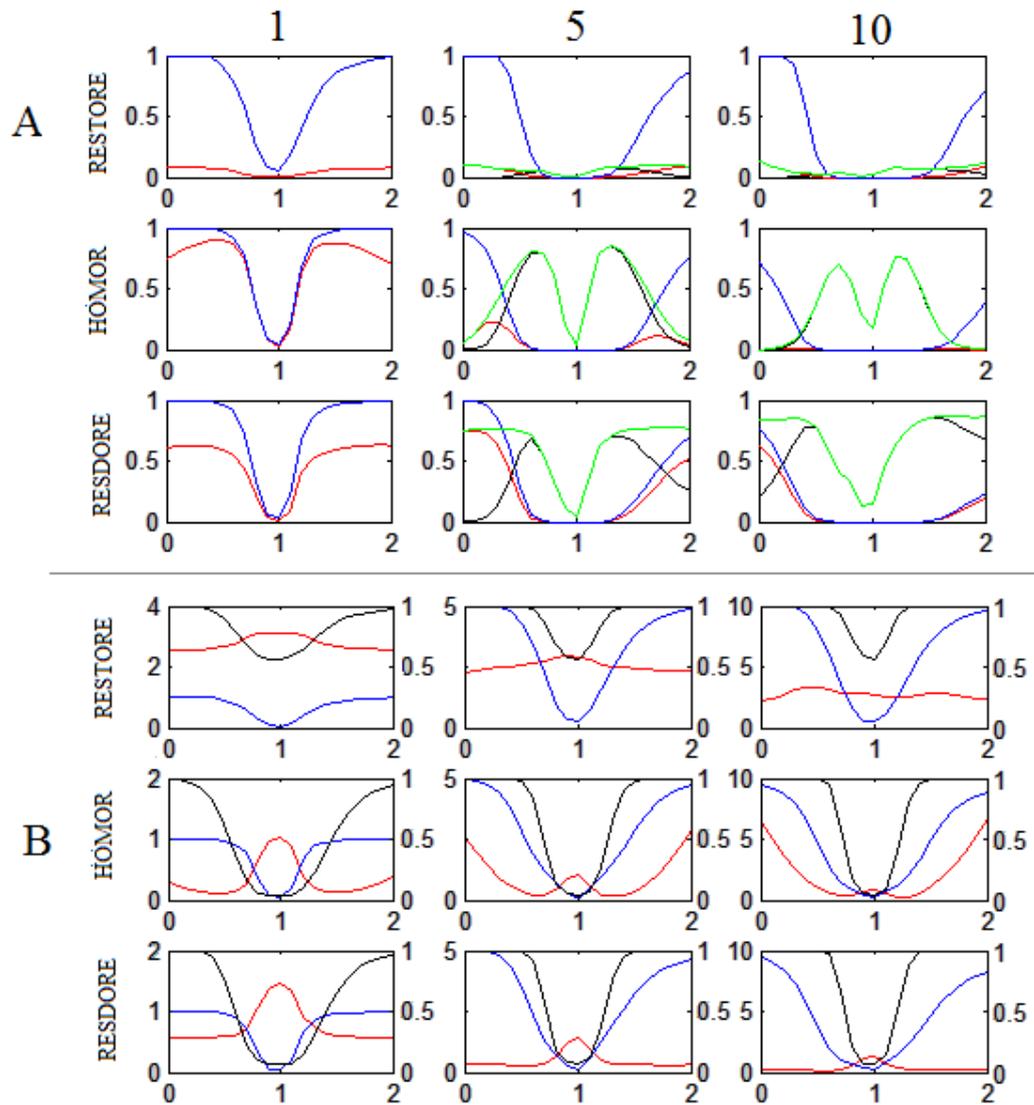
FIGURE 3.18: Single voxel crossing fibre false positives. (A) $1000s/mm^2$ (B) $3000s/mm^2$ (E) $10:1$ SNR (F) $50:1$. Single voxel crossing fibre set recovery (C) $1000s/mm^2$ (D) $3000s/mm^2$ (G) $10:1$ SNR (H) $50:1$ SNR. Green - HOMOR, Blue - RESDORE.

Moving on to the false positive and set completion results (Figures 3.17b and 3.18b), we observe similar patterns to those described in the previous paragraph. Beginning with the identification of corrupted elements (blue line, scaled to left axis), results again display an apparent improvement in sensitivity compared to Section 3.3.3 equivalents (Figures 3.10e and 3.12f respectively). The reason, as before, is simply due to bias introduced by the labelling phase. Those signals with more corrupted elements oriented along ADC maxima are more likely to be positively labelled and, subsequently, will provide an 'easier' dataset for elimination step to correctly process (again, a bias introduced by averaging results across the positively labelled set). False positive results (red line, scaled to left axis) are slightly more interesting; notice that at all methods now produce a spike in false positives around the $1\times$ corruption multiple (i.e. no corruption) – something that was only the case for RESDORE in Section 3.3.3 (Figures 3.10b and 3.12c). Taking the crossing fibre simulation as an example, we observe that (on average) at the $1\times$ multiple, HOMOR now produces 1.05 false positives per elimination tested voxel while RESDORE produces 1.40, representing an increase of ≈ 1 (compared to Section 3.3.3) in both cases. The reason is again linked to the signal labelling profile (black line, scaled to right axis). Note that at the $1\times$ multiple very few signals are being positively identified as containing corruption. Those signals that were labelled are likely to have been the result of some model fitting error resulting in one or more elements with residuals $\geq 3 \times SD_{noise}$. If we then pass this small set of voxels to the elimination algorithms they will most likely identify the outlying elements and flag them as corrupt, generating the observed false positive identifications. In reality this is of course a very minor issue, returning to Section 3.3.4 (Figures 3.10b and 3.12c) we see that model fit related issues (at least at the $1\times$ multiple) have virtually no effect. Again, the apparent increase in false positives is simply as a result of bias introduced by averaging over the output of the labelling phase.

While little genuinely new information is presented in this section, bringing together the labelling and elimination phases in this way does serve to highlight one key point; short of exhaustively testing every signal, the output of any elimination algorithms is strongly influenced by the performance of the chosen labelling method. Regardless of elimination algorithm performance, if the labelling step cannot see a particular form of corruption then the elimination algorithm will not have a chance to remove it. This is particularly important to both RESTORE and RESDORE since, unlike HOMOR, labelling and elimination are not integrally connected – one can theoretically pick and choose labelling methods depending on the target data (This is not the case for the stock implementation of HOMOR since the labelling phase is essentially identical to the first iteration of the elimination).

For further reference, a limited number of additional combined results are presented in Appendix A.

3.3.5   Single Voxel Simulation Overview

Beginning with RESTORE, practical performance is largely determined by diffusion weighting and, to a lesser extent, the target tissue. The reason we have not included the presence of crossing fibres/crossing angle in this list is that, recalling Jeurissen et al. (2013), crossing fibres comprise a large portion of white matter voxels. In practical terms, therefore, avoiding crossing fibre is simply impossible (unless targeting very specific areas/structures) and thus the only conditions under which RESTORE can be expected to function reliably are those in which both the diffusion weighting and component anisotropy of the target tissue are low – i.e. where the differences between a single and crossing fibre signal are minimised, allowing the diffusion tensor model to provide a 'reasonable' fit. As shown by Figures 3.8 and 3.10/3.12, where this is not the case RESTORE is highly likely to both falsely flag a voxel as containing outliers and then falsely eliminate a large portion of that voxels signal elements in order to force-fit an inadequate model. Where those constraints do fit, however, performance is reasonably competitive, offering high sensitivity to the corrupted elements and generally adequate accuracy.

In comparison, neither HOMOR nor RESDORE are limited in this manner, both having a basis in high angular resolution methods. Making a choice between the two methods is likely to be driven by the quality of the target data. A common theme throughout Section 3.3.3 and 3.3.4 is that, where corruption is light (particularly in terms of magnitude), HOMOR is likely to be a reasonable choice. As corruption increases, however, RESDORE begins to take the lead, being on average more likely to detect the full corrupted set and less likely to highlight false positives.

3.4   In Vivo Application

With thanks to Dr Pannek, we were able to acquire a portion of the data used in the HOMOR publication (Pannek et al. (2012)) for comparison of the two methods. More specifically, we were provided access to the $n = 19$ dual $1000s/mm^2/3000s/mm^2$ images which, paraphrasing the relevant text descriptions, include 18 participants of which 11 were male, all aged between 18 and 65 years (mean 46) with 10 suffering from mild/moderate traumatic brain injury (TBI), 4 with incidental findings and 4 normal healthy controls (the 19th image appears to be a repeat of one of the TBI subjects, there is severe slice-wide drop-out in the first image that may explain the second acquisition). All images were acquired along 64 diffusion weighted directions (Jones et al. (1999)), 1 b0 image, 2.5mm isotropic resolution and TR/TE separately optimised for each diffusion weighting. Images were then corrected for susceptibility distortions through field map un-warping (FSL FUGUE, Jenkinson et al. (2012)) and motion corrected using the ExploreDTI package (Leemans et al. (2009), a step that

differs from the original HOMOR publication which used FMAM, Bai and Alexander (2008), to achieve a similar result) with appropriate b-matrix rotations (Leemans and Jones (2009a)) and N3 spatial inhomogeneity corrections (Sled et al. (1998)). Finally, the mean diffusion weighted intensity was then normalised across all participants.

Without access to the larger 99 subject corpus used by Pannek et al. (2012) to generate their template image, or sufficient normal/healthy data to create one, the MNI152 space was used as a stand-in. Co-registration was achieved by first applying an affine warp (FSL FLIRT, Jenkinson et al. (2012)) to each B0 image in order to achieve general positioning over then template (using normalised mutual information cost measures), followed by a non-linear registration (FSL FNIRT) for improved correspondence. All subsequent co-registrations were achieved by applying the learned affine translation and warp-field to the image data of interest.

3.4.1 Voxel Labelling

As with the simulation section, we must first begin with a brief examination of voxel labelling results. Using the methods described by Walker et al. (2011), estimates for SD_{noise} may be calculated as:

$$SD_{noise} = \sqrt{\alpha(v) \times median(S_{ijk}^2)} \tag{3.4}$$

where S_{ijk}^2 is, for the region of interest, the ratio of the residual sum of squares to the model fit to the degrees of freedom provided by that model and the target data; and $\alpha(v)$ is a scaling parameter designed to account for the use of the median instead of the mean, though in practice this is negligibly different from 1 and as such ignored.

In their study, Pannek et al. (2012) demonstrated that, even in corruption free data, the probability of voxel undergoing elimination varies according to the position of that voxel within the brain. One potential reason for this observation is that since the brain is not homogeneous, variations in fibre geometry (or indeed, lack of fibre geometry in grey matter) could, as shown in Section 3.3.2, affect the likelihood of voxel labelling which, as the precursor to elimination, would of course bias the end result. To assess this possibility, Figure 3.19 displays – for each labelling method and b-value – voxel labelling probability maps produced by averaging the binary labelling maps resulting from the application of labelling each method, as outlined in Section 3.3.2, to the 19 image corpus with method and image specific SD_{noise} estimates calculated according to Equation 3.4. As expected, there is indeed a spatial variation of labelling probabilities, with increased likelihood observed within voxels containing white matter (particularly complex white-matter, note that in the red-labelled region containing crossing fibres, RESTORE

displays a high probability of rejection, where as HOMOR/RESDORE probabilities, being non-Gaussian methods, are comparatively lower) and across the lower/mid brain in general. While variations in geometry might explain the bias towards labelling complex white matter, reasons for the secondary bias towards increased labelling of lower/mid brain regions are less clear and thus will be discussed, after further examination, in the latter half of Section 3.4.2.
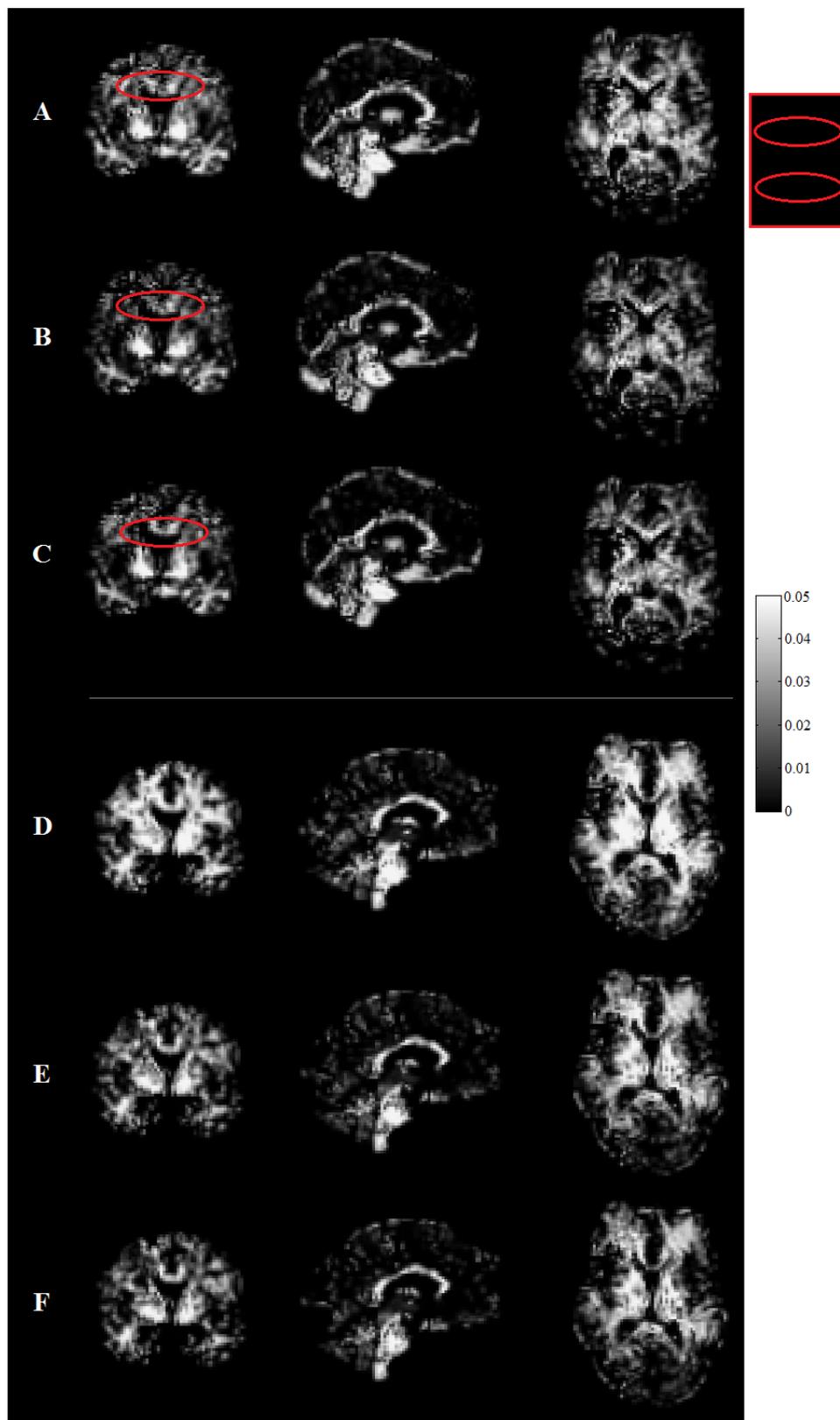
FIGURE 3.19: Voxel labelling probability maps. A-C: $1000s/mm^2$ RESTORE, HOMOR and RESDORE respectively. D-F: $3000s/mm^2$ RESTORE, HOMOR and RESDORE.

3.4.2   Element Elimination

In reality, the positional variations in element elimination are of course not solely dependant on variations in voxel labelling, as shown by Section 3.3.3, the elimination step itself may also be affected by properties of the target signal. It is therefore, perhaps, more sensible to look at the overall result (combined labelling and elimination). This information will be presented in two ways. Firstly, for comparison with Pannek et al. (2012), Figure. 3.20 displays outlier reject probability (ORP) maps (Walker et al. (2011), essentially the mean number of rejected signal elements normalised by the degrees of freedom of the model used) for both the RESTORE and HOMOR methods. RESDORE is omitted at this stage because (1) there would be no comparable reference image and (2) the number of degrees of freedom to the Richardson-Lucy deconvolution is not static (it varies with, for example, current iteration count White and Hanisch (1993)) which makes production of an ORP map (normalised by the degrees of freedom) a non-trivial undertaking – comparison between RESDORE and RESTORE/HOMOR will take another form to be explained shortly.

Examination of Figure 3.20 reveals evidence of patterns suggested by the previous sections simulations. Despite the similarities in labelling maps (Fig. 3.19), there are pronounced dissimilarities between HOMOR and RESTORE when it comes to elimination. At all diffusion weightings, signals processed through RESTORE will have the larger probability of element rejection. The key point, however, is that at low diffusion weightings (Fig. 3.20a-b), this difference is relatively small (as all simulations to this point suggested) while as the diffusion weighting increases, the differences become significantly more apparent, particularly in regions of crossing fibre (for example the areas circled red) – an observation matching those of Pannek. Examining more general trends, with both methods there exists a high probability of rejection across the lower surfaces of the forebrain and within the cerebellum with reductions as one moves upwards (note there is also a rim across the upper surface, though this is largely due to partial volume with untrimmed cerebral spinal fluid – an edge-case artefact of the motion/distortion correction process). Again, these observations are comparable with those made by Pannek et al. (2012) and, while our results may appear noisier than those presented in the HOMOR publication (largely owing to a 19 image average in comparison with 100 simulated whole volume datasets), this suggests that our implementation of each algorithm is behaving as expected. One interesting point for further exploration, however, is that both RESTORE and HOMOR possess an increased probability of detection at higher diffusion weightings. For RESTORE this is understandable (the tensor model is no longer appropriate), but for HOMOR, simulations suggest that increased diffusion weightings correlate with decreases in both true positive (Fig. 3.15a-b) and false positive (Fig. 3.16a-b) elimination likelihoods, two factors that would suggest a decrease in ORP at higher diffusion weightings, rather than the observed increase. In order to understand what is occurring, a more detailed examination of the data will be required.
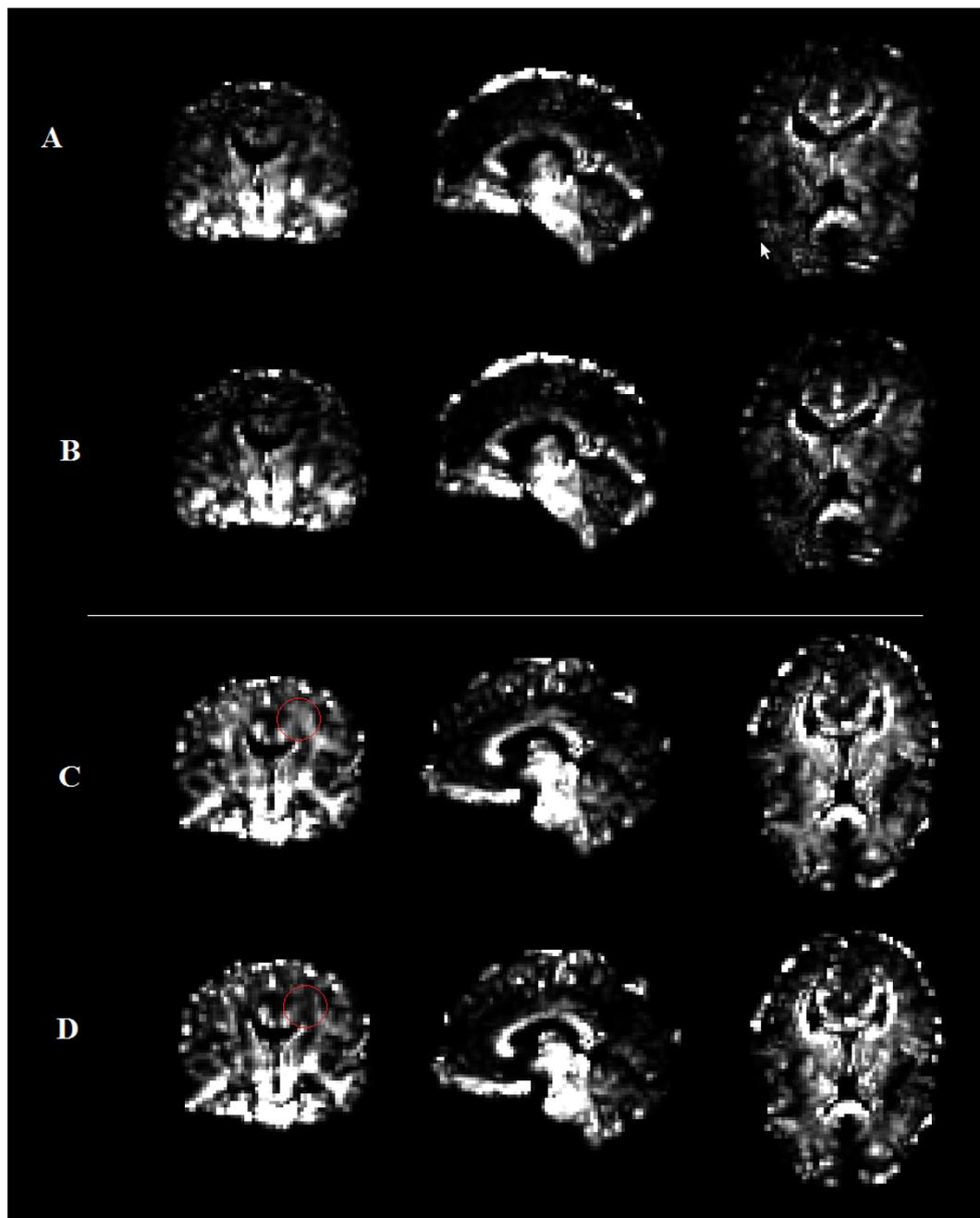
FIGURE 3.20: ORP maps. (A) $1000s/mm^2$ RESTORE. (B) $1000s/mm^2$ HOMOR. (C) $3000s/mm^2$ RESTORE. (D). $3000s/mm^2$ HOMOR. Red circle indicates an area of crossing fibre.
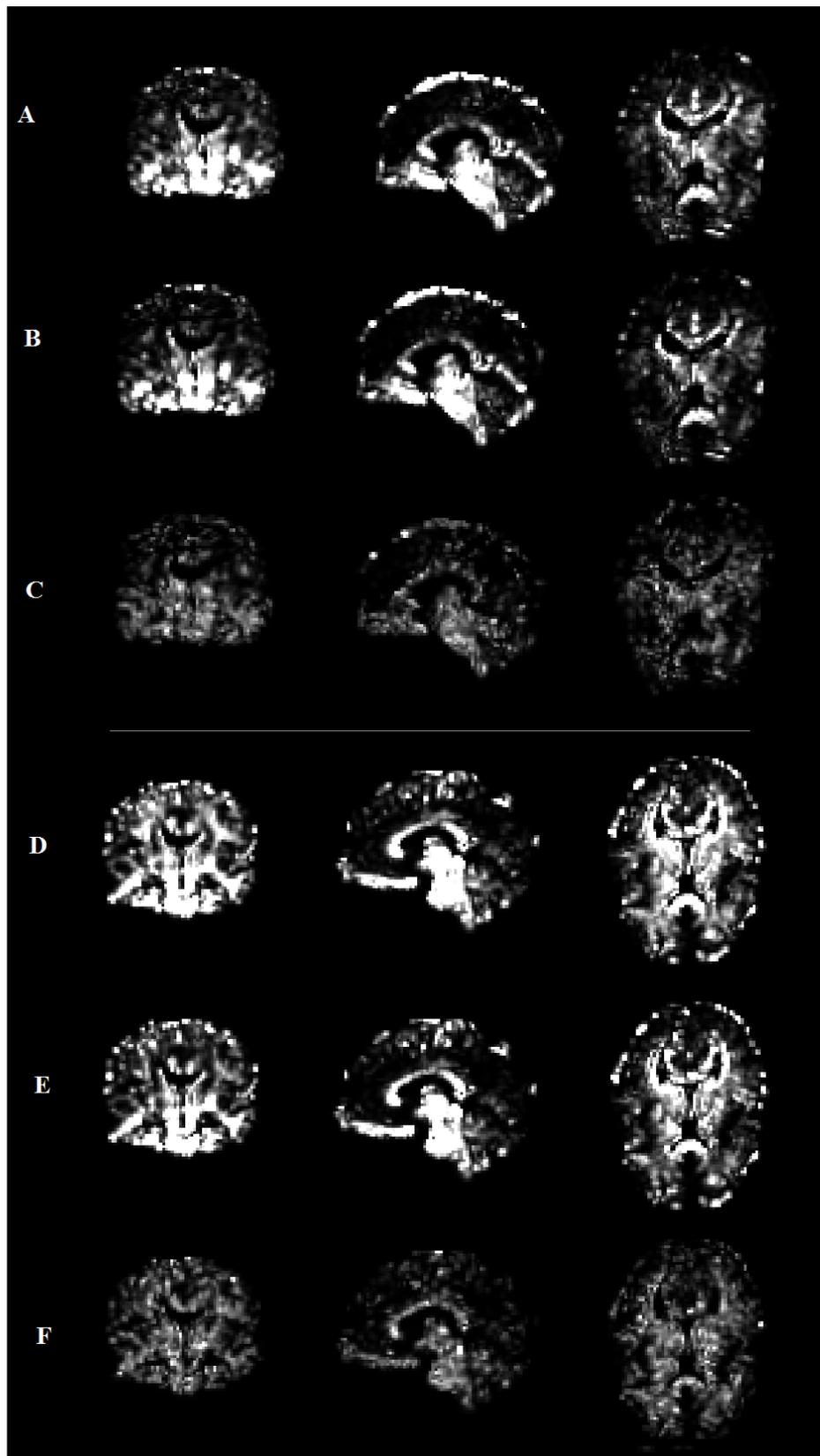
FIGURE 3.21: Maps of mean element eliminations per voxel. (A) $1000s/mm^2$ RE-STORE, (B) HOMOR, (C) RESDORE. (D) $3000s/mm^2$ RESTORE, (E) HOMOR, (F) RESDORE.

Figure 3.21 provides mean elimination maps for each of the three techniques. Unlike the previous ORP maps, these are not normalised by the degrees of freedom of the method used. Our arguments for this are twofold: firstly, and as previously discussed, due to the manner in which RESDORE processes signals the number of degrees of freedom is difficult to ascertain and thus ORP maps impractical to construct. Secondly, and most importantly, we believe the complexity of the model used is irrelevant to the end result; either a signal element is eliminated or it is not; any subsequent processing will simply be presented with the remaining set of signal elements and, as discussed by Chang et al. (2012) (and also Jones (2004), who discusses the minimum elements required for rotational invariance), eliminating too many elements – regardless of the method used – can result in signal bias just as easily as eliminating too few. It is thus only important that whatever technique is used eliminates the correct signal elements and therefore a non-normalised comparison will be more relevant to real-world performance.

Examining Figure 3.21, two points are easily seen: (1) The previously described relationships between HOMOR and RESTORE results is further reinforced in the absence of ORP normalisation (note, for example, that the greater difference in contrast in crossing fibre areas at high b-value). (2) RESDORE results, while comparable (to HOMOR) in midbrain regions, possess significantly lower image map intensity across in low and forebrain regions or areas typically affected by partial volume with CSF (necessitating the observed truncation of dynamic range causing lower regions of the RESTORE/HOMOR result to appear white).

The reason for the stark differences between RESDORE and RESTORE/HOMOR is down to the mechanisms by which each algorithm responds to signals of abnormal intensity, localised variations in SNR or very severe corruptions (beyond the 10 elements depicted in the previous sections simulations). To demonstrate what this means in practice, it is perhaps easiest to briefly study a small set of individual image slices. To begin, Figure 3.22a provides a directionally encoded colour map for a midbrain slice through an individual $b = 1000s/mm^2$ image. The abnormal green colouration (principal diffusion vector oriented along the x axis) of this image indicates the presence of at least one slice-wide artefact which, through visual inspection of the images (a method for using RESDORE to achieve the same task will be detailed in the following section) was identified and is displayed (with normal intensity images either side) in Figure 3.22b. Placed in the context of Section 3.3.3, this drop-out represents, on average, a single corrupted signal element with a $0.35\times$ intensity multiplier which, according to simulation results (Figure 3.14 being particularly relevant), should be well within the ability of both HOMOR and RESDORE to resolve accurately with minimal false positives (we will exclude RESTORE in these comparisons since, for reasons that will shortly be explained, it behaves in a manner very similar to HOMOR under the conditions we wish to explore).
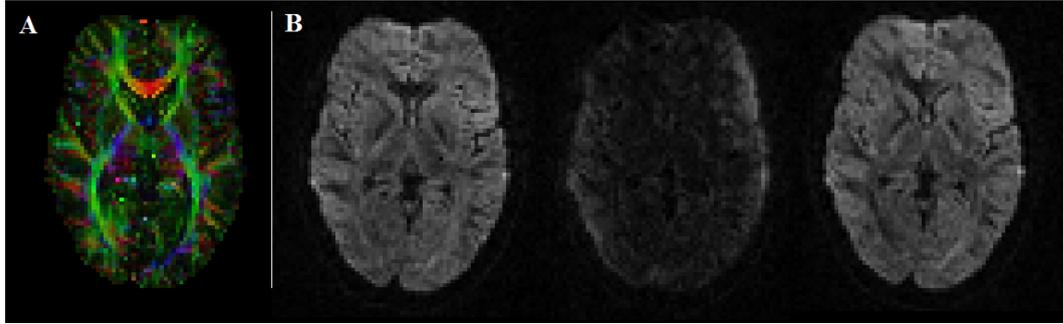
FIGURE 3.22: Slice wide drop out example. (A) Slice DEC map. (B) Left and right: images of correct intensity. Middle: image experiencing signal drop-out.

Using the HOMOR labelling method as described in Section 3.3.2 (with SD_{noise} estimates derived from the whole volume using the method described in Equation 3.4), Figure 3.23 provides voxel labelling results for (a): the target slice with the corrupted gradient image included and (b): the target slice with the corrupted image manually removed. Note that in Figure 3.23a, the majority of the slice is labelled as containing an outlier (as expected) while in 3.23b (with the main source of outliers removed) these detections are both greatly reduced in number. If we then calculate the difference between these two images (Figure. 3.23c), the highlighted voxels should be singularly corrupted when the errant gradient image is present and nominally uncorrupted (though still noisy) when that image is removed. Figure 3.24 displays the outcomes of both HOMOR and RESDORE element eliminations across the (Figure. 3.23c) highlighted voxels both with and without the corrupted gradient image. As expected, with a single corrupted element the majority of tested voxels for both HOMOR and RESDORE indicate a single corrupted element (averaging 1.18 and 1.27 eliminations per voxel respectively). Following removal of the corrupted image, HOMOR detects no outliers (again, this is expected since the labelling method is identical to the first HOMOR iteration and thus if nothing is labelled, nothing would be eliminated) while RESDORE, now applied to essentially uncorrupted data, detects outliers with a mean of 0.34 elements per voxel (recalling Section 3.3.3, this is close to the Empirical rule predicted performance under such conditions).

The purpose of the previous experiment was to examine the performance of both algorithms under optimal conditions and against a minimally corrupted target slice. As the simulations predicted, under these circumstances the performance of both algorithms is relatively similar which, returning to Figure 3.21, explains the similarity within mid-brain regions of the image. The question, then, is why is there such a difference in intensity across the lower brain? Part of the answer can be found in Figure 3.25 which plots: (a) slice wise SNR estimates (Calculated according to Dietrich et al. (2007)), (b) the number of (brain containing) voxels per slice, (c) the fraction of (brain containing) voxels, per slice, labelled as outliers and (d) the mean number of elements eliminated pet labelled voxel per slice using the HOMOR method. What this Figure shows is that
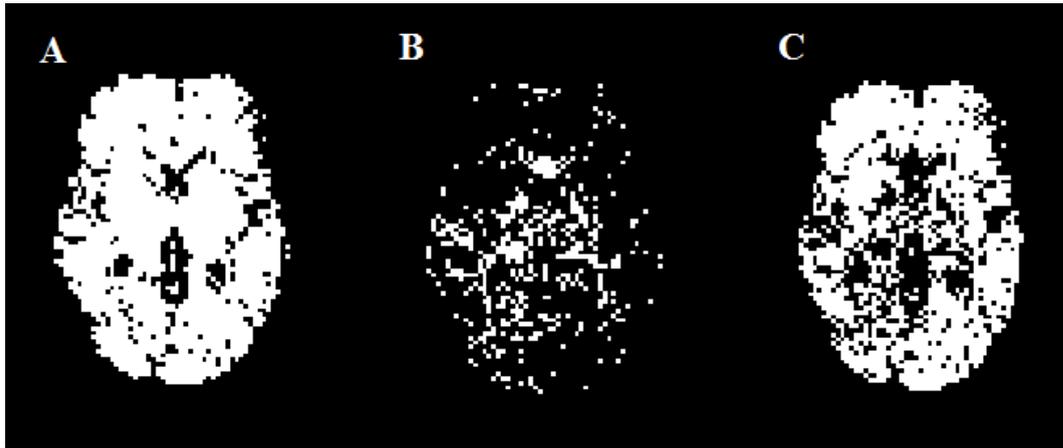
FIGURE 3.23: Slice wide drop out example. (A) Labelling with drop out included. (B). Labelling with drop out manually excluded. (C) A-B Difference map.
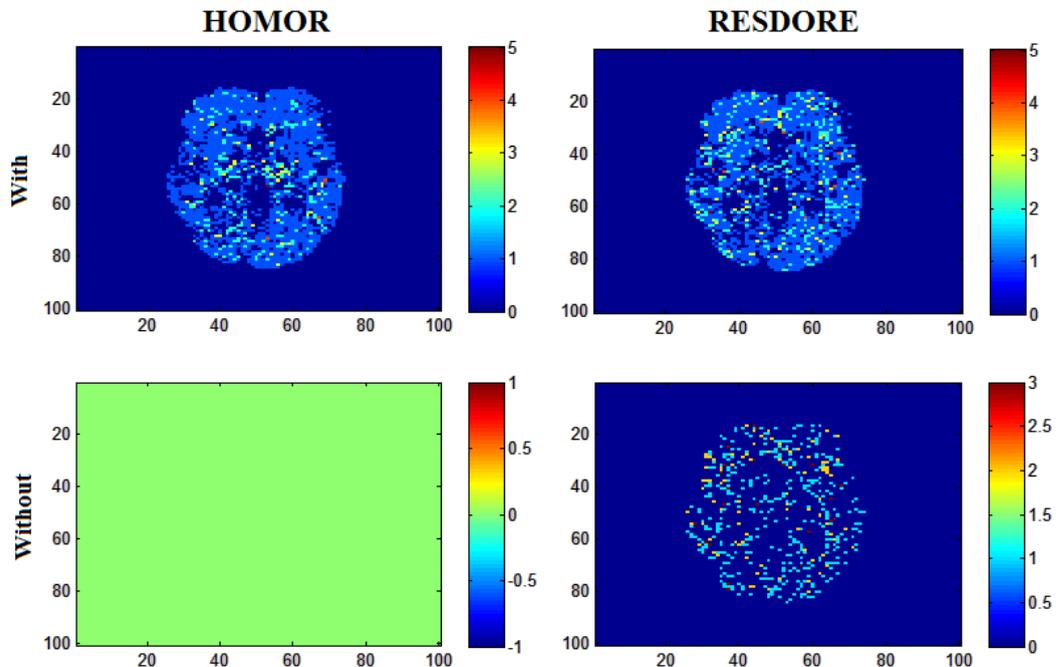
FIGURE 3.24: Slice wide drop out example, elimination results. Left HOMOR, right RESDORE. Upper results with drop out included, lower with drop out excluded.

SNR, as is well understood, does not remain constant throughout the imaged volume. The implication for HOMOR (and since it uses a similar method, RESTORE) is that since SD_{noise} is inversely linked to SNR, one volume wide estimate for SD_{noise} (as stipulated by Pannek et al. (2012)) will not sufficiently characterise the entire volume. Looking at Figure 2.25b, the majority of brain-containing voxels lie within the central (midbrain), high SNR region of the volume. Given the method by which SD_{noise} is then calculated (using an amalgam of residuals from the entire volume), the resulting estimate will naturally be biased by the most frequently occurring residuals which, according to Figures 3.25a and b, will be generated by the high SNR midbrain regions. As we proceed downwards through the image and into the lower SNR regions, noisy – yet

not necessarily corrupt - signal elements will increasingly fall foul of the restrictive high-SNR biased SD_{noise} estimate and this, in theory, would result in increasing numbers of false positive eliminations. If we then examine Figures 3.25c and d, this is exactly what appears to be happening – decreased SNR correlates with an increase in both the proportion of voxels labelled as containing outliers and, from those voxels, an increase in the mean number of rejected elements. Unfortunately, while certainly contributing to the differing intensities, this is only part of a larger mechanism. Paying closer attention to the upper and lower extremes of the image we see that according to Figure 3.25a they have approximately the same SNR but, at the lower extreme, a larger proportion of the voxels are identified as containing outliers, while at the higher extreme those that are identified seem to produce (on average) more eliminations per voxel. To understand why this should be the case, we will need to examine individual image slices.
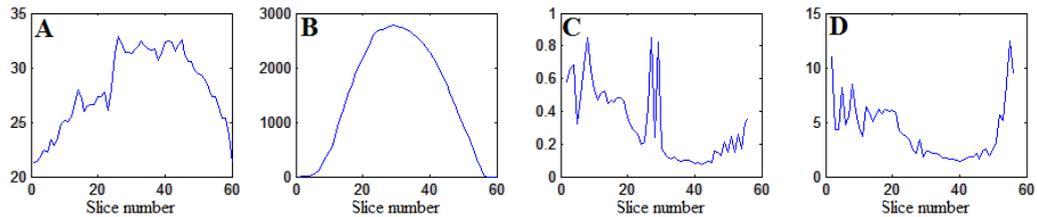
FIGURE 3.25: (A) Slice wise SNR estimates. (B) Number of brain-containing voxels per slice. (C) Slice wise SD_{noise} estimates. (D) Mean number of rejections (per labelled voxel) per slice.

Before proceeding to that examination, however, it would first make sense to demonstrate how RESDORE responds under the same conditions. Figure 3.26 contains this information (for later reference, we also overlay the Figure 3.25d HOMOR result and an equivalent HOMOR result with slice-wise SD_{noise} estimation). Note that with RESDORE, the mean detections per labelled voxel (Figure 3.26a) remain comparatively steady at approximately two eliminations per voxel which, while matching HOMOR in its high SNR regions (slices 30-50) is significantly lower towards either extreme. Interestingly however, if we compared the median (Fig. 3.26b) or mode (Fig. 3.26c) eliminations per voxel, results across the methods begin to stabilise at 1-2 rejections per voxel. The question, then, is why is there such a difference in mean result? Figure 3.27, which provides an example slice through lower brain (slice 18, blue line in Figure 3.27a) that holds part of the answer. If we examine Figure 3.27b, which displays a histogram of voxel wise elimination numbers, we observe that for a small fraction of examined voxels, HOMOR proceeds to eliminate ≥ 20 elements, while the majority lie in the 1-10 range similar to RESDORE. If we then examine Figures 3.27a and 3.27c, we observe that these high elimination voxels lie in regions which the signal (in this case a non intensity corrected b0, but the effects are equally apparent in diffusion weighted images) is below average intensity. While intensity correction can then normalise signal intensities and make the images *appear* homogeneous, in reality individual signal to noise ratios remain unchanged by the re-scaling multiplications and, since the same noise has greater effect on lower intensity signals, this can cause highly localised variations in

appropriate SD_{noise} values which cannot be represented by a globally defined estimate – leading to the artefactually high eliminations seen in the lower regions. It may seem reasonable, then, to define SD_{noise} on a per-slice basis. Indeed, the results presented in Figure 3.26 (black line) would seem to support this conclusion. However, the problem with operating in this manner is that should there be a slice-wide artefact, sensitivity to that artefact – particularly in fibrous tissue aligned to the corrupted gradient image (Fig. 3.28) – can be significantly reduced.
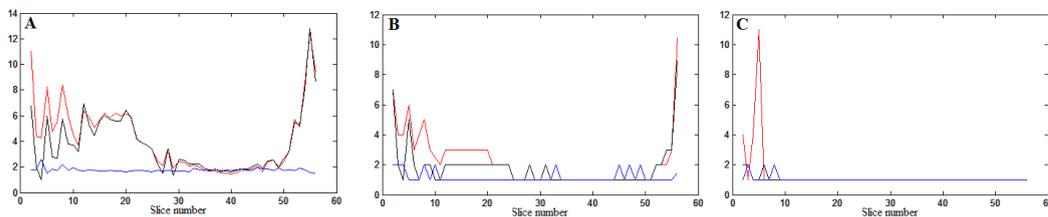
FIGURE 3.26: (A) Mean rejections per labelled voxel per slice. (B) Median. (C). Mode. Colour coding: Red – HOMOR using global SD_{noise} estimate. Black – HOMOR using slice-wide SD_{noise} estimate. Blue – RESDORE.
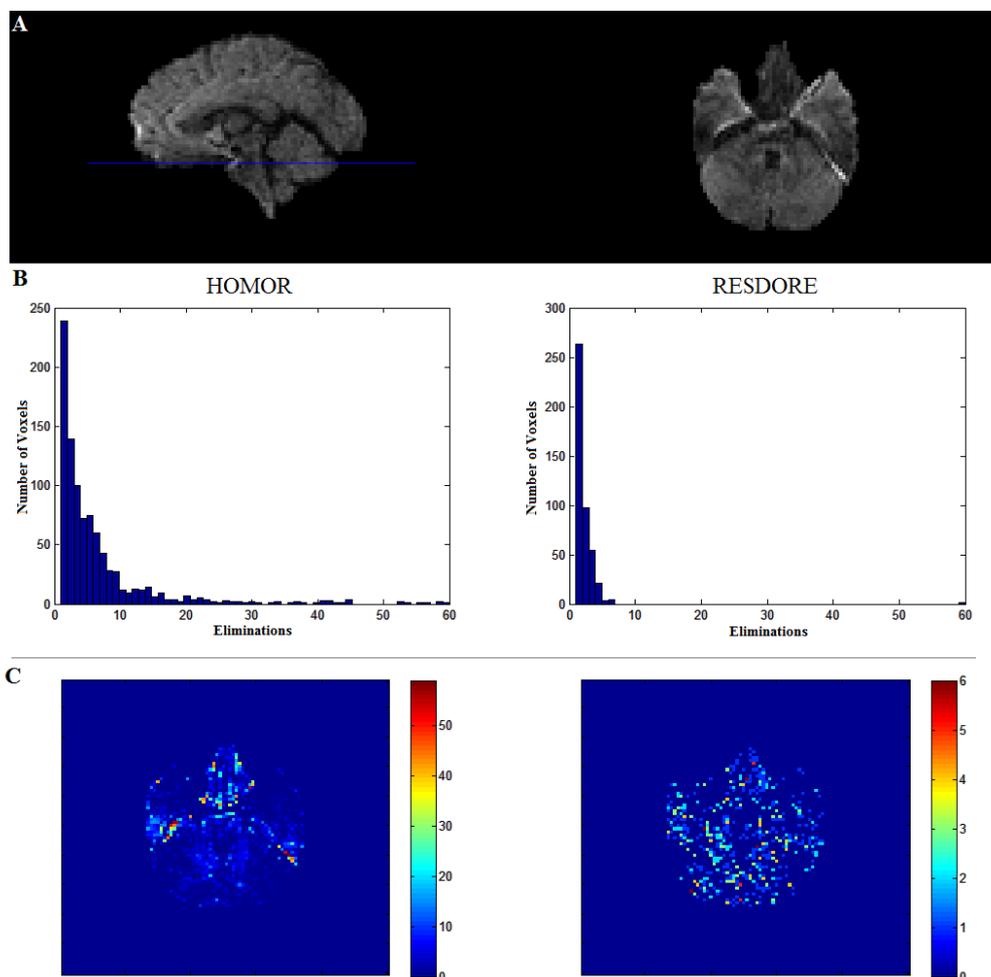
FIGURE 3.27: Slice through the lower brain. (A) B0 image, blue line represents the vertical position. (B). Histogram of rejections/labelled voxel. (c) Map of rejections/labelled voxel.
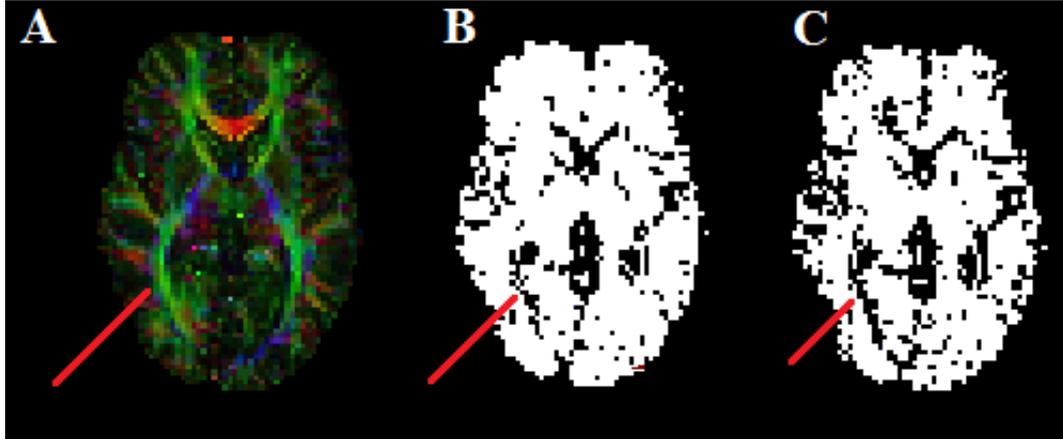
FIGURE 3.28: Labelling maps. (A). DEC map, note the labelled fibre pathway with orientation aligned to the corrupted gradient. (B) Labelling map using a global SD_{noise} estimate, note difficulty detecting corruption along the aligned fibre pathway. (C) Labelling using slice calibrated SD_{noise} estimate, note absence of detection along the aligned pathway. Red line indicates a nominally X-axis aligned fibre bundle.

RESDORE does not suffer from this particular issue, but that is not to say it is without fault. Using the variance of the signal itself provides a degree of insensitivity to regional SNR changes but, as shown in Figure 3.16, where SNR falls too low, the resulting large variances can make detection of corrupted data difficult. An example of this may be found in Figure 3.29 in which, due to a combination of inadequate field un-warping and errors in the motion/distortion correction, a combination of striating drop-out and linear stretching (down the phase encode direction) can be observed in the b0 image and SNR in the remaining DW-MRI images is generally poor. The way each algorithm responds to this situation is characteristic of their worst case performance. Beginning with HOMOR (Figure 3.29b), performance once again tends to be linked to corruption of the b0 image (Figure 3.29a), where b0 signal is low or absent, a large portion of that voxels signal elements are eliminated, conversely, where signal is high (even where geometrically distorted – note the alternating high/low patternation where high/low intensity signals have been stretched along the phase encode direction) rejections drop off sharply. Regardless of this failing, HOMOR is at least picking up some of the corruptions within this region. RESDORE, on the other hand, eliminates very little in comparison.

For visual demonstration, Figure 3.30 provides two examples. Each example consists of a b0 normalised DW-MRI signal, the residuals to the first iteration of the HOMOR fit/eliminate process and the distortions to the isotropic component of the RESDORE dictionary. Note that in the latter two, red lines indicate the elimination thresholds for each algorithm. The first example is taken from a extremely noisy/corrupted voxel in the region described in the previous paragraph. Note that the normalised DW-MRI signal exceeds one in several places which, remembering the $S = S_0 e^{-bD}$ equation covered in Chapter 2, is indicative of a signal loss artefact in the b0 image (S – the measured DW-MRI signal – should never exceed the unweighted S_0 signal intensity). It is through examination of this signal that, while very noisy, some signal elements are corrupted
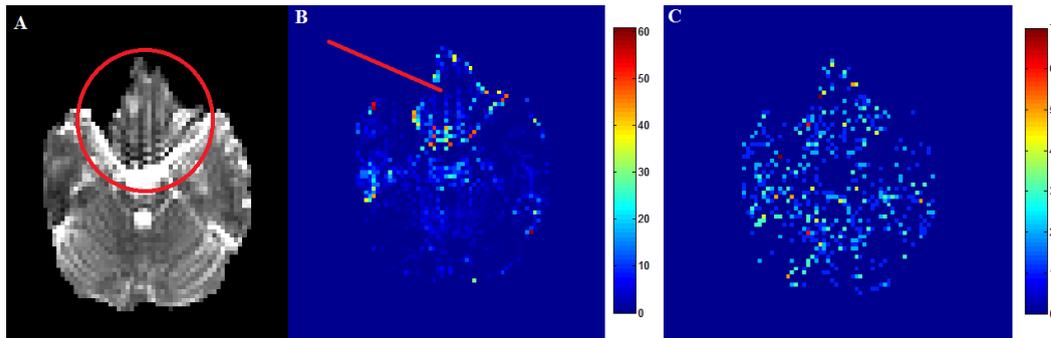
FIGURE 3.29: Corruption example. (A). B0 image detailing b0 corruption/distortion in the lower brain. (B) Elimination/labelled voxel map calculated using HOMOR, note the striation patterns matching that of the b0 drop-out. (C) Elimination/labelled voxel map calculated using RESDORE.
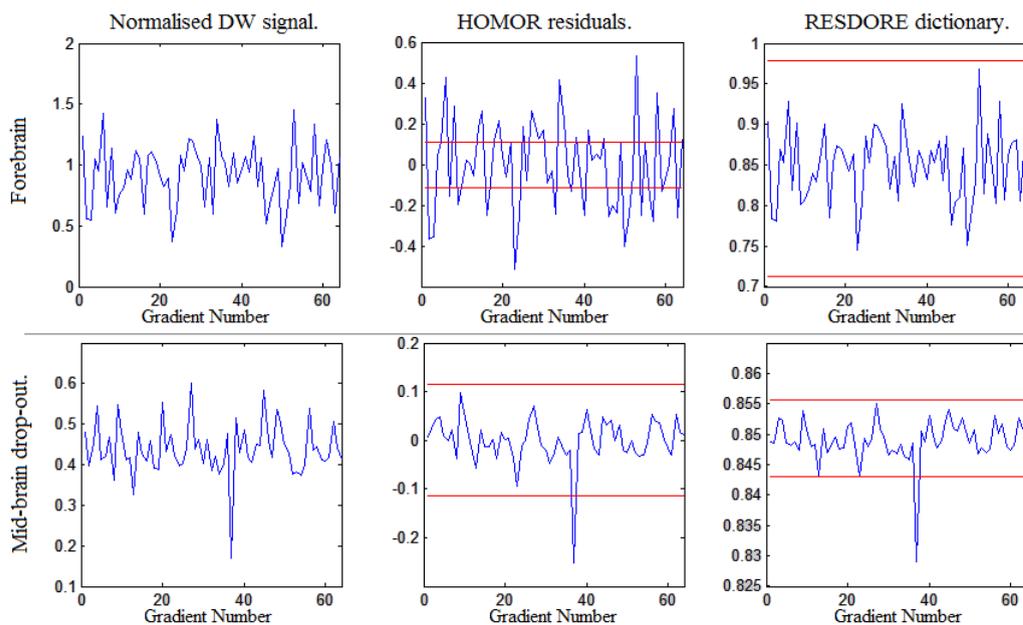
FIGURE 3.30: Modes of failure example. Upper: Noisy signal taken from the forebrain. Lower: High SNR signal taken from the mid brain, single element undergoing drop out. Red lines indicate elimination thresholds.

further than others – for example gradients 23, 50 and 53 show large residuals in both the HOMOR and RESDORE result and so, if possible, it would be preferable to eliminate these elements without rejecting the remainder. Beginning with HOMOR, the volume-wise SD_{noise} estimate is simply too restrictive. While all severely corrupted elements are eliminated, it also selects the majority of the remaining elements (43 after all iterations are completed). RESDORE elimination thresholds, on the other hand, are dominated by the high background noise in such a way that the rejection criteria is pushed beyond the residual spread and no elements are selected. Clearly, neither situation is desirable.

In the second example, we have selected a signal from slice 29 (the slice-wide drop out example covered in Figures 3.23 and 3.24). At this point in the image the overall SNR is significantly higher than in the previous example. As such, a single large drop out is

easily identified by both algorithms with no error. Thus, as we would expect, it is far easier for each algorithm to operate in regions of (comparatively) high SNR.

Accepting that in the majority of cases both algorithms will produce a similar result (within 1-2 eliminated elements), deciding on a preferable mechanism of edge case failure is difficult and, amongst other points soon to be discussed, could largely depend on how you choose to define corruption. In absolute terms, the mean squared residual to any model fit to the DW signal data in the first example will be higher than residuals fit to the second, even including the obvious slice wide artefact. By this definition, the first signal is far more corrupt than the second and HOMOR is correct in rejecting the majority of its elements. This kind of outcome could be useful if, for example, you are searching for voxel specific errors in otherwise clean data. Another way to look at corruption, however, is to accept that signals may be noisy and that noise will vary over the image. In this case, one might choose to define corruption as a signal element that is significantly different from its peers, rather than the volume as a whole. Now, in the case of the second example, either definition will make no difference, the single large drop out is both significantly different from its peers *and* significantly different from the volume as a whole and thus will be identified in both cases. The difference is observed in the first example; while there are elements in that signal that may be different from their peers, they are not *significantly* different and thus, by this definition not corrupt (or at least no more corrupt than their peers). By this definition, RESDORE performs the correct action.

Ultimately then, the decision between HOMOR and RESDORE can be condensed to three factors. Firstly, if you can guarantee a consistent SNR and mean signal intensity, HOMOR, provides a quick (averaging 2 minutes per volume) and potentially powerful method for the identification of outliers. If you are examining data equivalent to those we have seen thus far – high quality, high redundancy – either method can be used depending on your preferred definition of corruption. Finally, if examining data with highly variable SNR or low redundancy, RESDORE (while slower, averaging approximately 8 minutes per volume) is more likely to provide a consistent result while minimising eliminations that may otherwise result in an ill-conditioned matrix problem for future processing. RESTORE, or preferably iRESTORE which aims to reduce possibility of excessive eliminations, should only be applied to low angular resolution data for which there would be insufficient data points for either HOMOR or RESDORE to operate – perhaps images with $\leq 20 - 25$ gradient directions, remembering that while 4th order harmonics only require 15 data points to fit, points will be eliminated at each iteration of the HOMOR algorithm and could quickly lead to a ill-determined harmonic fit which would in itself cause issues.

Finally, returning to the the original observation of a counter intuitive increase in rejections across the upper brain as b-values rise, a possible generative mechanism can be seen in Figure 3.31. Figure 3.31a plots mean slice-wide estimates of SD_{noise} normalised
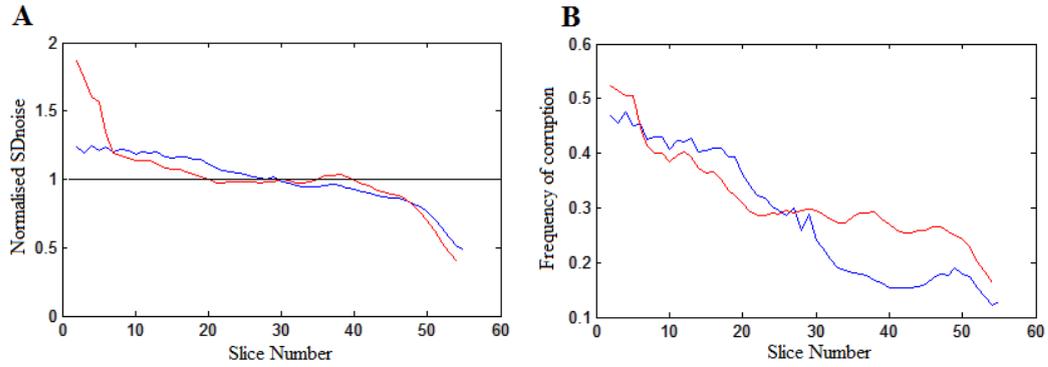
FIGURE 3.31: Differences between images at each diffusion weighting. (A) mean slice-wide SD_{noise} estimate (normalised by global estimate). (B) Mean fraction of brain-containing voxels labelled as potentially corrupt. Blue: $b = 1000s/mm^2$. Red: $b = 3000s/mm^2$.

by the mean volume-wise SD_{noise} while Figure 3.31b plots fraction of voxels (as a portion of total "brain-tissue" voxels in that slice) identified as containing corruption. The results provided here are from the HOMOR algorithm, but, since all selected labelling techniques use similar SD_{noise} estimation methods, the pattern is equally descriptive of RESTORE and RESDORE. Notice that in Figure 3.31a, the blue line (representing $b = 1000s/mm^2$) gradually transitions from the global SD_{noise} estimation underestimating the local value to overestimating the local value with an inflection point at about slice 30. If we then examine the labelled fractions, this point of inflection corresponds with a sharp decline in voxel labelling which, by extension, would decrease the probability of rejections in that mid-high brain region (in fact, returning to Figure 3.21, there is a distinct change in intensity at approximately slice 30 that supports this observation). With the higher diffusion weightings (red line, representing $b = 3000s/mm^2$), differences between local and global estimates are less pronounced across the majority of the signal volume, leading to a plateau in labelled fraction from approximately slice 20 to slice 50. With (comparatively) more voxels labelled, it would be natural for the observed increase in rejection probability. While the precise cause of the improved SD_{noise} estimation at high diffusion weightings is unknown (with the limited information we were provided with the image data it is difficult to make conclusions), this is essentially just another facet of the previously described regionally varying SD_{noise} related issues which, for HOMOR and RESTORE, is further compounded since that same estimate then becomes an integral component of the elimination process.

Section 3.4.2 Summary

In summary, this section begins with a simple validation of our implementations of the HOMOR and RESTORE algorithms by comparing the resultant ORP maps with those provided by Pannek et al. (2012) (Figure 3.20). As expected, our ORP maps proved consistent, indicating that our implementation of each algorithm is correct.

Moving forwards, mean voxel-wise elimination maps were then produced for direct comparison of the three algorithms. A key observation resulting from this comparison was that those algorithms based on a global SD_noise measurement (i.e. RESTORE and HOMOR) tended to produce a greater number of eliminations across the lower brain regions when compared to RESDORE. While investigating this discrepancy, two key findings were made.

Firstly, as suggested by the simulations (Section 3.3.3), RESDORE and RESTORE/HOMOR have two very different modes of failure when confronted with extreme corruption. While in both cases extreme corruption would result in unusually high residuals to signal elements (corrupted or otherwise) when subjected to the required model fit, the difference appears in how these residuals translated into rejections. For HOMOR/RESTORE, the strict global SD_{noise} is often simply too low to distinguish between genuine corruption and a higher than normal residual produced by a poorly fitting model. As a result, significant numbers of both corrupted and uncorrupted signal elements will be rejected in order to force a model fit with a lower than reasonable residual threshold, resulting in extremely high false-positive rejection rates in highly corrupted areas (Figure 3.30). RESDORE, on the other hand, will tend produce false-negatives once corruption becomes pervasive since increasing residuals on non-corrupt elements tends to reduce the relative difference between corrupt and non-corrupt residuals, making the latter harder to detect under the data-driven three standard deviation exclusion criterion.

Secondly, since image SNR tends to vary throughout an image, the global SD_{noise} estimate will consequently tend to either over or under estimate the local SD_{noise}. Where local SNR is lower (and thus SD_{noise} higher) than that modelled by the global value, non-corrupt signal elements are frequently rejected by an overly restrictive threshold (high false positives), whereas if local SNR is high (and thus SD_{noise} low), residuals arising from true corrupted elements can fall beneath the relatively large global SD_{noise} threshold and thus fail to be detected (high false negatives). With this in mind, examining Figure 3.25 allows us to observe a correlation between slice-wise SD_{noise} estimates and slice-wise rejection rates that explains our initially observed bias towards increased rejection in lower image regions – simply put, while there is indeed corruption within these areas, the SNR is also substantially lower than the main body of imaged volume and thus, as a result of higher than expected SD_{noise}, a large number of false-positive rejections are observed in the lower image regions. Since RESDORE uses an alternate rejection criteria it does not suffer from similar SNR variation issues and thus is more robust in their presence – a significant advantage.

Finally, while preliminary investigation of setting local SD_{noise} thresholds was briefly examined within this subsection, the following subsection (Section 3.4.3) will address this topic in greater detail.

3.4.3 Novel Data

All image data processed to this point are, with few exceptions, high quality, high redundancy and relatively artefact free. While analysing such image data has provided useful insights into each algorithm's behaviour, they are too "clean" to gauge performance against genuine and more severe corruption. To that end we will now present results for a specifically designed novel dataset. The image data in question comprises a $1000s/mm^2$, 48 direction (truncated from 60 due to premature termination of the scanning session, Cook et al. (2007)), 6 b0 scan recorded on a GE Signa HDx system. During this scan the subject was instructed to move, talk and and perform other actions that would simulate the degree of corruption one might expect when imaging a difficult subject. Unlike images used in the previous subsection, this image was captured across 30 slices spanning the mid and upper brain only (Fig 3.32a), ensuring an approximately constant (b0) SNR throughout the volume (Fig 3.32b). The implication for the examined algorithms is that, ignoring artefacts, estimated SD_{noise} would be relatively constant throughout the image, eliminating the previously observed biases towards over-detection/elimination in the lower/forebrain and under-detection/elimination in the mid/upper regions. As was previously stated, however, this image is not artefact free, in fact, almost every slice is affected by at least 3 major artefacts (two volume wide inter/intra slice intensity artefacts (Fig. 3.33a, b) and one volume wide drop-out (Fig. 3.33c, note that slices 20-26, while still experiencing drop out, are less severely affected) on top of numerous mis-registrations and motion correction faults. A brief summary of which (calculated in ExploreDTI from residuals to the diffusion tensor fit) can be found in Figure 3.34; peaks at DW image 7 and 23 correspond to the drop out and most significant inter-slice artefact, respectively, with the smaller peak at image 41 corresponding to the second inter-slice intensity artefact.
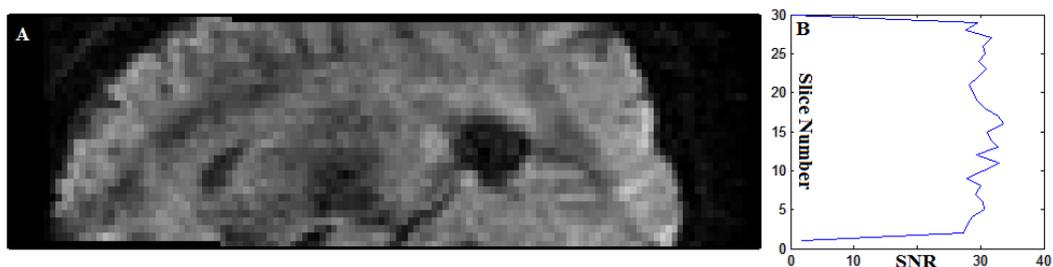
FIGURE 3.32: (A) 30 slice volume collected during 'difficult subject' scan. (B) b0 SNR estimate at each slice.
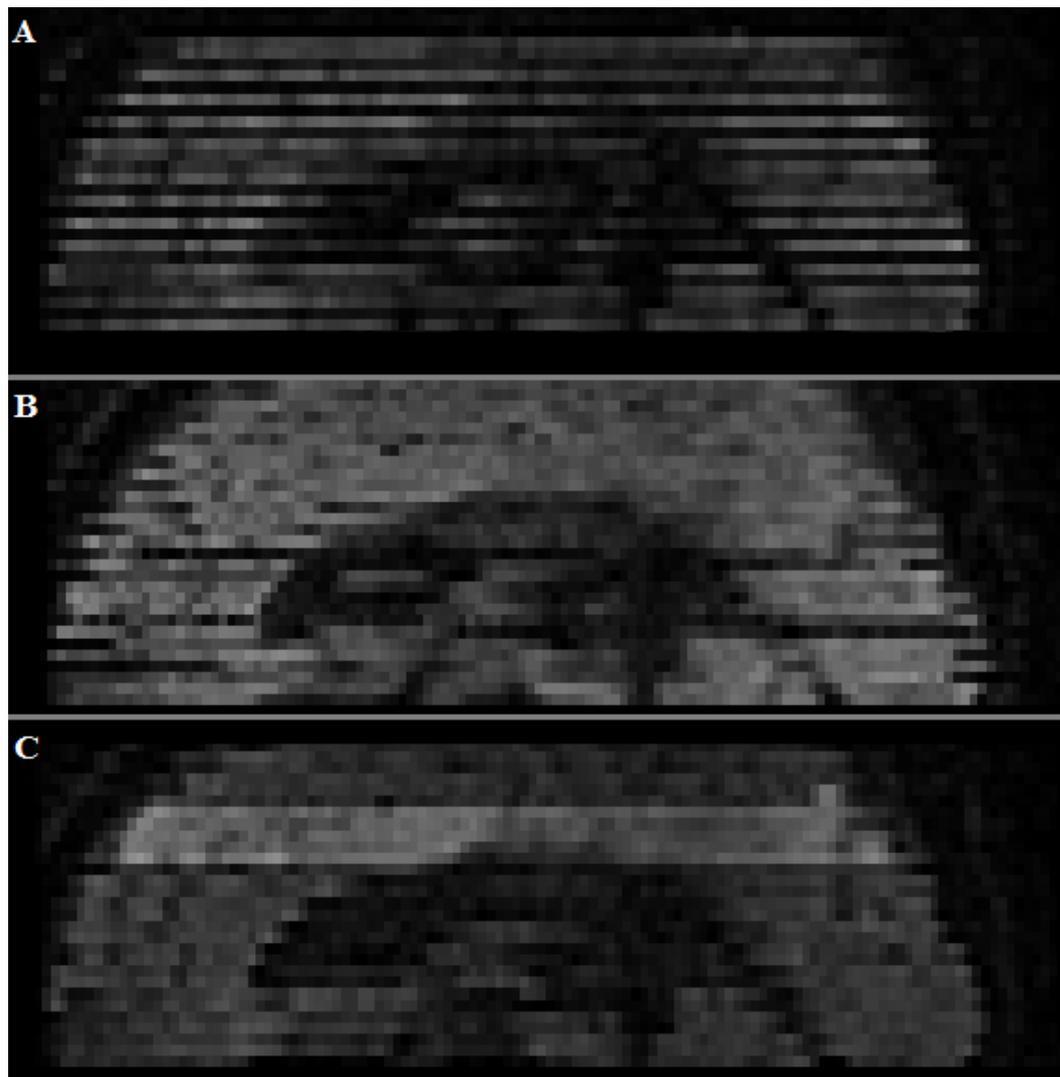
FIGURE 3.33: Example volume-wide errors. (A) Severe inter-slice intensity artefact. (B) Moderate inter-slice intensity artefact. (C) Drop out artefact.
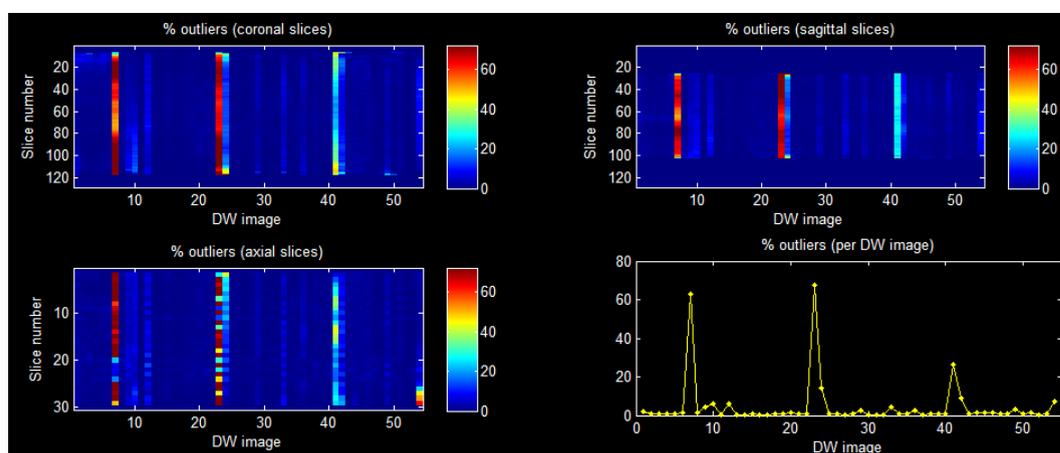
FIGURE 3.34: ExploreDTI data quality summary indicating outliers to the diffusion tensor fit.

