The development and application of a simulation system for diffusion-weighted MRI

Mark Graham

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Centre for Medical Image Computing & Department of Computer Science

University College London

April 22, 2018

I, Mark Graham, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Diffusion-weighted MRI (DW-MRI) is a powerful, non-invasive imaging technique that allows us to infer the structure of biological tissue. It is particularly well suited to the brain, and is used by clinicians and researchers studying its structure in health and disease. High quality data is required to accurately characterise tissue structure with DW-MRI. Obtaining such data requires the careful optimisation of the image acquisition and processing pipeline, in order to maximise image quality and minimise artefacts.

This thesis extends an existing MRI simulator to create a simulation system capable of producing realistic DW-MR data, with artefacts, and applies it to improve the acquisition and processing of such data. The simulator is applied in three main ways. Firstly, a novel framework for evaluating post-processing techniques is proposed and applied to assess commonly used strategies for the correction of motion, eddy-current and susceptibility artefacts. Secondly, it is used to explore the often overlooked susceptibility-movement interaction. It is demonstrated that this adversely impacts analysis of DW-MRI data, and a simple modification to the acquisition scheme is suggested to mitigate its impact. Finally, the simulation is applied to develop a new tool to perform automatic quality control. Simulated data is used to train a classifier to detect movement artefacts in data, with performance approaching that of a classifier trained on real data whilst requiring much less manually-labelled training data.

It is hoped that both the findings in this thesis and the simulation tool itself will benefit the DW-MRI community. To this end, the tool is made freely

available online to aid the development and validation of methods for acquiring and processing DW-MRI data.

Acknowledgements

I owe thanks to my supervisors, Gary and Ivana. Gary's time, attention to detail, and high standards greatly improved the quality of my PhD work. Ivana was an invaluable source of knowledge on MR physics and POSSUM, and provided a great amount of advice, academic and otherwise, throughout my time at UCL. I enjoyed a fruitful collaboration with Jesper and MJ from FMRIB, Oxford - thanks for your enthusiasm for the simulator, and the chance to contribute to eddy's development.

MIG provided a great environment for research. I'm glad to have been part of a research group with so many students: thanks for all discussions, cookies and conference fun. I was also fortunate to meet Alex, Eliza, Felix, and Zach: thanks for all the laughter, and ensuring I never took anything too seriously.

I am grateful to all my friends and family. Thank you Mum, Dad, and Helen, for all your love, support, and tolerating me never wanting to talk about my thesis! So many friends have made the last four years so fun, but special mention is owed to the residents of the Deptford Commune: Emily, George, Javal, Sinead, and Prev.

Lastly, to Lindsay. You've provided love, encouragement, and the ideal location for me to start writing this thesis: a hammock in a house in Zanzibar, jealously guarded by a very silly cat.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Humans have long sought to understand the functioning of the brain, in both health and disease. The earliest known published study of the brain dates back to Egypt in 1700BC. It records a number of surgical investigations of the damaged brain and describes how these injuries affected patients' abilities to walk, talk and see. In the 1890s Santiago Ramón y Cajal laid the foundations for modern neuroscience, enabling us to provide mechanistic explanations for the ancient Egyptians' observations. He used staining to produce intricate drawings of cells in the brain and proposed the neuron doctrine: the theory that the nervous system is made up of discrete individual cells. These neurons form the building blocks of the nervous system; information is transmitted along their axons (white matter) and processed by their cell bodies (mostly concentrated in regions referred to as grey matter). Understanding the structure and organisation of neurons and supporting cells provides insight into the brain's function.

Historically, our understanding of the structural organisation of the brain has been furthered by invasive procedures such as tracer studies and analysis of post-mortem tissue samples. Recently, diffusion-weighted magnetic resonance imaging (DW-MRI) has emerged as our best candidate for non-invasively studying the brain's building blocks. The technique measures the patterns of

water diffusion in biological tissue; since tissue restricts diffusion and shapes these patterns the measurements can be used to infer tissue structure. The technique is well suited to probing the structure of tissue over the order of a micrometre,[1] the relevant length-scale for exploring neurons. DW-MRI can provide information about the location, orientation and integrity of these cells; estimates of brain connectivity can be obtained by tracking white-matter fibres using information about their local orientation.

High-quality data is required to accurately characterise microstructure with DW-MRI. Some features of high-quality data include sufficient spatial resolution for the study being performed, sufficiently detailed measurements of the diffusion patterns, and an absence of image artefacts. Obtaining such data relies on the careful optimisation of the acquisition, given constraints such as available scan-time, whilst minimising artefacts, followed by the considered choice of processing pipelines that are designed to further remove any residual artefacts and ready the data for analysis. The choices at both stages are numerous and the process of MR image acquisition is sophisticated, meaning issues in both acquisition and processing can be difficult to predict, or even detect; it can also be challenging to understand the cause of problems that do arise. This problem is becoming more pronounced as recent advances in scanner technology and data acquisition provide data of unprecedented spatial and angular resolution. These datasets offer the potential for very rich characterisations of microstructure, but their acquisition presents new challenges and they require increasingly sophisticated processing strategies.

It is important that we are able to fully understand and carefully test data acquisition and processing pipelines, so that we may maximise data quality and understand, detect and correct problems. One of the most powerful tools for the careful testing and validation of a system is simulation. A simulator could enable us to better understand problems that arise in DW-MR datasets, as well as support the testing and development of methods for dealing with these

[1] 1 micrometre (1 µm) is one millionth of a metre. For reference, the width of a human hair ranges from 10-200 µm.

problems. Whilst some simulators exist in DW-MR, there are none that model the physics of MR acquisition and produce DW-MR images of the brain with sufficient realism to enable the testing and development of these methods.

1.2 Problem statement

There is a need for a DW-MRI simulator that is sufficiently realistic to assist the testing and development of methods for the acquisition and processing of DW-MR datasets.

1.3 Project aims and scope

The aims of this thesis are twofold:

1. To develop a DW-MR simulator that is sufficiently realistic to enable the testing and development of acquisition and processing strategies.

2. To apply the simulator to improve DW-MR acquisition and processing.

There are a huge number of elements of the real-world scanning process that can be incorporated into a simulator, governing its degree of 'realism'. In this thesis the choice of features implemented (aim 1) was governed by the degree to which they are useful for assessing and validating acquisition and processing techniques (aim 2).

1.4 Thesis overview and contributions

Most of the contributions in this PhD have occurred as developments to the simulator (meeting aim 1) accompanied by an application of the simulator (meeting aim 2) that required this development. As a result the thesis is best served by a chronological account of the work carried out during this PhD.

In chapter 2, I summarise the relevant background theory of MRI and DW-MRI, and undertake a critical analysis of the work to date in developing simulators for DW-MR.

Chapters 3 and 4 outline work that enables assessment of methods for dealing with two of the most commonly encountered problems in DW-MR: subject

motion and eddy-currents. Chapter 3 describes the simulator development necessary to achieve this. I discuss the decision to produce the DW-MR simulator by extending the open-source POSSUM fMRI simulator, a key strength of which is its ability to simulate subject movement. I describe the implementation of two features to enable assessment of motion and eddy-current techniques. The first is realistic DW-MR contrast — particularly important because the complex contrast in DW-MR datasets is one of the reasons their processing is so challenging. The second is the presence of eddy-current artefacts. This chapter also proposes a novel method for evaluation of image processing techniques; motivated by limitations found in existing methods in the literature. It involves the comparison of displacement fields produced by processing techniques to the ground-truth fields produced by the simulator. In chapter 4, the simulation framework is applied to assessing processing techniques for correcting motion and eddy-currents. Data with high sensitisation to diffusion (summarised by the b-value) is simulated, because such data is being acquired with increasing frequency. The study finds that many of the most frequently used processing methods are unable to deal with such data. It also shows that some of the techniques for evaluating methods that are used in the literature can give misleading results, demonstrating the strength of the evaluation framework proposed in chapter 3.

Chapters 5 and 6 give an account of work done to investigate the susceptibility artefact in diffusion MRI. One of the limitations of the framework is its inability to simulate the spin-echo pulse sequence, which is vital if the susceptibility artefact is to be simulated faithfully. Chapter 5 describes the implementation of spin-echo in POSSUM, and its validation. In chapter 6 I compare the three classes of processing technique used to assess the susceptibility artefact. I further apply the framework to explore the interaction between susceptibility and movement, something that is rarely addressed during image acquisition and processing, and difficult to study without simulation. I demonstrate this interaction affects analysis of DW-MR data, suggesting the need for

further methods development in this area, and suggest a simple adaptation to acquisition that could help mitigate the artefact.

In chapter 7 I demonstrate the application of the simulator to the development of new tools. A classifier is built to identify movement artefacts in DW-MR data. Rather than train the classifier on real data, which requires time-consuming manual labelling of a training set, it is trained on simulated data which comes 'ready-labelled'. I demonstrate the performance on this simulation-trained classifier is close to that of a classifier trained on real data, and suggest the simulator could be useful in other scenarios where machine learning can provide good results but training data is expensive or time-consuming to acquire.

Chapter 2

Background

The aim of my work is to develop and apply a simulation system for DW-MR, capable of producing realistic datasets along with their artefacts. This section will give a concise introduction to the relevant background in MRI (Section 2.1), and its extension to DW-MR (Section 2.2), with a view to understanding the artefacts that occur in DW-MR (Section 2.3). Finally I review existing simulation systems (Section 2.4), focusing on the current state-of-the-art and a discussion of its limitations.

2.1 Magnetic resonance imaging

In this section we provide a discussion of the main steps involved in the formation of an MR image. MRI is a non-invasive imaging technique that combines the nuclear magnetic resonance (NMR) phenomena first observed by Bloch and Purcell in 1946 [1, 2] and spatial encoding principles developed by Lauterbur in 1973 [3]. MR images are formed by recording the radiofrequency (RF) signal produced by a sample while a sequence of magnetic fields and RF pulses are applied across it. We begin with a classical view of protons in a magnetic field (Section 2.1.1), then discuss how the magnetization of these protons is manipulated by applied RF pulses in order to produce a signal (Section 2.1.2). We then discuss how this signal is encoded with spatial information then transformed to produce an image (Section 2.1.3). Finally we discuss some practicalities of image formation and the different types of contrast that may be achieved

(Section 2.1.4).

2.1.1 Protons in a magnetic field

The nuclear magnetic moment, $\vec{\mu}$, is a property of nuclei related to the spin $\vec{J}$, or angular momentum, of their protons and neutrons. Whilst these properties arise from a quantum mechanical treatment of the nuclei, we deal with the collective behaviour of large numbers of nuclei and so the principles we will discuss can be accurately described using a classical model. In MRI we are generally concerned with the manipulation and detection of the magnetic moments of hydrogen nuclei (single protons).

The magnetic moment is a vector quantity and thus needs to be specified by both its magnitude and direction. The magnitude is fixed, and can be determined from the relation

$$\vec{\mu} = \gamma \vec{J} \tag{2.1}$$

where γ is the gyromagnetic ratio, a physical constant that depends on the nucleus. Under normal conditions the direction of $\vec{\mu}$ is completely random and a population of hydrogen nuclei has no net magnetic moment. In MRI we create a net moment through the application of an external magnetic field, $\vec{B_0}$, which we will define, according to convention, as pointing along the z-axis. This field causes the the z-component of $\vec{\mu}$ to take on a fixed magnitude and align either parallel or anti-parallel with $\hat{z}$ [4]. The transverse (xy) components of each nuclei's magnetic moment experience a torque from the external field. The torque causes this component to precess about the z-axis with a frequency given by the Larmor equation:

$$\vec{\omega_0} = \gamma \vec{B_0} \tag{2.2}$$

where γ is $42.58\,\mathrm{MHz\,T^{-1}}$ for hydrogen. The net magnetic moment of this

sample is the vector sum of the n individual moments:

$$\vec{M} = \sum_{i=1}^{n} \vec{\mu_i} \tag{2.3}$$

The transverse components of each nuclei's magnetic moment precess at the same frequency but with random initial phase, so the net transverse component of $\vec{M}$ is zero. These parallel and anti-parallel states differ in energy, causing slightly more nuclei to occupy the lower energy parallel state than anti-parallel.[1] This means $\vec{M}$ has a non-zero value that points fully along the positive z-axis. From herein we discuss MR signal formation in terms of manipulation and detection of this net magnetization.

2.1.2 Detection of the magnetization

The formation of a non-zero $\vec{M}$ is not sufficient to form a detectable signal. $\vec{M}$ must be tipped into a transverse plane in order to set it into precession. It is this precessing magnetization that, according to Faraday's law, will induce electric currents in nearby coils, which constitutes the signal that we measure.

The tipping of $\vec{M}$ is achieved by the application of a magnetic field oscillating in the RF range, known as an RF pulse.[2] The RF pulse takes the following form:

$$\vec{B_1}(t) = B_1^e(t) \left[\cos\left(\omega_{rf} t\right) \vec{i} - \sin\left(\omega_{rf} t\right) \vec{j} \right] \tag{2.4}$$

where B_1^e is known as the envelope function, and ω_{rf} is the oscillation frequency. B_1^e is typically much smaller than B_0, on the order of µT.

The action of the pulse can be understood using the Bloch equations [5], which describe the temporal behaviour of magnetisation in an applied external

[1] The difference in population of the two states is related to their energy difference. This can be calculated using the Boltzman relationship and yields the expression for the net magnetisation for a sample of protons: $\frac{\gamma^2 \hbar^2 B_0 n}{4KT}$ where $\hbar$ is Planck's constant, K is Boltzmann's constant and T is the temperature of the sample.

[2] As we will soon see, these pulses need to have frequencies close to the Larmor frequency in order to manipulate $\vec{M}$. MR magnets tend to be in the range 1.5-13 T, which places ω_0 in the range 60-550 MHz, which corresponds to radio frequencies.

field. Here we write them in an MRI specific form:

$$\frac{d\vec{M}}{dt} = \gamma\vec{M} \times \vec{B} - \frac{M_x\vec{i} + M_y\vec{j}}{T_2} - \frac{\left(M_z - M_z^0\right)\vec{k}}{T_1} \tag{2.5}$$

where T_1 and T_2 are relaxation constants arising from spin-lattice and spin-spin interactions respectively, and M_z^0 is the value of M_z at thermal equilibrium. During excitation the relaxation terms can be neglected as the RF pulse duration is much less than T_1 and T_2: typical RF pulses are on the order of 1 ms, whilst T_1 ∼1000 ms and T_2 ∼100 ms for biological tissue [6]. The Bloch equation can then be rewritten in a rotating frame of reference [7, Chapter 3.2]:

$$\frac{\partial\vec{M}_{\text{rot}}}{dt} = \gamma\vec{M}_{\text{rot}} \times \vec{B}_{\text{eff}} \tag{2.6}$$

where $\vec{B}_{\text{eff}}$ is the effective magnetic field experienced in this frame. If the frame is rotating at the Larmor frequency, ω_0, and the RF field is applied with $\omega_{rf} = \omega_0$, then we find $\vec{B}_{\text{eff}} = B_1^e(t)\,\vec{i}$. The solution to this gives:

$$M_x = 0 \tag{2.7}$$

$$M_y = M_z^0 \sin\left(\int_0^t \gamma B_1^e(t)\,dt\right) \tag{2.8}$$

$$M_z = M_z^0 \cos\left(\int_0^t \gamma B_1^e(t)\,dt\right) \tag{2.9}$$

That is, in the rotating frame the magnetization precesses about the x-axis. By controlling the form of B_1 and the time it is applied for, we can control the angle of this rotation. Often we select these to achieve a 90° flip. In the lab frame, this flip is observed as a spiralling descent — see Fig 2.1.

2.1.3 Spatial encoding

After the RF pulse, $\vec{M}$ is precessing in a transverse plane, or 'excited', producing a measurable signal in the RF receive coils. In order to proceed from NMR to MRI, we need to spatially localise this signal to produce an image.

(a) Rotating frame **(b)** Lab frame

Figure 2.1: Motion of $\vec{M}$ during an 90° pulse in the rotating and lab frames.

This is typically done in a two-step process: the selective excitation of a 2D slab of protons, known as a slice, followed by the spatial encoding of the signal within that slice.

Slice selection is achieved by applying a magnetic field gradient, $\vec{G}$, during the RF pulse. In MRI, gradients are magnetic fields with a z-component[3] that varies linearly along the direction of $\vec{G}$. This causes the precessing frequency of the sample to vary spatially:

$$\omega(\vec{r}) = \gamma\left(B_0 + \vec{G}\cdot\vec{r}\right) \tag{2.10}$$

If we assume that $\vec{G}$ varies along z, the sample's Larmor frequency varies as $\omega(z)$. If we could apply an RF pulse that only contained frequencies in the interval $\omega(z_0 \pm \delta z)$, we only excite protons in a slice of thickness $2\delta z$, centred about z_0. This can be achieved by applying an RF pulse with an envelope function equal to a sinc function with appropriately chosen parameters; the frequency content of such a function, obtained from its Fourier transform, has the shape of a boxcar with uniform amplitude over the desired range, leading to the equal excitation of all protons in this frequency range.[4]

[3]The gradient field must also contain components in the x- and y-directions, in order to satisfy Gauss' law for magnetism, $\nabla \cdot B = 0$. However these are small compared to B_0 and can usually be ignored.

[4]To create an RF pulse with an envelope function with the exact form of a sinc would take infinite time. In practice, more complicated envelope functions designed to give reasonable approximations to the boxcar with shorter RF pulses are used, with the result that the excitation profile is not uniform across the slice.

It is now necessary to spatially encode the signal produced by the excited protons in this slice so that we may form an image. Conceptually, it is easiest to think in terms of a net transverse magnetization within a small volume element at each location $\vec{r}$ in the slice:

$$\vec{M}_{xy}(\vec{r}) = \int \vec{\mu}_{xy}(\vec{r})\,\mathrm{d}^3\vec{r} \tag{2.11}$$

From equation 2.10, we see that we can induce spatially varying phase in $\vec{M}_{xy}$:

$$\phi(\vec{r},t) = \int -\gamma\left(\mathrm{B}_0 + \vec{G}\cdot\vec{r}\right)\mathrm{d}t \tag{2.12}$$

where the minus sign accounts for the fact that the precession occurs in a clockwise direction. The signal we measure in our receiver coil is proportional to the sum of the contributions of the magnetisation at each location which we can express as an integral [7, Section 3.4]. Adopting a complex representation for $\vec{M}_{xy}$ we have:

$$S(t) = \int M_{xy}(\vec{r}) \exp\left(-i\gamma\left(\mathrm{B}_0 t + \int \vec{G}\cdot\vec{r}\,\mathrm{d}t\right)\right)\mathrm{d}\vec{r} \tag{2.13}$$

where we have neglected relaxation terms and assumed a homogeneous reception field over the sample. If we define the reciprocal space vector:

$$\vec{k} = \gamma \int \vec{G}\,\mathrm{d}t \tag{2.14}$$

then we can rewrite Equation 2.13, leaving out the constant phase shift induced by the main field B_0, to find a Fourier relationship between the magnetization and the measured signal

$$S(\vec{k}) = \int M_{xy}(\vec{r}) \exp\left(-i\vec{k}\cdot\vec{r}\right)\mathrm{d}\vec{r} \tag{2.15}$$

We can now see that if we record our signal in 'k-space', we can obtain M_{xy}

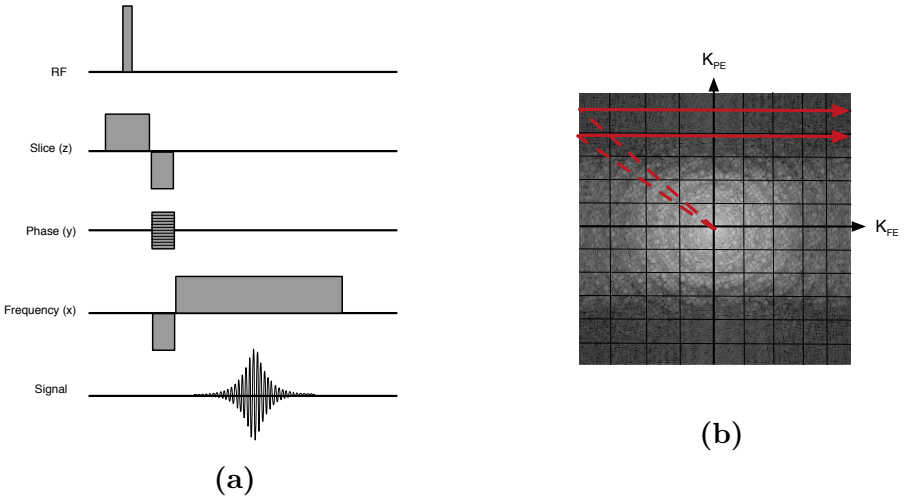


Figure 2.2: The gradient-echo pulse sequence. After excitation, the phase- and frequency-encoding gradients are used to navigate to the start of the k-space line, then frequency encoding and signal readout begins. Fig (a) shows the pulse sequence diagram and (b) shows the corresponding trajectory through k-space for the acquisition of two successive lines.

from a Fourier inversion of this signal:

$$M_{xy}(\vec{r}) = \mathcal{F}^{-1}\left(S(\vec{k})\right) \tag{2.16}$$

Imaging now becomes a matter of using carefully controlled sequences of applied RF pulses and gradients to sample $S(\vec{k})$.

A simple method of sampling k-space involves acquiring data one line at a time. Two gradients are used, called frequency encoding (G_{FE}) and phase encoding (G_{PE}), which are chosen here to lie along the x and y axes respectively. G_{PE} is applied for a time τ then G_{FE} is applied continuously for a time t whilst the signal is recorded at periodic intervals. This process is repeated for each line, and illustrated graphically in Fig 2.2. This type of sequence is called a gradient-echo. The traversal to the far left of k-space in Figure 2.2 corresponds to a gradient-induced dephasing of the spins, and the subsequent movement towards the right in k-space is a gradient-induced rephasing, such that the spins are maximally rephased (and thus the signal is largest) at $K_{\mathrm{FE}} = 0$, the point known as the echo. The signal is given by

$$S(k_{\mathrm{FE}}, k_{\mathrm{PE}}) = \iint M_{xy}(x, y) \cdot \exp(-ixk_{\mathrm{FE}}) \cdot \exp(-iyk_{\mathrm{PE}})\,\mathrm{d}x\,\mathrm{d}y \tag{2.17}$$

with $k_{\mathrm{FE}} = \gamma \int G_{\mathrm{FE}}\,\mathrm{d}t$ and $k_{\mathrm{PE}} = \gamma \int G_{\mathrm{PE}}\,\mathrm{d}\tau$. The two orthogonal components of the transverse magnetisation are usually detected in quadrature and represented as a magnitude and a phase image.

2.1.4 Image formation and contrast

We have seen that the formation of an image of M_{xy} involves the Fourier transform of a signal acquired in k-space. Until now we have implicitly assumed our k-space signal is continuous and extends to infinity, but in practice digitising the signal requires that we sample it discretely, and time requirements constrain the extent of the sampled k-space region. We can understand the effect of this sampling by considering the Fourier transform of an infinite signal multiplied by two functions; a comb-function to represent the discrete sampling and a boxcar that represents truncation. Working in one-dimension to simplify notation, we have:

$$\hat{M}_{xy}(x) = \mathcal{F}^{-1}\left(S(k_x)C(k_x)B(k_x)\right) \tag{2.18}$$

where $\hat{M}_{xy}$ is the discrete approximation of the continuous function, M_{xy}, the comb-function is defined as:

$$C(k_x) = \sum_{n=-\infty}^{\infty} \delta(n\Delta k) \tag{2.19}$$

where δ is the Dirac delta function and the boxcar:

$$B(k_x) = \begin{cases} 1 & |k_x| \leq K/2 \\ 0 & |k_x| > K/2 \end{cases} \tag{2.20}$$

Using the convolution theorem and substituting for Eq 2.16, we find that:

$$\hat{M}_{xy}(x) = M_{xy}(x) * \mathcal{F}^{-1}\left(C(k_x)\right) * \mathcal{F}^{-1}\left(B(k_x)\right)) \tag{2.21}$$

That is, our image is the convolution of a perfect, continuous image and the Fourier inversions of the comb and boxcar functions. The Fourier transform of a comb with spacing Δk is itself a comb with spacing $L = 1/\Delta k$ and the transform of the boxcar is $K\,\mathrm{sinc}(\pi K x)$. Fig 2.3 depicts the combined effect of these two convolutions on the reconstructed image, $\hat{M}_{xy}(x)$.

Discrete, finite sampling can introduce two artefacts into the final image. Consider the size of the object being imaged along a given dimension, W. If this is larger than the period of repeat caused by finite sampling, L, multiple copies of the image will overlap in the final image, known as aliasing. The condition for this not to occur in the image is known as the Nyquist criterion:

$$L > W \text{ or } \Delta k < \frac{1}{W} \tag{2.22}$$

and places a lower bound on the sampling frequency in k-space. The second artefact is the blurring introduced by the finite sampling. As K decreases (corresponding to sampling less of the high frequency components on the edges of k-space) the sinc function broadens and the image resolution is decreased.

Until now we have focused on forming an image of M_{xy}. In practice, M_{xy} is a function of proton density and the relaxation constants T_1 and T_2. By using different sequences of RF pulses and gradients, it is possible to sensitise M_{xy} to produce various image contrasts. This versatility is a key advantage of MRI when compared to other imaging technologies, such as CT. We can understand how signal contrast may be influenced by deriving the signal in a single voxel for the case of the gradient-echo sequence discussed earlier and shown in Fig 2.2. Rewriting the Bloch equations (Eq. 2.5) in transverse and

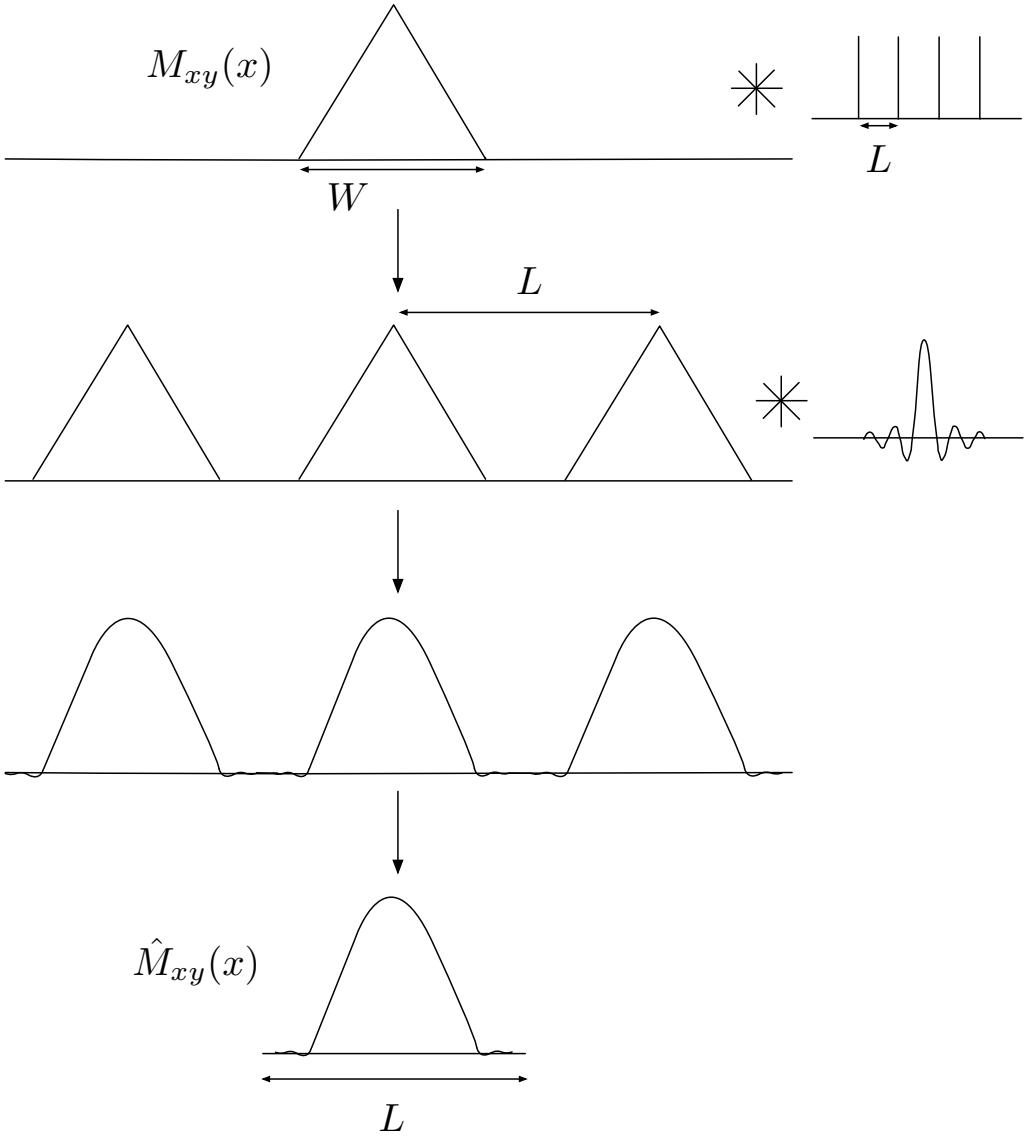


Figure 2.3: Image formation in the case of discrete, finite sampling. The actual image, M_{xy} is convolved with a sampling function to produce periodic repeats of the image. Further convolution with a sinc function leads to blurring. One of the image repeats is selected to produce the reconstructed image, $\hat{M}_{xy}$.

longitudinal form:

$$\frac{dM_{xy}}{dt} = -\frac{M_{xy}}{T_2^*} \tag{2.23}$$

$$\frac{dM_z}{dt} = -\frac{M_z^0 - M_z}{T_1} \tag{2.24}$$

where we have expressed the equations a frame rotating at the Larmor frequency, and replaced T_2 by T_2^*. T_2^* is an adjusted relaxation constant related to T_2 by the expression $1/T_2^* = 1/T_2 + 1/T_2'$, where T_2' accounts for dephasing caused by field inhomogeneities and accelerates the rate of transverse signal decay. These equations have the solutions:

$$M_{xy}(t) = M_{xy}(0)e^{-t/T_2^*} \tag{2.25}$$

$$M_z(t) = M_z^0\left(1 - e^{-t/T_1}\right) + M_z(0)e^{-t/T_1} \tag{2.26}$$

where $M_z(0)$ and $M_{xy}(0)$ are the initial conditions and M_z^0 is the magnetisation at equilibrium. After the first 90° RF of the gradient-echo sequence, we have $M_z(0) = 0,\ M_{xy}(0) = M_z^0$. If the spacing between RF pulses is T_R then just before the second RF pulse the magnetisation is:

$$M_{xy}(t = T_R) = M_z^0 e^{-T_R/T_2^*} \tag{2.27}$$

$$M_z(t = T_R) = M_z^0\left(1 - e^{-T_R/T_1}\right) \tag{2.28}$$

Assuming that $T_R \gg T_2^*$ so that the transverse magnetisation has decayed completely before the next RF pulse is applied, we can solve Eqs 2.25 and 2.26 with new initial conditions $M_z(0) = 0,\ M_{xy}(0) = M_z^0\left(1 - e^{-T_R/T_1}\right)$ to find the signal at $2T_R$, just before the third RF pulse:

$$M_{xy}(t = 2T_R) = M_z^0\left(1 - e^{-T_R/T_1}\right)e^{-T_R/T_2^*} \tag{2.29}$$

$$M_z(t = 2T_R) = M_z^0\left(1 - e^{-T_R/T_1}\right) \tag{2.30}$$

A pattern emerges: after the initial RF excitation, every subsequent RF pulse

leads to a total available transverse magnetisation of $M_z^0\left(1-e^{-T_R/T_1}\right)$. In the gradient-echo experiment the signal peak for each line occurs as the centre of k-space is crossed at $t = T_E$, and at this point we have:

$$M_{xy}(t = T_E) = M_z^0\left(1-e^{-T_R/T_1}\right)e^{-T_E/T_2^*} \tag{2.31}$$

Finally, noting that the magnetisation at equilibrium, M_z^0 is proportional to the proton density ρ we have:

$$M_{xy}(t = T_E) \propto \rho\left(1-e^{-T_R/T_1}\right)e^{-T_E/T_2^*} \tag{2.32}$$

We can now see that though careful choice of pulse-sequence timings T_R and T_E we can sensitise the available signal to different underlying tissue features. For large values of T_R, $(1-e^{-T_R/T_1}) \to 1$ and we lose sensitivity to differences in T_1 values, and for small values of T_E we have $e^{-T_E/T_2^*} \to 1$ and sensitivity to T_2 differences is lost. Images with long T_R and long T_E are commonly referred to as T_2-weighted, and images with short T_E and short T_R are known as T_1-weighted. Image contrast is always influenced by ρ, but sensitivity to density can be maximised with large values of T_R and small values of T_E. Fig 2.4 displays how these parameters influence images of the brain.