The effects of this corruption, particularly the inter-slice intensity artefacts (Fig 3.33a), can be clearly observed in plots of slice-wide SD_{noise} estimates (Fig. 3.35a). As intensity varies from marginally to significantly below the expected magnitude, so varies the slice wise SD_{noise} estimates from below to above average. If we then take the volume wide estimate (red line) and perform voxel labelling, predictably, the fraction of voxels labelled in each slice is strongly correlated with increases/decreases in slice-wise SD_{noise} (Fig. 3.35b). Clearly, global estimation does not function well in such heavily corrupted data. As final reinforcement of this fact, Figure 3.36 displays rejection maps for two consecutive slices (19 and 20). The first slice experiences severe drop out in the 7th and 23rd images, with further minor loss in the 41st resulting in a high slice SD_{noise}. The second – and one of the least corrupted slice in the set – has slightly decreased intensity in image 7 and then larger drop outs in the 23rd and 41st images, resulting in a comparatively low SD_{noise}. Both are then further affected by two (comparatively) minor motion/distortion correction artefacts (images 24 and 42, discussed in Appendix B).
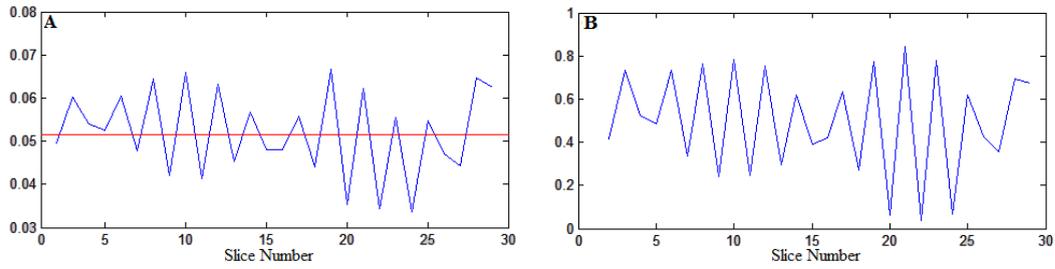
FIGURE 3.35: (A) Slice wise SD_{noise} estimates. (B) Slice wise fraction of brain-containing voxels labelled as potentially corrupt.

Beginning with HOMOR, on the surface the high SD_{noise} slice result appears reasonable. Excepting regions with high b0 intensity (e.g. the ventricles and grey matter, leading to previously described failures), the majority of the slice has been correctly labelled as corrupted. However, on further examination the number of elements rejected is almost uniform; this is confirmed by the histogram of rejected images (Fig. 3.36c) which shows frequent rejection of the two major sources of corruption (drop out in images 7 and 23), but virtually no sensitivity to the slice 41 artefact or any of the localised sources of corruption. For reasons that will be explained shortly, even though the global SD_{noise} estimate is below the slice-wide artefact, it is still simply too large for any sensitivity to smaller corruption. If we then examine the second slice with low SD_{noise}, the only voxels highlighted are those belonging to regions with naturally low b0 intensity – again, the result of a previously described failure.
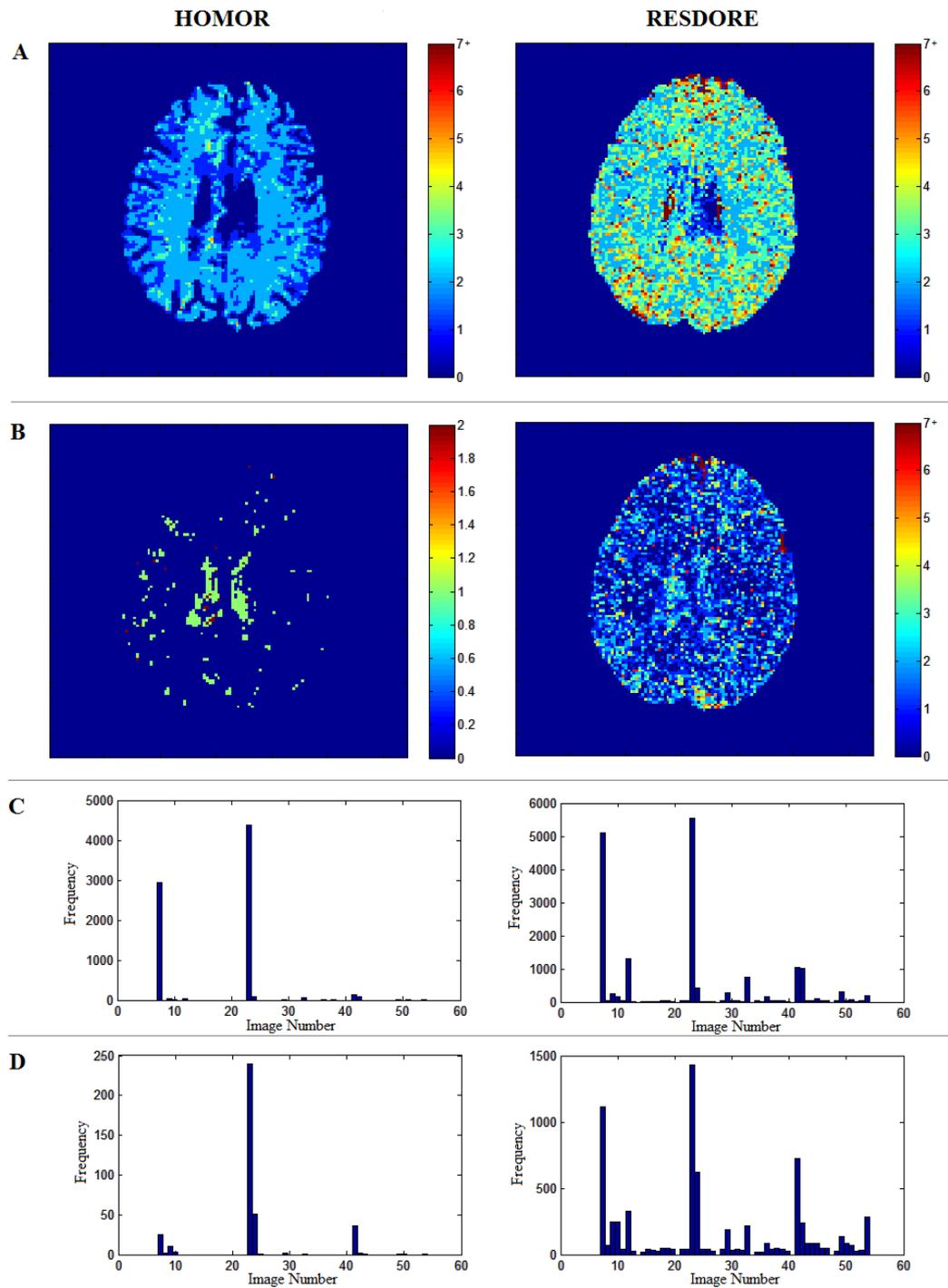
FIGURE 3.36: Varying degrees of corruption example. (A) Elimination map for slice 19. (B) Elimination map for slice 20. (C) Histogram of DW signal elements eliminated in slice 19. (D) Histogram of DW signal elements eliminated in slice 20.

Proceeding to RESDORE; we know from visual inspection that virtually every voxel in the image (and by extension the two slice test cases) is corrupted along at least one gradient direction. This knowledge allows us to forgo any voxel labelling step and simply apply the algorithm to all brain-containing voxels, negating the need for any SD_{noise} calculation (this is not a possibility with HOMOR since the first iteration of the elimination phase is identical to its labelling technique and thus still dependant on SD_{noise}). As a result, a far greater portion of the voxels in both images are correctly labelled as containing outliers. Starting with the more corrupted image (Fig. 3.36a), RESDORE's elimination of the two major artefacts is more effective than HOMOR – while identification of the image 23 artefact is more or less equal (accepting differences in the fraction of voxels tested), HOMOR struggled to identify image 7 in regions of high b0 intensity – something that is not an issue with RESDORE (Fig. 3.36c). Furthermore, RESDORE is also more likely to eliminate the image 41 artefact and the additional minor artefacts (e.g. image 42, Appendix B). Differences within the less corrupted slice are even more pronounced. While the entire image is no longer labelled (approximately 65% of brain-containing voxels), this is still a significantly larger fraction than managed by HOMOR. As expected, however, the reduced severity of the image 7 in slice 20 has proportionately decreased the likelihood of detection (compare the relative heights of images 7 and 23 in the Figure 3.36c and 3.36d RESDORE histograms, note the decrease in 3.36d). One point of concern is the increase in eliminations of non-obviously corrupt signal elements. Some of these may be attributable to very localised errors (e.g. artefacts in the motion/distortion correction, particularly at the image edges or boundaries between tissue types) but, another explanation is that given corruption is overall lower in this volume, we are actually seeing real false positives identifications. Recalling the simulations, we observed that where the number of true corrupted elements was close to one (and particularly if the degree of corruption was low), there is a chance of false positive detection based on spikes in the background noise. While difficult to prove one way or the other (there being no gold standard for this dataset), we would not be surprised if either were true in this case, particularly with the lack of a labelling phase for this experiment.
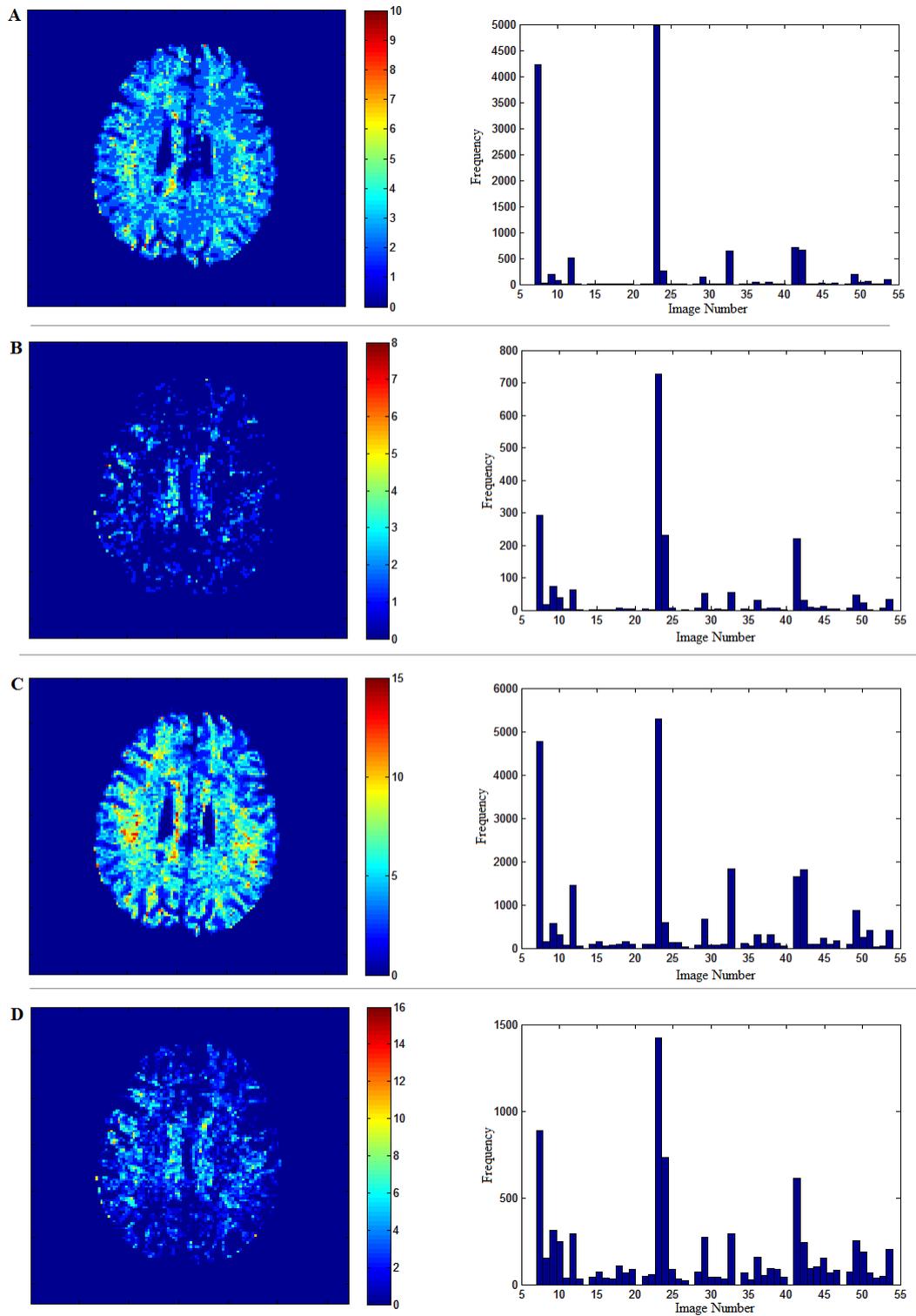
FIGURE 3.37: Varying degrees of corruption example. (A) Repeat of slice 19 using slice specific SD_{noise} estimate. (B) Repeat of slice 20 using slice specific SD_{noise} estimate. (C) Repeat of slice 19 with corrected SD_{noise} estimate. (D) Repeat of slice 20 with corrected SD_{noise} estimate.

This experiment, more than any so far, shows the true difference between the RESDORE method of elimination and that of HOMOR/RESTORE. The problem with estimating a parameter like SD_{noise} in the manner prescribed by Pannek et al. (2012) is that there is an implicit assumption that the signal data in each slice is comparable and that, at minumum, 50% of that data remains uncorrupted (Walker et al. (2011)). This image breaks those assumptions on two accounts; firstly, data quality varies wildly from slice to slice and, while one might consider pairing the high and low quality slices together for similarity, they would still break the 50% corruption assumption, greatly reducing the effectiveness of any detection. To explain, Figure 3.37 repeats the previous two slice experiment, this time using the lowest slicewise SD_{noise} estimate (0.0351 compared with global value 0.0516). If we examine the slice 19 elimination map (Fig. 3.37a), processing appears greatly improved, with results now largely matching those of RESDORE. Unfortunately, however, slice 20 results are still far from correct. The reason is fairly simple; even the lowest SD_{noise} estimate derived from this image was derived from 100% corrupted signal data. Given the previously discussed 50% assumption, the SD_{noise} calculation technique will provide a result such that 50% of the exemplar data falls below the threshold. Now, where corruption is more prevalent than the exemplar data, e.g. slice 19, this is not an issue – the residuals of the corrupted elements to the spherical harmonic fit will be larger than the (relatively low) "noise" threshold and thus the majority will be correctly detected and eliminated. The problem comes when we return to examine data of the same quality used to derive the SD_{noise} estimate which is specifically formulated to allow approximately 50% of the data to pass – and this is exactly what happens. It is simply not possible to derive an adequate measure of SD_{noise} from wholly corrupted signal data. The only way around this is to know, *a priori*, which signal elements are most corrupted and exclude those from the SD_{noise} fitting process and, if this is the case, the reason for running such an algorithm is largely negated. That said, Figures 3.37c and d repeat the experiment one last time with such an SD_{noise} estimate. While labelling and eliminations in slice 20 (Fig. 3.37d) significantly improve within in the white matter, elsewhere (or anywhere in slice 19, Fig. 3.37c), eliminations begin to increase significantly. What we are observing here is similar to the lower/forebrain areas in the previous subsection. In the less corrupt slice 20, we observe that any region with low b0 intensity experiences above-average eliminations more likely to be driven by a difference in SNR than true detection of corruption. Where the corruption is more severe, i.e. slice 19, the situation worsens. As predicted by the simulations, a combination of a restrictive SD_{noise} estimate and residual "drift" resulting from the, in some places significant (approaching 10% of expected intensity), signal drop outs conspire to produce eliminations that go far beyond what visual inspection of the data can support.

3.4.4   Voxel Wise Elimination: A conclusion

To close this subsection and tie in lessons learned through both simulation and analysis of in vivo images, the selection of outlier rejection algorithms largely depends on the quality of the target data or, more precisely, the applicability of a volume-wide SD_{noise} estimate. If one simply wishes to quickly check otherwise high-quality data for small or local artefacts, HOMOR is probably the method of choice, being far simpler (computationally) and therefore faster than either RESTORE or RESDORE while still providing adequate elimination results. In light of the high angular resolution techniques, uses for RESTORE are limited to either images with very low diffusion weightings or those with an insufficient number of gradient samplings to support more complex models – in most other situations, either HOMOR or RESDORE should provide a superior result. RESDORE's usefulness comes in two parts; firstly, it is more able to cope with localised variations in SNR than methods based on a fixed volume (or even slice) wide noise estimate, allowing greater data retention in noisy yet not necessarily corrupt areas. Secondly, where the data is heavily corrupted, limitations of the SD_{noise} estimation process severely hamper RESTORE/HOMOR operation while RESDORE, which operates independently from such parameters, is more likely to provide a reasonable result without undue variation as the degree of corruption varies. Adding to this information gathered through simulation, particularly the low false positive production which was then supported by real image data examinations, we can say with some certainty that RESDORE has achieved the stated goals, being able to match or exceed competing techniques detection capabilities, while failing in a way that preserves rather than expends individual data points.

3.5   Multi-Voxel Signal Elimination

We have mentioned at several points that the RESDORE algorithm can be modified to detect corruption on a multi voxel scale (e.g. across slices). As a final experiment before concluding this chapter, we will now demonstrate that capability. Recalling the equations for fODF fitting:

$$S^{t+} = S^t.*((A^{t+1})^T X).*((A^{t+1})^T S^t + \delta) \tag{3.5}$$

and dictionary update:

$$A(:,1) = A(:,1) - \mu * ((A*S - X)*(S(1,:)')) \tag{3.6}$$

where X and S are column vectors describing the DW-MRI signal fODF, and A is a matrix describing the set of dictionary elements; adaptation of the single-voxel RESDORE algorithm to multi-voxel processing is as simple as replacing X and S with matrices in which each column describes, respectively, one of the target DW-MRI signals and a related fODF approximation (all of which are initialised to a constant value). The calculations themselves, being matrix operations implemented in matlab, required no alteration.

Figure 3.39 displays dictionary modifications for 9 simulated $FA = 0.7$, $b = 1000s/mm^2$, $SNR = 10 : 1$ DW-MRI signals comprised, at random, of one or two fibrous components with randomly generated orientation. We then corrupt the second gradient direction (again, chosen at random from 60 availible choices) by multiples of 0.4, 0.6 and 0.8 (Fig. 3.38a-c respectively). In the left column, we observe the dictionary modifications for one of the 9 individual elements and on the right, the dictionary output from simultaneous processing of all 9 signals. Red lines indicate the elimination thresholds. As in previous simulations, the single-voxel application has difficulty detecting the imposed drop-out at this SNR given that (especially at high multiples) it may lie within the noise induced margin of error. Once we process the voxels simultaneously, however, there is no correlation amongst the noise along any one gradient – the only correlated error that can be corrected by modifying the dictionary is the second gradient corruption that we ourselves introduced. As such, when we examine the dictionary modification profile, the deliberately corrupted element begins to stand out, even (just barely) surpassing the elimination threshold at a 0.8× corruption multiple (completely undetectable when processing one voxel at a time). The fact that we are able to see such improvements while still only processing a relatively small sample (when compared to an entire slice) speaks strongly to the power we can add by processing in this manner. Future work will examine the utility of this method for elimination of highly localised clusters of corrupted data, such as arising from cardiac pulsation effects.

Before we conclude with a demonstration against genuine slice-wide corruption, the following should serve to demonstrate the sensitivity of this method when applied to a full slice. Simulating the previously observed inter-slice intensity artefacts, we have selected (at random, from a pool manually inspected slices confirmed to lack any major pre-existing slice-wide artefacts) 10 $b = 1000s/mm^2$ midbrain image slices from the HOMOR publication dataset. At random, we then corrupt 100 sets of 1, 2 and 3 gradient images by multiplying their intensities by values ranging from 0 to 2 at a 0.1 increment. Examining the result (Figure 3.39), we can observe that in comparison with single-voxel processing, simultaneous multi-voxel processing is extremely sensitive to shared artefacts, achieving in all cases over 90% elimination for corruptions as small as ±10%

Finally, as a demonstration of applicability to real-world data we return to the purposefully corrupted 'difficult subject' dataset. Figure 3.40 contains dictionary modification
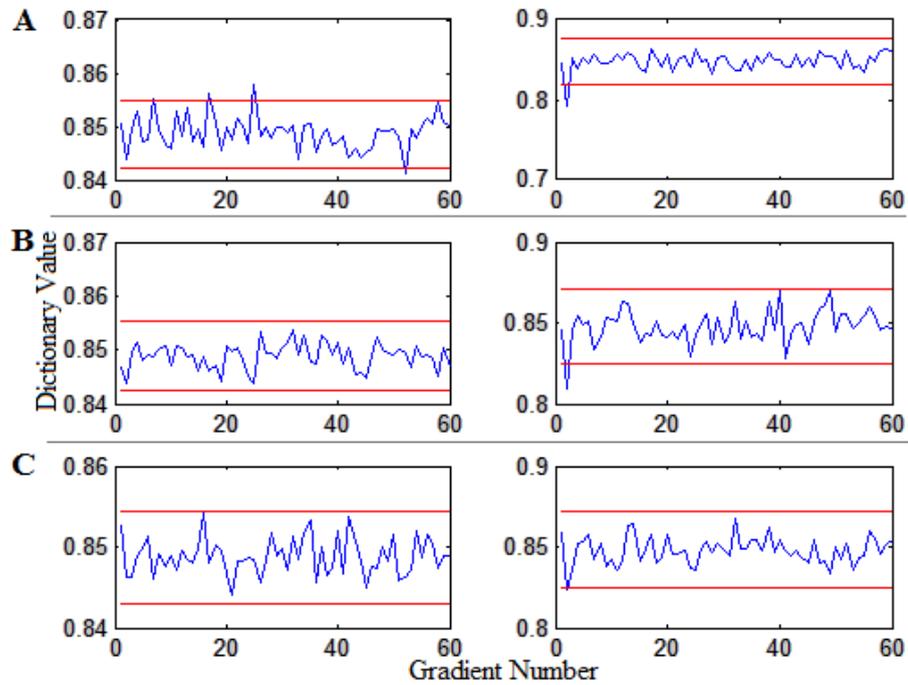
FIGURE 3.38: Multi-voxel processing example. Left: single voxel result. Right: 9 voxel result. (A) 0.4× multiple. (B) 0.6× multiple. (C) 0.8× multiple.
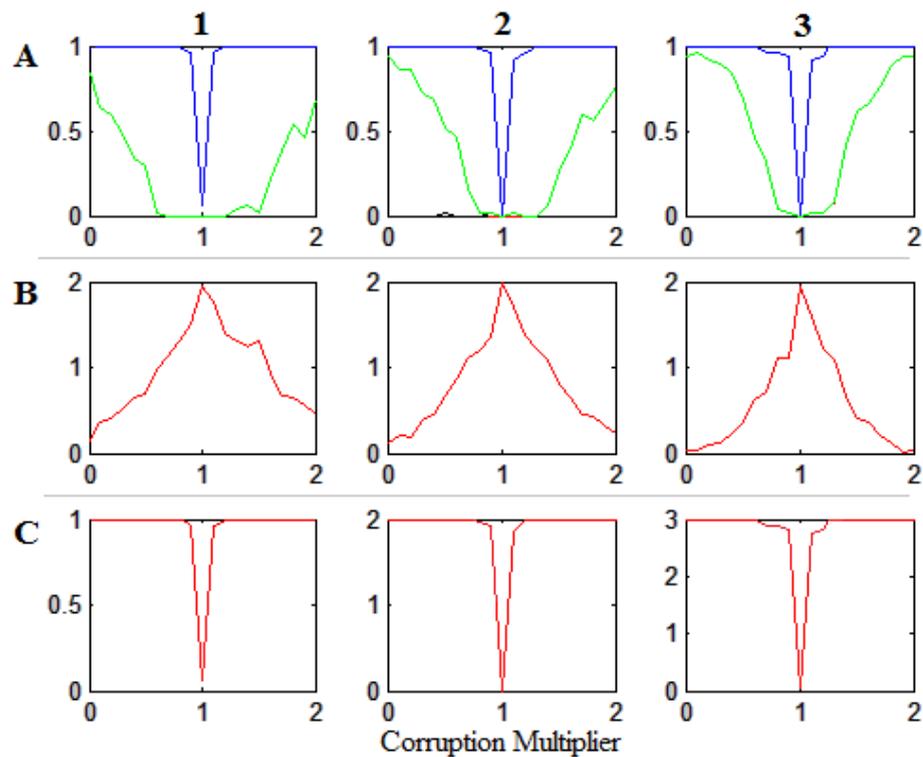
FIGURE 3.39: Slice wide corruption simulation. (A) Rejection results, colour coded according to Table 3.1. (B) False positives. (C) Probability of whole-set recovery.

traces for the previously examined slices 19 and 20, with red lines indicating the rejection thresholds. For reference, the eliminated DW gradient images are 1, 6, 17, 29, 35 and
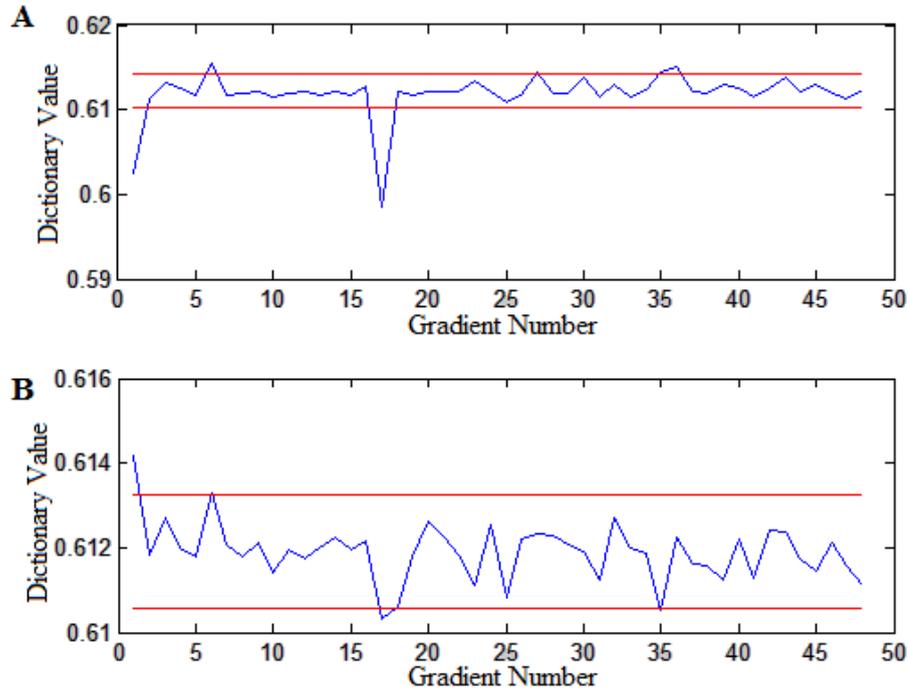
FIGURE 3.40: Multi-voxel processing example, elimination of slice-wide corruption in real data. (A) Dictionary modifications resulting from 'difficult subject' slice 19. (B) Dictionary modifications resulting from difficult subject slice 20. Red lines indicate elimination thresholds.

36 for slice 19 and 1, 6, 17 and 35 for slice 20. If we re-examine Figures 3.36 and 3.37, we see that these rejections (accounting for a shift in numbering due to exclusion of the first 6 b0 images) precisely match some of the major peaks in their respective voxel-wise elimination histograms, indicating good agreement between results (remembering that this methods is not intended to detect localised artefacts responsible for some of the other peaks in the Figure 3.36-3.37 histograms). As an additional benefit, due to re-duced number of dictionary update and copying operations (dictionary is modified once per iteration per slice, rather than once per voxel per iteration per slice), processing time is greatly reduced to approximately 1 minute per 60 slice image – though one must remember that application to uncorrupted slices is likely to result in a false positive result. For heavily corrupted datasets like the one used in this example, however, a whole-volume slice-wide RESDORE procedure might make a sensible first step. Recall-ing the previous section, the main restriction to applying HOMOR was that pervasive slice-wide corruption prevented accurate SD_{noise} estimation, using this method to first eliminate those slice-wide artefacts might well reintroduce HOMOR as a valid tool for detecting more localised errors.

3.6   Discussion and Conclusion

In this chapter we have presented a method for detection and elimination of corrupted DW-MRI signal elements. Through a combination of simulation and application to real subject image data we have demonstrated that RESDORE is in most cases competitive and, in heavily corrupted images, clearly superior to competing techniques, with the added bonus that it is easily scalable for eliminating (shared) corruption from a single voxel up to entire slices with matching increases in sensitivity.

The major advantage of RESDORE over competing methods is that it is not dependent (beyond an optional labelling phase) on elimination thresholds external to the voxel/set of voxels currently being processed. Operating in this manner provides a degree of insensitivity to localised variations in signal quality but, depending on the definition of corruption used, can have the potentially undesirable failure-to-detect side effects if the local SNR drops below a certain threshold ($\approx 10 : 1$). That said, Figures 3.15 and 3.16 indicate that were the whole image to display equally low SNR, both HOMOR and RESTORE would have similar problems.

Overall, and repeating previous recommendations, deciding between HOMOR, RESDORE and RESTORE can be aided categorising image data into one of four classes. Firstly, if it can be guaranteed that captured image data displays a constant SNR and that there is a low probability of pervasive corruption, HOMOR provides a quick and powerful method for the identification of minor outliers; while RESDORE is still competitive in such situations, the increase in processing time is probably not worthwhile. If SNR is variable, but corruption remains low, both HOMOR or RESDORE can be used so long as one acknowledges and accounts for the product of those variations – i.e. over/underestimation of the degree of corruption as local SD_{noises} varies with respect to global averages. Where data is heavily corrupted, or SNR varies and there is minimal signal redundancy, RESDORE is clearly the method of choice. Unlike HOMOR, RESDORE has been shown to be able to eliminate the majority of corrupted elements with a minimum of false positives/false negatives arising from slice wise variations. Finally, if examining signal data with low directionality (i.e. $\leq 20 - 25$ gradient sampling directions), there is really no other choice than RESTORE or iRESTORE. While not explicitly examined within this chapter, personal experience has shown that the Richardson-Lucy deconvolution (upon which RESDORE is based) has difficulty resolving an accurate fODF with fewer than 30 gradient sampling directions and, while the 4th order spherical harmonics employed by HOMOR technically require only 15 data points to fit, one must remember that the elimination process involves removing corrupted elements and then refitting to the remaining set – a process that could easily result in martix condition issues if the 15 point minimum is not intially exceeded by a healthy margin.

In closing, we have outlined a method for the elimination of corrupted DW-MRI signal elements. In doing so we have demonstrated favourable performance against other comparable techniques and achieved the original goal of designing an elimination with a minimal propensity for false positive rejection.

Chapter 4

A Pitfall in the Reconstruction of Fibre ODFs Using Spherical Deconvolution

Prologue

This chapter covers the applicability of two high angular resoltion methods to fibres of varying anisotropy, namely constrained spherical harmonic deconvolution (CSHD, Tournier et al. (2007)) and damped Richardson-Lucy deconvolution (dRL, Dell'Acqua et al. (2010)). The basis for this work stems from two observations drawn from a preliminary maxillofacial tractography study: (1) While diffusion tensor imaging (DTI, Basser et al. (1994a), commonly used for muscle fibre tractography) was sufficient to reconstruct larger muscles such as the masseter, tractography through smaller (though equally important) muscles where thickness/width approaches voxel dimensions (e.g. the orbicularis oris) was regularly disrupted due to unavoidable partial volume effects (Alexander et al. (2005)) known to confound DTI. The obvious conclusion was therefore that higher order models must be used within such regions in order to recover meaningful orientations throughout the entire maxillofacial region. (2) Application of CSHD (a commonly used higher order technique) to muscle tissue, while accurately resolving crossing fibres in places, also produced spurious fibre orientation density function (fODF) peaks and bias in indicated orientations whose frequency and magnitude appeared to correlate with decreasing fractional anisotropy of the target tissue. After examining the default parameters it appeared differences between the estimated fibre response function (derived from the image by averaging $0.8 < FA < 1$ responses in accordance with recommendations, Tournier et al. (2004)) and the average muscle response (corresponding to an $FA \approx 0.3$) was the probable cause.

The work presented below is comprised of a simulation study in which the link between calibration discrepancy (defined as the difference between estimated and actual single fibre response functions) and accuracy of resultant fODFs is explored; the results of which clearly demonstrate that both calibration and indeed selection deconvolution method must be tailored to the target tissue in order to achieve optimal results – especially within diseased or non-typical (e.g. muscle) tissue (Parker et al. (2013a)). The following chapter will then go on to provide an example of the real world relevance of these findings, presenting a case study in the application of CSD to maxillofacial muscle, demonstrating the effects of (in)appropriate calibration and, for the first time, anatomically plausible reconstructions of numerous surgically relevant facial muscles. In combination these two chapters provide a theoretical basis for, and practical example of, the second stage in the envisioned robust, automated diffusion processing pipeline.

The material presented in this chapter was published as Parker et al. (2013a).

4.1 Introduction

In recent years diffusion weighted MRI (DW-MRI) (LeBihan et al. (1986)) has become a valuable tool for clinical and experimental neuroscience research being the only methodology for characterising tissue microstructure in vivo. To date, diffusion tensor MRI (DT-MRI) (Basser et al. (1994a)) is the most widely applied technique, providing useful quantitative microstructural indices such as mean diffusivity (MD) and fractional anisotropy (FA) (Pierpaoli and Basser (1996a)), and basic tract reconstruction schemes based on the orientation of the principal eigenvector (Basser et al. (2000)). However, the utility of DT-MRI is limited since the assumption of an ellipsoidal diffusion profile (the tensor model) prohibits resolution of more than one fibre orientation per voxel (Alexander et al. (2005)). Given that the majority of voxels at typical 'clinical' image resolutions (e.g. voxels with dimensions of 2-3 mm), contain multiple fibre populations (Behrens et al. (2007), Jeurissen et al. (2013)), this is a severe limitation.

Alternative methods such as diffusion spectrum imaging (DSI) (Wedeen et al. (2005)), Q-ball imaging (Tuch (2004)), persistent angular structure MRI (PAS-MRI) (Jansons and Alexander (2003)) and spherical deconvolution (SD) (Tournier et al. (2004)) have all been proposed to overcome the DT-MRI limitation. Data acquisition times for DSI are prohibitive, since sufficient data are required to reconstruct the full diffusion propagator. Q-ball imaging is a variant of DSI that, through the use of the Funk-Radon transform, allows peaks in the diffusion orientation distribution function (dODF) to be found from data acquired at a single b-value. However, peaks in the dODF are not very sharp in comparison to the fibre orientation distribution function (Alexander et al. (2010b)); PAS-MRI and spherical deconvolution methods aim to obtain sharper estimates of the

peak in fibre orientation, the former requiring extensive computation time due to non-linear estimation requirements while the latter class of techniques facilitate fairly rapid fODF retrieval, though peak finding may take considerably longer depending on the desired accuracy/speed trade-offs. As spherical deconvolution approaches: (1) attempt to recover the fODF directly (as opposed to the dODF); (2) have acquisition requirements similar to that required for robust DT-MRI (Jones (2004)); and (3) have reasonable computation costs, they have become the methods of choice in our laboratory.

The assumption underpinning most spherical deconvolution techniques is that an observed DW signal is the result of the spherical convolution of an underlying fODF with a response function that characterises the diffusion-weighted signal from a single-fibre population. Response functions may either be estimated on a voxel-by-voxel basis (Anderson (2005)) or, for simplification, assumed to be a constant across the image – i.e., a single response 'calibration' step is performed. The two most common calibration techniques are: (i) averaging measured signals in a region of high FA (e.g. $FA \geq 0.8$ Tournier et al. (2004)); or (ii) by simulation of an idealised signal. Once a calibrated single fibre response function has been obtained, associated SD approaches proceed by deconvolving this response from the sampled DW signals (referred to here as the 'target') to estimate the underlying fODF. However, reconstruction of this fODF by spherical deconvolution is ill-posed and multiple solutions may exist, some of which may be physically implausible (such as negative peaks in the fODF). To address this, non-negativity (and typically non-small) constraints are often placed upon the deconvolution (Alexander (2005), Dell'Acqua et al. (2007), Tournier et al. (2007)) to improve fODF reliability.

Spherical harmonic deconvolution (SHD), and its constrained version (CSHD) are implementations of spherical deconvolution (Tournier et al. (2004), Tournier et al. (2007) respectively), in which it is assumed that both the observed DW signal and the single fibre response functions may be adequately represented by a (truncated) linear combination of spherical harmonic basis functions. Typically the series is truncated at the 8th order harmonic, requiring a total of 45 coefficients to be estimated (Tournier et al. (2009)). This both reduces the number of parameters that need be stored (reducing memory requirements) and, more importantly, reduces the deconvolution process to a computationally trivial operation in SH space (analogous to convolution/deconvolution in Fourier space), reducing overall processing requirements.

Where the assumption that all single fibre populations within an image volume exhibit the same diffusion profile is valid, CSHD can be shown to provide accurate estimates of fibre orientation (Tournier et al. (2008)). In practice, however, there are many situations in which this fundamental assumption is not valid. For example, within healthy white matter, neuro-development (Suzuki et al. (2003)) and normal inter-regional, inter-tract and intra-tract variations (Jones et al. (2005)) can result in significant differences between diffusion profiles within an imaged volume. While a change in measured FA

due to partial volume contamination will not necessarily alter the underlying diffusion profile, and a single profile assumption may still be applicable, regional variation of axon properties such as diameter/density (Abolitiz et al. (1992), Barazany et al. (2009), Alexander et al. (2010a)) or demyelination (Beaulieu (2002), Gulani et al. (2001), Song et al. (2002)) can result in genuine changes in the diffusion profile anisotropy (through increased radial diffusivity) of approximately 20%. This raises several important questions: Is it appropriate to use a single canonical fibre response function through the entire image volume? If not – what are the consequences of deviations from the assumed fibre response? While the assumption provides computational expediency, if there are deleterious consequences, should we consider sacrificing efficiency in favour of being robust against these deviations?

To address these questions, we first evaluate the performance of CSHD in the presence of a mismatch between the calibrated single fibre response and the fibre-response of the constituent fibre populations. We then compare the CSHD algorithm to dampened Richardson-Lucy deconvolution (dRL) (Dell'Acqua et al. (2010)) which is one of a number of alternative SD techniques that has identical acquisition requirements to CSHD, but rather than using a spherical harmonic basis representation, dRL performs deconvolution across a super-sampled (interpolated) fODF. The potential advantage of the Richardson-Lucy algorithm is that, although a single fibre response calibration is still required, the Richardson-Lucy deconvolution framework is known to tolerate imprecision in the initial response function estimates (Dell'Acqua et al. (2007)), potentially reducing the impact of imperfect response function calibrations.

This chapter therefore consists of a performance comparison between the two algorithms in the presence of a discrepancy between the calibrated and 'target' fibre responses (i.e. the DW-signal profile of the tissue in which the deconvolution process is to be applied). Specifically, we examine the impact of introducing a discrepancy between the anisotropy of the tissue used for calibration and the anisotropy of the fibres in the target tissue; examining signals representing both single and crossing fibre targets. While the motivation for the inclusion of crossing-fibre data should be self evident, the reasons for including 'simple' single fibre configuration are two-fold: Firstly, each time a new HARDI technique is developed, substantial efforts are expended to demonstrate its ability to recover the ODF in crossing-fibre configurations. Higher precision in the estimates of the orientation of constituent fibres and the ability to resolve smaller inter-population angles are the usual markers of efficacy. However, perhaps unsurprisingly, rarely do these investigations return to the seemingly trivial, yet essential, problem of reconstructing the ODF in single fibre populations. Such simple configurations are easily resolved by the techniques they aim to supersede (for example, the Gaussian tensor model performs adequately in single fibre populations). Secondly, characterisation of any errors in the resolution of a single fibre orientation will provide an easier way of

identifying any underlying systematic problems inherent in certain techniques, allowing us to diagnose their cause.

4.2 Methods

4.2.1 Data Simulation

Datasets were simulated (zero-mean Gaussian test function - Cook et al. (2006)) assuming a typical 60 direction sampling scheme (Jones et al. (1999)), $b = 2000s/mm^2$, and for four signal-to-noise ratios (SNR): infinite (i.e., noise-free), 50:1, 30:1 and 10:1 (consistent with Tournier et al. (2004)). The sixty diffusion-weighted signals were computed for:

i A single, prolate, axially-symmetric tensor with fixed mean diffusivity ($0.7{\times}10^{-3}mm^2/s$) and varying fractional anisotropy (FA) aligned along the x-axis

ii A pair of "crossing fibres" simulated as the signal produced by two prolate, axially-symmetric tensors the same FA, varying crossing angle (10-90° , 5° interval, one aligned along the x-axis, the other rotating about the z-axis) and signal contribution ratios (20/80% to 80/20%, 5% interval, first number represents signal contribution from the axially aligned fibre, second from rotated fibre)

iii A single, prolate, axially-symmetric tensor with fixed mean diffusivity ($0.7{\times}10^{-3}mm^2/s$), fixed FA (0.3) and aligned along each of 752 axes evenly distributed on the unit sphere.

In simulation (i), when the SNR was infinite, the FA was varied between 0.1 and 0.9 at a 0.01 interval (81 individual profiles). For other SNR's in simulations (i) and (ii), the FA interval was increased to 0.1 and 500 repetitions of the noisy DW-signals were calculated per FA value (9×500 total). In simulation (iii), 500 noisy repetitions at SNR=50:1 were generated for each of the 752 orientations (752×500 total). Using single-fibre 'calibration' responses from idealised (prolate axially symmetric) tensors with FA between 0.1 and 0.9, the peaks in the fODF and their associated magnitudes (fODFmag) were then extracted from all the simulated signals using both CSHD and dRL.

The noise-free data provides a 'high resolution' (81×81) array of fODFs, to gain a clear overview of the onset of any systematic failures that are not attributable to noise, while the noisy data allow assessment of the reproducibility of such artefacts at different SNR levels

CSHD Analysis: All CSHD analysis was performed using the algorithm exactly described by Tournier et al. (2007) and as implemented in the ExploreDTI software package (Leemans et al. (2009); see also Jeurissen et al. (2011)). The harmonic series were truncated at the 4th, 6th and 8th orders (referred to as Lmax), requiring 15, 28 and 45 free parameters to fit, respectively, allowing us to explore impact of different truncations.

dRL analysis: All dRL analysis was performed using the algorithm described by Dell'Acqua et al. (2010)) with deconvolution performed over 752 uniformly distributed fODF interpolation points (optional SH or radial basis function interpolation was not applied, and peaks were retrieved through an exhaustive search). A simple $(\cos 2\theta)$ response function with longitudinal and radial diffusivities of an axially symmetric prolate tensor with $MD = 0.7 \times 10^{-3} mm^2/s$ and the desired FA was used. Deconvolution was allowed to iterate 200, 300 and 400 times to explore the trade off between improving the angular resolution and over-fitting to noise within a runtime comparable to that of CSHD. Geometric and regularization threshold parameters (Dell'Acqua et al. (2010)) were set to 8 and 0.04 respectively.

DTI Analysis: The principal eigenvector of the best-fit single tensor was derived from the raw DW signal data according to Basser et al. (1994a).

4.2.2 Data Analysis - Single–Fibre, Axially-Oriented

The performance of the two SD algorithms was assessed according to three criteria: (1) the deviation between the peak of the reconstructed fODF and the principal eigenvector of the simulated tensor; (2) the magnitude and orientation of spurious peaks in the fODF; and (3) the expected 'failure' rate of implied fODFs if used by tractography algorithms.

4.2.2.1 Error in Fibre Orientation Estimates

For a single fibre population, the orientation of the largest peak in the fODF (the "primary peak") should provide the best estimate of the true fibre orientation. We therefore characterised the angular deviation between the orientation of the primary peak and the simulated fibre orientation for all calibration/target (C/T) pairings; defining the entire set of pairings as C/T space. For the noise-free data, the comparison was straightforward. For the noisy data, for each SNR and C/T pairing, the average orientation of the 500 estimates of primary peak orientation was computed by taking the principal eigenvector of the associated scatter matrix along with the 95% confidence interval (Jones (2003)) to quantify orientational uncertainty.

4.2.2.2 Spurious Peak Manifestation and Orientation

When using SD techniques to recover the fODF, the reconstruction of 'spurious' peaks is a common problem which is particularly problematic for tractography algorithms, i.e. to distinguish between artefactual and 'real' trajectories. To address this, the standard practice is to apply some form of threshold on the magnitude of the fODF peaks, so that only 'large' peaks are retained.

For the noise-free data (Simulation (i)), results were sufficiently dense (i.e., 81×81 pairings) that the maps of the magnitude of the spurious peaks were amenable to gradient edge detection, for which the Canny operator (Canny (1986)) was employed to identify and highlight boundaries in the C/T space.

To look for patterns in the relative orientation of spurious peaks with respect to the primary peak, for each C/T pairing, we first thresholded spurious peaks by magnitude ($fODFmag > 0.1$ for CSHD in keeping with standard literature practice, $fODFmag > 10\%$ of maximum for dRL). The spurious peaks were then visualised on the unit sphere and, as discussed below, were seen to fall into discrete clusters. k-means clustering (where k was selected on the basis of a visual inspection) was subsequently applied to extract mean angular orientations.

4.2.2.3 Expected Tractography Failure Rate

To illustrate the practical implications of inappropriate calibration, the frequency of occurence of a 'severe' tractography failure was derived for a standard 'Euler-like tractography algorithm (Basser et al. (2000)). Here failure was defined as: (1) an incoming fibre trajectory subtending an angle with a supra-threshold spurious peak that is below any angular threshold used as a termination criterion (so that the tracking algorithm follows this spurious peak); (2) false negative or (3) false positive results where bias in fibre orientation causes erroneous fibre reconstruction.

To derive the frequency of these occurrences, 100 uniformly distributed (Jones et al. (1999)) axes were considered as simulated incoming fibre trajectories. For each C/T/SNR combination (excluding noiseless data), the percentage of incoming trajectories resulting in a failure (as defined above) was defined for each of the 500 repetitions, and the mean computed. The angular threshold (i.e., the maximum angle through which the reconstructed streamline can turn between successive steps) was set to $35°$.

4.2.3 Data Analysis - Single–Fibre, Spherically Distributed

The data from simulation (iii) allowed determination of whether patterns in observed spurious fODF peaks (see subsection 4.2.2.2) are orientationally variant. The target FA

(T) was simulated at a fixed value of $FA = 0.3$, and calibration FA (C) was fixed to $FA = 0.9$ – a combination known to produce spurious fODF peaks within the axially-aligned data. For each simulated fibre orientation the 95% cone of uncertainty in primary peak orientation was estimated along with the mean number of spurious fODF peaks and the orientations of spurious peaks ($fODFmag > 0.1$) relative to the primary peak. These measurements provide insight into the orientational dependence of C/T discrepancy-driven uncertainty in the primary fibre orientation, and allow determination of whether the manifestation of spurious peak patterns is a general phenomenon, or whether they appear more frequently for a given orientation (such as the orientation selected in simulation (i)).

4.2.4 Data Analysis – Crossing Fibres

In the crossing-fibre case, performance was assessed according to two criteria: (1) the effect of miscalibration on angular resolution; (2) the effect of miscalibration on sensitivity to differential fibre signal contribution (i.e., changing the relative compartmental volume fractions).

4.2.4.1 Angular Resolution

A widely used performance metric for comparison of HARDI techniques is the minimum angle at which two distinct fibre populations may be reliably distinguished given comparable acquisition parameters, i.e., the angular resolution. To isolate the effects of C/T discrepancy we examined the subset of crossing-fibre data in which both fibres have equal FA and contribution to the diffusion-weighted signal, limiting the simulation variables to crossing angle, a single target anisotropy and calibration anisotropy. Here succesful trials are regarded as resolving two supra-threshold fODF peaks (0.1 absolute magnitude CSHD, 10% of maxima dRL) within a 20° (Alexander (2005)) cone of error about the expected orientations.

4.2.4.2 Volume Fraction

In the previous subsection it was assumed that both fibre populations contribute equally to the DW-signal. In reality this is unlikely to be the case and so performance of the two algorithms was examined as compartmental volume fractions were varied. To simplify results, the examined fibre configuration consisted of an orthogonal crossing with both fibres having the same anisotropy. The variable parameters were the relative volume fractions, the single target FA and the calibration FA. The same success criteria defined in Section 4.2.4.1 was used.

4.2.5   A Word on fODF Thresholding

Note that in some cases an fDOF threshold of 0.1 was applied to CSHD while no threshold was applied to dRL. From one perspective, this may be considered to be creating an unfair comparison (dRL may appear artificially worse). However, this practice was adopted for two reasons: (i) Under miscalibration CSHD produces a large number of spurious peaks, however, a proportion of these peaks are of such insignificant magnitude that they will not survive the fODF thresholds typically applied in tractography algorithms, and would therefore be unlikely to cause any practical problems in tract reconstruction. Thus to present a fairer picture of CSHD for its real-world application in tractography, it is entirely appropriate to apply an fODF threshold;. (ii) Conversely, almost all spurious peaks produced by dRL (against a single-fibre target) are insignificant in magnitude. By eliminating the threshold we are able to highlight any pattern of errors occurring with the dRL algorithm. For clarity, it is clearly stated where differential thresholds are applied.

4.3   Results

4.3.1   Single Fibre Orientation Error: Noise Free Data

Despite the absence of noise in the first set of simulated data (simulation (i)), both SD approaches produced erroneous results. The damped Richardson-Lucy algorithm results in a constant bias (angle between primary peak of the fODF and the simulated fibre orientation) of $1.13°$ (Fig.4.1b) regardless of the number of iterations used in the deconvolution. This is consistent with the angle subtended between the simulated fibre orientation and the closest interpolated sampling point on the sphere, and is therefore an artefact attributable to the finite angular resolution afforded by discrete sampling (as compared to continuous differentiable functions, such as those used in CSHD).

Constrained spherical harmonic deconvolution, on the other hand, produces inconsistent errors, whose occurrence is dependent on the discrepancy between the calibration and target response functions. (Fig. 4.1a). The majority of C/T pairings produce no error; it is only those pairings that lie between two distinct linear boundaries that lead to a significant bias in estimated fibre orientation (attenuated by reducing Lmax from 8 to 4). Note that instead of a horizontal boundary – which would imply that the bias is related to the target anisotropy alone, the bias in fibre orientation increases as the calibration discrepancy increases (i.e., calibration FA exceeding target FA).
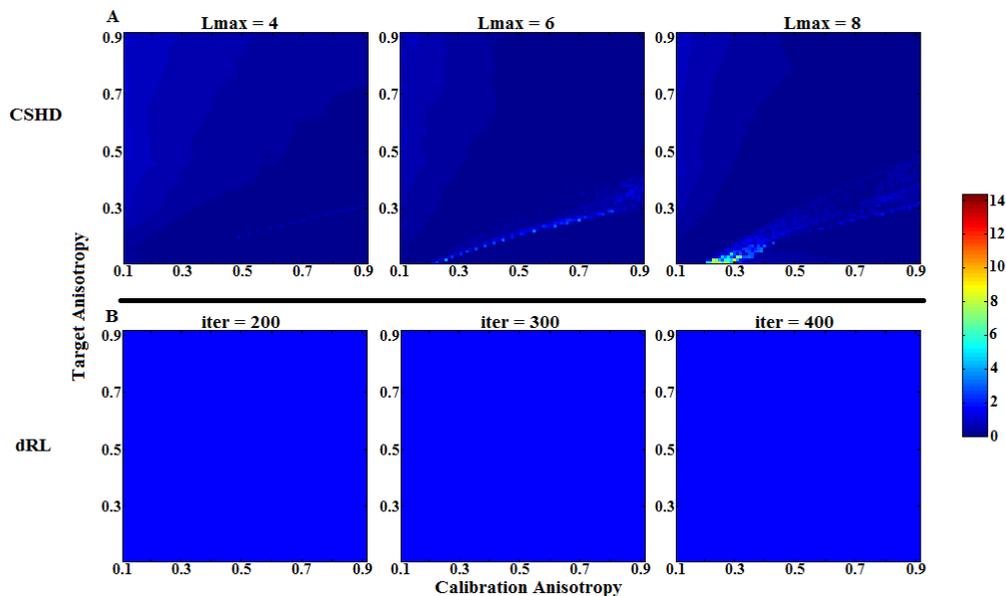
FIGURE 4.1: Angular bias (degrees) between the primary fODF peaks and simulated fibre orientations for (A) CSHD and (B) dRL. Note dRL possesses a constant non-zero bias.

TABLE 4.1: Mean Band Orientations

SNR 10	Cluster 1	Cluster 2	Cluster 3
Lmax 4	90°		
Lmax 6	71.4°	108.12°	
Lmax 8	59.49°	90°	120.06°
SNR 30	Cluster 1	Cluster 2	Cluster 3
Lmax 4	90°		
Lmax 6	69.8°	109.8°	
Lmax 8	56.77°	90°	122.46°
SNR 50	Cluster 1	Cluster 2	Cluster 3
Lmax 4	90°		
Lmax 6	68.75°	110.53°	
Lmax 8	54.44°	90°	124.97°
Predictions	Cluster 1	Cluster 2	Cluster 3
Lmax 4	90°		
Lmax 6	65.13°	114.89°	
Lmax 8	51.24°	90°	128.78°

4.3.2 Single Fibre Orientation Error: Noisy Data

Both SD algorithms produce noisy estimates of fibre orientation that are symmetrically distributed about the expected (simulated) orientation (Fig.4.2a-b), suggesting minimal to no systematic bias in the estimate of primary peak orientation. However, CSHD (Fig. 4.2c) results do exhibit increasing uncertainty as the calibration disparity increases, and the uncertainty increases with an increase in Lmax. In contrast to the noise-free data, the number of C/T pairings was insufficient to facilitate a meaningful linear regression

to identify boundaries in the results; however, visual inspection suggests that there is a similar linear boundary between affected and unaffected C/T space regions.

dRL results (Fig. 4.2d) show a similar trend towards increased uncertainty as calibration disparity increases, although the border between affected and unaffected regions does not appear linear and, perhaps more importantly, dRL appears to recover more quickly as SNR improves. It is interesting to note that, unlike CSHD, dRL results show an increased mean bias (though not uncertainty) in regions where T exceeds C and SNR remains high.

DT-MRI results (Fig. 4.4a-b) vary only with SNR and fibre anisotropy (reflecting the widely understood 'noise bias' (Pierpaoli and Basser (1996b), Jones (2010)). It is worth noting that at higher SNRs, in areas containing only single fibre populations with low anisotropy ($0.1 < FA < 0.4$), the uncertainty in estimates of principal fibre orientation is lower with DT-MRI than with either CSHD or dRL when the calibration discrepancy is large.
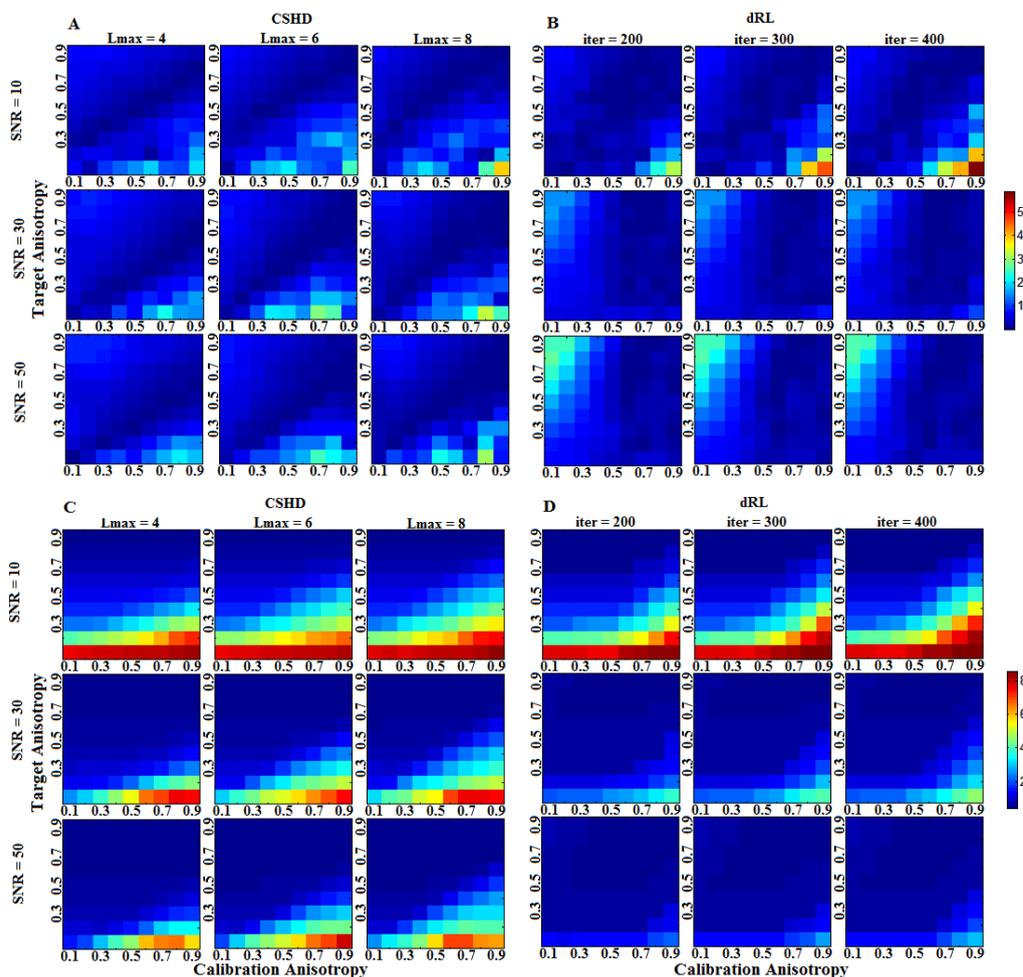
FIGURE 4.2: Upper: Bias in primary peak orientation of (A) CSHD and (B) dRL derived fibre orientation estimates. Lower: 95% confidence intervals in (C) CSHD and (D) dRL fibre orientation estimates.
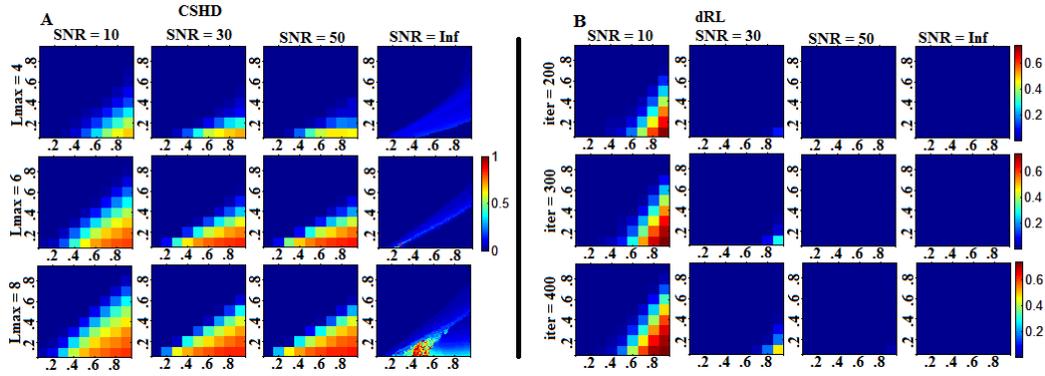
FIGURE 4.3: Mean magnitude of largest spurious peak across a range of SNR for fODF estimates produced by (A) CSHD and (B) dRL. To facilitate an easier comparison, spurious peak magnitudes have been normalised by that of the related primary peaks.

4.3.3 Single Fibre Spurious Peak Magnitude

As might be expected, the appearance of spurious peaks in the CSHD-derived fODFs (Fig. 4.3a, $SNR = \infty$) coincides with the deviations in primary peak orientation (Fig. 4.1a). Regression of linear boundaries (at $Lmax = 8$) suggests that the magnitude of the secondary peak falls into three domains: (1) C/T pairings where $T > 0.667C$ - spurious peaks exist but none are of sufficient magnitude to disrupt tractography algorithms; (2) C/T pairings where $0.667C > T > 0.5C$ - spurious peaks are discernible but mean magnitudes remain sub-threshold; (3) C/T pairings where $T < 0.5C$ - a substantial number of spurious peaks is produced (Fig. 4.8a) with sufficient magnitude to significantly corrupt results from tractography. The addition of noise (Fig. 4.3a) demonstrates that, as SNR decreases, the onset (with respect to calibration discrepancy) and magnitude of secondary peaks increases while following a similar pattern (linear boundaries between affected and unaffected regions). Lowering Lmax from 8 to 4 reduces the number of affected C/T pairings and the resulting spurious fODF peak magnitude (in both absolute and normalised terms – note Fig. 4.3 represents fODF data normalised by the primary peak magnitude).

Under noise-free conditions (Fig. 4.3b, $SNR = \infty$) dRL produces no spurious peaks. At high SNR, dRL will only produce spurious peaks under extreme C/T discrepancies, but, as SNR decreases, spurious peaks begin to form in regions where $C > T$. Increasing the number of algorithm iterations increases the number of C/T pairings which produce spurious fODF peaks and increases existing spurious peak magnitudes.
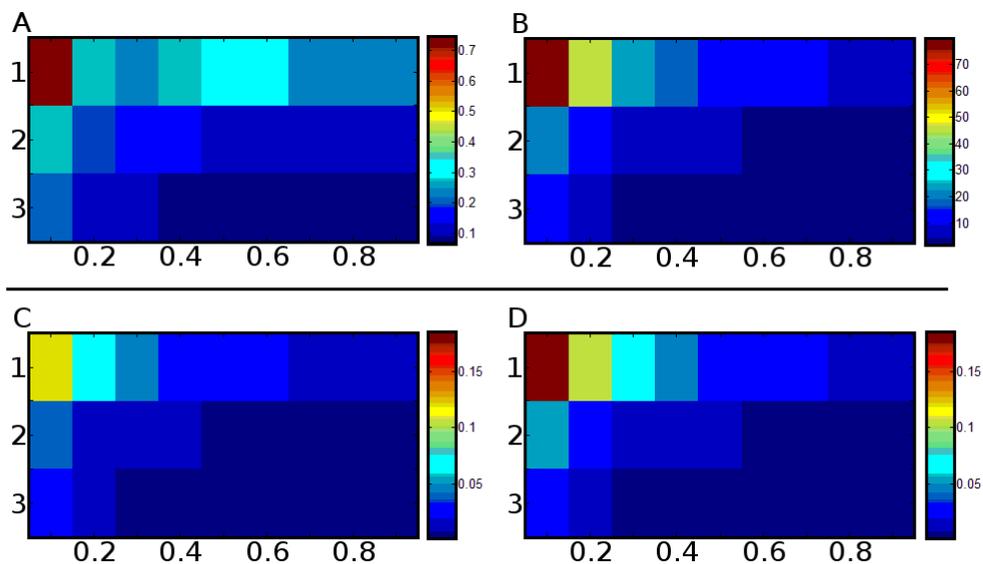
FIGURE 4.4: DTI performance. Rows 1, 2 and 3 correlate with SNRs of 10, 30 and 50 respectively. (A) Error (degrees) in mean orientation estimate. (B) 95% confidence intervals about mean orientation estimate. (C) Expected tractography error rate (percentage [0 1]) at a 25° angular threshold. (D) Repeat of (C) at 35° threshold.

4.3.4 Single Fibre Spurious Peak Orientations

Figure 4.5a shows the elevation angles of CSHD-derived spurious peaks (thresholded $fODFmag > 0.1, SNR = 50$) plotted relative to primary peak orientation. These aggregated peaks (harvested from all C/T pairings) are not randomly distributed over the sphere, but instead demonstrate 'banding' with the number and location of each band dependant on both Lmax and, to a lesser degree, SNR (Table 4.1) – observations which also hold true for individual C/T pairings (Fig. 4.6a). It is important to note, however, that azimuthal values (Fig. 4.6b) do appear to be random in distribution.

In comparison to CSHD, dRL does not produce consistent banding patterns (Fig. 4.5b). While there is a slight preponderance for spurious peaks to occur at a 90° elevation at the lowest SNR ($SNR = 10$), the number of spurious peaks is greatly reduced as SNR improves, suggesting that these are noise-induced (rather than algorithm-induced). The reduction in number of spurious peaks with increasing SNR becomes more marked as the number of iterations in dRL is increased. It is important to remember that the dRL results were not filtered through an fODFmag threshold, which would eliminate the majority of the observed spurious peaks.
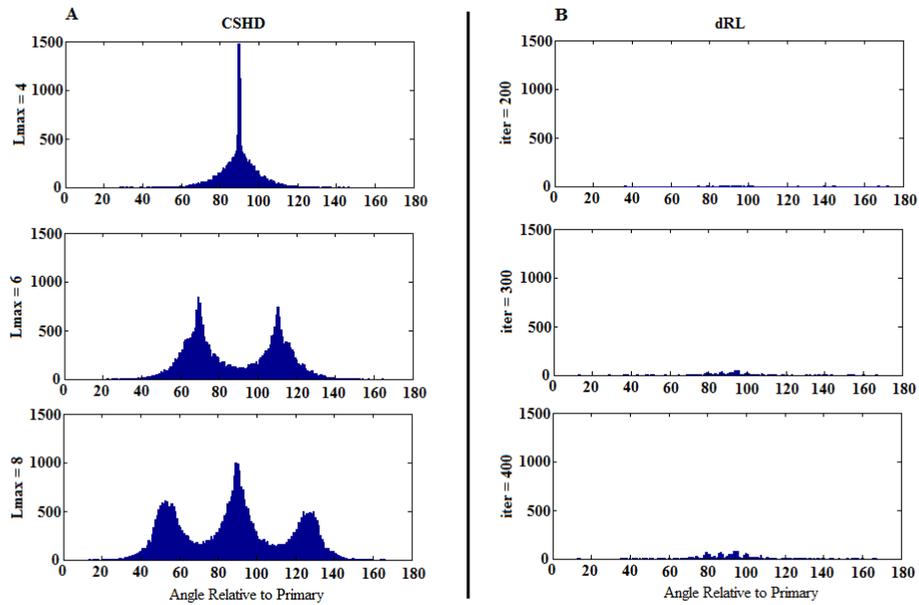
FIGURE 4.5: Aggregated elevation angles of artefactual peaks in $SNR = 50$ data. Left: CSHD. Right: dRL. Note the distinct structure within spurious CSHD peaks that shifts in relation to Lmax. For this figure CSHD peaks have been thresholded ($fODFmag > 0.1$), dRL results have not.
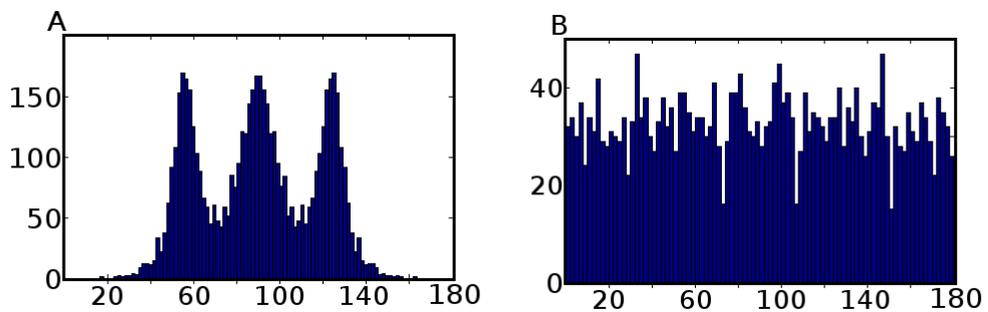
FIGURE 4.6: A: Distribution of spurious peak elevations ($fODFmag > 0.1$) relative to primary peak orientation ($C/T = 0.7/0.3, SNR = 50, Lmax = 8$). Note the concentrations about 54° , 90° and 127°. B: Azimuthal values across the same data, notice that the distribution remains approximately even across the entire range (180–360° symmetry omitted).

4.3.5   Single Fibre Expected Tractography Failure Rate

As might be expected, reducing the SNR increases the frequency of occurrence of severe failures in tractography (Fig. 4.7). Interestingly, however, while CSHD results show that an increase in SNR leads to a decrease in the number of C/T pairings likely to cause significant error (in line with previous results), the frequency of occurrence within affected regions of C/T space increases substantially as SNR increases, which appears counter-intuitive. However, while a lower SNR produces spurious peaks which are larger in magnitude (Fig. 4.3a), the absolute number of peaks per C/T pairing, on average, is reduced (Fig. 4.8a).

Assuming a roughly uniform distribution of spurious peaks (with respect to azimuthal angle, Fig. 4.6b), the 'spherical surface area' covered by a set of spurious peaks (and their associated 35° angular threshold cone) will be directly related to the size of the set. Thus, as the number of spurious peaks increases, so will the error rate. As Lmax is reduced, both the number of spurious peaks (Fig. 4.8a) and the number of elevational bands in which they reside also decreases (to a single sharp band at 90° at $Lmax = 4$), reducing the potential surface area coverage and therefore explaining the reduction in error rates compared to those observed at Lmax=8.
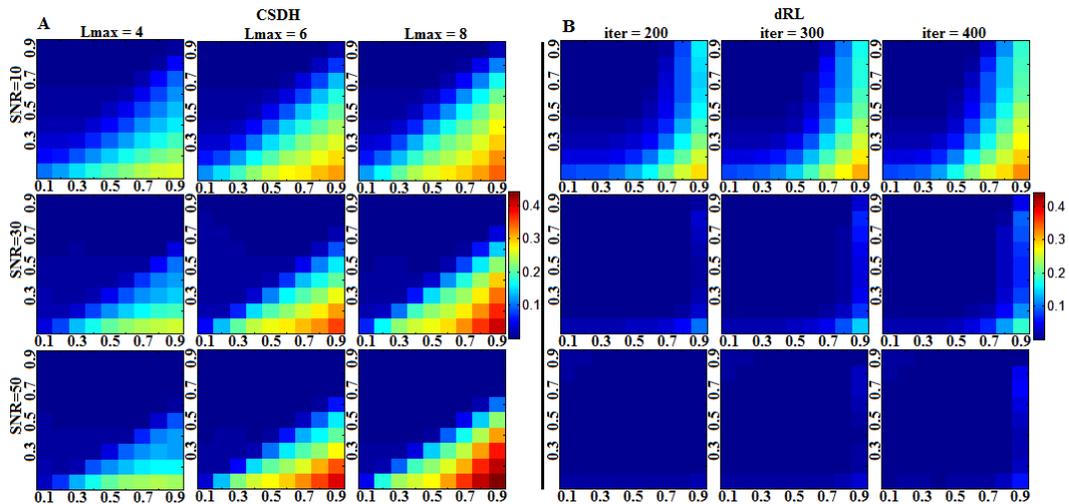
FIGURE 4.7: Approximate probability that an arbitrary incoming fibre trajectory will result in erroneous tractography through CSHD (A) and dRL (B) derived fODF estimates at a 35° angular tracking threshold. For this figure CSHD peaks were subject to a $fODFmag > 0.1$, dRL was unfiltered.

A different pattern is seen with dRL which, as previously shown, recovers to near noise-free results as the SNR improves, i.e. spurious peaks occur infrequently, with low magnitude (Fig. 4.3b, 4.5b, 4.8b), and with low uncertainty in the primary peak orientation (Fig. 4.2b). Even without thresholding, the expected error rates are substantially lower for dRL compared to CSHD (for comparable C/T pairings) in all but the lowest SNR case. Once again, however, increasing the number of algorithm iterations in dRL leads to a poorer result at high ($C = 0.9$) anisotropy calibration.

Given that the production of multiple spurious peaks through DT-MRI is an impossibility, (since there is only one principal eigenvector), failure rates (Fig. 4.4c,d) depend solely on the error in primary peak orientation. In this single fibre population case, DT-MRI results compare favourably with the HARDI alternatives, outperforming CSHD at low SNR/anisotropy once calibration discrepancy is introduced.
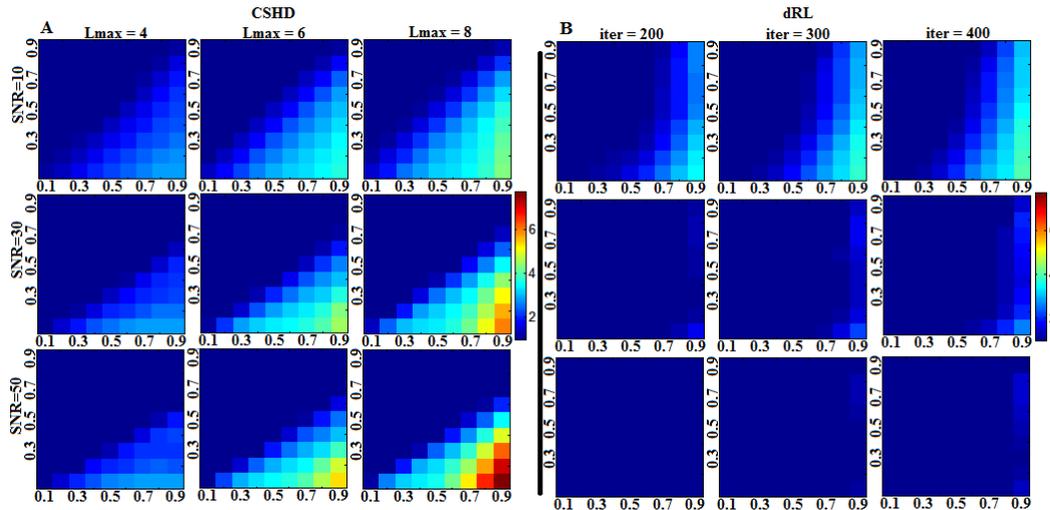
FIGURE 4.8: Mean number of supra-threshold ($CSHD \geq 0.1$, $dRL \geq 10\%$) spurious fODF peaks using (A) CSHD and (B) dRL.

4.3.6 Single Fibre Orientational Variability

Figure 4.9a indicates that for the majority of simulated fibre orientations, primary peak confidence intervals remain consistent with values predicted by similar axially aligned data (Fig. 4.3a). Note, however, that there is a marked increase in uncertainty for fibres aligned to the z-axis. This pattern is echoed closely by the mean spurious peak counts (Fig. 4.9b), the observed numbers closely match the axially-aligned data (Fig. 4.8) and variation (though small) appears to follow primary peak uncertainty.

Shifting focus to the relative distribution of spurious peaks (Fig. 4.9c-e); Figure 4.9c contains the aggregate of all spurious peak orientations and clearly shows the same characteristic Lmax dependent banding as seen in the axially-aligned data described in Section 4.3.5 (Fig. 4.5), suggesting that the orientation of spurious peaks relative to the main peak orientation is independent from the simulated peak orientation itself. This can be confirmed by examining individual cases such as Figure 4.9d - an example with typical primary peak confidence (fibres simulated along a [0.5774, -0.5774, 0.5774] direction vector) - and Figure 4.9e - a 'high uncertainty' example with z-axis alignment [0, -0.1114, 0.9938]). In general the described banding patterns appear stable across all simulated fibre orientations, although there may be some orientational dependence of intra-band dispersion (i.e., bands appearing wider at certain orientations), the reason for this is unclear and remains a question for future study.
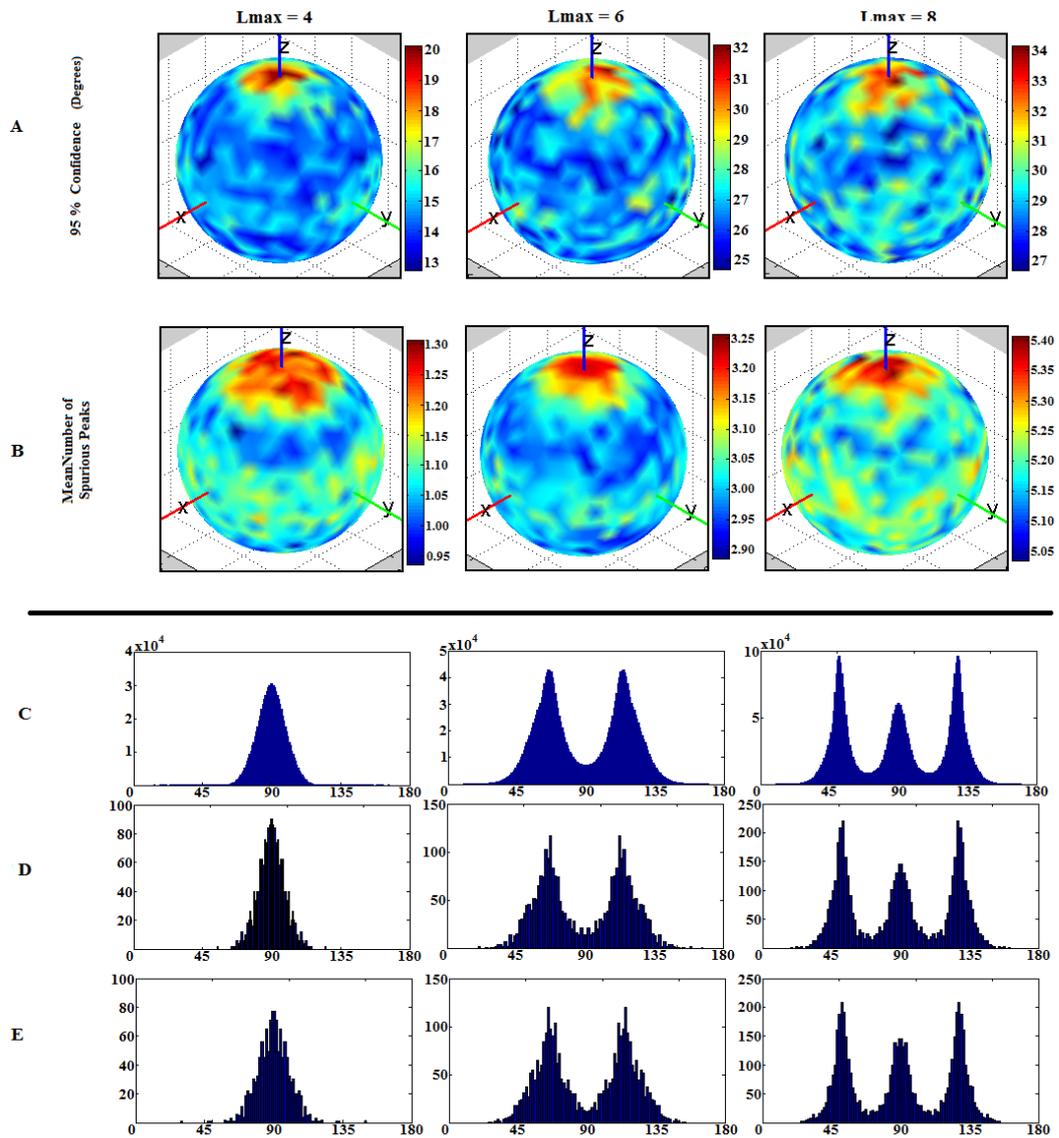
FIGURE 4.9: Impact of varying fibre orientation (simulation (iii)) on CSHD. (A) 95% confidence intervals (degrees) in primary peak orientation. (B) Mean number of spurious peaks. (C) Aggregated distribution of all spurious peak orientations relative to their primary peak. (D) Distribution of spurious peaks selected from a single fibre orientation chosen for its 'typical' 95% confidence interval and mean spurious peak count. (E) Distribution of spurious peaks from an a-typical region.
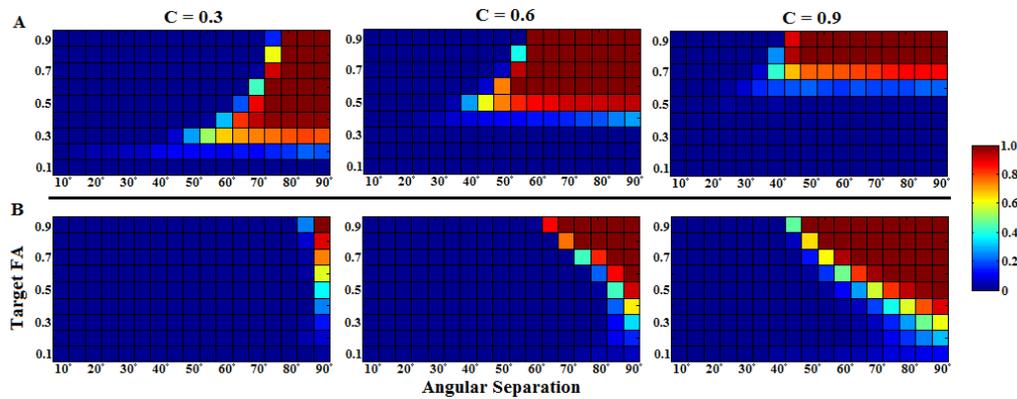
FIGURE 4.10: Crossing fibre angular resolution: plot of crossing-angle (horizontal axis) against target FA (vertical) for $SNR = 50 : 1$ data over a range ($C = 0.3$, $C = 0.6$ and $C = 0.9$) of calibrations; colour indicates probability of successful fODF peak retrieval (scaled 01). (A) CSHD results, $Lmax = 8$. (B) dRL results, $iter = 200$.

4.3.7 Crossing Fibre Angular Resolution

Two inferences can be drawn from the CSHD data (Fig. 4.10a, 4.11a-c): (1) Optimal results are generally achieved by selecting a calibration that closely matches the target fibre (Fig. 4.10a). Underestimating target FA results in a gradual decline in angular resolution, while overestimation results in a sharp fall. There are exceptions, particularly at low SNR (Fig. 4.11a, SNR=10, T=0.9) where, for all but $Lmax = 4$, calibrations slightly lower than the target anisotropy yield better angular resolution; (2) Increasing Lmax yields the expected increase in angular resolution, but does so at the cost of greater sensitivity to miscalibration. For example, at $Lmax = 4$ a calibration of 0.9 produces far fewer spurious results than at an Lmax of 6 or 8 (Fig. 4.11b). This is not because fibre orientations are unresolved, but - as in the single-fibre case- $C > T$ pairings produce an abundance of spurious fODF peaks that foul the strict two-peak success criterion. Temporarily disabling this constraint (Fig. 4.13a) demonstrates this point more clearly as success rates for overestimated calibrations duly increase.

Figures 4.11d-f highlight the differences between CSHD and dRL. As with the single-fibre configurations, a highly anisotropic calibration response function produces the optimal results in most situations. The one exception is at low SNR (Fig. 4.11d) where a (slightly) lower anisotropy provides more resilience to noise. Unlike CSHD, spurious peak formation only accounts for a low proportion of observed failures with dRL (generally this only happens at extreme calibration discrepancies at low SNR. Note that Fig. 4.13b shows improvement at $SNR = 10$) and is more likely to be due to failure to resolve more than one fODF peak or recovery of two peaks with large biases (Fig. 4.13c, but also notice minimal improvements in Fig. 4.13b - $T = 0.6, SNR = 30$). Finally, it is possible to improve the angular resolution of medium FA calibrations through increasing the number of algorithm iterations (Fig. 4.11d, e).

Overall, CSHD appears to be the superior choice for low FA targets (compare Fig. 4.11c with f), low SNR performance appears similar (Fig. 4.11a, d) and as SNRs and target anisotropies increase, the difference between best-case results is minimal. However, it must be noted that CSHD will require careful monitoring of target anisotropy to recover optimal results, while dRL generally will not. The figures referenced in this subsection are exemplars to demonstrate key results. A comprehensive set of results covering a wider range of C/T/SNR tuples is available in Appendix C.
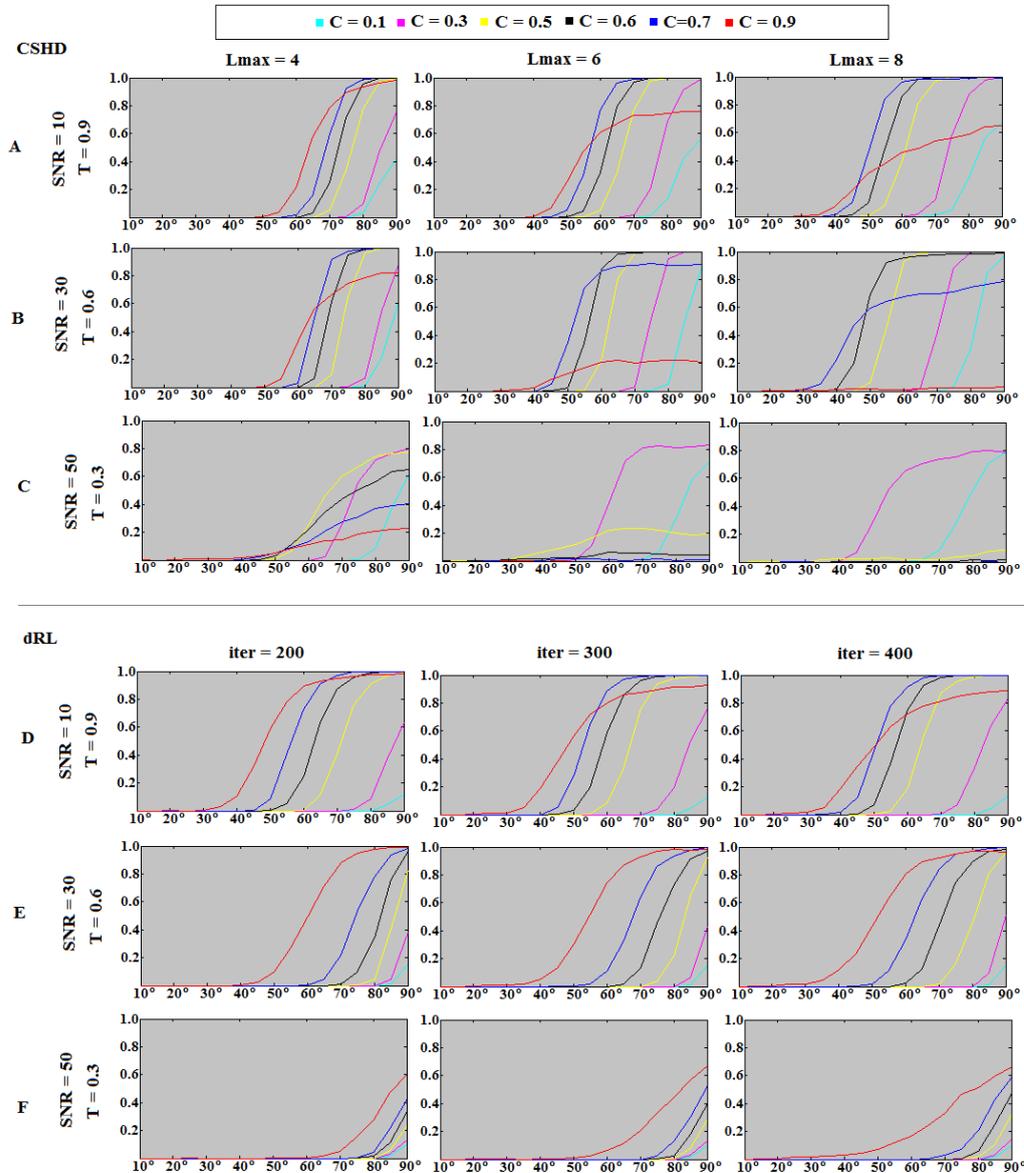
FIGURE 4.11: Crossing fibre angular resolution: fraction of successful trials (vertical, parametrised 0–1) vs. angular separation (horizontal, degrees). (A-C) Selection of CHSD results. (D-F) Selection of dRL, test conditions identical to corresponding CSHD plots.

4.3.8 Crossing Fibre Volume Fractions

The results from varying the volume fraction are presented in Fig. 4.12. For CSHD (Fig. 4.12a-c), the best results are achieved through matching calibration to target anisotropy with the caveat that, at low SNR, a slightly reduced calibration anisotropy or lower Lmax may result in improvements (Fig. 4.12a, notice how $C = 0.9$ performance deteriorates as Lmax increases while the $C = 0.7$ calibration improves). Furthermore, increasing Lmax again allows one to resolve increasingly smaller signal contributions at the expense of requiring a more target specific calibration (compare $C = 0.7/0.9$ results at $Lmax = 4$ to $Lmax = 8$, Fig. 4.12b).

At low anisotropies, the performance of dRL again appears to be inferior to that of CSHD (Fig. 4.12d-f). This may be most clearly observed by comparing Figures 4.11c and f, but subtle differences in performance are also discernible in Fig. 11R, where dRL performance is inferior in edge case performance (20/80, 30/70 volume fraction ratios) against correctly calibrated CSD (Fig. 4.11b). While low SNR performance remains comparable with CSHD, as with angular resolution, this is the one case in which a lower FA calibration may be preferable (Fig. 4.11D). As both SNR and target anisotropy increase it is possible to achieve best-case results through use of a single $C = 0.9$ calibration. Though as with angular resolution, if necessary it is possible to achieve improvements using lower FA calibrations (Fig. 4.12d) or against lower FA targets by increasing the number of algorithm iterations (Fig. 4.12E, examine C=0.9 as iteration count increases). Again, a more comprehensive result set is provided in Appendix C.

4.4 Discussion

Results show that under the majority of circumstances, both CSHD and dRL will produce incorrect fODF estimates when calibrated to an inappropriate fibre response function (the exception being noise-free single-fibre dRL). It is clear that dRL, however, excepting the noisiest data or most extreme calibration discrepancies should not produce spurious peaks in sufficient quantity (Fig. 4.8) or magnitude (Fig. 4.3) to result in significant errors in tractography in regions containing single fibre populations (Fig. 4.7). However, in regions containing crossing fibres, dRL performance deteriorates when the target FA is low. CSHD, on the other hand, performs well across a range of target FA values in regions containing crossing fibres, but is prone to producing spurious fODF peaks in all cases as the calibration anisotropy begins to exceed that of the target and a loss of angular resolution where calibration underestimates the target.
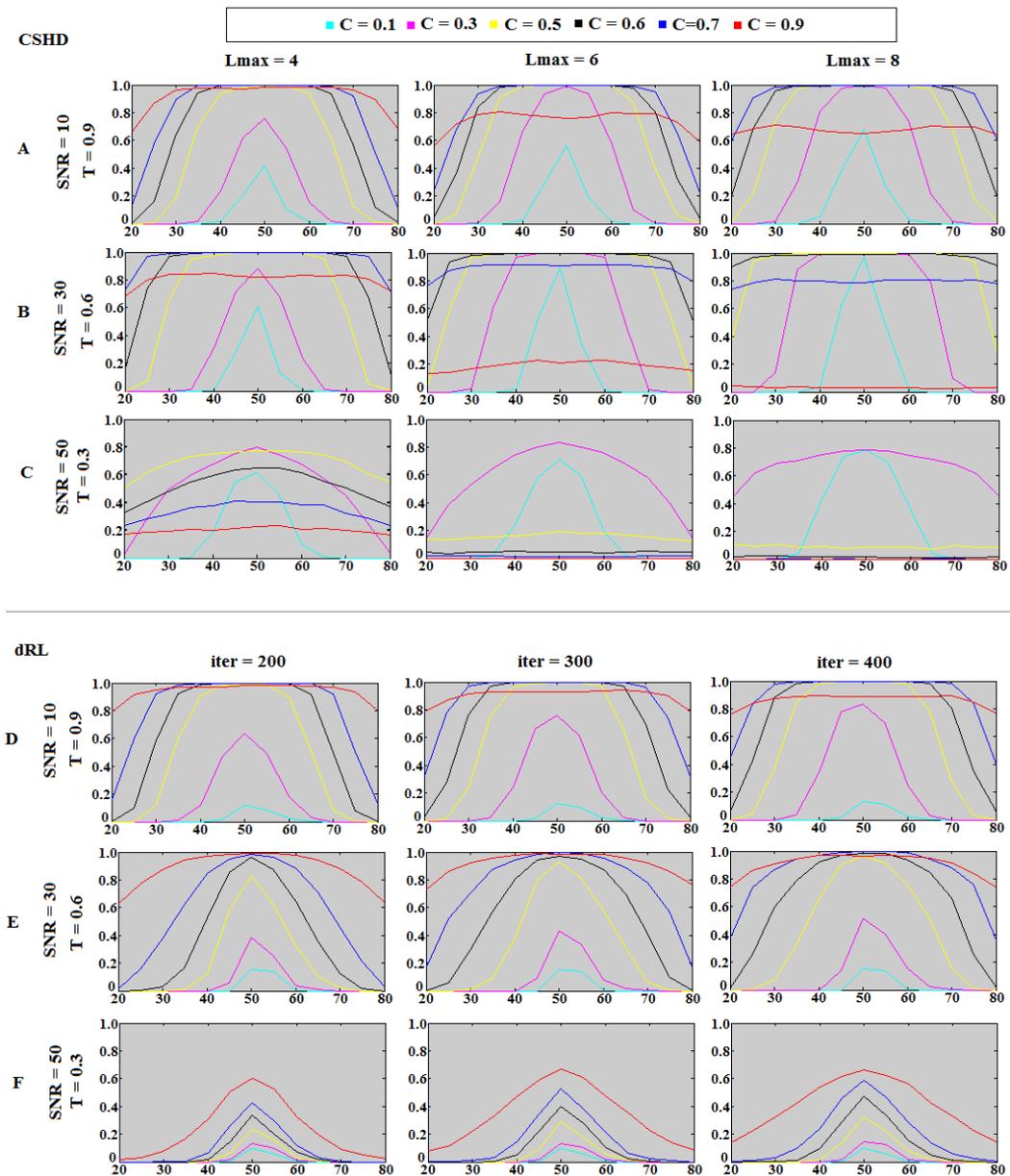
FIGURE 4.12: Crossing fibre volume fraction results: percentage of successful trials (vertical) against relative contribution (horizontal axis, ticks denote percentage of signal provided by axially aligned fibre component). (A-C) Selection of CSHD results. (D-F) Corresponding selection of dRL results.

Figure 4.1, derived from noise-free CSHD data, shows a distinct linear boundary ($T = 0.66C$, $Lmax = 8$) between those C/T pairings that produce artefacts and those that do not. With the addition of noise (Fig. 4.2) it is seen that, on average, mean orientations are consistent with our expectation (simulated fibre orientation), and that individual estimates are distributed about the expected axis with no systematic bias. The result of calibration discrepancy therefore appears to be increased uncertainty in fibre orientation related to both SNR and the magnitude of the discrepancy, To illustrate the impact of varying the SNR in isolation, you might consider a toy example where the FA of

calibration and target are identical ($C = T$), but the SNR varies. For $C = T = 0.6$ ($Lmax = 8$), the 95% confidence intervals in fibre orientation at SNR's of 10, 30 and 50 are 6.0°, 2.8° and 1.3° respectively. For $C = 0.9, T = 0.6$, placing the C/T pairing on the $T = 0.66C$ border observed in noiseless data, the respective 95% confidence intervals are 14.4°, 8.5° and 5.7°, representing an approximate threefold increase over the noise-induced error. Thus the likelihood of experiencing calibration related artefacts at comparatively low calibration discrepancies is dependent on SNR. The presence of biases, however small, in the noise free data indicate that rather than C/T discrepancy simply amplifying noise-induced uncertainty, there are other systematic issues at hand.
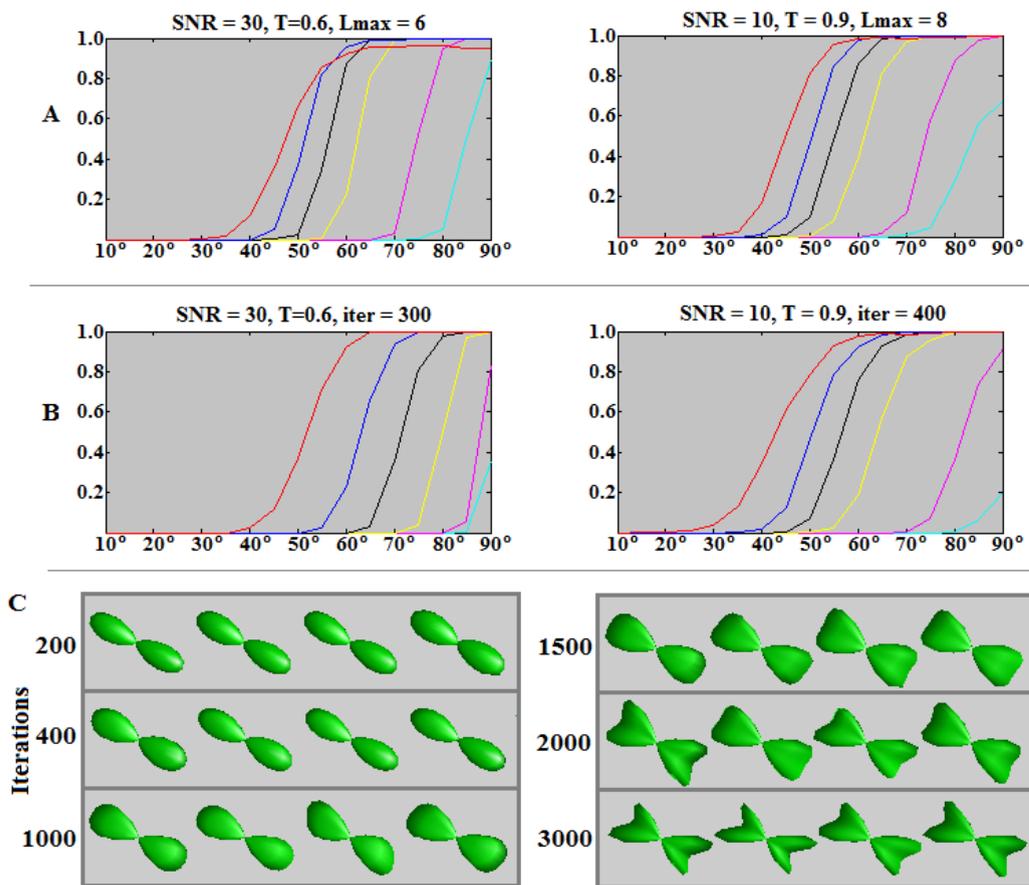
FIGURE 4.13: Angular resolution results if disregarding the strict "two peak only" success criteria. (A) CSHD: note improvement in high calibration FA results. (B) dRL: Improvements at low SNR, but at moderate/high SNR little change is observed. (C) dRL convergence example, four instances of a $T = 0.9$, 60°, 50/50 crossing fibre configuration at 50:1 SNR, varying calibration anisotropies.

Erroneous peaks lie in two distinct regions of interest in C/T space (Fig. 4.3c). Noiseless data predict that the first region is bounded by $0.667C > T > 0.5$ (for $Lmax = 8$), here peak magnitudes/quantities should be lower (compared to the second region) and not seriously affect nave tractography implementations. The second region defined by $T < 0.5C$ should contain larger peaks in greater quantities and thus have more significant detrimental effects. The limited resolution afforded by the noisy data does not

permit these same assertions; however, a general trend is seen where a reduction in SNR leads to an increase in the maximum magnitude and affected area (C/T combinations). Reconsider the $T = 0.6$ border case with ideal $C = 0.6$ calibration ($Lmax = 8$), it is seen that spurious peaks produced through noise have magnitudes of 0.055, 0.0291 and 0.0288 (for SNRs of 10, 30 and 50 respectively). Selecting a ($C = 0.9, T = 0.6$) miscalibration increases these magnitudes to 0.199, 0.126 and 0.088, placing two of the three peak amplitudes above the commonly-used 0.1 tracking threshold. We must therefore conclude that while underlying calibration related complications are responsible for the bulk of the artefacts (as evidenced by noiseless results and the persistence of artefacts at high SNR), calibration disparity may amplify pre-existing (yet harmless) artefactual peaks to disruptive magnitudes.

Finally, when examining expected tractography failure rates (Fig. 4.8), it must be remembered that the approach used is naïve, in that only local effects are considered. For example, the 'success' criterion for an incoming streamline within an angular threshold of the expected orientation to continue is that the tangent to the streamline minimally subtends the best estimate of fibre orientation. What is not considered here is the impact that calibration-induced errors on the best estimate (up to 35°) will have on the following step; e.g. given even mild tract curvature, following such a trajectory may incorrectly cause a premature termination of the streamline by exceeding angular thresholds at the next step, or, initiate tracking down an erroneous path. Extending beyond the immediate locality, it is clear that calibration induced errors will have a more detrimental impact on tractography results than presented here.

4.4.1 Orientation of Artefactual Peaks: A Potential Cause

To Determine the cause of spurious peak formation, it is instructive to consider the basic mathematics underpinning CSHD. CSHD assumes that the fODF, single-fibre response function and DW-signal can be adequately approximated by a linear combination of spherical harmonics (Legendre polynomials). However, in order to reduce the computational load and the number of unique diffusion-weighted images that must be acquired, we tend to truncate the harmonic series to a predetermined maximum order – the property referred to as Lmax. While this is a prudent step to reduce data acquisition requirements and maximise model parsimony, there are some drawbacks. Although it is possible to sufficiently characterise the relatively smooth DW-signals and fibre response functions at low Lmax, characterisation of the sharper fODF often requires contributions from higher order harmonics whose omission leads to well known ringing effects (Anderson (2005)). In the single-fibre case, the ideal fODF response is a unit impulse with orientation along the fibre axis. However, attempting to describe this impulse with truncated harmonics results in a ringing artefact that generates both positive and negative spurious fODF lobes at predictable orientations relative to the impulse. By

differentiating the responses at various Lmax, it is possible to compute the orientation of the positive ringing artefacts (note that negative lobes would be culled by the non-negativity constraint implemented in CSHD). Table 1 and Figure 5 show that the spurious fODF lobes observed in our empirical data are very similar to these theoretical values. We therefore consider that this simple artefact, exacerbated by miscalibration, is the root cause of the issues we are investigating. However, as the SNR decreases, there is an increased bias away from these predicted values, which may be explained by the 'squashed peanut' phenomenon arising from the Rician noise floor, as described elsewhere (Jones and Basser (2004b)).

Tournier et al., (2004) discuss the effects of deconvolving a wider diffusion profile from a narrower diffusion profile (in essence, $C < T$), stipulating that it may lead to production of negative terms in the diffusion fODF (although data were not provided). Essentially, we are exploring the opposite. By deconvolving a narrow diffusion profile from a wider one ($C > T$) we are erroneously adding, or increasing, positive terms in the resultant fODF thereby increasing the magnitude of the ringing induced fODF peaks beyond typical non-small/negative and fODF threshold constraints. It is not necessary to consider the original $C < T$ case, since negative lobes are dealt with through the non-negativity constraint in the CSHD approach (Tournier et al. (2007)), which explains the acceptable performance in these $C < T$ C/T space regions.

It is also well-known that higher order harmonic representations are more sensitive to noise since the magnitudes of individual SH coefficients are reduced (Anderson (2005)); this may also explain variations with respect to Lmax. Notice that as Lmax is decreased, resilience to miscalibration improves. It may therefore be reasonable to consider miscalibration as analogous to noise; by performing a mismatched deconvolution we add or subtract from the coefficients of the ideal fODF result. Where ideal coefficients are large, miscalibration 'noise' will be small in comparison and thus only large miscalibrations will produce noticeable results ($Lmax = 4$). Where these coefficients are smaller, the 'noisy' contributions will have a larger relative impact, producing spurious fODF peaks at smaller miscalibrations ($Lmax = 8$).

Extending this argument allows us to begin to explain the counter-intuitive increase in spurious peak numbers with increasing SNR (Fig. 4.8). As calibration anisotropy varies with respect to the target diffusion profile, differences in diffusion profile width transition smoothly from negative to positive and, more importantly, distribution of such differences remains approximately equal at any given elevation – in effect creating a constant radial residual. The only way to describe this constant residual would be multiple 'evenly' distributed fODF peaks at elevations biased by the previously discussed ringing artefacts.

With the addition of noise, residuals at any given elevation are no longer constant, a diffusion weighted signal may indicate a higher than expected ADC along one orientation

and a lower than expected ADC along another. The result (before the non-negativity constraint is applied) might be an fODF containing both negative and positive coefficients with much greater amplitudes than the comparable noiseless C/T pairing. In real terms this can mean that C/T pairs where $T > C$ are able to produce positive/large fODF peaks (explaining the earlier appearance of spurious peaks at low SNR) and pairings where $C > T$ are able to produce negative (explaining reduced numbers of spurious peaks) and larger positive (hence higher peak magnitudes at lower SNR) erroneous coefficients at the same time.

4.4.2   A Porential Alternative: dRL

In practice, the assumption that the fibre response function can be adequately described through a truncated series is, while computationally elegant, potentially problematic. This is not to say that CSHD should be abandoned; in fact, crossing-fibre resolution performs well provided the calibration step is performed carefully. If, however, a narrow range of single-fibre responses cannot be guaranteed (breaking the assumption of a single canonical fibre response function), then alternative approaches may be of interest – of which dRL represents one possibility.

Single-fibre resolution through dRL is largely insensitive to calibration anisotropy and its performance should rarely be significantly worse than DT-MRI under similar noise conditions. Problems do however exist in the resolution of crossing-fibres. Results show that in general dRL performs best with a high FA calibration, $C = 0.9$ providing the best results across the target anisotropy gamut. However, unlike the single fibre case, dRL's inbuilt damping is not always sufficient to suppress the formation of spurious peaks, leading to poor performance against low anisotropy, low SNR crossing-fibre targets at higher iteration counts. Unfortunately, unlike CSHD, one cannot simply select a matched ($C = T$) calibration. Much as a $C < T$ calibration with CSHD will result in the loss of angular resolution, so will $C < 0.9$ with dRL (Fig. 4.10, 4.11); the resultant fODF's remain overly rounded, leading to either a large imprecision in orientational estimates (biased towards the mean orientation) or a single peak (obviously this is not a problem for single-fibre targets).

The main issue for dRL seems to be one of convergence. The algorithm is believed to return the best results shortly before the fODF deconvolution fully converges (Dell'Acqua et al. (2010)), while actually reaching convergence can often result in an over-representation signals noise component; which can be most readily seen in single-fibre data (Fig. 4.7, $C = 0.8/0.9$, or the low SNR improvements Figs. 4.11d and 4.13b) where spurious peak formation causes error rates to increase with iteration count. The number of iterations required to reach the convergence 'sweet spot' varies as a function of both target and calibration anisotropies; as either decreases the number of iterations required to achieve an optimal solution increases – rapidly approaching computationally impractical levels

(Fig. 4.13c) and becomes increasingly sensitive to target SNR in the process. This also explains the steady performance drop-offs with decreasing target anisotropy for a fixed number of iterations. As a toy example, it appears near impossible to reliably resolve $T = 0.3$ crossing-fibre fibres at $C = 0.3$ with an SNR of 10-20 regardless of the iteration count, while at higher SNRs it simply takes a long time.

To attempt to use dRL in this way also seems rather self-defeating. Committing to a $C < 0.9$ calibration will require an increased number of algorithm iterations which (ignoring the extra processing time which would likely make tailored CSHD a more tempting choice) will also likely preclude the ability to resolve higher anisotropy target fibres, crossing or otherwise, due to differential convergence points – i.e. a higher FA target will require fewer iterations to reach an optimal solution, the remaining additional iterations will most likely lead to spurious fODF peak formation due to previously discussed over-convergence effects.

In some ways it may be helpful to imagine the calibration properties of dRL as being 'inverse' to those of CSHD. In CSHD, a high anisotropy calibration response function will lead to spurious fODF peaks forming due to ringing artefacts, thus calibration at lower anisotropy is a better default position (avoiding the worst-case outcome and providing a graceful performance degradation as target FA increases). In dRL, however, a calibration response function with low anisotropy causes the spurious peak formation and a highly anisotropic calibration response function ensures the maximum number of iterations used will not exceed the over-convergence threshold for any possible target fibre, providing a graceful degradation as target anisotropy drops. Conveniently however, this default 'high' ($C = 0.9$) anisotropy is typically above anisotropies expected of 'typical' white matter ($T > \approx 0.5$) and thus, with practical algorithmic parameters, dRL can provide useful resolution of fibre orientations throughout the entire brain without the need for a carefully tailored calibration.

4.5   Conclusion

The resolution of a single fibre orientation through SD methods is not as trivial as one may initially believe. While the addition of noise to a DW signal will obviously produce uncertainty in the estimation of fibre orientation and, if sufficiently severe, the appearance of spurious fODF peaks, we have also shown a further confounding effect in the form of errors produced through inappropriate calibration. While dRL is exceptionally resilient to calibration error in the single-fibre case, CSHD is particularly vulnerable to overestimation of the target diffusion profile regardless of DW-signal SNR. The truncated harmonic series response representation harbours descriptive deficiencies (ringing effects) which, as the magnitude of the calibration error increases, results in both predictable and unavoidable spurious fODF peaks.

That said, resolution of crossing fibres is equally important, but selection of an appropriate algorithm, calibration and running parameters is not a simple matter; trade-offs must be made depending on knowledge of the target image. For a white-matter target ($FA >\approx 0.5$) with no other a priori information, CSHD and dRL can be considered approximately equal. With CSHD it should be possible to achieve reasonable results with a moderate FA calibration in the 0.7-0.8 range (dropping Lmax if necessary) offering the best balance between preventing spurious fODF peaks at the lower FA range and maintaining angular resolution in the upper range. For dRL calibration is relatively straightforward, with $C = 0.9$ generally being the optimal choice and, while angular resolution is lost as the target FA decreases, resolution of single-fibre orientations will remain robust across the FA window- there are also other fringe benefits to dRL such as resilience to isotropic partial volume effects. If the range of target profiles is known, then the choice of algorithm is likely to be based upon the location of the damaged/low FA tissue of interest. If this occurs within regions of crossing-fibres, CSHD appears to be the solution of choice since: through tailored low FA calibration, it is possible to retrieve the highest angular resolution within the region of interest with accurate resolution of non-crossing fibres at a cost of reduced angular resolution/sensitivity to volume fractions in higher FA regions that may not be of interest. If, on the other hand, low FA tissue is non-crossing, then dRL would be the preferred choice. Resolution of crossing-fibre in high FA regions will be unaffected and, with a single-fibre target, low FA angular resolution issues are not a concern. For images where the diffusion profile remains constant, choice is largely dominated by the shape of that profile. For highly anisotropic diffusion there is little difference between dRL and CSHD, however, as profile anisotropy falls (e.g. to 0.3 for muscle, Parker and Jones, 2011), dRLs inability to refine a sharp fODF makes CSHD the preferred method, especially considering that it is possible to reduce Lmax for increased robustness at low SNR with little penalty in angular resolution at these anisotropies.

The real-world implications of these findings will probably not be felt by those studying healthy white-matter. However in white matter degradation or studies of other fibrous tissues, our results have shown that choice of algorithm and calibration can have marked effects on the end result. This should be of particular interest to those currently complying with Tournier's (2004) calibration recommendation for CSHD, i.e. scanning the image raster-fashion for high anisotropy voxels (e.g. $FA > 0.8$) upon which to base an average fit calibration, where our results clearly demonstrate that a more considered calibration might lead to vast improvements in results.

In summary, the aim of this work was to highlight the effects of initial calibration on the veracity of two SD-derived fODF estimates. When presenting or reviewing a new HARDI technique, authors often (rightly) stress crossing fibre resolution performance but fail to address the seemingly trivial single fibre case; here, we have shown the pitfalls of such a strategy. Not only can the single fibre problem provide insight into a

techniques strengths and weaknesses, it is also representative of a significant proportion of human white matter. Comparing DT-MRI to CSHD in particular, only a small C/T discrepancy is necessary for the performance of CSHD to significantly degrade compared to a 'benchmark' DT-MRI comparison (Figs. 4.2, 4.4, 4.7). While we are not arguing that DT-MRI is superior to CSHD, DT-MRI's simple model and lack of an inherent a priori anisotropy assumptions does give it an inherent advantage with all single fibre signals (regardless of anisotropy) that must be closed through appropriate calibration. In essence, the advantages of crossing-fibre resolution are a definite boon, but utterly useless if single fibre resolution is incorrect (consider the consequences for the resulting tractography!). In practice it would best if we could do away with these a priori image-wide assumptions in HARDI; indeed some SD techniques do in fact achieve this through voxel-wise modelling (Anderson (2005)). However, for dRL and CSHD several key recommendations can be made: Provided diffusion profile anisotropy remains high (throughout this manuscript we have used FA as a parametrisation), both dRL and CSHD should provide accurate measurements of fibre orientation; dRL simplifies calibration choice ($C = 0.9$ is always optimal) while CSHD can be used to target specific locations with a tailored calibration. At profile anisotropies below 0.5, CSHD becomes the only choice, but the selected calibration anisotropy should be equal to, or slightly lower than, the lowest expected profile anisotropy; while this may cost in terms of angular resolution, it will avoid the ringing related artefacts that we have explored and single fibre estimates will remain robust.

Chapter 5

Maxillofacial Imaging: A Case Study

Prologue

Following on from the previous chapter, this work serves two functions. Firstly, we will demonstrate the real world consequences of inappropriate calibration. Secondly, in the larger context of the overall project, we will demonstrate progress towards maxillofacial imaging and tractography. While, for reasons that will shortly be discussed, imaging in the maxillofacial region is problematic with currently available equipment – and as such later work will transition onto white-matter or dual purpose research areas - the advancements and insights gained through this study clearly illuminate the path forwards. As evidence for this, we conclude the chapter with the first anatomically plausible in vivo reconstructions of several important maxillofacial muscles – a key step on the path towards subject-specific maxillofacial modelling.

5.1 Introduction

It is known that the manner in which a muscle delivers mechanical force depends strongly on its fibrous architecture (Bovendeerd et al. (1994)). In order to better understand subject-specific biomechanics, it is necessary to obtain accurate measurements of muscle fibre orientation in vivo. The most commonly used method for retrieving such information is currently DT-MRI (Basser et al. (1994a)) which has achieved considerable success in the reconstruction of large skeletal (Damon et al. (2002), Heemskerk et al. (2010), Kan et al. (2009)) and even cardiac muscle tissue (Zhukov and Barr (2003), Kim et al. (2005)); facilitating recovery of important biomechanical properties including pennation

angle (Heemskerk et al. (2005)) which can be shown as consistent with data obtained through anatomical dissection (Damon et al. (2002)).

The principal limitation of DT-MRI, however, is that the tensor model is incapable of providing a meaningful description of the kinds of non-Gaussian diffusion arising from multiple fibre components contributing to a single DW-MRI signal - an occurance known as the partial volume effect (PVE, Alexander et al. (2005)). This limitation is extremely relevant since a significant part of the challenge of muscle fibre reconstruction is its relatively short T_2 ($33 - 45ms$ depending on muscle activity (Remeur et al. (1994)) in comparison to that of more commonly imaged white matter ($\approx 80ms$, Wansapura et al. (1999)). As covered in Chapter 2, a reduction in T_2 will increase the rate at which the signal attenuates and so impose greater limits on voxel dimensions and diffusion weighting if one wishes to ensure that DW-signal intensities remain above the Rician noise floor for the majority of the recorded echo train - in effect placing relatively low thresholds on both voxel and angular resolutions. Typical DT-MRI acquisitions in muscle use a $3 \times 3 \times 3 - 10mm$ voxel with b-value of $400 - 600s/mm^2$ (Damon et al. (2002)) compared to the 1-2mm isotopic/1-3000s/mm^2 acquisitions in comparable white matter studies – though it is worth noting that the negative effects of a reduced b-value, while still felt, may be less than expected given the higher diffusivities (Table 5.1) in muscle tissue which serve to increase related attenuations. Given a set of muscles whose diameters far exceed that of a single voxel, such as the previously mentioned skeletal/cardiac muscles, these restrictions are not necessarily an issue. The vast majority of intersected voxels will lie wholly within a single muscle and, as such, single fibre diffusion assumptions should apply. Considering the entire volume, this means that the effects of PVE on resultant reconstructions should be minimal and as such DT-MRI is generally considered reliable – in fact, in some applications it is common place to discard potential PVE voxels completely by eroding the outer surface voxels of individual muscles which, due to their size (relative to the voxels), can be assumed to have minimal effects on the usefulness of the retrieved information (Heemskerk et al. (2010)). However, in regions where muscle fibre bundle diameter approaches that of a voxel, the frequency of voxels containing PVE contributions will naturally increase, necessitating the retainment of PVE voxels since their occurrence as a portion of total voxels becomes too large to ignore and, by extension, draws the validity of DT-MRI derived results into question.

		Muscle	White Matter	Percentage Difference
λ_1		$1.93 \times 10^{-3} mm^2/s$	$1.75 \times 10^{-3} mm^2/s$	-9.24%
λ_2		$1.34 \times 10^{-3} mm^2/s$	$0.44 \times 10^{-3} mm^2/s$	-67.46%
λ_3		$1.09 \times 10^{-3} mm^2/s$	$0.34 \times 10^{-3} mm^2/s$	-69.25%

TABLE 5.1: Average eigenvalues sampled from non-PVE neck muscle and the genu of the corpus callosum (white matter). Note the large difference in radial diffusivities.

A suitable example to reinforce this argument is the tongue. Where groups have attempted to image the tongue in vivo e.g. Gilbert and Napadow (2005), SNR considerations forced a low spatial resolution, $\approx 3 \times 3 \times 6 - 8mm$, which meant that the majority of the lingual core – comprised of numerous interwoven muscle fibre groups – was captured in partial volume. This then forced the authors to abandon any orientational information and instead attempt to infer voxel content based on spatial location relative to classical anatomy, using anisotropy – with the assumption that low anisotropy implied inhomogeneous (crossing/fanning) tissue – as a means of identifying tissue boundaries (Fig. 5.1). Compared to more successful ex vivo works which are able to reduce the SNR concerns by eliminating movement, using surface coils, longer scan times and excluding common sources of susceptibility artefacts around the mouth/sinuses, high resolution DT-MRI (Kim et al. (2005)) and even DSI (Wedeen et al. (2005), Gilbert et al. (2006)) make it clear that in vivo DT-MRI lacks the fidelity to provide accurate biometric information for more complex, compact muscle tissue. Unfortunately, the high b-values required for DSI ($5500 - 8500s/mm^2$), the only other commonly employed method within muscle publications, are unlikely to be plausible in vivo.
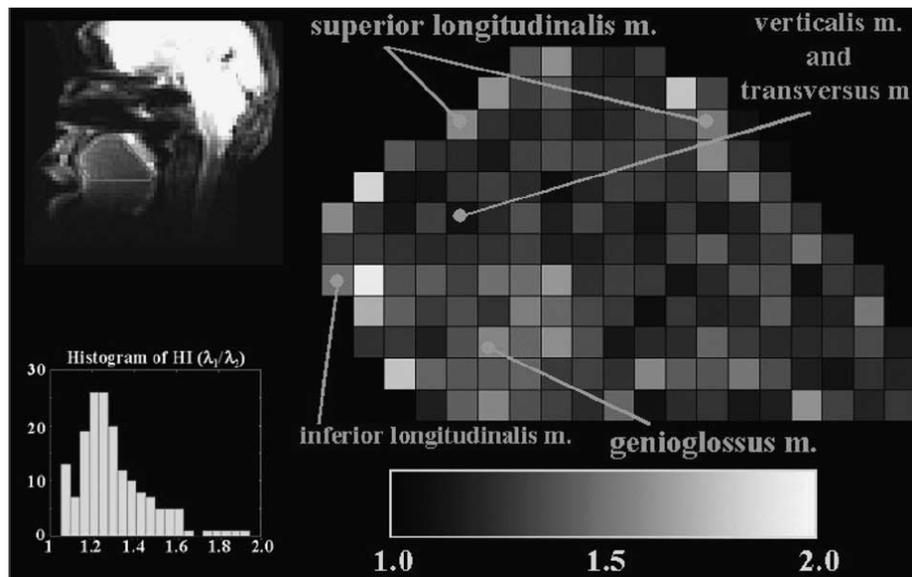
FIGURE 5.1: Reproduction of Gilbert and Napadow (2005), Figure 6. Diffusion anisotropy of the tongue with (λ_1/λ_2) was displayed as a grey-scale map.

Constrained spherical harmonic deconvolution (CSHD, Tournier et al. (2007)) provides one possible solution. CSHD is one of a number of HARDI techniques designed to alleviate the PVE problem through resolution of multiple fibre compartment orientations per voxel. As covered in chapters two and four, CSHD assumes the orientation and contribution of all fibres modulating the diffusion weighted signal (its fODF) can be retrieved through deconvolving an idealised single fibre response function from the measured DW-MRI signal. Furthermore, the previous simulations indicate that CSHD, unlike dRL (Dell'Acqua et al. (2010), the other method explored in Chapter 4), may be able to resolve crossing fibres with muscle-like anisotropy (in comparison with white

FIGURE 5.2: Cropped reproduction Gilbert et al. (2006), Figure 5. Image plane roughly equivalent to Figure 3.1 though the bovine tongue subject is dissimilarly shaped compared to the previous human tongue. Note the three distinct, mutually orthogonal, fibre populations.

matter, muscles radial diffusivity is increased while axial diffusivity remains similar, creating a rounded diffusion profile with low anisotropy), albeit at far higher diffusion weightings than are likely to be practical in vivo – a matter that will be examined shortly through simulations.

Diffusion weightings aside, selection of a tailored response function (calibration) for muscle tissue is not a trivial mater. At lower anisotropies, e.g. muscle ($FA = 0.3$), previous simulations suggest that even small overestimations during the calibration phase can generate spurious fODF peaks while, conversely, small underestimations can reduce angular resolution to the extent that resolution of crossing fibres is near impossible beyond the orthogonal – defeating the purpose of using a HARDI method. There are also the well known effects of harmonic truncation (Lmax) to consider (Alexander (2005)); recalling Chapter 2, CSHD performs deconvolution through manipulation of a truncated set of spherical harmonics, the order at which truncation occurs determines the representable maximum angular frequency and thus angular resolution of the resulting deconvolution. The drawbacks however are two-fold. (1) Increasing Lmax requires an ever increasing number of unique data-points (directional images) in order to achieve a fit – significantly increasing acquisition time for only the chance of a real-terms increase in angular resolution (if the diffusion weighting does not support the angular resolution, increasing harmonic order will have very few benefits). (2) An increased sensitivity to noise in the DW-MRI signal.

In order to access the suitability of CHSD to in vivo muscle applications we must therefore address three important points: (1) While previous simulations show that resolution of low anisotropy crossing fibre may be possible, these are conducted at a b-value of

$2000s/mm^2$; given that this is unlikely to provide reliable data in vivo, it is necessary to determine whether resolution of crossing fibres is feasible within practical diffusion weighting and, if so, how to balance the required diffusion weightings with spatial resolution in order to maintain a workable SNR. (2) In conjunction with point one, given that we expect to face numerous SNR issues and low diffusion weightings, are there any benefits to an increased Lmax? At which point is the best balance between noise sensitivity, angular resolution and acquisition time achieved? (3) Given that simulations predict target specific calibration appears to be a fundamental component of an accurate CSHD fODF estimation, does this bare out in practical circumstances? If so, what can be done to ensure that an appropriate calibration is reliably achieved?

By addressing each of the highlighted points, we believe it should be possible to arrive at a well reasoned set of acquisition requirements and processing parameters for producing reliable results within a research environment. To achieve this goal two datasets are examined. Firstly, a series of simulations are conducted to examine the effects of b-value, Lmax and SNR on effective angular resolution within muscle tissue, analysis of which should provide ballpark estimates for acquisition and processing parameters. Secondly, to test the validity of the simulation results, human subject data were acquired with (as close as was feasible) optimised parameters allowing segmentation of muscle tissue (to demonstrate current progress) and provide a vehicle for discussion of various artefacts that are commonly encountered and potential solutions to these problems. The end result is effectively a primer on CSHD based exploration of the maxillofacial region, providing a strong starting point for future work towards the eventual aims of this project.

5.2 Muscle groups

In order to achieve the goals we have laid out, it is first necessary to define the two regions of interest (and the muscle contained within) upon which this study will focus.

To pre-empt the following subsections, owing to technical difficulties explained in Section 5.4, we are not yet able to meet the minimum requirements (Section 5.3) required for high quality DW-MRI imaging. Fortunately, however, owing to their position and composition, two maxillofacial regions commonly come close to these requirement and, furthermore, contain muscle tissue that provides the crossing fibres needed to validate the previous chapters CSHD results.

5.2.1 Muscle Group 1

The first reliably imaged region is centred about the temporomandibular joint and contains the temporalis, masseter, buccinator, resorius and medial pterygoid muscles.

1. The temporalis is a muscle of mastication originating from the temporal fossa and deep temporal facia, passing under the zygomatic arch and inserting into the coronoid process and anterior aspect of the ramus of the mandible

2. The masseter is a muscle of mastication consisting of overlapping layers:(a) A superficial component arising from the zygomatic process of the maxilla and anterior two third of the lower border of the zygomatic arch. (b) a central component originating from the deep surface of the anterior two thirds of the lower border of the posterior third of the zygomatic arch, and (c) a deep component arising from the deep surface of the zygomatic arch. All three insert into the lateral aspect of the ramus of the mandible.

3. The buccinator is a thin quadrilateral muscle occupying the interval between the maxilla and mandible. Originating from the anterior margin of the pterygomandibular raphe and alveolar margin of the maxilla and mandible; diverging into three thinly connected bands that traverse the cheek and insert into one of two locations. The upper band inserts into the lower orbicularis oris (upper lip); the central band twists at the modiolus such that the upper component inserts into the lower orbicularis oris and the lower into the upper; finally the lower band inserts into the lower oribicularis oris.

4. Finally, the risorius is an extremely thin muscle of facial expression originating from the connective tissue overlaying the parotid gland, passing orthogonally across the external surface of the masseter, through the cheek and into the modiolus (corner of the mouth).

5. The medial pterygoid is a thick quadrilateral muscle of mastication, the bulk of the muscle originates from above the medial surface of the lateral pterygoid plate while the smaller superficial head originates from the maxillary tuberosity and pyramidal process of the palatine bone. The muscle inserts at the lower and rear surface of the ramus and angle of the mandible.

To place these descriptions in context, a visual representation of each muscle is available in Figure 5.3. For more detail consult Norton (2011).

While the majoirity of these muscles are relatively large in comparison with the $3 \times 3 \times 3mm$ voxel, and thus may be (at least partially) recovered through standard DT-MRI, there are interesting points to consider. The masseter and buccinator, for example, consist of multiple layered components with each possessing their own origin/insertion points and differing trajectories to match. This creates the possibility for intra-muscle PVE which will be later be examined. Furthermore, some of the muscles (masseter and buccinator/risorius) come into direct – near orthogonal – contact, providing an ideal testbed for resolution of crossing muscle fibre in vivo, away from the major sources of image distortion.

FIGURE 5.3: Visual depictions of the group 1 muscle cluster. A: The temporalis and masseter muscles in situ. B: Depiction of the internal face of the buccinator, note the three distinct fibrous bands. C: External view of the buccinator demonstrating position relative to masseter. D: Risorius, extremely thin muscle generally existing in partial volume. E: External view of the medial and lateral pterygoid muscles.

5.2.2 Muscle Group 2

The second region centres upon the tongue and is mostly limited to areas immediately surrounding the medial spline – likely due to the surrounding homogeneous (muscle) tissue shielding the core from the various imaging artefacts. Fortunately, while limited, this still permits direct comparison with other work – namely the ex vivo DSI publication (Gilbert et al. (2006)). For this reason, muscles of interest (as named in that publication) are the genioglossus, superior and inferior longitudinal fibres and transversus muscles – three nominally orthogonal fibre groups (Fig. 5.4).

FIGURE 5.4: Visual depictions of the group 2 muscles. (A) Cross-section of the medial spline depicting the genioglossus, transverse and superior longitudinal muscles in situ. (B) A more extrinsic view of the tongue depicting the inferior longitudinal muscles (with relationship to the genioglossus), note that fibres may merge with the styloglossus at the posterior of the tongue.