2.2 Diffusion MRI

We have seen how images in MRI are formed through the excitation of signal and its ensuing spatial localisation, and further explored how careful choices of the pulse sequences can produce images sensitive to contrast in ρ, T_1 and T_2. In this section we describe how the MR acquisition process can be extended to produce contrast that depends on the local diffusion of water molecules, once again with an view towards understanding the artefacts that arise in this technique. We begin with a discussion of the phenomenon of diffusion (Section 2.2.1), then discuss how MRI can be used to measure diffusion (2.2.2). Finally we touch on models used to infer microstructure from these measure-

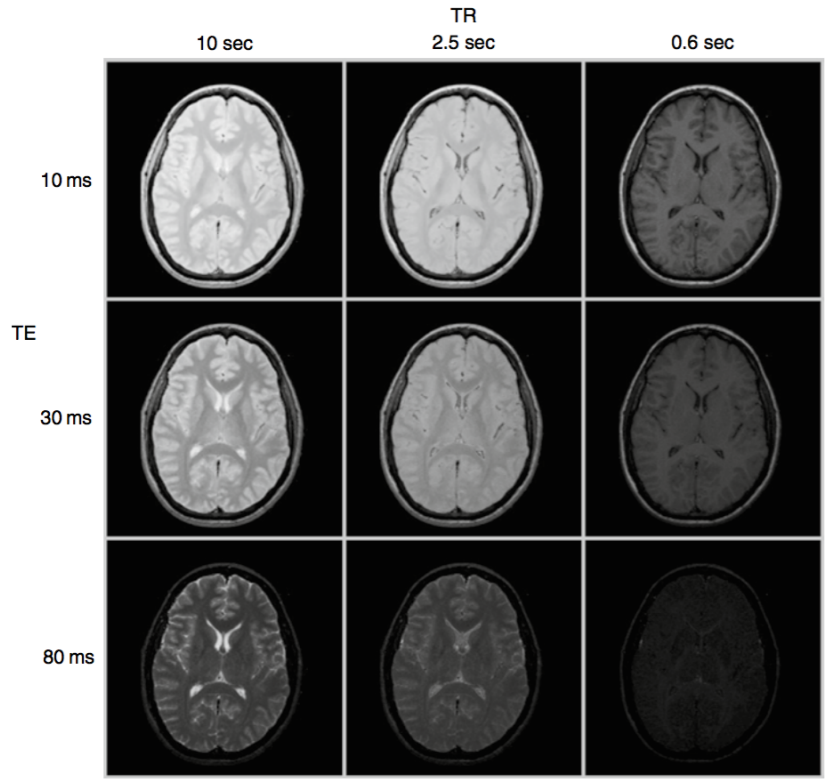


Figure 2.4: Image contrast as a function of TR and TE. The top-right image would be referred to as T_1-weighted, and the bottom left T_2-weighted. Figure reproduced with permission from Elsevier.

ments in Section 2.2.3.

2.2.1 Diffusion

Diffusion is the transport of mass without requiring bulk motion, thus differentiating it from other forms of mass transport such as convection and advection. We can derive the diffusion equation by taking a phenomenological approach first used by Fick, though it is also possible to arrive at the same result by taking a microscopic view of diffusion, as Einstein did in 1905 [8].

We start with Fick's first law, which relates the flux of a substance, $\vec{J}$, to its concentration, ϕ. In the biological systems we typically study with diffusion MRI we consider self-diffusion, that is the diffusion of water in water. Thus the concept of concentration becomes poorly defined, and we replace it with $P(\vec{r_0}, \vec{r_1}, t)$[5], the probability that a particle will move from position $\vec{r_0}$ to $\vec{r_1}$ at

[5]It isn't obvious we can replace concentration with this quantity. See [9] for an explana-

time t. Fick's law is thus given by:

$$\vec{J} = -D\nabla P \tag{2.33}$$

where D is the diffusion coefficient, a property of the diffusing particles, the substance they are in, and the temperature. We can progress further by imposing conservation of mass, assuming there are no sinks or sources of our substance in the region we are considering:

$$\nabla \cdot \vec{J} + \frac{\partial P}{\partial t} = 0 \tag{2.34}$$

Combining the two yields the diffusion equation:

$$\frac{\partial P}{\partial t} = \nabla \cdot (D\nabla P) \tag{2.35}$$

We can solve this equation for a given set of boundary conditions and initial conditions to find $P(\vec{r_0}, \vec{r_1}, t)$, which characterises diffusion for our system. It is important to understand this equation does not model the behaviour of any single particle. Our system is comprised of many particles undergoing random walks, and $P(\vec{r_0}, \vec{r_1}, t)$ describes the net behaviour of this ensemble of particles.

The diffusion of molecules is influenced by their environment. For example in the brain the diffusion of water within a long, cylindrical axon is impeded perpendicular to the axon's main axis, and much freer along this axis — this is illustrated in Fig 2.5. This means local structure influences the diffusion propagator, P. Moseley et. al. were the first to demonstrate it is possible to measure this anisotropy in the nervous system using MRI [10, 11]. The aim of DW-MR is to acquire a series of images that allow us make inferences about the local structure.

---

tion.

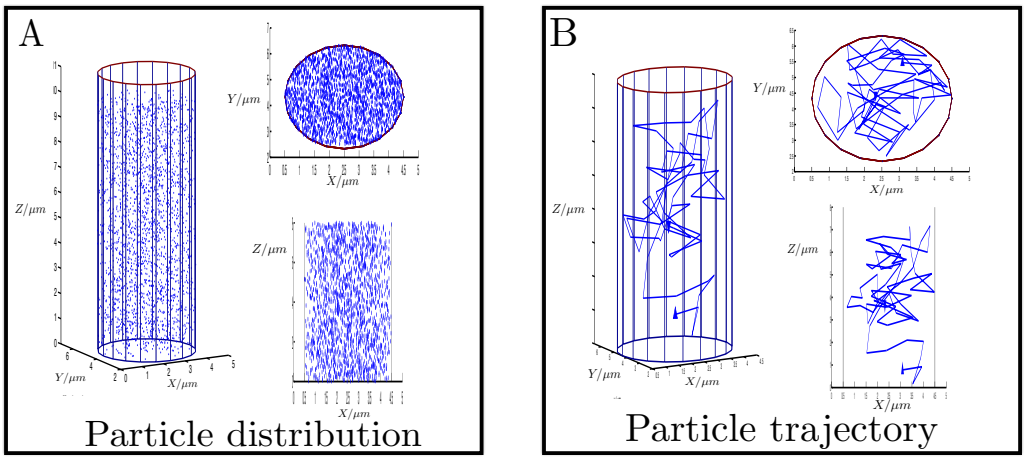


Figure 2.5: Diffusion in a cylindrical geometry. Figure A shows a distribution of spins inside a cylinder at a time t_0, from three different views. Figure B shows the trajectory of a single spin from t_0 to a time t_1. The particle is freer to diffuse along the axis of the cylinder (here z) than it is perpendicular to the axis (the $x-y$ plane), which will be reflected in the diffusion propagator P.

2.2.2 Measuring diffusion with MRI

To illustrate the extension of MRI to measuring diffusion we need to introduce three new concepts. We will first cover diffusion preparation gradients and the spin-echo (SE), which combine to form one of the most popular diffusion encoding sequences, the pulsed-gradient spin-echo (PGSE) [12]. We then discuss a fast imaging technique, echo-planar imaging (EPI), which is crucial for DW-MR.

2.2.2.1 Diffusion preparation

The MR signal can be sensitised to reflect the diffusion of its constituent protons by the application of a set of two or more gradients. As we have seen in Section 2.1.3, if we apply a gradient in a given direction for a time period δ, we introduce a phase shift into each spin:

$$\phi = \int_0^{\delta} \gamma \left(\mathrm{B}_0 + \vec{G} \cdot \vec{r}(t) \right) \mathrm{d}t \tag{2.36}$$

If, at time Δ, we apply a second gradient along the same direction and length as the first but with the opposite sign then the total phase of each spin is:

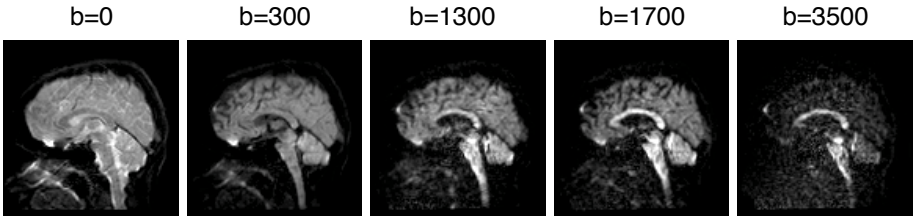


Figure 2.6: Variation in contrast with b-value, for a fixed direction of diffusion-weighting. Units are $\mathrm{s\,mm^{-2}}$. Intensities scales of each image have been individually adjusted; it is contrast rather than absolute intensity that is relevant here.

$$\phi = \gamma \left(\int_0^{\delta} \vec{G} \cdot \vec{r}(t) - \int_{\Delta}^{\Delta+\delta} \vec{G} \cdot \vec{r}(t) \right) \mathrm{d}t \tag{2.37}$$

If the spin moves parallel to the direction of the applied gradients in this time then it will end up with a non-zero phase. A population of spins diffusing along the direction of the gradients will end up with a distribution of phases. This causes a reduction in $\vec{M}_{xy}$ in a local region, and if this is done after an excitation pulse (commonly chosen to be 90° in DW-MR) but before the imaging is performed it will cause a reduction in the measured signal from this region. The more that spins are free to diffuse along the direction of the gradient, the broader the distribution of spins and the greater the reduction in measured signal. Stejskal showed that in the case of free diffusion, the MR signal S can be expressed as:

$$S = S_0 e^{-D\left(\gamma\delta|\vec{G}|\right)^2(\Delta-\delta/3)} \tag{2.38}$$

where D is the diffusion coefficient along the direction $\vec{G}$, S_0 is the MR signal when no diffusion gradients are applied and the b-value $b = \gamma^2\delta^2|\vec{G}|^2\,(\Delta - \delta/3)$ summarises the strength of diffusion weighting — increasing this quantity increases the amount of signal attenuation. Figure 2.6 shows how image contrast is affected by the b-value.

2.2.2.2 Spin-echo

The use of a spin-echo enables us to measure more of the available signal, increasing our signal-to-noise ratio (SNR). We have seen that the transverse MR signal decays exponentially as a result of two factors, one intrinsic to a tissue type and caused by random processes on the molecular level, T_2, and an additional factor caused by field distortions and inhomogeneities, T_2'. A spin-echo allows us to correct for this additional factor, producing more signal. This is particularly important in DW-MR because the diffusion preparation gradients increase the amount of time between excitation and the recording of the signal, T_E.

The spin-echo consists of an 180° RF pulse applied in the transverse plane at time $T_E/2$, which inverts the phase of each spin. Spins in regions of positive field distortion, which have accrued extra phase as a result of these distortions, will now make up this extra phase between times $T_E/2$ and T_E, so that by time T_E there will be no phase differences between spins caused by constant field distortions. This will maximise the net magnetization available for measurement. We are able to recover signal lost due to T_2' effects because they remain constant in time and thus the amount of rephasing after the 180° pulse equals the amount of dephasing beforehand; T_2 effects are not recoverable in the same way because the inhomogeneities they are caused by do not stay fixed in time.

In PGSE, the typical DW-MR measurement, the spin-echo is applied between the two diffusion preparation gradients (Fig 2.7). Note that the use of a spin-echo also means that the second diffusion gradient must now have the same sign as the first.

2.2.2.3 EPI

Together, diffusion preparation and spin echo provide an appropriately sensitised signal with sufficient SNR. However the method for recording this signal outlined in 2.1.3, based on a line-by-line acquisition of k-space, is slow because an RF excitation is required for every single line in the slice. In DW-MR we

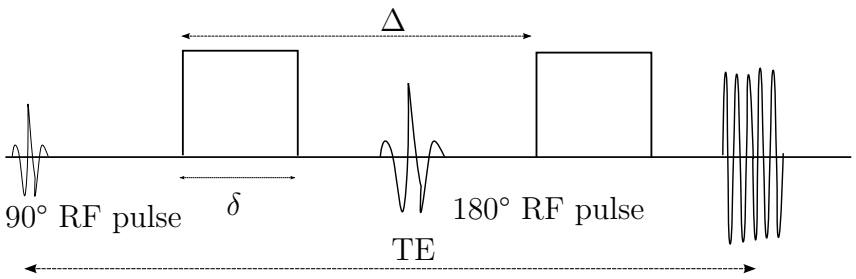


Figure 2.7: Schematic of the Stejskal-Tanner pulse sequence

typically acquire a number of volumes, and a faster imaging method is needed to prevent subject movement corrupting the images.

In single-shot EPI the k-space for a single slice is fully traversed in a single excitation [13], greatly increasing the speed of acquisition. After a line is acquired with a FE gradient, a PE gradient is 'blipped' to increment the phase of the sample, and then the next line is acquired using a FE gradient of reversed polarity, see Fig 2.8. This is repeated until the full slice has been acquired. Multi-shot variants exist in which different portions of k-space are sampled across several excitations, then combined to produce the total k-space. As with the line-by-line readout, 3D volumes are produced by exciting and acquiring successive 2D slices and stacking them.

2.2.3 Models of diffusion

Once a series of measurements of local diffusion have been performed, it is desirable to make inferences about the microstructure that gave rise to this pattern of diffusion. Historically one of the most popular models has been the diffusion tensor (DT) model [14]; due it its simplicity and ability to capture anisotropic diffusion.

$$\mathbb{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} \tag{2.39}$$

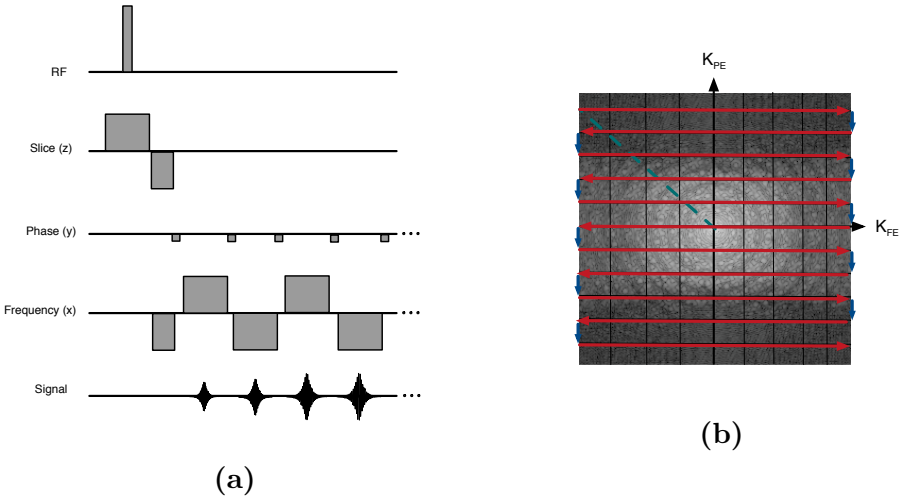


Figure 2.8: A single-shot EPI scheme. Lines are acquired sequentially without returning to the origin of k-space between each, a much faster process than line by line acquisition. (a) pulse sequence and (b) corresponding trajectory through k-space shown.

D_{xx}, D_{yy} and D_{zz} represent diffusivity along three orthogonal axes, and the off-diagonal components represent correlations between displacements along these orthogonal axes. The signal from the DT model can be written:

$$S = S_0 e^{-b\hat{G}^T \mathbb{D} \hat{G}} \tag{2.40}$$

The diffusion tensor contains six unique elements, meaning that seven unique measurements are needed to estimate both $\mathbb{D}$ and S_0.

The DT is able to approximate the signal arising from brain tissue at sufficiently low b-values. For example, in the case of a bundle of parallel axons we would see high diffusivity along the bundle's axis and reduced diffusivity parallel to it. The diffusion tensor would reflect this structure with $D_{zz} > D_{xx}, D_{yy}$ (assuming the bundle's axis lies along z). Metrics such as fractional anisotropy (FA) [15], mean diffusivity (MD), and radial diffusivity (RD) may be calculated and can provide insights into the underlying microstructure. The tensor's principal eigenvector describes the direction of maximum diffusivity and can be used to perform tractography [16].

Despite its uses, the DT has some limitations. It is predicated on the as-

sumption of Gaussian, or free, diffusion [17]; whilst it performs well for moderate b-values it does not describe data well for higher b-values [18]. Furthermore the DT cannot model more complex microstructural features such as crossing white-matter fibres or dispersion within fibre bundles. These limitations have led to the development of a number of more sophisticated models. Broadly, these models can be divided into two classes: signal and compartment models.

Signal models aim to capture the main characteristics of the signal with a limited number of parameters, without necessarily requiring these parameters have a direct biological interpretation. The DT is arguably a signal model; another example of a signal model is diffusion kurtosis imaging (DKI), which relaxes the DT's assumption of Gaussian diffusion by expanding the signal in terms of a cumulant expansion [19]. Some signal models directly try to reconstruct the underlying of white matter fibres that generated the signal, known as the fibre orientation distribution function (fODF). Spherical convolution [20] and its more robust successor, constrained spherical deconvolution [21] are popular examples of this.

Compartment models try to geometrically model the microstructure of biological tissue with the aim of having an easily interpretable set of parameters that directly relate to microstructural features of interest. They typically aim to model the three main compartments in biological brain tissue, intra- and extra-cellular space, and the cerebrospinal fluid (CSF), and can be represented as a taxonomy with different geometric models being used for each compartment [22]. Models such as CHARMED [23] and the ball-and-stick [24] focus on estimating neurite density. AxCaliber [25] and ActiveAx [26] aim to estimate the diameter of axons. The ball-and-rackets model [27] and NODDI [28] are attempts to measure the dispersion of a set of neurites. Models such as VERDICT [29] aim to model tumour cells to provide clinically useful markers of disease.

A commonality between these more advanced models is that they require richer characterisations of the diffusion signal than the DT. They tend to

require higher angular resolution and multiple b-values; these b-values are typically higher (up to b=3000 s mm^{-2}). These richer acquisitions typically take more time and are more demanding for the scanners, increasing the prevalence and severity of image artefacts.

2.3 Image artefacts

We have seen how images are acquired in MRI and how the process can be extended to produce images that are sensitive to local microstructure. The process is susceptible to a number of artefacts that can compromise the quality of these measurements. In this section we discuss the main set of artefacts that occur in DW-MRI. This list is not intended to be exhaustive as MR scanners are extremely complex and the number of things that can go wrong is extremely large; rather this aims to cover the set of artefacts that are frequently encountered in DW-MR images and that, ideally, a realistic DW-MR simulator should be able to reproduce. General references for these artefacts are [30–34].

- Motion

  Patient movement in the scanner is inevitable. If the movement occurs between the acquisition of successive volumes then the imaged object will be offset by a rigid transformation. In reality, patients often move during the acquisition of a single volume. As the volume is acquired as a (typically interleaved) stack of slices, this can cause these slices to incorrectly stack, leading to misalignment that cannot be straightforwardly corrected by a rigid transformation. This can also lead to portions of the brain being missed in an image volume, e.g. translation of the head along the negative slice-select axis between acquisition of slices can lead to some slices never being excited. The opposite may also happen: transformation along the positive slice-select axis can lead to slices being acquired twice. In this case, the slice is re-excited before the expected T_R, leading to the availability of less transverse magnetisation (as discussed in Section 2.1.4) and thus the production of less signal than expected; this

is referred to as the spin-history artefact. In addition to rigid movement of the head, non-rigid motion occurs due to the pulsation of the brain as a result of the cardiac cycle, particularly in regions near the ventricles (cavities filled with CSF) [35]. In addition to physical displacement of the measured signal, both rigid and non-rigid movement can lead to a loss of signal, or dropout.

- Signal dropouts

  Signal dropouts are losses of signal from a slice, caused by rotational movement of the head during the diffusion preparation gradients [36], localised movements due to cardiac pulsation, or even vibration of the scanner [37]. This movement means the second diffusion gradient only partially rephases the sample which causes a large, linear phase shift across the sample. This leads to a translation of the k-space image which is often large enough to translate its centre outside of the FOV, leading to very little signal in the resultant image. The dropout can occur for a whole slice if caused by bulk movement of the head, or just part of a slice if the source of motion is localised.

- Eddy currents

  Eddy currents are produced in conducting components of the scanner by the switching of the magnetic gradients, as described by Faraday's law of induction. These currents in turn lead to undesired, time-varying magnetic fields which add to the imaging gradients and interfere with the spatial encoding of the image. The fields are usually small but the fast-switching gradients used for diffusion sensitisation tend to have longer time periods between the switching on and off of the gradients. This means the induced eddy-currents do not cancel out as much, producing larger fields which leads to more significant artefacts. Such fields, and other undesired magnetic fields in MRI, are sometimes referred to as off-resonance fields because they cause spins to precess at a frequency

different from that expected from a perfect B_0 and a set of linear imaging gradients.

Disruption to imaging gradients is particularly problematic when the EPI sequence is used for spatial localisation. To understand why, we can first consider the gradient in the FE direction. Typically, when a frequency-encoding gradient is applied protons that are separated by 1 mm along the FE axis have a difference in frequency of 500-1000 Hz. A typical gradient inhomogeneity caused by an eddy-current will induce a frequency different of $\sim$100 Hz mm^{-1} [34]. Since these protons are spatially localised according to their frequency, this inhomogeneity could cause an apparent shift in location between the protons of 0.1-0.2 mm; this is considered negligible in MRI. We can think of the protons in the PE direction as being separated by an effective frequency, too, equal to the difference in the amount their phase changes with each EPI 'blip' divided by the time between blips. The amount that the phase may change per blip is constrained: in order to be able to distinguish between protons at the top and bottom of the PE axis, the phase change across the sample must be less than 2π. This constraint means that for a typical scan protons 1 mm apart have an effective frequency difference of $\sim$10 Hz, and so the eddy-current inhomogeneity will cause signal displacements of 10 mm along the phase-encode axis, which is much more significant.

To a first order approximation the EC-induced fields are linear gradients along the $x-$, $y-$ and $z-$ axes. When superposed on the spatial encoding gradients, these gradient lead to skews, scaling and translations of the image along the PE axis. In reality the induced fields are non-linear and thus cause non-linear distortion [38]. Furthermore these currents temporally decay; if this happens significantly over the course of a slice acquisition this can lead to blurring of the resultant images [39].

- Susceptibility

Materials have different responses to being placed in magnetic fields. The magnetisation of an object $\vec{M}$ can be related to the applied field, $\vec{B}$ by a scalar quantity called the material's susceptibility, χ:

$$\vec{M} = \chi \vec{B} \tag{2.41}$$

The magnetisation $\vec{M}$ in turn produces its own additional magnetic field, which sums with the applied field to create a total field. When two objects with different susceptibilities are placed in a magnetic field the resultant field around the boundary is complicated and non-homogeneous. Air and tissue have different susceptibilities, so these fields are created around the sinuses and ear canals when a head is placed in the scanner. These fields affect the acquired image in three ways. Firstly, tissue can be incorrectly excited by the slice-select gradient, causing it to appear in the wrong slice. Secondly, the frequency alterations cause geometric distortions along the PE axis. These differ from eddy-current distortions in that the inhomogeneities and thus the distortions are spatially localised, rather than varying slowly across the image. Finally, if the field varies significantly compared to the length-scale of a voxel, local dephasing of the transverse magnetization can occur. This causes signal loss in GE sequences, but in SE sequences the spins are rephased and signal loss does not occur. The susceptibility field can change over time, for example due to changes in the shape of the chest cavity during respiration, or due to rotation of the head about an axis that is not parallel to B_0 [40].

- Noise

  The noise in the complex MR signal is normally distributed and we typically take the magnitude of these complex images. For a single coil this causes the noise to have a Rice distribution [41], but in general when signal is combined from multiple coils the data has a noncentral-χ distribution [42]. Both of these distributions lead to a bias in the expectation

of the signal value compared to the real value. In diffusion MRI high diffusivity is reflected by low signal in an image; this bias can increase the measured signal for high diffusivity leading to this diffusivity to be underestimated, which can cause a reduction in measured anisotropy [43].

- Gibbs ringing

  As we have seen in Section 2.1.4, k-space is truncated in practice. Higher values of k-space describe higher frequency components and without these the Fourier series cannot accurately represent high-frequency image components such as sharp boundaries. This leads to a series of light and dark lines parallel to this boundary, representing the overshoot and undershoot from the truncated Fourier reconstruction.

- Ghosting

  The Nyquist artefact, or ghosting, is caused by mismatches between alternately acquired lines of k-space in the EPI sequence. The centre of each line is an echo caused by the rephasing action of the FE gradient. Because alternate lines are acquired when traversing opposite directions of k-space, any small delay introduced into an echo will cause offsets between the echoes in odd and even lines. These delays can be caused by eddy-currents, gradient coil heating, inaccurate timing of the sampling relative to the applied gradients, and patient motion. For a single-shot EPI sequence the offset causes displacements of intensity by half of the field-of-view (FOV) in the PE direction of the reconstructed image, which are known as ghosts.

- RF artefacts

  In Section 2.1.2 we saw that the flip-angle was controlled by the magnitude of the applied RF field, B_1^e and the time it was applied for. Non-uniformity in $B_1^e(t)$ across the brain causes non-uniform flip angles that reduce the available transverse magnetization, M_{xy}, and thus signal in

certain areas, leading to bright and dark patches in the resulting images. This problem is exacerbated at high field imaging (e.g. B_0=7 T) because the shorter RF wavelengths needed to achieve the relevant Larmor frequency are more similar to the size of the head which increases wave interference. Similarly, inhomogeneities in the RF receive coil's sensitivity profiles cause spatially varying signal in the resultant image, often noticeable as darker patches in regions of the brain. Coil sensitivity tends to be highest around the edges of the brain, so these darker patches tend to occur in the middle. Another potential artefact is caused by bursts of RF noise, which lead to high intensity spikes in k-space. These manifest themselves as stripes across the image, with frequency and orientation determined by the location in k-space that the noise was recorded.

- Chemical shift

  The resonant frequency of nuclei is affected by their local molecular environment. For example, fat protons are shielded from external magnetic fields by their surrounding electrons, causing them to precess at a slightly lower frequency than protons in water. This causes the fat signal to be spatially shifted along the PE axis in EPI acquisitions.

2.4 Simulation systems

2.4.1 MRI

The field of MR simulation is extremely large and well developed. Every simulator is a simplification of the physical environment, and the choices and assumptions made in designing a simulator are typically motivated by the intended application of the system. For example, one sub-field focuses on RF simulation to assist the design of pulses with more desirable properties, such as sharper spatial profiles and decreased power deposition [44–46]. Another example is NMR simulation, which tends to focus on detailed modelling of quantum mechanical effects, such as spin-coupling, which are important for generating realistic chemical spectra [47, 48]. In this survey we focus our

attention on systems that aim to simulate the MR acquisition process and produce realistic, full-brain MR images, along with the artefacts that are often present in these images.

The simplest tools for MR simulation use equations that relate parameters such as T_1, T_2 to spatial signal intensity [49, 50]. By applying these equations to an input that describes the spatial variation of these parameters in the brain, an output MR image may be produced. Whilst computationally very efficient, this approach fails to model the process of image acquisition, involving the recording of a signal in k-space that is subsequently Fourier transformed to produce a spatial image, which is vital if realistic images and their artefacts are to be simulated.

An early method that made use of k-space signals was the 'k-space formalism' [51]. It involved Fourier transforming a high-resolution map of tissue type and proton density to produce a k-space signal, then simulating the acquisition for a given pulse sequence by selecting the relevant elements from this dense k-space signal. Relaxation parameters and simple artefacts such as eddy currents are taken into account during the k-space selection. This method is fast but precludes the inclusion of more complex artefacts, such as movement of the object during the acquisition of a slice.

Another popular method for simulation analytically relates properties of an input object to the imaged signal. Early work [52, 53] focused on finding a closed-form expression for the k-space signal generated from ellipsoids such as those in the famous Shepp-Logan phantom [54]. Guerquin et al [55] extended this work by finding the analytical relationship between piecewise polynomial functions and their k-space signal, while Ngo [56] developed an expression for arbitrary polyhedra, both of which enable the simulation of more realistic phantoms. The analytical method allows for very realistic representations of scanner objects, but they don't directly model the physics of the MR acquisition process, i.e. the application of gradients to manipulate magnetic moments, which prevents the inclusion of a number of artefacts such as move-

ment and spin-history effects. In order to realistically model such artefacts a simulator must solve the Bloch equations.

The simplest Bloch-based simulators solve the equations once for each tissue type — white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) — and combine the resultant signals in each voxel using proportions determined by a tissue template, as seen in the extremely popular BrainWeb tool [57, 58]. This method is fast because the equations need only be solved once per tissue type. However, the method means each spin experiences the same magnetic field history which prevents artefacts such as B_0 inhomogeneities being modelled.

Methods that solve the Bloch equations over a grid with spatially varying values of ρ, T_1 and T_2, representing the object being imaged, allow for the most realistic simulations. A key challenge with such approaches is that a large number of spins per voxel must be simulated to accurately capture effects such as intra-voxel dephasing or spin-echoes [59]. Early examples of this by [60, 61] used only a few spins per voxel, preventing the simulation of dephasing effects. Olsson [62] addressed this by using a model-based approach to correctly account for dephasing amongst a limited number of spins, enabling the modelling of T_2^* dephasing effects. These approaches are all 1D or 2D; the SIMRI project [63] extended simulations to 3D. SIMRI only simulates a few isochromats per voxel, and implements the model based-approach to dephasing similar to Olsson and also proposes a model based-approach to handle spin-echoes, essentially by modifying the magnitude of the magnetisation vector as dictated by the spin-echo. SIMRI enables reproduction of many of the artefacts discussed in Section 2.3, including ECs, susceptibility, Gibbs ringing, ghosting and chemical shift. SIMRI requires a susceptibility-induced off-resonance field to be specified — Yoder proposed a method to calculate realistic susceptibility-induced fields from a known input object and incorporate these into simulations [64]. A notable feature lacking from all these Bloch-based simulators is support for complex motion. The POSSUM simu-

lator allows for this, providing support for arbitrary movement of the object throughout the acquisition of an image [65]. Like Yoder, POSSUM can calculate susceptibility-fields from the input object, but it further extends this to account for the changing of the susceptibility-induced field with object movement, as well as being able to account for time-varying off-resonance fields, such as those caused by the patient breathing [66].