5.3 Simulation

DW-MRI signal data were simulated (using methods discussed in chapter three) assuming a typical 60 direction sampling scheme (Jones et al. (1999)) with b-values varying from $100s/mm^2$ to $1000s/mm^2$ at $50s/mm^2$ intervals with signal to noise ratios of between 10:1 and 90:1 in 20:1 step intervals. Each signal comprised two equally contributing crossing fibre components consistent with a fibre posessing $2 \times 10^{-3}mm^2/s$ axial and $1.2 \times 10^{-3}mm^2/s$ radial diffusivites (mean values across a sample of masseter muscle, $FA \approx 0.3$, $MD = 1.43 \times 10^{-3}mm^2/s$), one component axially aligned and the other rotated in-plane to provide an inter-component angle of between 0° and 90° at 5° intervals. Five hundred unique repetitions were conducted per b-value/SNR/crossing-angle tuple.

CSHD was then applied to each signal tuple in turn with calibrations ranging from $FA = 0.1$ to $FA = 0.9$ at a 0.1 interval (axial/radial diffusivities calculated according to a fixed $1.43 \times 10^{-3}mm^2/s$ MD) and all resultant fODF peaks/magnitudes recorded. For evaluation purposes a successful trial is considered to be one that produces two, and only two, supra-threshold ($fODFmag > 0.1$) fODF peaks, each falling within a 20° or 30° cone of error about the expected fibre orientations.

5.3.1 Simulation Results

To address the simplest question first, the simulations (Fig. 5.5, SNR constant at 50 : 1) suggest that CSHD should be able to resolve crossing fibre of muscle-like anisotropies using b-values within the $\geq 600s/mm^2$ range which are frequently used in vivo (Fig 5.5, row 2). Going further, response function underestimation ($C = 0.2$) will result in loss of angular resolution through failure to delineate crossing fibre and overestimation ($C \geq$

0.4), while slightly improving the range of angular separations that *may* be delineated (particularly at $Lmax = 4$), does so at an increasing cost to reliability (note that the "width" of the non-zero error areas increases, but colours within corresponding voxels trend towards the blue end of the spectrum, indicating a relative loss of reliability). In agreement with the previous chapter, it therefore appears that optimal calibration still coincides with the response function that most closely matches the target tissue.
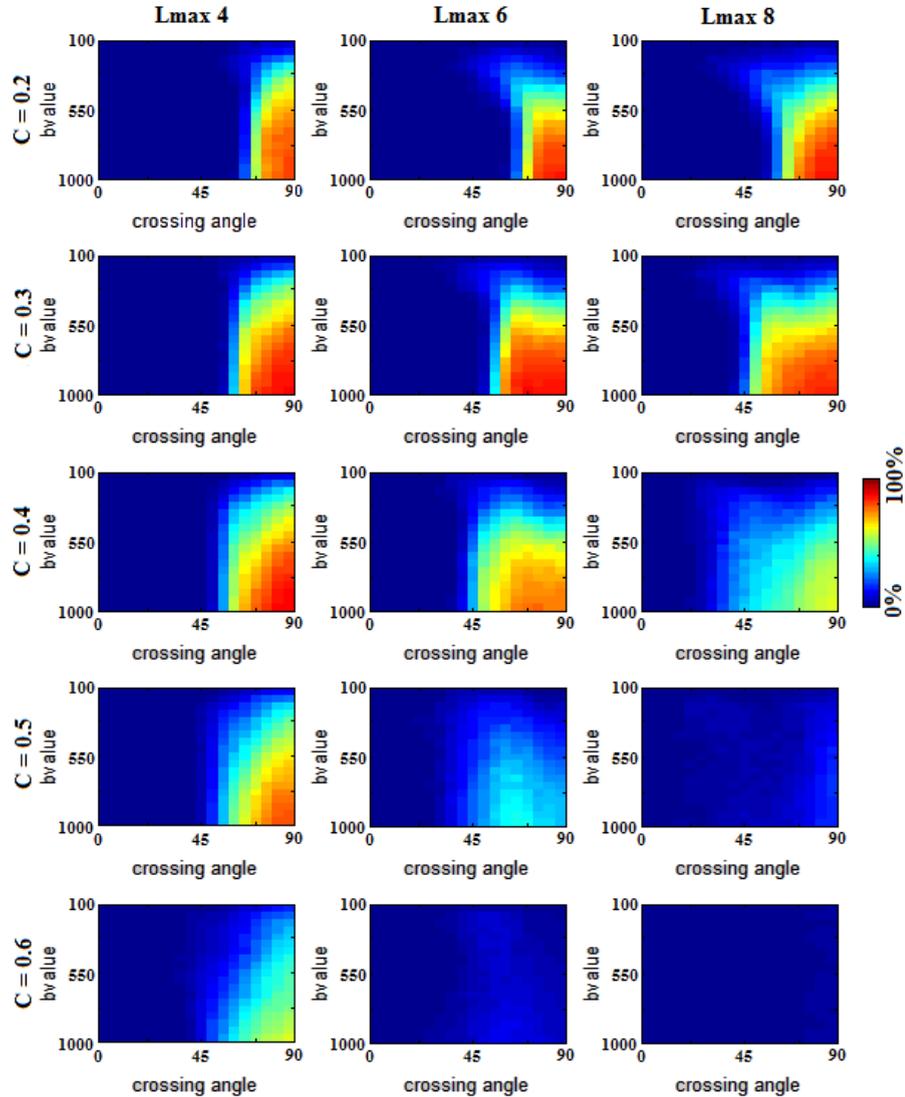
FIGURE 5.5: Plots of angular resolution with respect to diffusion weighting (s/mm^2). Note that while calibration anisotropy increases down the rows, SNR remains fixed at 50:1

Continuing to Lmax, results presented in Figure 5.6 ($C = 0.3$) show that at very low SNRs ($\approx 10 : 1$), lowering Lmax will have little to no effect on solution quality. As SNR improves however, a lower Lmax tends to allow most accurate resolution of crossing fibres at lower b-value, albeit within a lower range of angular separations. The overall result is that at lower b-values and SNRs, $Lmax = 4$ CSHD appears to be the optimal choice, though as SNR/b-value increases, it may be beneficial to switch to a $Lmax = 6$

deconvolution – $Lmax = 8$ does not appear beneficial at any point within the selected range of diffusion weighting. Again, these observations coincide with both the previous chapters simulations and common knowledge regarding the effects of lmax truncation on the CSHD algorithm
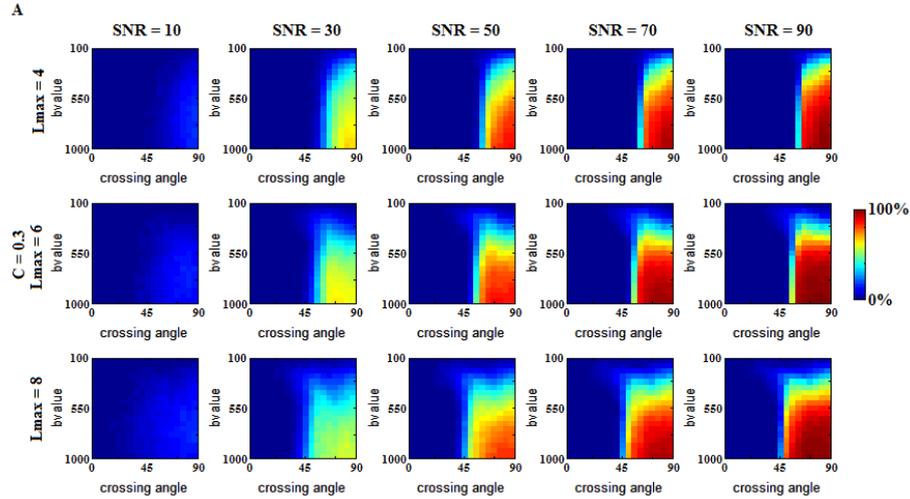
FIGURE 5.6: Plots of angular resolution with respect to diffusion weighting. Calibration anisotropy fixed at $C = 0.3$, SNR increasing across the columns

Bringing these results together, for a $\geq 80\%$ chance of resolving non-orthogonal crossing-fibres to within $20°$ of the true orientation, the simulations suggest that b-value must exceed $750 s/mm^2$ and $SNR \geq 50 : 1$ (Fig. 5.7). Simply put and as the next section will show, this was unachievable with current equipment. The problem is two-fold. Firstly, at diminished diffusion weightings there is an increased tendency for recovered fibre orientations to display bias towards a central point (nominally the mean) which, as the weighting decreases, will naturally extend until the once separate fODF lobes merge to indicate a single artefactual fibre orientation. Given the b-values under which we must operate and the limited extent of the discriminating attenuations produced by the low anisotropy fibre components, it is safe to assume such effects have a large outcome on any recovered fODF. This is then compounded by the disproportionately large effects of a reduced SNR on low anisotropy signals (already discussed in detail in Chapter 4) which not only adds further bias, but also has the tendency (regardless of calibration) to cause the formation of additional spurious fODF lobes, particularly at higher Lmax which are known to be more susceptible to noise. To this end, marked improvements can be observed if we are prepared to either: (1) Accept a wider margin of error in recovered fibre orientations. Figure 5.8 displays results were the angular acceptance criterion increased to $30°$ or (2) accept spurious fODF components. Figure 5.9 displays results given a maximum of three supra-threshold peaks, provided two remain within $20°$ of the expected orientations. Note that, as might be expected, the largest improvements are felt at lower SNR.
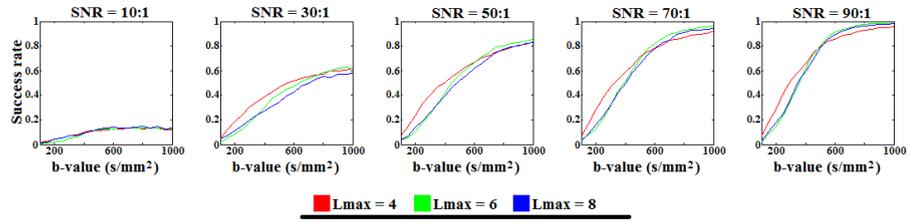
FIGURE 5.7: Plot of successful fODF recovery against diffusion weighting for a fixed 75° intersection and $C = 0.3$ calibration
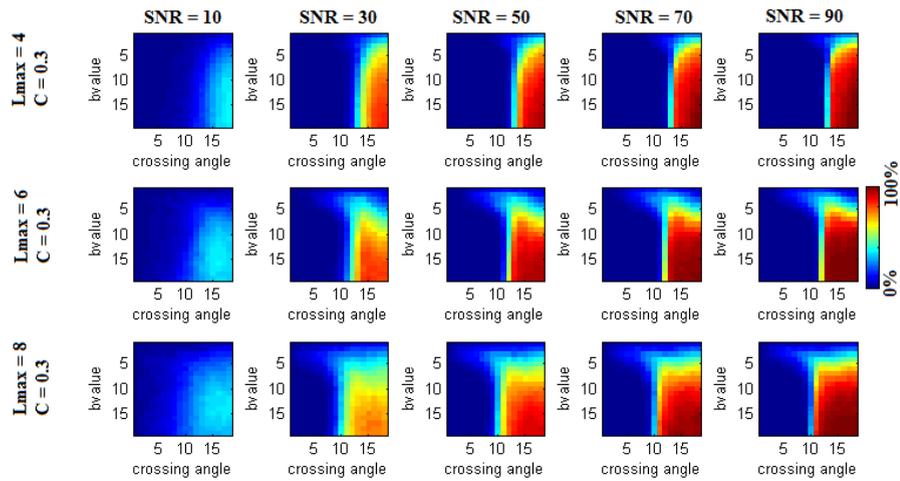
FIGURE 5.8: Plots of angular resolution with respect to diffusion weighting for optimally calibrated CSHD. Angular threshold increased from 20° to 30°
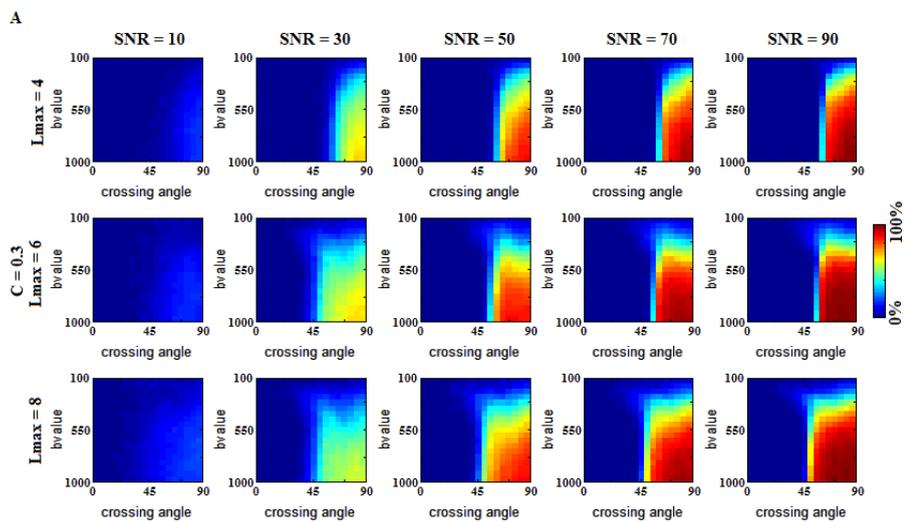
FIGURE 5.9: Plots of angular resolution with respect to diffusion weighting. 20° acceptance criterion allowing for a 3rd peak approximately normal to the main orientations

Given the impracticality of reaching the predicted SNR and b-value requirements for white matter like angular resolution/repeatibility in vivo (Figure 5.10 displays the expected success rates at the $450s/mm^2$ bvalue using: (A) a 20° and (B) a 30° angular threshold. Note that even at 30°, success barely exceeds 75% for an orthogonal crossing), it is perhaps beneficial to rephrase the angular resolution question – for a given SNR, b-value and angular separation (1) is resolution of crossing fibre possible and (2) if so, what degree of confidence may be expected? If one then re-examines the figures presented within this subsection, they essentially provide a set of look-up tables for this very information (Figure 5.6 columns 2 and 3 are particularly relevant to this work).
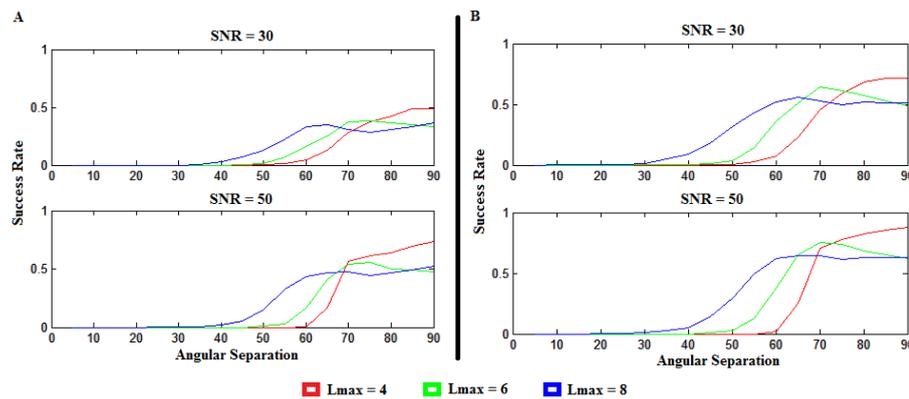
FIGURE 5.10: Plot of angular resolution at SNRs of 30:1 and 50:1. (A) Angular resolution at a 20° threshold. (B) Angular resolution at a 30° threshold.

5.4 Maxillofacial Imaging

As previously stated, the low T_2 of muscle tissue leads to rapid relaxation with negative impacts on potential SNR. Unfortunately, the inherent difficulties of imaging muscular tissue were further compounded by: (1) Using a bucket type 8 channel brain imaging coil. While comfortably the "best" selection of the available hardware (having attempted to adapt a variety of other coils intended for other uses) and capable of providing usable image quality for upper sections of the maxillofacial musculature, limitations in its design cause sensitivity to drop as one proceeds downwards, resulting in signal roll-off across the lower portions of any maxillofacial image (Fig. 5.11a). (2) The unusual anatomical construction of the maxillofacial region. Most common imaging targets tend to exist within a bulk of paramagnetically homogeneous material (e.g. grey/white matter within the brain or large muscle complexes) which leads to homogeneous distortion of the B_0 field and thus, excepting the boundaries of the tissue group, susceptibility has little consequence on the resultant image. Within the maxillofacial region however there are a large number of materials with differing paramagnetic qualities (Table 2, Collins et al. (2002)) that interface comparatively frequently within a small volume (e.g. skin/muscle, muscle/bone, skin/enamel, air/skin, air/enamel, enamel/metal (fillings) etc), leading to

frequent and severe distortions of the B_0 field (Fig. 5.12) which by extension break the positional encoding assumptions, leading to image distortion along the phase code direction (Examples of these artefacts are presented in Fig. 5.11b). While attempts are made to correct this image distortion using field map based methods (e.g. FSL FUGUE, Jenkinson et al. (2012)). In the majority of cases this has proved insufficient (the frontal surface of the face is more or less irrecoverable) and, due to limitations in the scanners software, more aggressive techniques are unavailable (this topic and its implications will be discussed further within later chapters).

Material	μ_r
Free space	1.00000000
Air	1.00000040
Water	0.99999096
Fat	0.99999221
Bone	0.99999156
Blood	0.99999153
Grey Matter	0.99999103
White Matter	0.99999120

TABLE 5.2: Relative magnetic permiability μ_r of a subset of the materials visible within a maxillofacial image. Table derived from information published by Collins et al. (2002).

With this information in mind, the principal dataset evaluated during the following case study was produced as follows: a single 24 year old male subject was firmly secured within the scanner in a head tilted back position (bringing as much of the jaw as possible into the region of higher RF coil sensitivity) and instructed to breathe slowly and shallowly, maintain a mouth-closed position and refrain as much as possible from moving the tongue in order reduce motion artefacts. Four repeat images were then captured during the single imaging session with the intention of providing sufficient repeated measurements to partially overcome defects in any individual image. Each image was acquired with identical parameters consisting of an optimized 60 direction imaging scheme (Jones et al. (1999)) with 6 b0 images, $b = 450s/mm^2$ and a $3 \times 3 \times 3mm$ isotropic resolution. EPI distortion correction (FSL FUGUE) was applied to each imagine separately before images were then co-registered and concatenated through the ExploreDTI concatenation tool (Leemans et al. (2009)), taking care to rotate the b-matrices (Leemans and Jones (2009a)) where necessary. Finally, corrupt signal rejection was applied, both slice wide and locally, according to the method outlined in Chapter 3. The end result was a single image with four repeated measurements per approximate gradient direction (remembering the b-matrix rotation will cause each repeat to differ slightly from the intended orientation).

Additional imaging was attempted both previous to and following this acquisition. A wide array of diffusion weightings were attempted to try and improve potential angular resolution, though SNR dropped sharply in response to any increase (as would be expected). We then attempted repeat images at the $450s/mm^2$ level to provision for
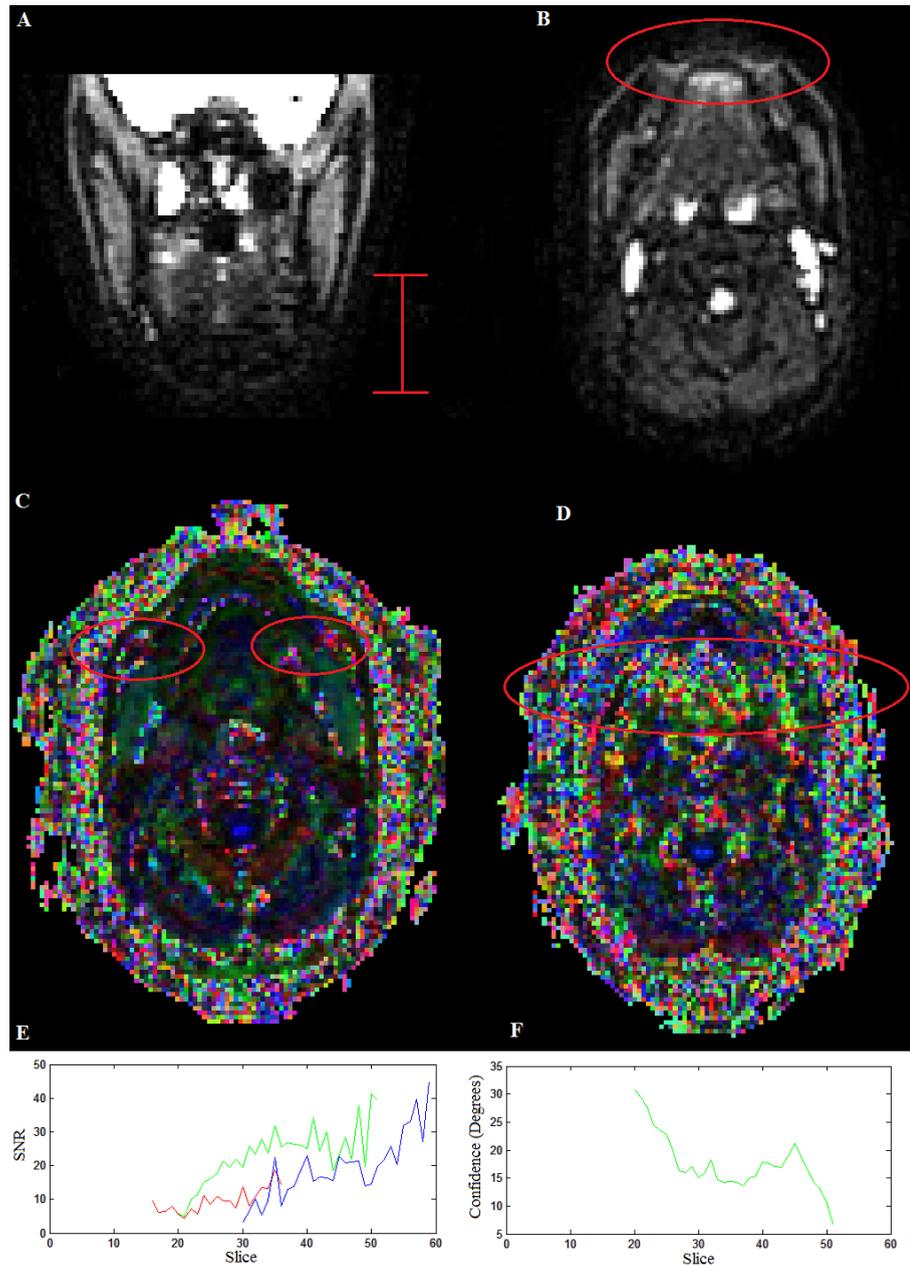
FIGURE 5.11: Examples of common artefacts. (A) Signal roll-off, note declining intensities across the labelled region. (B) EPI distortions, see lips/left masseter (labelled red). (C) and (D) Compairson of images at $b = 450s/mm^2$ and $b = 600s/mm^2$ respectively, corrupted regions labelled red. (E) Plot of mean SNR relative to slice position for three key muscles: the genioglossus (red), masseter (green) and temporalis (blue). Note that as one increases the Z axis coordinate, SNR rises sharply, providing further evidence for signal roll off. (F) Slice-by-slice mean bootstrapped estimate of the 95% cone of uncertainty in 'gold-standard' (see later bootstrap subsection) primary peak orientation across the masseter – note negative correlation with SNR.

a reproducibility study, but the quality of these images was extremely variable. Eventually we resolved to postpone imaging for the arrival of new hardware (specifically a custom built RF coil) which, due to various delays in production, did not arrive in time for use in this project. As a visual example, Figure 5.11c-d display side-by-side
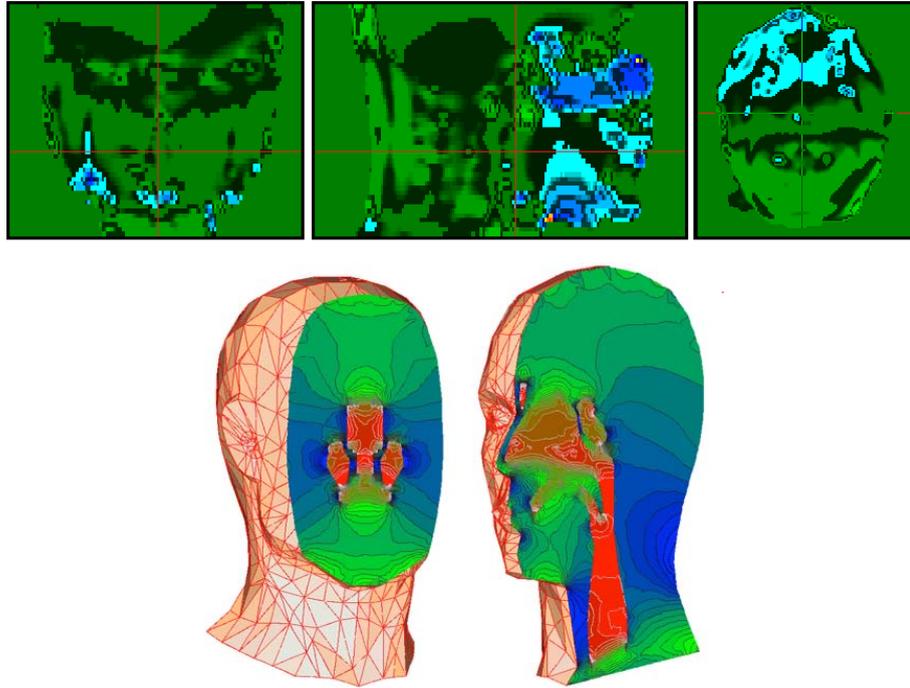
FIGURE 5.12: Upper: Phase map corresponding with the DW-MRI image displayed in Figure 5.11c. Note areas of rapid phase change correspond to previously described EPI distortions. Lower: (Modelled) Static field within the human head (Collins et al. (2002)). Note the high frequency isolines within the maxillofacial region.

comparisons of $450s/mm^2$ and $600s/mm^2$ image results (with otherwise comparable parameters and of the same subject). Note how the inconsistency already present in the cheeks at $450s/mm^2$ (visible as speckling in the DEC map) propagates to the larger muscle masses as b-value increases, prohibiting retrieval of any meaningfully accurate information. This is corroborated by a small sample of the recorded SNR data presented in Table 3.3 (calculated according to Dietrich et al. (2007)) which displays the mean SNR across a multi-layer region of interest positioned over a consistent region within the right masseter muscle – note that (as expected) SNR sharply decreases as the b-value increases.

b-value	mean masseter SNR
$1200s/mm^2$ (3mm isotropic)	7.27:1
$900s/mm^2$ (3mm isotropic)	11.13:1
$600s/mm^2$ (3mm isotropic)	15.43
$450s/mm^2$ (2mm isotropic)	9.71:1
$450s/mm^2$ (3mm isotropic)	24.31:1
$450s/mm^2$ (3mm isotropic)	25:21:1

TABLE 5.3: Mean SNR measured across the right masseter muscle with respect to the b-value used. Note that the final entry in this table was measured from one of the images used in this case study.

While the SNR in case study image is not significantly better (see the final entry in Table 3.4, also Table 3.4 provides mean measurements across 3 key muscles at varying positions in the vertical plane), it does approach the point at which simulations predict crossing fibre resolution may be possible at near orthogonal angles and wider margins of error – conditions that should suffice for this proof of concept work (also, the four repeated measurements should alleviate some of the effects). Finally, if we examine the SNRs with respect to vertical position within the Table 3.3 muscles (Fig. 5.11e), a general trend is seen as one moves upwards, with SNR reaching far higher levels towards the upper edges with a welcome corresponding decrease in CSHD fODF uncertainty (Figure 5.11f, the methods used to produce this figure will be explained shortly). The reason for this is fairly simple; as we proceed upwards we enter the RF coil's region of sensitivity. While not a great result for this work it *does* mean that the MR scanner itself is capable of simulating far stronger echoes than those we have been recording from the majority of these images (owing to the distance between the receiver coils and signal source) and thus when we bring the new coil on-line, there should (barring any design defects) be an immediate and substantial improvement in measured signal quality.

Muscle	Mean SNR
Masseter	25.21:1
Temporalis	24.16:1
Genioglossus	9.62:1

TABLE 5.4: Mean SNR across three key muscles.

Despite the compromises and corrections, the smaller muscles of facial expression are simply out of reach – corresponding image regions are generally close to air/tissue boundaries and as a result are too distorted to reliably recover any useful information. The two regions defined in Section 5.2, however, fair reasonably well within most recorded images. Both areas contain (relatively) large muscles of mastication which, at certain points, intersect at angles that simulations predict are within our grasp to resolve, providing a useful testbed for exploring the application of CSHD to muscle tissue.

5.4.1 Uncertainty In Vivo: A bootstrap analysis

Before proceeding to muscle segmentation, the nature of the image used in this case study opens up a limited opportunity for more quantitative analysis. The four repeated images – allowing for small variations due to the b-matrix rotation applied during motion correction – provide four repeat measurements along each gradient direction. This opens up the opportunity for a bootstrapping experiment (very similar to Jones (2003)) that will (a) provide the opportunity to provide an in vivo analogue to the previous simulation experiments, demonstrating the real world effects of calibration on a low anisotropy CSHD result and (b) gauge the levels of uncertainty within currently plausible image

acquisitions and, by extension, gain insight into the likelihood of reliable crossing fibre resolution in vivo.

Experimental set up was relatively simple. Firstly, 6 non diffusion weighted images were selected at random (with replacement), from the 24 image pool; then, for each for the 60 general gradient directions, one of the four repeat measurements was selected (at random) and general gradient vector replaced with the corresponding rotated (motion corrected) counterpart. This process was repeated 500 times to generate 500 potential instances of a 60 direction, 6 b0 scan result. Each image instance was then subjected to six individual CSHD deconvolutions (explained shortly) and all supra-threshold ($fODF_{mag} \geq 0.1$) fODF peak orientations and magnitudes recorded. For each of the six deconvolutions the following were calculated (a) the mean number of supra-threshold fODF peaks per voxel and (b) voxel-wise 95% confidence intervals (Jones (2003)) in the primary peak orientations. For reasons that will become apparent, the primary peak orientation is defined in two ways, producing two separate confidence results. (1) As the supra-threshold fODF peak most closely subtending the largest fODF peak produced in the corresponding voxel of the full (24 b0, 240 DW) concatenated image which is acting as a gold standard in this context; and (2) the largest fODF peak regardless of orientation.

5.4.2 Data Driven Calibration in vivo: Six Deconvolutions

The selection of exemplar signals for data driven CSHD calibration, as covered in the previous chapter, is generally achieved by selecting a threshold anisotropy, above which all matching voxels can be assumed to contain a non PVE example of the tissue type of interest. For white matter, this assumption appears to work reasonably well, selecting a threshold of 0.8 (Tournier et al. (2004)) will ensure that only straight, isolated, white matter bundles are isolated, providing exemplars of the CSHD's response function requirements. The problem, unfortunately, is that it only works when the fibre type of interest constitutes the most anisotropic signals likely to be produced within the imaged region. For white matter this is true, the only voxels likely to produce higher anisotropy than straight non-PVE bundles are those in the background image regions which are easily excluded when using tools such as BET (Smith (2002)) to eliminate non-brain image content. Muscle on the other hand is a low anisotropy fibre type, even in isolation and following a perfectly straight trajectory, its anisotropy rarely exceeds ≈ 0.35, placing it squarely within the realm of, for example, crossing white-matter fibres. Even filtering to a narrow band in the 0.3 region is insufficient to reliably extract only straight non-PVE muscle fibre exemplars (e.g. Fig 5.14d), thus new data-driven calibration methods must be devised.

The simplest option is to manually highlight exemplars. To achieve this one should first manually examine the image for large (multi-voxel width) straight fibrous structures with diffusive properties similar to the fibre bundles of interest. Once such structures are located, DT-MRI seed way-gates should then be positioned over the widest orthogonal cross-section and highly constrained ($\leq 10°$ angular threshold) DT-MRI streamlines allowed to propagate. The combination of a tight angular threshold and DT-MRI's inherent sensitivity to PVE will then ensure that streamlines only propagate core non-PVE regions of the selected structure, meaning that any intersected voxel should provide an excellent exemplar for response function calibration.

The first two de-convolutions use exactly this method for response function selection; waypoints were manually positioned over the splenius muscles (Fig. 5.13a) which, while not technically a maxillofacial muscle (part of the neck muscle complex), displays equivalent diffusive properties (FA and MD of $0.321/1.45 \times 10^{-3} mm^2/s$ respectively compared to $0.317/1.51 \times 10^{-3} mm^2/s$ matching a core sample of the temporalis muscle) and possesses the added benefit that its relative size and position (back of the neck) keep it away from major image artefacts. Using the intersected voxels as an exemplar, response function estimations were then performed at Lmax 4 and 6. Taking guidance from the previous simulations and preliminary experimentation, an $Lmax = 8$ calibration/deconvolution was deemed unnecessary since no significant benefits were observable.
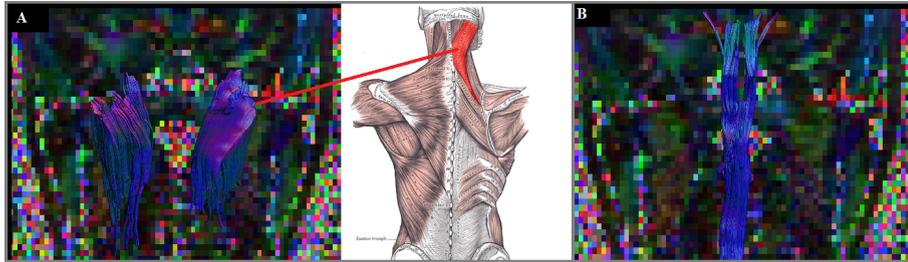
FIGURE 5.13: Manually segmented exemplar data (A) the splenius muscles (with accompanying illustration) and (B) the spinal cord

The disadvantage of this manual calibration is simply that it *is* manual; users must spend time labelling suitable structures and checking the validity of selective DT-MRI tractography. It would be preferable to design an automated system (like the white-matter threshold) to achieve a similar task with minimal input. To this end, an automated calibration scheme has been devised, combining a voxel coherence measure (dubbed V_c) and general thresholds based on a priori knowledge of the target's fractional anisotropy and mean diffusivity (MD). While FA and MD are standard measurements (calculations described in chapter two), V_c is defined as follows:

$$V_c = k \sum_{n=1}^{26} w(D_t \cdot \mathcal{D}_n) \tag{5.1}$$

where D_t is the DT-MRI principal eigenvector for the target voxel and D_n is the equivalent for one of its n neighbours. In this formulation w provides a weight signifying the importance of each dot product based on the distance between the centre points of D_t and D_n:

$$w = \begin{cases} 1 & \text{if voxels share face} \\ 1/\sqrt{2} & \text{if voxels share edge} \\ 1/\sqrt{3} & \text{if voxels share vertex} \end{cases} \quad (5.2)$$

and k (the inverse of the sum of the 26 applied wieghts) scales the summation to between 0 (where D_t is perpendicular to all D_n) and 1 (where D_t is parallel to all D_n)

$$k = \frac{\sqrt{6}}{\sqrt{216} + \sqrt{432} + \sqrt{128}} \approx 0.0523 \quad (5.3)$$

thus V_c is essentially a normalised pairwise weighted sum of vector dot products between a target voxel and its neighbours, returning 1 if the candidate is perfectly aligned and 0 when orthogonal. Selecting a suitable V_c threshold will thus highlight voxels from within locally consistent regions (i.e. the core of a straight fibre bundle), but provides no distinction between different *types* of fibre – which is then achieved by taking the unison with a second mask highlighting voxels with the correct diffusive properties.

To generate a muscle fibre response, thresholds were set at $V_c \geq 0.9$, $FA = 0.32 \pm 0.13$ and $MD = 1.43 \times 10^{-3} \pm 0.49 \times 10^{-3} mm^2/s$ (the latter two being mean values $\pm$ two standard deviations across several samples of masseter muscle). An example of each stage of the masking process is presented in Figure 5.14. Figure 5.14b shows the V_c map which is then thresholded to produce the mask in Figure 5.14c, note that the spinal cord (highlighted red) passes this stage. Figure 5.14d then displays the output of the MD/FA threshold, note now that the spinal cord is absent. Finally, on unison of the two masks (Fig. 5.14e), the result is a concise mask of each large (non-complex) muscle structure passing through the slice (masseter, medial pytergoid, rectus capitis, spinalis capitis and trapezius, labelled 1-5 respectively) while avoiding the inclusion of (potentially PVE) voxels at their various intersections. Again, CSHD was calibrated at Lmax 4 and 6 using the exemplars highlighted by this process. An interesting point of exploration will be the difference between this and the previous manual calibration technique, should the results appear similar it should bode well for the V_c method as an automated alternative.

Finally, in the interests of comparison and to demonstrate the in vivo effects of response function overestimation, a more highly anisotropic response function was generated using the manual method centered over the spinal cord (with mean ≈ 0.53 FA).
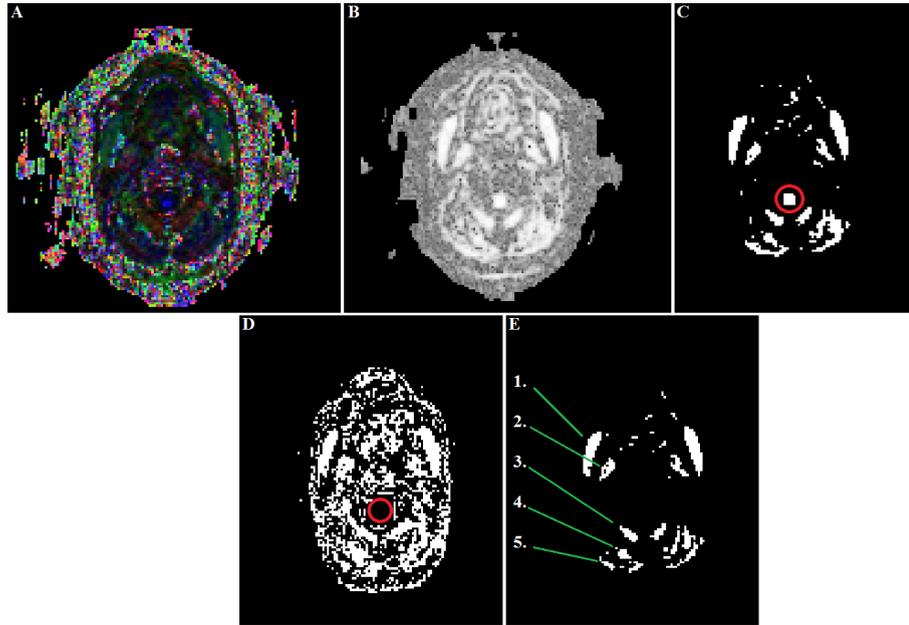
FIGURE 5.14: Illustration of the automated exemplar selection process. (A) DEC map of the target slice. (B) Corresponding V_c map. (C) Thresholded ($V_c \geq 0.9$) V_c map, undesirable spinal cord highlighted (red). (D) Thresholded MD/FA mask, spinal cord removed. (E) Final product ($C \cap D$), labels 1-5 correspond (respectively) to the masseter, medial pytergoid, rectus capitis, spinal capitis and trapezius.

5.4.3 Bootstrap results

Figure 5.15 displays statistics regarding the number of supra-threshold fODF lobes produced by deconvolutions calibrated (a) automatically, (b) manually and (c) manually against spinal tissue at $lmax = 6$. The upper panels depict the voxel-wise mean, middle panels (b and c only) depict absolute difference from (a) and the lower panels provide a histogram representation of the whole volume count. Note that for both tissue specific calibrations, the indicated number of fibres in known low-PVE regions (such as the core of the masseter and spinal muscles – labelled in slot (a) of the second row) remain close to the expected value of one; whereas in regions where crossing fibre is known to exist, e.g. the tongue, this figure rises to approximately two, again matching expectations. Encouragingly, the mean only exceeds two in either corrupted regions (e.g. the frontmost air/skin interfaces) or null-signal regions (outside the head, or within the signal void left by the mandible) where the outcome is more or less inconsequential.

Comparing these results to fODFs originating from the spinal calibration, the difference is immediately noticeable. Examining the upper column, almost all regions within the image indicate four or more fibre populations which, being obviously impossible, coincides well with the predicted appearance of spurious fODF lobes at high calibration discrepancies. The difference becomes even more apparent when examining the middle and lower rows. Beginning with the middle row, if visualised on the same scales (as is the case for this figure) the absolute difference between manual and automated tissue
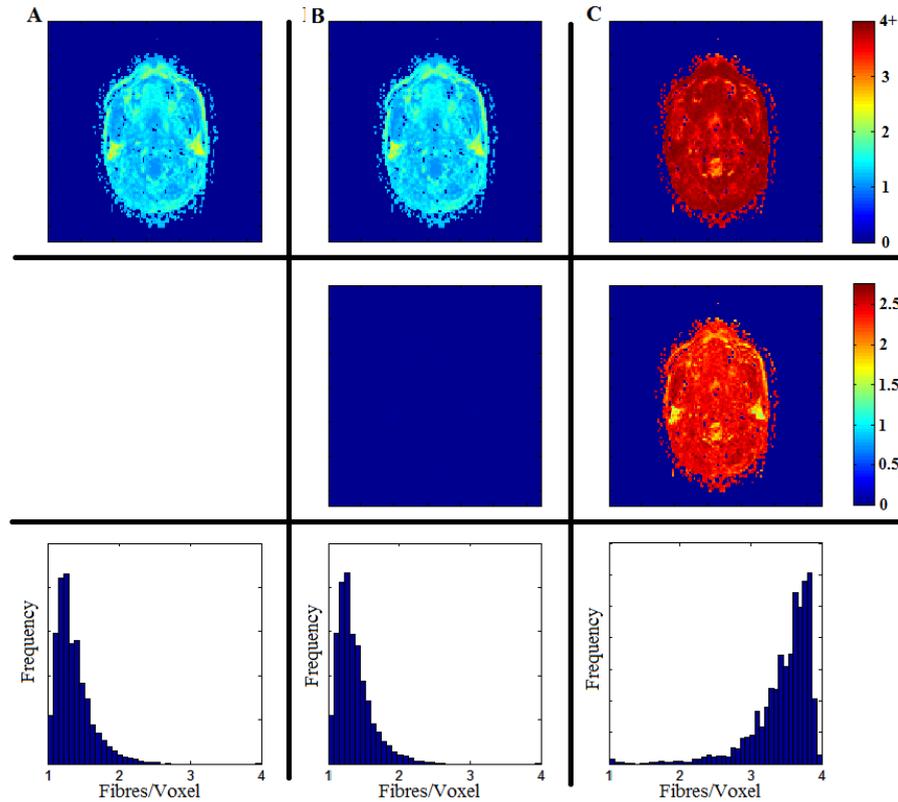
FIGURE 5.15: Bootstrap fibre count results at $lmax = 6$ for (A) automatically calibrated CSHD (B) manually calibrated CSHD and (C) spinally calibrated CSHD. Upper row displays mean fibre count over the 500 repetitions (capping at 4+ fibres). Middle row demonstrates absolute difference between automated and manual/spinal respectively. Lower row displays a volume wide histogram of fibre counts.

muscle-specific calibrations is almost imperceivable – though for the record, the mean voxel-wise difference in fibre count is 0.0015 with a standard deviation of 0.02, indicating excellent agreement – whereas differences between the automated and spinal result are far more apparent (averaging 2.37 fibres per voxel with a standard deviation of 0.18). Examining the histograms emphasises the matter further; with the brain extracted, only the very core of the spinal cord allows results in single-fibre indications from the spinal calibration (as would be expected, since this is exactly the tissue from which it is derived), with the majority of the other voxels strongly skewed towards ≥ 4 fibres in stark contrast with both appropriately calibrated results.

Proceeding to 95% confidence intervals, Figure 5.16 displays results for the automated, manual and spinally calibrated de-convolutions (organised as before). Beginning with the upper panels showing confidence intervals for bootstrapped peaks minimally subtending the 'gold standard' orientation; initial observations show that within large predominantly single fibre structures (e.g. within the masseter or neck group) automated and manual muscle-specific calibrations possess relatively low confidence intervals and again agree strongly with one another. Interestingly, average confidence intervals across spinally calibrated CSHD appear far lower than that of tissue specific results (though

notably higher within low PVE muscle regions). The reason for this is actually fairly simple, a large number of these voxels contain multiple spurious fODF peaks which, one of which (by power of numbers) will almost certainly subtend any given orientation at a low angle. This can be verified by examining the second row of Figure 5.16 which depicts stability of the largest peak orientation throughout the bootstrap iterations. Note that in this case the confidence intervals across the spinally calibrated result are almost universally high ($70 - 90°$ range) whereas tissue-specific calibrations still produce acceptable confidence intervals within well defined muscle regions. For reference, the mean voxel-wise difference between manual and automated muscle calibrations was an $0.82°$ expansion of the 95% interval for manual over automated with $2.30°$ standard deviation (indicating that the automated result does indeed perform well); while spinal calibration is consistently worse, increasing confidence intervals by an image-wide average of $19.38°$ with $14.56°$ standard deviation.
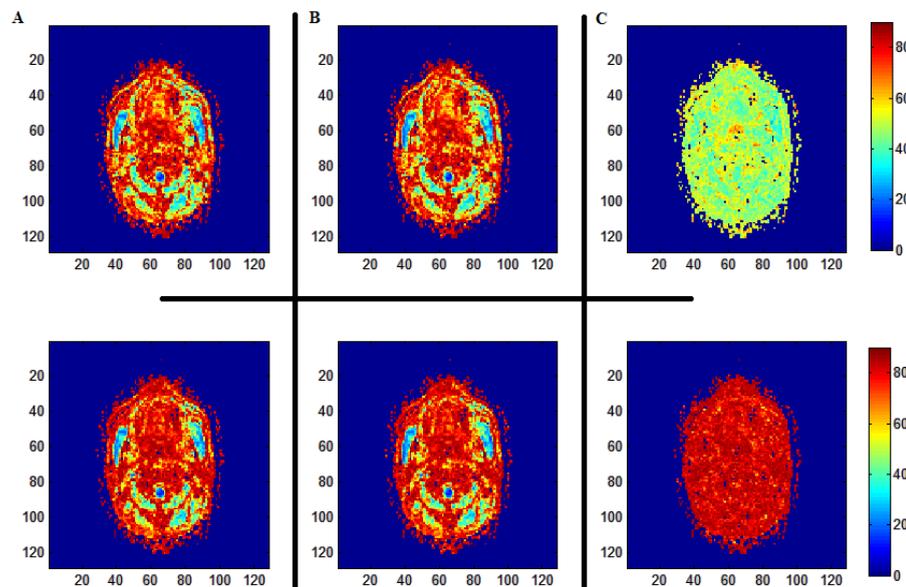
FIGURE 5.16: Bootstrap 95% confidence intervals (degrees) at $lmax = 6$ for (A) automatically calibrated CSHD (B) manually calibrated CSHD and (C) spinally calibrated CSHD. Upper row demonstrates confidence in the fODF peak most closly subtending the "gold standard" result while the lower row demonstrates confidence in the largest fODF peak.

If we then examine results at $lmax = 4$ (omitting manual calibration due to its repeated similarity with the automated result), two points become readily apparent. Firstly, spinal calibration continues to produce large numbers of spurious fODF peaks (column 5.17b, upper panel), resulting in near identical performance to the previous $lmax = 6$ experiment. Secondly, if one examines the masseter for example, the $lmax = 4$ deconvolution being less sensitive to noise produces a larger portion of sub $15°$ confidence intervals (Fig. 5.17a, rows 2 and 3) in comparison to the $\approx 20°$ intervals produced in equivalent voxels at $Lmax = 6$ (Fig. 5.16a,b row 1, mean $4.3°$, standard deviation

increasing by 1.3°). While the increased confidence is encouraging, it comes at the expense of a reduction in the number of indicated fibre populations within the tongue and regions of the masseter that *may* correspond with boundaries between the three components. On this data alone it is difficult to predict whether this is genuine loss of angular resolution, or if the perceived reduction is due to spurious fODF peaks in the $Lmax = 6$ result that are eliminated in the more resilient $Lmax = 4$ deconvolution. While the optimal lmax choice remains open at this point (to be examined in greater detail in the following section), one conclusion is certain – overestimating the anisotropy of the response function definitely will produce spurious fODF peaks and thus earlier simulated results can be confirmed.
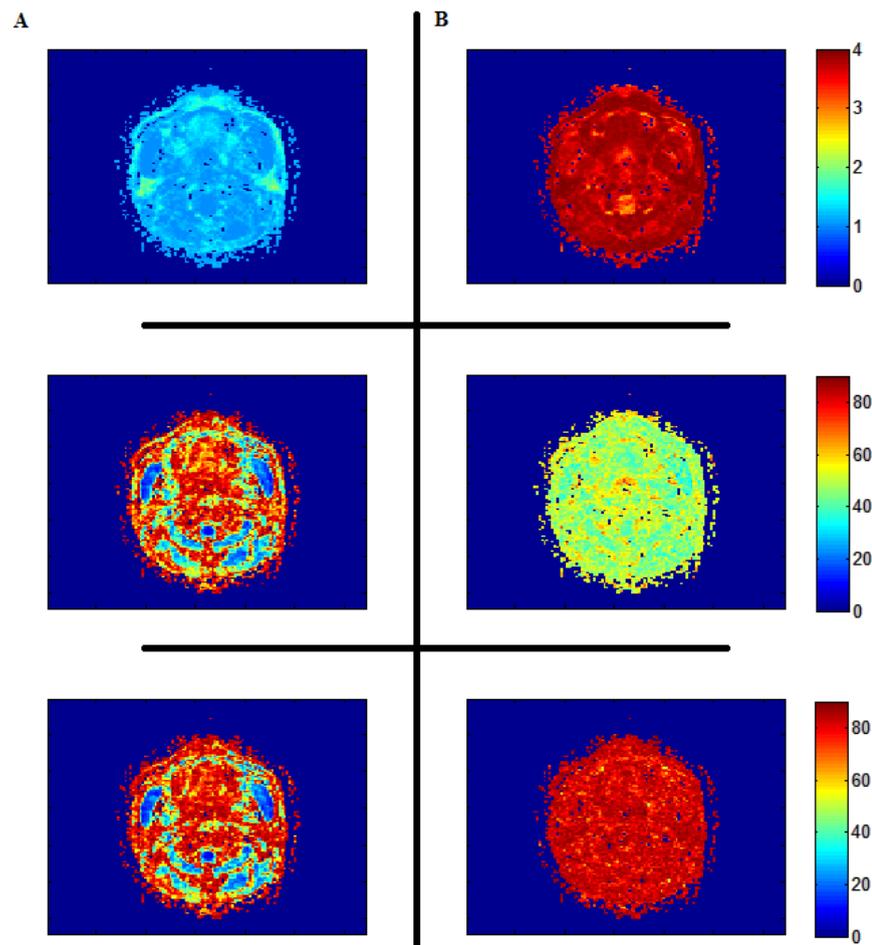
FIGURE 5.17: Bootstrap results at $lmax = 4$ for (A) automatically calibrated CSHD and (B) spinally calibrated CSHD. Upper row displays mean fibre count, middle row displays confidence about gold standard and lower row displays confidence in the highest magnitude peaks.

5.5 In Vivo Tractography and Dissection

All that remains now is to perform tractography and demonstrate the extent to which maxillofacial musculature may currently be reproduced. To ensure the maximum chance of success the full 24 b0, 240 diffusion weighted direction dataset was used and the primary calibration conducted through the automated method (with previously discussed parameters) at Lmax 4 and 6, though we also briefly demonstrate tractography results using the manual muscle and spinal response functions to visually confirm the bootstrap (dis)similarities. Aside from the varied calibrations, CSHD tractography parameters were as follows:

1. Seed points spaced over a 2mm isotropic grid.

2. 0.5mm step size.

3. 30° angular threshold.

4. $fODF_{mag} \geq 0.1$ peak magnitude threshold.

all of which (where applicable) are consistent with other muscle-related work. While the resulting analysis is purely qualitative, since no gold-standard is obtainable without dissection, the anatomical plausibility of each reconstruction has been discussed with a clinical practitioner with knowledge of the subjects maxillofacial structure.

5.5.1 Dissection: Muscle Group 1

Beginning with the larger muscles of mastication surrounding the temporomandibilar joint, it is first important to discover whether CSHD can be expected to achieve reliable results across 'trivial' muscle fibre targets. As previously explained, prime candidates for this assessment are the temporalis (Fig. 5.18) and masseter (Fig. 5.19). Beginning with the temporalis; Figure 5.18 displays the result of three segmentation using (a) automated calibration at $Lmax = 4$, (b) automated $Lmax = 6$ and (c) manual (muscle) $Lmax = 6$. All reconstructions bare good resemblance to known anatomy. While recovery of the upper fanned portion of the muscle is restricted by reaching the vertical limit of the image, the vast majority of visible muscle tissue is recovered with, particularly at $Lmax = 4$, excellent fidelity as the muscle passes the zygomatic arch and down into insertion at the coronoid process (the accompanying illustration provides an approximation of the recovered muscle mass). Compared to $Lmax = 4$, $Lmax = 6$ deconvolutions (Fig. 5.18b,c) demonstrate visible gaps along the upper image boundary and again as the muscle turns at the zygomatic arch. If we examine the DW-MRI image at these locations (highlighted red), there are clear indications of distortion/corruption that may explain the discrepancy since it is well known (with simulations/bootstrap experiments

in this and previous chapters confirming) that lower order harmonic truncations are less susceptible to noise. Comparing the two $Lmax = 6$ results directly, there is little significant difference – indicating that as predicted by the bootstrap experiment, out automated calibration method provides a working alternative to manual labelling.
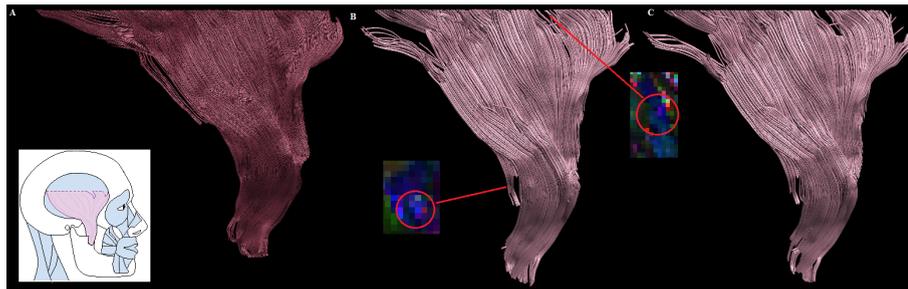
FIGURE 5.18: CSHD reconstructions of the temporalis muscle using (A) an automated $Lmax = 4$ calibration. (B) an automated $Lmax = 6$ calibration and (C) a manual (muscle specific) $Lmax = 6$ calibration.
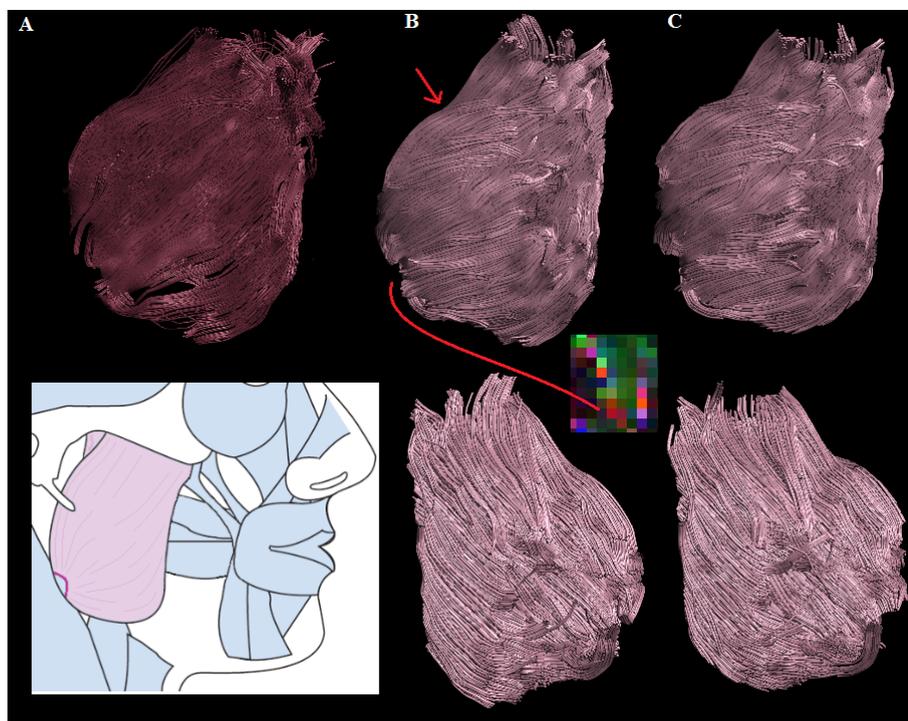
FIGURE 5.19: CSHD reconstructions of the masseter muscle using (A) an automated $Lmax = 4$ calibration. (B) an automated $Lmax = 6$ calibration and (C) a manual (muscle specific) $Lmax = 6$ calibration.

Masseter reconstructions (Fig. 5.19) are again consistent with anatomical knowledge; all three segmentations exhibit the expected shape (with the exception of a small portion of the lower trailing edge attributable to an image artefact, see illustration) however, this time $Lmax = 6$ appears to produce the superior result, providing clearer definition between the external and mid fibrous components (juncture labelled with red arrow). It would appear that due to the sheltered position of the masseter (away from the main sources of distortion), the above-average local image quality is sufficient to realise the

improved angular resolution afforded by *Lmax=6* – something that was hinted at during the bootstrap analysis. Again, both *Lmax=6* calibration results are equivalent and, for this reason, only the automated results will be shown from this point forwards.

If we then examine the medial pytergoids (Fig. 5.20), the relative simplicity of the muscle structures permits equivalent reconstructions from both 4th and 6th order deconvolution thus, as might be expected, where neither SNR nor complex architecture are an issue, Lmax has little effect on solution outcome.
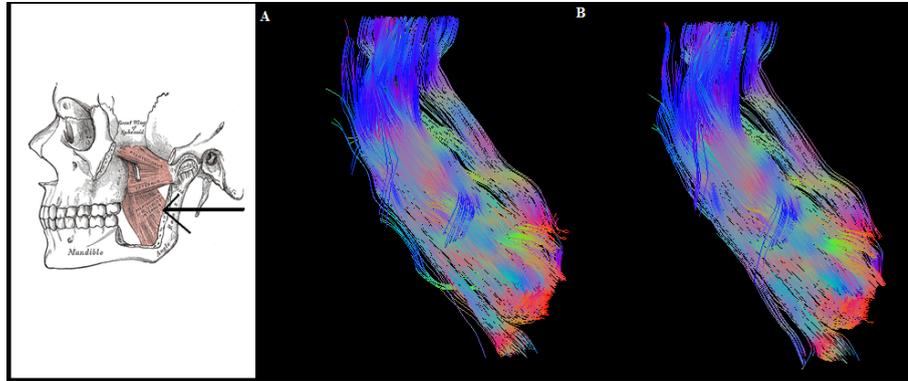
FIGURE 5.20: CSHD reconstructions of the medial pterygoid muscle. Note that for this relatively simple, equivalent reconstruction is achieved at both (A) $Lmax = 4$ and (B) $Lmax = 6$
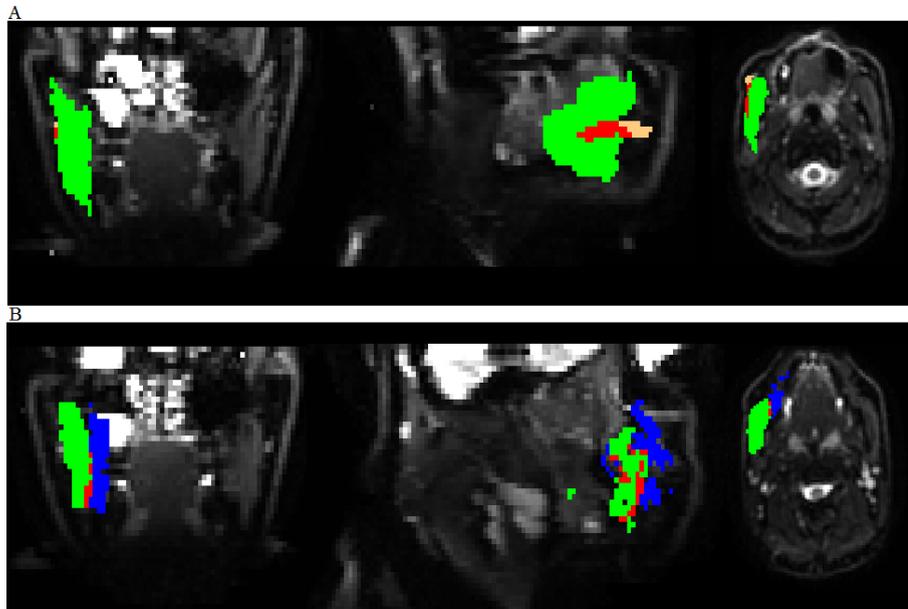
FIGURE 5.21: Muscle intersections (red, automated $Lmax = 6$ deconvolution) superimposed over the subjects b0 image. (A) Masseter-risorius (green-brown) intersection. (B) Masseter-buccinator (green-blue) intersection. Further intersections exist outside of the three planes shown.

Proceeding to the final muscles in group one, reconstruction of the buccinator and risorius proves to be far more problematic than previous muscles, mainly because reliable imaging of the cheek, and by extension muscle passing through it (i.e. the two muscles in question), is not possible with current imaging arrangements. The principal effect of this restriction is that reconstructions are limited to the rearwards portion of each muscle, however, as these are the regions of interest (contact with the masseter, Fig. 5.21, red indicating a shared voxel), examination is still warranted.
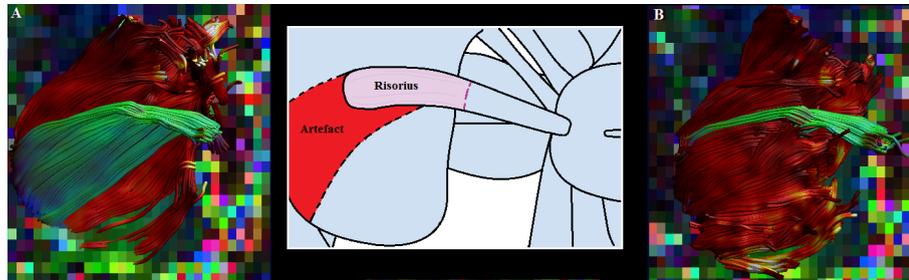
FIGURE 5.22: CSHD reconstructions of the risorius using (A) an automated $Lmax = 4$ calibration and (B) an automated $Lmax = 6$ calibration. For visual reference, sections of the corresponding masseter reconstruction are also included (risorius green, masseter red) – note that at the origin (left side) the risorius incorrectly blends into the masseter, creating the appearance of a larger muscle. The central caricature depicts the true shape of the risorius, labelling an approximations of the recovered portion and artefactual extension in pink and red respectively.

Beginning with the risorius, at both $Lmax = 4$ and $Lmax = 6$ there are clear indications that the muscle exists within the recovered fODFs. Streamlines are visible departing from the masseter along trajectories consistent with that of the risorius (Fig. 5.22) though in both cases these streamlines then artefactually merge with the masseter at locations consistent with the muscle origin – put simply the risorius appears to be emerging from the masseter rather than sitting on top of it, an artefact caused in equal parts due to the low angular separation and large voxel dimensions. The encouraging factor here, though, is that in both cases the crossing fibre configuration was resolved for a limited portion of the streamline length.

Finally, like the masseter, the buccinaor consists of three interwoven components and, like the masseter segmentations, the $Lmax = 4$ result (Fig. 5.23a) appears to lack the angular resolution required to correctly reproduce component interactions. In comparison with the $Lmax = 6$ result (Fig. 5.23b) there is arguably an improvement in recovered muscle volume, noticeably so towards the lips (again, the lower lmax results in better tracking through poor data) but if we examine the middle bundle there is a tell-tale difference. Returning to the descriptions provided in Section 5.2 – "the central band of the buccinator twists at the modiolus (corner of the mouth) such that the upper component inserts into the lower orbicularis oris (lower lip) and the lower into the upper". This twisting is clearly visible in the $Lmax = 6$ result (labelled red) but distinctly lacking in the fourth order truncation.
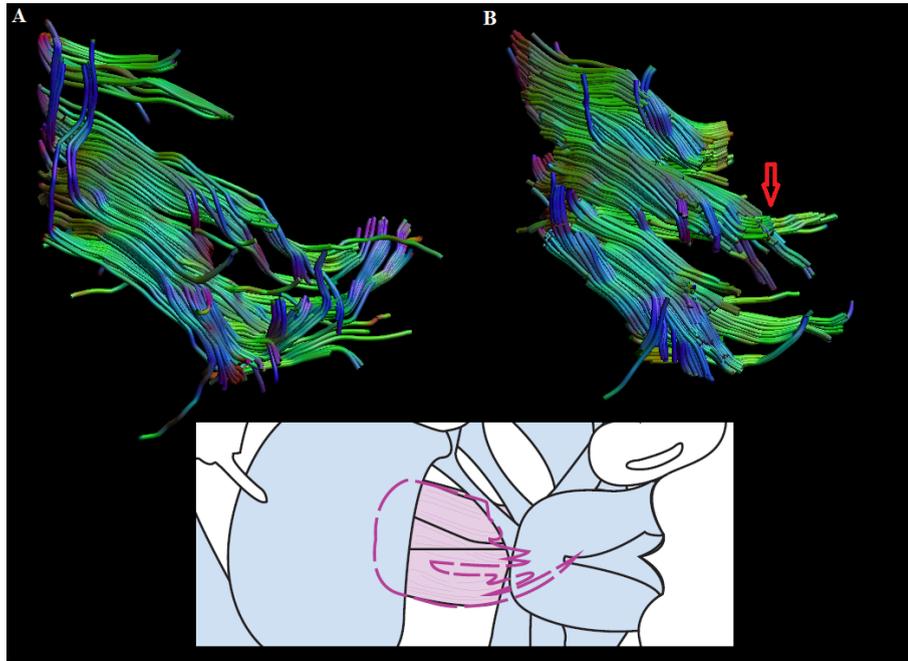
FIGURE 5.23: CSHD reconstructions of the buccinator using (A) an automated $Lmax = 4$ calibration and (B) an automated $Lmax = 6$ calibration. Note that the lower order truncation results in a failure to resolve the twist in the central band.