These solvers are computationally expensive, and recently emphasis has been placed on parallelised implementations that enable the simulation of many spins per voxel. This means a model-based approach for T_2 dephasing and spin-echo rephasing is no longer required. The computational speed-ups also enable the addressing of another weakness of all the previously mentioned frameworks: the reliance of analytically-derived descriptions for the actions of RF pulses on isochromats. The JEMRIS simulator [67] enables the Bloch equations to be solved numerically for arbitrary RF pulses for which analytical solutions may not exist; as well as providing support for multi-coil transmission and reception. MRISIMUL offers similar features to JEMRIS but allows for GPU computations that can offer speedups of up to two orders of magnitude compared with code executed on a CPU [68]. MRiLab [69, 70] also relies on GPUs to accelerate computation, and provides support for modelling protons residing in different compartments within a voxel and exchanging between these compartments, enabling the simulation of MR techniques such as magnetisation transfer [71].

2.4.2 Diffusion MRI

There has been substantial work on the simulation of diffusion MR signals inside a single voxel, without the need for spatial encoding. These approaches vary in flexibility. The simplest generate this signal using a model that assumes a particular underlying microstructure and pulse sequence (most often PGSE) [72–74]. Approaches such as Callaghan's matrix formulation [75] enable DW-MR signal to be generated for simple geometries but more complex pulse sequences [76]. The most flexible use a Monte-Carlo approach to simu-

late the diffusion of particles within a single voxel, for arbitrary microstructural environments and pulse sequence [77–79].

Work on simulation of full-brain images with diffusion contrast is much more limited. The simplest methods simply assign a single value of diffusion-weighting contrast to each of the three tissue types, then combines these signals using probabilistic segmentations to form an image [80, 81]. Not only do these methods fails to represent the complexity of the diffusion signal, which cannot be simplified to a single intensity per tissue type, but they do not simulate the process of MR acquisition. Some work has combined this simple model of diffusion contrast with a simulation of the full process of MR acquisition [82].

Much of the recent work in DW-MR simulation has focused on providing data that can help assess tractography algorithms [83]. Leemans [84] provided a mathematical description of the diffusion signal by combing a physical description of the spatial extent of a WM bundle with the DT model. Close provided a tool to generate random structures of densely packed fibre bundles, using the DT model to simulate the signal from these bundles [85]. The Phantomas software tool used the CHARMED model [23] to generate WM signal and incorporated MR relaxation parameters [86]; [87] developed a similar simulator but used multiple DTs to model crossing fibres. All these tractography tools produce the signal directly in image space. The FiberFox simulator is able to produce full-brain images based on an input tractogram [88] and is able to simulate the process of image acquisition in k-space, though it is simple when compared to the MRI simulations discussed in Section 2.4.1: it does not simulate the process of RF excitation and slice-selection, and is unable to model the movement of the object during image acquisition. FiberFox uses a zeppelin model of diffusion in the WM (a DT with the two eigenvalues perpendicular to the principal diffusion direction constrained to be equal) and an isotropic tensor to model the rest of the tissue. The D-BRAIN phantom uses similar k-space acquisition, but extends the diffusion model to contain two different isotropic tensors to model both the GM and CSF [89].

These models rely on an underlying description of WM structures to generate the signal, which means they are unable to produce full-brain images that realistically model the signal in the GM and CSF. Work by Du has used Monte-Carlo methods [90] to produce simulations with realistic full-brain contrast, but does not faithfully model the physics of MR acquisition. [91] tries to bridge this gap by providing both realistic diffusion-contrast from a Monte-Carlo simulation coupled by a modelling of the MR acquisition process, but the computation time required to simulate full-brain volumes with this approach is prohibitive.

2.4.3 Critical analysis

In order to be suitable for the testing of acquisition and post-processing techniques, a DW-MR simulator must be able to reproduce the artefacts detailed in Section 2.3, as these are the problems that these techniques typically try to address. Furthermore, it must be able to produce realistic DW-MR contrast, as it is the large variations in contrast across DW-MR datasets that can make their processing so challenging. Whilst DW-MR simulation systems exist, none are able to produce realistic enough datasets, complete with artefacts, to be used for these purposes. Specifically, existing systems all exhibit at least one of the following limitations:

1. The failure to model the physics of DW-MR image acquisition

   MR acquisition involves the application of a sequence of magnetic fields across a sample, producing a signal in 'frequency space' that is then transformed to generate an image of the sample. Many of the artefacts found in MRI are introduced during acquisition and/or transformation, and failing to faithfully model this process can preclude the realistic simulation of many artefacts.

2. The failure to produce realistic diffusion-weighted contrast.

   The large variations in contrast between volumes in a DW-MR dataset are part of the reason that correcting artefacts can be so challenging,

and so these variations must be captured if we are to fairly assess post-processing techniques. Furthermore, some post-processing techniques use models of diffusion contrast to form the basis of their correction [92, 93] and so the use of a particular model to simulate the diffusion contrast could cause some circularity that would bias their assessment.

3. The inability to produce full-brain images

   Many of the existing simulation systems for DW-MR produce data for just the brain's WM [88], for the purposes of evaluating WM tracking methods, or even just in a single voxel [77], for validating models of diffusion. In order to be useful for testing correction techniques, a simulator needs to be able to produce full brain images.

Chapter 3

A simulation system for diffusion-weighted MRI

3.1 Overview

This chapter introduces a simulator designed to address the limitations of existing frameworks discussed in Section 2.4.3. It is able to produce realistic full-brain contrast and models the physics of MR acquisition, enabling many of the artefacts discussed in Section 2.3 to be simulated. A novel framework for evaluating post-processing techniques is also introduced, with a view towards applying the framework to assessing techniques for correcting motion and eddy-current artefacts in Chapter 4.

3.1.1 Research dissemination

An early prototype of the simulation framework was presented at IPMI. The framework was extended, addressing several limitations of this prototype, and published in NeuroImage. Code for the simulator has been made available at `https://github.com/marksgraham/DW-POSSUM`.

- A simulation framework for quantitative validation of artefact correction in diffusion MRI. MS Graham, I Drobnjak, H Zhang. In *International Conference on Information Processing in Medical Imaging*, volume 9123 of *Lecture Notes in Computer Science*, pages 638-649, 2015.

- Realistic simulation of artefacts in diffusion MRI for validating post-processing correction techniques. MS Graham, I Drobnjak, H Zhang. In *NeuroImage* 125, 1079-1094, 2016.

3.2 Introduction

As we have seen in Section 2.3, DW-MR data can include a large number of artefacts. Techniques for dealing with these artefacts can broadly be divided into those implemented at acquisition time [94–96], involving either some modification to the acquisition process or the collection of supplementary data, and post-processing methods implemented after acquisition time [97, 98]. Post-processing techniques are very popular, as they have several advantages: they can be applied retrospectively to already acquired data, a user can revert to the original data if the technique does not work as hoped, and they don't require additional scan-time, which is often expensive.

The literature contains a vast body of post-processing techniques and software packages for correcting artefacts in DW-MRI [96, 98, 99]. Ideally their corrections would be validated by comparison to the ground truth, i.e. a map of the spatial deformations caused by the artefacts, but these cannot be obtained for real data. As a result the literature relies on either qualitative visual assessments of image alignment [100, 101], or quantitative assessments of surrogate measures of alignment, such as tract length [102], fractional anisotropy (FA) values [103] or reduced residuals from fits to microstructural models [92, 104]. The lack of an objective ground truth means existing techniques cannot be systematically assessed, preventing end-users from making an informed choice. The development of new methods is also hindered, as any improvements over existing ones are difficult to demonstrate.

A simulation framework is able to provide a direct, quantitative way of assessing post-processing techniques. Such a framework must have two key features. Firstly, it must be able to produce realistic DW-MR data, along with artefacts. It is particularly important that the simulated data contains

a realistic representation of diffusion contrast, because it is the large contrast variations with a DW-MR dataset that makes it uniquely difficult to process. This is because many processing techniques rely on registration, which makes use of image similarity. Images acquired with very different b-values do not look similar at all, making registration challenging. Secondly, the simulation framework must provide a quantitative means of assessing the success of a post-processing algorithm. As post-processing algorithms try to estimate a displacement field to resample the data from distorted to undistorted space, the most direct way of assessing a technique is to assess the errors in the predicted displacement field. Section 3.3 describes the development of a simulation framework that possesses both of these key features.

3.3 Methods

In this section we describe our simulation framework for producing realistic DW-MR images along with the displacement fields that map them into undistorted space. Section 3.3.1 provides an overview of the framework. Details of the implementation of the framework follow in 3.3.2, 3.3.3, and 3.3.4, where we describe how we produce both the DWIs, how artefacts are incorporated and how the ground truth displacement fields are generated.

3.3.1 Framework overview

As seen in the discussion of simulation systems in Sections 2.4.1, MRI simulators exist which are able to faithfully model MR physics, producing the artefacts that arise during data acquisition, but none exist which also produce DW-MR contrast. In Section 2.4.2 we saw the available DW-MR simulators tend to have simplified MR physics and do not provide realistic, full-brain diffusion contrast. We have addressed these issues by combining a fully-featured MR simulator with a model-free representation of diffusion obtained directly from a real dataset.

Figure 3.1 presents a conceptual overview of how an MR simulator may be adapted to provide realistic diffusion contrast. The framework takes four

main inputs. The first is a geometric object that specifies the proton density and location of WM, GM and CSF along with their T1 and T2 values. The second is a representation of diffusion-weighting. The third is a pulsed-gradient spin-echo (PGSE) sequence, detailing RF pulses and gradients. The first two inputs are combined with diffusion parameters extracted from the third (direction and magnitude of diffusion weighting) to produce a geometric object with its proton density reduced by a diffusion attenuation factor - this is how diffusion-weighting is introduced into the simulated DWIs. The attenuation factor is defined as $A_b\left(\hat{b}\right) = S_b\left(\hat{b}\right)/S_0$, the ratio between the diffusion-weighted signal $\left(S_b\left(\hat{b}\right)\right)$ and the signal without diffusion weighting (S_0) in a given voxel, defined for a b-value, b, and direction of diffusion weighting, $\hat{b}$. The attenuation factor is a dimensionless quantity that takes values between 0 and 1. The PGSE sequence is converted to a standard echo-planar imaging (EPI) sequence for simulation, so the diffusion attenuation is introduced solely through the input object. The fourth input is any details that will lead to the simulation of artefacts, such as motion parameters. The effects of eddy currents arising from the diffusion gradients are included in the EPI sequence that is passed to the simulator.

The framework creates two outputs. The first is a DWI. The MR simulator takes the attenuated object, pulse sequence and details pertaining to artefacts, and solves Bloch's and Maxwell's equations at each point in the object, summing the resultant signal in order to generate the k-space measurements. This is Fourier transformed to produce the output DWI. The second output is a displacement field that describes the mapping of this DWI from a distorted to undistorted space.

3.3.2 Producing DWIs

We chose to use POSSUM to provide the core MR simulation [66, 105]. POSSUM has a number of features that make it ideal for use here. As discussed in Section 2.4.1, its a fully featured MR simulator that solves the Bloch equations for an isochromat at each voxel in the input object over a 3D grid, enabling

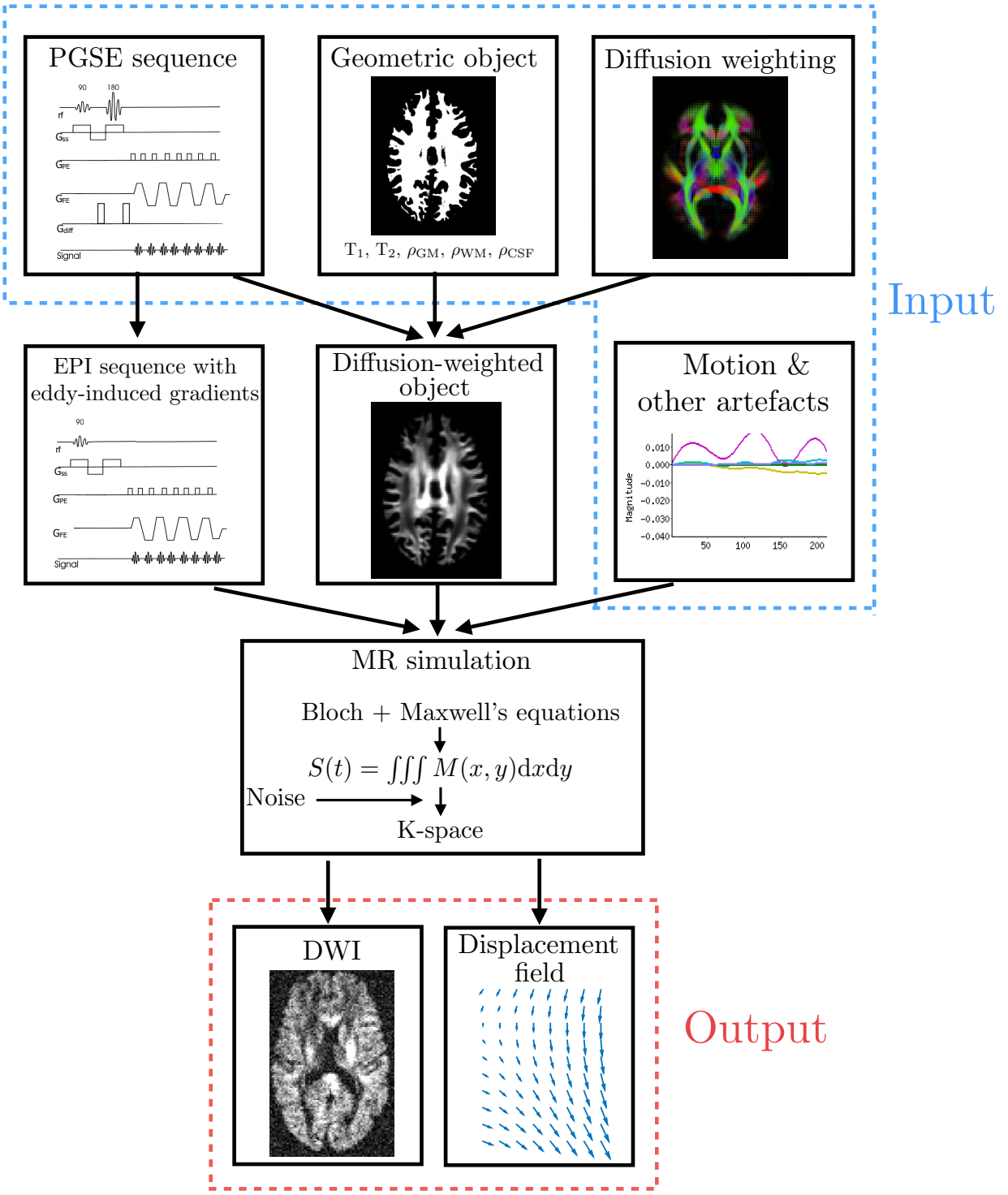


Figure 3.1: The pipeline for simulating DWIs. Details of DWI, image artefact and displacement field generation are in Sections 3.3.2, 3.3.3, and 3.3.4.

it to realistically model many of the sources of artefacts in MRI. A key advantage POSSUM has over other advanced simulators discussed is its strong support for modelling motion artefacts, including motion during read-out periods. It also has the ability to model the susceptibility-movement interactions, a feature not offered by any other simulator. POSSUM is designed for the simulation of MRI/fMRI data, so a number of extensions were made to enable DW-MRI simulations. These include the introduction of diffusion weighting and the inclusion of spin-echo contrast.

The input object and representation of diffusion weighted used for simulation were generated from data from a single subject of the WU-Minn Human Connectome Project (HCP) dataset [106]. This data was chosen because of its extremely high quality, providing high-resolution structural images and diffusion data, enabling the generation of high-quality inputs. A full-brain segmentation was used as POSSUM's geometric object input, created with T1- and T2-weighted images from the HCP subject using FSL's FAST [107]. The representation of diffusion weighting was achieved using a voxel-wise spherical harmonic (SH) fit to the subject's diffusion data [108]. An order $n = 8$ fit was used to fully capture the angular information available in these datasets [109], with constraints placed on the coefficients to ensure the signal is real and exhibits antipodal symmetry. Separate fits were performed on the b=1000 s/mm^2 and b=2000 s/mm^2 shells. These SH fits can be viewed as an interpolation function in q-space, enabling us to predict the diffusion-weighting at each voxel along any gradient direction, $\hat{b}$. In general, the fit to the b=1000 s/mm^2 shell was used to predict signal attenuation for simulated DWIs with b=1000 s/mm^2, and the b=2000 s/mm^2 fit to predict attenuation for simulations at b=2000 s/mm^2.

For some experiments it is desirable to simulate weighting with $b <$ 1000 s/mm^2. A mono-exponential assumption was used to predict the signal attenuation in these cases. Along the direction $\hat{b}$, the attenuation for a value $b = s$, $A_s(\hat{b})$ can be obtained from the predicted attenuation at b=1000 s/mm^2,

Tissue	T_1/ms	T_2/ms	ρ
Grey matter	1331	75	0.86
White matter	832	70	0.77
CSF	3700	500	1

Table 3.1: Tissue parameters used for the DW-MR simulations. Proton density ρ is in arbitrary units.

$A_{1000}(\hat{b})$:

$$A_s\left(\hat{b}\right) = \exp\left(\frac{s}{1000}\ln\left(A_{1000}(\hat{b})\right)\right) \tag{3.1}$$

The prediction of diffusion-weighting at this new value assumes Gaussian diffusion, this assumption is reasonable in practice for $b < 1000\,\mathrm{s/mm^2}$ [110].

POSSUM simulates gradient-echo EPI sequences, and is not able to simulate the spin-echo typically used in DW-MR. To enable the production of images with T_2 contrast, we used a GE EPI sequence and replace the default tissue-specific T_2* values with their corresponding T_2 values: 75 ms and 70 ms in the GM and WM, respectively. These values were obtained by adjusting the average estimates from the literature [6] until our simulations best matched the contrast seen in a real dataset. The full set of tissue parameters used are shown in Table 3.1. The inability of POSSUM to simulate spin-echo sequences is addressed in Chapter 5.

3.3.3 Incorporating artefacts

The framework is able to reproduce many of the key DW-MR artefacts that are discussed in Section 2.3. Motion, noise, Gibbs ringing, ghosting, and chemical shift artefacts have all been demonstrated in POSSUM [105], and can be readily included in our simulations. One limitation of the movement model is that it describes the rigid motion of the whole input object, and so non-rigid effects such as pulsatile movement cannot be simulated.

One key artefact that we have added to the framework is eddy-currents. Eddy-induced gradients were added to the EPI pulse sequence using the spatially linear model in [82], by superposing a sum of decaying exponentials on

each gradient field. In this model, the eddy-currents are assumed to only arise from the gradients that provide diffusion weighting. The EC-induced gradients act along the same direction as the diffusion gradient. Mathematically, the EC-induced gradient along each axis (G_x^{E},G_y^{E},G_z^{E}) is described by:

$$G_{x,y,z}^{\mathrm{E}} = \sum_i \pm\varepsilon\, G_{x,y,z}^{\mathrm{diff}} \exp\left[-(t-t_i)/\tau\right] \tag{3.2}$$

where t_i corresponds to the time each diffusion gradient is turned on or off (determined by the pulse width δ and diffusion time Δ, obtained from the input PGSE sequence), τ is the decay time, $G_{x,y,z}^{\mathrm{diff}}$ is the strength of the diffusion-weighted gradient along x, y or z, ε is a constant determining the relationship between the strength of eddy and diffusion gradients and a + or - is selected depending on whether the gradient is being turned on or off.

We performed simulations with a maximum value of $G^{\mathrm{diff}} = 40\,\mathrm{mT\ m^{-1}}$, and selected $\varepsilon = 0.009$ and τ=100 ms to represent typical values found in a clinical scanner [111].

Whilst POSSUM supports the susceptibility artefact, the lack of a spin-echo pulse sequences means the artefact causes both geometric distortion and signal loss, rather than just geometric distortion. This limitation is addressed in Chapter 5. The framework does not support signal dropouts — this is addressed in Chapter 7.

3.3.4 Ground truth displacement fields

The framework generates the ground truth displacement field for each DWI, detailing its mapping from distorted to undistorted space, as comparison of predicted and ground truth spatial displacement fields is the most direct way to evaluate post-processing methods that aim to estimate a corrective displacement field. Here we describe how we obtain the combined displacement field that corrects for off-resonance and motion artefacts in our simulations. In the following we calculate them for eddy-currents and inter-volume rigid motion artefacts, with a view towards the assessment of EC and motion correction

schemes in the next chapter, but the equations shown can be extended to other off-resonance artefacts straightforwardly.

The EC displacement field can be calculated from a knowledge of the off-resonance frequency that the EC give rise to. In the presence of any off-resonance conditions, the spin frequency can be written:

$$\omega(\mathbf{r},t) = \omega_0 + f(\mathbf{r},t) \tag{3.3}$$

where ω_0 is the system frequency and $f(\mathbf{r},t)$ is the off-resonance term, that may vary with spatial location $\mathbf{r}$ and time t. When the off-resonance term is time-invariant, i.e. $f(\mathbf{r},t) = f(\mathbf{r})$, we can express the voxel displacement field caused by these off-resonance effects, $\boldsymbol{\psi}^O$, as [112]:

$$\boldsymbol{\psi}^O(\mathbf{r}) = t_s N f(\mathbf{r}) \hat{\mathbf{p}} \tag{3.4}$$

where t_s is the echo spacing (in seconds), N is the number of phase-encode lines, $\hat{\mathbf{p}}$ is a dimensionless vector pointing in the phase-encoding direction, $f(\mathbf{r})$ is defined in Hz.

In the case of linear EC gradients assumed in our simulations, we can express the off-resonance term:

$$f(\mathbf{r},\mathbf{t}) = \gamma \mathbf{G}^E(t) \cdot \mathbf{r} \tag{3.5}$$

where $\mathbf{G}^E = \left(G_x^E, G_y^E, G_z^E\right)$ and γ is the gyromagnetic ratio. These gradients are time-varying according to Eq. 3.2. We evaluate this off-resonance term at the centre of k-space, $t = T_E$, as this dominates the gross structure of the image:

$$f(\mathbf{r}) = \gamma \mathbf{G}^E(T_E) \cdot \mathbf{r} \tag{3.6}$$

We can now combine Eqs. 3.4 and 3.6 to find the displacement field:

$$\boldsymbol{\psi}^O(\mathbf{r}) = \gamma t_s N \left(\mathbf{G}^E(T_E) \cdot \mathbf{r}\right) \hat{\mathbf{p}} \tag{3.7}$$

The motion displacement field, $\boldsymbol{\psi}^M(\mathbf{r})$ is straightforwardly obtained from the rigid transformation $\mathbf{R}$:

$$\boldsymbol{\psi}^M(\mathbf{r}) = \mathbf{R}\mathbf{r} - \mathbf{r} \tag{3.8}$$

We can obtain the total displacement field, $\boldsymbol{\psi}^T(\mathbf{r})$, from a composition of the two fields:

$$\boldsymbol{\psi}^T(\mathbf{r}) = \boldsymbol{\psi}^O \circ \boldsymbol{\psi}^M(\mathbf{r}) \tag{3.9}$$

The ground truth displacement field gives us our mapping from undistorted to distorted space. An artefact correction method predicts a displacement field, $\boldsymbol{\psi}^P(\mathbf{r})$, that attempts to map a volume from distorted to undistorted space. We define the error field, $\boldsymbol{\psi}^E(\mathbf{r})$, as the displacement field that describes the mapping of each voxel in undistorted space into corrected space:

$$\boldsymbol{\psi}^E(\mathbf{r}) = \boldsymbol{\psi}^P \circ \boldsymbol{\psi}^T(\mathbf{r}) \tag{3.10}$$

A zero error displacement field indicates perfect correction has been achieved. In this work we make use of these error fields to quantitatively assess the effectiveness of artefact correction schemes.

3.4 Simulator demonstration

In this section we assess the ability of our framework to produce DW-MR data containing realistic contrast and demonstrate the inclusion of EC artefacts in the framework.

3.4.1 Image contrast

We first assess how well the simulated images capture the most important characteristics of real images. POSSUM has been shown to provide realistic MR simulation without diffusion weighting [66, 105], so here we focus on assessing the simulation of diffusion weighting. In the case of DW-MR the key characteristic is the variation in contrast as the strength and direction of diffusion

weighting changes. To test this we compared a real and simulated dataset with identical parameters: a 3T scanner with three shells, b=300/700/2000 s/mm^2, TR/TE = 7500/109 ms, voxel size 2.5 mm isotropic. Figure 3.2a compares the changes in contrast with varying b-value. Figure 3.2b compares changes in contrast with varying direction of b-vector.

The results demonstrate that our simulation framework is able to capture both the increased attenuation with b-value and the variation in contrast with varying b-vector that are present in real data. Note that the real data and the data used to generate the input object for simulation were obtained from different subjects, so the slices are not perfectly matched, which may account for some of the differences in the appearance of WM tracts and the differences in the overall shape of the brain.

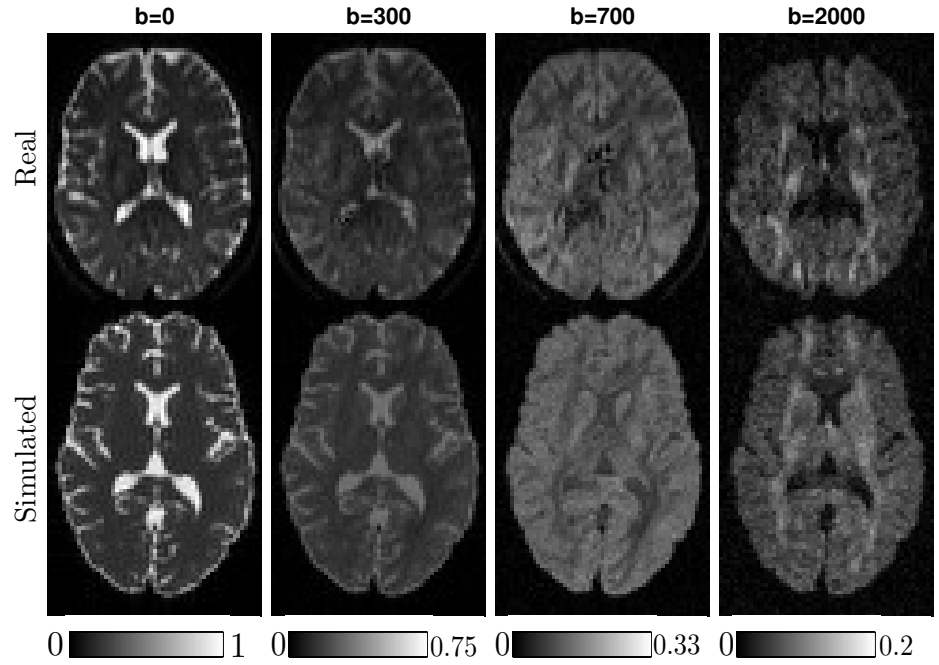
We also compare our simulations with the current state-of-the-art in model-based approaches for simulating DW-MR data, the FiberFox simulator in Figure 3.3. As discussed in Section 2.4.2, FiberFox uses a zeppelin to model diffusion in the WM and isotropic tensors to model signal in the GM and CSF.

3.4.2 Eddy-current artefacts

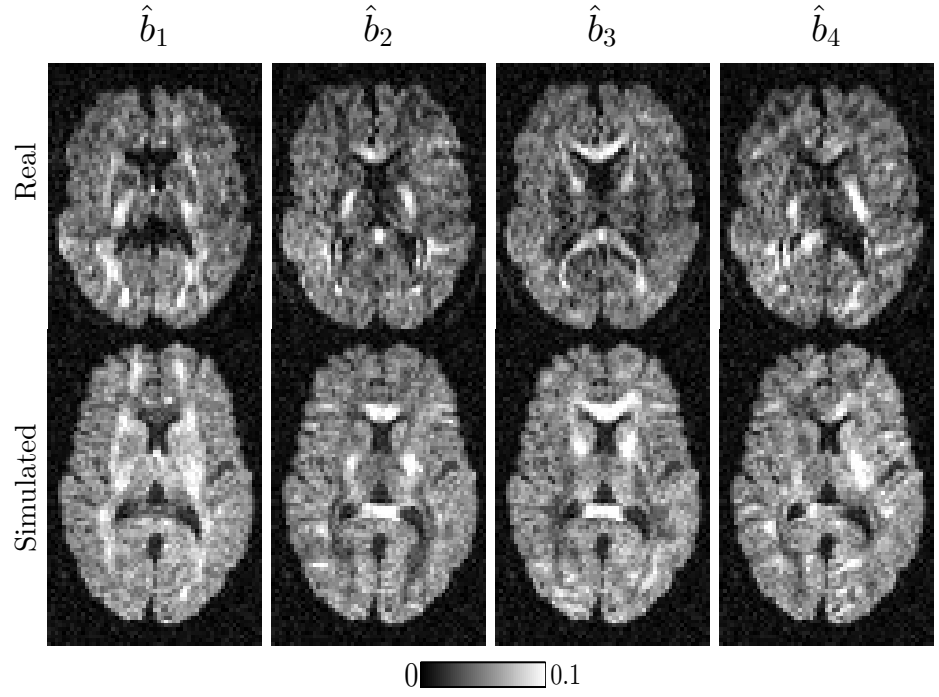
Finally we show that we are able to reproduce EC distortions seen in real data. For the case of linear ECs, EC-induced gradients along the x-, y- and z-axes should lead to shears, scaling and translations in the images respectively (assuming here that y is the PE direction and z is slice-select). In our model, an applied diffusion gradient in a given direction gives rise to linear EC gradients in the same direction, so we expect to see pure shears, scaling and translation in the three examples in Fig 3.4, which we do.

3.5 Discussion

We have presented a framework that allows for the simulation of full-brain DW-MR datasets with artefacts, along with their corresponding ground truth displacement fields, providing an objective and quantitative means of assessing



(a) Variation in contrast with respect to b-value. Both real and simulated datasets normalised against their respective b=0 images. The direction of diffusion weighing is the same for both datasets.



(b) Variation in contrast with respect to direction of diffusion weighting. Each column represents an image acquired at b=2000 s mm^{-2} with a different b-vector, $\hat{b}_1$-$\hat{b}_4$. Real and simulated datasets are normalised against their respective b=0 images and shown on one intensity scale.

Figure 3.2: Comparison of real and simulated data.

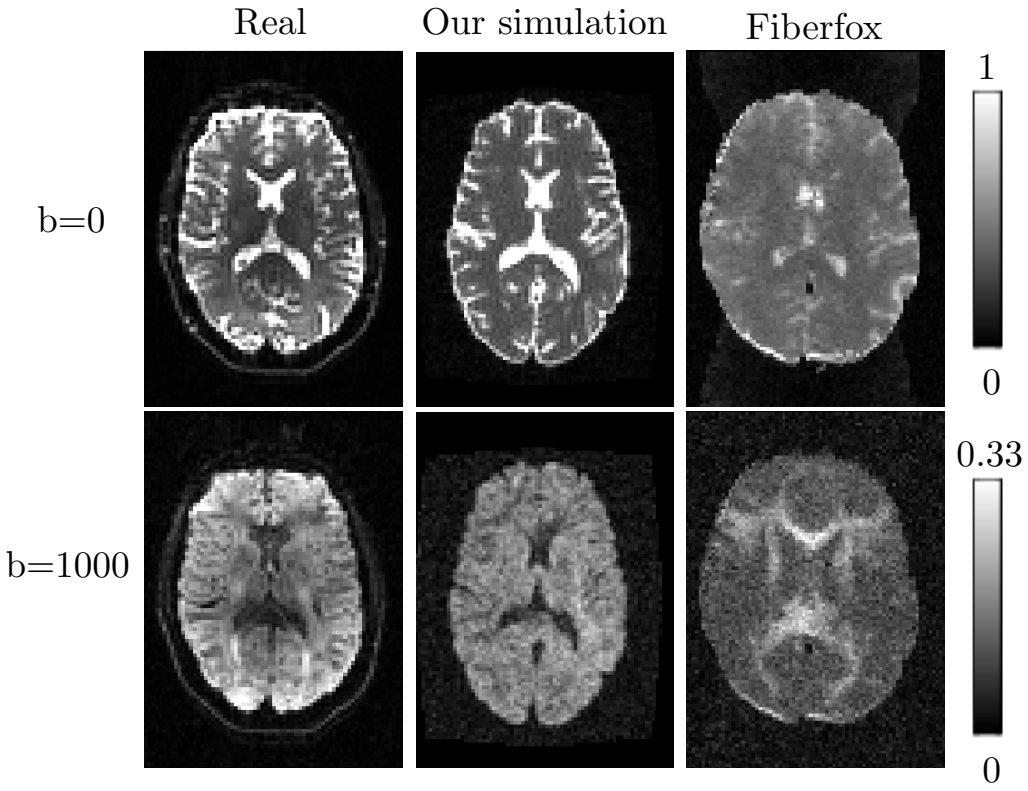


Figure 3.3: Comparison of our simulated data to state-of-the-art. Fiberfox data taken from the 2015 ISMRM tractography challenge.

post-processing techniques. Our framework combines two elements necessary for the simulation of realistic full-brain DWIs. Firstly, our simulated images are able to provide a realistic representation of the contrast differences found across DW-MR datasets. We achieve this using a model-free approach that obtains the signal from real data. Previous simulations have used a single representative mean diffusivity or diffusion tensor for each tissue type to provide diffusion-weighting [80, 82], which leads to vastly oversimplified contrast, or have used underlying WM structures to generate the signal [85, 86, 88], preventing realistic contrast from being achieved in non-WM regions of the brain. The second element is the modelling of the MR acquisition process. Without simulating the full image generation process certain artefacts cannot be introduced.

We chose to use POSSUM as the core MR simulator. Whilst POSSUM is

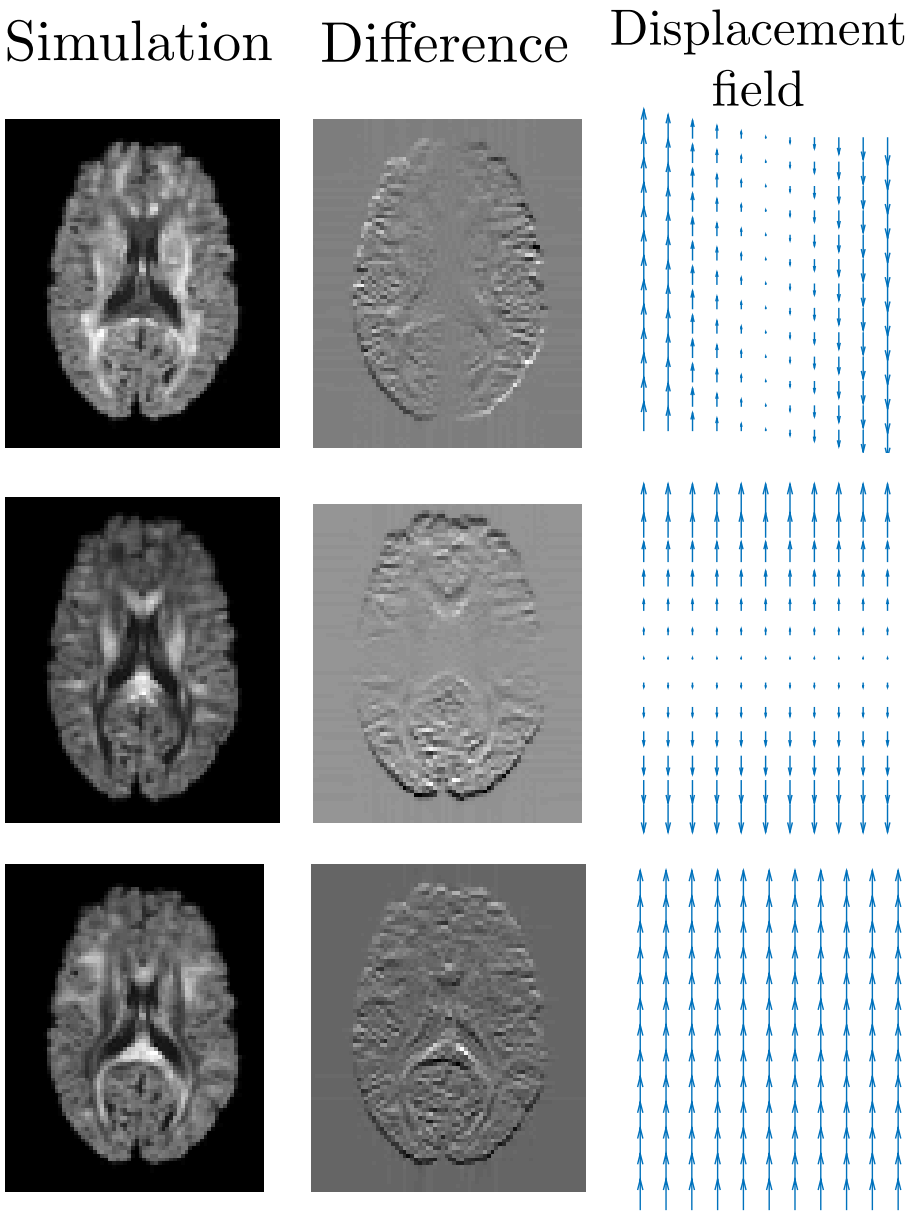


Figure 3.4: Demonstration of DWIs simulated with EC artefacts. Left column are DWIs simulated at b=2000 s/mm^2, with gradients pointing along the x-, y- and z-axes. Middle column shows difference images between these DWIs and their counterparts simulated without EC distortion. Right column shows displacement fields, representing the transformations from undistorted to distorted space. Displacement fields downsampled for clarity.

a powerful tool, some limitations must be acknowledged. Simulations assume a single coil with uniform sensitivity. This means we cannot currently model transmit and receive bias fields, and the noise does not spatially vary across the brain. This also means that parallel imaging techniques such as SENSE [113] or GRAPPA [114] cannot be implemented. POSSUM solves the Bloch equations over a 3D grid for one isochromat at each voxel in the input object. This is not dense enough to enable 'true' modelling of the T_2 dephasing: in the simulations used here, the input grid had resolution $1 \times 1 \times 0.5$mm, and the output simulations were 2.5 mm isotropic. This means POSSUM uses a model-based approach to model this dephasing, trading a small amount of realism for decreased computation time. POSSUM also assumes instantaneous RF pulses, allowing the effect of the pulse to be modelled as a rotation around a specified axis; a slice profile can be specified, which controls the excitation angles across the slice. Fully solving the Bloch equation's over the excitation period would allow for more detailed modelling of the RF pulses

Using a real dataset to determine the signal attenuation allows us to achieve more realistic contrast throughout the brain than is currently possible with model-based techniques, but this approach has two limitations. The spherical harmonic approach means we can only predict attenuation for b-values we already have data for — in our HCP dataset this is b=1000, 2000 and 3000 s/mm^2. Whilst we were able to predict attenuation at b <1000 s/mm^2 using data from the b= 1000 s/mm^2 shell and a mono-exponential assumption, this is only appropriate at low b-values and could not be used to predict attenuation at higher b-values. Employing an approach such as MAP-MRI [115] could allow us to simulate datasets over a wider range of b-values. Furthermore we used a single value of T_1 and T_2 for each tissue type, whilst they have been shown to vary spatially across the brain [6]. Failure to model this will account for the lack of subtlety in the contrast compared to real data, noticeable in e.g. the comparison at b=700 s/mm^2 in Fig 3.2a. Further work could incorporate spatial maps that will allow for variations within tissue types

Furthermore, any artefacts such as ghosting, ringing and noise that are present in the input data will be projected into our simulations. We tried to minimise this effect by using data from the HCP, which is acquired using bespoke scanners and sequences, then carefully processed to produce high quality data. We also visually inspected the dataset to check for the presence of artefacts. We believe that artefacts that were not detected by these checks will have minimal impact on the experiments performed.

3.6 Conclusions

I introduced a simulator designed to address the limitations of existing frameworks discussed in Section 2.4.3. It is able to produce realistic full-brain contrast and models the physics of MR acquisition, enabling many of the artefacts discussed in Section 2.3 to be simulated. I also introduced a novel framework for evaluating post-processing techniques, with a view towards applying the framework to assessing techniques for correcting motion and eddy-current artefacts in the next chapter.

Chapter 4

Application I: assessing motion and eddy-current correction techniques

4.1 Overview

In this chapter we apply the simulation framework to assessing popular post-processing techniques for correcting motion and eddy-current artefacts. We find that one of the most popular post-processing techniques performs poorly, highlighting the usefulness of being able to directly and quantitatively evaluate techniques. We investigate the performance of the best performing technique as a function of the acquisition protocol (number and distribution of diffusion directions), in order to make practical recommendations for its use.

4.1.1 Research dissemination

These results have been published in NeuroImage, and an extension to them presented at ISMRM.

- Realistic simulation of artefacts in diffusion MRI for validating post-processing correction techniques. MS Graham, I Drobnjak, H Zhang. In *NeuroImage* 125, 1079-1094, 2016.

- Quantitative evaluation of eddy-current and motion correction tech-

niques for diffusion-weighted MRI. MS Graham, I Drobnjak, H Zhang. In, *Proceedings of the International Society for Magnetic Resonance in Medicine*, 2016.

4.2 Introduction

Eddy current and motion artefacts are nearly always present in DW-MR data. Eddy currents vary according to the diffusion gradient, which causes each acquired volume to contain different distortions. Motion leads to misalignment between (and within) volumes. Both of these artefacts lead to violation of the assumption that a given image voxel corresponds to the same spatial location across the dataset, and undermines subsequent analysis of the data. The growing trend towards acquiring richer datasets, with increased angular-resolution and higher b-values is increasing the severity of these artefacts, by increasing scan time, which makes it more likely the patient will move, and increasing the strength of the diffusion gradients, which increases the severity of the eddy currents.

Some acquisition-time techniques can reduce these artefacts. For example, prospective motion correction can detect head movement and realign the imaging gradients in real time in order to acquire the correct slice [116], and eddy-currents can be mitigated using adjusted sequences such as the twice-refocused spin-echo (TRSE) [94]. However prospective motion correction isn't available on all scanners, and the TRSE is a longer sequence than PGSE and requires T_E to be increased, which can affect SNR. As discussed in the previous section, in practice post-processing techniques are very popular, as they have several advantages: they can be applied retrospectively to already acquired data, a user can revert to the original data if the technique does not work as hoped, and they don't require additional scan-time, which is often expensive. However these techniques are difficult to evaluate; as a result the literature relies on either qualitative visual assessments of image alignment [100, 101], or quantitative assessments of surrogate measures of alignment, such as tract

length [102], fractional anisotropy (FA) values [103] or reduced residuals from fits to microstructural models [92, 104]. The lack of a suitable, quantitative evaluation makes it difficult to understand which techniques should be used.

In this chapter, we apply the simulation framework introduced in Chapter 3 to compare two of the most commonly used post-processing techniques for correcting EC and motion artefacts. FSL's `eddy_correct` represents what until recently has been standard practice, the registration of each DW-MR volume to a $b = 0$ volume, whilst `eddy` is a newer tool that registers each volume to a model-free prediction of how it should look in undistorted space [38]. We then turn the framework to a more thorough investigation of the most promising of these techniques, `eddy`. We investigate how `eddy's` performance depends on the data acquisition: both the number of directions acquired, and the choice of full-shell vs half-shell protocols, in order to make practical recommendations for its use. Finally we look at the impact of correction quality by examining microstructure fits to real and simulated data.

4.3 Experiments and Results

We firstly undertake a comparison of two artefact correction methods, `eddy_correct` and `eddy` (Section 4.3.1), before evaluating the performance of `eddy` as a function of the quality of the dataset (Section 4.3.2). Finally we assess the impact of correction by fitting models of microstructure to real and simulated data (Section 4.3.3).

4.3.1 Comparison of correction methods

In this section we compare the quality of correction obtained from two post-processing techniques. The first is FSL's `eddy_correct`, which performs an affine (12 degrees-of-freedom) registration of each volume in a dataset to a b=0 image in order to simultaneously correct for motion and EC distortions. We also test a more sophisticated method, FSL's `eddy`, which registers each volume to a model-free prediction of how it should look in undistorted space [38]. Firstly we compare them quantitatively using our simulation framework,

and then we compare them qualitatively on a real dataset to demonstrate that our findings in simulation are present in real data.

The simulated dataset consists of two shells, b=700/2000 s/mm^2, 32/64 directions with 12 b=0 images, TR/TE = 7500/109 ms, 72×86×55 with isotropic voxel size 2.5 mm. All simulations were performed at 3 T. K-space was acquired with a readout bandwidth of 100 kHz, using a linear-ordered, cartesian sampling trajectory with full coverage. We used a matrix size of 72×86, which was chosen along with the image voxel size (2.5 mm isotropic) to strike a balance between minimising computation time and ensuring full-brain coverage. K-space was apodized using a Hamming window, and no zero-filling was performed. Diffusion directions were distributed isotropically on the sphere. EC gradients were added to the pulse sequence according to the model in Section 3.3.3. One dataset was simulated with just EC artefacts, and one was created with both EC and motion artefacts. In this dataset, a translation along each axis was selected for each volume randomly from the range -5 to 5 mm, in addition to a rotation about each axis taken from the range -5 to 5°. Ground truth displacement fields were obtained for each volume as described in Section 3.3.4. Normally distributed noise $\mathcal{N}\left(\mu,\sigma^2\right)$ with $\mu=0$ and a spatially constant σ was added to the real and imaginary channels of the signal, so that the magnitude images contained Rician noise. Two different values of σ were selected to produce two datasets, one with SNR=10 and one with SNR=20. The SNR was defined as $A/\sigma_{measured}$, where A was the mean signal value in the $b=0$ images across a region of interest (ROI) delineated in the WM (the centrum semiovale), and $\sigma_{measured}$ was obtained by taking the standard deviation of the signal in a large region of the image background, drawn to exclude any brain voxels.

The simulated datasets were corrected using `eddy_correct` and `eddy`. Default settings were used for `eddy_correct`: correlation ratio as the similarity measure and trilinear interpolation. Default settings were mostly used for `eddy`: 1000 voxels for estimating the Gaussian Process hyperparameter, spline

interpolation, quadratic first-level modelling of the EC, and no second-level modelling. However we used 10 iterations rather than the default 5, because we sometimes found 5 was not sufficient to ensure convergence. We also performed correction using `eddy_correct` with normalised mutual information (NMI) as a cost function, to test the claim that it is more robust than other cost-functions [117]. Results for correction of the datasets with EC artefacts are shown in Fig. 4.1a, and that dataset with both EC and motion artefacts are in Fig. 4.1b. Figure 4.2 shows how these displacement field errors are spatially distributed across the brain.

The real dataset was acquired on a Siemens PET-MR 3T with similar parameters to the simulated dataset: two shells, b=700/2000 s/mm^2, 32/64 directions with 4/8 b=0 images, TR/TE = 7500/103 ms, isotropic voxels of size 2.5 mm. The only differences were the dimensions, here 96×96×55, and the TE (103 ms here vs 109 ms in the simulations). The SNR of the data was 25, measured on the b=0 images in an ROI in the centrum semiovale. This dataset was also corrected using `eddy_correct` and `eddy` with the default settings. Figure 4.3 shows the results.

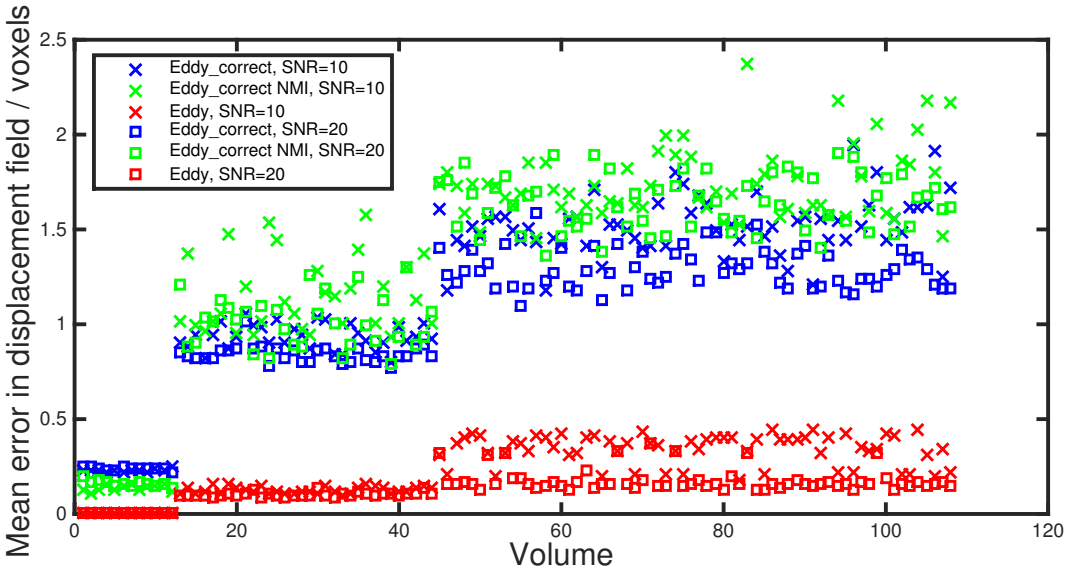
Figure 4.1 demonstrates that `eddy_correct` is unable to correct the data well, even for DWIs acquired with b=700 s/mm^2, whilst `eddy` is able to provide good correction across the dataset. Volumes corrected with `eddy_correct` have average errors of one voxel at b=700 s/mm^2, rising to 1.5 voxels at b=1000 s/mm^2. This seems to be caused by the increasing contrast differences between the DWI and b=0 volumes as b-value is increased, which makes direct registration progressively worse. These findings are in agreement with previous work [80] which found that DWIs can only be successfully corrected by registration to b=0 for $b \leq 300$ s/mm^2. By contrast, `eddy` is able to correct with errors of less than 0.2 voxels across the dataset at SNR=20, and 0.5 voxels at SNR=10. The spatial distribution of errors in Figure 4.2 show that `eddy_correct` consistently over-scales the data. This is likely caused by the attenuation of the CSF rim around the brain in DWIs. This makes the DWIs

look smaller then the b=0 images, which causes registration to enlarge them. Interestingly, Fig 4.1 shows that at SNR=10 `eddy` provides better correction for the dataset with EC and motion than the dataset with EC and no motion. For the dataset with no motion, `eddy` was not able to detect the EC artefacts, and essentially left the data uncorrected, whilst for the dataset with motion `eddy` was able to accurately estimate both the motion and EC artefacts. At SNR=20 `eddy` was able to estimate the correct parameters for the dataset with EC and no motion. Thus it seems that the combination of low SNR and no motion artefacts caused `eddy's` optimisation process to fail. We found that both methods showed little sensitivity to noise for SNR$>$ 20: results on a dataset with SNR=30 (not shown) were very similar to those found on the SNR=20 data.

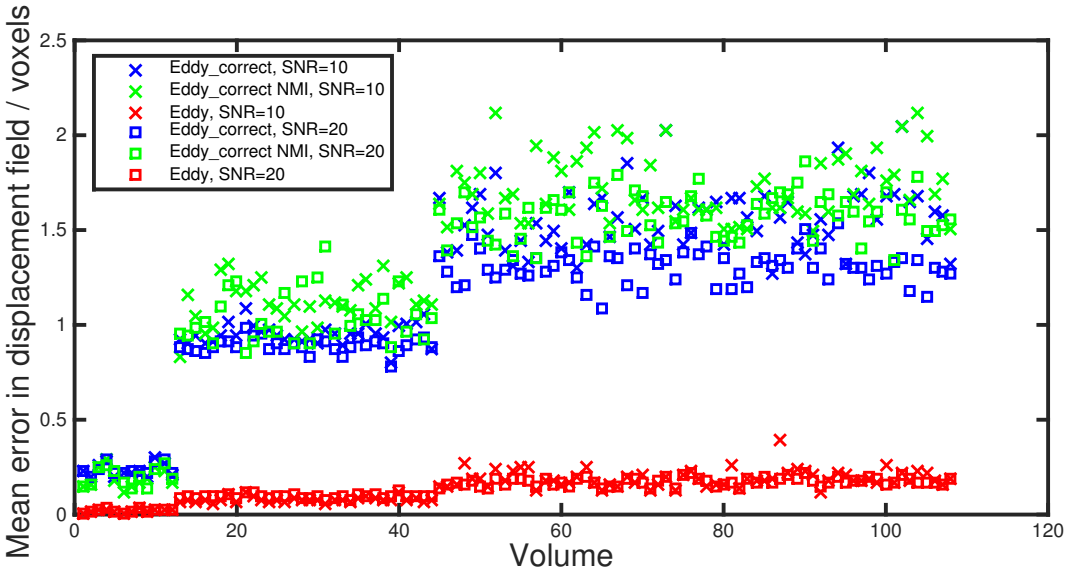
The results for real data corroborate with our findings for simulation. Figure 4.3 demonstrates an over-scaling of data corrected by `eddy_correct`, noticeable at b=700 s/mm^2 and extremely clear at b=2000 s/mm^2. The datasets corrected by eddy are much better aligned with the b=0 outlines. This figure also serves to highlight the difficulty in the application of qualitative methods to the assessment of artefact correction: the results are sensitive to exactly how the outline is drawn on the b=0, which involves a subjective judgement of how much of the CSF to exclude as the b-value varies.

4.3.2 Evaluation of `eddy`

We investigated the dependence of `eddy's` performance on the quality of the dataset being corrected, both by varying the number of DW directions and by comparing the recommended full-shell acquisition scheme to a half-shell scheme. Eddy makes use of information from 'similar' volumes in order to create a registration target for each volume in the dataset, and so, unlike `eddy_correct`, its ability to correct each volume is dependent on the full dataset. Thus it is recommended that datasets have a sufficient number of diffusion directions and are sampled on the full-sphere (or alternatively, with a blip-up blip-down acquisition). However there is no information available on



(a) Eddy currents.



(b) Eddy currents and motion.

Figure 4.1: Mean error in displacement field across the brain. The first 12 volumes are b=0, the next 32 are b=700s/mm^2 and the remaining 64 b=2000s/mm^2.

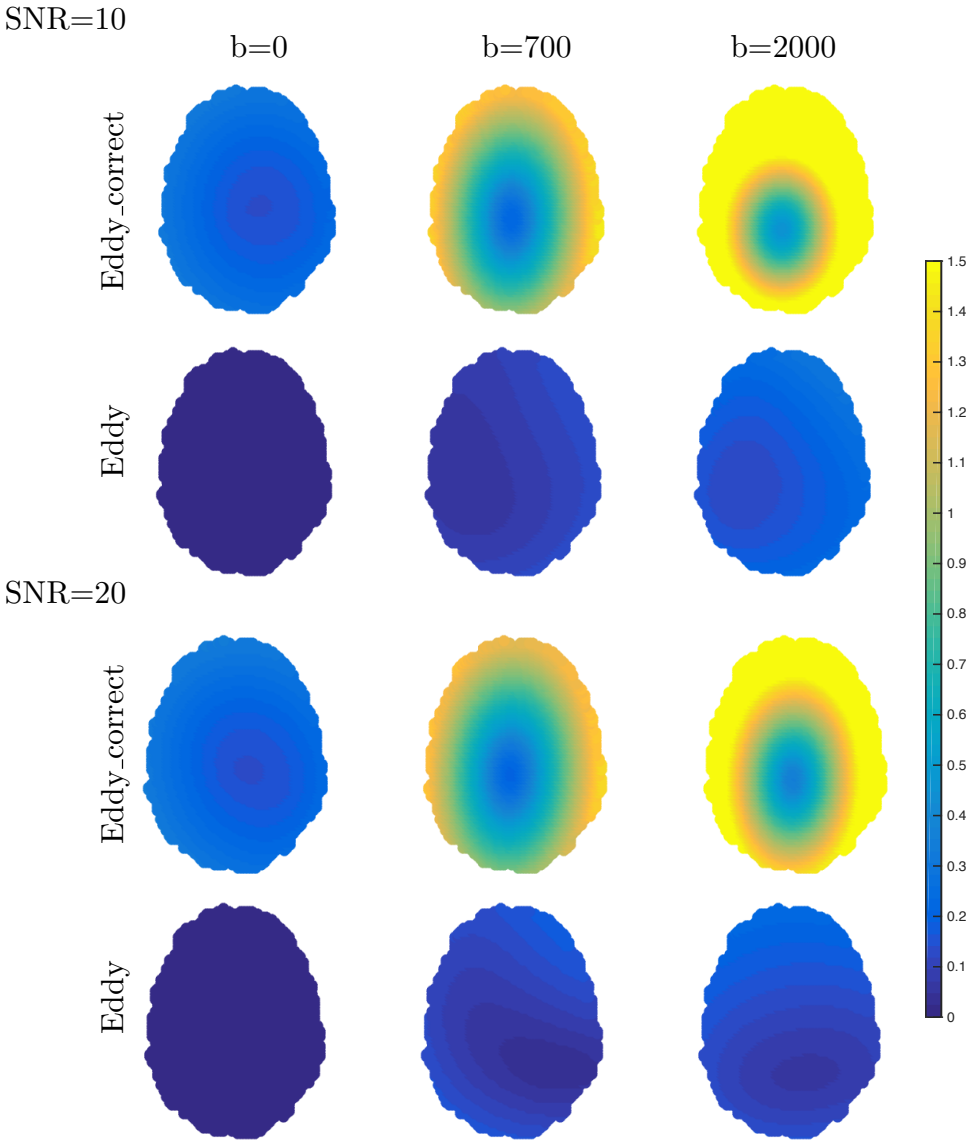


Figure 4.2: Spatial errors in displacement field, in voxels (each 2.5 mm isotropic), shown across one slice of the brain. The errors are a mean across all of the volumes with the same *b*-value. Results shown for `eddy_correct` were obtained using the default cost function.

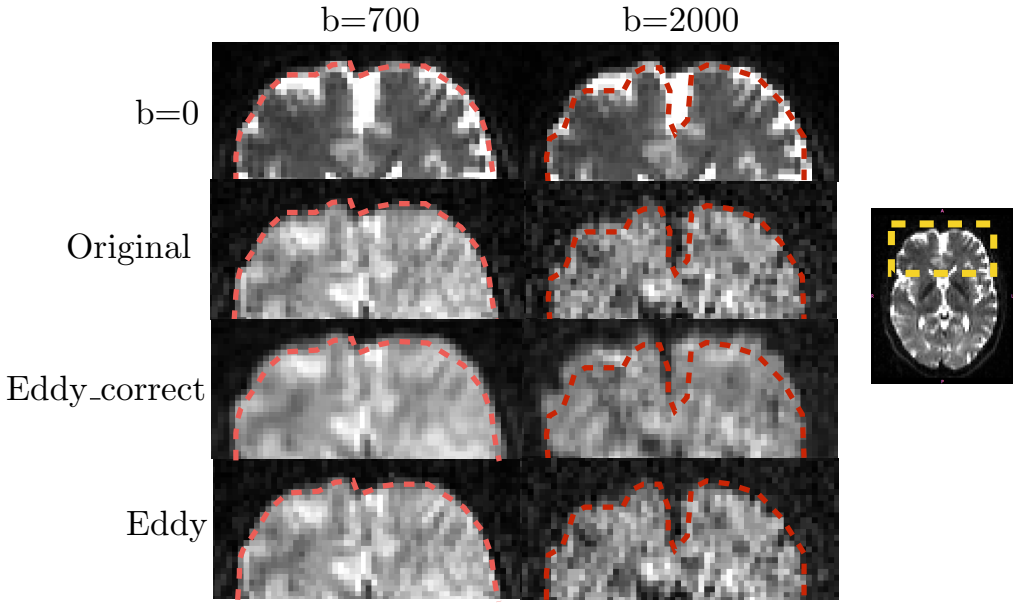


Figure 4.3: Correction errors on real data. Anterior portion of the brain in an axial slice is shown, corresponding to the yellow region on the inset image. An outline was drawn around the undistorted b=0 image, then superposed on a DWI. The boundaries of an undistorted DWI should align with this outline. Different outlines were drawn on the b=0 for use on the b=700 and b=2000 volumes, to account for the different amounts of CSF attenuation present.

the minimum number of diffusion directions required for effective correction, or on the performance penalty incurred when correcting data acquired on the half-sphere.

To test `eddy's` dependence on the number of DW directions, a number of datasets were simulated to represent common diffusion tensor imaging (DTI) and high angular resolution (HARDI) acquisition protocols. Single-shell datasets with 16, 32, 48 and 64 diffusion directions were generated at both b=1000 s/mm^2 and b=2000 s/mm^2, with one b=0 image for every 8 DWIs. Multi-shell datasets were made by combining the single-shell acquisitions, to create sets with 16/16, 16/32, 32/32, 32/64 and 64/64 directions in the b=1000 s/mm^2 and b=2000 s/mm^2 shells respectively. Diffusion directions were obtained from a minimisation of electrostatic energy as implemented in Camino [118, 119], and optimised on the full-sphere. Each dataset had TR/TE = 7500/109ms, 72×86×55 with isotropic voxel size 2.5 mm, and Rician noise

was added to produce datasets with both SNR=10 and 20. EC and motion artefacts were added in the manner described in Section 4.3.1. Each dataset was corrected using `eddy` with the same settings described in Section 4.3.1. The results are reported in Fig. 4.4.

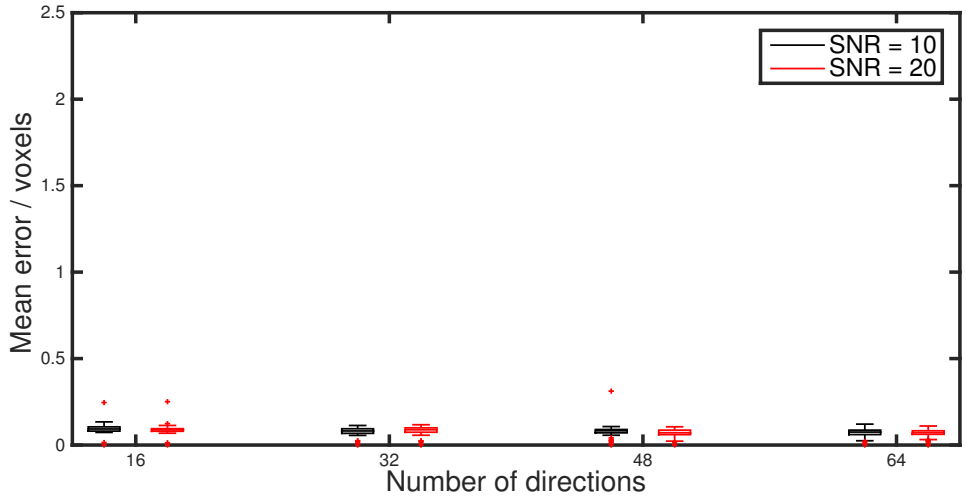
To test `eddy's` ability to cope with half-shell datasets, half-sphere sampling schemes were obtained from the full-sphere schemes mentioned previously, by negating each b-vector with a z-component less than 0. Datasets were generated from these schemes with the same acquisition parameters as their full shell counterparts. Rician noise was added to each dataset to create an SNR of 20. Each dataset was corrected three times using `eddy`: the first using the default settings as described in Section 4.3.1, and the second and third times using linear and quadratic models that relate the parameters that define the EC distortion field to the applied b-vector (the default places no constraints on the relationship between these parameters and the b-vector). The results are shown in Fig 4.5.

The results indicate that `eddy` is able to provide good correction down to 16 diffusion directions at SNR = 20, at both b=1000 s/mm^2 and b=2000 s/mm^2. At SNR=10 good correction is achieved for the b=1000 s/mm^2 datasets, but `eddy` struggles to correct the b=2000 s/mm^2 single-shell datasets. The addition of more information might improve correction at low SNR -results for the combined 64/64 dataset are better than the 64 direction dataset at b=2000 s/mm^2.