5.5.2 Dissection: Muscle Group 2
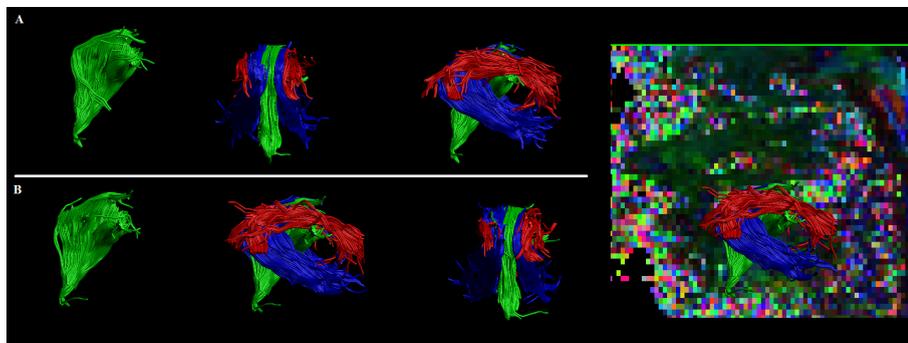
FIGURE 5.24: CSHD reconstructions of the tongue region using (A) an automated $Lmax = 4$ calibration and (B) an automated $Lmax = 6$ calibration. Column one contains reconstuctions of the genioglossus muscle in isolation (green colour, given that it the view is disrupted in situ), columns 2 and 3 display the muscle group in situ from two axially aligned view points and column 4 displays the $Lmax = 6$ result framed within the colour encoded DEC map to aid understanding of the spatial positioning within the volume.

Figure 5.24 depicts reconstructions of the genioglossus and superior/inferior longitudinal muscle fibres constituting a significant portion of the tongue. On visual inspection, reconstructions at either Lmax do not appear significantly different, with the possible caveat that at $Lmax = 6$ (lower row), the streamlines representing superior longitudinal fibres (red) appear slightly less smooth – most likely related to the increased uncertainty in primary fibre orientation predicted by the bootstrap result. Examining the genioglossus, both tractographies demonstrate the appropriate fanning structure with

good convergence at the pointed insertion into the origin at the medial spline of the masseter. The only major defect in genioglossus reconstruction (for either Lmax) is that the fanning portion should most likely extend a further $\approx 20°$ in the anticlockwise direction, though recalling the image quality in this area (Fig. 5.9) this deficiency is unsurprising. Omitting the obvious differences between bovine and human tongue and the missing genioglossus fraction, these reproductions compare well with the ex vivo DSI results published by Gilbert et al. (2006) (reproduced in Fig. 5.2). The only remaining significant difference is the lack of transversus fibres, however, if one places a waypoint directly over the medial spline of the tongue (Fig. 5.25, $Lmax = 6$), evidence of their existence within the fODF results becomes apparent, though the resultant streamlines tend to either terminate prematurely after a few millimetres (i.e. not crossing the full extent of the tongue) or prematurely merge with the longitudinal fibre groups due to a combination of noise induced bias and/or lack of angular resolution.
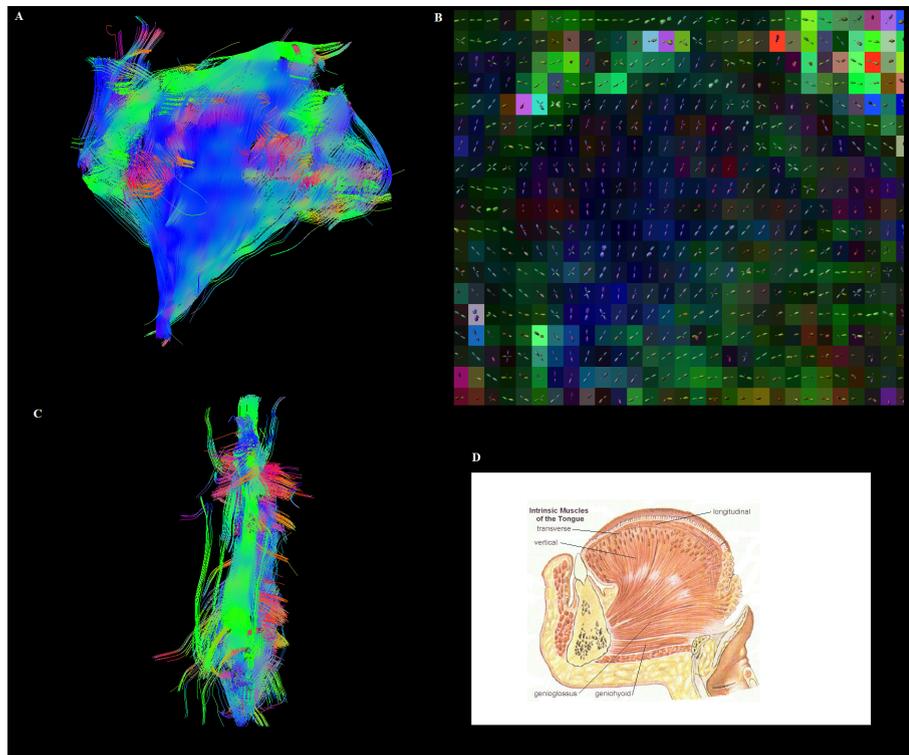
FIGURE 5.25: CSHD reconstructions across the medial spline of the tongue at $Lmax = 6$. Red streamlines/fODF lobes are indicative of partially resolved transvesus muscle fibres. Comparing subfigures (A) and (B) to the anatomical drawing in (D) the correlation between the origin of partially complete transverse streamlines and positioning of the transverse fibres should become apparent. Subfigure (C) provides a top-down view of (A) demonstrating the premature termination in the transverse domain.

5.5.3 Miscellaneous Segmentations

As a conclusion to this section two further interesting results are presented. The first, Figure 5.26, demonstrates the effect of miscalibration on CSHD tractography. Figure 5.26a outlines an area of non-PVE temporalis muscle tissue; if we zoom in and plot the CSHD fODF's for this region (Fig. 5.26b), an appropriate calibration produces (as expected) largely single peak fODFs with orientations matching those of the temporalis in this region. If we then overestimate the CSHD response (Fig. 5.265b lower), spurious fODF lobes appear as expected. The global result (Figure 5.26c) is that the spurious fODF peaks produced by the overestimation lead to numerous "run away" streamlines propagating along artefactual trajectories as one spurious fODF lobe minimally subtends another (an effect postulated in the previous chapter) while the appropriately calibrated result generates a reasonable reconstruction of the muscle in question. Along with the previous simulations and bootstrap work, it should now be beyond question that appropriate calibration is the key to accurate fODF retrieval.
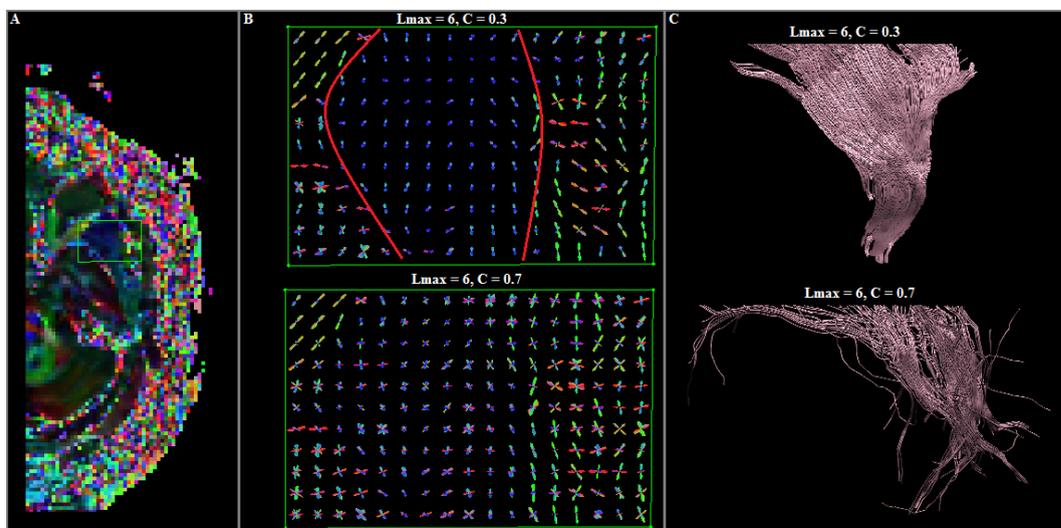
FIGURE 5.26: Reconstructions of the temporalis muscle. fODFs in (B) are drawn from the signals originating in the green boxed area of (A) – a core sample of the temporalis muscle (red labelling in upper image shows approximate boundaries). Examining the resultant tractographies (C), the global effects of spurious fODF peaks become obvious.

Secondly, to provide a final overview, Figure 5.27 displays all segmented muscles in situ. While obviously lacking the more complex muscles of expression that were initially part of the target for this work, given the materials available the quality of these segmentations bode well for future progress.
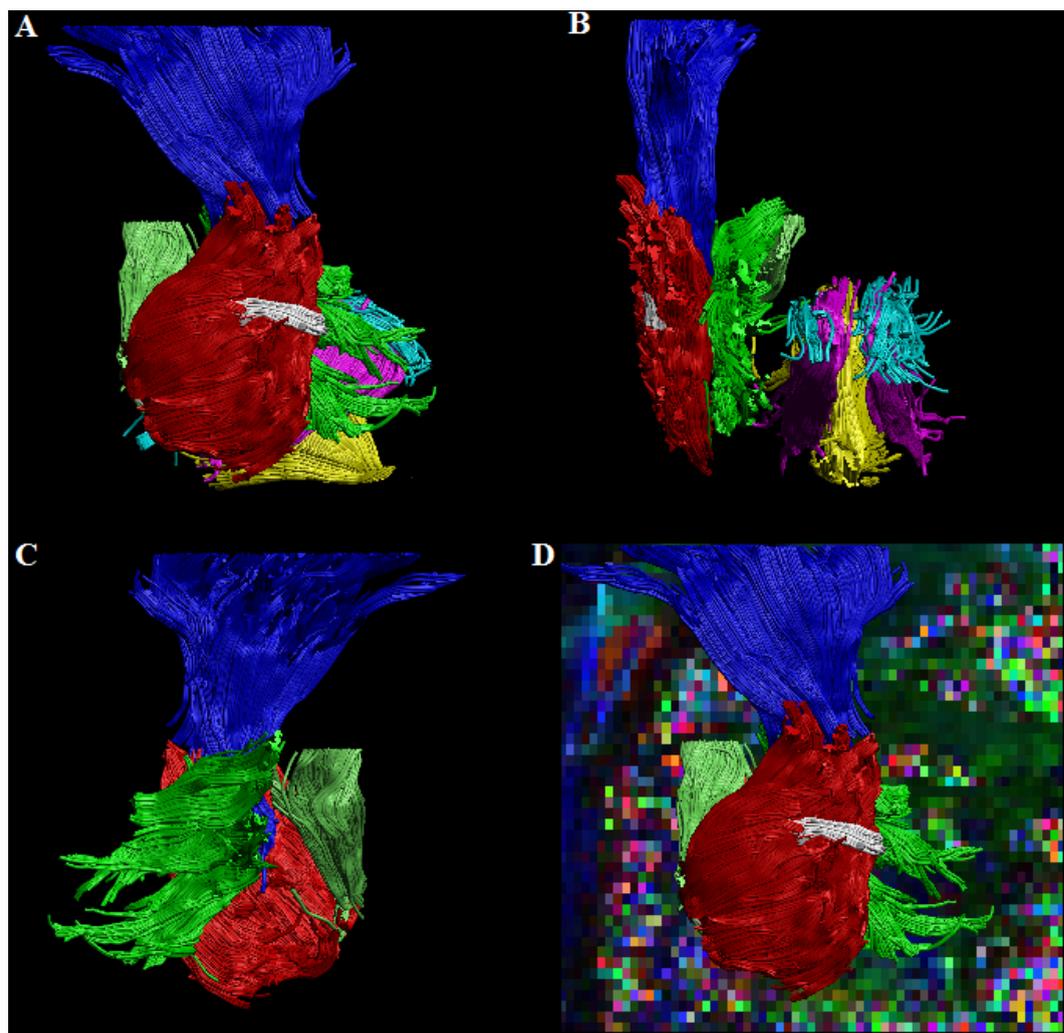
FIGURE 5.27: An in situ representation of all muscle reconstructions discussed within this section. Red: Masseter. Blue: Temporalis. Green: Buccinator. Light Green: Pterygoid. White: Risorius. Orange: Buccinator. Yellow: Genioglossus. Cyan: upper longitudinal fibres. Magenta: Lower longitudinal fibres. (A) External view (right side) (B) Frontal view. (C) Left view (tongue complex removed). Note that the Pterygoid muscle shown in this illustration is the right pterygoid, previous illustrations concern the left pterygoid which is better preserved within the image data. (D) External view of the Group 1 muscles with background to aid spatial localisation.

5.6 Discussion

5.6.1 Imaging

Both the simulations and in vivo imaging stress the importance of two key factors for accurate resolution of crossing muscle fibre in vivo. Firstly, and most importantly, maintaining a high SNR should be a primary concern for any group attempting to apply CSHD to muscle tissue. Secondly, it is unlikely that given the levels of diffusion weighting used in current DT-MRI based studies will provide more than $20°$ of reliable angular resolution either side of an orthogonal crossing for a two fibre system. Thus, to advance in vivo muscular DW-MRI, a revision of the current acquisition norms is required.

A well known problem in DT-MRI is the increased in noise induced bias as fractional anisotropy decreases, i.e. for a given SNR, the perturbations of parameters gained from a low anisotropy signal will be disproportionately larger than the perturbations of parameters gained from high anisotropy signals. The reason for this is fairly well known, at low anisotropies there is (by definition) a smaller difference between axial and radial signal attenuations and by extension the resulting DT-MRI eigenvalues. With this being the case, small changes (with respect to overall signal amplitude) can have large effects on the resultant model fit – a small artefactual increase in indicated diffusivity along some direction may be sufficient to bias the diffusion tensor away from its true orientation. Where these directionally dependant attenuation differences are larger, i.e. high anisotropy signals, an equally small change will have lower effect relative to the overall diffusion profile and thus incur less bias. As a result, for some given SNR, the reliability of any derived estimates is likely to correlate with the underlying anisotropy of the fibrous tissue (and effect known as sorting bias, Martin et al. (1999), Basser and Jones (2002)).

This is in essence the same issue that affects our ability to resolve crossing muscle fibre. The combination of low diffusion weightings and low diffusion anisotropy limit the dynamic range of the observed signal attenuations across a muscle fibre's diffusion profile. As an example, at $450s/mm^2$, the signal generated by two muscle fibres (simulated as two $FA \approx 0.3$ compartments separated by $90°$) is, at least by the eye, barely distinguishable from a sphere (Fig. 5.28a). At higher anisotropy, e.g. $FA \approx 0.9$, this is not the case – while still not appearing to match the classic 'combination of diffusion peanuts' shape (Jones and Basser (2004b)), there is at least readily visible anisotropy in the diffusion signal (Fig. 5.28b). As shown, CSHD is highly dependant on the assumption that an observed signal was generated by convolution of some fODF with a *known* fibre response function; however, by adding noise to an observed signal this assumption understandably begins to break down and, since the response function is fixed a priori, can only result in a spurious fODF recovery containing biased fibre orientation and potentially even spurious fODF peaks. While this is true of all fibrous targets, the diffusion properties of

muscle result in fairly indistinct diffusion weighted signals and thus, as with DT-MRI, the effects of noise on fODFs derived from muscle tissue will be disproportionately higher than those derived from other more isotropic tissue.
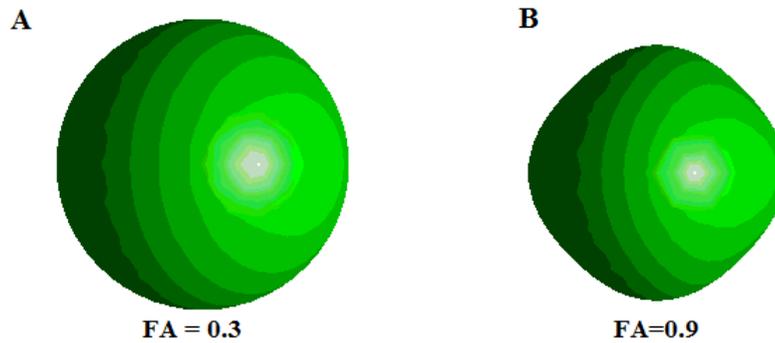
FIGURE 5.28: Visual representation of the DW-MRI signals produced by crossing fibres at anisitropies of (A) $FA = 0.3$ and (B) $FA = 0.9$ at low diffusion weightings $(b = 450s/mm^2)$.

The most commonly acknowledged method for sidestepping this problem would be to perform deconvolutions at the lowest possible Lmax. As shown by Anderson (2005) and already discussed in detail within this and the previous chapter, reducing Lmax has the effect of increasing the coefficients on the remaining spherical harmonics (since the signal magnitude does not change and must now be described by fewer basis functions) and, much in the same way that higher anisotropies benefit DT-MRI, improves the deconvolution's resilience to small perturbations in signal magnitude. Unfortunately, while sound in theory (and reduction from $Lmax = 8$ to $Lmax = 6$ or $Lmax = 4$ shows the expected benefits in simulation), both the bootstrap results and qualitative tractography analysis show clear indications (reduction in multiple fibre orientations in the tongue between Figures 5.15 and 5.17, failure to resolve the more complex fibre interactions in the masseter, Figure 5.19, or buccinator, 5.23) that even within large and relatively simple muscles, $Lmax \geq 6$ is the bare minimum permissible truncation point if one wishes to achieve the required angular resolution for accurate in vivo dissection/modelling. If we then accept $Lmax = 6$ as that minimum, simulations suggest that $\approx 750s/mm^2$ with an $\geq 50:1$ SNR would be required to achieve even 80% reliability within a 20° threshold. If we then recall that these simulations only cover a perfect 50/50 signal contribution split, and that the previous chapter shows that ability to resolve crossing fibre diminishes as contributions become asymmetric, the real minimum requirements are likely higher still – far exceeding the current $450s/mm^2$, $\approx 25:1$ output.

Given the nature of DW-MRI pulse/gradient sequences, the easiest option for increasing SNR with current hardware would be to reduce the echo time, T_E by bringing forwards the 180° RF pulse. Recalling chapter two, any reduction in T_E shortens the time between spin excitation and the peak in the NMR echo, thus reducing the time available for T_2 relaxation to effect undesirable non diffusion related signal attenuations – a particular issue given the short T_2 of muscle. The reduced attenuation leads to an increase

in signal magnitude and by extension an increase in SNR. Unfortunately, as a counter to this approach, any reduction in T_E necessitates proportional decreases in the duration of diffusion weighted gradient applications (δ) and mixing time Δ which, given that we cannot increase gradient strength (G) to compensate (the coils are already run at maximum), has the negative effect of reducing the effective diffusion weighting (remembering $b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3)$, where δ and Δ are reduced while G remains static). This simple and well known relationship establishes an inverse correlation between diffusion weighting and SNR meaning that, at least for the current hardware and acquisition strategy, meeting the image quality requirements is simply impossible without increasing voxel resolution which, for reasons of spatial accuracy, cannot be altered.

Moving away from SNR/b-values; other key issues affecting the quality of DW-MRI imaging in the maxillofacial region are: (1) the frequently mentioned susceptibility artefacts, a common example of which is visible in Figure 5.10 where air/tissue and tissue/air/metal interfaces have distorted the upper surface of the tongue/left side of the mandible and (2) motion artefacts which, while minimised by co-operation from an informed subject, cannot be entirely eliminated due to various involuntary reflexes (breathing, swallowing etc.) inherent to the imaged region.

Beginning with susceptibility artefacts; there are areas in which we are simply forced to admit defeat. While there are known methods to combat the air/tissue interfaces in the mouth and at external surfaces (e.g. tailored pulse sequences Bowtell et al. (1994), Andersson et al. (2003), Morgan et al. (2004), field unwarping Jezzard and Balaban (1995) and susceptibility matching/offset Komlosh et al. (2007), Lee et al. (2010)), there are ethical barriers to, for example, invasive procedures that might eliminate sinus related distortion and, even if that were possible, large metal dental fillings are simply too different (with respect to susceptibility) to counter easily. Likewise, with movement the majority of the related artefacts (post correction) appear to exist within the tongue and mandible. This makes sense; not only are these areas most prone to involuntary movement, they are also articulated and deform non-rigidly, making slice-to-slice registration particularly difficult without a suitable model for the deformations. Motion correction based on the assumption of a rigid body, while a not unreasonable assumption for the brain, will simply not suffice – new techniques must be developed or adapted.

While this subsection may appear quite negative overall, it simply represents a frank discussion of the issues inherent to the *current* maxillofacial imaging regime. In identifying the issues here, the groundwork has been laid out for avenues of improvement, an in depth discussion of which may be found in Chapter 7 (conclusions and future work). Without pre-empting those conclusions, we have acquired new hardware (the RF coil discussed in Section 5.4), software and know-how since acquiring the data used within this chapter. While we did not have the chance to bring these advances on-line within the time frame of this PhD project, they have the potential to drastically improve our

maxillofacial imaging prospects and we intend to take the opportunity to explore these ideas in future work.

5.6.2   Maxillofacial Reconstructions

Putting aside the previous subsection, the reconstructions presented within this chapter are very promising. Beginning with the temporomandibular joint cluster, admittedly a number of the muscles presented here could be reconstructed (at least in part) through DT-MRI. The masseter, temporalis and pytergoid certainly approach the size at which (ignoring internal structural overlaps), the core of the muscle lies free of external partial volume. This is, however, irrelevant. As shown by Gilbert and Napadow (2005) (Fig. 5.1), reconstruction of more complex muscle tissue at currently typical voxel resolutions is simply impossible using DT-MRI. Thus, while DT-MRI may indeed provide reasonable reconstructions of the larger muscles, the fact that CSHD is able to match this result is of far greater importance.

The work presented here comprises some of the first (Parker and Jones (2011)) to demonstrate plausible application of HARDI technologies to in vivo muscle data. Not only have we shown the ability to at least match DT-MRI reconstruction of large muscles, but we have also demonstrated limited yet plausible reconstruction of crossing muscle fibres in vivo – demonstrated within this muscle cluster by the retrieval of the risorius, delineation of individual masseter components and of the crossing/fanning portion of the buccinator.

Moving onto the tongue, this is where the promise of CSHD really comes into its own. For reasons that have already been discussed we are not yet able to completely match the quality of Gilbert et al. (2006) ex vivo DSI result (Fig. 5.2). However, the images presented in Figures 5.24 and 5.25 show clear and convincing agreement, a result that does not appear to be matched anywhere else in vivo. While probably the least aesthetically pleasing of the reconstructions, Figure 5.25 may well also be the most important. This figure demonstrates, without doubt, anatomically viable reconstructions of two interwoven fibre groups crossing, as simulations predicted would be viable, with near orthogonal trajectories. Accepting that these streamlines do not propagate particularly far – either terminating shortly after the seed or merging with other groups – the fact that they were visible at all tells us that the deconvolution is at least close to matching our simulation expectations. If this holds true, and we are able to garner the improvements predicted by an acquisition improvement, then there is a strong reason to believe that the work laid out within this chapter provides a solid foundation for future progress.

5.7   Conclusion

In closing, over the course of this and the previous chapter we have laid out the ground-work for both adaptation of the CSHD algorithm to muscle tissue and the requirements for retrieval of accurate, subject specific muscle fibre orientations. While technical limitations have prevented us from thus far meeting the initial project goal (i.e. full retrieval, reliability testing and model construction) we have, over the course of this chapter, been able to provide answers to key components of the puzzle. Returning to the original questions set in the introduction to this chapter (Section 5.1), we have shown through both simulation and practical demonstration that resolution of crossing muscle fibre, however limited, is indeed possible in vivo. Expanding upon this we have also confirmed the results of the previous chapter simulations, demonstrating without doubt that the deconvolution artefacts occurring within the simulations are not inherent to the simulated data itself, but represent a real and quantifiable phenomenon. Finally, thanks to the knowledge acquired over the course of this work, we can safely say that the segmentation result in Figure 5.27 represents current state-of-the-art in maxillofacial DW-MRI and that, given the changes to be outlined in Chapter 7, we can only see this improving further in the future.

Chapter 6

Automatic Streamline Segmentation

Prologue

This chapter discusses techniques for the automated segmentation of whole-volume DW-MRI streamline tractography results into bundles corresponding with individual anatomical features, a key and time consuming preparatory step in, for example, construction of maxillofacial models or analysis of tract-specific white matter metrics. Over the course of this chapter we lay out a novel technique, achieving streamline recognition (and thus segmentation) through a combination of shape (represented as a coordinate in a learned and pre-segmented feature space) and position discriminators. Attacking the problem in this manner allows us to avoid the computationally expensive inter-streamline comparisons commonly employed by competing methods, resulting in low and linearly scaling processing costs (with respect to dataset size), with no apparent detriment to solution quality. Owing to difficulties described in the previous chapter, there were an insufficient number of maxillofacial images (and by extension tractography results) to provide a sufficient corpus for quantitative analysis and thus we primarily focus on the topic of white matter segmentation as a more accessible analogue.

In the wider context of the proposed DW-MRI processing pipeline, the work presented in this chapter represents the final processing step. Over the course of this thesis we have laid out methods for both robust error detection and robust fODF retrieval – two core steps the production of reliable whole volume tractography results. Building upon a whole volume tractography result, the last non-specialised step remaining in any tractography based study is to segment the whole volume dataset into individual bundles of interest which may then be subjected to further analysis/processing (depending on the studies aims). As stipulated in the previous paragraph, the purpose of this chapter is to discuss the topic of streamline segmentation and present our own novel method for

achieving this task, thus completing our envisioned robust processing pipeline which, in its entirety, provides the tools required to complete the 'generic' portion of most single-shell DW-MRI tractography based studies.

A condensed version of this work was presented at the 2013 annual ISMRM meeting (Parker et al. (2013b)), receiving the *summa cum laude* award.

6.1 Introduction

As consistently stated throughout this thesis, DW-MRI has become a valuable tool for clinical and experimental research, being the only methodology able to characterise fibrous tissue microstructure in vivo. An important property that may be elucidated through DW-MRI, using methods discussed in Chapter 2, are three dimensional estimates of fibrous tissue orientation (e.g. Basser et al. (1994a), Tournier et al. (2004), Wedeen et al. (2005)) . Building upon these estimates, a class of techniques known as tractography (Basser et al. (2000), Mori et al. (1999), Mori and van Zijl (2002)) attempts to reconstruct underlying fibre geometry by tracing streamlines through consistently oriented voxels allowing, for example, the study of white matter connectivity.

An issue with current tractography methods is that it is often difficult to know exactly where to place seeds in order to reconstruct a specific fibre bundle, a common workaround is to seed across the vertices of a regular grid super-imposed over the image space, creating a whole volume tractography within which the bundle(s) of interest should be represented. The drawback, however, is that these whole volume datasets then require segmentation into those individual bundles in order to retrieve useful information.

One of the most common (and earliest) segmentation methods is manual segmentation, achieved by positioning Boolean waypoints at strategic locations (e.g. bundle end points plus key locations along the bundles path such as sharp turns, Conturo et al. (1999), Catani et al. (2002)) and then filtering streamlines based upon their intersections with such waypoints (Fig. 6.1a). Unfortunately, there are no guarantees that a single set of way-gates will produce similar results across multiple subjects since anatomical variation, errors in image co-registration and even simple image noise can alter streamline trajectory (Lori et al. (2002), Lin et al. (2001), Lazar and Alexander (2001)), leading to mismatches between streamline/way-gate positions (in essence selecting/filtering the wrong streamlines) or the appearance of spurious streamlines occurring from any point along the target bundle, requiring additional pruning that the standard way-gate set may not provide. As such, manual segmentation is typically performed on a per-subject basis, requiring expert anatomical knowledge, experience and a large time commitment in order to produce results that may still vary between operator/sessions due to the subjective judgements required (Wakana et al. (2007)).
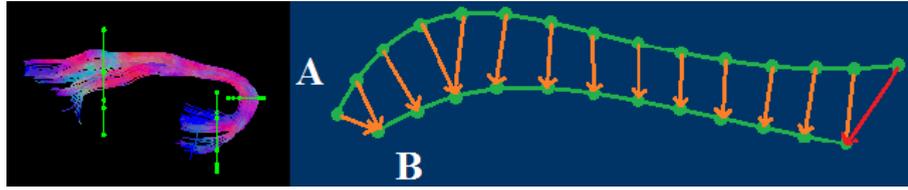
FIGURE 6.1: A: Positioning three Boolean AND way-gates (green) to select the intersecting subset of streamlines from a whole volume tractography result. In this case the streamlines depict the Uncinate Fasciculus. B: Hausdorff distance calculation. Comparing streamline A to B, Arrows represent minimum distances between all points on A and points on B, red arrow represents the maximum of the minimums.

The automated segmentation of tractography has become a popular research topic in recent years. Initial work began with the creation of standardised image/coordinate spaces (e.g. Talairach and Tournoux (1988)) which, while initially designed to aid image segmentation/co-registration, were later adapted (e.g. by warping streamlines to a standard atlas) to provide an automated steamline segmentation results (e.g Lebel et al. (2008), Lawes et al. (2008)). The drawback with these naive atlases was that their accuracy was strongly tied to the accuracy of the image co-registration and, particularly in the case of end-point atlases, did not cope well with spurious streamlines (e.g. given a streamline that starts and ends at the 'correct' positions but takes an artefactual route between these points, a naive end point atlas might incorrectly assign that streamline as belonging to the target bundle).

Representing an increase in complexity over basic atlas methods, clustering has also gained popularity. The initial goal was merely to organise whole volume datasets into bundles of similarly shaped/positioned streamlines; as such, several methods were proposed that were based on implementations of K-means (Ding et al. (2003)), agglomerative hierarchical (Zhang and Laidlaw (2004)), fuzzy c-means (Shimony et al. (2001)), spectral (Jonasson et al. (2005)) and expectational maximisation (M. Maddah (2006) clustering algorithms. While nominally successful, there are several downsides to such approaches. Firstly, in order to achieve optimal results it is necessary to know the optimal number of clusters a priori, given that this may not necessarily match the number of unique anatomical features, a precise number is often difficult to determine and thus may easily result in over/under segmentation – note that approaches based on affinity propagation (Leemans and Jones (2009b)) and data-stream (Smith and Nayak (2010)) clustering have been proposed to overcome this particular limitation by allowing the algorithm to determine an appropriate number of clusters on-the-fly. Secondly, while streamlines may be clustered into anatomically plausible bundles, few such naive methods attempt to apply corresponding anatomical labels, making extraction of particular white matter structures from within the clustered result difficult. Finally, in order to perform such clustering, it is generally necessary to perform exhaustive pair-wise streamline comparisons in order to create the affinity matrix, this results in a minimum $O(N^2)$

computational complexity (where N is the number of streamlines), thus limiting (compared to techniques that will soon be discussed) the size of the dataset that may be processed within reasonable time/memory thresholds, even before any actual clustering takes place (the exception is Smith and Nayak (2010) which incorporated other optimisations specifically for processing larger datasets).

More recent works attempt to overcome these drawbacks through hybridization of the atlas and clustering concepts. A good example of this approach is work by O'Donnell and Westin (2007) in which the author describes creation of a "high-dimensional atlas". This atlas consists of a reference image (for atlas co-registration), a set of labelled cluster centroids determined through clustering of training data, normalisation and embedding vectors derived from said clustering, and a random subset of training data streamlines. The core idea is that by comparing the affinity between novel streamlines and the randomly selected training data subset, it should be possible to embed novel data into the same space in which the original training data was clustered and thus estimate which of the predefined clusters each novel streamline would have belonged to if they were clustered with the original training data. The benefits here are threefold: (1) Linear scaling of processing time with respect to the number of streamlines. Rather than exhaustive pairwise comparisons [$O(N^2)$], each novel streamline is compared to a fixed subset of exemplar streamlines [$O(N)$]. (2) Elimination of on-line clustering. By projecting novel data onto the training space it is no longer necessary to compute a new clustering solution each time a novel dataset is processed. This has the added benefit of (3) consistent labelling. Given a stable projection, similar streamlines across multiple datasets will be consistently mapped to similar cluster centroids. By determining which structures these cluster centroids are likely to represent, they may be labelled according to anatomical content and then pass that label onto any assigned streamline, simplifying identification of individual bundles.

There are however downsides to this approach that are shared with many older and competing techniques. Throughout the literature, a substantial portion of clustering based works (including O'Donnell's) calculate their affinities through some variation on the Hausdorff distance (Rucklidge (1996)). In the simplest implementation, this may be described as the maximum of the minimum distances between each point on the candidate streamline and all points on the streamline to which it is being compared, essentially an exhaustive comparison of points (Fig. 6.1b). Given the recent trend towards lower step sizes, this again introduces a non-linear scaling in computational complexity, $O(M^2)$ (M representing the average number of points on a streamline) which, while trivial when comparing individual streamlines, becomes rather less so when considering the number of streamlines that must be compared. As an example, the datasets used for this work contain approximately 55 thousand streamlines with an average length of 74mm. At a 1mm step size, comparing one streamline to all of the others will require approximately 301 million euclidean distance comparisons, while at 0.5mm step size, this rises

to 1.2 billion ($M^2 \times 55000$), both of which are computationally costly when scaled to a complete dataset. O'Donnell attempts to reduce this burden in two ways; as previously discussed, each streamline is only compared to a limited set of exemplars (approximately 4000 in total, reducing distance calculations to 22 and 87 million per streamline for 1mm and 0.5mm step sizes respectively) and also sub-sampling of each streamline to 15 approximately equidistant points; fixing the number of distance measurements required for affinity calculation to 900 thousand per streamline. While this is undoubtedly preferable to the more naive approaches, we feel that it might be beneficial to adopt a new embedding strategy – eliminating the need for explicit inter-streamline comparisons – rather than working around the limitations of an existing one. To that end, the work presented within this chapter is intended to make the case for re-casting the streamline segmentation problem as a modified shape recognition/classification task, facilitated by trivial embedding of streamlines into pre-segmented, dimensionally reduced feature space.

Shape representation as a means of dimensional reduction is not a new concept within the streamline segmentation field. Work by Brun et al. (2004) discusses the feasibility of representing streamlines as a nine element vector comprised of mean streamline position and the six lower diagonal elements of a (3×3) point-wise covariance matrix, the benefits of which should be easily apparent when you consider that the O'Donnell method (for example) describes shape similarity (rather than shape itself) in a $\approx 4000D$ feature space. While the eventual use of these descriptors is again an exhaustive comparison to permit application of normalised cuts, the actual comparisons are far simpler than those based on the Hausdorff distance – reducing the problem to a single Euclidean distance measure between two 9D points – and, while $O(N^2)$, may actually be more efficient than complex linear-time methods within the bounds of a "reasonable" number of streamlines. Regardless of whether or not this is the case, Brun's work does indicate clear advantages for methods working in low dimensional shape spaces over those working in high dimensional shape similarity spaces, even if it does not go that far itself.

Covariance matrices are only one of a number of techniques used to build shape descriptors. Work by M. Maddah (2006), for example, approaches the task using the coefficients of spline curve. Perhaps one of the most interesting papers on this topic, however, is the publication by Batchelor et al. (2006). In short, this paper discusses the relative advantages and disadvantages of three methods for shape description and comparison when applied to streamline tractography results. (1) A Frenet framework (do Carmo (1976)) – describing streamlines through their curvature and torsion. (2) Fourier descriptors (Zhan and Roskies (1972)) – describing the frequency of change along the streamline x, y and z element vectors, and (3) Principal Component Analysis (PCA, Pearson (1901)) – describing streamlines by their variations from a mean shape. While Batchelor appears to favour Fourier descriptors for shape representation and later segmentation, owing to their rotational invariance (a property shared by the Frenet method) and general insensitivity to noise (shared by PCA but not Frenet); we would argue that this conclusion

is more of a result of the restrictive exemplar data used and that, in the real world, a modified PCA method may be more robust given a complete dataset. To explain, we would first argue that while rotational invariance is certainly beneficial given Batchelor's test data – the horizontal "U shaped" middle cerebellar peduncle and vertical, roughly linear, cortico spinal tracts – in reality, shapes repeat throughout the brain. Taking the U shape as an example, such a shape may also be found throughout corpus callosum. Taken in isolation, it may be difficult to tell whether a streamline belongs to the front, middle or rear of the corpus callosum; however, once orientational information becomes available, the decision becomes trivial – the end points will 'point' forwards at the front, up in the middle and rearwards at the back. On a whole brain scale therefore, perhaps orientation (after co-registration to a standard space) is a useful component of "shape" and thus, while using rotationally invariant methods may provide a purer description of the shape in isolation, practically speaking they might be disregarding useful information.

To progress with PCA based streamline segmentation, it is then useful to examine work by Cootes et al. (1995). In this publication, Cootes describes the creation of a PCA based Active Shape Model (ASM) capable of learning and then testing against permissible variations of a shape class to automatically determine whether a candidate belongs to the modelled set. In essence, training data is labelled by positioning a fixed number of sample points at consistent locations along the shapes outline (Fig. 6.2 – we will refer to the process of consistent labelling as producing a 'Normalisation'), the positions of these sample points are then concatenated to form a set of 1D feature descriptors to which PCA is applied to learn a model of accepted modes (described through PCA eigenvectors) and magnitudes (PCA eigenvalues) variation for the particular shape class. To conduct a comparison, one must simply normalise the novel data in the same manner as the training data, calculate the difference between the mean training data example and the each novel datum and then project (Eq. 6.8) those differences onto the axes described by a selected subset of the PCA eigenvalues. If the projected points lie within the bounds of variation described by the corresponding eigenvalues (conventionally $\pm 3\sqrt{\lambda_n}$, where $\sqrt{\lambda_n}$ corresponds to one standard deviation along the n'th axis) then the corresponding novel data is regarded as belonging to the modelled shape.

While we will show that active shape models may be applied (almost exactly as described) to segment individual streamline bundles, there are limitations to this method. The principal limitation of the traditional ASM is that each model is essentially describing its own shape space (a unique set of eigenvectors) in which only one shape class may be described with accuracy. In order to segment an entire brain, hundreds of ASM models may be required, each requiring its own projection and even, potentially, a unique normalisation. As such, while the purest form of ASM may be excellent for segmenting a very particular fibre bundle, they are not practical for segmenting an entire volume
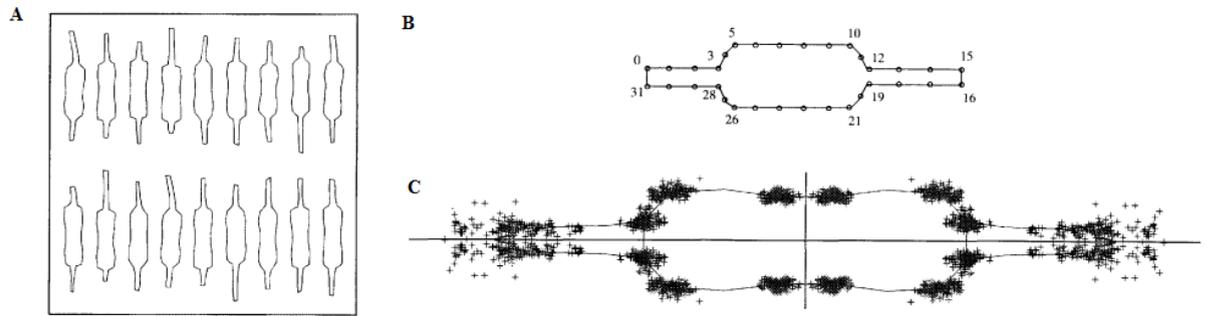
FIGURE 6.2: A scheme for consistently labelling electronic resistor outlines. Image constructed from extracts from Cootes et al. (1995): Figures 3, 4 and 5. A: Example resistors. B: Consistent labelling scheme, each point must be positioned in the corresponding location for each example. C: Plot of the label positions across the exemplar set.

because, in doing so, you may simply exchange expensive inter-streamline comparisons for equally expensive exhaustive model comparisons.

Our novel solution, therefore, is to think of the entire whole volume tractography as a single shape. If we then apply the ASM method to entire volume, instead of learning the modes of variation of a single shape, the eigenvectors should begin to describe the methods of variation required to transform any plausible (within the confines of the human brain) streamline into any other streamline, in essence determining a set of shape basis functions. Given that these basis functions are described by a set of independant mutually orthogonal eigenvectors they may also, as before, be considered to form the basis of a high dimensional brain feature space in which a streamline may be represented as a single point corresponding with the linear combination of basis vectors required to produce that particular shape.

Returning to O'Donnell's work, one of the biggest weaknesses of the described technique was the poor efficiency with which a single streamline could be embedded into a known space due to the required inter-streamline comparisons. If that could be overcome, i.e. through a trivial embedding into a PCA derived shape space, then it should be possible to employ similar methods to produce a whole volume segmentation tool that operates in low linear time. The challenges are therefore threefold. (1) To determine whether PCA based methods are capable of streamline segmentation: as already suggested, we will demonstrate this through a simple adaptation of the method described by Cootes. (2) To determine whether PCA derived whole volume shape spaces are stable; that is to say whether the calculated eigenvectors remain similar across multiple brains or groups of brains. (3) Finally, to determine whether the embedding into the feature space provides meaningful information, i.e. do similar shapes map to similar locations as we would hope or, more pressingly, do corresponding streamlines from different brains map to similar locations? Once the answer to these questions has been determined we can then outline methods for and the results of whole volume segmentations through projection into a

PCA derived feature space, providing both quantitative and qualitative analysis of the results.

6.2 Principal Component Analysis as a Tool for Dimensional Reduction

Since PCA is core to this chapter it is prudent to cover the relevant theory with particular focus on its use as a dimensional reduction technique.

A core assumption of PCA is that the modelled variable $v \in \mathbb{R}_n$ takes a multivariate normal distribution, and thus its PDF may be described as the multivariate Gaussian:

$$G(x, \mu, C_{n \times n}) = \frac{1}{2\pi^{\frac{n}{2}} |C_{n \times n}|^{\frac{1}{2}}} e^{(-\frac{1}{2}(x-\mu)^T C_{n \times n}^{-1}(x-\mu))} \tag{6.1}$$

where x is any point in the n dimensional space and the Gaussian itself is characterised by its μ and covarance matrix $C_{n \times n}$.

Extending upon this idea, Eigenmodels, Ξ, can be build such that:

$$\Xi = (\mu, U_{n \times n}, L_{n \times n}) \tag{6.2}$$

where μ represents the origin of the data in its original space, $\mu \in \mathbb{R}_n$; $U_{n \times n}$ represents a spanning basis of the Eigenspace in which columns define a set of orthonormal basis vectors (i.e. Eigenvectors) describing the axes of the Gaussian and $L_{n \times n}$ provides a diagonalised representation of the n eigenvalues specifying the extent of the Gaussian along their corresponding eigenvector axes.

Fortunately, fitting such a model, i.e. principal component analysis Pearson (1901), is fairly simple. Given a set of observations, $O_{n,k}$, they are first centralised by subtracting the mean at each datapoint:

$$\tilde{O}_{n,k} = O_{n,k} - \overline{O}_k \tag{6.3}$$

and then manipulated to produce a (centralised) co-variance matrix:

$$C_{n \times n} = \frac{1}{k} \tilde{O} \tilde{O}^T \tag{6.4}$$

which is then subjected to eigenvalue decomposition, solving:

$$\tilde{O} = ULU^T \tag{6.5}$$

providing the U and L components referenced in Eq. 6.2 which, when combined with the mean $(\overline{O})$, comprise the components of an eigenmodel representing the set of observations O.

Consideration of the Empirical Rule (Ross (1987)), which states that approximately 99.7% of normally distributed data lies within three standard deviations of the mean, leads to two implications for the PCA derived eigenmodel. Firstly, following projection of an observation, x into the eigenspace:

$$P = U^T(x - \mu) \tag{6.6}$$

coordinates with a component outside three times the square root of the corresponding eigenvalue (which provides an estimate of variance along a particular axis of the Gaussian representation) can be deemed with some confidence to not belong to the same set of observations upon which the model was constructed – providing a convenient recognition constraint for PCA based methods such as Cootes et al. (1995) active shape models.

Furthermore if extended to variance as a whole, the full set of n eigenvectors (assumed to be sorted in descending order of corresponding Eigenvalues) can be truncated to some smaller set m where $m \leq n$ by discarding higher modes of variation whose eigenvalues indicate reduced contribution to the overall description. Keeping in mind the Empirical Rule, one potential method of truncation is to assume the sum of eigenvalues (describing the extent of variation along a particular axis) provides a reasonable estimate of the total variance within a dataset and thus truncate at the point at which the cumulative sum reaches $\approx 99.7\%$ of that total.

Regardless of the exact method used, the purpose of this truncation is to retain the set of eigenvectors which contribute significantly to the overall description of the observed data while simultaneously eliminating those which, to within some tolerance, are superfluous – thus reducing the dimensionality of the projected space while not significantly affecting the accuracy of its representations.

For completeness, given a reduced basis, U_m, projection remains similar to Equation 6.6, simply incorporating a reduced basis set:

$$P = U_m^T(x - \mu) \tag{6.7}$$

which, unless otherwise stated, will be what is meant by projection or embedding for the remainder of this work. Additionally, it is also possible to approximate the original data (O) from its projected coordinate through the equation:

$$O = U_m P + \mu \tag{6.8}$$

which will later be used to estimate the error at various truncation points.

The key point to take from this section is that where the variance in a dataset can be decomposed onto relatively few key axes compared to the dimensionality of the observations, significant dimensional reductions are possible through the PCA method while minimising the loss of precision inherent to any truncation. Fortunately, as will shortly be demonstrated, this is exactly the case for the representation of streamlines comprising fibre bundles and may be successfully exploited to achieve segmentation tasks.

6.3 Image Acquisition, Processing and Tractography

As a final aside before continuing to the novel portions of this chapter it is necessary to mention image acquisition and processing parameters. Diffusion weighted images were acquired for 30 healthy adult volunteers using a GE 3T Signa HDx system with SS-EPI sequence, b-value of $1200s/mm^2$ along 60 optimal directions (Jones et al. (1999)) with $1.8mm$ isotropic resolution producing a $128 \times 128 \times 60$ voxel image. Whole volume deterministic tractography) was then conducted using the damped Richardson-Lucy (dRL) fODF retrieval algorithm (Dell'Acqua et al. (2010)) calibrated to an alpha value of approximately $1.5mm/s^2$ with seed points spaced across a $2mm$ isotropic grid. Step size, angular threshold, fODF threshold and streamline length constraints were set to $0.5mm$, $45°$, 0.1 and $[50\ 250]mm$ respectively.

6.4 Active Shape Models: A Proof of Concept

To provide proof of concept (addressing point one of the three criteria that we have outlined) we will begin by demonstrating the applicability of a Cootes-type ASM to the streamline segmentation task. Principally, our proposed method is composed of two components – model construction and model application – with the option of an additional iterative region growing that we will outline shortly. The basic premise is that a model of streamline shape may be learned through PCA of carefully filtered manually segmented training data and, once learned, may be applied to unseen data - selecting matching streamlines (i.e. those conforming to the model) from a supplied candidate pool. While these selected streamlines will often provide a suitably accurate portrayal

of the overall shape, limitations of the modelling technique limits recognition to "full length" streamlines, those that terminate prematurely will not demonstrate the correct shape and are thus not accepted. Should the inclusion of sub-length streamlines be required, the region growing algorithm may be employed to ensure their inclusion.

6.4.1   ASM: Model Construction

A direct implementation of the ASM technique requires consistently labelled training data. The process for modelling a particular white-matter structure begins by deciding whether it may be described as a single shape or if it may be better described as a collection of shapes, each requiring an individual model, that may be recombined into the desired structure. For simple structures, such as the Uncinate Fasciculus (Fig. 6.3a), all complete streamlines follow a consistent path and thus, as we will show, it is plausible to capture its shape variations within a single model. Moving onto more complex bundle, we will also cover the Fornix (Fig. 6.3b) which for the simple implementation used must be broken down into two models covering the left and right hemisphere separately (Fig. 6.3c) - note that while this would not be necessary if a hemispherical mirroring scheme were used (e.g. O'Donnell and Westin (2007)) we choose not to in this case in order to provide a 'toy' example for multi-bundle segmentations.
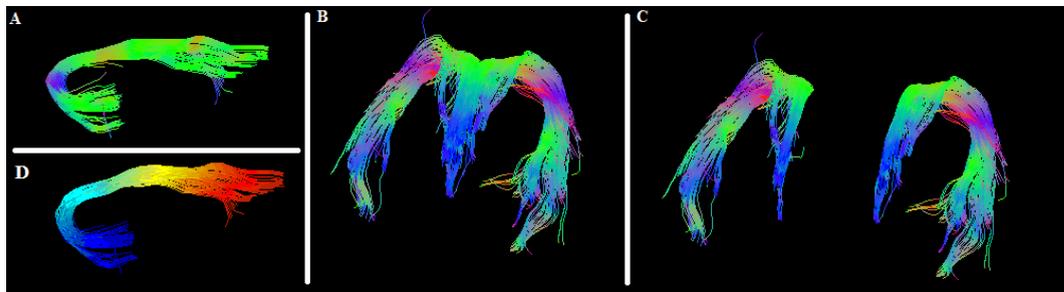
FIGURE 6.3: A: Example streamlines representing the Uncinate Fasciculus. B: Example streamlines representing the Fornix. C: Parcellation of the Fornix into left and right hemisphere bundles. D: Labelled Uncinate Fasciculus, colours encode label numbers from 1 to 30.

The process for modelling individual bundles begins with manual segmentation of the corresponding exemplar streamlines. From here, individual streamlines are first 're-ordered' (explained shortly) according to an external reference point, filtered by length (mean length $\pm$ two standard deviations) to exclude abnormally long (e.g. if the tractography were to double back as is sometimes possible) or prematurely terminated streamlines and the interpolated through cubic b-splines to a fixed 30 point representation in order to provide a consistent labelling. Re-ordering is necessary to reduce unnecessary variation caused as a by-product of streamline representation. While two streamlines may seem identical when visualised, their 3D point representation (remembering from chapter 2 that streamlines are most often represented as a series of 3D points calculated

at each tractography integration step) may be significantly different since this representation includes an inherent 'direction' of streamline propagation. As an explanatory example, imagine a straight horizontal line represented by a series of points. These points may travel from right to left or left to right and thus while visually describing the same shape, representationally describe the inverse of the other. A simple solution to this problem is to choose some external reference point and re-order the streamline representations such that the first (or last) point is closest to that external reference, e.g. for the Corpus Callosum it might make sense to order streamlines so that the first point is in the left hemisphere.

In order to further reduce non shape related variances caused by position, scale and rotation, a final iterative transformation is applied to all training streamlines. Firstly, each streamline is translated to the origin by centre of mass subtraction, such that for a given streamline $X_{n,m}$ consisting of n points in $m = 3$ dimensions, its translated position X' becomes:

$$X'_{n,m} = X_{n,m} - \overline{X}_m \tag{6.9}$$

performed uniquely for each streamline. Following this translation a single streamline from the ensemble is selected at random to provide a template for the following iterative process:

1. Warp all streamlines onto the template using the Procrustes transformation.

2. Choose the mean streamline label position as the next template.

3. IF the new template matches the previous template: Convergence achieved - template found.

4. ELSE repeat from step one.

which is essentially an implementation of generalised Procrustes analysis (Gower (1975)) involving a combination of translation, rotation and uniform scaling optimising which, while altering size, position and orientation to best match a template, will not distort the shape of a geometric object. While achieving full convergence may take a sizeable amount of time, practically speaking 4-5 iterations of this process appears to produce a workable result.

As a final step the 3D coordinates are concatenated, forming 90 element descriptors, which are then combined into an observation matrix for PCA. In order to reduce the dimensionality of the feature space, a cumulative sum of the resultant eigenvalues is calculated and then the corresponding eigenvectors truncated at the point at which the cumulative sum exceeds 98% of the total. Depending on the target structure this has

typically resulted in a 5-10 dimension feature space – a significant reduction compared to the original 90. Note that a more detailed exploration of the effects of truncation will follow in Section 6.5.2.

While this process may appear to lack any explicitly enforced point-wise correspondence, the manner in which training data are selected (ending at similar locations with similar lengths and trajectories) implicitly leads to an approximate correspondence (Fig. 6.3d) which, after subsequent processing steps results in only minor errors that should not adversely affect our observed results. Should it become necessary to explicitly enforce correspondence, Colby et al. (2012) discusses one possible solution.

6.4.2   ASM: Shape Model Application

To select the initial set of candidate streamlines a simple masking technique is used. Before manual bundle segmentation, each training image is first affinely (Jenkinson and Smith (2001), Jenkinson et al. (2002)) co-registered to a standard image space (MNI FA). Once bundles are segmented, those same affine transformations are applied to each streamline, bringing the native-space tractography into a standard space, after which all intersected voxels are recorded. Once all streamlines have been processed in such a manner, a mask is generated by retaining all intersected voxels with a visitation frequency greater than 10% of the maximum frequency to eliminate contributions from spurious streamlines. Following this the mask is then re-enlarged by 5mm in all directions, creating a deliberately coarse representation of the core bundle position within the standardised space, intended to reduce any false negatives that may result from poor future registrations. Novel whole volume datasets are then co-registered to this same standard space, following which the inverse transform is applied, moving the mask from MNI to native, allowing selection of candidate streamlines by intersection with the masked region.

To facilitate streamline shape comparison, candidate streamlines are first re-ordered (as explained in the previous section, with particular reference point selected on a per-bundle basis), interpolated to 30 data points and then transformed (centre of mass followed by Procrustes) onto the iteratively acquired mean template for the bundle in question. Particular attention must be paid to rotations during this phase since, as discussed in the introductory section, orientation with respect to the cardinal axes of the brain can be considered an important aspect of 'shape'.

In practice this is fairly simple; we first calculate the Euler angles of the Procrustes matrix (Slabaugh (1999)), giving an account of the rotations required to match the streamline with the template from which we subtract the bulk rotations caused by head positioning (Euler angles of the affine transformation matrix) to provide an estimate of the overall rotation within a brain centred reference frame. It is then fairly trivial to

place thresholds on the extent of allowable rotations since, quite obviously, a perfect match is useless if it requires a 180° rotation to achieve. All streamlines passing this test are then converted to a 90 element feature descriptor and projected (Eq. 6.8) into the PCA model space. Those streamlines whose coordinates fall within three standard deviations (square root of corresponding eigenvalue) on all axes are accepted as belonging to the modelled shape class (and by extension the bundle of interest).

6.4.3   ASM: Iterative Region Growing

The described shape models are deliberately designed to retrieve only full-length, full-shape streamlines. The reasoning for this is two-fold. Firstly, in numerous cases the full length streamlines are the streamlines of interest - e.g. there are graph theory methods that weight edges based on the number of streamlines connecting the nodes, partially complete streamlines would not make the connection and thus are unimportant. Secondly; depending on the position of the partial streamline, the point(s) at which it breaks prematurely and the source of that breakage (e.g. an overly restrictive angular threshold may provide consistent breaks, but random noise may not), the potential for unique shapes is almost limitless. It is simply infeasible to create models for all but the most commonly reproducible sub-length streamline shapes. Should these incomplete streamlines contain potentially useful information a secondary technique must be devised in order to retrieve it.

Taking streamlines recovered through an ASM as a reference, it should be reasonable to infer that unrecovered streamlines, whether that be due to premature termination or variations unseen in the training data, should remain consistent with the recovered portion. This is to say that for the majority of their length they should remain in close proximity to the recovered portion and not deviate significantly from its overall trajectory. The following algorithm therefore facilitates an approximate enforcement of these constraints (within a discrete framework) as part of an iterative region growing.

1. Resample the original image to a coarse (2-3mm isotropic) resolution.

2. FOR each voxel intersected by the segmented set :
   2.1 Calculte mean streamline orientations (E.q. 6.10)

3. FOR each streamline not in the included set:
   3.1. IF streamline segments intersects segmented set for $\geq$ threshold percentage of its duration:
   3.1.1 Estimate orientation at each point.
   3.1.2. Calculate mean deviation from voxel-wise estimates.
   3.1.3 IF below threshold, include in next iteration of segmented set.

4. IF current segmented set matches previous segmented set:
   4.1. Convergence: result found.

5. ELSEIF maximum iterations reached.
   5.1. Accept current segmented set.

6. ELSE:
   6.1. GOTO 2.

In this formulation proximity is implicitly enforced by setting a minimum percentage of streamline duration that is within the set of voxels intersected by the current segmented set, being the only voxels about which reasonable assumptions of bundle orientation may be made (or even if the bundle intersects the voxel at all). Similarity of trajectory is then enforced by comparing point-wise estimates of streamline trajectory (approximated by the vector connecting the current point to the next) to the mean orientation of the segmented set within the current voxel, approximated (according to Jones (2003)) as the first eigenvector of the mean dyadic tensor:

$$d_{mean} = eig_1\{\frac{1}{n}\sum_{i=1}^{n} d_i d_i^T\} \tag{6.10}$$

where d_n represents the orientation of the n'th intersecting member of the segmented set as it passes through the target voxel. The advantage of performing the calculation this being that unlike other methods, the dyadic tensor naturally resolves the issue of antipodal symmetries.

Since we can only definitively know orientations within a short radius of the ASM recovered framework, a low iteration count is generally preferable to prevent unwarranted run-away expansion (note that the known area and approximations of orientation are updated after each iteration, allowing the algorithm to 'learn' subject specific variation); that said, if more is known about the underlying geometry – particularly if there is smooth variation in length and orientation across the structure with minimal outside

intersection, as is the case for some muscles (fig. 6.4b) – it may be possible to achieve full segmentation by selection of a single full length seed streamline and iterating until convergence.
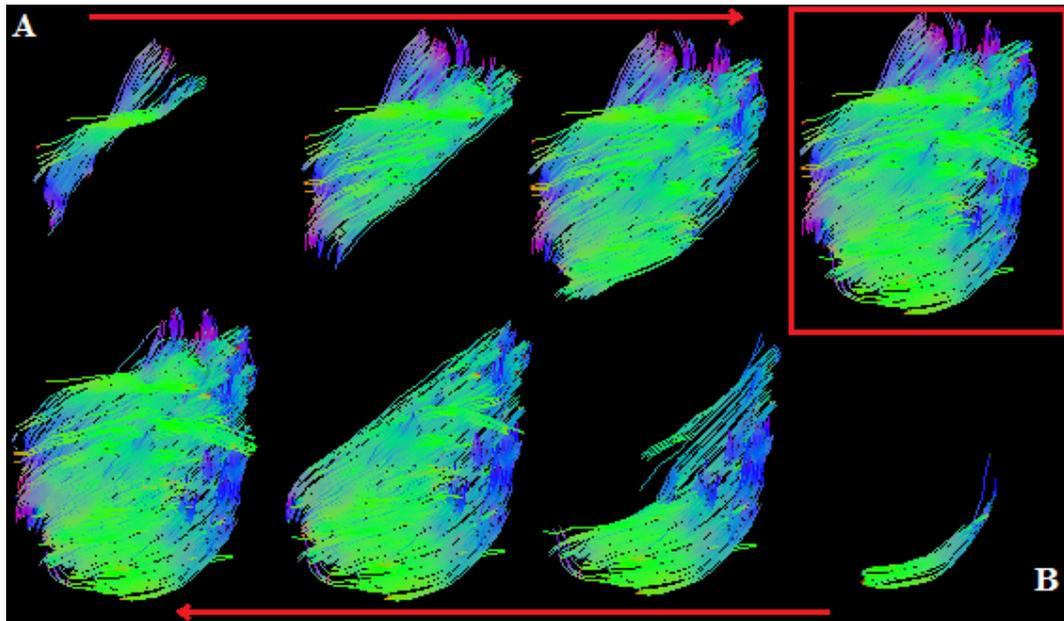
FIGURE 6.4: A: Pictorial representation of the region growing algorithm. B: An Example of region-growing as a segmentation tool in its own right – recovering a masseter muscle from two seed points converging to a result with $\approx 98\%$ streamline commonality.

6.4.4 ASM: Results

Using the methods described above, 15 datasets were selected at random and used (14 at a time) to construct 15 active shape models/model sets in a leave-one-out experiment. For the purpose of this experiment the uncinate fasciculus (right hemisphere only) and fornix (left and right hemisphered modelled separately) were selected for modelling/segmentation with reordering such that the uncinate streamlines travel front to back and the fornix from left to right. Region growing parameters were 3mm isotropic re-sampling, $20°$ mean angular deviation threshold, 70% length threshold and a maximum of 5 permissible iterations.
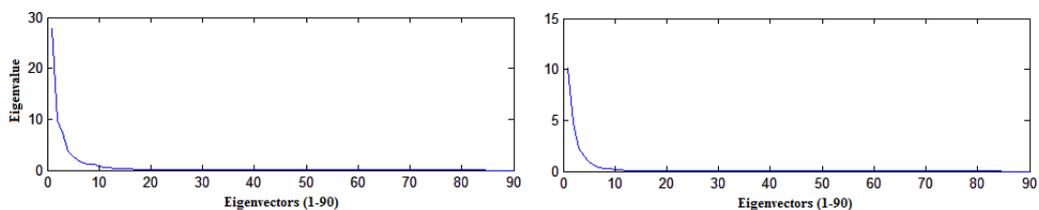
FIGURE 6.5: A: Eigenvalue profile of the Uncinate Fasciculus. B: Eigenvalue profile of the right hemisphere Fornix.
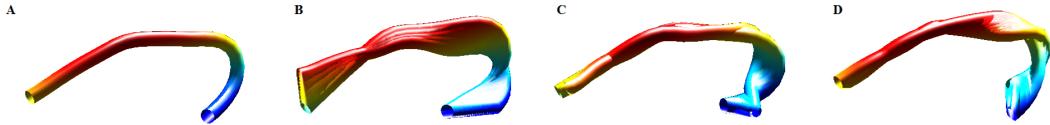
FIGURE 6.6: Modes of variation of the Uncinate. A: Mean shape. B: $\pm 2SD$ along first mode of variation. C: Second mode. D: Third mode.

Beginning with the models themselves, Figure 6.5 demonstrates typical eigenvalue profiles for (a) the uncinate and (b) right hemisphere fornix component. Note that only the first 10-12 modes of variation describe any significant variance in the overall shape and so, while this matter will be examined in greater detail at a later point, for this purpose of this preliminary proof of concept dimensional reduction was performed using the cumulative 98% method described in section 2 - leading to (on average, given that each of the 15 models is subtly different) a 14 dimensional uncinate and 10 dimensional fornix representation. For visual reference, Figure 6.6 then demonstrates the first few modes of variation described by the uncinate model; with Figure 6.6a demonstrating the mean shape and Figure 6.6b-d describing ± 2 standard deviations of variation along the 1st, 2nd, and 3rd modes of variation respectively. While difficult to visualise in a static image, the first mode of variation captures the "sharpness" of the bend, the second mode captures curvature of the rear extension as it either splays outwards away from or turns inwards and under the main body and the third mode captures a general torsion of the shape.