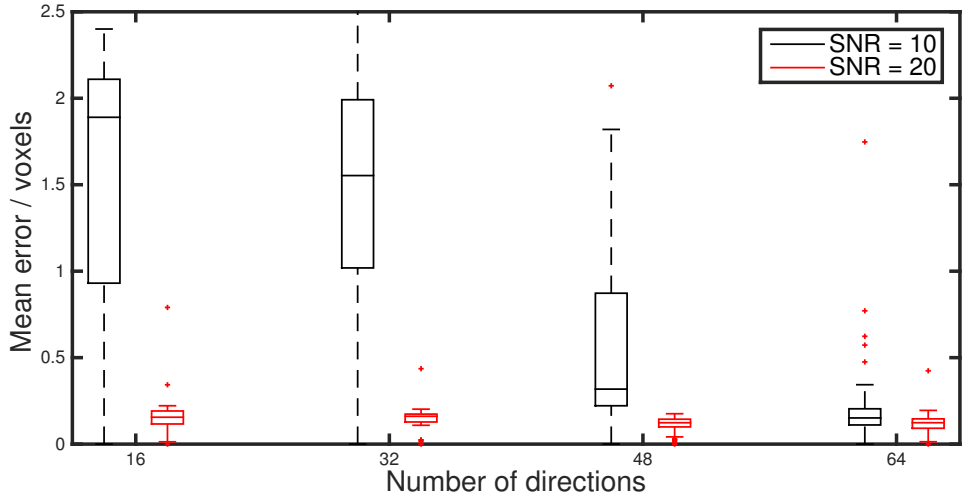
Results also indicate that, whilst full-shell sampling is optimal, it is still possible to obtain good correction on datasets acquired on the half-sphere. For most datasets the correction could be marginally improved by enforcing linear second-level modelling of the EC artefacts, which is to be expected as our simulations use a linear EC model.

4.3.3 Impact of correction on microstructure estimation

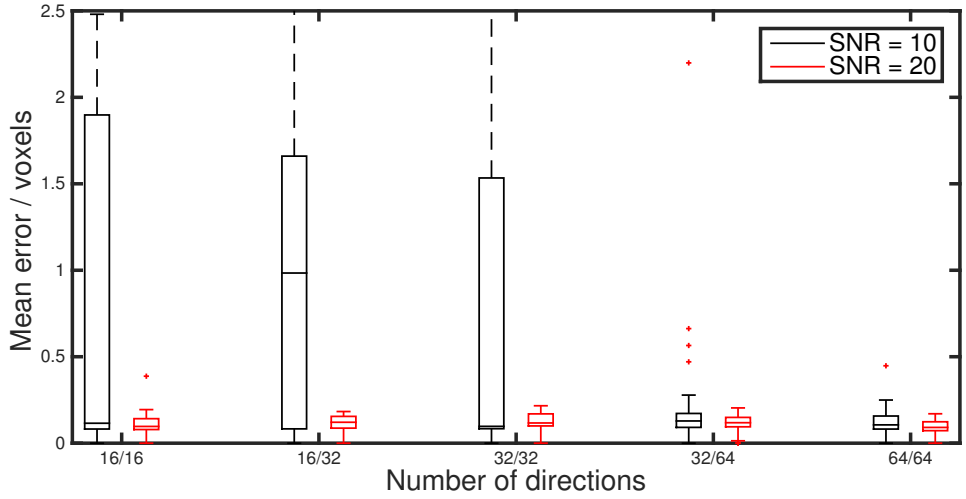
In this section we investigate the impact of artefact correction on the estimation of microstructural features from diffusion data. Firstly we fit two models,



(a) Single shell, b=1000 s/mm^2

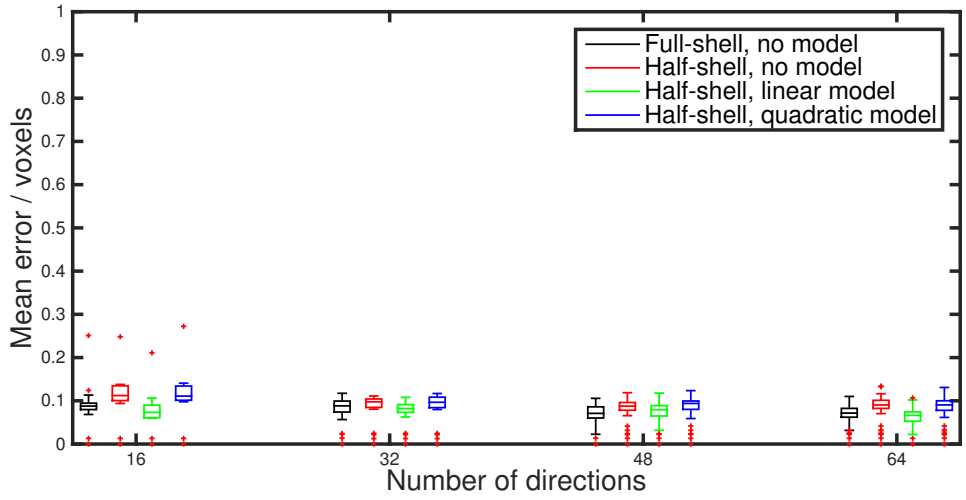


(b) Single shell, b=2000 s/mm^2

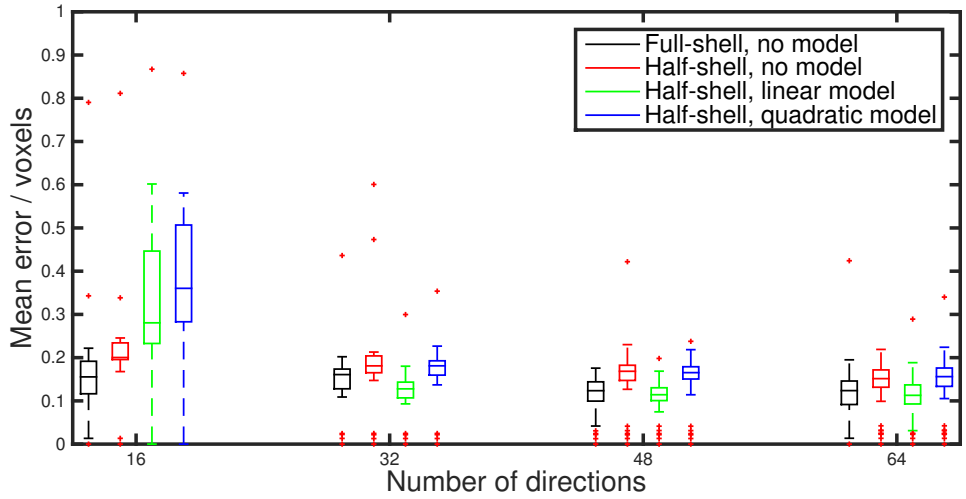


(c) Multi-shell, b=1000 s/mm^2 and 2000 s/mm^2

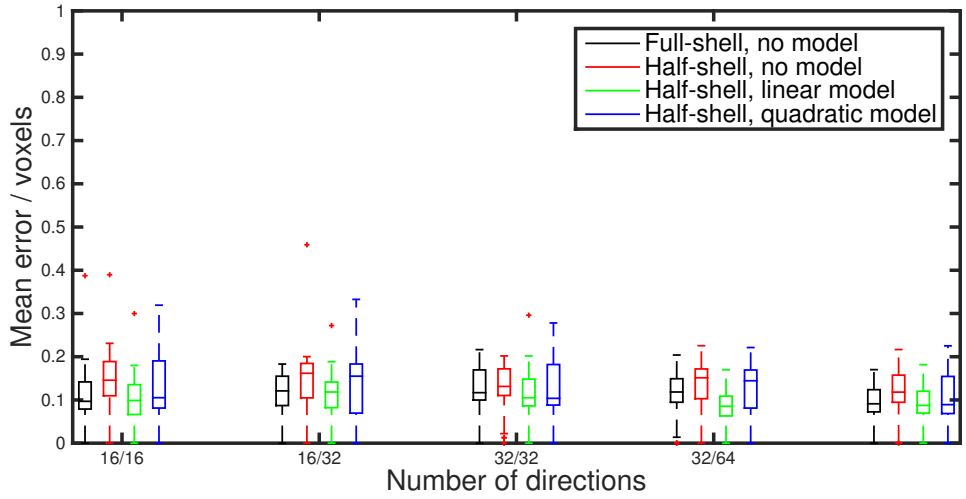
Figure 4.4: Errors in the displacement fields in datasets corrected with `eddy`, as the number of directions in the acquisition is varied. Each data point is a mean over the voxels in the brain for a volume, boxplots show the distribution of these means across the dataset.



(a) Single shell, b=1000 s/mm^2



(b) Single shell, b=2000 s/mm^2



(c) Multi-shell, b=1000 s/mm^2 and 2000 s/mm^2

Figure 4.5: Errors in the displacement fields in datasets corrected with `eddy` for full-shell and half-shell acquisitions. Each data point is a mean over the voxels in the brain for a volume, boxplots show the distribution of these means across the dataset. The model refers to the relationship between the applied $b-$vector and the parameters that determine the EC field. All datasets have SNR=20.

the diffusion tensor (DT) and NODDI, to simulated datasets before and after correction. The advantage of simulations is that we can compare the results to a 'ground truth' obtained by fitting these models to a dataset simulated without artefacts. Secondly we fit these models to a real dataset to demonstrate consistency with our findings on simulation. Finally, we examine the use of fitting residuals as a surrogate marker of improved correction.

The simulated dataset was the same as the one used in Section 4.3.1: two shells with b=700/2000 s/mm^2, 32/64 directions and isotropic voxel size 2.5 mm. We fit the DT to the b=700 s/mm^2 shell of five datasets: the ground truth, both with and without noise, one distorted with motion and EC artefacts, and this distorted dataset corrected by `eddy` and `eddy_correct`. All datasets had SNR=20. We changed the interpolation used by `eddy_correct` to spline, to match that used by `eddy`. FSL's DTIFIT was used to fit the tensor. We fit NODDI to both shells of each of these datasets, using the NODDI Matlab Toolbox. Figure 4.6 shows the resulting FA and sum-squared error (SSE) of the model prediction's residual maps, and Fig. 4.7 shows the NODDI parameter maps.

The real dataset was the same used in Section 4.3.1: two shells, b=700/2000 s/mm^2, 32/64 directions with isotropic size 2.5 mm. We fit the DT to the b=700 s/mm^2 shell of the original data, and the data after correction by `eddy` and `eddy_correct`. We fit NODDI to the multi-shell dataset. Figure 4.8 show the resulting parameter and residual maps.

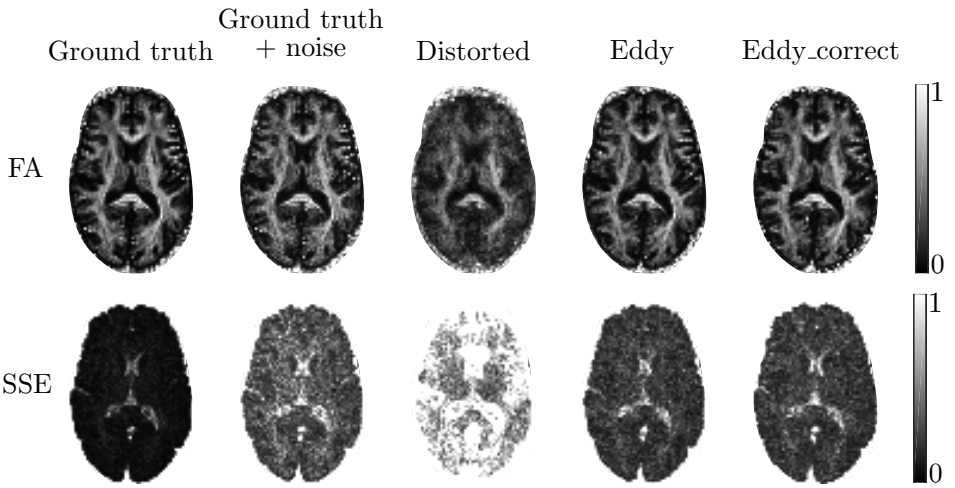
The results from simulation in Figs 4.6 and 4.7 demonstrate the impact of these post-processing techniques on estimating microstructure. Whilst FA maps in Fig 4.6 are hard to distinguish, the difference maps in Fig 4.6b are more informative. They show anatomical structure in the data corrected by `eddy_correct` which is not apparent in the data corrected by `eddy`. The NODDI maps in Fig 4.7a show a smoothing of the parameters when data is corrected with `eddy_correct`, particularly noticeable in the orientation dispersion index (ODI). There is overestimation of the intra-cellular volume fraction

(V_{IC}) in the dataset corrected by `eddy_correct`, particularly in GM regions, and an underestimation of the isotropic volume fraction (V_{ISO}) around the edges of the brain, likely caused by the overscaling of the DWIs. The difference maps in Fig 4.7b make these problems more clear. There is noticeable structure in the difference maps for `eddy_correct`. By contrast, the parameter maps estimated from data corrected by `eddy` are much more similar to the ground truth maps.

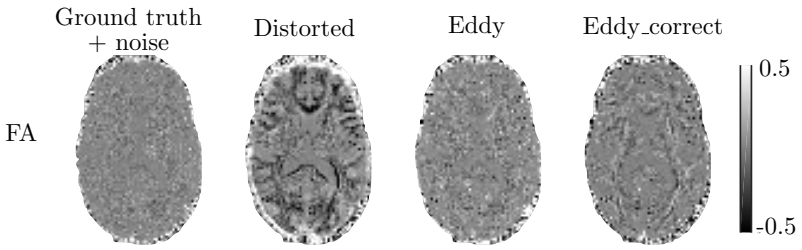
Whilst there is no GT available for the fits to real data, the parameter maps show features consistent with those found on simulation. Compared to the data corrected by `eddy`, the dataset corrected by `eddy_correct` shows a loss of sharpness in some of the FA structures in Fig 4.8a. There is also an overestimation of the V_{IC} and underestimation of V_{ISO} relative to the results from `eddy`, indicated on the figure. The general smoothing of the ODI map is also noticeable.

The results allow us to investigate the effectiveness of using fitting residuals as a surrogate marker of image alignment. Both corrected datasets show lower residuals than the original data. Residuals from the two correction methods appear similar, and are best compared using difference images, shown in Fig 4.9. Despite results from Section 4.3.1 showing that `eddy` provides better correction, `eddy_correct` gives lower residuals from DT fits in a rim around the brain, in both real and simulated data. There are also some regions in the middle of the brain where `eddy_correct` shows lower residuals from DT fits. We speculate this is caused by the smoothing that results from the overscaling of the DWIs, which makes the signal easier to fit to. It seems that the smoothing from interpolation also has an effect: in Fig 4.7a the residuals from the corrected datasets are lower than those in the GT + noise data.

Figure 4.9 shows that residuals from NODDI fits are lower for data corrected with `eddy` than data corrected with `eddy_correct`. The increased residuals for data corrected by `eddy_correct` seem to be due to the multi-shell nature of the data. Results in Section 4.3.1 show that `eddy_correct` overscales



(a) FA and signal residual maps. Residuals are the sum-squared difference between actual and predicted signal.



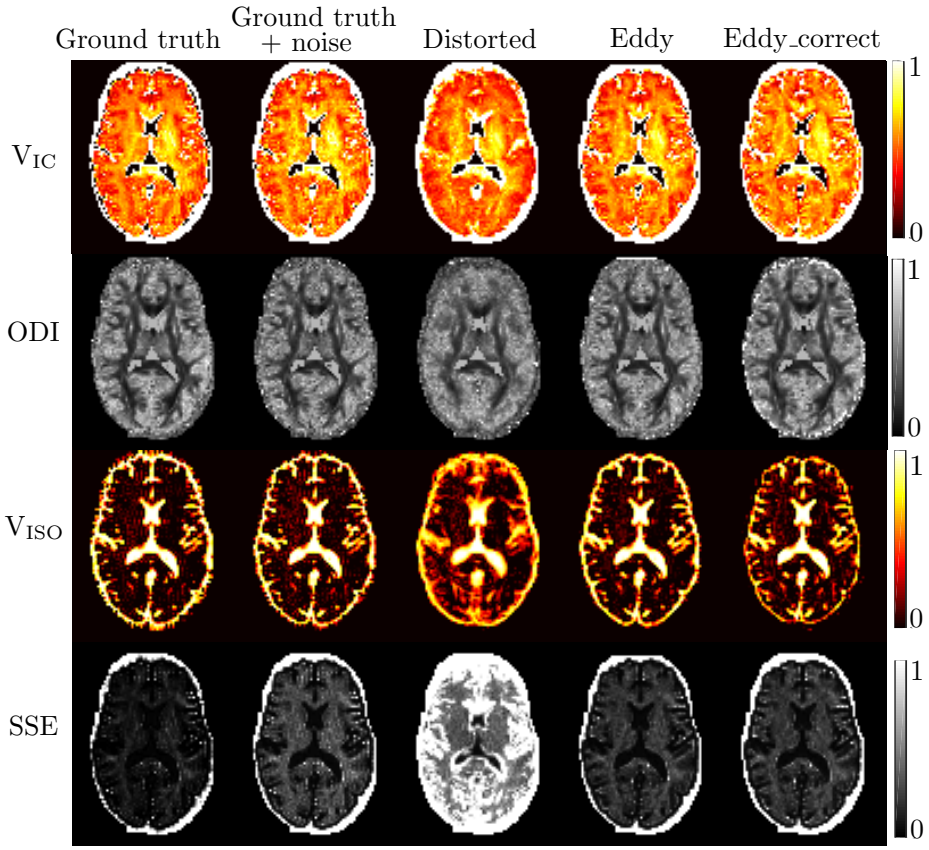
(b) Difference in FA compared to ground truth estimates.

Figure 4.6: FA maps resulting from fits to ground truth data (i.e. data simulated with no distortions), ground truth data with added noise, data distorted with both EC + motion, and distorted data corrected by `eddy` and `eddy_correct`. SNR=20 for all noisy data.

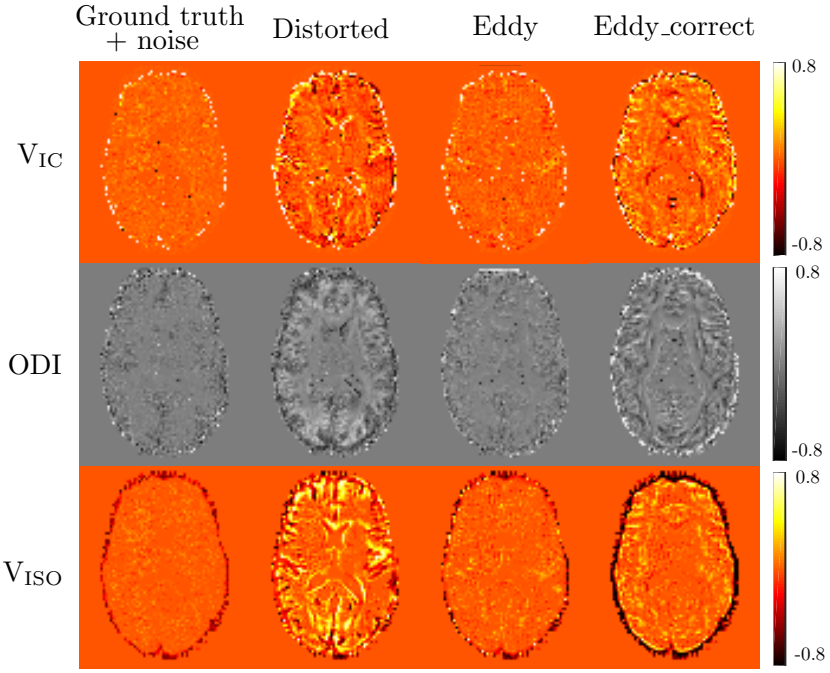
DWIs by increasing amounts with increasing b-values, so this internal misalignment in the dataset is likely the reason that NODDI fits data corrected using `eddy_correct` badly. These results indicate that reduced residuals from model fits can be indicative of increased image alignment, but can also be confounded by other factors, such as image smoothing and expansion into background regions, that mean these measures need to be interpreted carefully.

4.4 Discussion

The simulation framework allowed us to quantitatively compare a widely used technique for correcting EC and motion artefacts in DWI, `eddy_correct`, and its recently proposed alternative, `eddy`. We were able to provide quantitative

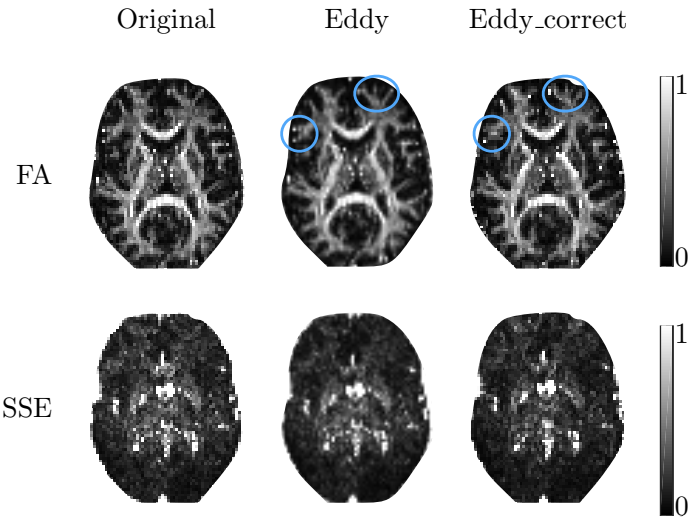


(a) NODDI parameter and signal residual maps. Residuals are the sum-squared difference between actual and predicted signal.

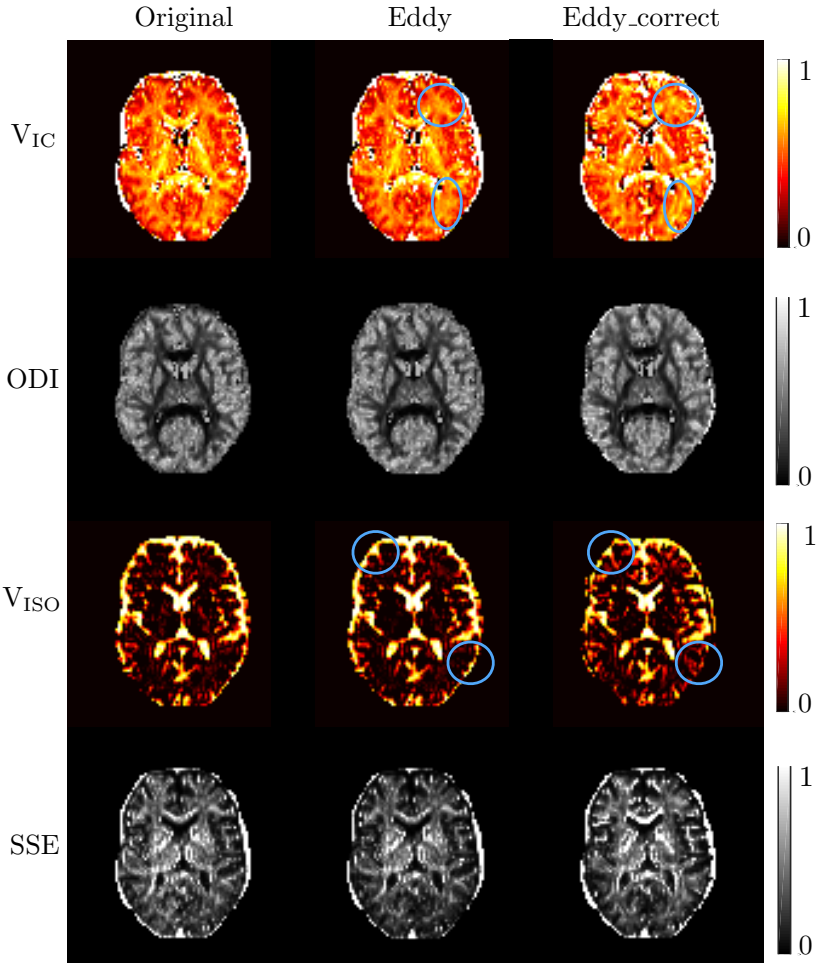


(b) Difference in NODDI parameters compared to ground truth estimates.

Figure 4.7: NODDI parameters resulting from fits to ground truth data (i.e. data simulated with no distortions), ground truth data with added noise, data distorted with both EC + motion, and distorted data corrected by `eddy` and `eddy_correct`. SNR=20 for all noisy data. Parameters are: V_{IC} - intracellular volume fraction, ODI - orientation dispersion index, V_{ISO} - isotropic volume fraction.



(a) FA and SSE residual maps. Blue regions highlight blurring of FA structures by `eddy_correct` compared to `eddy`.



(b) NODDI parameter and SSE residual maps. Blue regions highlight increase in V_{IC} and decrease in V_{ISO} in data corrected by `eddy_correct` when compared to `eddy`. Also note `eddy_correct` causes smoothing of the ODI map.

Figure 4.8: Parameter maps resulting from fits to real data.

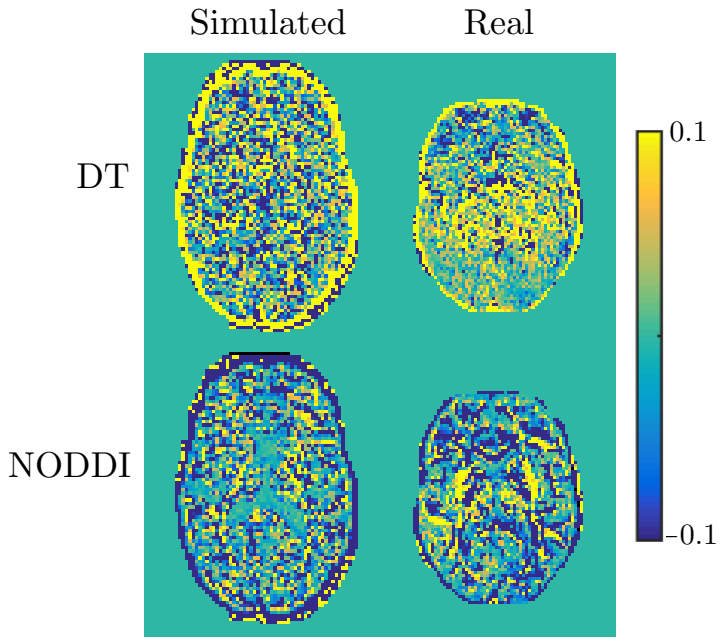


Figure 4.9: Difference in residuals from `eddy` and `eddy_correct`: SSE_{eddy} - $SSE_{eddy\_correct}$, meaning positive values indicate voxels where data corrected using `eddy_correct` had lower residuals.

evidence that `eddy_correct` systematically overscales DWIs, which corroborates with findings in the literature [120]. We also demonstrated that `eddy` is able to provide significantly better correction. These findings corroborate with a previous study [121], which applied the two methods to real datasets and assessed them using a combination of visual inspection and comparison of FA values.

An important question is whether other, similar correction techniques exhibit the tendency of `eddy_correct` to overscale the data. In an extension to the results already shown in this chapter, we investigated this for two other correction techniques: `ACID` performs a constrained 9 degrees-of-freedom (DOF) registration to a $b = 0$ image [122] and `ExploreDTI` registers to b=0 in order to optimise the parameters of an EC and motion specific model [117]. We applied these techniques to the motion and EC dataset described in Section 4.3.1. Figures 4.10 and 4.11 show the error fields for these new methods, along with those of `eddy_correct` and `eddy`. These other methods perform better than

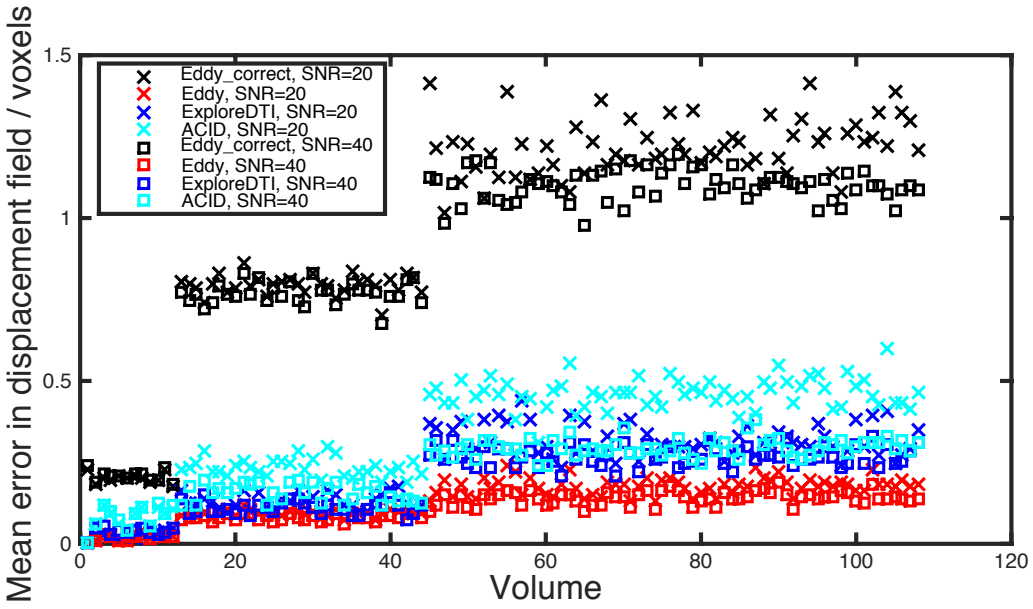


Figure 4.10: Mean error field over the brain, evaluated for each volume. Volumes are arranged by b-value so that the first 12 are b=0, next 32 are b=700 and final 64 are b=2000.

`eddy_correct`, likely because they use more constrained methods for registering each DWI to the $b = 0$ image, but are still outperformed by `eddy`, which avoids registration to $b = 0$.

We were able to investigate the quality of dataset needed to ensure good correction with `eddy`. For data with SNR=20, good correction can be obtained on as few as 16 directions, though denser sampling is needed for data with lower SNR, such as b=2000 s/mm^2 data at SNR=10. We also demonstrated that, whilst a full-sphere acquisition scheme is ideal, it is still possible to achieve good correction on half-sphere datasets. These findings are useful for retrospective studies where full-shell acquisition has not been considered. The datasets simulated contained particularly severe artefacts, including large, random movements between each volume and large EC distortions, so these findings could be considered to be an upper-bound on the error that can be expected when correcting DW datasets. However, our findings indicate that the performance of `eddy` is robust to the severity of artefact, as supported by the similar corrections achieved for a dataset with just EC artefacts and

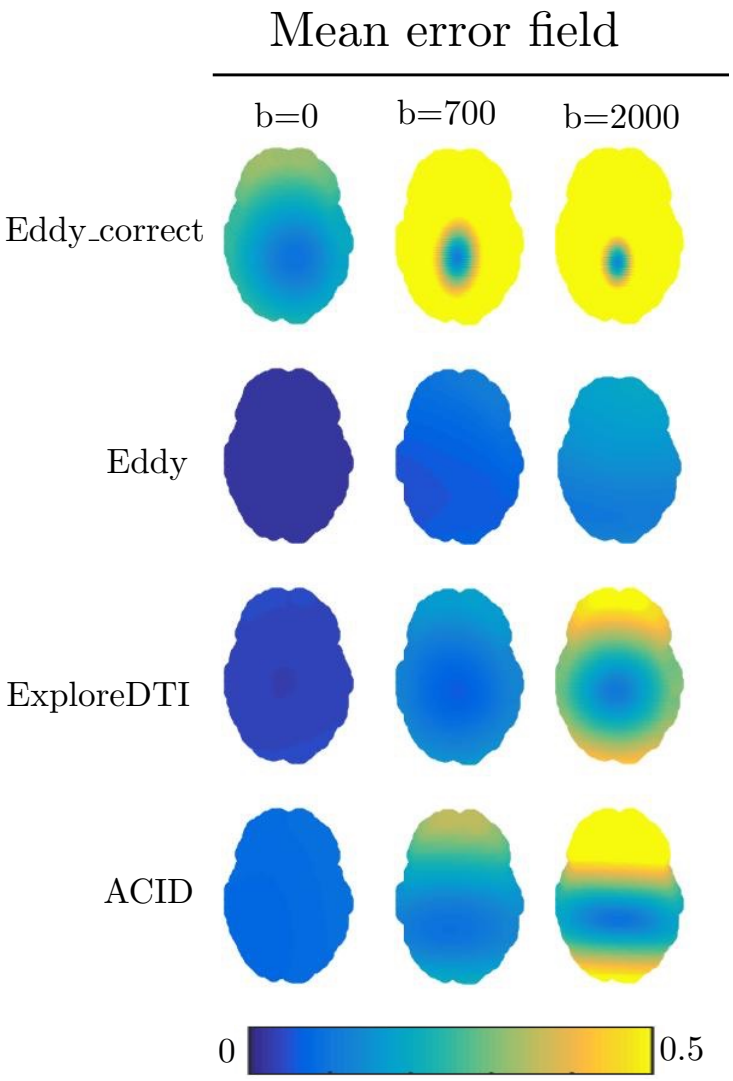


Figure 4.11: Spatial plots of the mean error field over every volume at a given b-value for SNR=20, shown for one axial slice. Errors are in voxels (2.5 mm isotropic).

a dataset with both EC and severe motion in Fig 4.1. It may be that the severity of these artefacts explains why we found more than the recommended 5 iterations were necessary to achieve good correction with `eddy`.