Figures 6.7 and 6.8 demonstrate a selection of automated reconstruction results. For reference, column 1 displays the results of manual segmentation, while the second column displays the result of a purely ASM driven segmentation using the methods described in Section 6.4.2. Finally, the third column displays the results achieved by augmenting the pure ASM with an additional region growing step as described in Section 6.4.3. Qualitatively speaking, the automated results demonstrate excellence resemblance to the equivalent manual segmentations with column 3 (ASM plus region-growing) in particular appearing to produce a "fuller" segmentation, likely owing to the recovery of short streamlines filtered out during manual segmentation (too short to pass through the specified way-gates) and deliberately excluded from the active shape model (the ASM being intended to model full-length streamlines only). Recall and precision statistics (Figure 6.9) calculated as:

$$recall = \frac{t_p}{t_p + f_n} \tag{6.11}$$

and

$$precision = \frac{t_p}{t_p + f_p} \tag{6.12}$$
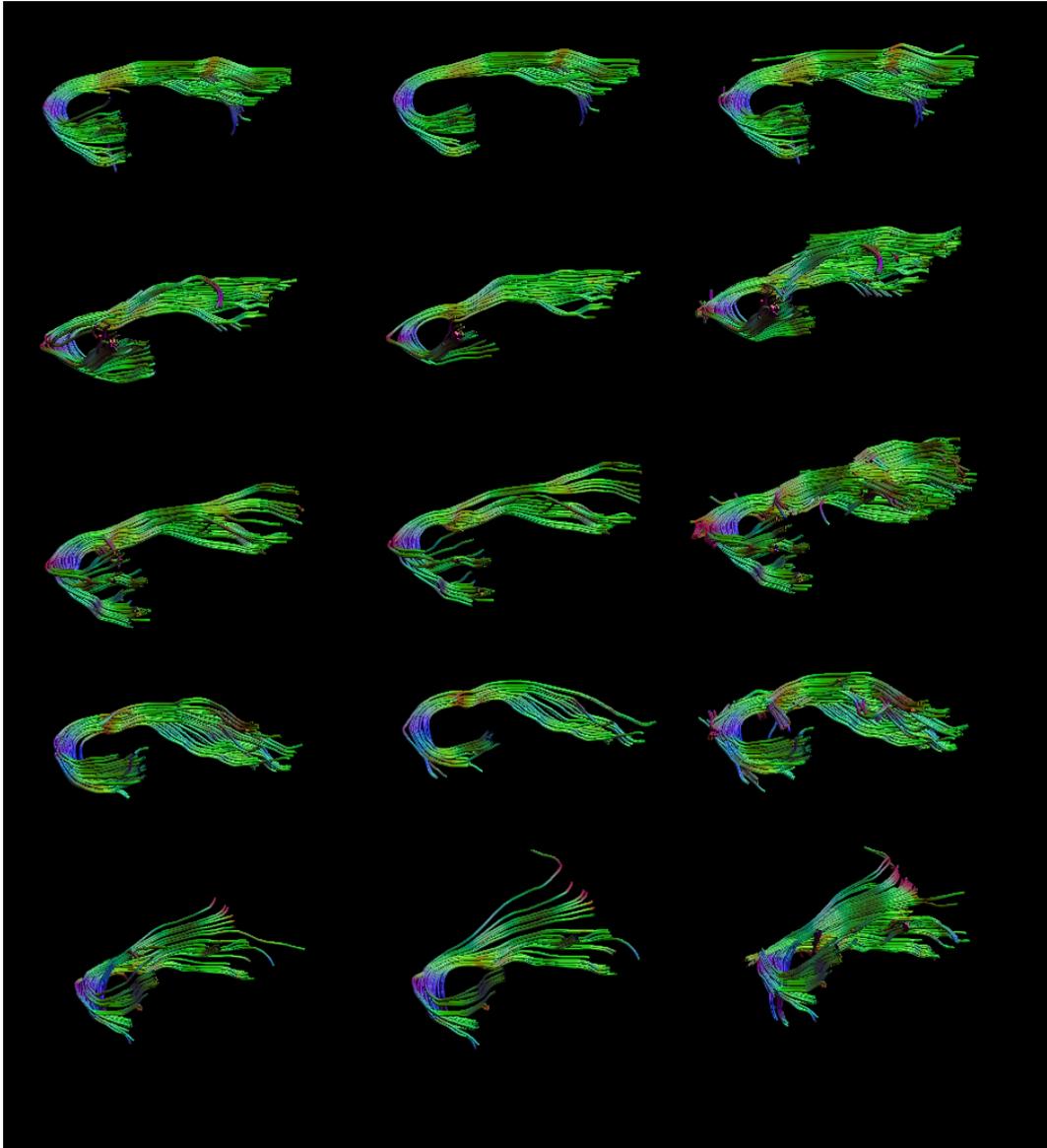
FIGURE 6.7: Example segmentations of the Uncinate Fasciculus. Column 1: manual segmentation. Column 2: ASM result before growth. Column 3: ASM following region growing.

where t_p, f_p and f_n are the number of true positives, false positives and false negatives respectively; reinforce this qualitative agreement. Using the combined ASM and region-growing results as a point of comparison, recall (column one) confirms that the fraction of the manually segmented streamlines recovered by the automated method remains universally high for both the uncinate and fornix reconstructions, indicating excellent retrieval characteristics. Precision (column two, labelled "Prevision (overall)") however, initially appears comparatively low, indicating the presence of a large number of false positive retrievals in the automated result. The reason for this is fairly simple; as previously explained, the region-growing process is intended to recover short (though valid) streamlines that were deliberately excluded from the manual segmentations. As a result, these short streamlines will always present as false positives, lowering the perceived
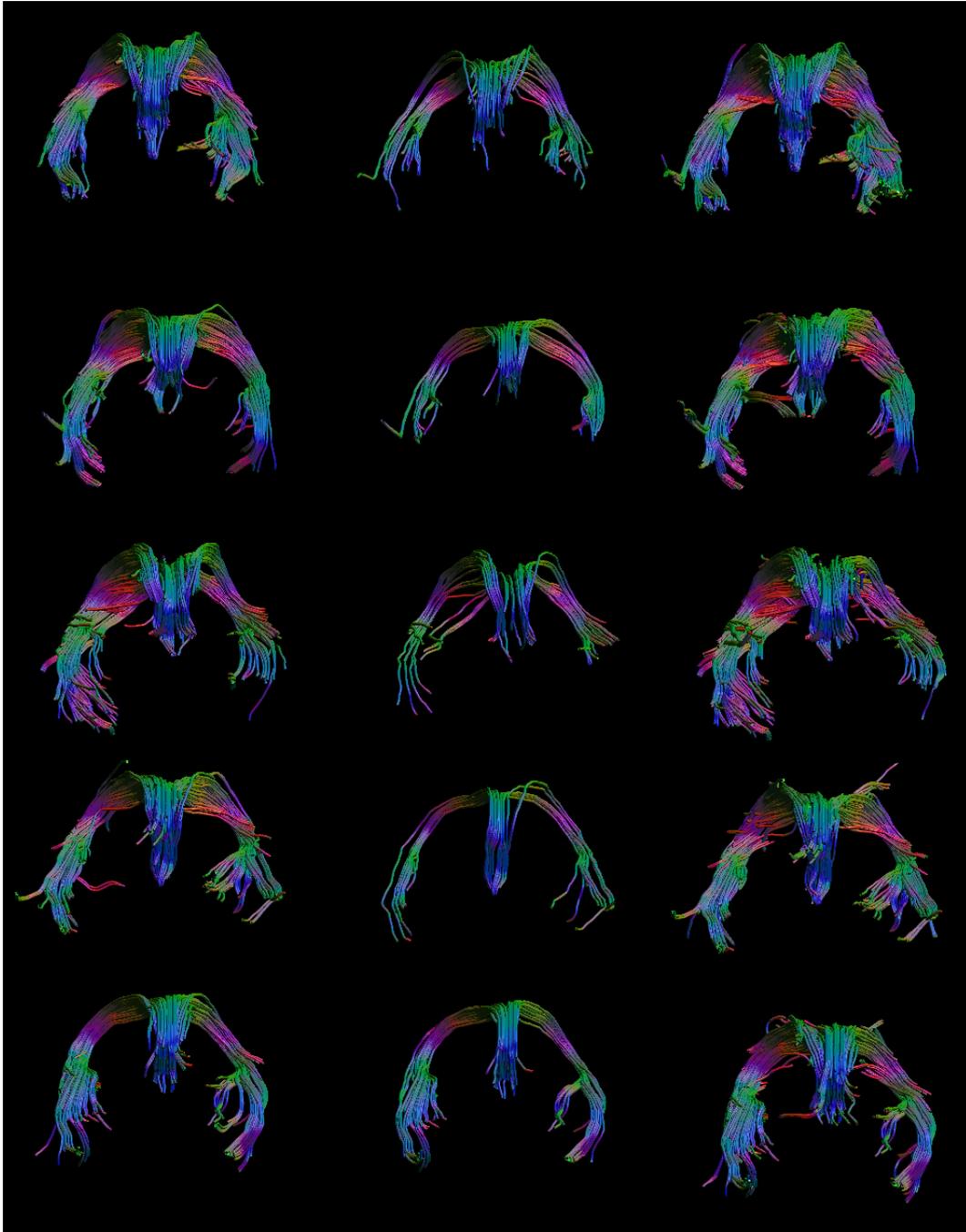
FIGURE 6.8: Example segmentations of the Fornix. Columns 1-3 consistent with Figure 6.7.

precision. If, however, we compensate for this mismatch by momentarily excluding the shorter streamlines from the automated segmentation result (essentially falling back to the pure ASM), precision statistics (column 3, labelled Precision (Core)) increase to the expected levels, indicating that the automated method suitably replicates the manual streamline selection process.
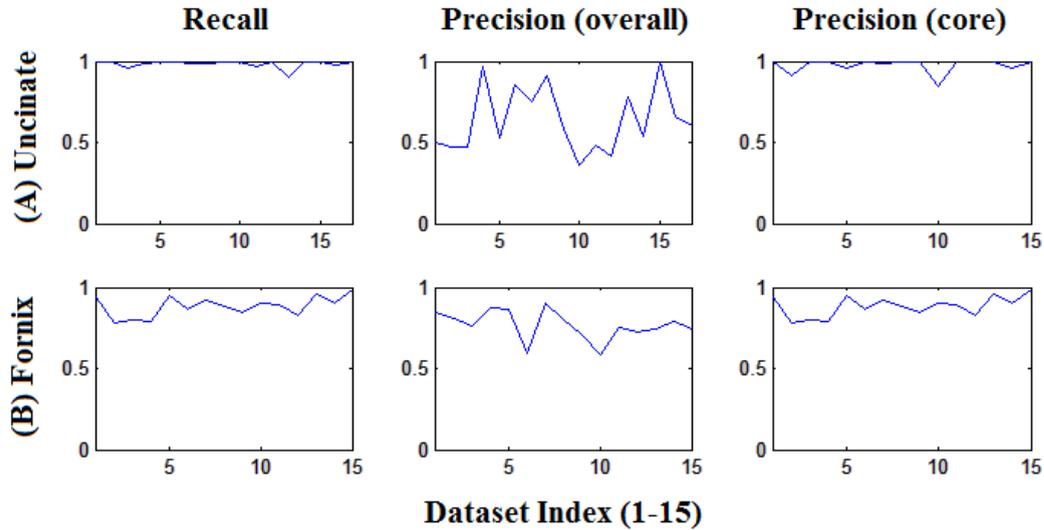
FIGURE 6.9: Comparison between automated and manual reconstructions.

While by no means an exhaustive study, the high precision/recall statistics provide concrete evidence that a shape based segmentation *may* work within the human white matter tractography. The agreement shown by the length thresholded precision results importantly indicates that the ASM type model is capable of retrieving the same type of streamlines on which the model was trained with minimal overlapping with neighbouring structures. In answer to point one laid out in the introduction, these experiments clearly show that ASM is a segmentation and, as a more practical point, that the adaptations implemented for streamline normalisation appear to be suitable for the job at hand, thus, when moving to a whole volume segmentation they might be easily adapted.

6.5 Whole Volume Modelling

As discussed within the introduction, the principal drawback of traditional active shape models is that they are in essence bundle-specific, requiring the creation of individual models for every conceivable shape – a time consuming proposition with respect to both training an segmentation. The solution therefore is to construct a single space in which the numerous streamline shapes present within the brain may be represented.

While a more detailed discussion and explanation will appear in the following subsections (6.5.1-6.5.7), in short, this is achieved through relatively minor modifications to the previously described (Section 6.4) method. Rather than attempting to describe each bundle separately, we discovered that PCA could be applied to whole volume datasets (following co-registration and streamline normalisation) with the results essentially defining a set of independent 'shape basis functions' which, in linear combination, could reasonably approximate any plausible streamline shape – the logical extension being that any streamline can be described by a point in a space whose axes are defined by these basis

vectors and that (as will be discussed in Section 6.5.3) streamlines with similar shape (i.e. belonging to to particular bundles) will tend to reproducibly cluster in certain regions of that space. With this in mind, the following subsections will lay out detailed mechanisms for construction of such a space (Section 6.5.1) and, importantly, determining which sub-regions map to individual bundles of interest, using both K-means (Section 6.5.5) and support vector machine (Section 6.5.6) methods, for the purpose of segmenting novel whole volume datasets (Section 6.5.2).

6.5.1 Whole Volume Model Creation

The process for building a whole volume feature space is similar to that of single bundle modelling, though there are several key differences. While minimising non-shape variance is a fairly trivial task when dealing with a single bundle, e.g. the Procrustes transformation may be used to eliminate rotation, position and scale once a suitable template has been selected, there are limitations to the manipulation of whole volume data that necessitate the inclusion of other variances within the 'shape' model. The first task is simply to decide which of these variances are acceptable (or potentially even beneficial) and which must be eliminated regardless of expense.

Beginning with the Procrustes transformation; while looking at a particular bundle, all streamlines present a similar shape and thus the generalised Procrustes analysis process works reasonably well, generating similar templates across repeat attempts. The issue with the whole volume is that it doesn't make sense to select a streamline at random when given the vast range of potential shapes. While it may seem that this can be easily replaced by setting the initial template to the whole volume average shape, the problem is that this average (due to later streamline mirroring) tends towards a singular point (Fig. 6.10) which can lead to instability in the resultant transformations - i.e. streamlines with similar shapes may experience strongly dissimilar rotations due to minor differences in their label positions. The solution is actually fairly simple; as discussed within the introduction, sensitivity to streamline orientation within brain-centric reference frame may actually be beneficial in discerning between similarly shaped though distinct fibre bundles. In order to approximate streamline position within this reference frame we simply apply the transformation matrix produced during affine registration of the native space image to a standard (MNI FA) to each streamline - in effect bringing them into the standard space. This has two effects: (1) The rotations applied during the transformation eliminate rotations due to subject head position, moving distinct images into an (approximately) consistent space. (2) Rescaling of the brain (and by extension the resultant tractography) to a standard volume reduces size discrepancies across corresponding structures, thus decreasing (though not eliminating) subject intra-bundle size variations while preserving useful inter-bundle contrast - i.e. a basic scale invariance.
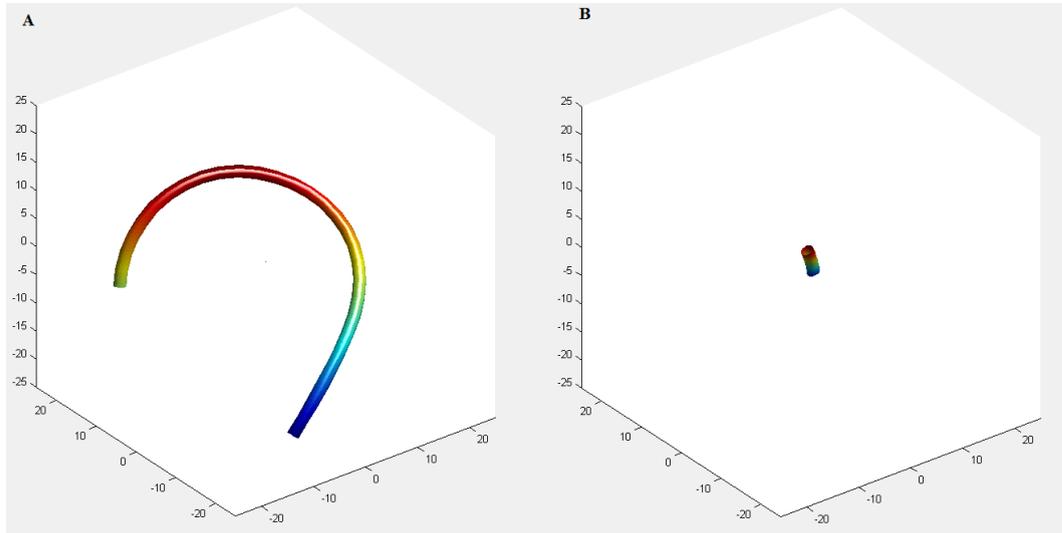
FIGURE 6.10: Comparison of mean shapes. A: The mean shape for the left hemisphere Fornix component. B: Mean shape for the whole volume.

Unlike single bundles, when examining a whole volume tractography there is also no clear reference point/direction according to which streamlines may be re-ordered to produce directionally consistent labellings. The solution is to simply discard the streamline re-ordering aspect of the previous ASM approach, instead we attempt to create a 'symmetrical' feature space in which both potential representations of plausible streamline shapes may be equally represented. To achieve this – and to remove potential bias in the training data – all datasets are first subjected to streamline duplication by appending a second set of reversed streamlines to the original tractography result.

With these variations in mind, creation of the feature space then proceeds much as before. Individual streamlines are first interpolated to a 30 point representation using cubic b-splines. From there, streamlines are translated to the origin using a centre of mass subtraction (removing variance due to position) and then the x, y and z label components concatenated to form a 90 element shape descriptor. Once all streamlines have been processed, a principal component analysis is conducted and the resultant eigenvectors, eigenvalues and mean shape recorded.

6.5.2 Feature Space Stability

Before proceeding to streamline segmentation methods, it is first important at this juncture to determine the stability of the constructed feature space. The most logical place to begin is to determine the "re-usability" of a whole volume space. While it may be possible to create a feature space that adequately describes all streamlines belonging to a single volume, there is no guarantee that the shape basis functions derived from one volume, or set of volumes, would be sufficient to describe (and therefore encode) streamline shapes present within unseen data.

Since the eigenvectors derived from a single dataset are likely to be in some way dependant on that individuals brain morphology, it makes sense when building a feature space to include example streamlines covering a range of brain morphologies such that the result would, one hopes, represent a more "universal" feature space capable of representing a wider array of streamlines. The first step in creating this universal space is therefore to determine the number of datasets required, on average, to ensure that the resultant eigenvector orientations (i.e. the feature space axes) can be considered consistent regardless of the selected datasets.

To address this question a bootstrap experiment was conducted in which sets of one to twenty datasets were selected (from a 30 dataset corpus) with 50 repetitions per increment (i.e. 50 permutations of a two dataset combination followed by 50 permutations of a three dataset combination and so on) and without replacement. Within each increment the 50 permutations were then individually combined (creating 50 large datasets) and the resultant set of streamlines subjected to principal component analysis. Finally, the 50 eigenvector sets were then exhaustively evaluated for relative stability using the Frobenius matrix norm method:

$$F_n = \frac{1}{N_e}||U^T V||^2 \tag{6.13}$$

where N_e represents the number of dimensions to which the compared eigenvector matrices U and V are truncated. The resultant stability measurements (2500 per increment in total) were then averaged to produce a single (per increment) characterising statistic. Note that this method is similar to that employed by O'Donnell and Westin (2007), though the formulation used in that publication incorrectly omits the squared term stipulated by the referenced material (Fowlkes et al. (2004)). For reference, we will examine stability at eigenvector truncations of 5, 10, 15 and 20 dimensions.

Figure 6.11 contains the result of this experiment. As one might expect, the number of datasets required to produce a stable result strongly depends on the dimensionality of the described feature space. While at no point are the matrices particularly unstable (measuring 0.951 at the lowest), low dimensional truncations (5 and 10 dimensions, red and green respectively) show rapid improvement after just two datasets are included and then increase in stability until plateauing at or around 8 datasets. As dimensionality increases (e.g. 15 and 20 dimensions, blue and black respectively) initial stability is far poorer, though again as the number of datasets used increases they once again improve to a plateau after 8 or, with 20 dimensions, ≈ 18 datasets are included within the shape space creation.

An obvious conclusion from this result is that there are major modes of shape variation that remain relatively constant across human brains. Evidence for this can be seen in the rapid convergence of the first five, i.e. highest variance, eigenvectors towards
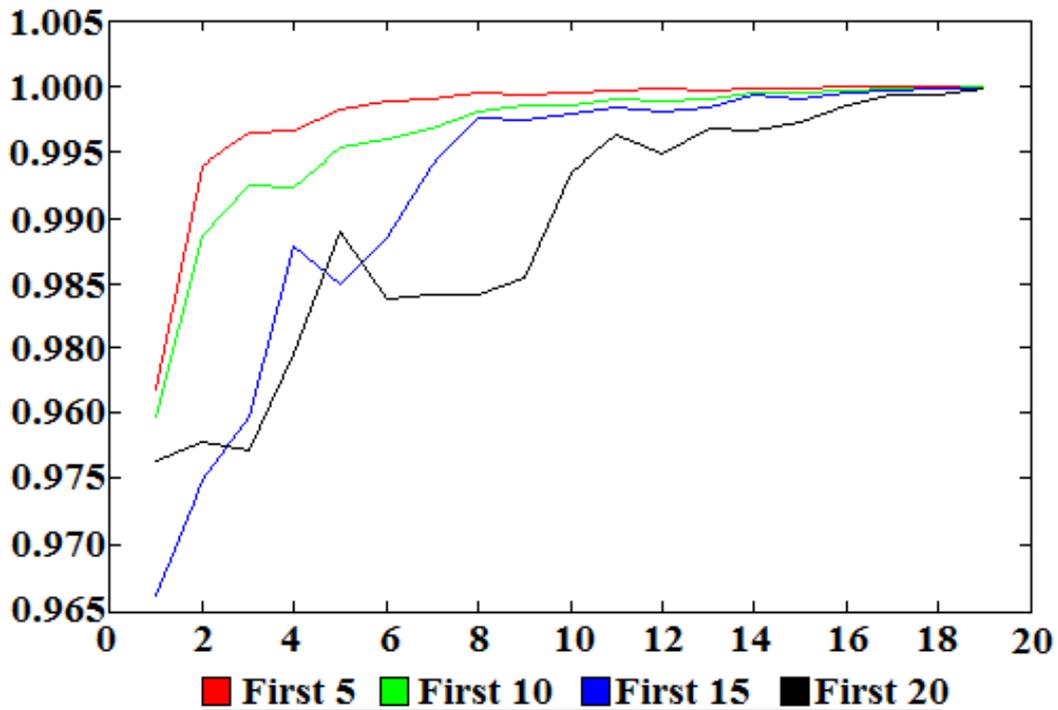
FIGURE 6.11: Mean Frobenius matrix norm bootstrap result.

relative stability as only a few additional datasets are added. What appears to separate the individual brains is then not the overall shapes but, as might seem obvious, the more subtle variations on those shapes as morphology varies from person to person, as evidenced by the slower convergence as finer modes of variation are taking into account. Thankfully, for the purpose of this work the effects of eigenvector instability appear to be a moot point, for reasons that will be discussed in the next few paragraphs we will use a 10D truncation which, given the leave-one-out experiment planned, should be very stable after a 29 principal component analysis.

The next point of interest is whether eigenvector stability translates to stable streamline descriptivity across unique datasets. For 20 datasets in turn, streamlines were projected into feature space and then a matching approximation created using the coordinate-to-streamline projection operation defined in Equation 6.9. As an approximation of the error incurred during the projection, each streamline was compared (point-to-matching-point) with its simulated counterpart and the mean distances (mm) accumulated across each dataset. Examining Figure 6.12a we see that, as was expected, the mean error remains consistent across the 20 dataset sample , indicating a parity of descriptivity – notice that increasing the number of eigenvectors available for shape representation has the expected effect of decreasing overall error. If we examine this result in a slightly different manner (Fig. 6.12b), the average error across all datasets at a given eigenvector truncation reveals a sharp decline in errors as the initial few eigenvectors are added, followed by a period of steady decline after the inclusion of the 7th eigenvector. The
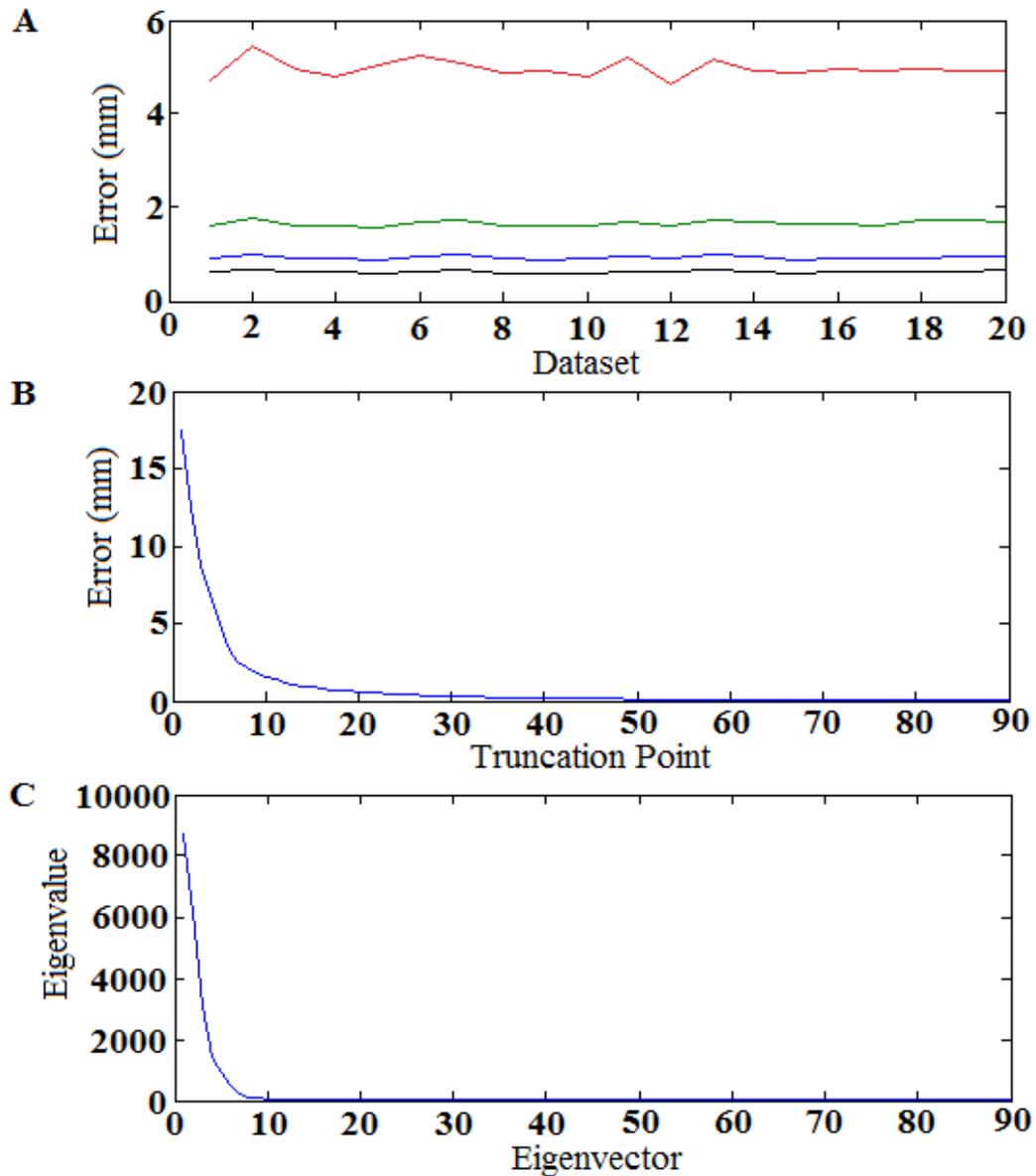
FIGURE 6.12: Analysis of error with respect to eigenvector truncation point. A: Error across a 20 dataset sample, note that the mean error remains consistent across the datasets indicating equal descriptivity. B: Mean error across all datasets, as expected error falls as truncation point increases. C: Eigenvalue profile for the feature space.

profile of these error reductions strongly correlates with the PCA eigenvalues (Fig. 6.12c) – an expected result given that these describe, in descending order, the variation along each axis - once again indicating the presence of a core group of shape basis functions responsible for defining general shape, with addition functions providing the 'fine tuning' required to describe each individual.

When selecting a truncation point, it is beneficial to decide the level of representational accuracy required. The soon-to-be-specified methods rely on a two-pronged approach, combining a shape label with a spatial location in order to identify a streamline as belonging to some structure. Given the motivation of this particular experiment – bulk

segmentation of large complete structures rather than exquisite subdivision of individual component's – total accuracy in shape representation is not required. Instead, we simply require a general concept of shape, i.e. a rough classification. If we examine the results related to Figures 6.11 and 6.12, there are repeated indications that a truncation point between 7 and 10 dimensions should provide a sufficiently descriptive basis set to describe most streamline shapes (rapid fall then levelling of error, Fig. 6.12b), and that proceeding beyond this point is shown to provide both rapidly diminishing returns in representational accuracy (Fig. 6.12b, c) and eigenvalue stability (Fig. 6.11). Curiously, if we apply the 98% variance threshold discussed in the PCA dimensional reduction section, a similar recommendation of 9 dimensions is retrieved. Thus, given that the data suggests that any point within the 7-10 dimension region should be sufficient and that there is indeed a need to choose one, we have selected 10 dimensions for the remainder of this work.

6.5.3 Feature Space Exploration

With a feature space now defined (rotating 29 datasets, 10 dimension truncation), it is now important to check how the various structures present within the brain are delineated and distributed. The simplest method (Figure 6.13) is to calculate distances between the points belonging to various shape classes, note that while we will use the Euclidean distance, others are equally valid (e.g. Mahalanobis distance). Figure 6.13a, for example, displays distances between approximately 16500 unsorted examples of streamlines belonging to the Uncinate Fasciculus. While the raw data does not appear very revealing in this form, upon (spectral) re-ordering (Fig. 6.13b), there is a clear delineation of the data into two clusters which, as was predicted, correspond to the two possible directional encodings of the streamline shape this results is particularly important for later support vector machine (SVM, Cortes and Vapnik (1995)) based segmentations (see section 6.5.6 for further discussion on this point)..
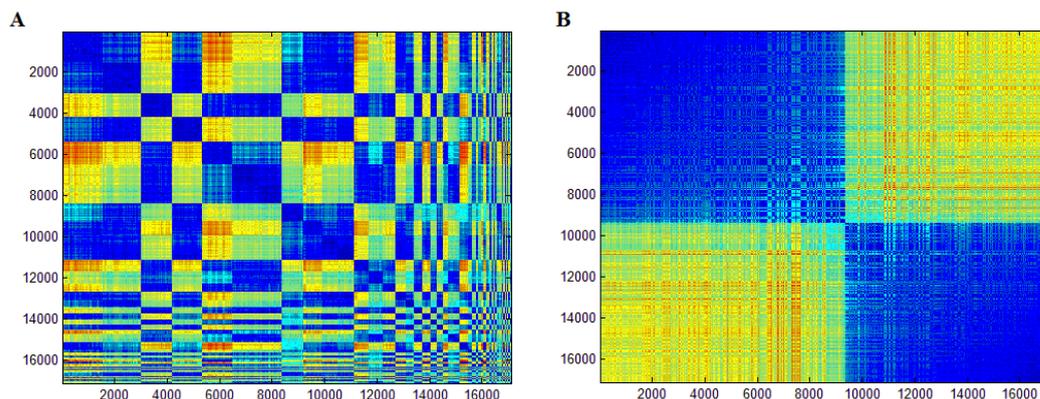
FIGURE 6.13: A: Distances within feature space between streamlines belonging to the Uncinate Fasciculus. B: Spectral re-ordering of A, note that the distances divide neatly into two clusters corresponding to the two possible orders for streamline representation.

Ignoring the issue of symmetry for a moment, the problem with distance calculations is that similarly shaped streamlines appear to be spread over concise but elongated sections of feature space. For example, if we measure the intra-uncinate and intra-arcuate streamline differences, the mean distance does not appear to be significantly lower than the mean SLF to Uncinate distance, however if we then visualise that data (Fig. 6.14, first three dimensions), while there is admittedly mild overlap due to commonality of shape at the edge cases, the majority of streamlines project to distinct and easily separated regions, boding well for later segmentation techniques.
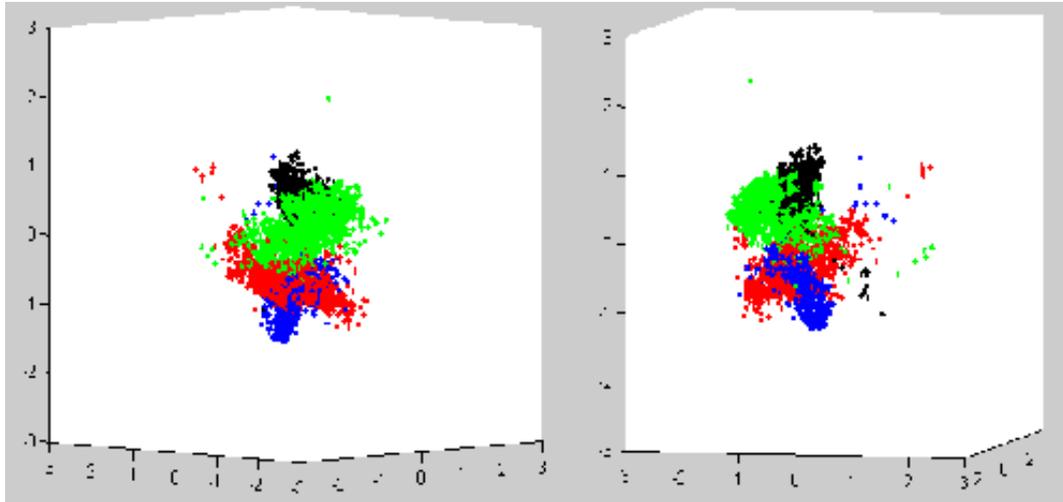
FIGURE 6.14: Visualisation of the first three dimensions of feature space. Projected points correspond to the left Arcuate (green), right Arcuate (red), left Uncinate (black) and right uncinate (blue)

Another desirable quality for the feature space is a smooth transition between shapes across the regions, any discontinuities or sharp changes in gradient would greatly complicate segmentation. While this is difficult to quantify, Figure 6.15 displays the result of following a straight line trajectory between points representing the Arcuate Fasciculus (1) and Corpus Callosum (10) with visualisation of the modelled shape at 8 equidistant points. Note the smooth logical transition in shape and size as the change progresses, indicating feature space consistency the given trajectory. While this of course does not imply consistency across the entire feature space, additional visualisations provided in Appendix D (covering a range of alternative trajectories) demonstrate similar traits, providing additional evidence in favor of this conclusion.

6.5.4 Feature Space Parcellation

There are two readily implementable methods for feature space parcellation, the first involves clustering and the second an array of support vector machines. While the eventual implementations of each method differ, both begin by projecting the original training data into the newly created feature space. As before, training data is concatenated with
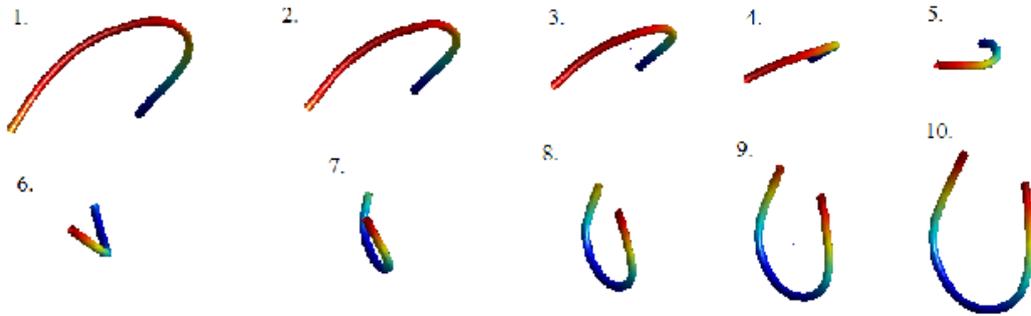
FIGURE 6.15: Visualisations of a linear path through feature space beginning at the arcuate (1) and progressing with equidistant steps (2-9) towards the Corpus Callosum (10).

its reversed duplicate, transformed into a standard image space, reduced to a 30 point equidistant knot representation and then translated on to the origin (i.e. the normalisation process). Following this, streamlines are then embedded into the truncated 10 dimensional feature space using Equation 6.7.

6.5.5 Feature Space Clustering

Cluster based segmentation operates by parcellating the feature space using the results of a k-means clustering of the embedded training data coordinates. The obvious drawback to K-means is that the number of clusters, or in this case shapes, would ideally be known a priori. Examining related literature, the estimated number of distinct white matter bundles within the brain ranges from 1-200 (O'Donnell and Westin (2007), Westin (2010)) which, assuming for a moment that each bundle possesses a unique shape, suggests a minimum of 2-400 clusters will be required for a suitable parcellation of shape space, remembering that each shape may be represented with two-point symmetry. Fortunately, absolute precision does not appear necessary with the method that will be outlined, though we will return to discussing the number of clusters shortly (it is beneficial to the discussion to present the general method first).

Following clustering, the result of primary interest is the location of each cluster centroid, to which novel data will later be assigned based on a closest (Euclidean) distance measure from their point of projection – in essence forming a de-facto Voronoi tesselation (Voronoi (1908)) of the feature space with each cluster centroid defining a region falling under a common label. The principal task for this method of segmentation is then to determine which cluster centroids correspond with which streamline shapes and thus, by extension, which labels correspond to which white matter structures.

Achieving this task begins with manual segmentation of the corresponding streamlines within the training datasets, however, unlike the previous ASM method it is no longer necessary to break larger structures into their component bundles since the structures
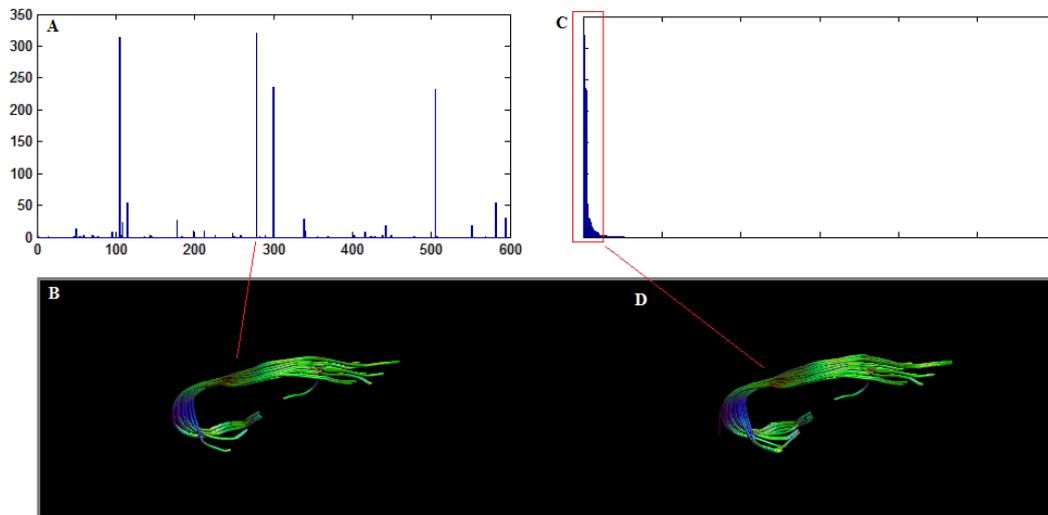
FIGURE 6.16: A: Histogram of cluster intersect frequencies (clusters labeled 1-600). B: Example of streamlines assigned to cluster 279 (single dataset), note that the main shape is captured. C: Rearrangement of histogram A, sorted in descending order of frequency. D: Contents of the first 90% of clusters (cumulative sum) by visitation frequency. Note that the core shape retrieved in B is expanded to include finer details.

will be defined by the collection of cluster centroids incorporating the different individual shapes. Following segmentation, streamlines are normalised as before and then projected into the feature space. Once projected, each streamline is assigned to the nearest cluster centroid and a histogram of cluster encounters constructed (Fig. 6.16a). Each encountered cluster is then manually inspected (Fig. 6.16b); this involves visualising all streamlines listed as belonging to the cluster (using the k-means result, not the projection of the segmented data) and, through expert knowledge/experience, determining if a sufficient portion of the visualised material belongs to the structure of interest. As will be discussed shortly, small overlap with other structures is reasonably likely, though not an issue once spatial masking is applied. The end result should be a set of related clusters centroids covering feature space regions describing shapes belonging to the structure of interest. As an aside, while maximum accuracy will always be achieved by manual inspection of all clusters, as a general rule of thumb, if one arranges the visitation histogram into descending order (Fig. 6.16c), the set covering the first 90% of cumulative visitations can generally be accepted on faith (Fig. 6.16d), reducing manual inspection workloads to covering the remaining 10%.

Segmentation of new data is then a trivial matter. Candidate streamlines are normalised and projected into the feature space defined by the training data. Once in feature space, each streamline is assigned a label according to the nearest cluster centroid. As previously stipulated, while shape space parcellations do define single shape classes, shapes can be repeated throughout the brain and thus a single cluster may be affiliated with multiple (as a worse case, usually only 1-2 affiliations – discussion on this point

soon) and thus cluster labels alone are insufficiently unique to determine a streamlines affiliation.

Returning to the feature space creation, one of the variances we chose to eliminate was streamline position within the brain. This was for two reasons: (1) Eliminating position as a mode of variation increases the significance of variance caused by shape, thus increasing the sensitivity to smaller changes in shape and, by extension, reducing reliance on the following method. (2) Once registered to MNI space, estimating a streamline's position within the brain is simple. As before, the manually segmented streamlines used for cluster content identification are used to create a coarse positional mask which, in combination with a shape classification, is sufficient to resolve the feature space ambiguity (Fig. 6.17).
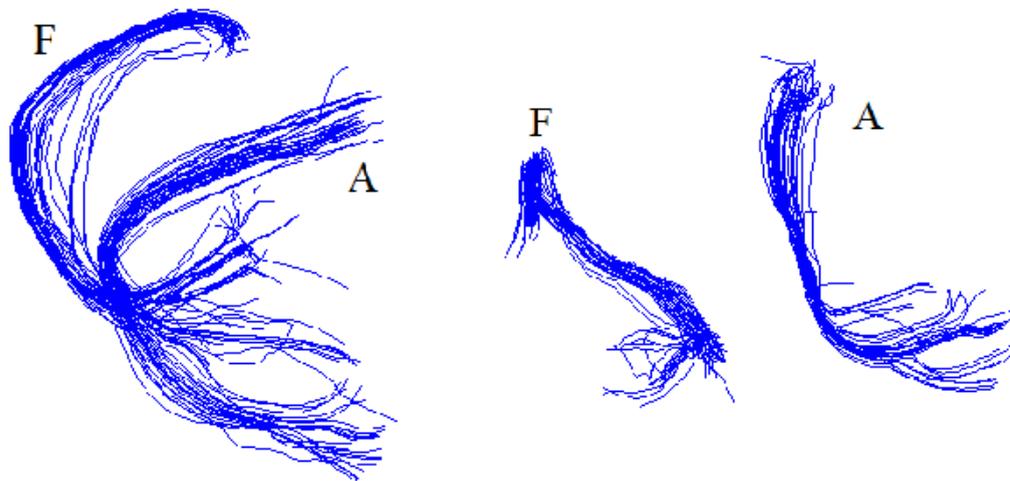
FIGURE 6.17: Streamlines belonging to one cluster in the 600 cluster space; containing streamline shapes shared between the right Fornix (Labelled F) and right Arcuate Fasciculus (Labelled A). Note that while the shapes are similar, there is sufficient distance (within Image space) between each group for spatial masking to easily resolve the shape label ambiguity.

The remaining question is to determine a suitable a priori estimate for the number of clusters. To achieve this, projected whole volume training data were first clustered with a priori estimates ranging from 200 to 1000 at 100 cluster intervals. Following this, manually segmented training data pertaining to the Uncinate (left and right hemisphere), Fornix, Corpus Callosum, Cingulum (left/right), Middle Cerebellar Peduncle, Cortico Spinal Tracts (left/right) and Arcuate Fasciculus (left/right) were then projected into the clustered space and assigned a label according to the nearest cluster centroid. Given an ideal clustering, the resultant labellings would be such that each centroid is visited by streamlines belonging to precisely one (or none, given that other shapes exist which are not covered by the given sample) of the represented samples. Unfortunately, without extremely high cluster counts this level of fidelity is unlikely to be achievable, in large part due to a combination of between-subject anatomical variations and premature streamline termination creating an unavoidable similarity/overlap of shapes between structures

sharing a similar fibrous architecture. Thus, given a practical number of clusters (200-1000), the best we might hope for is that the majority of clusters are single-structure and shared labelling is kept to a minimum.
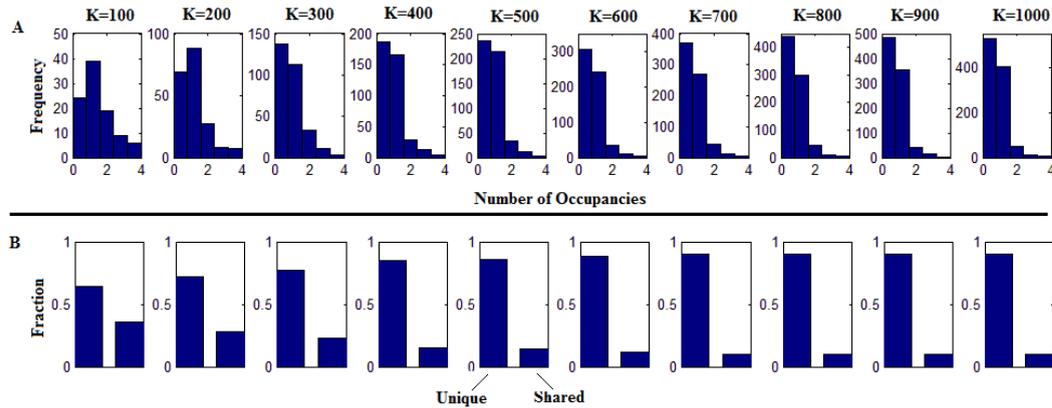
FIGURE 6.18: A: Histogram of the number of unique structures intersecting individual clusters. K-means clustering performed from $K = 100$ to $K = 1000$ from left to right. B: The fraction of manually labelled streamlines belonging to (left) clusters shared only with streamlines from the same structure and (right) clusters shared with multiple structures.

Examining Figure 6.18, we see that this prediction holds true. Increasing the number of clusters does indeed reduce the portion of clusters that are shared across two or more structures (Fig. 6.18a) but, at least within the examined range, will not eliminate them completely. As a balance to this, we also observe that as K increases, so does the portion of unoccupied clusters (represented as 0 occupancy in Fig. 6.18a) which, in fact, becomes the dominant state by $K = 300$. While a large portion of these empty clusters would of course be due to the limited set of structures involved in the test (inclusion of the Optic Radiations, for example, would fill a segment of the empty set), we would also speculate that others are simply being assigned to increasingly small divisions of "waste" streamlines, e.g. the various association fibres that are of little interest to the current work, or streamlines that break prematurely due to image specific artefacts, creating a range of arbitrary shapes sufficiently dissimilar from the general corpus of streamlines to demand an disproportionate number of the additional clusters created by increasing K. On the positive side, the assignment of waste streamlines to their own clusters would separate them from "useful" parcellations, thus improving the feature space segmentation result and decreasing the reliance on spatial masking. However, examining the negative consequences, notice that between $K = 600$ and $K = 1000$ the ratios between the non-zero occupancies barely changes while the zero occupancy continues to increase. What this suggests is that although finer and finer subdivision is occurring, the shape similarity overlap is not really being resolved - as discussed there are cases where streamlines from different structures are simply the same shape. This is backed up by Figure 6.18b which displays the fraction of streamlines (from the segmented set) which belong to a cluster attributed solely to its own structure, note that while there are steady improvements from $K = 100$ to $K = 400$ this plateaus as

one progresses further. Weighing up this information, it appears there is a consistent region of overlap between these structures which, at least when using K-means, cannot be resolved within a reasonable number of clusters, thus increasing K beyond $K \approx 600$ is unlikely to increase segmentation fidelity significantly enough to justify the expense of checking projected coordinates against hundreds of additional cluster centroids. For this reason, the following cluster based segmentation experiments will be conducted at $K = 600$ which provides a reasonable compromise between fidelity and speed.

Actual segmentation results will be displayed in conjunction with those of the following subsection.

6.5.6 Support Vector Machines

One of the drawbacks of standard K-means clustering (or any unsupervised method) is that it is difficult to incorporate a priori information about known boundaries between entities. Examining the toy example provided in Figure 6.19, using a priori information it is clear that there are two groups existing with the defined region of space (Fig. 6.19a, red/blue labelling). Once that information is removed (Fig. 6.19b), i.e. as K-means would see the data, there are no longer any obvious hints as to where that boundary may lie. If we then cluster the data using the correct value for K (Fig. 6.19c, 2 in this case) there are no guarantees that the output of the distance minimisation process will produce a correct labelling, potentially resulting in a misclassification of spatially similar, though distinct points. (Fig. 6.19d).
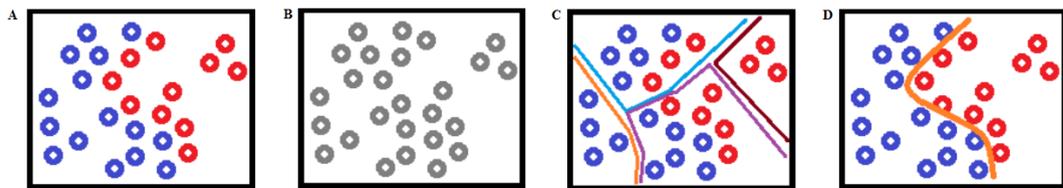
FIGURE 6.19: Toy example highlighting the drawbacks of an unsupervised segmentation. A: The boundary between classes is clearly visible with a priori information. B: Without the a priori information, selecting the boundary becomes difficult. C: A possible K-means result. K=2. D: Notice how this result, while minimising distances between cluster centroids and thier members, misclassifies the underlying data.

Practically speaking, this is not a significant issue - cluster based results will show that a simple spatial mask provides sufficient additional information to resolve the shape space ambiguity. However, to reduce this drawback or achieve a more precise segmentation, supervised learning methods may be of interest. Of the available techniques, support vector machines are an appealing option due to their wide use within segmentation literature; achieving efficient segmentation of data through projection into a higher dimension space, in which a maximum margin hyperplane (or other function) may be defined, providing delineation between the class of interest and all other data points.

The drawback of the SVM is that for most implementations classification is once again binary – either a streamline belongs to the class of interest or it belongs to some other unknown class. Like the ASM method, the solution is to revert to a bank of support vector machines with one SVM trained for each bundle of interest. Though unlike the ASM method, these support vector machines could operate on streamline coordinates from a single feature space.

Defining individual bundles may be achieved in two ways; either manually or semi-automatically. The manual method is very similar to that described for ASM training with the exception that, since streamline directionality is preserved in the whole brain space, bundles must be further sub-divided into two groups, one for each directed representation. To segment an entire volume in this manner would obviously require an unreasonable amount of time/effort, thus to alleviate this problem we have designed a second method based around affinity propagation (AP, Frey and Dueck (2007)) clustering. As briefly discussed within the introduction to this chapter, AP clustering uses a form of message passing (the affinity propagation) to achieve an optimal clustering of the supplied data which, if no a priori information is provided, includes autonomous determination of the number of clusters present. To achieve AP clustering, structures are segmented in bulk (no subdivision) and the corresponding streamlines normalised and projected into the whole volume feature space. From here, a square distance matrix is created by calculating the Cartesian distance between each coordinate pair. This matrix is then fed into the AP clustering software (matlab code provided by Frey and Dueck), with data-point preferences – values that may be provided to influence the number and position of cluster centroids - set to an even value (median inter-streamline distance), indicating that no a priori information is available. Figure 6.20a provides the output of AP derived bundling of the right hemisphere Fornix branch.

Regardless of bundling method, SVM production then requires a 'negative' dataset, providing an array of points not within the set of interest. To provide example data we normalised and projected 10 whole volume datasets and then selected, at random, an equivalent number of points corresponding to the number of streamlines within the target bundle – the main caveat being that streamline coordinates falling within the mean minimum distance between bundle coordinates (i.e. those which would conform to an overlapping shape, potentially arising from the same streamlines in the 10 whole volumes) are excluded as potential candidates for the negative example set – to include them would cause training issues. Figure 6.20b visual representation of the training data used to construct support vector machines for the four sub-bundles comprising the right hemisphere Fornix branch in Figure 6.20a. Note how each sub bundle falls into a distinct feature space region.

Finally, for each training data group, support vector machines are constructed using the libsvm toolkit (Chang and Lin (2011)) with a radial basis function kernel and cost/gamma parameters specified by the provided setup tool (libsvm "easy.py", tool
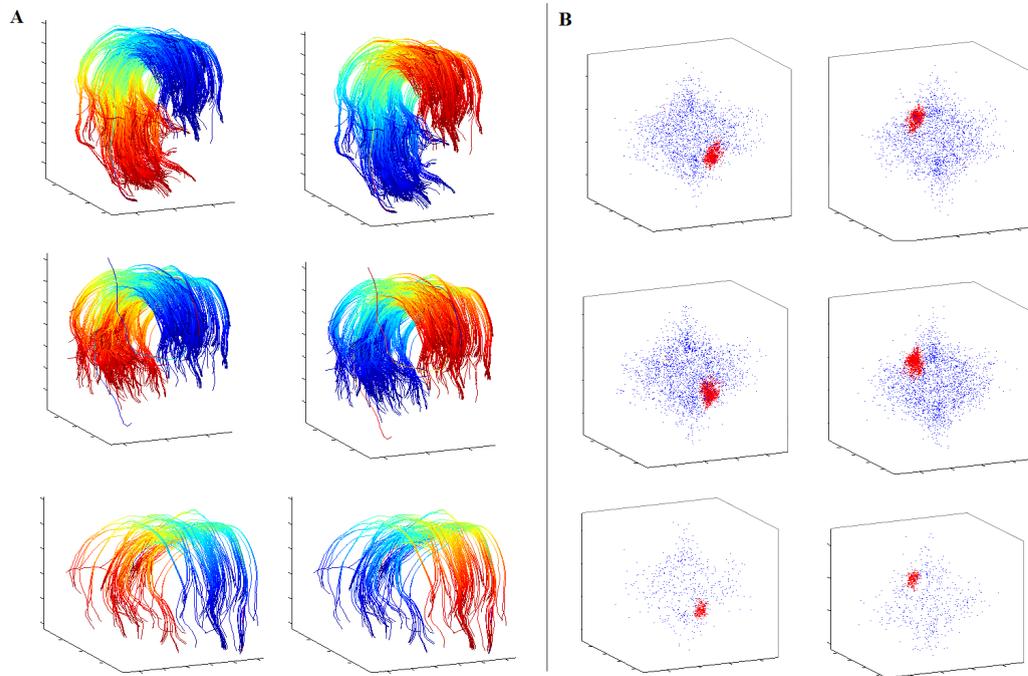
FIGURE 6.20: A: Fornix bundle segmentation through affinity propagation clustering. Note that each subsequent cluster depicts a shorter length of streamline (from full length to short artefactual truncations) or its directional opposite. B: Visual depiction of the SVM training data (first 3 feature space coordinates), red dots correspond to the Fornix streamline coordinates, blue to the additional training data.

searches the cost/gamma space for the optimal combination for a given training dataset, using cross-referencing to ensure accuracy). The segmentation itself is then a simple matter. Novel streamlines are embedded into the feature space where each streamline is then tested, in turn, against the assembled bank of one-vs-all SVM classifiers. Should a positive identification occur, the streamlines image space position is cross-referenced with a corresponding spatial mask and assigned to a bundle or sent for further testing accordingly.

6.5.7   Whole Volume Segmentation: Results

As an initial demonstration, a selection of in situ whole volume segmentation results (alternately produced by clustering and SVM segmentation methods) are presented in Figure 6.21 with individual structures colour coded according to table one. These are joined in Figures 6.22-6.25 by demonstrations of white matter segmentation (with accompanying manual segmentations for qualitative comparison) for the Uncinate, Cortico spinal tracts, Middle Cerebellar Peduncle and Cingulum.

For quantitative analysis, examples of the Fornix and Arcuate Fasciculus were manualy segmented from the full 30 dataset corpus. Then, using the methods described above, both cluster and SVM based segmentations were used to automatically segment these
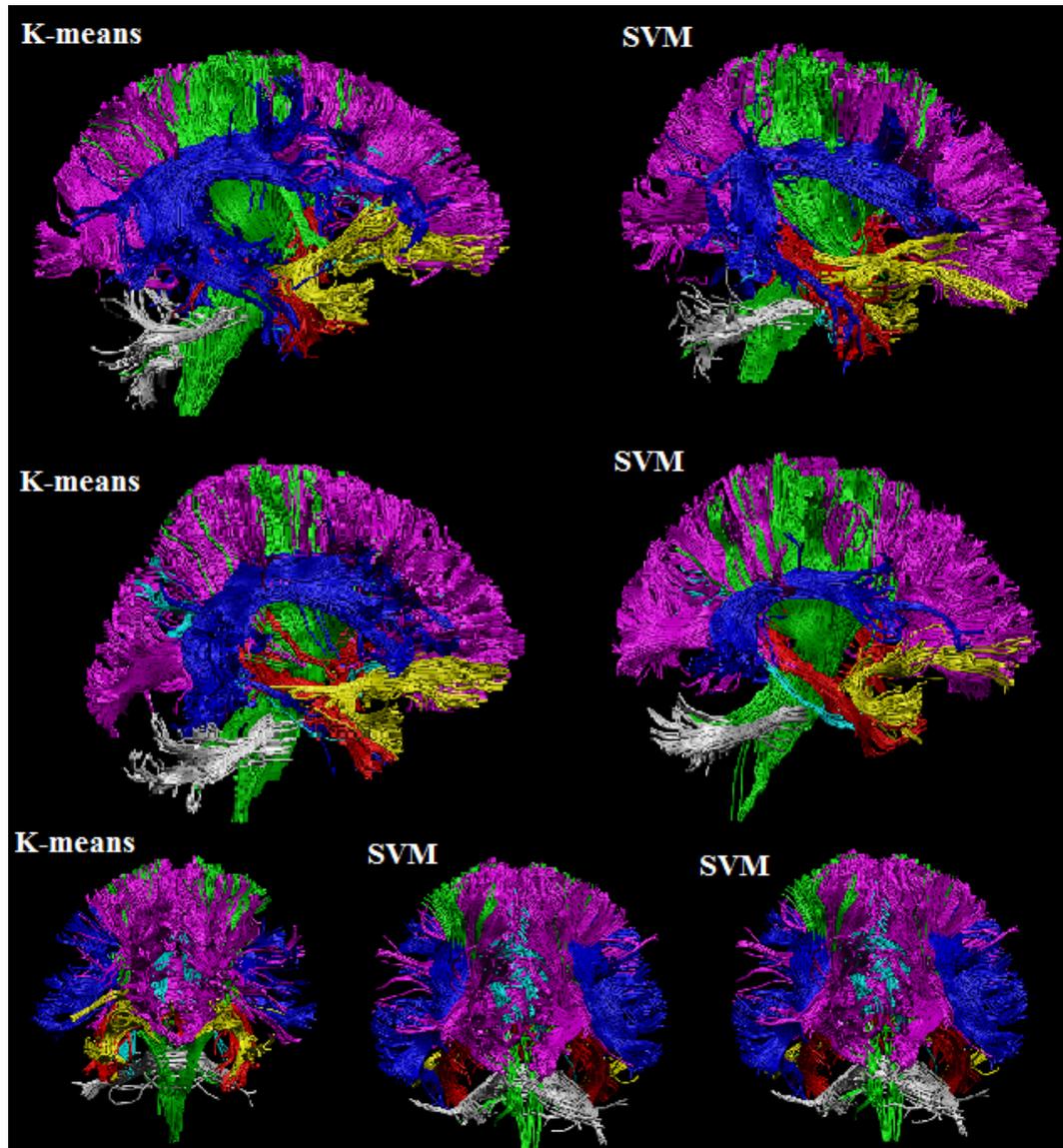
FIGURE 6.21: Selection of whole volume segmentations. Structures colour coded according to table 1.

structures from the same 30 datasets using a leave-one-out training regime (i.e. if segmenting dataset one, manual segmentations for two through thirty would be used to provide the training material).

For a more quantitative approach, using the full 30 dataset corpus, 30 unique example of the Fornix and Arcuate Fasciculus were manually segmented. Then, using the methods described above, both cluster and SVM based segmentations were used to automatically segment these structures from the same 30 datasets using a leave-one-out training regime (i.e. if segmenting dataset 1, manual segmentations for datasets 2 through 30 would be used to provide the training material).

Using the manual segmentations as a gold standard, Figure 6.26a-b displays recall (red) and precision (blue) statistics for the two structures using both cluster and SVM based

TABLE 6.1: Colour Coding

Structure	Colour
Arcuate Fasciculus	Dark Blue
Corpus Callosum	Purple
Cortico-Spinal Tracts	Green
Middle Cerebellar Peduncle	White
Cingulum	Cyan
Uncinate Fasciculus	Yellow
Fornix	Red
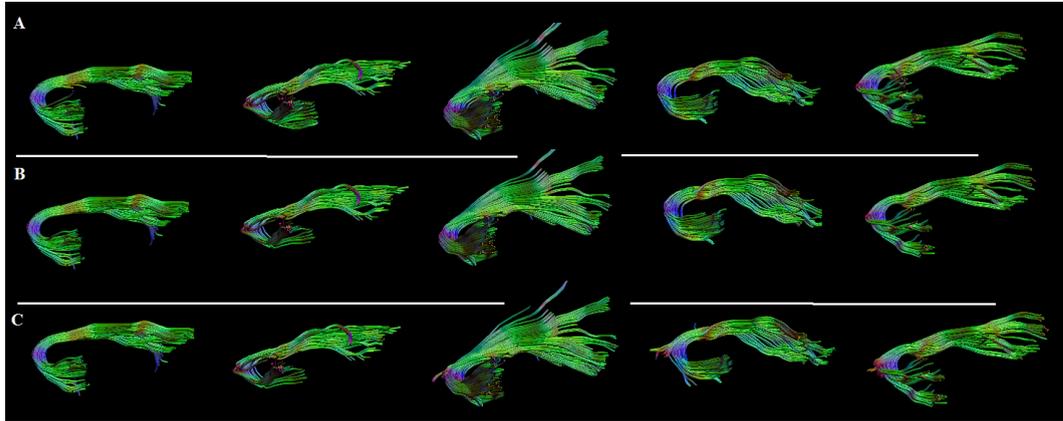
FIGURE 6.22: Selection of Uncinate Fasciculus segmentations. Row A: Manually segmented. Row B: Cluster segmentation. Row C: SVM segmentation.

segmentations. As we can see, agreement between the manual and automated segmentations is generally high across the datasets (labelled 1-30 on the X axis) with recall consistently indicating $\geq 85\%$ retrieval and precision, while lower (in part due to inconsistencies within the manual segmentations, a point that will be discussed shortly) averaging 75%. If we then examine agreement between the two automated segmentations, Dice coefficient scores (Fig. 6.26c, calculated according to Equation 6.14) are encouraging – averaging 93% commonality for the Arcuate (Red) and 91% for the Fornix (blue).

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \tag{6.14}$$

This numerical similarity is then born out during visual inspection. Figures 6.27 and 6.28 contain some of the 'poorest' segmentation results with respect to precision, though upon visual inspection there appear to be few if any streamlines not corresponding to the Arcuate or Fornix respectively, the automated results simply provide a "fuller" reconstruction.