These results are important in the context of techniques that use DW-MR data. Data is most commonly acquired at b=1000 s/mm^2, and our results indicate we can expect errors of more than 1 voxel in such images if they are corrected using registration to b=0. These are enough to cause anatomical misalignment in regions of partial volume, such as the boundaries between GM and CSF which will compromise any information on microstructure obtained from such data, as demonstrated by the FA maps in Fig 4.6. Figure 4.7 shows this effect is even more severe for data acquired at b= 2000 s/mm^2, which is becoming more common with the increasing popularity of HARDI techniques.

In this work a linear EC model was used, which causes simple scalings, shearings and translations of each DWI. This has been found to be inadequate for describing the EC fields in some scanners. Rather they are better described by a second or third order polynomial [117]. We also modelled motion as occurring instantaneously between the acquisition of volumes, enabling it to be modelled as a simple rigid-body deformation. This matches the assumption that most post-processing techniques make when trying to correct the data, but in practice motion can occur at any point in the acquisition. Future work could assess the impact that both higher-order ECs and movement throughout the acquisition have on attempted correction. We also assumed that only EC and movement artefacts were present in the data. Perfect fat saturation was assumed - the framework is currently unable to simulate this, as the input object used does not contain information about fat. Other artefacts that can be included in the simulations (e.g. ghosting, RF spikes) were excluded to allow us to focus solely on EC and motion artefacts, but they can be included in future work to see how their presence affects EC/motion correction.

We use comparison of displacement fields in order to assess the effectiveness of correction techniques. This directly measures the desired outcome of

such techniques, i.e. the mapping of all images into an undistorted space. We contrast this with some of the quantitative surrogate metrics that have been used to assess techniques, such as FA values in WM tracts [121], the length of tracts obtained from tractography [102] or the size of the residuals from a model-based fit to the data [92, 104]. These methods are often the best available when testing correction on real datasets where a GT is not available, but they are indirect metrics: increased image alignment is only one of many factors that can affect them. For example, the results indicate that residuals from a DT fit can be lowered by smoothing of the data due to over-scaling of the DWIs. Current methods only attempt to correct for geometric distortions, so we predicted geometric displacement fields for assessment. Future methods may also try to correct for non-geometric effects such as the blurring due to decay of ECs, and our framework can be extended to provide GT estimates of how these may be corrected, which could assist both the development and testing of such methods.

4.5 Conclusions

I applied the simulation framework to assessing popular post-processing techniques for correcting motion and eddy-current artefacts. I was able to quantitatively demonstrate that `eddy_correct` performs poorly, which is significant because it is frequently used by the community. The results also demonstrated that `eddy` performs well, and I was able to investigate the type of acquisition required to enable `eddy` to correct datasets, in order to make practical recommendations for its use. This work highlights the importance of careful evaluation of acquisition and processing pipelines, and demonstrates how the simulation framework can aid such evaluation.

Chapter 5

Simulating the spin-echo

5.1 Overview

In this section we extend the simulator to address one of its key limitations discussed in Chapter 3, the inability to simulate spin-echo pulse sequences.

5.1.1 Research dissemination

This work described here is published in PLOS ONE. The implementation of spin-echo in POSSUM will be available in the next release of the FSL toolbox.

- Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI. MS Graham, I Drobnjak, M Jenkinson, H Zhang. In *PLOS ONE* 12 (10), e0185647

5.2 Introduction

One of the key limitations of the simulation framework is its inability to simulate spin-echo (SE) pulse sequences. Whilst it has been possible to simulate the contrast found in spin-echo images by using a GE-EPI sequence and replacing T_2^* with T_2 values, the lack of a SE means the simulator is not able to faithfully model the effects of a susceptibility field in DW-MR data. This is because the susceptibility-induced field causes both geometric distortions and signal loss in GE-EPI data, whilst in SE-EPI data it only causes geometric distortions. In this chapter we describe work to implement SE simulation in POSSUM.

Tissue	T_1/ms	T_2^*/ms	T_2/ms	ρ
Grey matter	1331	51	75	0.86
White matter	832	44	70	0.77
CSF	3700	500	500	1

Table 5.1: Tissue parameters used in the spin-echo version of POSSUM. Proton density ρ is in arbitrary units.

5.3 Methods

We first detail our implementation of the spin-echo within POSSUM, and then discuss how our SE version of POSSUM is incorporated into the simulation framework.

5.3.1 Implementation of spin-echo in POSSUM

The implementation required several changes to POSSUM — an overview of how POSSUM works is shown in Fig 5.1 with all the changes made highlighted in red. Two of the changes are straightforward. Firstly, T_2 values for each tissue type are now passed to the simulator, along with the T_1 and T_2^* values passed in the previous version — the full set of tissue parameters used is shown in Table 5.1. Secondly, new functionality for generating pulse-sequences with a spin-echo is included.

The main change is the modelling of the action of the spin-echo on the magnetisation. A spin-echo consists of an 180° RF pulse that rotates the magnetisation of each spin about an axis in the transverse plane, inverting the magnetisation's phase. Ideally the modelling of spin-echo would simply involve modelling the RF action on each of these spins. However, POSSUM represents the dense collection of spins in a voxel as a single isochromat. This means we must take a model-based approach to implementing the spin-echo, similar to that described in [63], to simulate the effects of the spin-echo on each isochromat.

The 180° RF pulse affects both the magnitude and phase of the isochromat. After excitation, the isochromat's magnitude is reduced by the dephasing of its spins. The spin-echo reverses any loss of magnetisation that is caused by

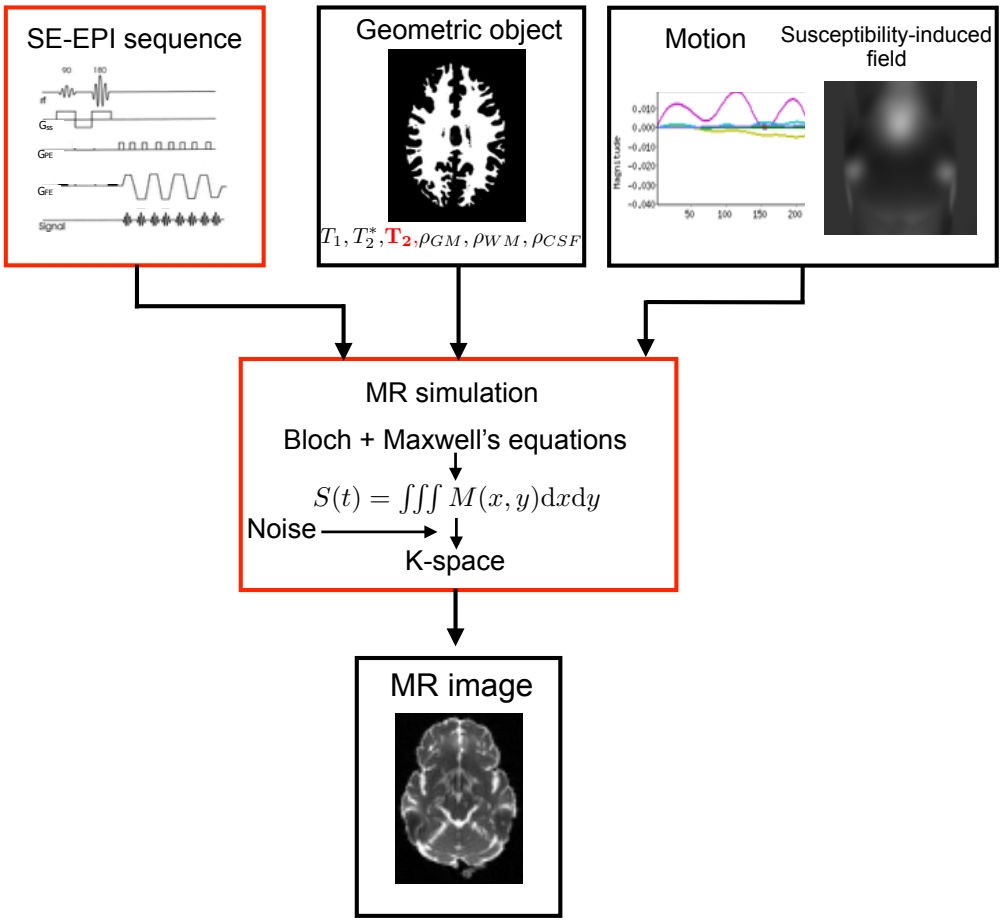


Figure 5.1: Flowchart describing the POSSUM simulator. Changes made in order to implement spin-echo in red. These changes are: addition of a 180 pulse to the input pulse sequence, inclusion of T2 tissue values as an input with the geometric object and modelling of the effects of the 180 pulse on the magnetisation in the main MR simulation phase.

dephasing which is time-invariant, i.e. that from T_2' effects and susceptibility-induced gradients. The overall phase of the isochromat is altered by any magnetic fields that change the frequency of its precession: this is caused by both intended fields (e.g. the applied imaging gradients) and undesired fields (e.g. those induced by eddy-currents and susceptibility). The spin-echo reverses any phase accumulated by the isochromat between excitation and the 180° RF, leading to the cancellation of any additional phase accrual caused by undesired, time-invariant fields at the echo time.

In the original POSSUM, T_2^* values were supplied for each tissue type, and the relaxation of the transverse magnetisation was calculated according to:

$$m(t) = m(0)\exp\left(-\frac{t}{T_2^*}\right) \tag{5.1}$$

where t is the time since excitation of the isochromat. Using the relation $1/T_2^* = 1/T_2 + 1/T_2^{'}$, this may be rewritten:

$$m(t) = m(0)\exp\left(-\frac{t}{T_2}\right)\exp\left(-\frac{t}{T_2^{'}}\right) \tag{5.2}$$

To handle the effects of the spin-echo on $T_2^{'}$-induced magnetisation loss the magnetisation calculation is replaced with:

$$m(t) = m(0)\exp\left(-\frac{t}{T_2}\right)\exp\left(-\frac{|t_2 - RF_{dist}|}{T_2^{'}}\right) \tag{5.3}$$

where t_2 is the time since the last RF pulse and RF_{dist} is the time between the 90 and 180 pulses (0 if no spin-echo pulse has occurred). The effect of the $|t_2 - RF_{dist}|$ term can be understood by examining its behaviour in three regimes: before the spin-echo, after the spin-echo and after the echo itself:

$$|t_2 - RF_{dist}| = \begin{cases} t & t < T_E/2 \\ T_E - t & T_E/2 < t < T_E \\ t - T_E & T_E < t \end{cases} \tag{5.4}$$

This formulation is such that rephasing of $T_2^{'}$ begins at $t = T_E/2$ and is complete ($|t_2 - RF_{dist}| = 0$) at $t = T_E$; dephasing continues after this point. For a gradient echo sequence, $t_2 - RF_{dist} = t$ for all values of t and we observe the expected T_2^* decay.

The second factor leading to loss of longitudinal magnetisation is gradient-induced dephasing across the voxel. POSSUM handles this by evaluating an analytical function of the time-integral of these gradients [105]. The effects of the spin-echo here are modelled by reversing the sign of integrals at the 180 RF

pulse; effectively causing a cancelling of their contribution at the spin-echo.

The effects of the spin-echo on the isochromat's phase are accounted for straightforwardly, by reversing any phase accumulated when the 180° RF occurs.

5.3.2 Incorporation of POSSUM into the DW-MR framework

The overall framework for simulating DW-MR data, incorporating the SE-enabled version of POSSUM, is demonstrated in Figure 5.2. The figure shows that susceptibility-induced fields are now included as an input, a SE-EPI sequence is passed to the simulator rather than a GE-EPI sequence and T_2 values are also passed to the simulator.

The output displacement fields now also account for the geometric distortions caused by susceptibility-induced off-resonance fields. These are calculated according to [112]:

$$\boldsymbol{\psi}(\mathbf{r}) = t_s N f(\mathbf{r}) \hat{\mathbf{p}} \tag{5.5}$$

where $\boldsymbol{\psi}(\mathbf{r})$ is the field of spatial displacements in each DWI that result from off-resonance fields (in voxels), defined at each location in the image $\mathbf{r}$, t_s is the echo spacing (in seconds), N is the number of phase-encode lines, $f(\mathbf{r})$ is the susceptibility-induced off-resonance field (in Hz), and $\hat{\mathbf{p}}$ is a dimensionless unit-vector that points along the phase-encode direction. The term $t_s N$ is also known as the readout time, and is the reciprocal of the bandwidth per pixel in the PE direction. The bandwidth per pixel is equal to the number of Hz in the off-resonance field that leads to a one-voxel displacement of signal along the PE direction. $f(\mathbf{r})$ now includes effects from susceptibility-induced fields, in addition to the eddy-current gradients discussed in Chapter 3.

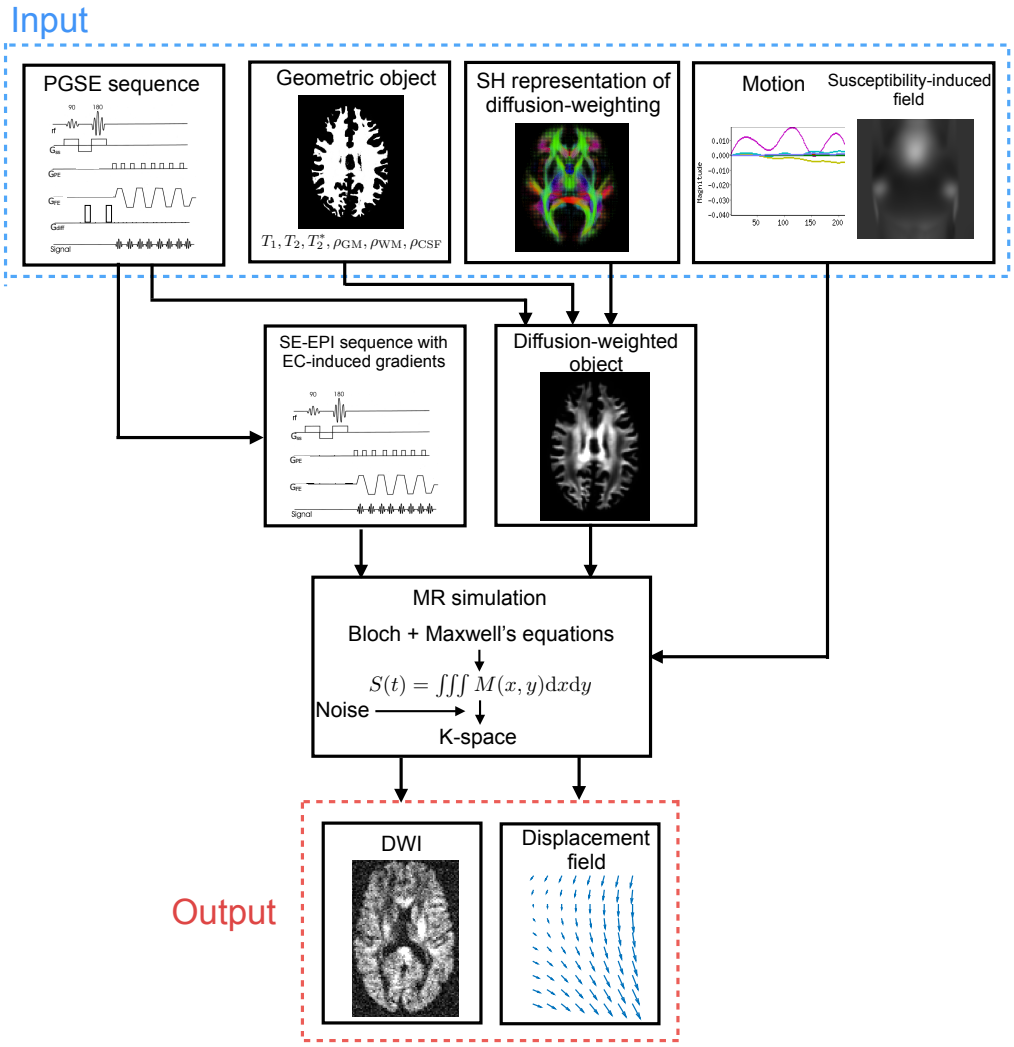
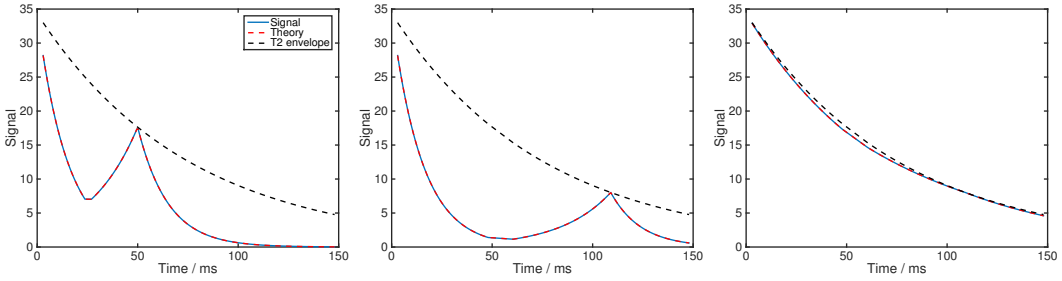


Figure 5.2: The new framework for simulating DW-MR data. Note susceptibility-induced fields are now included as an input, a SE-EPI sequence is passed to the simulator rather than a GE-EPI sequence and T_2 values are also passed to the simulator.

5.4 Validation

As discussed in the previous section, one of the main ways a spin-echo affects the magnetisation is by recovering signal lost to $T_2^{'}$ dephasing. We first examined this for the simple FID experiment, where signal is excited and followed by a spin-echo at time $T_E/2$, because an analytical solution for the signal in this situation is known. Figure 5.3 shows the agreement between theory and simulation for three spin-echo FID experiments performed on a homogeneous sample. This demonstrates that $T_2^{'}$ dephasing is being correctly reversed by



(a) T_E=50 ms, T_2^*=15 ms **(b)** T_E=109 ms, T_2^*=15 ms **(c)** T_E=109 ms, T_2^*=70 ms

Figure 5.3: Comparison of simulation with theory for a spin-echo free-induction decay experiment, with varying values of T_E and T_2^*. Spin-echo occurs at $T_E/2$. T_2 fixed at 75 ms.

the spin-echo in the simulations. Figure 5.4 demonstrates this qualitatively for the simulation of a full brain, by comparing GE and SE simulations over a whole brain slice; the increased signal in the GM and WM is evident.

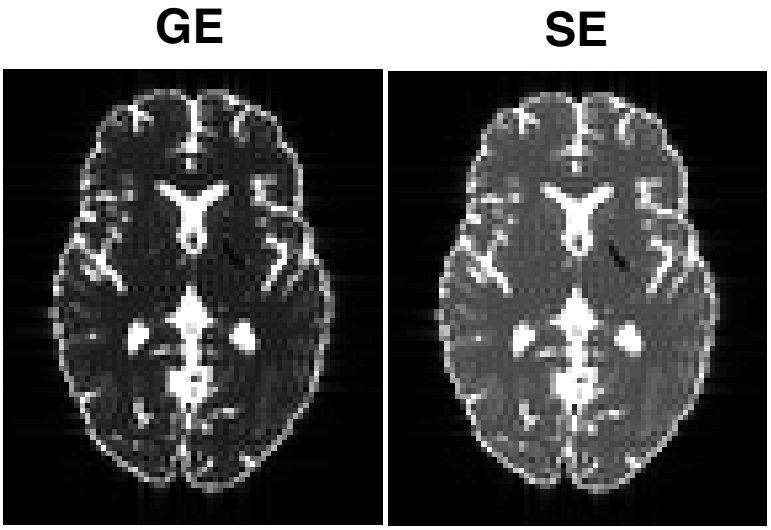


Figure 5.4: Comparison between GE-EPI and SE-EPI images simulated in POSSUM, shown on the same arbitrary intensity scale.

The second source of signal loss that the spin-echo is able to recover is that caused by gradient-induced dephasing across the voxel. To demonstrate the effects of this, we performed simulations both with and without a B_0 inhomogeneity field that varied linearly across the field of view, so that it had a constant gradient. We modified the simulator to prevent B_0 fields causing spatial offsets in image space, so that the effect of the gradient could be studied

in isolation. Figure 5.5 shows that this gradient causes signal loss for the GE-EPI sequence, but the SE-EPI sequence completely reverses this signal loss, causing the resultant image to be indistinguishable from the image produced in the case where no B_0 gradient exists. Figure 5.6 compares GE and SE sequences with spatial offsets re-enabled, and with a realistic B_0 inhomogeneity map. The figure shows spatial offsets similar to those expected in real data (the correctness of the spatial offsets simulated by POSSUM has already been extensively validated in [105]), as well as signal loss in the GE image that is not present in the SE image, as expected.

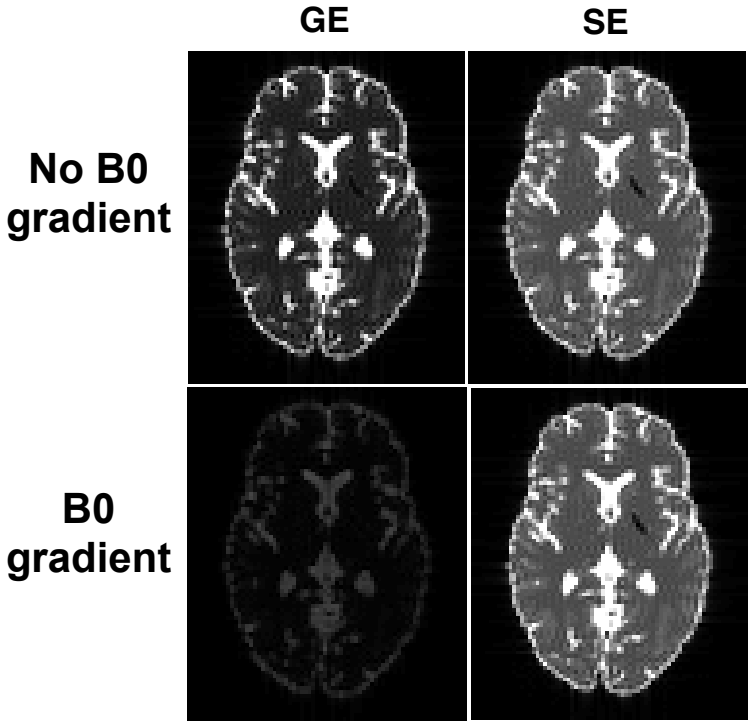


Figure 5.5: Simulations both with and without a linear B_0 gradient across the FOV; which should cause signal reduction the GE-EPI sequences but not SE-EPI.

5.5 Discussion

We have presented an extension of POSSUM that enables the simulation of spin-echo pulse sequences. The work enables us to incorporate the susceptibility artefact into the DW-MR simulation framework, which is important as

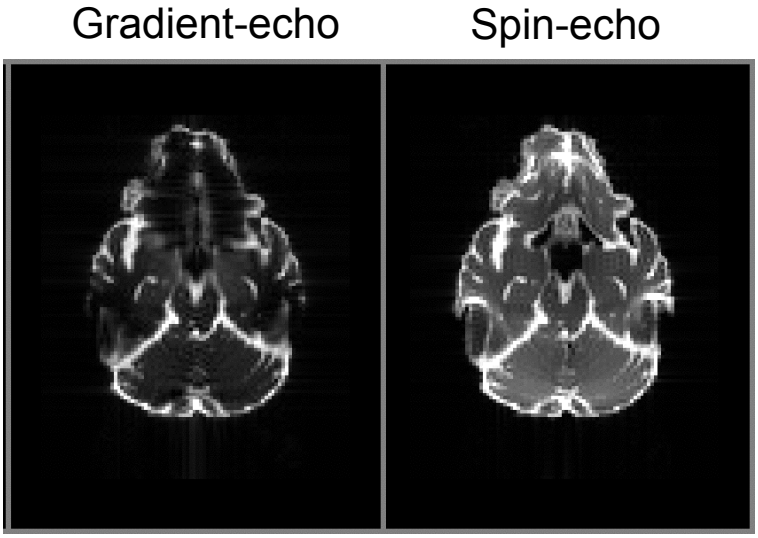


Figure 5.6: Comparison between GE-EPI and SE-EPI images simulated in POSSUM in the presence of a susceptibility-induced field. Both images are displayed with the same maximum intensity. All acquisition parameters matched. Compared to the SE image, the GE image has less signal (due to T_2^* decay) and dropout in regions of large susceptibility induced-fields. A different slice to that in Figure 5.5 has been chosen here to highlight the effects of the realistic B_0 field.

this artefact is always present in acquired data.

There are some limitations to the work presented here. The need to explicitly model the effects of a spin-echo on a single isochromat, rather than a dense group of spins, places some limitations on the type of sequences that can be simulated. For example, the model we implemented here would not be suitable for a pulse sequence with multiple 180 pulses within a single excitation, such as RARE [123]. The model would need to be modified in order to be made suitable for new pulse sequences. We also modelled the actions of a perfect 180 degree pulse. Imperfect pulses can lead to artefacts such as the line artefact that are typically dealt with by gradient crushers around the pulse [4, pp. 807–809]. Our assumption of perfect pulses mean that these crushers were not required.

Chapter 6

Application II: investigating the susceptibility artefact

6.1 Overview

In this chapter, we use our simulation framework to assess the three most commonly used post-processing methods for correcting the susceptibility artefact. We also investigate the interaction between susceptibility and movement, which is not accounted for by any of the commonly used methods, and show that it adversely impacts analysis of DW-MR data. We suggest its impact can be reduced by a simple adjustment of the acquisition, but that this could also be a valuable area for future methods development.

6.1.1 Research dissemination

This work described here is published in PLOS ONE.

- Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI. MS Graham, I Drobnjak, M Jenkinson, H Zhang. In *PLOS ONE* 12 (10), e0185647

The datasets simulated with movement-susceptibility interaction have also been used to validate a novel method for correcting this artefact, but this work is not discussed in this chapter:

- Susceptibility-induced distortion that varies due to motion: Correction

in diffusion MR without acquiring additional data, JLR Andersson, MS Graham, I Drobnjak, H Zhang, J Campbell. In *NeuroImage*, In press.

6.2 Introduction

DW-MR data is confounded by the presence of the susceptibility artefact, caused by an off-resonance field induced by differences in magnetic susceptibility at the air-tissue interface. When images are acquired with the spin-echo (SE) echo-planar imaging (EPI) sequence typically used in DW-MRI [124] this field causes geometric distortions in the data. If the subject remains static during acquisition these geometric distortions will be the same for each volume, resulting in diffusion datasets that are internally consistent (every volume contains the same distortions) but not anatomically faithful (the volumes do not match the subject's true anatomy). This has been shown to preclude accurate alignment to anatomically faithful structural data [125], a step that is often necessary for localising fine structures in the diffusion data, and to also introduce bias into results obtained from tractography [126–128]. We refer to this situation as the static susceptibility case. If the subject moves during acquisition the susceptibility field itself changes [40], altering the geometric distortions in the data, meaning that even after rigid realignment to correct for motion the diffusion datasets are both geometrically distorted and internally inconsistent (DW-MR volumes are misaligned relative to each other). In this case even analysis of the data that is not dependent on anatomical faithfulness, such as voxelwise fits to the data, will suffer from increased variability. We refer to this movement-induced change to the susceptibility field as the dynamic portion of the susceptibility artefact.

Recent trends in DW-MR are making it increasingly important that we have robust, well validated techniques for correcting this artefact. In the recent past, it has been common to reduce the impact of the susceptibility artefact at scan-time, by using in-plane parallel imaging techniques to reduce the number of phase-encoding (PE) steps. However, recently a number of high-profile

studies such as the Human Connectome Project (HCP), HCP lifespan and the UK Biobank have chosen to forego the use of these techniques in favour of SMS methods [129, 130], citing instabilities in reconstruction when the two are employed together [131, 132]. As these acquisition choices filter down to more 'everyday' studies there will be a concurrent increase in the severity of the susceptibility artefact. Furthermore, there is a trend, partly facilitated by the ability to image faster, towards acquiring datasets in more 'difficult' populations such as in the developing HCP [133]. These populations tend to move more in the scanner, further exacerbating problems caused by the interaction between susceptibility and motion.

There are a number of techniques available for correcting the susceptibility artefact. Correction is usually undertaken using post-processing strategies that may require the collection of some additional data. Broadly, these techniques can be divided into three types. The first involves registration of the data to a geometrically correct structural image [134–141]. The second type estimates a map of the B_0 inhomogeneities from acquired gradient-echo scans, and uses this along with some information about the diffusion acquisition protocol to correct for the distortions [95, 142–145]. The third estimates the underlying distortions using additional EPI data that is acquired with different phase-encoding (PE) and thus contains different distortions [96, 146–151]. This last class of technique offers the additional opportunity to accurately recover lost signal information if the full dataset has been acquired with reversed phase-encoding.

There are two classes of correction technique that rely on specialised pulse sequences not currently available on most scanners. Multi-reference approaches [112, 152, 153] are similar to fieldmap-based methods in that they involve the acquisition of additional reference scans to measure the geometric distortions in the data and correct in post-processing. Scan-time correction schemes estimate the fieldmap in real-time and correct it using gradient shims [154, 155]. These techniques are less commonly used and not examined in this chapter.

The majority of post-processing techniques assume a single susceptibility-induced field for the diffusion data and it is difficult to assess the impact of this assumption on the analysis of diffusion data. The problem has been investigated in the context of fMRI [95, 125], and a number of post-processing techniques suggested for its correction [156, 157] but they cannot be used to correct DW-MR data. This is because these methods either assume the off-resonance field can be measured from the phase [157], which is not true for diffusion data where the weighting can alter the phase, or because they assume all undistorted images in a time-series should have the same shape [156], which is not true if the images have different diffusion-weighting, which can alter the apparent location of the brain's outer surface.

It is important that we have available careful comparisons of susceptibility correction strategies, so that we are able to select the best for our processing pipelines. It is also vital that we are aware of the impact that their inability to correct for the dynamic portion of the artefact has on data analysis. To date, there are no systematic comparisons of existing methods for susceptibility correction and their limitations. A key reason for this is the difficulty in evaluating correction techniques. When validating on real data, the lack of any ground truth means evaluations are typically indirect [136, 150, 158] or qualitative [95, 96, 148]. Furthermore evaluations are often confounded by features in the data that are not of interest, such as other artefacts. Simulation can provide a ground-truth that enables direct, quantitative evaluation, and further allows for the careful design of experiments that enable the direct testing of the artefact of interest, without confounds.