As a final comparison, we will now briefly demonstrate the unmasked (shape only) segmentation results. While the final results for both cluster and SVM based segmentations
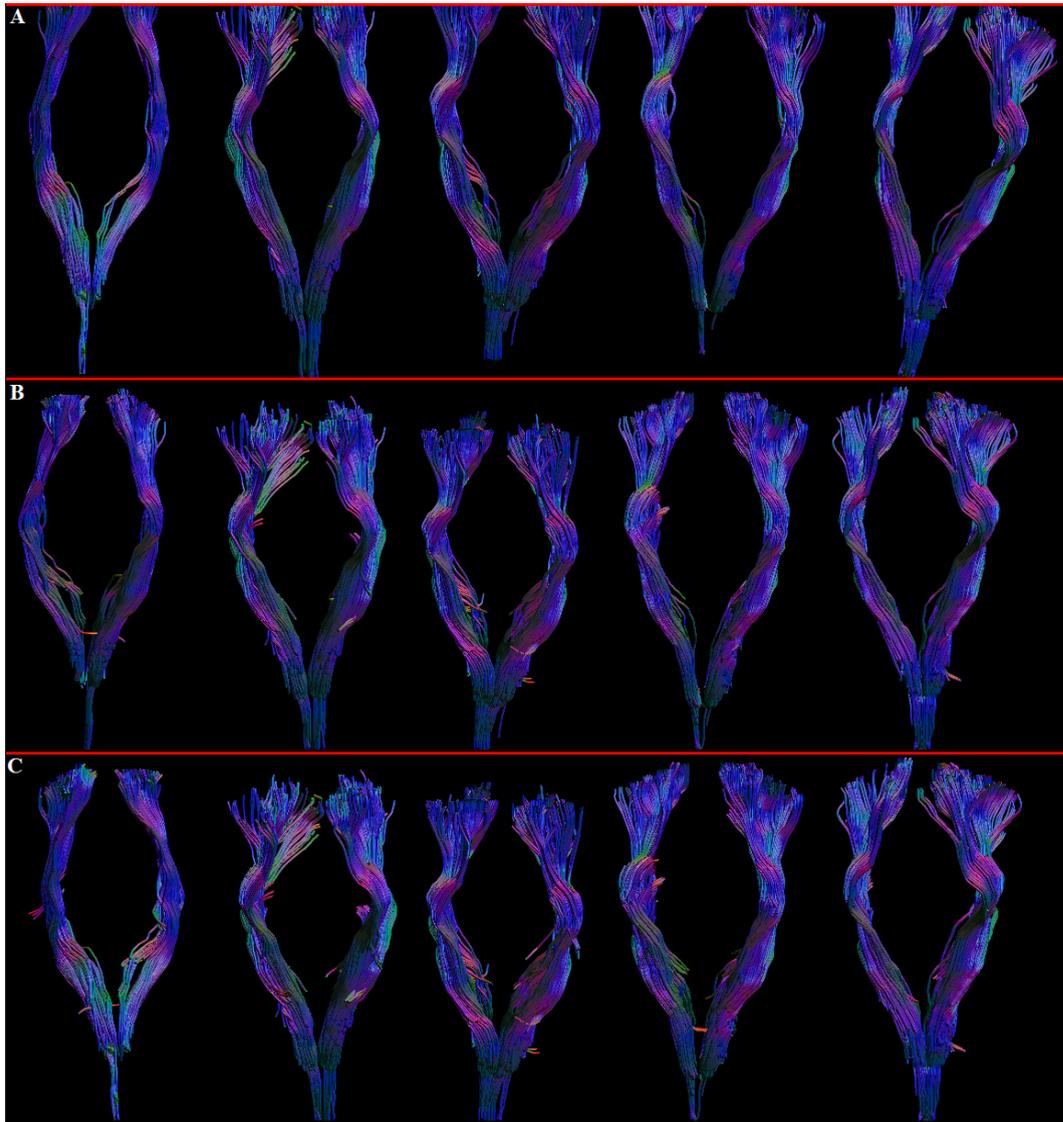
FIGURE 6.23: Selection of Cortico Spinal Tract segmentations. Row A: Manually segmented. Row B: Cluster segmentation. Row C: SVM segmentation.

appear very similar, the unmasked segmentations can differ significantly. Figure 6.29 for example demonstrates an Arcuate segmentation. Beginning with the clustering method, the most pressing issue with this form of feature space parcellation is that similarities between edge-case shapes can lead to differing structures sharing entire clusters, amplifying the effect of true shape commonality throughout the brain. The result of this can be seen in Figure 6.29a where edge-case shape similarity has lead to cluster commonality causing the acceptance of streamlines belonging to both the Cingulum and Uncinate Fasciculus. In comparison, the SVM result (Fig. 6.29b), while still providing some false positives, does so in greatly reduced quantities. The implications of this are twofold: (1) There is significantly greater fidelity available within the feature space than is accessible within a practical number of clusters (remembering the experimentation in section 6.5.5) and (2) Dependant on further work, support vector machines appear to a viable option
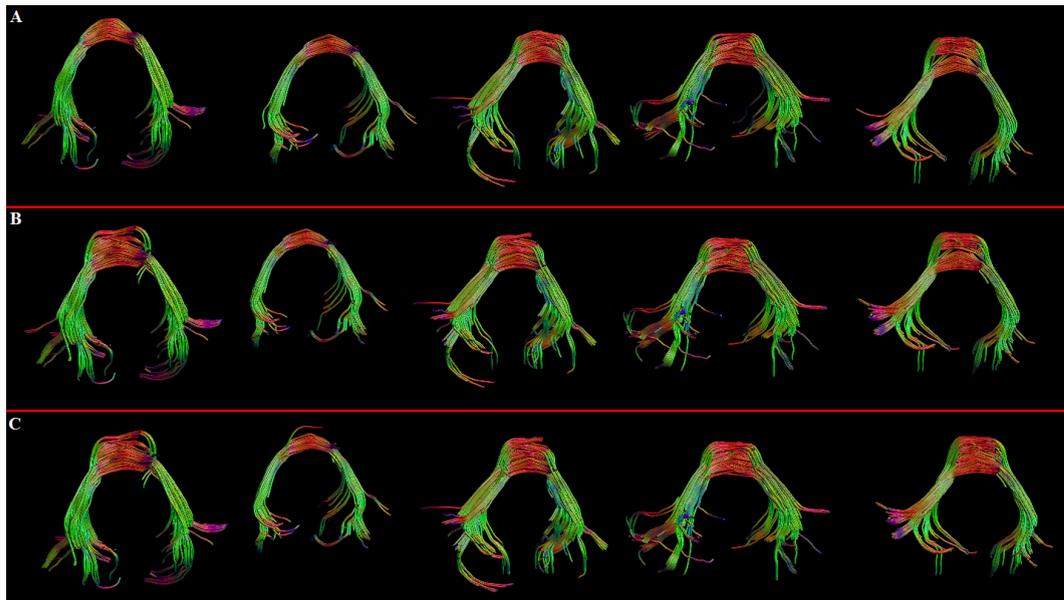
FIGURE 6.24: Selection of Middle Cerebellar Peduncle segmentations. Row A: Manually segmented. Row B: Cluster segmentation. Row C: SVM segmentation.
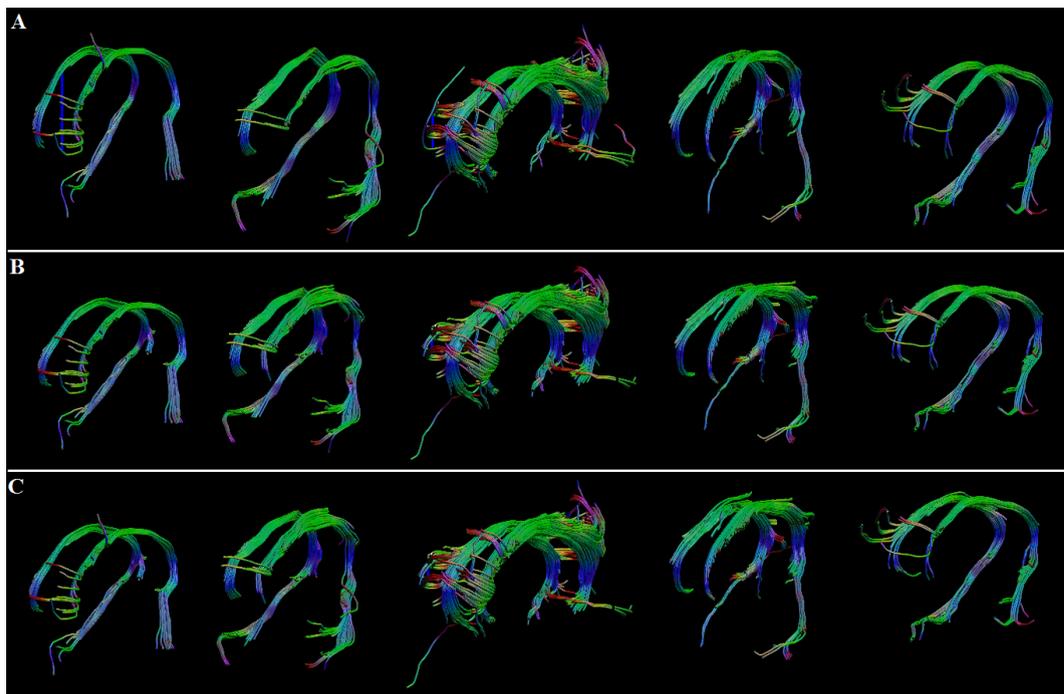
FIGURE 6.25: Selection of Cingulum segmentations. Row A: Manually segmented. Row B: Cluster segmentation. Row C: SVM segmentation.

for accessing this fidelity, perhaps providing a method for more exquisite sub-divisions. This all being said, however, if we take the union of either set of streamlines with the set specified by an appropriate spatial mask (Fig. 6.29c), both results are entirely sufficient to perform the bulk segmentation task examined in this work (Fig. 6.29d).
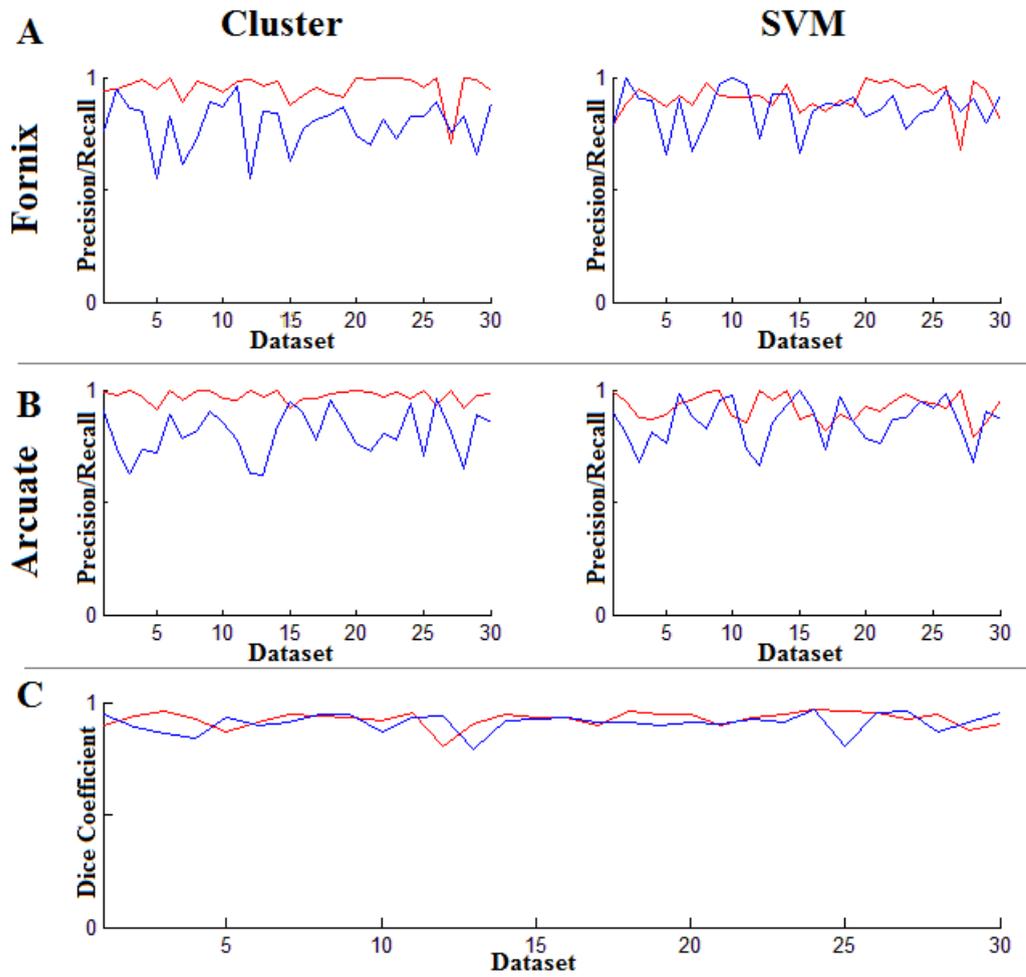
FIGURE 6.26: Quantitative statistics. A/B: Recall (red) and precision (blue) statistics across the 30 datasets using manual segmentations as a gold standard. C: Dice coefficient scores comparing the two automated segmentation methods. Red indicates the Arcuate and blue the Fornix.
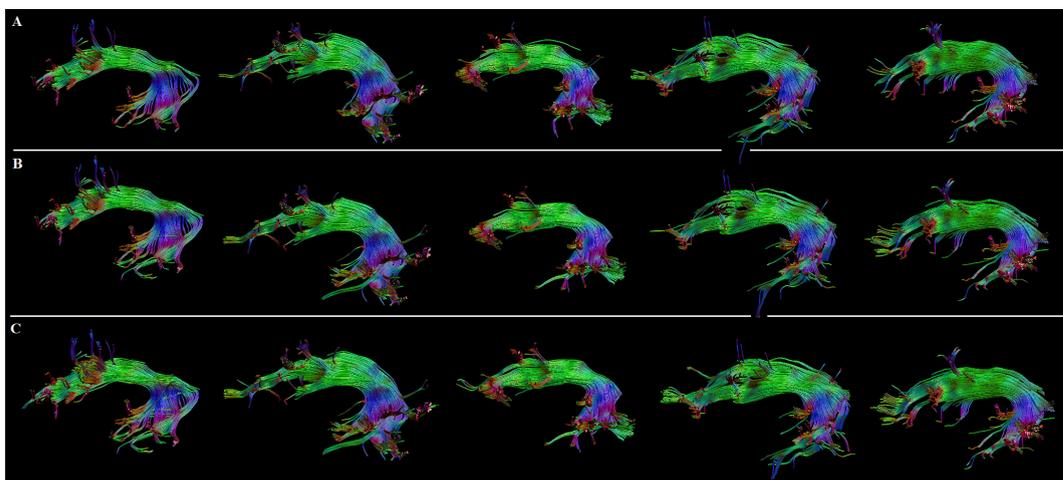
FIGURE 6.27: Selection of Arcuate Fasciculus segmentations, particularly those with lower than average precision. Row A: Manually segmented. Row B: Cluster segmentation. Row C: SVM segmentation.
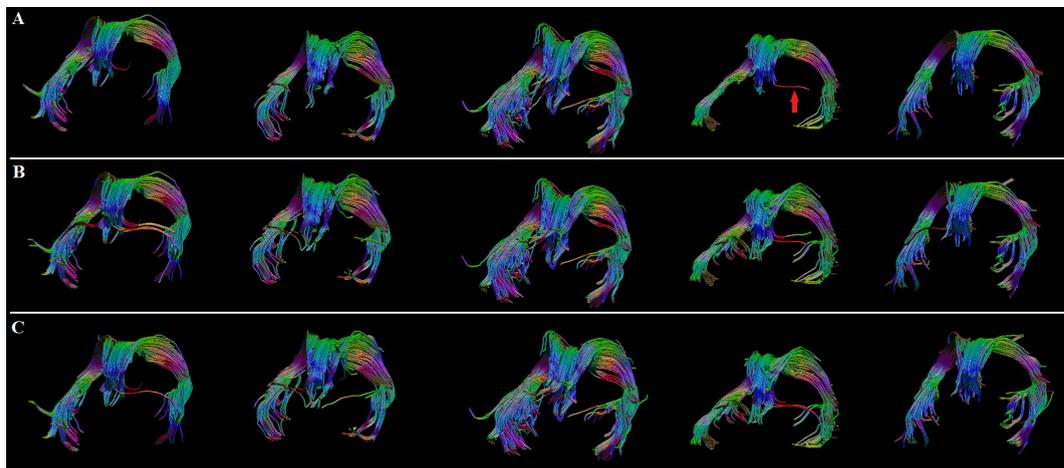
FIGURE 6.28: Selection of Fornix segmentations, particularly those with lower than average precision. Row A: Manually segmented. Row B: Cluster segmentation. Row C: SVM segmentation.
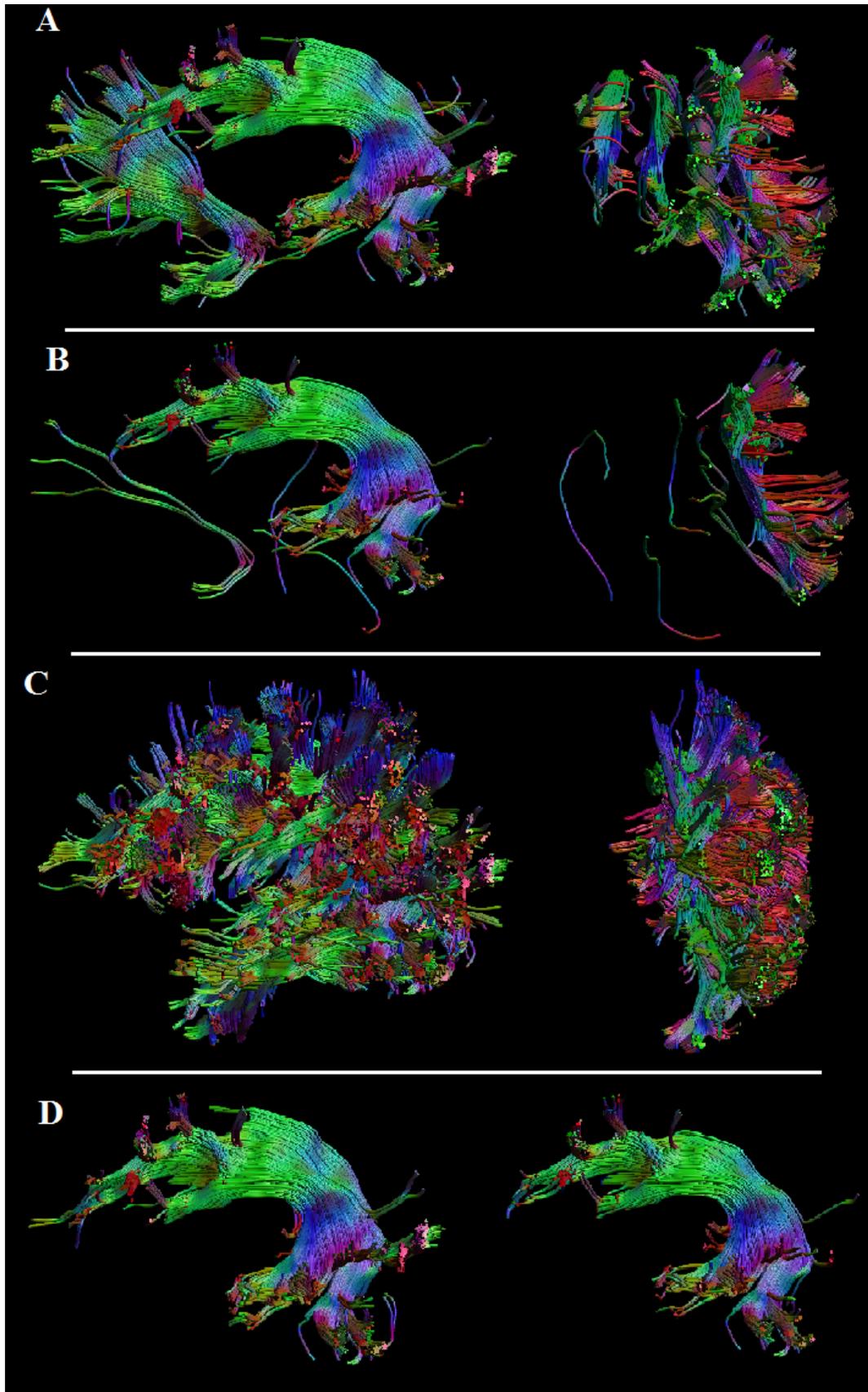
FIGURE 6.29: Example Arcuate Segmentation. A: Cluster segmentation without masking. B: SVM segmentation without masking. C: Streamlines passed by mask, note no Cingulum/Uncinate streamlines present. D: End results. Left - cluster. Right - SVM.

6.6   Discussion and Conclusion

The automated segmentation of whole volume tractography is a maturing topic within diffusion MRI research. While methods are improving, the vast majority are still constructed over a framework of expensive inter-streamline distance comparisons. Over the course of this chapter we have presented alternative methods for streamline which, eschewing the expensive comparisons, opt for a simpler and direct feature space embedding.

Before discussing each method in detail, it would first be prudent to cover the advantages and limitations of shape space analysis in general. To begin, as results have shown it is simply not possible to uniquely identify streamlines/bundles through shape alone. The realities of subject and tractography parameter variations ensure that structures, at least to some degree, will always share an overlapping set of shapes. Fortunately, this is not a terminal issue. As has been shown, provided the feature space is constructed and segmented in a way designed to reduce these overlaps to a minimum, the corresponding regions of shape overlap are sufficiently separate in native image space such that even an extremely naive masking technique can provide the additional information required to resolve the ambiguity.

The choice of individual method then depends on the application in mind. The basic active shape model, while a poor choice for whole volume work, allows absolute minimisation of non-shape variance within a specific segmentation task. This has the natural advantage of increasing shape sensitivity to a maximum and by extension should provide the ability to differentiate between subtle changes that may be imperceivable within the whole volume feature space (at stable eigenvector truncations) due to the portion of variances in such a space assigned to modelling streamline mirroring and rotation. In combination with the region growing algorithm (which incidentally also be applied to any segmentation result), the ASM method may be a useful tool for surgical planning/execution; facilitating the construction of specialised models for each state of the surgery in order to provide precise and targeted low turn-around automated segmentations.

Proceeding to whole volume segmentation and ignoring, for now, the differences between the individual techniques (i.e. cluster vs. SVM); there are several key changes from the active shape model. The first and most obvious differences is of course the ability to segment multiple structures from within a single feature space with the previously discussed trade-offs between segmentation efficiency (streamlines are only subject to one projection) against reduced shape fidelity. However, there are other, more subtle, differences that one must be aware of when using these methods which are largely concerned with the training phase. The whole volume segmentation methods described here possess what we will call multi-shape segmentation abilities. This is to say that, rather than manually supplying one bundle at a time for individual modelling, as would be the case for the ASM, these methods can accept an arbitrary collection of streamlines and

then automatically, through either assignment to a k-means derived cluster or affinity propagation clustering, divide those streamlines into similarly shaped bundles for use as training material, whose "lessons" will be applied to all future segmentations. Given this training procedure, one must be extremely careful about the kinds of streamline supplied during the training phase since any mistakes made here will propagate to all future segmentations. The best way in which to explain this point is to examine the Fornix segmentation results, a visual depiction of which is available in Figure 6.28. For various reasons, the number of streamlines representing a Fornix varies from dataset to dataset, even if tractography parameters remain constant. In some of the training data, the number of streamlines assigned to the Fornix was particularly low and so the decision was made (in order to boost the number of streamlines available for training) to include streamlines which divert in a horizontal manner for the anterior pillars (See red arrow in Fig. 6.28a) - a common artefact in Fornix tractography. With these examples now included within the training corpus, the recovery of similar shapes became, by design, acceptable during segmentation of novel datasets. If we then examine the automatically segmented results (Fig. 6.29b-c), we can observe that, as expected, such streamlines are recovered even when there is no need to do so, thus negatively affecting the corresponding precision metrics.

The lesson here is a simple yet important one. The described techniques are able to reproduce – with excellent accuracy – a consistent approximation of the segmentation style of the training operator. Unfortunately, this does not include leeway for the varying subjective judgements that an operator naturally makes over the course of a series of segmentations. Thus, for optimal performance, it is highly recommended that the operator decides a priori exactly which streamline variations are acceptable and then follows this precisely throughout the training sequence. Failure to do so, such as the Fornix example, will not necessarily *ruin* the output, but may lead to unexpected streamlines becoming present in subsequent automated segmentations.

With this caveat made and regardless of the exact implementation used, Figures 6.21 through to 6.28 clearly demonstrate that the described methods are able to perform the assigned segmentation tasks over a variety of white-matter structures. Making the decision between the cluster and SVM methods then depends on what the user most values. The first two rows of Table 6.2 provide the mean segmentation time required to produce a whole volume segmentation matching those in Figure 6.21. For both methods, a significant portion of that time was spent performing common tasks – the affine image co-registration ($\approx 1min$) and normalisation and projection of streamlines ($\approx 2 - 3min$) – however, there is still a substantial difference in speed between the cluster and SVM based segmentations. The reason for this is again fairly simple, classification through a bank of RBF SVM's is far more complex than the simple euclidean distance calculations required for the cluster method. Thus, for rapid preliminary segmentations the cluster segmentation is clearly superior. The real split comes when looking at unusual data or

attempting more sophisticated segmentations within the whole volume space. Clustering, and unsupervised segmentation as a whole, is a blunt tool. As previously discussed, shape fidelity is directly related to the number of clusters (see Section 6.5.5, Figure 6.18); with the downside that as you increase the number of clusters, you run the risk of over-segmenting the feature space, and thus greatly increasing the number of distance comparisons required (slowing things down) while observing only minor improvements in practical segmentation quality. The solution to this problem is more likely to be a tailored support vector machine. While the SVM's used in this worth thus far have been rather generalised for segmentation of large structures, the differences in Figure 6.29 provide a telling indication that through provision of tailored training data (and associated SVM parameters), SVM's may be capable of a significantly more precise feature space parcellation with the added benefits that may bring (e.g. improved potential for sub-division of large structures).

Finally, returning to the more general question of speed, rows three and four of Table 6.2 approximate the average time required to produce a distance matrix (and *only* a distance matrix) using both the method described by O'Donnell and Westin (2007) and an exhaustive Hausdorff calculation (as employed by numerous other algorithms) assuming the same ≈ 55000 streamline dataset and single threaded MATLAB environment used throughout this work. While in reality it is likely that these distance calculations would actually take place in a highly parallelised environment, thus greatly reducing the listed time, the fact that the proposed methods are so much faster within the same environment (while still producing an accurate segmentation) surely cements the case for the advantages of shape based feature spaces over the currently popular alternatives.

TABLE 6.2: Algorithm Runtime.

Method	Time elapsed
K-means	6.01 minutes
SVM	8.3 minutes
O'Donnell	17.2 hours
Exhaustive	87 hours

Chapter 7

Conclusions and Future Work

This thesis was intended to fulfil two functions, firstly to provide the groundwork for further diffusion weighted imaging in the maxillofacial region and secondly to create a robust, automated pipeline for the processing of DW-MRI signal data – muscle or white matter. With the novel material provided in the previous chapters we believe these criteria have been adequately met. In this final chapter we will therefore conclude by providing a brief overview of each contribution and then discussing ways in which the presented work could be expanded, either to further the goals of this project, or to provide increased contributions to the field of DW-MRI in general.

7.1 Robust Processing

With respect to robust processing of DW-MRI data, this thesis has, in order, provided a novel method for the automated identification of outliers/corruption in DW-MRI images, an analysis of spherical deconvolution methods (particularly optimal methods for CSHD calibration) and a novel shape based streamline segmentation tool. While these were all initially inspired by tasks required for the accurate segmentation of muscle tissue – compensation for the frequent signal drop outs, accurate resolution of crossing muscle fibres and an automated segmentation tool to lower clinician workload – their applicability extends beyond the immediate field and into DW-MRI techniques in general.

7.1.1 Rejection of Corrupted DW-MRI Signal Elements

Chapter 3 demonstrates, through simulations and in vivo experimentation, that RES-DORE either matches or exceeds the performance of comparable methods with respect to both the reliability of detection of corrupted signal elements and the minimisation of false positive eliminations. The major drawback to this method, however, is that

in voxel-wise applications the iterative update and fit process (which must be run one voxel at a time) is more complex than comparable methods (particularly HOMOR) and thus incurs a longer than desirable run-time (5-10min processing time for a typical $128 \times 128 \times 60$ voxel image). While not prohibitive for a low number of datasets, this could foreseeably become problematic for larger studies.

One solution would be to develop software that farms the task out to a high-performance multi-cpu cluster and thus reduces overall processing times through parallelisation of the required calculations. While this is indeed feasible in larger research establishments, wider access to such hardware is generally limited and thus a multi-CPU implementation may not be the best choice for widespread adoption. Fortunately however, recent developments in GP-GPU technology, for example Nvidia's CUDA (Nvidia, Santa Clara, CA, USA), have provided comparatively cheap access to another form of highly parallel processing that would be ideal for speeding up the RESDORE algorithm. With this in mind, and along with further testing and tuning of the algorithm itself, a GP-GPU based implementation could be an interesting option to persue for a future RESDORE publication.

7.1.2 Analysis of Spherical Deconvolution

The simulations in Chapters 4 and 5 both backed up our preliminary empirical findings (that CSHD may be suitable for resolution of crossing muscle fibres in vivo) and also provided an added exploration of the effects of calibration on the reliability of spherical deconvolution derived fODF approximations (results that are interesting to the wider diffusion MRI community). Thus, towards the goals of the project, we have discovered at least one method that is theoretically capable of performing the required fODF estimation, leaving the remaining task of demonstrating that capability in vivo as the most obvious future work (while results of the preliminary examination in Chapter 5 are certainly positive, CSHD applicability can only be conclusively proven once higher quality data is acquired and repeatability studies conducted). Thus, from a project point of view, this simulation aspect of this work is essentially complete; though it would certainly be interesting (and beneficial) to persue similar analysis of other HARDI methods, particularly those with the ability to perform per-voxel calibration (e.g. Anderson (2005)), to determine whether any exist with the capability to match or exceed CSHD's current performance within a reasonable execution time.

If other more suitable methods could not be found then the largest remaining question would be finding the exact cause of the spurious CSHD peaks. While we feel that we have made reasoned and supported deductions, we can provide no incontrovertible proof. From the project point of view, the simulations have shown when spurious peaks are likely to appear and demonstrated that appropriate calibration will prevent them, an outcome that is sufficient to achieve the desired goals. However, from an intellectual

point of view it would be interesting to further explore the matter. Discovering (or confirming) the precise cause could possibly provide the insights required to rectify the problem algorithmically (e.g. the non-negative constraints introduced to prevent other forms of spurious peak), leading to a more calibration independent approach which would be a highly desirable extension for this and other applications.

In conclusion, future work for this portion of the thesis largely depends on how CSHD performs once additional maxillofacial image data is acquired. For the time being it appears that CSHD should be sufficient to meet our immediate needs; though if once data quality improves this assertion proves incorrect, similar experimentation will likely be repeated for a wider range of methods in order to determine, and if necessary help adapt, other methods more suited to our aims.

7.1.3 Automated Segmentation of Whole Volume Tractography

With Chapter 6 we have demonstrated methods that, at least for white matter, can be shown to provide rapid, accurate segmentation of targeted streamlines from a whole volume corpus. With respect to the overall project aims, the obvious drawback is that this ability has not been demonstrated within maxillofacial tissues and thus the simplest avenue for future work would be to (when data is available) repeat the experiments across a suitable maxillofacial image corpus.

Expanding away from the maxillofacial region, there are multiple avenues for future work based on this technology. First amongst the possibilities is to conduct a wide scale repeatability experiment. As an example, CUBRIC is involved in the Avon Longitudinal Study of Parents and Children (ALSPAC) project which, annually, collects DW-MRI datasets covering a 250 subject cohort. Current processing is generally manual (with segmentation of tractography achieved using previously discussed way-point methods), presenting an ideal opportunity for a large scale statistical comparison between metrics derived from manual and automated segmentation procedures and, if these metrics are sufficiently similar, automated processing of large cohorts with a view to publication of any significant results (both with respect to the segmentation method and the retrieved data).

Away from large scale segmentation, it may also be interesting to study the ability of shape space methods to perform more delicate subdivisions of larger white matter structures. While the support vector machine based segmentations appeared promising in this respect, detailed analysis was not conducted as part of demonstrated work. Provided with more time, we believe that analysis of this aspect of the proposed method would provide interesting results.

While such studies may appear to be the most obvious options for expansion upon this work, there are other more esoteric applications that may also be considered. A good

example would be as an aid to graph theoretical research (Sporns et al., Bullmore and Sporns (2009)). Firstly, the number of streamlines connecting a pair of nodes (or some variation upon that metric, e.g. connection density) represents common source of edge weighting within graph theory publications (e.g. Caeyenberghs et al. (2012)). A common approach to retrieving these measures is to warp a white matter parcellation atlas to the native image space then simply count the number of streamlines that pass through, or end within, the two node regions of interest. The problem, however, is that this often requires manual intervention to determine whether a given streamline connecting the two nodes follows the correct path (i.e. is anatomically correct) or whether the perceived connection is artefactual and should be discarded. It should be easily apparent from this description that a shape based streamline filter could easily replace this manual task and thus reduce the user interaction/time required to generate reliable graph theoretical statistics.

A more adventurous adaptation, however, would be the reconstruction of missing graph links. This is to say that across a corpus of datasets there often exists a set of "transient" links between nodes – present and strong in one dataset, absent in the next (an example might be Meyer's loop, a part of the optic radiation that is known to be difficult to reconstruct through DT-MRI Miller (2005), Yamamoto et al. (2005)). Where these transient links are known to be anatomically correct, there may be a window for application of a shape based method to reconstruct the absent streamlines (which for example may have terminated prematurely due to length of the path, its curvature or a number of other imaging related artefacts). A naive example of this process might proceed as follows:

1. Select streamlines which complete the target link.

2. Using selected streamlines as a template, create a shape model.

3. Project the parcellation map onto a dataset with missing link.

4. Overlay the shape model onto the new dataset and simulate an array of likely streamline trajectories.

5. Using the region growing algorithm, recover streamlines that (for the duration of their length) follow the expected path provided by the simulated streamlines.

6. Amalgamate the recovered streamline portions into a usable edge weight, for example mean streamline density along the prescribed trajectory.

While this is unlikely to work exactly as prescribed, such reconstructions would serve as an ideal test bed for more advanced implementations of the shape space paradigm. For example, given that there is a finite amount of room within the brain, it is not unreasonable to expect that alterations in the brains shape, or alterations in the shape one

white matter component (e.g. the shape/size of the corpus callosum), might correlate with alterations of the shape of all white matter components. Should this assumption hold, it would not be unreasonable to expect that by knowing the coordinate of one particular structure within a shape space, we could alter the expected coordinates for other structures of interest according to a known correlation, and thus provide more accurate segmentation results (reducing false positives by eliminating variations unlikely to exist within the current dataset). With respect to the reconstruction of missing streamlines this could mean that, for example, if we knew the shape of adjacent structures we could confine the range of expected shapes that might join the nodes of interest and thus reduce the probability of false positive inclusions.

In conclusion then, while this subsection does not represent all possible avenues for future shape based work, it should be apparent from this short list that there are numerous ways in which the methods presented in Chapter 6 could be expanded, with relevance to both this project and DW-MRI in general.

7.2 Maxillofacial Imaging

As it stands, Chapter 5 represents state of the art in reconstruction of maxillofacial musculature in vivo, however, as discussed within that chapter, there are numerous technological and procedural limitations that are currently hindering progress along this front. In the remainder of this chapter we will therefore briefly discuss each point of difficulty and steps that could be undertaken in order to improve results and advance progress towards true subject-specific maxillofacial modelling.

7.2.1 RF Coils

It is common knowledge that a good signal to noise ratio is one of the defining characteristics for high quality DW-MRI signal data. If anything, Chapter 5 is a testament to this fact, made particularly relevant by demonstrations of the proportionately greater consequences of poor SNR on low anisotropy tissue (i.e. muscle). Along with problematic relaxation properties (discussed shortly) one of the major causes of poor SNR in our images thus far has been inappropriate RF coil hardware selection.

If we re-examine the slice-wise SNR traces provided in Figure 5.9e (reproduced here as Figure 7.1), there is a definite trend towards an improved SNR as we proceed upwards through the image, starting from particularly poor SNRs at the lower extremes ($\approx 10 : 1$, slices 10-20) to nearing the required SNRs (according to simulation results) of $\geq 50 : 1$ at the upper boundaries. The reason for this is simple, the 8 channel brain coil used in this work was designed for maximum sensitivity within the brain and, due to its "bucket" construction, could not be sufficiently lowered over the subjects head to provide

adequate sensitivity across the majority of the maxillofacial region. What Figue 7.1 tells us, however, is that where RF sensitivity is adequate (i.e. the upper slices), there is sufficient signal being produced to begin to provide accurate estimations of muscle fibre trajectory in vivo; thus, if sensitivity could be extended to the whole maxillofacial region, it stands to reason that the acquired images would be of significantly higher quality than those achieved with with the 8 channel brain coil.
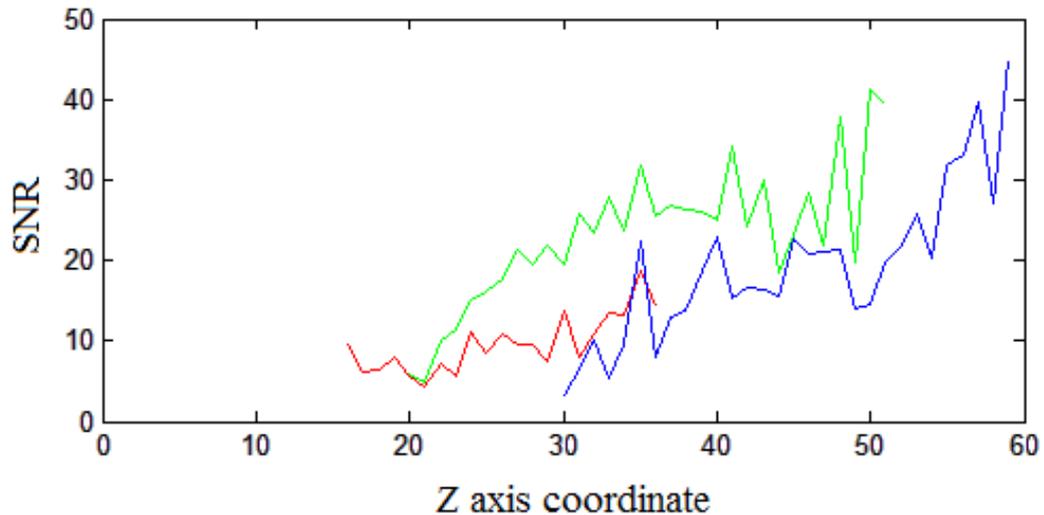
FIGURE 7.1: Reproduction of Chapter 5, Figure 5.9e. A plot of mean SNR relative to slice position for three key muscles: the genioglossus (red), masseter (green) and temporalis (blue). Note that as one increases the Z axis coordinate (i.e. moving upwards through the image), SNR rises sharply, providing further evidence for signal roll off.

While not operable over the time-line of this project, a purpose designed maxillofacial RF coil is now in our possession (Fig 7.2). Briefly, this new RF coil is held within a flexible mount designed to be positioned over the subjects face and, via manipulation of threaded rods, deform so that it closely follows the subjects skin surface – minimising the distance between each of the 8 individual coils and their target maxillofacial region (maximising sensitivity and thus SNR). Unfortunately, while we are technically able to acquire images using this coil, complications in its configuration (particularly acceleration through parallel imaging) are hindering its full adoption. Once this problem is corrected however, significant improvements in image quality are expected.

7.2.2 Magnetic Susceptibility

Magnetic susceptibility artefacts are another key contributor to poor image quality within the maxillofacial region. As previously discussed, the abundance of paramagnetically inhomogeneous organic (muscle/bone/skin) and inorganic (metal fillings, air, etc.) materials within the maxillofacial region create significant inhomogeneities within the B0 field and, by extension, distortions in the resultant image. In Chapter 5, the
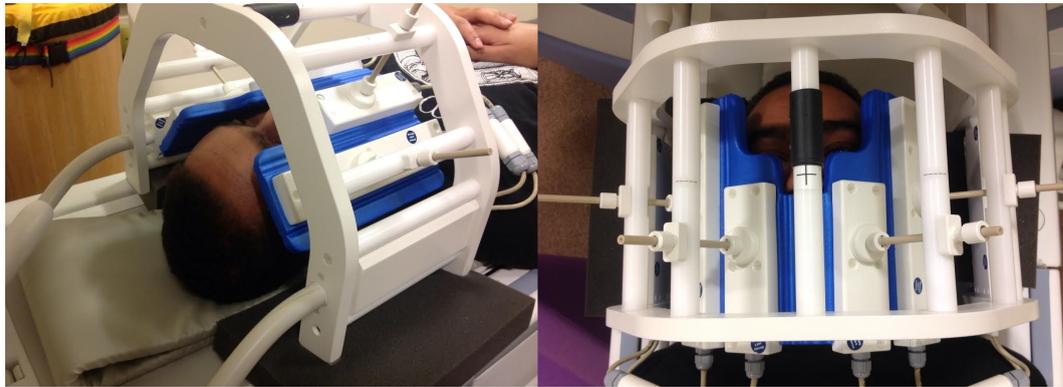
FIGURE 7.2: Photographs of the new 8 channel maxillofacial imaging coil.

procedure used to correct for these artefacts was to acquire a phase map with each DW-MRI acquisition and then, using software tools (Jezzard and Balaban (1995)), un-warp the diffusion weighted image. Unfortunately, this method was not entirely effective and, as such, distortion was still encountered across the lips, cheeks, sinuses and left side of the mandible in general. In this subsection we will therefore lay out methods that should alleviate this issue within future acquisitions.

Without altering acquisition parameters or collecting additional data, the only way to eliminate the susceptibility artefacts would be to eliminate the inhomogeneities that cause them. In keeping with this line of thought, a number of ex vivo works (Benveniste et al., D'Arceuil et al. (2007), D'Arceuil and de Crespigny (2007)) have highlighted one possible avenue for exploration. In these publications, ex vivo tissues are surrounded with a substance known as Fomblin LC/8 (Solvay Solexis Inc.) – a perfluorocarbon grease. An interesting property of Fomblin is that its paramagnetic susceptibility is close to various anatomical tissues while, due to its perfluorocarbon structure, there are no excitable protons (at the specified RF frequencies) to emit a measurable NMR echo. In practice then, the perflourocarbon grease acts as a barrier, simultaneously "pushing" the interface between inhomogeneous materials away from tissue surfaces and ensuring that, at the point of that inhomogeneity, no signals will be emitted above the level of normal background noise. With this in mind, we have acquired sufficient quantities of an equivalent substance, Fluorinert FC-70 (3M, St. Paul, Minnesota), for use in future imaging experiments (note Fluorinert has also seen ex vivo appliction similar to the use of Fomblin, Biton et al. (2006)). While ethical concerns prevent application of this substance in vivo (we could not, for example, fill the sinuses with FC-70 - despite its non-toxicity), it could be used to provide an external barrier. As an example, contained within a sufficiently thin bag the FC-70 liquid could be placed over the surfaces of the lips and cheeks and thus displace the inhomogeneous air/tissue interfaces away from target musculature – hopefully resulting in a significant reduction in the inhomogeneity artefacts within the associated regions.

Owing to the ethical restrictions, however, significant inhomogeneities may still remain following physical compensation, e.g. surrounding the previously mentioned sinuses or near large metal dental work. With this in mind, other *acquisition-side* methods must also be considered in order to provide adequate compensation for these artefacts. One of the more promising methods is to acquire an additional image with reversed k-space trajectory (Bowtell et al. (1994), Andersson et al. (2003)). Crudely speaking, images with opposing k-space trajectories will experience susceptibility induced geometric distortions of equal magnitude, but opposing directionality along the phase encode imaging axis. Combining these two images with a model for spin-echo EPI image formation would then allow us to estimate the underlying magnetic field and, by extension, an undistorted DW-MRI image. Unfortunately, while reverse k-space has shown promise in other works (and has been recommended in personal conversations with Dr Martijn Froeling, a researcher with numerous muscle-imaging publications, e.g. Froeling et al. (2010), Froeling et al. (2012)), limitations of the current imaging hardware (or more specifically, its current firmware version) are preventing its adoption. Pending the availability of new scanner hardware/firmware, other methods must be explored and, for the time being, appropriate responses to the susceptibility problem remain an open avenue of further research.

7.2.3 Relaxation

Returning to signal to noise ratios, a significant contributor to poor SNR in DW-MRI images of muscle tissue is the short (compared to white-matter) T_2 relaxation time inherent to muscle tissue. As discussed in Chapter 5, the easiest way to reduce undesirable T_2 signal attenuation (and thus increase SNR) is to reduce the echo time (T_E) of the acquisition sequence which will then, by extension, reduce the time over which T_2 relaxation is allowed to act. Unfortunately, in the majority of applicable sequences T_E is invariably linked to the duration of the applied diffusion encoding gradients and thus reducing T_E will have the undesirable effect of reducing the applied diffusion weighting. With this is mind, barring the installation of higher-performance gradient coils (which would reduce the gradient duration required for a specific diffusion weighting), very little can be done to improve T_E with the current SS-EPI imaging sequence.

One potentially interesting group of alternatives, however, are diffusion weighted variants (e.g. Rieseberg et al. (2005)) of the stimulated echo acquisition mode paradigm (STEAM, Frahm et al. (1985)). A theoretical advantage of STEAM methods is that their pulse formulation (typically $3 \times 90°$ pulses, compared to the $90°/180°$ combination employed by SS-EPI) allows for shorter effective echo times (since T_2 decay is offset between pulses 2 and 3) while simultaneously, due to a longer sequence, permitting longer application of the diffusion weighting gradients and thus higher b-values. The downside to STEAM acquisitions, however, is that the maximum obtainable signal magnitude is

at most one half of that available through a comparable EPI acquisition – creating its own source of SNR issues. An avenue for future work therefore could be to determine whether, for maxillofacial applications, the gains from a shorter apparent T_E counterbalance the inherent loss of signal. Should this prove to be the case, there are additional benefits to the paradigm that might improve image quality in other ways – key examples being a high degree of resilience to susceptibility induced geometric distortions (Rieseberg and Frahm (2004)) and, in "high speed" variants (Frahm et al. (1991)), motion artefacts.

7.2.4   Motion Correction

As the final component in this discussion, we must now cover the matter of subject motion. While we have already discussed various fast-imaging techniques designed to reduce the effect of subject motion during the pulse/gradient sequence, it is the effects of between-image movements that are potentially most troubling. Unlike the brain, whose motions can for the most part be modelled as that of a rigid body (barring pulsation artefacts which can be eliminated by cardiac gating, subject *movement* will rarely cause significant non-linear distortion of the brain), motion in the maxillfacial region cannot be so simply modelled – for example, the tongue is free to move and deform within the mouth, the mandible free to arcticulate from the skull and the entire head may articulate from the neck, all of which may be happening concurrently in any given image.

This is not to say that non-linear registration tools do not exist within the corpus of DW-MRI processing software. The ExploreDTI processing tool (Leemans et al. (2009)) used in Chapter 5, for example, provides an implementation of Elastix (Klein et al. (2010)) designed to perform this very task. Where the degree of motion remains relatively low, as it is in Chapter 5, such methods do in fact appear to produce acceptable results. Our concern, however, is that these existing methods will not be able to reconcile the frequent motion of patients in significant discomfort that cannot be restrained or reasonably expected to remain motionless (the goal, after all, is to use maxillofacial imaging to aid surgical planning).

As an example, imagine movements of the jaw. The effects of such articulation depends on the position and function of the muscle of interest. If the target muscles lie wholly within the mandible then articulation will simply result in a rotation/translation with magnitude dependant on position relative to the point of articulation (i.e. the temporomandibular joint). If the muscles connect to both the mandible and some external structure (e.g. the skull), however, fibres will experience rotation/translation at the mandibular insertion while remaining comparably stationary at the other, resulting in both a change of fibre trajectory and muscle shape/size as the fibres are stretched/compressed to the new configuration. To provide any accuracy in this situation, a targeted

co-registration algorithm would need to be able to cope with both trivial rotation/trans-lation/scale and complex (anatomical) function driven changes in fibre bundle positions, sizes and orientations. The only way in which this could be realistically achieved is to incorporate some model of the underlying anatomy into the algorithm itself.

In reality, this is likely to require (1) an accurate model for all likely maxillofacial deformations and (2) a method for estimating subject pose from DW-MRI image data. Unfortunately, this is unlikely to be as simple as it might originally sound. Depending on how an image is acquired, there is no guarantee that images of a given slice along different gradient directions, or images of neighbouring slices along the same direction will show any consistency in pose induced deformation or that deformation within a given slice can be easily represented. For example, imagine a slice though a rotated mandible. At the rear of the slice would be the neck, since this does not articulate with the mandible there is no reason to expect that it will have significantly deformed at all provided the skull has not moved. Progressing forwards through the slice, consecutive rows of voxels will correspond to segments of anatomy at different levels within the jaw. The rearward rows might correspond to the anatomy of the lower edge of the mandible (i.e. muscles connecting the mandible and larynx), moving forwards you could then observe the lingual core of the tongue - itself distorted due to some independent motion – followed finally by the upper surface of the tongue and lips (Figure 7.3 attempts to display the described trajectory). With very limited data, such an algorithm would, at minimum, need to (1) determine the position of the head, (2) determine the position of the mandible, (3) determine the position of the tongue, and (4) apply the correct deformation to each voxel, remembering that each intersected muscle would deform in a unique manner according to (2) and (3).

FIGURE 7.3: Cartoon depiction of a slice trajectory through the mandible. Left: Slice captured (green line) as the subject involuntarily opens his/her mouth. Right: Position of the intersected voxels (green line) in a pose-neutral position. Note that as expected the degree of displacement within the mandible depends on the position relative to the axis of rotation, though at similar displacements the unconnected tissues within the neck experience no displacement between pose conditions.

With this in mind, there are two principle avenues for progress. The first task is to deter-mine if and when the previously described co-registrations might be necessary. Recalling Chapter 5, it would appear that below a certain threshold, the described distortions can likely be ignored (i.e. corrected by 'standard' means) with little ill consequence. If we

were able to determine this threshold and, importantly, methods to decide when it has been breached, we can begin the decide on whether our concerns are truly justified – it might simply be that movement thresholds are breached so infrequently that, on the rare occasion it occurs, opting to sedate the subject and capture a repeat image would be a better choice.

Should, however, complex co-registration appear necessary, it is likely to represent a significant body of work. Speculating on avenues for potential exploration, the obvious first step would be to construct a model for plausible maxillofacial deformations. One way to achieve this would be by using a combination of real-time (Uecker et al. (2010)) or standard structural images and CT or black-bone MRI (Eley et al. (2012)) to capture the position of various muscular and skeletal structures across a range of plausible poses. Combining bone segmentation and the output of a point or contour tracking technique (e.g. Dias and Feurie (2007)) applied to the soft-tissues, one could then model the displacement/distortion of key samples of muscle tissue with respect to rigidly articulated skeletal reference points which, if repeated for a sufficient number of points and across sufficient number of subjects, could then provide the data required to build a useful deformation model. With the deformation model constructed, one would then need a co-registration algorithm capable of integrating and applying that information. How this might work is unclear but, if it were possible to estimate bone positions from DW-MRI images (e.g. by looking for regions of low signal intensity), then one could approximate the subjects pose and, using the deformation model, warp key voxels to a standard pose-neutral template, ready for use is subsequent maxillofacial modelling.

7.3 Conclusion

In conclusion, the work presented in this thesis comprises both a primer for application of diffusion weighted magnetic resonance imaging to maxillofacial and three consecutive links in an imagined robust signal processing pipeline. While the initial goal of creating a complete subject-specific maxillofacial musculature model was not reached, the work presented in these chapters comprises a series of novel contributions to the field of DW-MRI in general and maxillofacial imaging in particular – the highlights of which are the first anatomically plausible reconstruction of several key muscles of mastication and novel shape-based method for the accurate segmentation of whole volume tractography results.

Appendix A

Additional Combined Labelling and Elimination Results

As stipulated in Chapter 3 Section 3.3.5, this section of the appendix contains additional results for the combined labelling and elimination algorithms. Each result is listed with a description of the experiment and relevant Section 3.3.4 comparison figure.

A.1 Single Fibre Examples
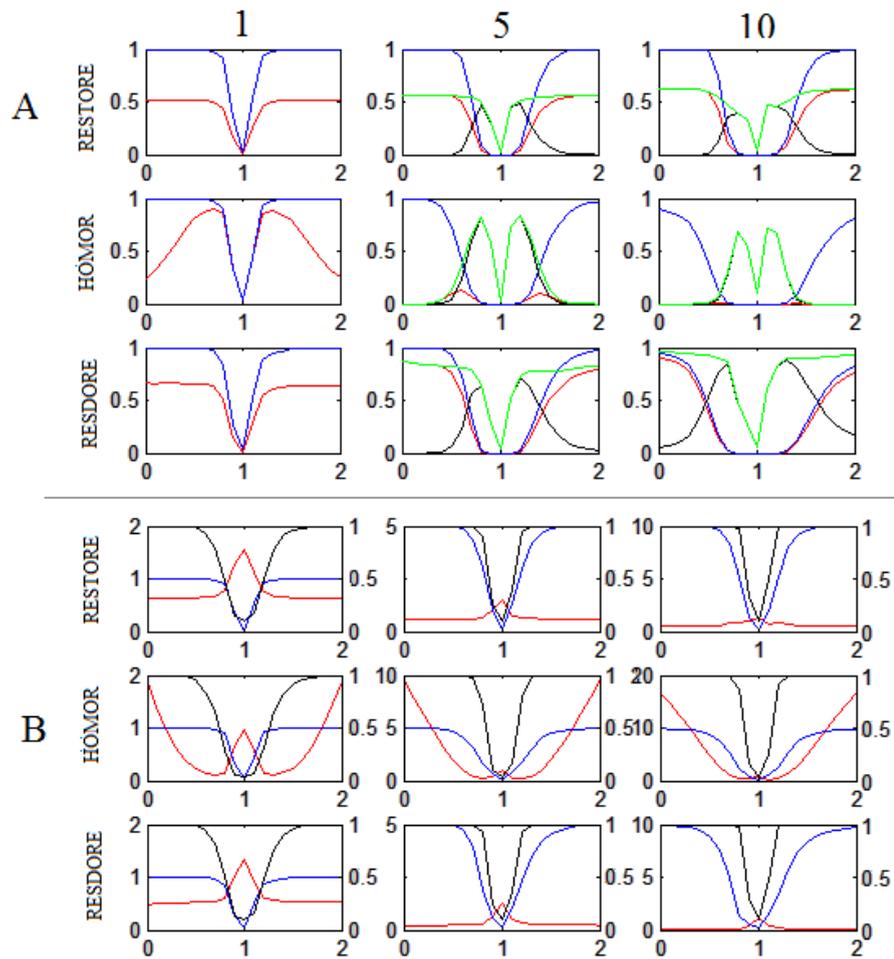
FIGURE A.1: Single fibre, $b = 1000s/mm^2$, $SNR = 30 : 1$, $FA = 0.7$. (A) Detections, see Fig. 3.9a for comparison. (B) False positive and set identification results, see Figures 3.10a and 3.10d.

FIGURE A.2: Single fibre, $b = 3000s/mm^2$, $SNR = 30 : 1$, $FA = 0.7$. (A) Detections, see Fig. 3.9c for comparison. (B) False positive and set identification results, see Figures 3.10c and 3.10f.

A.2 Crossing Fibre Examples: Angular Separation.

FIGURE A.3: Crossing fibre, $b = 2000s/mm^2$, $SNR = 30 : 1$, $FA = 0.7$, 30° intersection. (A) Detections, see Fig. 3.11a for comparison. (B) False positive and set identification results, see Figures 3.12a and 3.12d.

FIGURE A.4: Crossing fibre, $b = 2000s/mm^2$, $SNR = 30 : 1$, $FA = 0.7$, $60°$ intersection. (A) Detections, see Fig. 3.11b for comparison. (B) False positive and set identification results, see Figures 3.12b and 3.12e.

A.3 Crossing Fibre Examples: Component Anisotropy

FIGURE A.5: Crossing fibre, $b = 2000s/mm^2$, $SNR = 30 : 1$, $FA = 0.3$, 90° intersection. (A) Detections, see Fig. 3.13a for comparison. (B) False positive and set identification results, see Figures 3.14a and 3.14d.
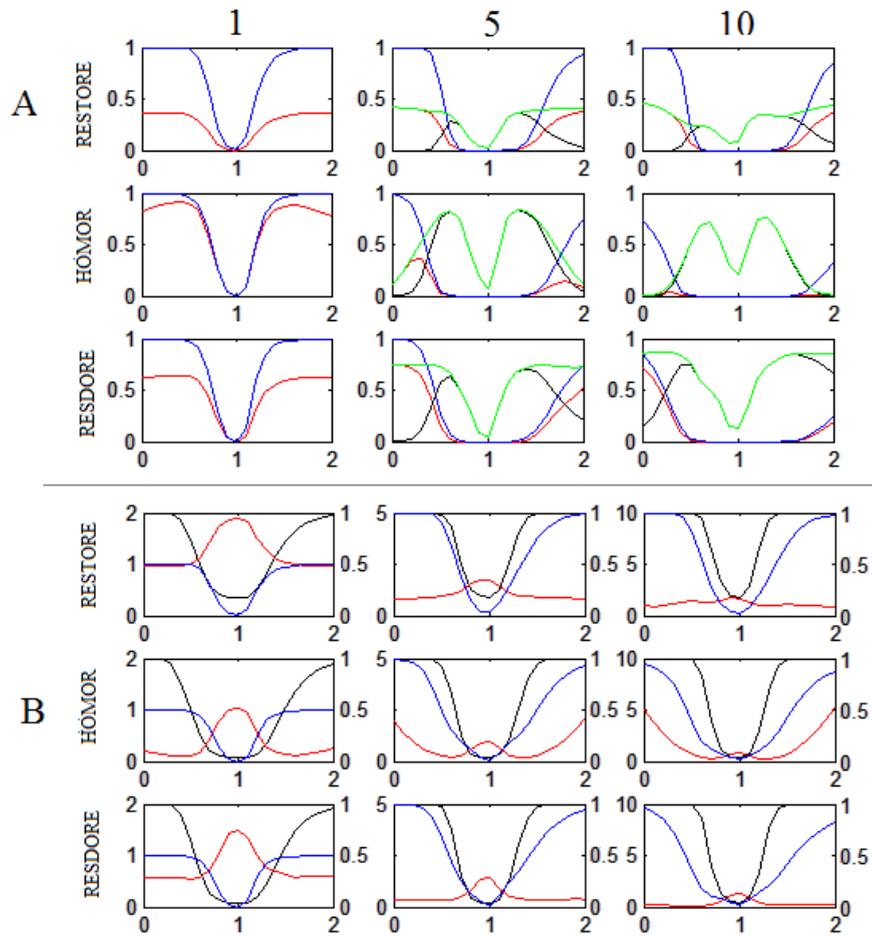
FIGURE A.6: Crossing fibre, $b = 2000s/mm^2$, $SNR = 30 : 1$, $FA = 0.5$, $90°$ inter-section. (A) Detections, see Fig. 3.13b for comparison. (B) False positive and set identification results, see Figures 3.14b and 3.14e.
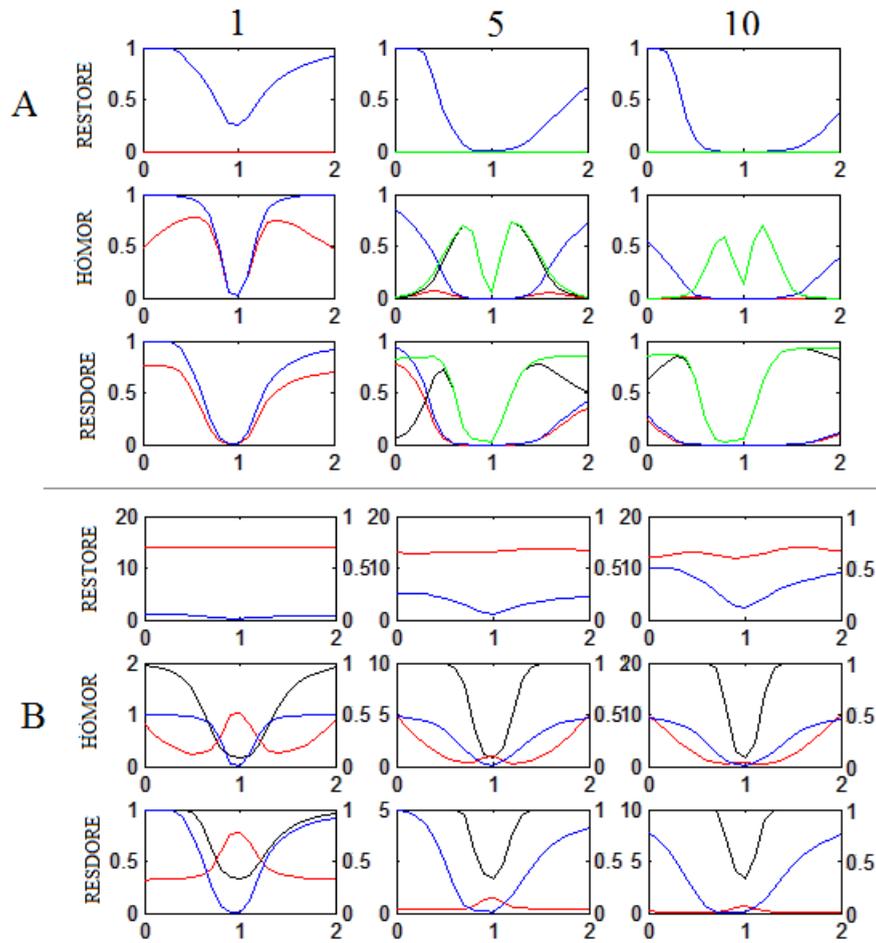
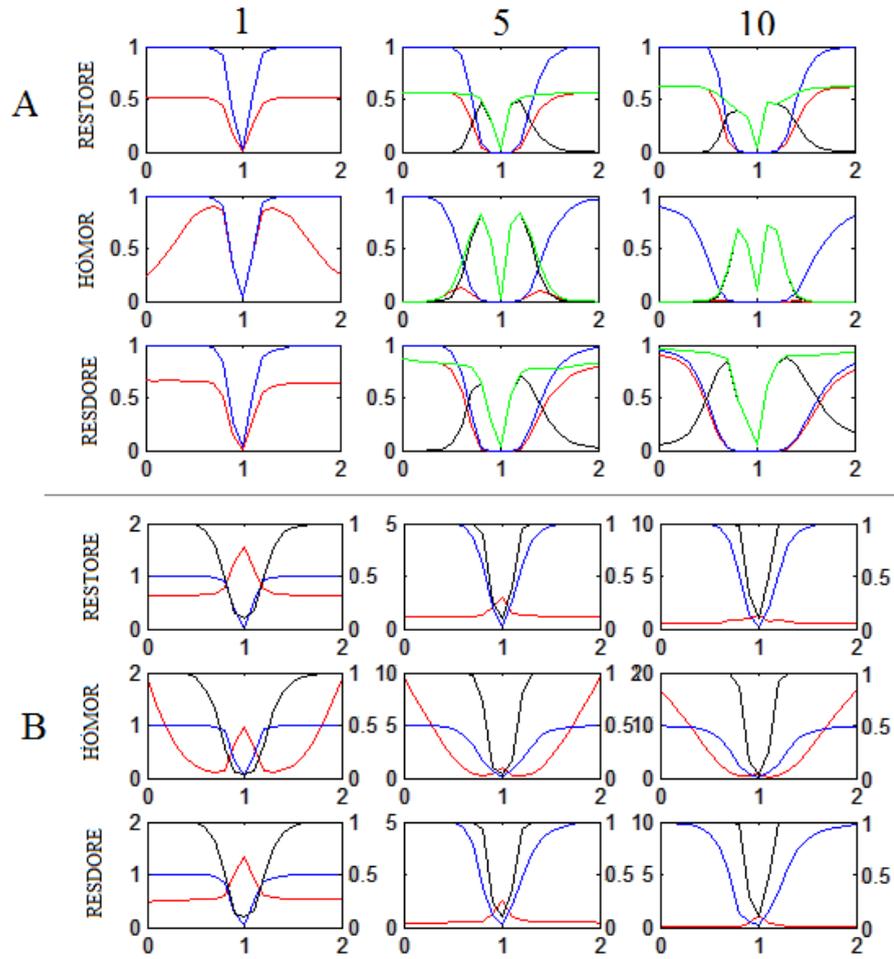FIGURE A.7: Crossing fibre, $b = 2000s/mm^2$, $SNR = 30 : 1$, $FA = 0.9$, $90°$ intersection. (A) Detections, see Fig. 3.13c for comparison. (B) False positive and set identification results, see Figures 3.14c and 3.14f.

A.4 Crossing Fibre Examples: SNR
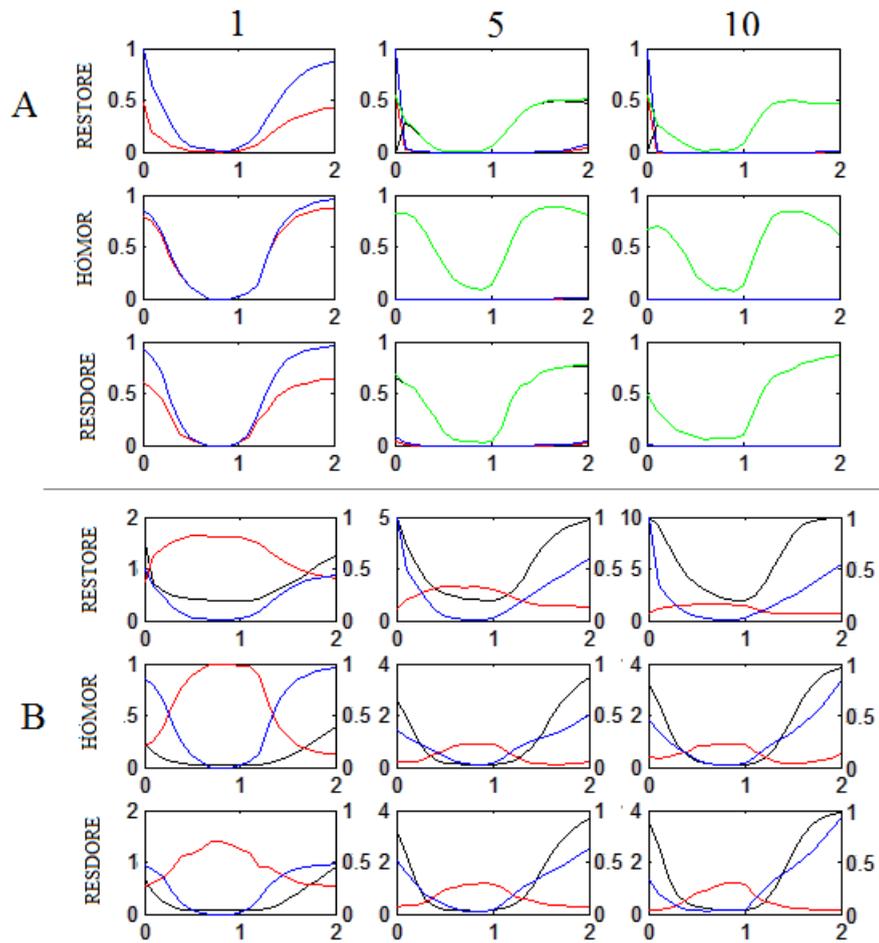
FIGURE A.8: Crossing fibre, $b = 2000s/mm^2$, $SNR = 10 : 1$, $FA = 0.7$, 90° intersection. (A) Detections, see Fig. 3.15c for comparison. (B) False positive and set identification results, see Figures 3.16c and 3.16g.
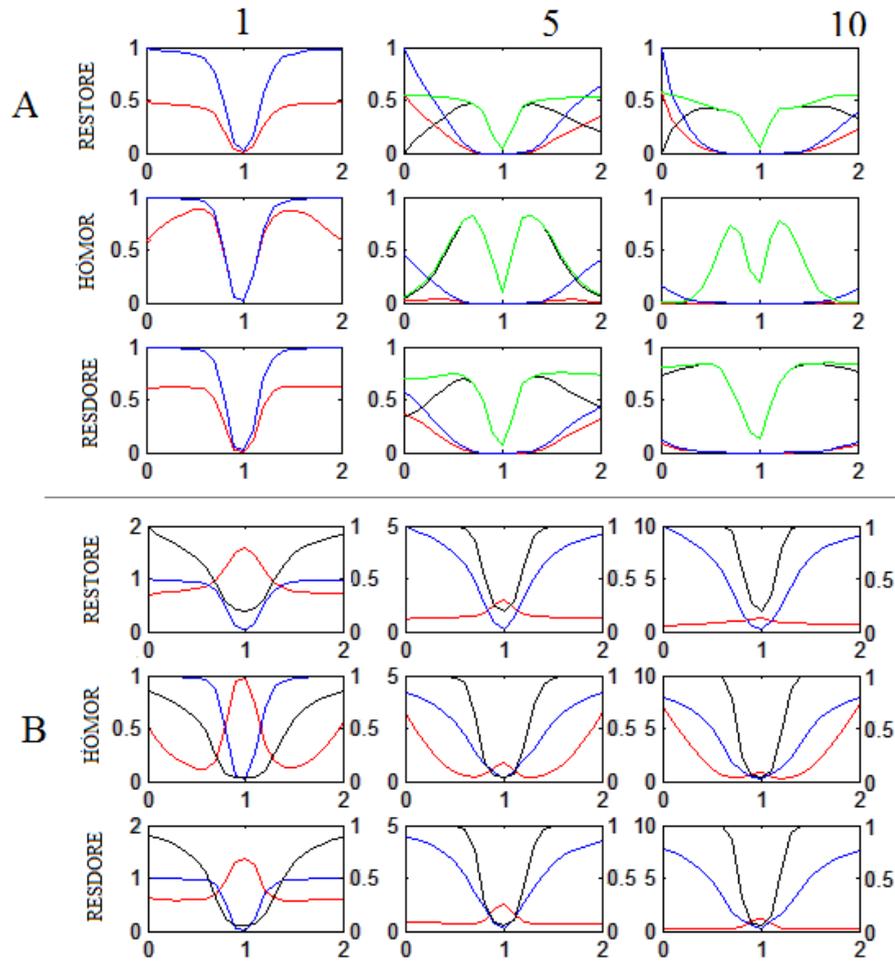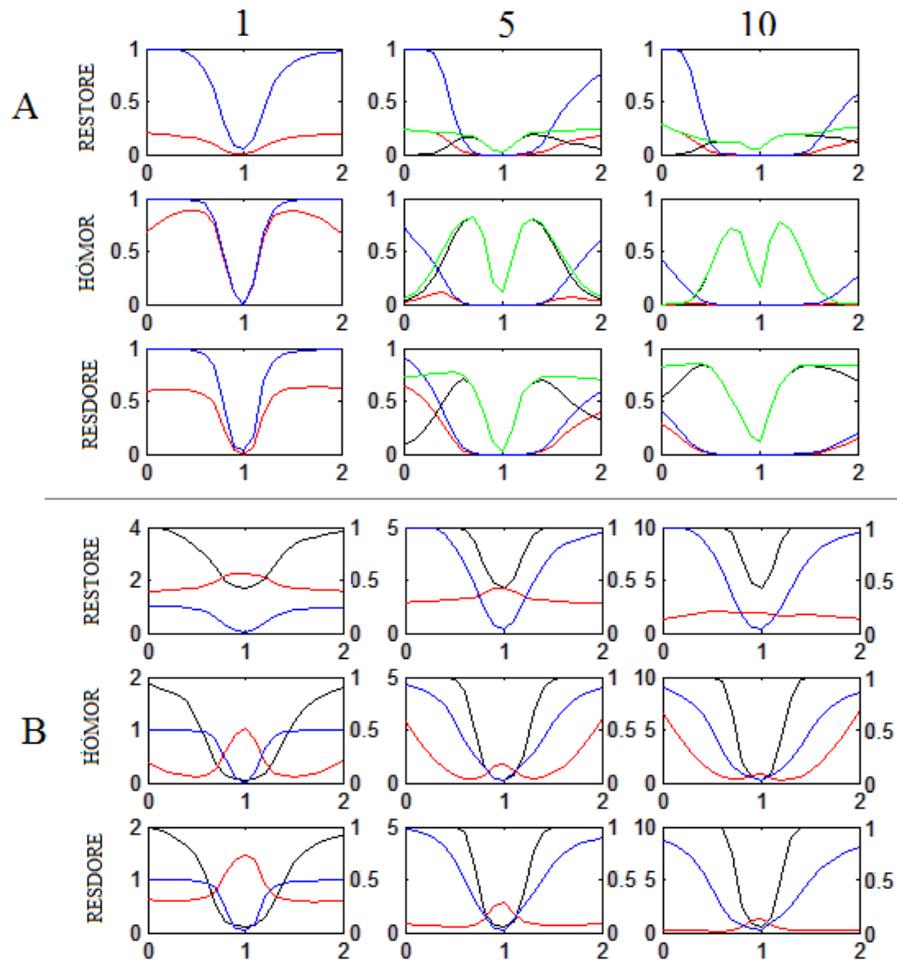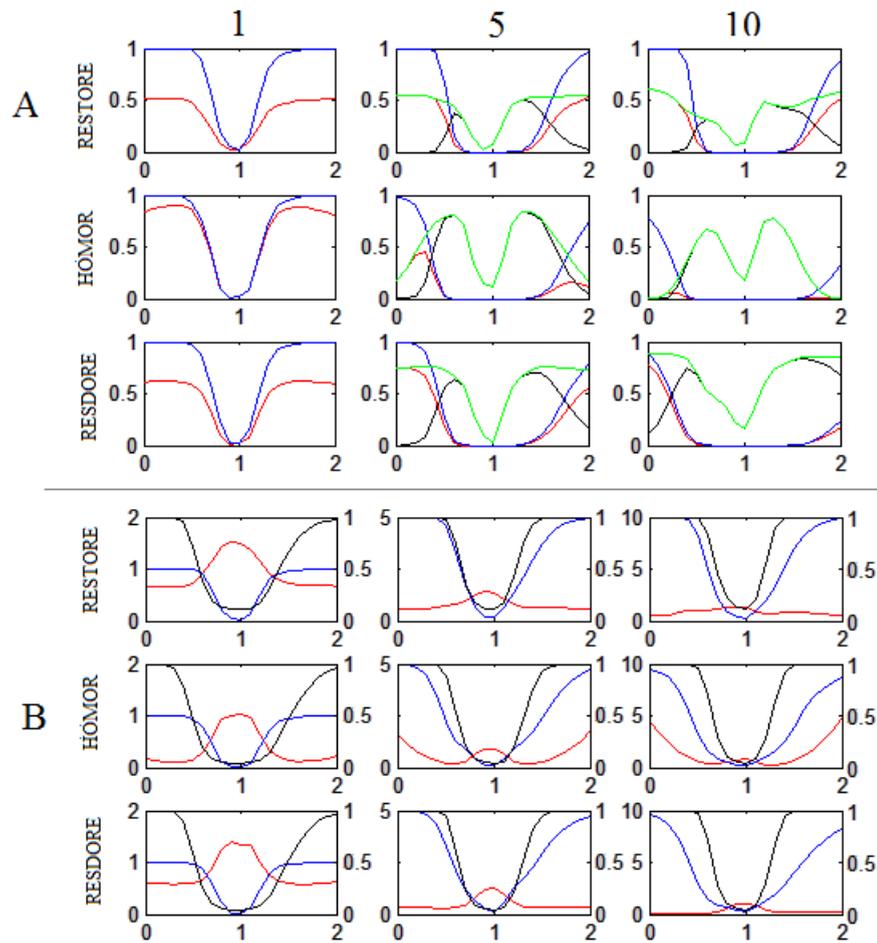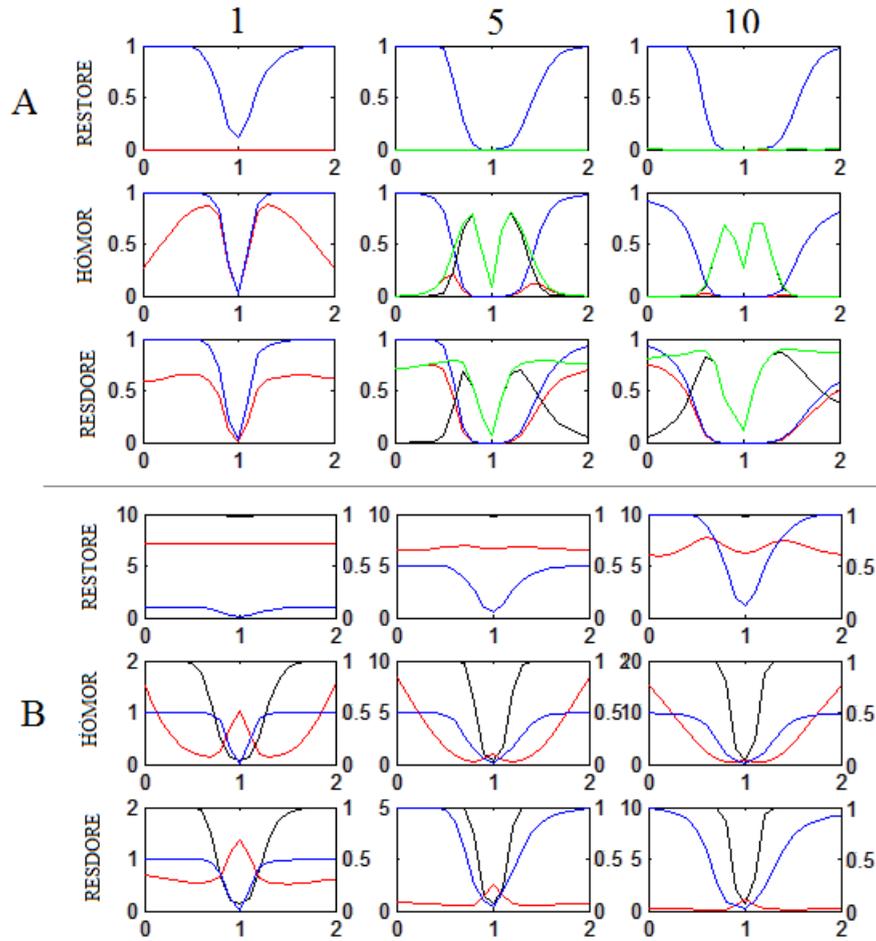
FIGURE A.9: Crossing fibre, $b = 2000s/mm^2$, $SNR = 50 : 1$, $FA = 0.7$, 90° intersection. (A) Detections, see Fig. 3.15d for comparison. (B) False positive and set identification results, see Figures 3.16d and 3.16h.

Appendix B

Additional Image Artefacts

As stipulated in Chapter 3 Section 3.4.3 the 'difficult subject' dataset was predominantly afflicted by three major volume-wide artefacts i.e. those detailed within the chapter - and an additional pair of motion/distortion correction artefacts to be described within this appendix.

Beginning with image 24; along with intensity and translation artefacts affecting the lower slices, there appears to volume wide Y axis rotation (Fig. B.1) that has gone uncorrected by the motion/distortion correction algorithm.
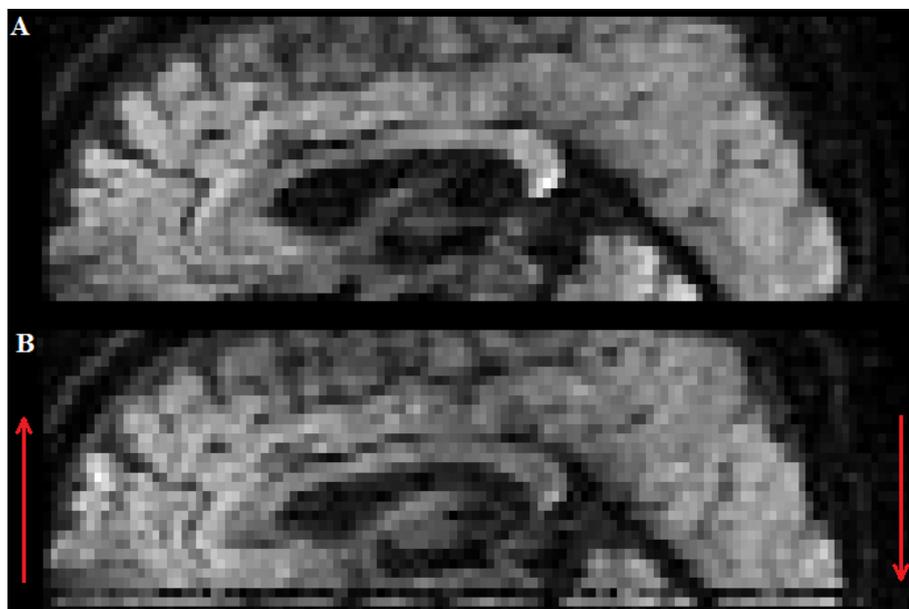
FIGURE B.1: Motion/distortion correction artefact. (A) An example of image correctly aligned with the rest of the volume. (B) Image 24, note in addition to the lower slice artefacts the entire image displays a slight clockwise rotation.

Similarly, if we then examine image 42 (Fig. B.2), we can observe evidence of an uncorrected rotation about the Z axis.
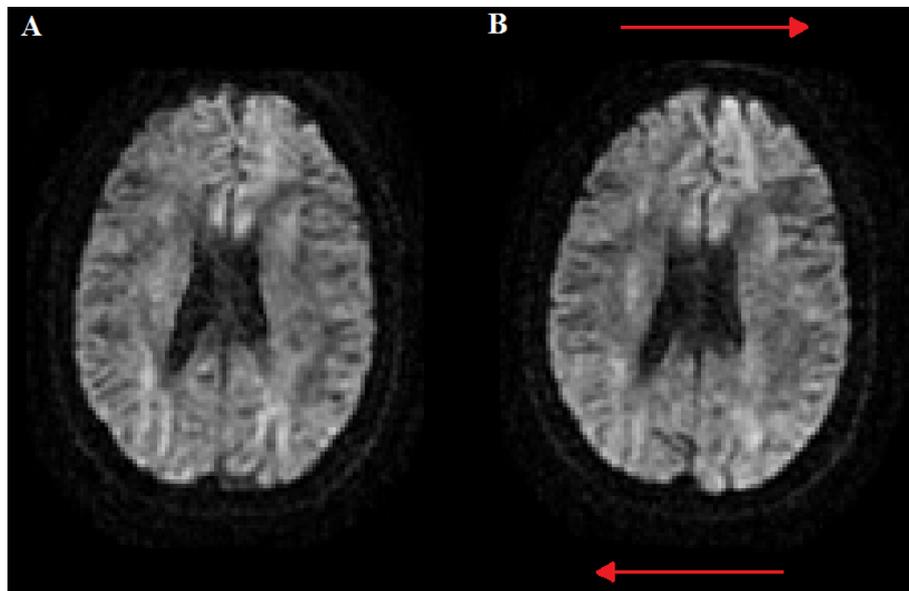
FIGURE B.2: Motion/distortion correction artefact. (A) Exemplar image. (B) Image 42, note the pronounced clockwise rotation about the Z axis.

Appendix C

Additional Calibration Results

As stipulated in Chapter 4, this appendix covers the result of changes in calibration anisotropy over a wider range of crossing fibre angles and volume fractions.
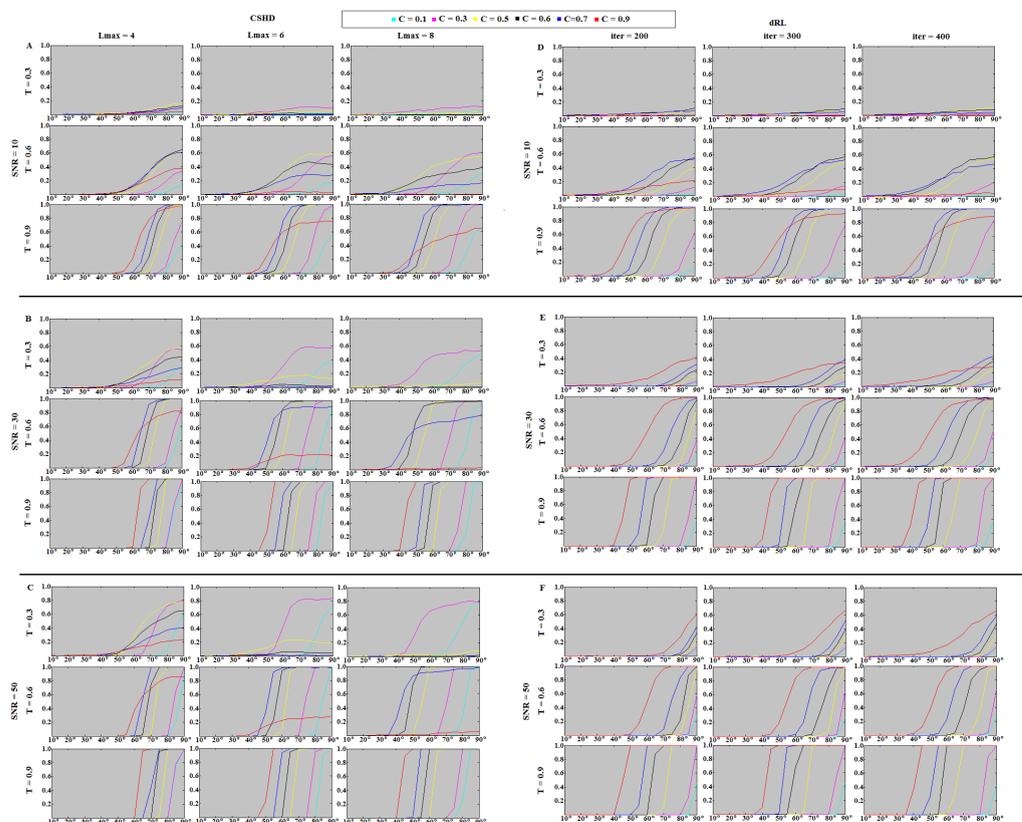
FIGURE C.1: Supplimental data, wider range of angular resolution results.
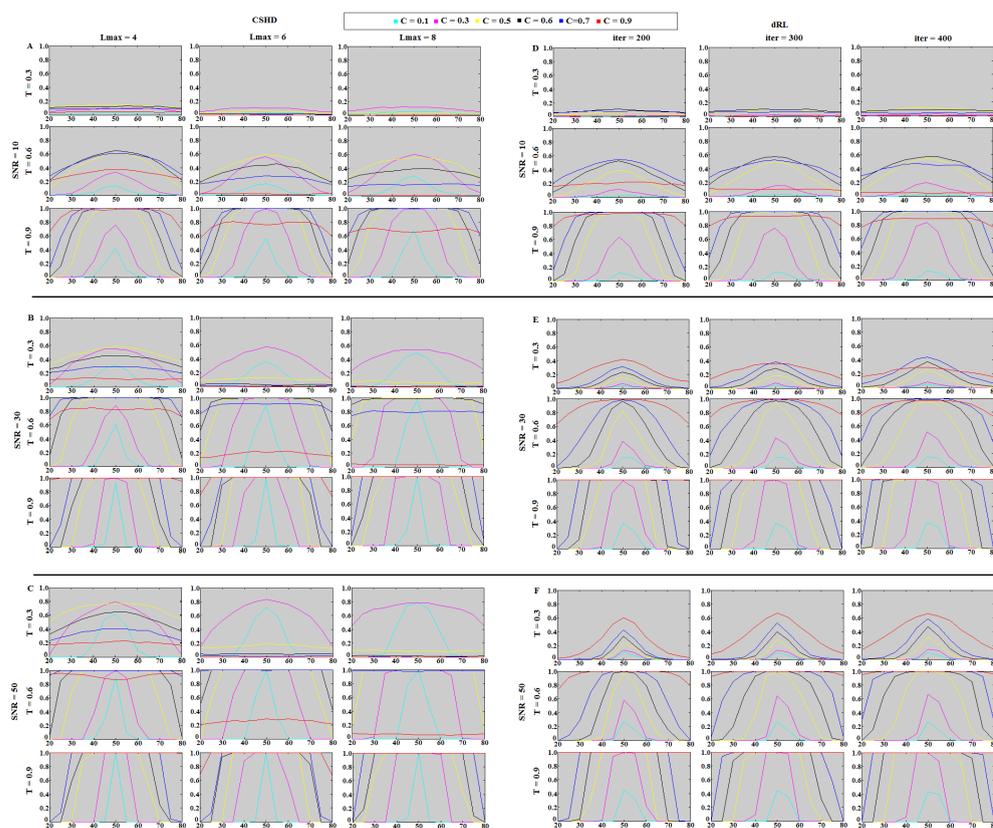
FIGURE C.2: Supplimental data, wider range of volume fraction results.

Appendix D

Additional Shape Space Trajectories

As stipulated in Chapter 6, this appendix provides visualisations of additional trajectories completing a closed loop through the whole volume feature space.
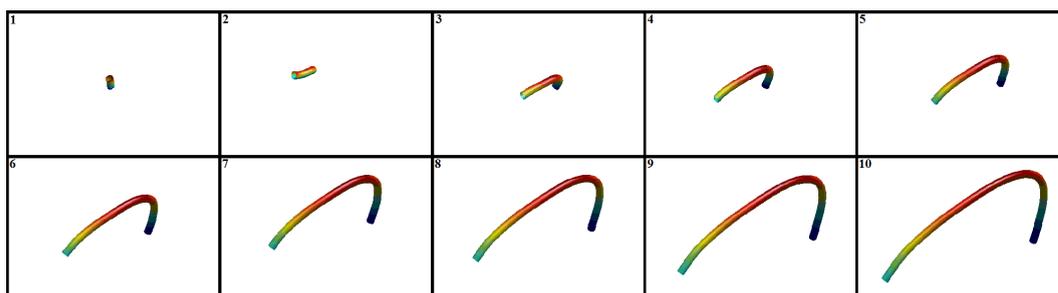
FIGURE D.1: Visualisation of the trajectory between the origin (mean shape) and the left Arcuate.
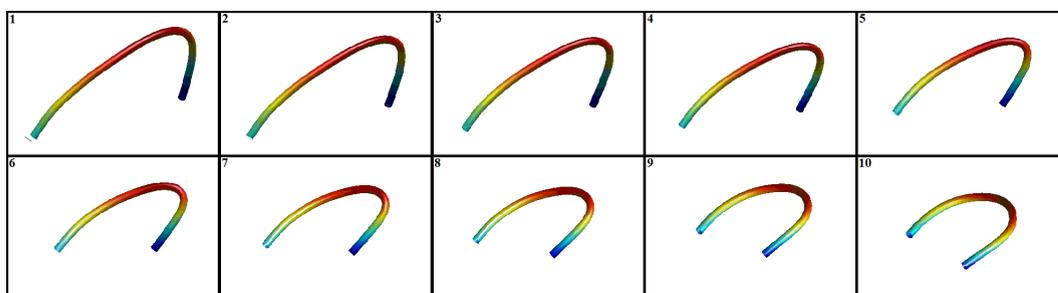
FIGURE D.2: Visualisation of the trajectory between the left Arcuate and the frontal Corpus Callosum.
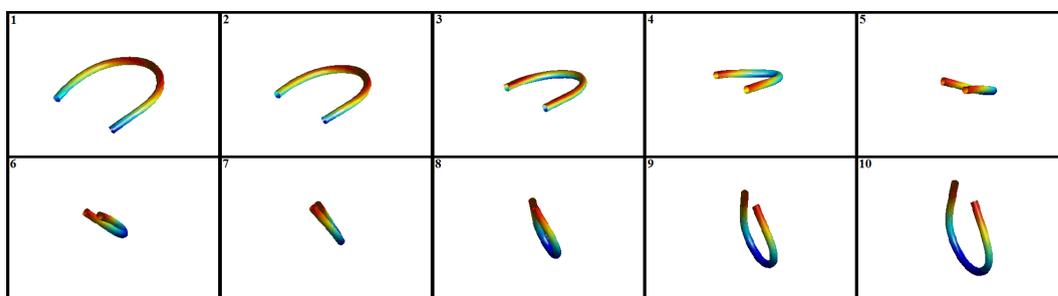
FIGURE D.3: Visualisation of the trajectory between the frontal Corpus Callosum and the medial Corpus Callosum.
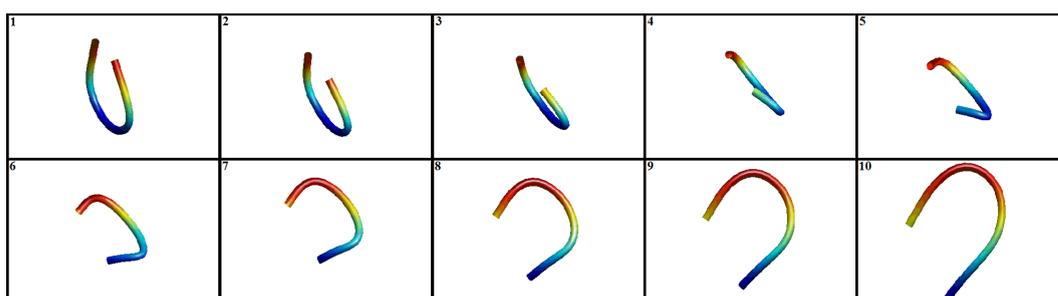
FIGURE D.4: Visualisation of the trajectory between the medial Corpus Callosum and the left Fornix.
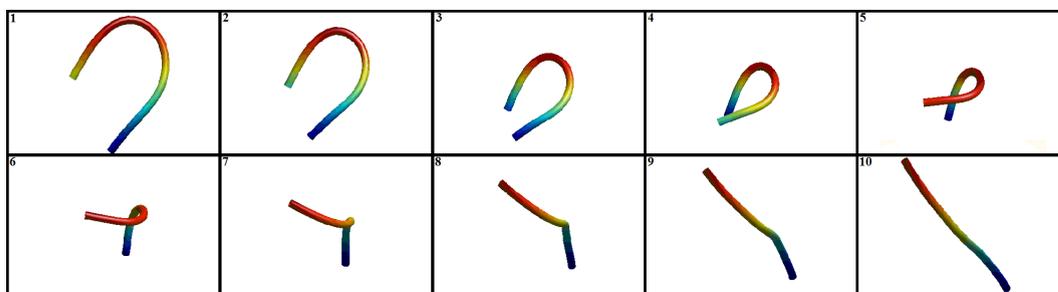
FIGURE D.5: Visualisation of the trajectory between the left Fornix and a Cortico Spinal Tract.
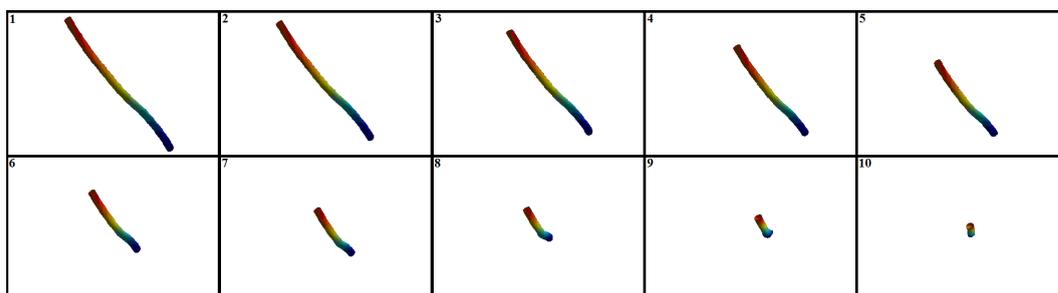
FIGURE D.6: Visualisation of the return trajectory from the Cortico Spinal Tract back to the origin.

Bibliography

F. Abolitiz, A.B. Scheibel, R.S. Fisher, and E. Zaidel. Fiber composition of the human corups callosum. *Brain Res*, 598(2):143–153, 1992.

A.L. Alexander, K.M. Hasan, M. Lazar, J.S. Tsuruda, and D.L. Parker. Analysis of partial volume effects in diffusion-tensor MRI. *Magn. Reson. Med.*, 19:76–87, 2005.

D.C. Alexander. Maximum entropy spherical deconvolution for diffusion MRI. *Image Process. Med. Imaging*, 19:76–87, 2005.

D.C. Alexander, G.J. Baker, and S.R. Arridge. Detection and modelling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magn. Reson. Med.*, 48 (2):331–340, 2002.

D.C. Alexander, P.L. Hubbard, M.G. Hall, M. Ptito, G.J. Parker, and T.B. Dyrby. Orientationally invarient indices of axon diameter and density from diffusion MRI. *NeuroImage*, 52(4):1374–1389, 2010a.

D.C. Alexander, K.K. Seunarine, and D.K. Jones (editor). *'Mathematics of Crossing Fibres' in* Diffusion MRI, Theory, Methods and Applications. Oxford University Press, Oxford, 2010b.

A.W. Anderson. Measurement of Fibre Orientation Distributions Using High Angular Resolution Diffusion Imaging. *Magn. Reson. Med.*, 54:1194–1206, 2005.

A.W. Anderson and J.C. Gore. Analysis and correction of motion artifacts in diffusion weighted imaging. *Magn Reson. Med*, 32:379–387, 1994.

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

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

S.C. Aung, R.C.K. Ngim, and S.T. Lee. Evaluation of the laser scanner as a surface measuring tool and its accuracy compared with direct facial anthropometric measurements. *British Journal of Plastic Surgery*, 48(8):551–558, 1995.

Y. Bai and D.C. Alexander. Model-based registration to correct for motion between acquisitions in diffusion MR imaging. In *IEEE Biomedical Imaging: From Nano to Macro: ISBI 2008*, page 947, 2008.

D. Barazany, P.J. Basser, and Y. Assaf. In vivo measurement of axon diameter distribution in the corpus callosum of rat brain. *Brain*, 132(5):1210–1220, 2009.

P.J. Basser and D.K. Jones. Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. *NMR in Biomedicine*, 15:456–467, 2002.

P.J. Basser, J. Mattiello, and D. LeBihan. Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance*, 103:247–254, 1994a.

P.J. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994b.

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

S. Basu, T. Fletcher, and R. Whitaker. Rician noise removal in diffusion tensor MRI. pages 117–125, 2006.

P.G. Batchelor, F. Calamante, J.-D. Tournier, D. Atkinson, D.L.G. Hill, and A. Connelly. Quantification of the shape of fiber tracts. *Magnetic Resonance in Medicine*, 55(4): 894–903, 2006.

C. Beaulieu. The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR in Biomedicine*, 15:435–455, 2002.

C. Beaulieu and P.S. Allen. Determinants of anisotropy water diffusion in nerves. *Magn. Reson. Med*, 31:394–400, 1994.

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

T.E.J. Behrens, M.W. Woolrich, M. Johansen-Berg, H. Nunes, R.G. Clare, S. Matthews, P.M. Brady, and J.M. Smith. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.*, 50:1077–1088, 2002.

L. Beldie, B. Walker, Y. Lu, S. Richmond, and J. Middleton. Finite element modelling of maxillofacial surgery and facial expressions – a preliminary study. *Int. Journal of Medical Robotics and Computer Assisted Surgery*, 6(4):422–430, 2010.

S. Bells, M. Cerigani, S. Deoni, Y. Assaf, O. Pasternak, C.J. Evans, A. Leemans, and D.K. Jones. "Tractometry": – Comprehensive Multi-Modal Quantitative Assessment of White Matter Along Specific Tracts. page 678, 2011.

H. Benveniste, K. Kim, L. Zhang, and G.A. Johnson. Magnetic resonance microscopy of the C57BL mouse brain. *NeuroImage*, 11:601–611.

M. Berghagen. Photogrammetric Principles Applied to Intra-oral Radiodontia. A method for diagnosis and therapy in odontology. *Springer: Stockholm*, 1951.

J. C. Bezdek, L.O. Hall, and L.P. Clarke. Review of MR image segmentaion technqiues using pattern recognition. *Int. J. of medical physics and research practice*, 20(1):1033, 1993.

D. Le Bihan, C. Poupon, A. Amadon, and F. Lethimonnier. Artifacts and pitfalls in diffusion MRI. *Journal of Magnetic Resonance Imaging*, 24:478–488, 2006.

I.E. Biton, I.D. Duncan, and Y. Cohen. High b-value q-space diffusion MRI in myelin-deficient rat spinal cords. *Magnetic Resonance Imaging*, 24:161–166, 2006.

F. Blais. Review of 20 years of range sensor development. *Journal of Electron Imaging*, 13(1):231–240, 2004.

F. Bloch. Nuclear Induction. *Physics Review*, 70(7):460–474, 1946.

P.H.M. Bovendeerd, J.M.R.J. Huyghe, T. Arts, D.H. van Campen, and R.S. Reneman. Influence of endocardial-epicardial crossover of muscle fibers on left ventricular wall mechanics. *J. Biomech.*, 27(7):942–951, 1994.

R. Bowtell, D.J.O. McIntyre, M.-J. Commandre, P.M. Glover, and P. Mansfield. Correction of geometric distortion in echo planar imaging. In *In Proc. 2nd Meeting of the Society of Magnetic Resonance*, page 411, 1994.

A. Brun, H. Knutsson, H.-J. Park, M.E. Shenton, and C.-F. Westin. Clustering fiber traces using normalised cuts. In *In Proc. MICCAI 2004*, pages 368–375, 2004.

E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10:186–198, 2009.

P.H Burke and F.H. Beard. Stereophotogrammetry of the face. A preliminary investigation into the accuracy of a simplified system evolved for contour mapping by photography. *American Journal of Orthodontics*, 51(10):769–782, 1967.

D. Leibfritz C. Meier, W. Dreher. Diffusion in compartmental systems: II. Diffusion weighted measurements of rat brain tissue in vivo and postmortem at very large b values. *Magn. Reson. Med.*, 50:510–514, 2003.

K. Caeyenberghs, A. Leemans, C. De Decker, M. Heitger, D. Drijkoningen, C. Vander Linder, S. Sunaert, and S.P. Swinnen. Brain connectivity and postural control in young traumatic brain injury patients: A diffusion MRI based network analysis. *NeuroImage: Clinical*, 1:106–115, 2012.

P. Callaghan, C. Eccles, and Y. Xia. NMR microscopy of dynamic displacements: k-space and q-space imaging. *Journal of Physics E: Scientific Instruments*, 21(8):820–822, 1988.

J. Canny. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

H.Y. Carr and E.M. Purcell. Effects of diffusion on free precession in magnetic resonance experiments. *Physical Review*, 94:630–638, 1954.

A.J. Caruso, S.J. Stanhope, and D.A. McGuire. New technique for acquiring three-dimensional orofacial nonspeech movements. *Dysphagia*, 4(2):127–132, 1989.

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

M. Chabanas, V. Luboz, and Y. Payan. Patient specific finite element model of the face soft tissues for computer assisted maxillofacial surgery. *Medical Image Analysis*, 7(2):131–151, 2003.

M. Chabanas, C. Marecaux, and Y. Payan. Models for planning and simulation in computer assisted orthognatic surgery. In *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention, Tokyo, Japan*, pages 315–322, 2002.

M. Chabanas and Y. Payan. A 3D finite element model of the face for simulation in plastic and maxillo-facial surgery. In *Proceedings of the 3rd International Conference on Medical Image Computing and Computer Assisted Interven tion, Pittsburgh, PA, USA*, pages 1068–1075, 2000.

C.C. Chang and C.J. Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent systems and technology*, 2(3):27, 2011.

L.C. Chang, D.K. Jones, and C. Pierpaoli. RESTORE: Robust Estimation of Tensors by Outlier Rejection. *Magn Reson Med*, 53:1088–1095, 2005.

L.C. Chang, L. Walker, and C. Pierpaoli. Making the robust tensor estimation approach "RESTORE" more robust. page 3557, 2009.

L.C. Chang, L. Walker, and C. Pierpaoli. Informed RESTORE: A method for robust estimation of diffusion tensor from low redundancy datasets in the presence of physiological noise artifacts. *Magn Reson Med*, 68:1654 – 1663, 2012.

T.L. Chenevert, J.A. Brunberg, and J.G. Pipe. Anisotropic diffusion in human white matter: demonstration with MR techniques in vivo. *Radiology*, 177(2):401–405, 1990.

L. P. Clarke, R.P. Velthuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, and M.L. Silbiger. MRI segmentation: Methods and applications. *Magnetic Resonance Imaging*, 13(3):343–368, 1995.

J.B. Colby, L. Soderberg, C. Lebel, I.D. Dinov, P.M. Thompson, and E.R. Sowell. Along-tract statistics allow for enhanced tractography analysis. *NeuroImage*, 59(4):3227–3242, 2012.

C.M. Collins, B. Yang, Q.X. Yang, and M.B. Smith. Neumerical calculations of the static magnetic field in three-dimensional multi-tissue models of the human head. *Magnetic Resonance Imaging*, 20:413–424, 2002.

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. *National Academy of Sciences*, 96(18):10422–10427, 1999.

P.A. Cook, Y. Bai, S. Nedjati-Gilani, K.K. Seunarine, M.G. Hall, G.J. Parker, and D.C. Alexander. Camino: Open-Source Diffusion-MRI Reconstruction and Processing. 2006.

P.A. Cook, M. Symms, P.A. Boulby, and D.C. Alexander. Optimal acquisition orders of diffusion-weighted MRI measurements. *Journal of Magn. Reson. Imaging*, 25(5):1051–1058, 2007.

T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

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

B.M. Damon, Z. Ding, A.W. Anderson, A.S. Freyer, and J.C. Gore. Validation of diffusion tensor MRI -based muscle fibre tracking. *Magnetic Resonance in Medicine*, 48:97–104, 2002.

H. D'Arceuil and A. de Crespigny. The effects of brain tissue decomposition on diffusion tensor imaging and tractography. *NeuroImage*, 36:64–68, 2007.

H. D'Arceuil, S. Westmoreland, and A. de Cespigny. An approach to high resolution diffusion tensor imaging in fixed primate brain. *NeuroImage*, 35:553–565, 2007.

R.L. DeLaPaz. Echo-planar Imaging. *Radiographics*, 14(5):1045–1058, 1994.

F. Dell'Acqua, G. Rizzo, P. Scifo, R.A. Clarke, G. Scotti, and F. Fazio. A Model-Based Deconvolution Approach to Solve Fiber Crossing in Diffusion-Weighted MR Imaging. *IEEE TMI*, 54(3):462–472, 2007.

F. Dell'Acqua, P. Scifo, G. Rizzo, M. Catani, A. Simmons, G. Scotti, and F. Fazio. A modified damped Richardson-Lucy algorithm to reduce isotropic background effects in spherical deconvolution. *NeuroImage*, 49(2):1446–1458, 2010.

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

O. Dietrich, S. Heiland, and K. Sartor. Noise correction for the exact determination of apparent diffusion coefficients at low SNR. *Magn. Reson. Med.*, 45:448–453, 2001.

O. Dietrich, J.G. Raya, S.B. Reeder, M.F. Reiser, and S.O. Schoenberg. Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. *Journal of Magnetic Resonance Imaging*, 26(2):375–385, 2007.

Z. Ding, J.C. Gore, and A.W. Anderson. Classification and Quantification of Neuronal Fiber Pathways Using Diffusion Tensor MRI. *Magn. Reson. Med.*, 49:716–721, 2003.

M.P. do Carmo, editor. Prentice Hall, New Jersey, 1976.

R.R. Edelman, P. Wielopolski, and F. Schmitt. Echo-planar MR imaging. *Radiology*, 192(3):600–612, 1994.

W. Edelstein, A. Bottomley, and L.M. Pfeifer. A signal-to-noise calibration procedure for NMR imaging systems. *Med Phys*, 11:180–185, 1984.

A. Einstein. On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat. *Annalen der Physik*, 17:549–560, 1905.

K.A. Eley, A.G. McIntyre, S.R. Watt-Smith, and S.J. Golding. "black bone" MRI: a partial flip angle technique for radiation reduction in craniofacial imaging. *British Journal of Radiology*, 85:272–278, 2012.

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

A. Fick. On liquid diffusion. *Journal of Membrane Science*, 100(1):33–38, 1995.

C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the Nystrom method. *IEEE trans. on pattern analysis*, 26(2):214–225, 2004.

J. Frahm, W. Hanicke, H. Brun, M.L. Gyngell, and K.D. merboldt. High-speed STEAM MRI of the human heart. *Magnetic Resonance in Medicine*, 21(1):133–142, 1991.

J. Frahm, K.D. Merboldt, and W. Hanicke. Stimulated Echo Imaging. *Journal of Magnetic Resonance (1969)*, 64(1):81–93, 1985.

L.R. Frank. Characterization of Anisotropy in High Angular Resolution Diffusion Weighted MRI. *Magn. Reson. Med.*, 47:1083–1099, 2002.

B.J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–976, 2007.

M. Froeling, A.J. Nederveen, D.F.R. Heijtel, A. Lataster, C. Bos, K. Nicolay, M. Maas, M.R. Drost, and G.J. Strijkers. Diffusion-tensor MRI reveals the complex muscle architecture of the human forearm. *Journal of Magnetic Resonance Imaging*, 36: 237–248, 2012.

M. Froeling, J. Oudeman, S. van der Berg, K. Nicolay, M. Maas, G.J. Strijkers, M.R. Drost, and A.J. Nederveen. Reproducibility of Diffusion Tensor Imaging in Human Forearm Muscles at 3.0T in a clinical setting. *Magnetic Resonance in Medicine*, 64: 1182–1190, 2010.

S. Geman and D.E. McClure. Statistical methods for tomographic image reconstruction. *Bull Int Stat Inst*, 52:5–21, 1987.

R.J. Gilbert and V.J. Napadow. Three-dimensional muscular architecture of the human tongue determined in vivo with diffusion magnetic resonance imaging. *Dysphagia*, 20: 1–7, 2005.

R.J. Gilbert, V.J. Wedeen, L.H. Magnusson, T. Benner, R. Wang, G. Dai, V.J. Napadow, and K.K. Roche. Three-dimensional myoarchitecture of the bovine tongue demonstrated by diffusion spectrum magnetic resonance imaging with tractography. *The Anatomical Record Part A.*, 288:1173–1182, 2006.

B. Gilles and N. Magnenat-Thalmann. Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations. *Medical Image Analysis*, 14: 291–302, 2010.

E. Gladilin and A. Roginsky. A framework for biomechanical simulation of cranio-maxillofacial surgery interventions. In *Proc. Intl. Symposium on Medical Simulation*, pages 287–294, 2004.

R.G. Gonzales, P.W. Schaefer, F.S. Buonanno, L.H. Schwamm, R.F. Budzik, G.R. Rordorf, B. Wang, A.G. Sorensen, and W.J. Koroshetz. Diffusion-weighted MR imaging: Diagnostic accuracy in patients imaged within 6 hours of stoke symptom onset. *Radiology*, 210(1):155–162, 1999.

J.C. Gower. Generalised Procrustes Analysis. *Psychometrika*, 40(1):31–51, 1975.

H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magn Reson Med*, 34:910–914, 1995.

V. Gulani, I.D. Duncan, and P.C. Lauterbur. Apparent diffusion tensor measurement in myeline-deficient rat spinal cords. *Magn. Reson. Med*, 45:191–195, 2001.

E. Hahn. Spin echoes. *Physical Review*, 80(4):580–594, 1950.

G.N. Haunsfield. In *A Method of and Apparatus for Examinations of a Body by Radiation Such as X- or Gamma- Radiation.* UK Patent No.1283915, 1972.

D.M. Healy, H. Hendriks, and P.T. Kim. Spherical Deconvolution. *Journal of Multivariate Analysis*, 67(1):1–22, 1998.

A.M. Heemskerk, T.K. Sinha, K.J. Wilson, Z. Ding, and B.M. Damon. Repeatability of DTI-based skeletal muscle fibre tracking. *NMR in Biomedicine*, 23(3):294–303, 2010.

A.M. Heemskerk, G.J. Strijkers, A. Vilanova, M.R. Drost, and K. Nicolary. Determination of mouse skeletal architecture using three-dimensional diffusion tensor imaging. *Magn. Reson. Med*, 53:1333–1340, 2005.

R.M. Henkelman. Measurement of signal intensities in the presence of noise in MR images. *Medical Physics*, 12:232, 1985.

R.M. Henkelman, G.J. Stanisz, J.K. Kim, and M.J. Bronskill. Anisotropy of NMR properties of tissues. *Magn. Reson. Med*, 32:592–601, 1994.

P. Hoyer. Non-negative sparse coding. pages 557–565, 2002.

A. Hyvarinen, P. Hoyer, and E. Oja. Image Denoising by Sparse Code Shrinkage. pages 473–479, 1999.

S.F. Ismail, J.P. Moss, and R. Hennessy. Three-dimensional assessment of the effects of extraction and nonextraction orthodontic treatment on the face. *Am. J. Orthodont Dentofac. Orthoped*, 121(3):244–256, 2002.

K.N. Jansons and D.C. Alexander. Persistent angular structure: new insights from diffusion magnetic resonance imaging. *Inverse Problems*, 19:1931–1946, 2003.

M. Jenkinson, P.R. Bannister, J.M. Brady, and S.M. Smith. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841, 2002.

M. Jenkinson, C.F. Beckman, T.E.J. Behrens, M.W. Woolrich, and S.M. Smith. Fsl. *NeuroImage*, 62:782–790, 2012.

M. Jenkinson and S.M. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Ananlysis*, 5(2):143–156, 2001.

B. Jeurissen, A. Leemans, D.K. Jones, J.-D. Tournier, and J. Sijbers. Probabilistic fiber tracking using the residual bootstrap with constrained spherical deconvolution. *Human Brain Mapping*, 32(3):461–470, 2011.

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

P. Jezzard and R.S. Balaban. Correction for geometric distortion in the echo-planar images from B0 field variations. *Magn. Reson. Med*, 34:65–73, 1995.

P. Jezzard, A.S. Barnett, and C. Pierpaoli. Characterisation of and correction for eddy current artefacts in echo planar diffusion imaging. *Magn. Reson. Med*, 39:801–812, 1998.

H. Jiang, M.C. Chou, P.C. van Zijl, and S. Mori. Outlier detection for diffusion tensor image by testing for ADC consistency. page 1404, 2009.

L. Jonasson, P. Hagmann, J.P. Thiran, and V.J. Wedeen. Fiber tracts of high angular resolution diffusion MRI are easily segmented with spectral clustering. In *In Proc. ISMRM 13*, page 1310, 2005.

D.K. Jones. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magn. Reson. Med*, 49:7–12, 2003.

D.K. Jones. The effect of gradient sampling scheme on measures derived from diffusion tensor MRI : A monte carlo study. *Magn. Reson. Med*, 51:807–815, 2004.

D.K. Jones. Tractography gone wild: Probabilistic fibre tracking using the wild bootstrap with diffusion tensor MRI. *IEEE TMI*, 27(9):1268–1274, 2006.

D.K. Jones. Precision and accuracy in diffusion tensor magnetic resonance imaging. *Magn. Reson. Med*, 21(2):87–99, 2010.

D.K. Jones and P.J. Basser. "squashing Peanutes and Smashing Pumpkins": How noise distorts diffusion weighted MR data. *Magn Reson Med*, 52:979–993, 2004a.

D.K. Jones and P.J. Basser. "squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. *Magn. Reson. Med*, 52(5):979–993, 2004b.

D.K. Jones, G. Eden, C. Pierpaoli, and P.J. Basser. PASTA: Pointwise assessment of streamline tractography. *Magn. Reson. Med*, 53(6):1462–1467, 2005.

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

J.H. Kan, M.H. Anneriet, Z. Ding, A. Gregrory, G. Mencio, K. Spindler, and B.M. Damon. DTI-based muscle fibre tracking of the quadriceps mechanism in lateral patellar dislocation. *Magnetic Resonance in Medicine*, 29(3):663–670, 2009.

C.H. Kau, S. Richmond, A. Incrapera, J. English, and J.J. Xia. Three-dimensional surface acquisition systems for the study of facial morphology and their application to maxillofacial surgery. *Int. Journal of Medical Robotics and Computer Assisted Surgery*, 3:97–110, 2007.

C.H. Kau, S. Richmond, J.M. Palomo, and M.G. Hans. Three-dimensional cone beam computerized tomography in orthodontics. *Journal of Orthodontics*, 32(4):282–293, 2005.

S. Kim, G. Chi-Fishman, A.S. Barnett, and C. Pierpaoli. Dependence on diffusion time of apparent diffusion tensor of ex vivo calf tongue and heart. *Magn. Reson. Med*, 54 (6):1287–1296, 2005.

S. Klein, M. Staring, K. Murphy, M.A. Viergever, and J.P.W. Pluim. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE trans. on Medical Imaging*, 29 (1):196–205, 2010.

M.E. Komlosh, F. Horkay, R.Z. Freidlin, U. Nevo, Y. Assaf, and P.J. Basser. Detection of microscopic anisotropy in grey matter and in a voxel tissue phantom using double Pulsed Greadient Spin Echo NMR. *Journal of Magnetic Resonance*, 189, 2007.

L.D. Landau, E.M. Lifshitz, and L.P. Pitaevskii, editors. *Statistical Physics*. Butterworth-Heinemann Ltd, Oxford, 1996.

P. Lanzer, C. Barta, E.H. Botvinick, H.U. Wiesendanger, G. Modin, and C.B. Higgins. ECG-synchronised cardiac MR imaging: method and evaluation. *Radiology*, 155(3): 681–686, 1985.

P. Lanzer, E.H. Botvinick, N.B. Schiller, L.E. Crooks, M. Arakawa, L. Kaufman, P.L. Davis, R. Herfkens, M.J. Lipton, and C.B. Higgins. Cardiac imaging using gated magnetic resonance. *Radiology*, 150(1):121–127, 1984.

I.N. Lawes, T.R. Barrick, V. Murugam, N. Spierings, D.R. Evans, M. Song, and C.A. Clark. Atlas-based segmentation of white matter tracts of the human brain using diffusion tractography and comparison with classical dissection. *NeuroImage*, 39:62–79, 2008.

M. Lazar and A.L. Alexander. Error analysis of white matter tracking algorithms (streamlines and tensorlines) for DT-MRI. In *In Proc. ISMRM, Glasgow, Scotland.*, page 506, 2001.

M. Lazar, D.M. Weinstein, J.S. Tsuruda, K.M. Hasan, M.E. Meyerand, B. Badie, H.A. Rowley, V. Haughton, A. Field, and A.L. Alexander. White matter tractrography using diffusion tensor deflection. *Human Brain Mapping*, 18(4):306–321, 2003.

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

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

G.C. Lee, P.W. Goodwill, K. Phuong, B.A. Inglis, G.C. Scott, B.A. Hargreaves, L. Li, A.C. Chen, R.N. Shah, and S.M. Conolly. Pyrolytic graphite foam: A passive magnetic susceptibility matching material. *Journal of Magnetic Resonance Imaging*, 32(3):684–691, 2010.

N. Lee and M. Singh. Compressed sensing base diffusion spectrum imaging. In *In Proc. ISMRM International Conference*, page 1697, 2010.

A. Leemans, B. Jeurissen, J. Sijbers, and D.K. Jones. ExploreDTI: A Graphical Toolbox for Processing, Analyzing, and Visualizing Diffusion MR Data. In *17th Annual Meeting of Intl Soc. Mag. Reson. Med, Hawaii*, page 3537, 2009.

A. Leemans and D.K. Jones. The B-matrix must be rotated when motion correcting diffusion tensor imaging data. *Magn. Reson. Med*, 61:1336–1349, 2009a.

A. Leemans and D.K. Jones. A new approach to fully automated fiber tract clustering using affinity propagation. page 855, 2009b.

C.P. Lin, W.Y. Tseng, and J.H. Chen. Validation of diffusion tensor magnetic resonance axonal fibre imaging with registered manganese enhanced optic tracts. *NeuroImage*, 14:1035–1–47, 2001.

N.F. Lori, J.S. Akbudak, T.S. Shimony, R.K. Snyder, and T.E. Conturo. Diffusion tensor fibre tracking of brain connectivity: reliability analysis and biological results. *NMR in biomed.*, 15:494–515, 2002.

M. Lustig, D. Donoho, and J.M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.*, 58:1182–1195, 2007.

S.K. Warfield M. Maddah, W.E.L. Grimson. Statistical modelling and EM clustering of white matter fiber tracts. In *In Proc. IEEE Int. Symp. Biomed. Imaging*, pages 53–56, 2006.

J.F. Mangin, C. Poupon, C. Clark, D. Le Bihan, and L. Bloch. Distortion correction and robust tensor estimation for MR diffusion imaging. *Med Image Anal*, 6:191–198, 2002.

K.M. Martin, N.G. Papadakis, C.L. Huang, L.D. Hall, and T.A. Carpenter. The reduction of the sorting bias in eigenvalues of the diffusion tensor. *Magn. Reson. Imag.*, 17(6):893–901, 1999.

N.R. Miller. Diffusion tensor imaging of the visual sensory pathway: Are we there yet? *American Journal of Ophthalmology*, 140:896–897, 2005.

P.S. Morgan, R.W. Bowtell, D.J.O. McIntyre, and B.S. Worthington. Correction of spatial distortion in EPI due to inhomogeneous static magnetic fields using the reversed gradient method. *Journal of Magnetic Resonance Imaging*, 19(4):499–507, 2004.

S. Mori, B.J. Crain, V.P. Chacko, and P.C.M. Van Zijl. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2):265–269, 1999.

S. Mori and P.C.M. van Zijl. Fiber tracking: principles and strategies – a technical review. *NMR in Biomedicine*, 15:468–480, 2002.

J.P. Moss. Northcroft revisited. *British Journal of Orthodontics*, 16(3):155–167, 1989.

C.H. Neuman. Spin echo of spins diffusing in a bounded medium. *J. Chem. Phys*, 60: 4508–4511, 1974.

C. Nguyen. 3D image construction of the craniofacial complex. Masters Thesis, Temple University, Philadelphia, PA. 1999.

D.G. Norris. Implications of Bulk motion for Diffusion-Weighted Imaging Experiments: Effects, Mechanisms and Solutions. *Journal of Magnetic Resonance Imaging*, 13:486–495, 2001.

N. Norton, editor. *Netter's Head and Neck Anatomy for Dentistry: ISBN: SBN: 978-1-4377-2663-3*. Saunders, Philadelphia, United States, 2011.

L.J. O'Donnell and C.F. Westin. Automated tractography segmentation using high-dimensional white matter atlas. *IEEE TMI*, 26(11):1562–1575, 2007.

B.A. Olshausen and D.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

S. Pajevic and C. Pierpaoli. Colour schemes to represent the orientation of anisotropic tissues from diffusion tensor data: application to white matter fibre tract mapping in the human brain. *Magn. Reson. Med*, 42:526–540, 1999.

K. Pannek, D. Raffelt, C. Bell, J.L. Mathias, and S.E. Rose. HOMOR: Higher Order Model Outlier Rejection for high b-value MR diffusion data. *NeuroImage*, 63:835–842, 2012.

G.D. Parker, D.Marshall, P.L. Rosin, N. Drage, S. Richmond, and D.K. Jones. A pitfall in the reconstruction of fibre ODFs using spherical deconvolution of diffusion MRI data. *NeuroImage*, 65(15):433–448, 2013a.

G.D. Parker and D.K. Jones. Towards Automated Modelling of Maxillofacial Musculature. page 1916, 2011.

G.D. Parker, D. Marshall, P.L. Rosin, N. Drage, S. Richmond, and D.K. Jones. Fast and fully automated clustering of whole volume brain tractography results using shape-space analysis. page 778, 2013b.

G.D. Parker, D. Marshall, P.L. Rosin, N. Drage, S. Richmond, and D.K. Jones. RES-DORE: Robust estimation in spherical deconvolution by outlier rejection. page 3148, 2013c.

G.J.M. Parker, H.A. Haroon, and C.A.M. Wheeler-Kingshott. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural inter-pretation of MRI diffusion measurements. *Journal of Magnetic Resonance Imaging*, 18(2):242–254, 2003.

K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901.

C. Pierpaoli and P.J. Basser. Microstructural and physiological features of tissues elu-cidated by quantitative diffusion tensor MRI. *J. Magnetic Resonance*, 111:209–219, 1996a.

C. Pierpaoli and P.J. Basser. Towards a quantiative assessment of diffusion anisotropy. *Magn. Reson. Med*, 26:893–906, 1996b.

C. Pierpaoli, S. Marenco, G. Rohde, D.K. Jones, and A.S. Barnett. Analyzing the contribution of cardiac pulsation to the variability of quantities derived from the diffusion tensor. page 70, 2003.

E. Pusey. Magnetic resonance imaging artifacts: mechanism and clinical significance. *Radiographics*, 6(5):891–911, 1986.

E.L.E. Remeur, F. Carre, A.M. Bernard, J.Y. Bandard, P. Rochcongar, and J.D. De Certaines. Multiparametric classification of muscle T1 and T2 relaxation times der-ermined by magnetic resonance imaging. The effects of dynamic exercise in trained and untrained subjects. *British Journal of Radiology*, 67:150–156, 1994.

S. Rieseberg and J. Frahm. Diffusion tensor imaging without geometric distortions using diffusion-weighted single-shot STEAM with partial Fourier acquisition. In *In Proc. ISMRM Kyoto*, page 1185, 2004.

S. Rieseberg, K.D. Merboldt, M. Kuntzel, and J. Frahm. Diffusion tensor imaging using partial Fourier STEAM MRI with projection onto convex subsets reconstruction. *Magnetic Resonance in Medicine*, 54(2):486–490, 2005.

S.M. Ross, editor. *Introduction to Probability and Statistics for Engineers and Scientists*. John Wiley and Sons, *ISBN:047181752X*, 1987.

M. Rowe, H.G. Zhang, N. Oxtoby, and D.C. Alexander. Beyond Crossing Fibres: Tractography Exploiting Sub-voxel Fibre Dispersion and Neighbourhood Structure. *Information Processing in Medical Imaging. Lecture Notes in Computer Science*, 7917: 402–413, 2013.

W. Rucklidge, editor. *The Hausdorff Distance*. Springer Berlin Heidelberg, Berlin, 1996.

P.G. Saffman and M. Delebruck. Brownian motion in biological membranes. *Biophysics*, 72(8):3111–3113, 1972.

K.K. Seunarine and D.C. Alexander. Linear Persistent Angular Structure MRI and nonlinear Spherical Deconvolution for Diffusion MRI. In *In proceedings ISMRM, Seattle, Washington, USA*, 2006.

A.J. Sherbondy, M.C. Rowe, and D.C. Alexander. MictoTrack: An algorithm for concurrent projectome and microstructure estimation. *Medical Image Computing and Computer-Assisted Intervention*, 6361:183–190, 2010.

J.S. Shimony, A.Z. Snyder, N. Lori, and T.E. Conturo. Automated fuzzy clustering of neuro-sonal pathways in diffusion tensor tracking. In *In Proc. ISMRM 10 Honolulu*, 2001.

S. Skare, M. Hedehus, M.E. Moseley, and T.Q. Li. Condition number as a measure of noise performance of diffusion tensor acquisition schemes with MRI. *J. Magn. Reson.*, 147:340–352, 2000.

G. Slabaugh. Computing Euler Angles from a Rotation Matrix, available at *http://www.soi.city.ac.uk/˜sbbh653/publications/euler.pdf*. 1999.

J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE TMI*, 17:87–97, 1998.

S.M. Smith. Fast Robust Automated Brain Extraction. *Human Brain Mapping*, 17: 143–155, 2002.

T.B. Smith and K.S. Nayak. MRI artifacts and correction strategies. *Imaging Medicine*, 2(4):445–457, 2010.

S.K. Song, S.W. Sun, M.J. Ramsbottom, C. Chang, J. Russel, and A.H. Cross. Dysmyelination revealed through MRI as increased radiial (but unchanged axial) diffusion of water. *NeuroImage*, 17:1429–1436, 2002.

O. Sporns, D.R. Chialvo, M. Kaiser, and C.C. Hilgetag. Organization, development and function of complex brain networks. *TRENDS in Cognitive Sciences*, 8(9):418–425.

E.O. Stejskal and J.E. Tanner. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *J. Chem. Phys*, 42(1):288–282, 1965.

Y. Suzuki, H. Matsuzawa, I.L. Kwee, and T. Nakada. Absolute eigenvalue diffusion tensor analysis for human brain maturation. *NMR in Biomedicine*, 16(5):257–260, 2003.

J. Talairach and P. Tournoux, editors. Thieme Medical Publishers, 1988.

J.-D. Tournier, F. Calamante, and A. Connelly. Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35:4591472, 2007.

J.-D. Tournier, F. Calamante, and A. Connelly. How many diffusion gradient directions are required for HARDI? In *Proc. ISMRM 2009*, page 358, 2009.

J.-D. Tournier, F. Calamante, and A. Connelly. MRtrix: diffusion tractography in crossing fibre regions. *Int. J. Imaging Syst. Technol.*, 22:53–66, 2012.

J.-D. Tournier, F. Calamante, D.G. Gadian, and A. Connelly. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23:1176–1185, 2004.

J.-D. Tournier, K.H. Cho, F. Calamante, C.H. Yeh, A. Connelly, and C.P. Lin. Resolving crossing fibres using constrained spherical deconvolution: Validation using diffusion-weighted imaging phantom data. *NeuroImage*, 42:617–625, 2008.

C.A. Trotman, M.M. Gross, and K. Moffatt. Reliability of a three-dimensional method for measuring facial animation: a case report. *Angle Orthodont*, 66(3):195–198, 1996.

T.P. Trouard, Y. Sabharwal, M.I. Altbach, and A.G Gmitro. Analysis and comparison of motion correction technique in diffusion-weighted in diffusion weighted imaging. *J. Magn Reson Imaging*, 6:925–935, 1996.

D.S. Tuch. Diffusion MRI of complex tissue structure. *PhD Thesis*, 1:115–130, 2002.

D.S. Tuch. Q-ball Imaging. *Magn. Reson. Med.*, 52(6):1358–1372, 2004.

O.C. Tuncay. Three-dimensional imaging and motion animation. *Semin. Orthodont*, 7 (4):244–250, 2001.

R. Turner and D. Le Bihan. Single-shot diffusion imaging at 2.0 Tesla. *Journal of Magnetic Resonance*, 86:445–452, 1990.

M. Uecker, S. Zhang, D. Voit, A. Karaus, K.D. Merboldt, and J. Frahm. Real-time MRI at a resolution of 20 ms. *NMR in Biomedicine*, 23(8):986–994, 2010.

W.E. Vinje and J.L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276, 2000.

G. Voronoi. Nouvelles applications des paramertres continus a la theorie des formes quadratiques. *Journal fr die Reine und Angewandte Mathematik*, 133:97–178, 1908.

S. Wakana, A. Caprihan, M.M. Panzenboeck, J.H. Fallon, M. Perry, R.L. Gollub, K. Hua, J. Zhang, H. Jiang, P. Dubey, A. Blitz, and P. van Zijl. Reproducibility of quantitative tractography methods applied to cerebral white matter. *NeuroImage*, 36(3):630–644, 2007.

L. Walker, L.C. Chang, C.G. Koay, N. Sharma, L. Cohen, R. Verma, and C. Pierpaoli. Effects of physiological noise in population analysis of diffusion tensor MRI data. *NeuroImage*, 54:1168–1177, 2011.

J.P. Wansapura, S.K. Hollans, R.S. Dunn, and W.S. Ball. NMR relaxation times in the human brain at 3.0 tesla. 9(4):531–538, 1999.

V. Wedeen, T. Reese, D. Tuch, M. Weigel, J. Dou, R. Weiskoff, and D. Chessler. Mapping fiber orientation spectra in cerebral white matter with fourier-transform diffusion MR. In *Paper Presented at: Proc. Intl. Soc. Mag. Res. Med. (Denver)*, page 82, 2000.

V.J. Wedeen, P. Hagmann, W.I. Tseng, T.G. Reese, and R.M. Weisskoff. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magn. Reson. Med*, 54:1337–1386, 2005.

R.L. White and R.J. Hanisch, editors. *Image restoration using the damped Richardson-Lucy method*. Space Telescope Institute, Baltimore MD, 1993.

T. Yamamoto, K. Yamada, T. Nishimura, and S. Kinoshita. Tractography to depict three layers of visual field trajectories to the calcarine gyri. *American Journal of Ophthamology*, 140(5):781–785, 2005.

C.T. Zhan and R.Z. Roskies. Fourier Descriptors for Plane Closed Curves. *IEEE Trans. on Computers*, 21(2):269–281, 1972.

S. Zhang and D.H. Laidlaw. DTI fiber clustering in the whole brain. In *In Proc. IEEE Visualization*, page 28, 2004.

Z. Zhou, W. Liu, J. Cui, X. Wang, D. Arias, Y. Wen, R. Bansal, X. Hao, Z. Wang, B.S. Peterson, and D. Xu. Automated artifact detection and removal for improved tensor estimation in motion-corrupted DTI datasets using the combination of local binary patterns and 2D partial least squares. *Magn. Reson. Imaging*, 29:230–242, 2011.

L. Zhukov and A.H. Barr. Heart-muscle fibre reconstruction from diffusion tensor MRI. In *Proceedings of the 14th IEEE Visualization Conference 0-7695-2030-8*, 2003.