In this work, we use simulation to undertake a comparison of the three classes of technique used for correction of the susceptibility artefact, and further characterise the impact of their inability to correct for the dynamic portion of the artefact. Our analysis directly measures the important outcomes for correction strategies: the ability to correctly estimate the underlying displacement field for correction, and the ability to recover information lost from

regions of signal compression. We also use the simulation framework to evaluate one of the most commonly used surrogate metrics for assessing susceptibility correction, the comparison of corrected datasets acquired with AP and LR phase-encoding, and use this surrogate to extend our comparison to real data. Finally, we quantify the increased variability in diffusion metrics caused by the dynamic susceptibility artefact.

6.3 Methods

The aim of this work is to assess the performance and limitations of existing methods for correcting the susceptibility artefact. Not only does the susceptibility-induced field produce geometric distortions in the data, but when the head rotates around an axis non-parallel to that of the main B_0 field, the susceptibility-induced field is altered and the artefact cannot be fully corrected using a field estimated before the head moved [40]. Whilst work has characterised [125] and attempted to correct [156, 157] this effect in fMRI, we are not aware of any available post-processing methods that address the issue in DW-MR. As a result we divide out analysis into three parts: in the first two, we compare existing methods for susceptibility correction on data with the susceptibility artefact but no head movement, using both simulated and real data. In the third, we characterise the impact of neglecting the movement-susceptibility interaction on the analysis on DW-MR data, using simulated data. In the following we describe the experiment design for each component of the assessment.

6.3.1 Assessment of existing techniques using simulated data

In this section we describe the comparison of existing methods for susceptibility correction on data with the susceptibility artefact but no head movement. In DW-MR the susceptibility artefact leads to geometric distortions of the data along the PE direction. The non-linear nature of these distortions mean they can cause redistribution of the signal which appears as either a compression

or stretch. In regions of compression some information is lost, and additional information is required in order to recover the true signal. An ideal susceptibility correction method will both correctly estimate the underlying geometric distortions and the true original signal.

There are three main classes of post-processing technique used for correcting the susceptibility artefact. Registration based (RB) techniques non-linearly register the distorted data to a non-distorted structural target, often a T2-weighted image due to its similar contrast to the b=0 volume. Fieldmap based (FMB) techniques estimate the off-resonance field from a series of images with different echo times, and then use this field to predict the underlying displacement field needed to correct for geometric distortions. Both RB and FMB techniques provide only a first-order correction of the signal changes, achieved by modulating the corrected image by the Jacobian of the local displacement field. Multiple phase-encoding based techniques (MPB) use multiple images acquired with different phase-encoding directions, and thus with different distortions, in order to estimate the underlying field needed to correct the data. If only a single DWI is acquired with multiple PE directions, the technique enables just the estimation of the field used to correct the dataset and employs the same first-order correction of signal intensity that is possible using RB and FMB techniques. If the full dataset is additionally acquired with multiple PE, these methods offer the added potential to recover the information lost from regions of signal compression, because these regions will instead be expanded in the reversed PE dataset. We refer to this special case of the MPB technique as full multiple phase-encoding based (MPB/F).

In order to assess these techniques, we simulated DW-MR datasets with susceptibility distortions, along with a T2-weighted structural image, field-mapping scans, and an additional DW-MR dataset with a reversed PE direction to enable application of the RB, FMB and MPB techniques, respectively. We designed the DW-MR datasets to contain levels of distortion similar to that found in recent high-end studies, such as the HCP [106] and UK Biobank,

which forego the use of in-plane acceleration techniques (IPAT) in favour of SMS techniques, in order to characterise the ability of these techniques to correct data in the 'worst case' scenario. The DW-MR data, shown in Fig 6.1, was simulated with 32 volumes b=1000 s/mm^2 and 4 b=0 volumes. We used a matrix size of 90×106, which was chosen along with the image voxel size (2 mm isotropic) and number of slices (68) to strike a balance between minimising computation time and ensuring full-brain coverage. The echo spacing was 1 ms, and no IPAT was used, leading to a PE bandwidth per pixel of 9.5 Hz, similar to values for data from the HCP project (9) and the HCP lifespan data acquired on a 3T Prisma scanner (10.4). Partial fourier was not used as POSSUM is currently unable to simulate it — this does not affect the level of susceptibility distortion in the data [159], but meant our TE of 109 ms is slightly higher than typical. K-space was apodized using a Hamming window, and no zero-filling was performed. Data was acquired with both posterior-anterior (PA) and anterior-posterior (AP) PE directions. Gaussian noise was added to the real and imaginary channels of the k-space data at two different levels, to produce datasets containing Rician noise with an average whole-brain SNRs of 40 and 20, as determined on the b=0 volume (as described in Chapter 4) — these represent the upper and lower bounds of SNR that we expect on modern scanners. Five realisations of each noise level were simulated, as well as a noise-free dataset. No other artefacts (e.g. eddy-currents, motion and concomitant fields) were included in the simulations. We also simulated a ground-truth set of DWIs, acquired with the same acquisition parameters but no input susceptibility field. The structural T2 was simulated with 1 mm isotropic resolution, dimensions 180×212×136 using a conventional spin-echo sequence with TE=110 ms, TR=2200 ms and a flip-angle of 90°. The field-mapping acquisition emulated the standard field-mapping scan found on a Siemens scanner, and involved the simulation of two gradient-echo images with the same voxel dimension and matrix size as the DW-MR scans, using a TR=700 ms, flip-angle of 60° and TE values of 4.92 ms and 7.38 ms. Noise was

added to these scans using the same standard deviation used in the DW-MR data, to simulate the same level of thermal noise in all datasets.

In this paper we tested a representative correction technique from each of the three classes. For the RB method we used the reg_f3d command from NiftyReg (Git commit bf926) [160], which uses a cubic b-spline deformation model. The first b=0 volume was registered to the T2-weighted structural, and the estimated transform applied to each DWI. We used default settings for the registration but constrained the deformation field along the PE axis, emulating standard practice for susceptibility correction [134, 136, 161], and after experimentation set the bending energy term to 0.01. For the FMB methods we used the following steps: 1) mask the first magnitude GE image; 2) erode the mask by one voxel; 3) estimate the fieldmap for all voxels inside the mask using FSL's PRELUDE (version 5.0.9) [162]; 4) apply the fieldmap to each DWI using FUGUE [163], smoothing the fieldmap using a 3D Gaussian kernel with sigma equal to 1 voxel (2 mm) as recommended in [125] and applying Jacobian modulation. For the MPB and MPB/F methods we used FSL's TOPUP (version 5.0.9) [96], using the default supplied configuration file. For the MPB/F case we changed the resampling from least-squares resampling (LSR) to Jacobian, after noticing that LSR introduced some slight 'ringing' artefacts into our corrected data. TOPUP combines each PE pair by averaging them after Jacobian modulation. After correction each dataset was transformed into the same space by rigidly registering its b=0 image to the noise-free, ground truth b=0 image using a 6 degrees-of-freedom (DOF) transform with NiftyReg's reg_aladin tool and then applying the estimated transform to each volume in the dataset.

The evaluation strategy is divided into three parts. Firstly, we assess the ability of each method to recover the correct underlying displacement field, and thus produce anatomically faithful data, by comparing each method's estimated field to the ground truth field obtained using Equation 5.5. Secondly, we assess the ability of each method to recover the correct intensity at each

voxel by computing difference maps between the corrected and ground truth images. Finally we investigate the impact of correction quality on subsequent analysis by comparing diffusion tensor (DT) fits in both corrected datasets and 'ground truth' datasets, simulated free of artefacts.

In addition to the experiments described, we investigated a surrogate metric for correction quality that is often used to assess performance on real data: the comparison of corrected datasets that have been acquired with both AP and LR phase-encoding [136, 150]. The expectation is that the greater the correction quality, the greater the similarity between the corrected datasets. We aimed to evaluate whether this is a suitable surrogate for the most direct measure of correction quality, i.e. the displacement field error. To enable this experiment we simulated additional DW-MR data with LR and RL phase-encoding. These datasets were correcting using the same methods described above, and then compared to the corrected datasets acquired with AP and PA phase-encoding.

6.3.2 Assessment of existing techniques using real data

In this section we extend our evaluation of existing techniques to real data. To enable this evaluation we used the surrogate metric described in the previous section, the comparison of corrected datasets with both AP and LR phase-encoding. We used ten subjects from the developing HCP project [164]. We selected this dataset because it provides DW-MR data acquired with four PE directions: AP, PA, LR and RL, fieldmaps and structural data, enabling correction using all the methods used in this paper and evaluation using the surrogate metric. The data was acquired on a 3T Philips Achieva, consisting of a spherically optimized set of directions on 4 shells (b0: 20, b400: 64, b1000: 88, b2600: 128) split into four PE subsets. It was acquired using an acceleration of MB 4, SENSE factor 1.2 and partial fourier 0.86, TR/TE 3800/90 ms. The acquired resolution is 1.5x1.5 mm, 3 mm slices with 1.5 mm overlap, reconstructed to give data of resolution 1.17x1.17x1.5 mm. The T2-weighted image had TR/TE 12000/156 ms with a reconstructed resolution of

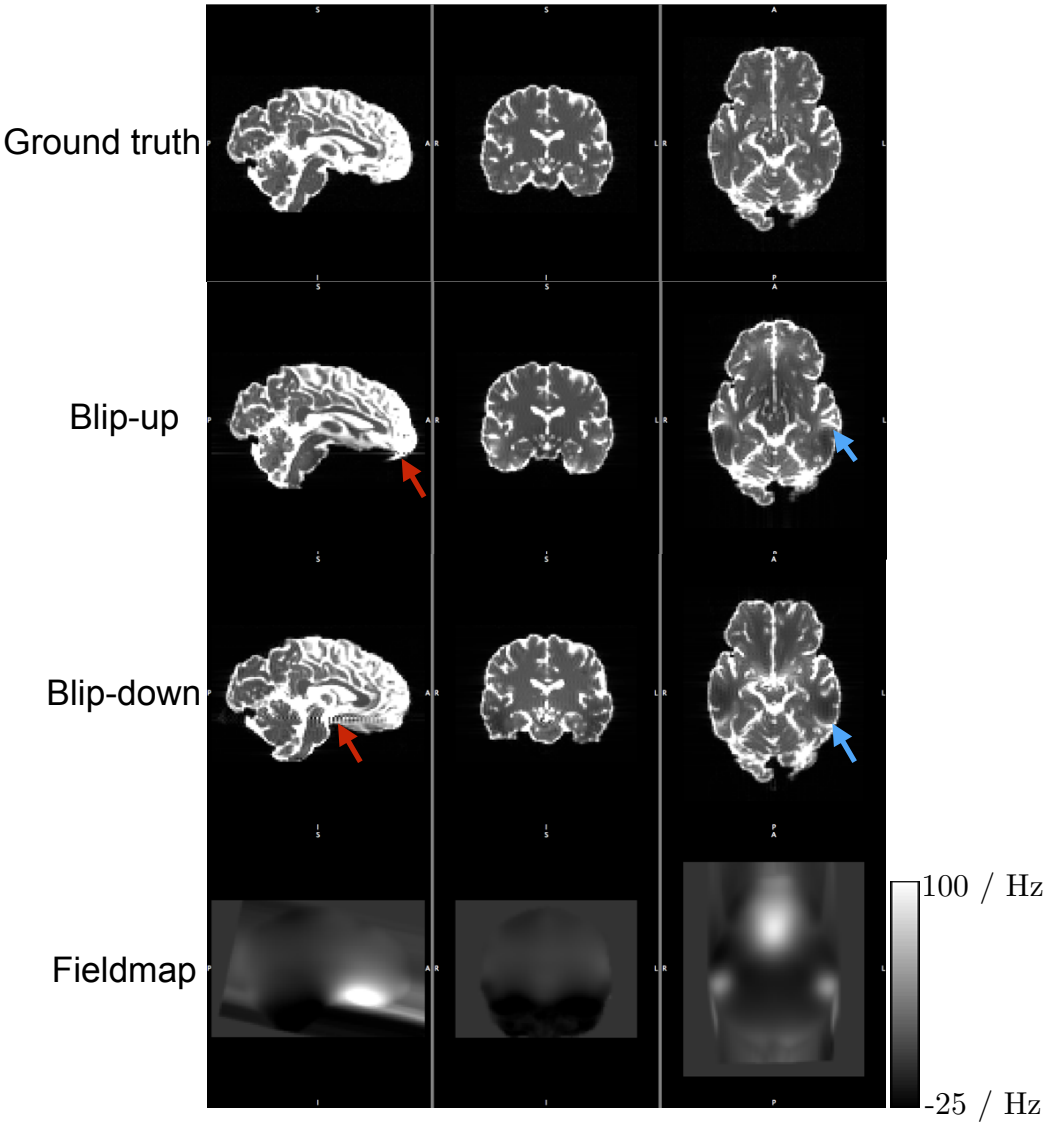


Figure 6.1: Simulated DWIs. Pairs of coloured arrows point to corresponding pairs of compression and expansion in the blip-up and blip-down images. The 'streaking' visible in the fieldmap is caused by the linear extrapolation to ensure a continuous field at the edge of the brain. The bounding box visible around the fieldmap is caused by its resampling into the space of POSSUM's input object; this is not a problem for the simulation as the fieldmap is smooth and defined over all brain voxels in input object. Some Gibbs ringing is visible in the sagittal views of the distorted data — this is induced by sharp boundaries in regions of signal pile-up.

$0.8\,\mathrm{mm}^3$. AP and LR datasets were corrected separately using the correction methods as described previously. Corrected b=0 images were rigidly registered to a T2-weighted image using a 6 degrees-of-freedom (DOF) transform with NiftyReg's reg_aladin tool and the similarity between AP and LR corrected images assessed.

6.3.3 Assessment of the susceptibility-movement interaction

In this section we describe the experiments performed to characterise the impact of the interaction between the susceptibility field and head motion. This is an effect that none of the commonly used post-processing correction strategies currently account for in DW-MR. To investigate the impact of this on analysis of data, we compare state-of-the-art correction on datasets simulated with and without a dynamic susceptibility field.

DW-MR data was simulated with the same parameters as in the previous section, using a full AP and PA acquisition totalling 72 volumes. A key difference to the previous section's simulations is that here we calculated the field from an air-tissue segmentation of POSSUM's input object using a perturbation method as described in [165]. This method takes as input a spatial map of susceptibility values throughout the brain, obtained from a tissue-air segmentation, and obtains a first-order solution to Maxwell's equations by considering the tissue susceptibility to be equal to the susceptibility of air plus a perturbation. The method is physically motivated, providing realistic fields, and provides a set of basis-functions that enable POSSUM to calculate how the susceptibility-induced field changes as the head moves. Movement was simulated during the data acquisition, using motion parameters measured from a healthy patient during an MRI exam, to emulate a scan with a normal level of motion. Movement was simulated to occur between the acquisition of volumes. A second dataset was simulated with the same set of parameters but the level of motion scaled up by a factor of three, to emulate a situation where a patient moves a lot. For all the simulations, the translation parameters were set to 0

as, on the assumption that the main field is entirely uniform, translations do not contribute to the dynamic susceptibility effect and their inclusion would require the imaging FOV to be increased, which would increase computation time. We also create an additional set of simulations to control for the effects of image interpolation on our correction. These control datasets were simulated assuming the static case where the susceptibility field remains unchanged as the head moves — this situation matches the assumptions of existing susceptibility correction techniques. All datasets were simulated without noise.

Each dataset was corrected to mimic 'state-of-the-art' correction, in which motion and the static portion of the susceptibility field are corrected for. Ground truth displacement fields for each volume were created from the motion parameters and the static portion of the field, both of which are known inputs to the simulation, and applied. Two sets of correction were performed: in the first, only the AP dataset was corrected, and in the second, joint correction of the AP and PA dataset was carried out to enable correction in regions of compression (this involves separate correction of the AP and PA images using Jacobian modulation to account for compression/stretching, followed by an averaging of the two resultant images). The impact of residual distortions caused by the dynamic portion of the susceptibility field was demonstrated using errors in displacement fields, and the impact of the dynamic field on subsequent analysis was measured by characterising the errors in estimated FA values in corrected data.

6.4 Results

6.4.1 Assessment of techniques with simulated data

Figure 6.2 shows the errors in the displacement fields estimated by the three methods across a representative slice of the brain (the MPB and MPB/F methods use the same displacement field, estimated from just the b=0 images with reversed PE). Full results for the five noise realisations are shown in Table 6.1a. It is immediately clear that the RB method is unable to accurately estimate

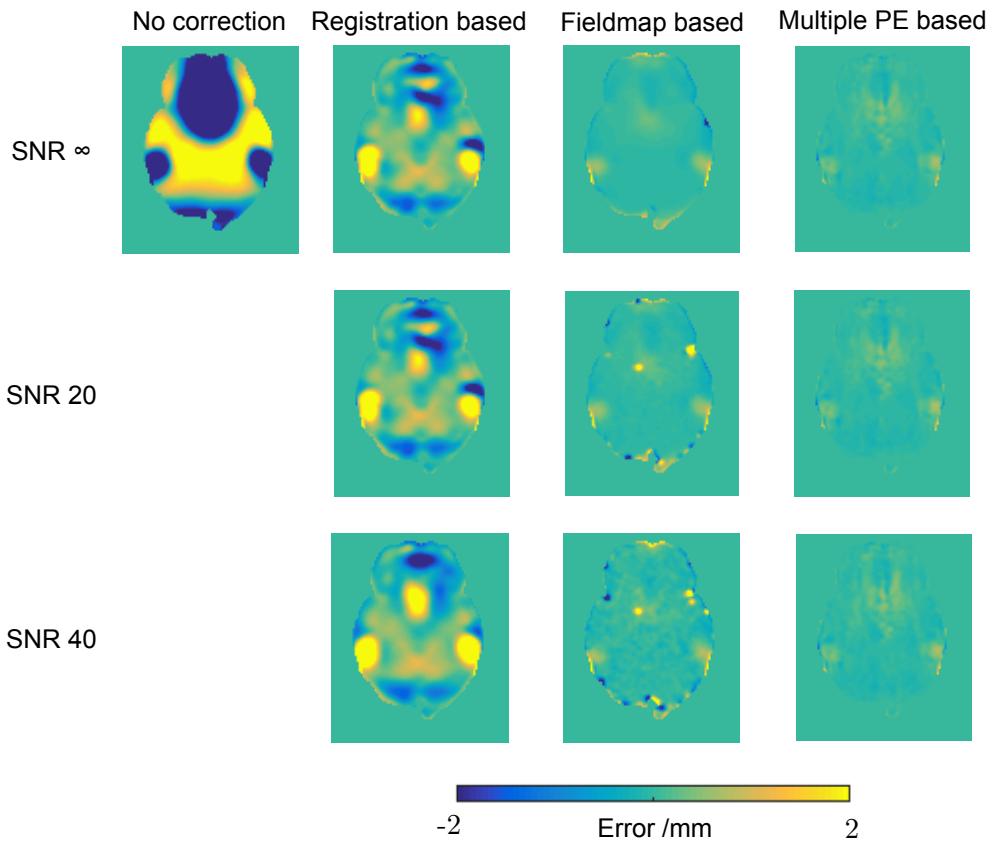


Figure 6.2: Displacement field errors. Error in displacement fields estimated by the three methods, assessed by subtraction from the ground truth field. One axial slice shown.

the underlying displacement field, whilst the FMB and MPB methods show better performance. The FMB method shows some errors in brain voxels that contain partial volume with air, around the edges of the brain, and these errors are introduced into brain voxels when the estimated fieldmap is smoothed. When the edge voxels are excluded using an eroded brain mask, the mean absolute errors per voxel reduce more for the FMB method than other methods (Table 6.1b). The FMB's difficulty estimating the field in edge voxels is exacerbated as the noise level increases, whilst the MPB method is relatively unaffected by noise.

The impact that the displacement field estimation has on the corrected data is demonstrated in Fig 6.3, which shows the b=0 images after correction,

Table 6.1: Error tables.

	Registration	Fieldmap	Multiple PE
SNR ∞	0.24	0.051	0.032
SNR 40	0.263 ± 0.011	0.071 ± 0.003	0.032 ± 0.000
SNR 20	0.272 ± 0.013	0.087 ± 0.002	0.036 ± 0.000

(a) Mean of absolute errors in displacement field across the brain. Values shown are the mean across the five noise realisations, and errors are the standard deviation of the mean value for each noise realisation. Note that the multiple phase-encode results cover both MPB and MPB/F methods.

	Registration	Fieldmap	Multiple PE
SNR ∞	0.24	0.024	0.029
SNR 40	0.249 ± 0.010	0.035 ± 0.001	0.029 ± 0.000
SNR 20	0.251 ± 0.012	0.048 ± 0.002	0.033 ± 0.000

(b) The same metrics in Table 6.1a but calculated over an eroded brain mask. Values shown are the mean across the five noise realisations, and errors are the standard deviation of the mean value for each noise realisation.

		GT + noise	RB	FMB	MPB	MPB/F
SNR ∞	b=0	0.00	2.41	1.36	0.73	0.59
	DWI	0.00	0.70	0.20	0.13	0.09
SNR 40	b=0	0.23 ± 0.00	2.55 ± 0.04	1.71 ± 0.04	0.78 ± 0.00	0.62 ± 0.00
	DWI	0.23 ± 0.00	0.75 ± 0.01	0.35 ± 0.01	0.27 ± 0.00	0.19 ± 0.00
SNR 20	b=0	0.70 ± 0.00	2.81 ± 0.03	2.09 ± 0.02	1.05 ± 0.00	0.80 ± 0.00
	DWI	0.70 ± 0.00	1.00 ± 0.00	0.69 ± 0.00	0.68 ± 0.00	0.49 ± 0.00

(c) Absolute errors in image intensity, averaged across the brain for all b=0 and DWI volumes. Values shown are the mean across the five noise realisations, and errors are the standard deviation of the mean value for each noise realisation. Units are arbitrary signal units.

		GT + noise	RB	FMB	MPB	MPB/F
SNR ∞	FA	0.000	0.017	0.014	0.010	0.008
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.000	0.077	0.073	0.048	0.040
	V1 / degrees	0.000	3.602	1.693	1.008	0.687
SNR 40	FA	0.016	0.027	0.021	0.020	0.014
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.022	0.087	0.090	0.055	0.046
	V1 / degrees	3.787	6.418	3.773	3.825	2.704
SNR 20	FA	0.051	0.053	0.044	0.049	0.033
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.078	0.134	0.122	0.096	0.078
	V1 / degrees	11.580	11.775	9.872	10.931	7.744

(d) Errors in diffusion metrics (FA, MD and the principle diffusion direction V1), averaged across the brain. Values shown are the mean across the five noise realisations. V1 errors were only calculated in voxels with a ground-truth FA >0.2. Errors (calculated as the standard deviation of the mean value for each noise realisation) not shown as they were all 0 to 3 decimal places.

along with error maps obtained by subtraction from the ground truth images. These results are shown in full in Table 6.1c. The effects of the RB method's poor estimation are clear in these results. The effect of the FMB method's poor displacement field estimates in edge voxels is apparent here. Both the MPB and MPB/F methods show better results. The figure demonstrates a region of higher error in the FMB and MPB methods, due to their inability to recover the correct signal from a region of compression, where the MPB/F does better, due to its ability to resample from the corresponding expanded region in the reversed PE image. There is some slight ringing noticeable in the error maps, particularly for the MPB and MPB/F methods. This is Gibbs ringing present in the ground-truth b=0 images, caused by the strong CSF rim. The corrected images have much reduced ringing because they have been smoothed by interpolation, so the ringing is visible upon subtraction. The ringing appears less visible for RB and FMB methods because it is obscured by larger errors caused by poorer correction.

Table 6.1c demonstrates an additional advantage of the MPB/F method: the SNR boost obtained for each corrected image by resampling from two images, in effect averaging over the noise. It causes the mean errors for the DW volumes to be lower than the errors in the noisy ground-truth images. The results entangle two effects: the ability to recover information from regions of compression, and an SNR boost from having twice as much data. These can be disentangled by examining the SNR infinite case in Table 6.1c, where the improvements from the MPB/F are solely due to improved signal recovery in regions of compression.

Fig 6.4 show FA maps estimated from the corrected datasets for one slice of the brain, as well as their errors, with full results for FA, MD and the principal diffusion direction (V1) shown in Table 6.1d. The downstream effect of information loss from areas of compression is appreciable in both the FMB and MPB maps, as is the ability of the MPB/F method to mitigate these errors. The SNR boost provided by the MPB/F method is visible in the figure,

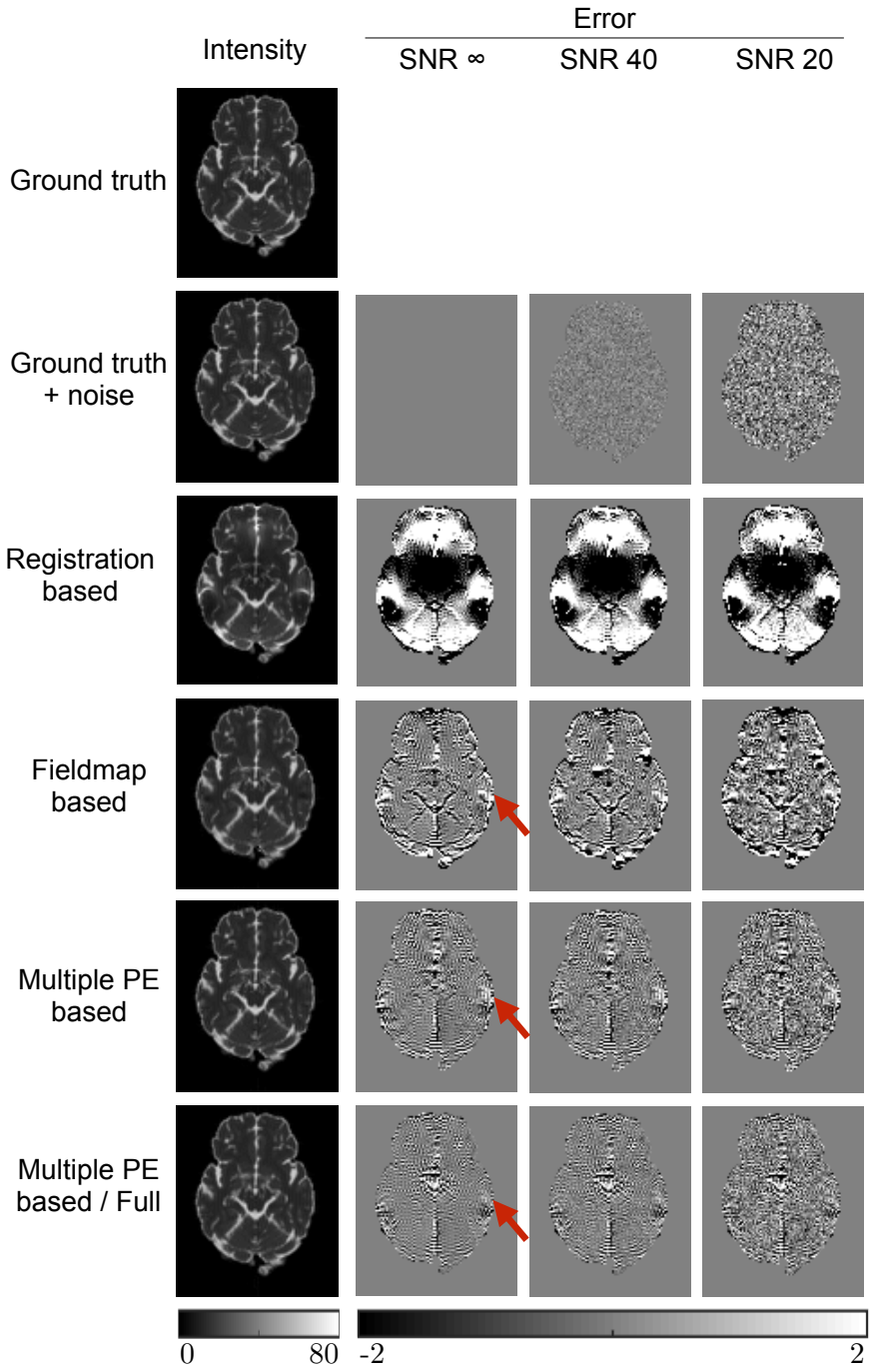


Figure 6.3: Errors in image intensity. b=0 images after correction by each method, along with the ground truth images shown both with and without noise. Intensity images shown for the infinite SNR case. Error maps are obtained by subtraction from the noise-free ground truth image. Units are arbitrary signal units. Red arrows highlight a region of signal compression that can only be corrected by the MPB/F method. Note the MPB/F method uses twice as much data as the other methods, increasing its effective SNR.

which shows some regions have lower error than the noisy ground truth maps, and is also clear from the tabulated results. The results in Table 6.1d also demonstrate that the amount of smoothing introduced by a method can affect results. For noisy data the errors for corrected datasets are sometimes smaller than the error for the 'ground truth + noise' datasets, which reflects the noise-reducing effects of the local smoothing introduced upon correction. This effect is particularly strong for the FMB method, likely due to the smoothing of the estimated fieldmap during processing, causing it to produce smaller average errors for FA and V1 than the MPB method, despite MPB estimating the underlying displacement field more accurately. Dividing the errors according to the size of the underlying distortion (Table 6.2) confirms that it is the smoothing that causes this; FMB tends to outperform MPB in areas of low distortion where errors are mostly controlled by the amount of noise, whilst MPB outperforms FMB in regions of large distortion where estimation of the correct underlying displacement field is important.

To investigate the suitability of AP-LR differences as a surrogate metric we plot the corrected AP and LR b=0 images, and their differences, for a representative slice in Figure 6.5. The whole-brain mean of the intensity difference for b=0 volumes was computed for every correction method, for each of the five noise realisations at each SNR — results are shown in Table 6.3a. The results show the largest errors for the RB method with MPB and MPB/F performing the best, consistent with previous results. A two-sample t-test without assuming equal variance was performed between these values for each pair of correction methods; all differences were significant at the $p<0.001$ level. These results demonstrate that the surrogate metric shows the same ordering of correction ability as more direct metrics, such as error in displacement field, indicating that it can be a useful metric. However, it should be noted that the metric does not give the same contrast between methods as displacement field error. Errors in displacement fields show large differences between the RB and FMB methods, and a much smaller difference between FMB and MPB

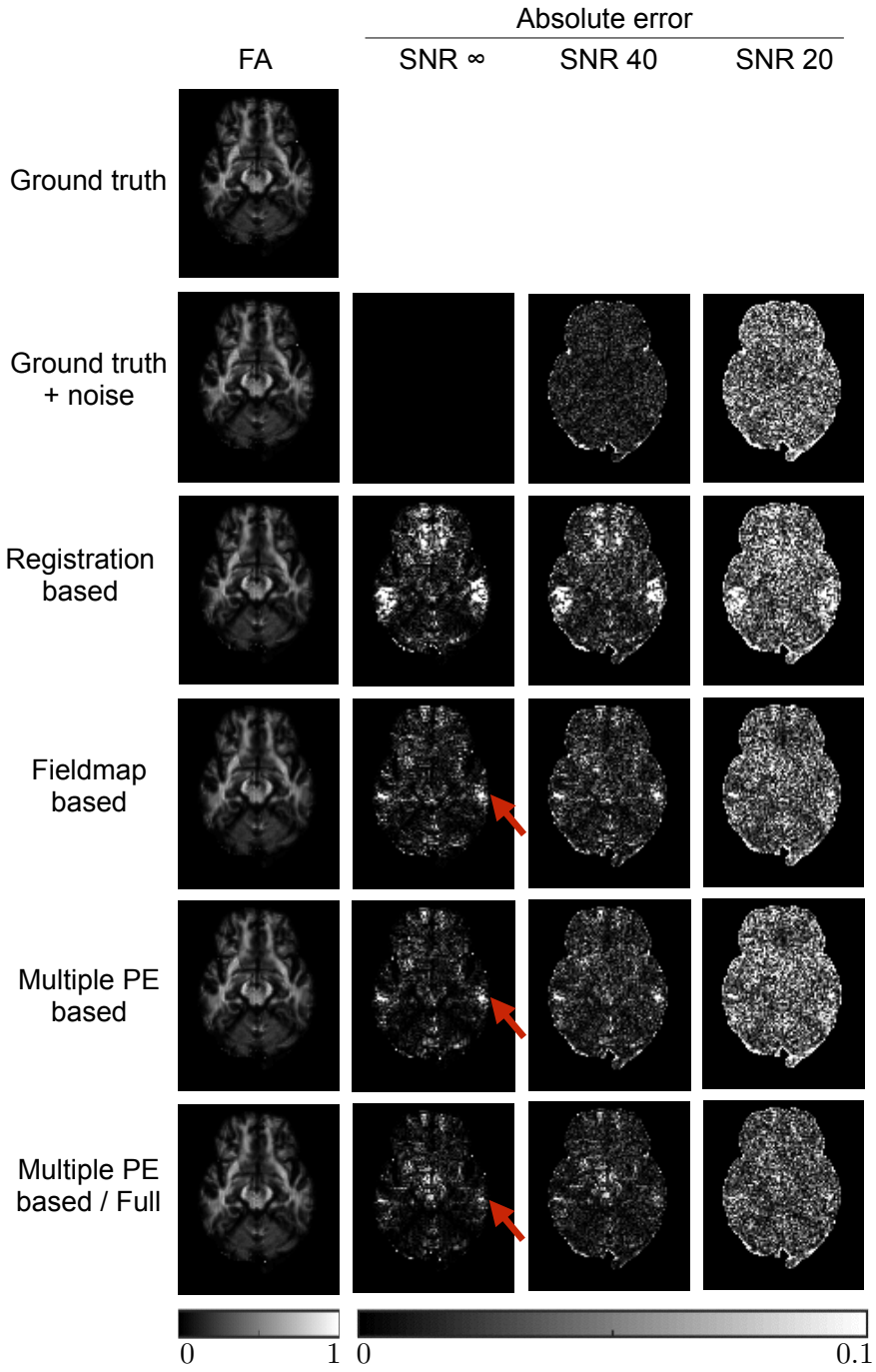


Figure 6.4: Errors in FA metrics. FA maps estimated from corrected and ground truth images, along with error maps obtained by subtraction from the noise-free ground truth estimate. FA map shown for SNR infinite case. Red arrows show regions of high error caused by signal pileup that could not be corrected by the RB and MPB methods, despite estimation of the correct displacement field. The MPB/F method is able to reduce errors in these regions. Note the MPB/F method uses twice as much data as the other methods, increasing its effective SNR.

		GT + noise	RB	FMB	MPB	MPB/F
SNR ∞	FA	0.000	0.049	0.024	0.019	0.016
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.000	0.077	0.073	0.048	0.040
	V1 / degrees	0.000	3.602	1.693	1.008	0.687
SNR 40	FA	0.015	0.054	0.031	0.026	0.020
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.015	0.154	0.110	0.078	0.072
	V1 / degrees	4.185	20.363	5.495	5.201	3.564
SNR 20	FA	0.049	0.083	0.054	0.052	0.036
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.056	0.193	0.129	0.103	0.087
	V1 / degrees	12.470	21.283	11.352	11.865	9.057

(a) Large distortion. Errors in diffusion metrics averaged across all brain voxels with >6 mm geometric distortion.

		GT + noise	RB	FMB	MPB	MPB/F
SNR ∞	FA	0.000	0.021	0.016	0.012	0.010
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.000	0.081	0.085	0.053	0.044
	V1 / degrees	0.000	4.923	1.919	1.276	0.826
SNR 40	FA	0.016	0.029	0.023	0.020	0.015
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.021	0.092	0.105	0.059	0.049
	V1 / degrees	3.892	7.783	3.918	3.970	2.833
SNR 20	FA	0.050	0.052	0.045	0.047	0.033
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.072	0.128	0.130	0.093	0.075
	V1 / degrees	11.736	12.663	9.858	10.976	7.875

(b) Medium distortion. Errors in diffusion metrics averaged across all brain voxels with >2 mm and <6 mm geometric distortion.

		GT + noise	RB	FMB	MPB	MPB/F
SNR ∞	FA	0.000	0.013	0.013	0.009	0.007
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.000	0.070	0.066	0.043	0.036
	V1 / degrees	0.000	2.299	1.470	0.762	0.558
SNR 40	FA	0.016	0.023	0.019	0.019	0.013
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.023	0.081	0.082	0.052	0.043
	V1 / degrees	3.711	4.836	3.589	3.665	2.586
SNR 20	FA	0.052	0.050	0.043	0.049	0.033
	MD / $10^{-3}\,\mathrm{mm^2\,s^{-1}}$	0.083	0.132	0.118	0.097	0.078
	V1 / degrees	11.448	10.716	9.777	10.846	7.593

(c) Small distortion. Errors in diffusion metrics averaged across all brain voxels with <2 mm geometric distortion.

Table 6.2: As in Table 1C, errors for FA, MD and the principle diffusion direction V1, but here divided into regions of interest based on the amount of distortion in the data. Values shown are the mean across the five noise realisations. V1 errors were only calculated in voxels with a ground-truth FA >0.2. Errors (calculated as the standard deviation of the mean value for each noise realisation) not shown as they were all 0 to 3 decimal places.

Table 6.3: Surrogate metrics. Table shows whole-brain-mean intensity differences between AP and LR corrected datasets (units are arbitrary signal units).

	RB	FMB	MPB	MPB/F
SNR 40	3.791 ± 0.044	2.818 ± 0.047	1.904 ± 0.100	1.498 ± 0.025
SNR 20	3.976 ± 0.026	3.230 ± 0.041	1.989 ± 0.022	1.688 ± 0.069

(a) Simulated data. Errors are the standard deviation of the means over the five noise realisations. Metrics show statistically significant differences between all methods at the $p<0.001$ level.

RB	FMB	MPB	MPB/F
6.463 ± 1.282	5.277 ± 1.278	3.579 ± 0.885	3.078 ± 0.965

(b) Real data. Errors are the standard deviation of the means over the ten subjects. Metrics show statistically significant differences between all methods at the $p<0.001$ level.

methods. The surrogate metric loses this contrast, indicating a roughly similar improvement going from RB and FMB methods as FMB to MPB methods. This is because there is not a simple relationship between displacement field error and intensity error; the size of intensity error depends on both the size and the location of the displacement error.

6.4.2 Assessment of techniques with real data

Figure 6.6 shows the differences between AP and LR correction for real data. These results resemble findings in simulated data, indicating that RB methods perform the worst and MPB/F best. Table 6.3b reports the whole-brain mean of the intensity difference for b=0 volumes. A paired t-test was performed between these intensity differences for each method, all differences were significant to $p<0.001$.

6.4.3 Interaction between susceptibility and movement

Fig 6.7a shows the rotation parameters used for the simulations. Fig 6.7b shows the residual errors in the first four DWIs after correction for motion and static susceptibility. We found that a 5° rotation about the y-axis caused changes in the susceptibility field of up to 30 Hz, corresponding to distortions of

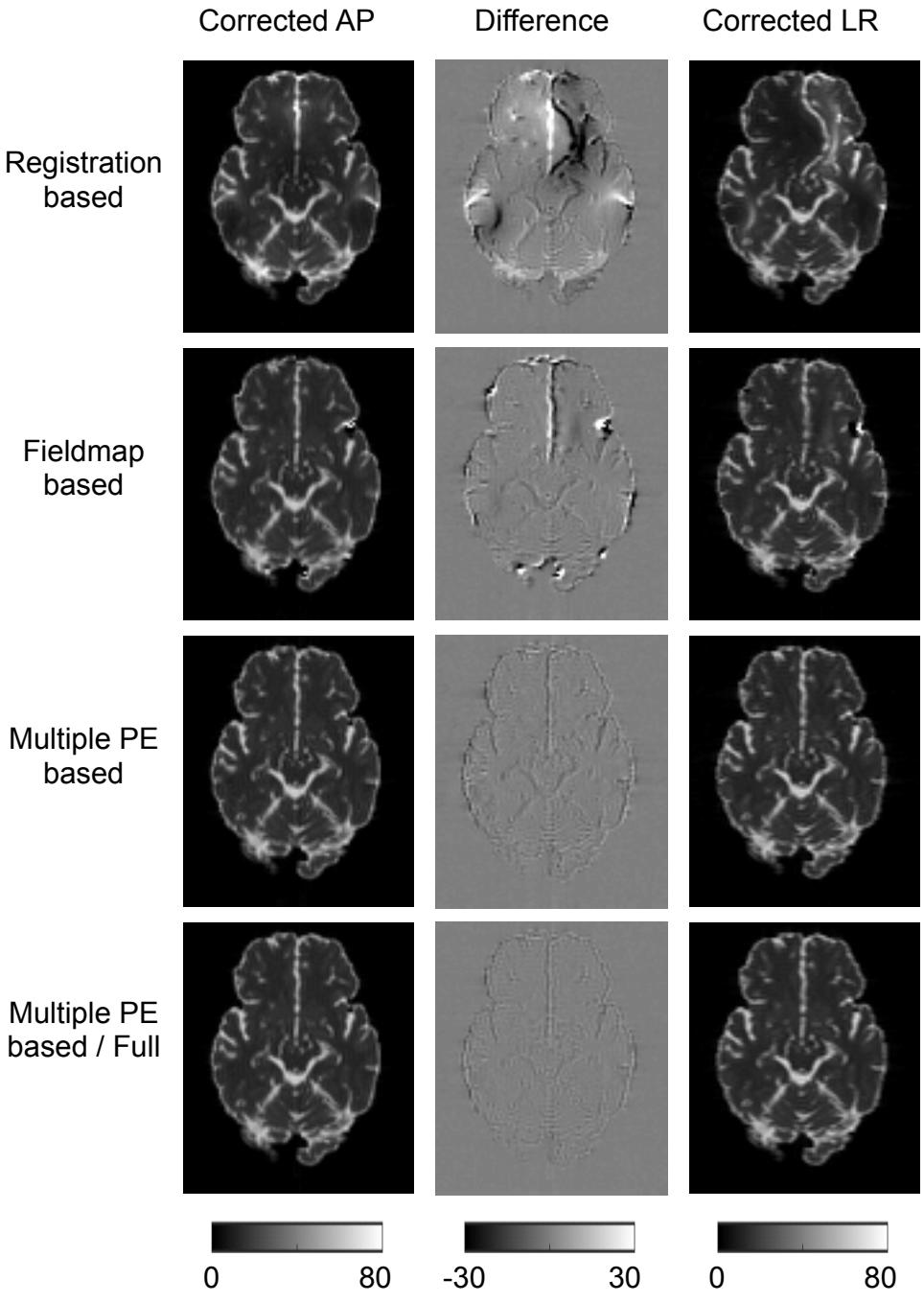


Figure 6.5: AP-LR comparison on simulated data. Figure shows corrected AP and LR b=0 images, and the intensity difference between them. SNR=40 dataset shown.

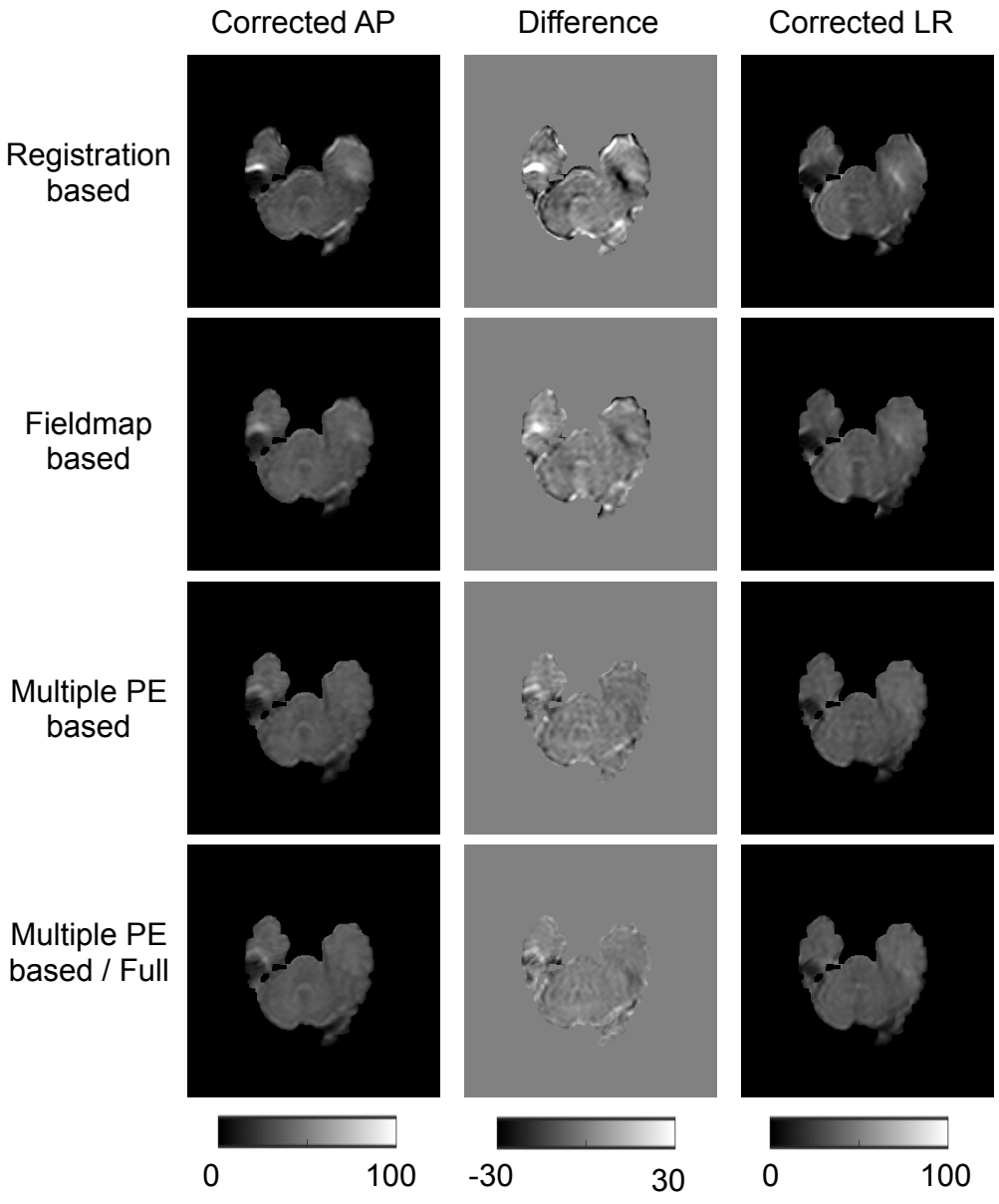


Figure 6.6: AP-LR comparison on real data. Figure shows corrected AP and LR b=0 images, and the intensity difference between them.

up to 6 mm for the acquisition protocol used. This is slightly smaller than the field changes measured in real data — [40] found changes of 50 Hz for similar rotations at 3 T — indicating our dynamic distortions are in a realistic range but may slightly underestimate the true size of the effect. Our simulations show a left-right asymmetry in dynamic displacement fields for rotations around y (Fig 6.7b, Volume 4) and left-right symmetry for rotations around x (Fig 6.7b, Volume 5) that matches observations made in real data [156].

Fig 6.8 shows the errors in FA for corrected data across an example slice. Comparing data simulated with static and dynamic susceptibility fields, we see increased errors in the data with dynamic effects. The differences in static and dynamic errors across the full brain are shown in Fig 6.9. It is worth noting that the errors for the data with static susceptibility fields are non-zero, despite the data being noise free and corrected with the ground truth fields. These errors are introduced by interpolation effects, and introduce similar levels of error into both static and dynamic data, so it is the difference in error between these two sets of data that reveals the errors introduced by the dynamic susceptibility fields alone. The results show that the dynamic field increases FA errors especially for the case of the subject with larger motion. The effect of the dynamic field was slightly smaller for data corrected using both AP and PA volumes than data corrected using only the AP set.

6.5 Discussion

In this paper we assessed the three main classes of technique used for correcting the susceptibility artefact, and investigated the impact of their inability to correct for the dynamic portion of the artefact. This work is particularly timely given that recent trends in acquisitions could lead to increased severity of the artefact, increasing the importance that the community has access to careful evaluations of the correction techniques available to them and their limitations. To enable our assessment, we extended an existing MR simulator and incorporated it into a framework that simulates realistic DW-MR datasets.

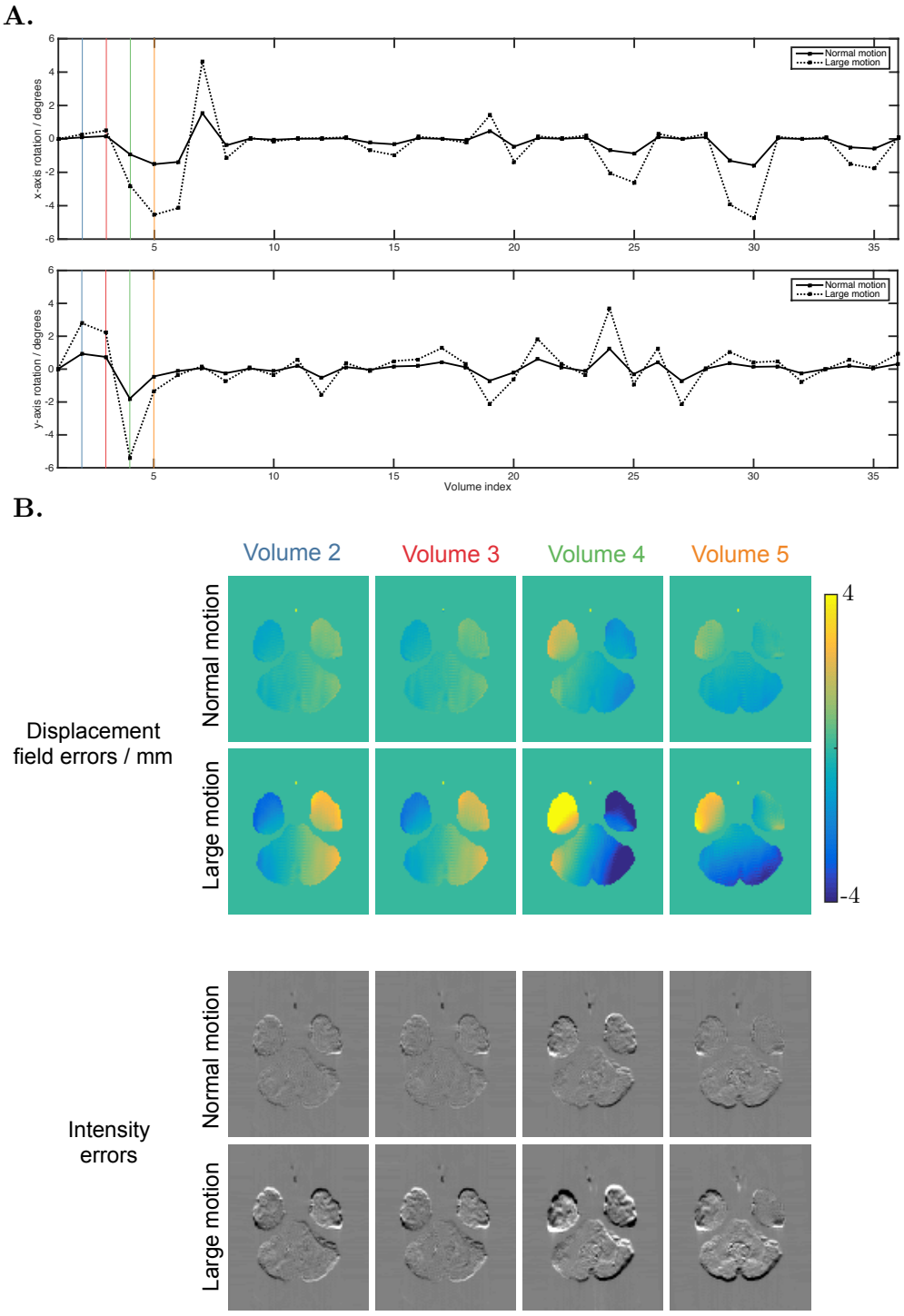


Figure 6.7: Susceptibility-movement interaction. A. The x- and y- rotation parameters used for the simulation of the first 36 volumes (z-rotations not shown because they do not contribute to the dynamic susceptibility effect, translations were all 0). The coloured vertical lines highlight the motion of the volumes depicted in plot B. B. Top two rows show the errors in displacement field caused by the dynamic portion of the susceptibility artefact, for volumes 2-5 of the acquisition — the motion these volumes experienced is highlighted with colour in plot A. Bottom two rows show the error in intensity of these volumes after they are corrected for motion and the static portion of the susceptibility field, obtained by subtraction from ground truth images.

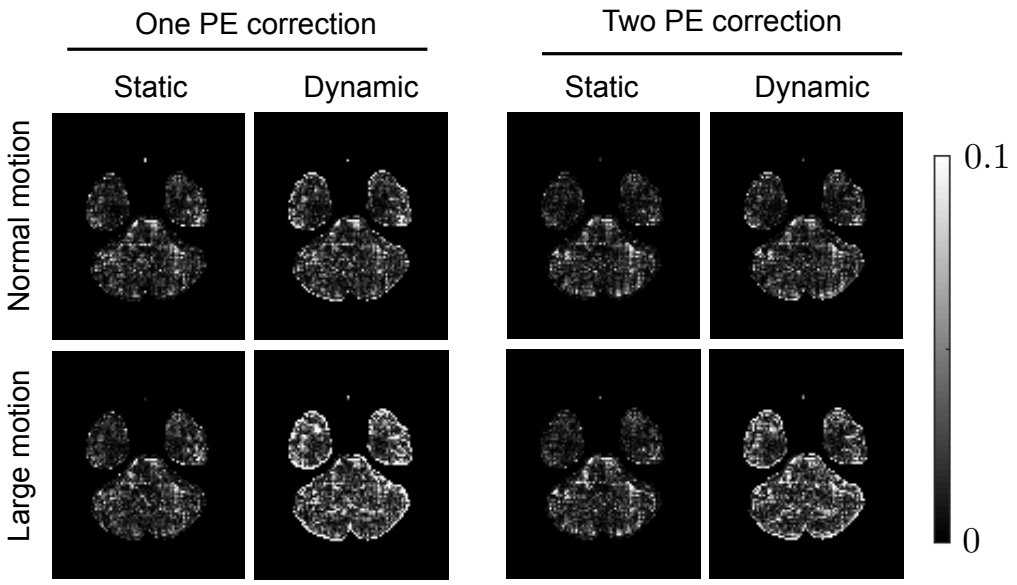


Figure 6.8: Errors introduced when failing to account for the susceptibility-movement interaction. Absolute errors in FA shown over one slice, for datasets corrected for motion and static susceptibility. Data in the static columns were simulated with only motion and static susceptibility artefacts, whilst the dynamic data contained motion and dynamic susceptibility. 'One PE correction' indicates only the AP data was used for correction, and 'two PE correction' indicates the AP and PA data were both used — note these are different from MPB and MPB/F, which are methods for both estimating and applying a displacement field, whilst in this case known ground-truth displacement fields have been applied.

This enabled us to directly and quantitatively assess the desired features of a susceptibility correction approach: the ability to both correct geometric distortions and recover the signal distribution lost in regions of compression. It further enabled us to carefully examine the impact that neglecting the dynamic susceptibility field has on analysis of diffusion data.

Our results showed that registration of distorted data to a structural T2-weighted volume was insufficient for fully correcting geometric distortions in the data. However, it offers improvement over performing no correction at all, and has the advantage that a structural volume is often acquired in a scan, making it suitable for retrospective studies where data required for FMB or MPB methods is not available (contrast inversion techniques may mean similar results can be obtained with a T1-weighted volume [141, 158]). FMB and

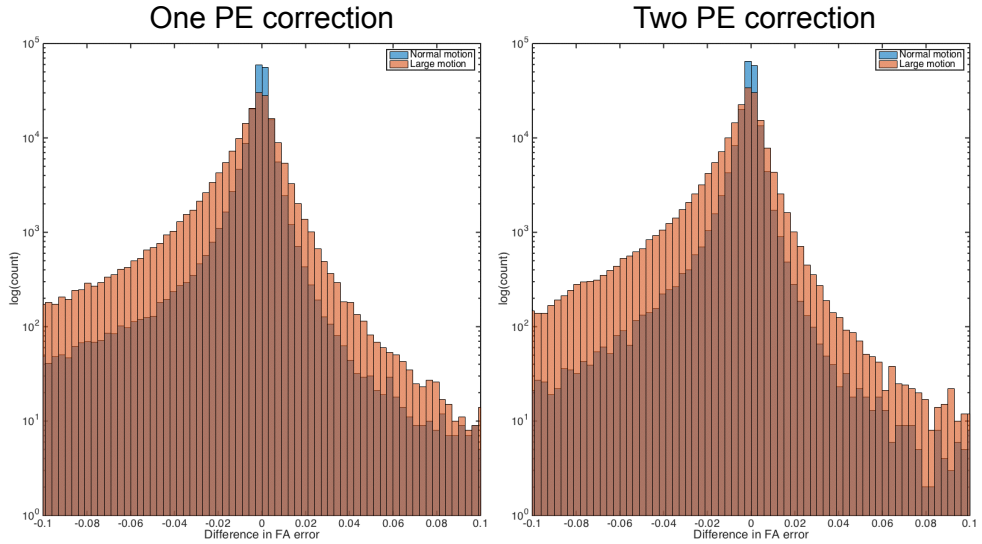


Figure 6.9: Error distribution. Histogram of the difference in absolute FA errors over the full brain, for datasets corrected for motion and static susceptibility: $|\Delta\mathrm{FA}_{\mathrm{static}}| - |\Delta\mathrm{FA}_{\mathrm{dynamic}}|$, so the heavy tail for negative values indicate higher errors for the dynamic case.

MPB techniques performed better, though the FMB method demonstrated sensitivity to partial volume with air, particularly as the noise level increased. This occurred despite following standard practice of only estimating the field for voxels within an eroded mask, to exclude voxels with significant partial volume, and smoothly extrapolating the estimated field outside the domain of the mask. It has been reported that fitting a set of 3D discrete cosine transformations can lead to improvement [125], but to our knowledge this is not implemented in any available software packages. In addition to providing better correction than the FMB technique, the MPB method has the additional advantage [147] of being able to correct for concomitant fields which cause translations of slices far from the isocentre along the PE direction [166]. Concomitant fields cannot be measured by field-mapping scans and thus are not corrected by them. However, it should be noted that this sensitivity to concomitant fields is less of an issue for modern scanners with field strengths of 3 T or above, as the effects of the fields scale with the inverse of the main magnetic field strength, and in a well-tuned system they are corrected for on the scanner itself [34].

Whilst FMB and MPB techniques were both able to estimate the underlying displacement field well, neither of them are able to provide the correct signal distribution in areas that have been compressed by the susceptibility artefact. This is an inherent limitation, which our results showed can be addressed using MPB/F. The downside of MPB/F is that each DWI needs to be acquired twice, doubling the acquisition time. Whether this is an acceptable trade-off is likely to be influenced by several factors such as the available acquisition time, desired analysis methods (e.g. compartment modelling, tractography), and the brain regions to be studied. It is worth noting that the MPB/F method increases SNR in the corrected images, so does not provide any time penalty if repeats are already being acquired to boost SNR through averaging. It should be noted that we used Jacobian modulation to correct for compression/stretching in each volume before combination through averaging. Whilst this still provides enhanced correction in regions of large distortion when compared to the MPB method, a least-squares reconstruction of the two images should be used for the best recovery of information lost in regions of compression [96].

Surrogate metrics are often used to evaluate correction quality on real data. Such metrics can sometimes be misleading; for example visual inspection of registration results, or use of similarity metrics, can be misleading because differences between source and reference images can appear small despite the underlying displacement field having been poorly estimated. In order to extend our analysis to real data, we used the simulation framework to validate one of the most promising surrogate metrics for correction quality, the difference between AP and LR corrected datasets. Our simulations demonstrate the metric can be a useful surrogate for the statistic of real interest — error in the ground truth displacement field. However, the results showed that the metric does not necessarily give the same contrast between methods as errors in displacement field, and it is possible to conceive of a situation where the metric gives a different ordering of correction quality than displacement

field error (for example, a small displacement field error in a region with high intensity-contrast). We suggest that any surrogate metric must be interpreted cautiously, and ideally should be accompanied with supporting evidence from simulated data where a ground-truth is available. It should be noted that the real data used to demonstrate the surrogate metric is relatively unusual, having been acquired in a postnatal cohort. This means the contrast differs to the adult brain and a specialised acquisition protocol was used with overlapping slices and interleaved b-values. For this reason the findings made will not necessarily apply to adult data and should be interpreted with some caution.

Our findings are in agreement with the existing literature, which largely made use of real data. [136] used real data to compare RB and FMB methods. They found FMB methods outperformed RB methods in all regions affected by susceptibility artefacts. Interestingly, they found RB methods outperformed FMB in the superior few slices of the brain. This is likely because the data was acquired at 1.5T and thus suffered from measurable concomitant field-induced shifts in these slices, which FMB methods cannot correct for. [167] also used surrogate metrics on real data and found that FMB outperformed RB methods. [148] used real data to compare FMB and MPB/F techniques, finding MPB/F methods to be superior. [168] used simulations to compare RB, FMB and MPB/F (but not MPB) techniques. Whilst they used simpler simulations and less direct metrics to assess correction efficiency, they observed the same ordering of the technique's effectiveness as we did. We are not aware of any published comparisons of FMB and MPB methods, likely the most relevant comparison for most researchers as they offer the potential for good correction of geometric distortions with only slightly increased acquisition times.

We also investigated the impact of failing to correct for the dynamic portion of the susceptibility field on the analysis of diffusion data. Our results highlight that even if a subject moves a little the dynamic field increases the errors in estimated diffusion metrics, but that the problem becomes much worse for subjects that move 'a lot'. This result is important in the context of pop-

ulation studies, where a group that moves a lot (due to e.g. age, disease) is compared to a healthy control group that is likely to move less. Interestingly, we found correcting data using both AP and PA acquisitions did not further increase the errors introduced by dynamic susceptibility, and in fact marginally reduced them. We hypothesised an increase in error would occur because each corrected volume was created by combining information from two differentially distorted volumes, thus creating corrected data that was even more 'wrong' than data obtained simply resampling all the PA volumes. A potential explanation for the result is that the motion trace we used is characterised by small rotations for most volumes and occasional spikes of larger motion for some. Thus, for each PA (or AP) image with a large dynamic susceptibility component caused by large motion, its corresponding AP (or PA) volume is likely to be much less distorted, and so producing a corrected volume from the two leads to an averaging effect that serves to reduce the total amount of distortion in the corrected data.

To our knowledge, none of the commonly used post-processing schemes correct for the dynamic susceptibility artefact in DW-MR. There have been attempts to deal with the problem using real-time auto-shimming [154, 155] but these require non-standard pulse sequences and are only able to correct for the linear terms of the dynamic field. Acquisition of a field-map for each volume is also possible, at the expense of increased scan-time [169]. Registration of every DW-MR volume to an undistorted structural target has been suggested in the past [137], but is inappropriate given the differences in contrast between diffusion-weighted and structural volumes, especially for modern acquisitions which tend to make use of higher b-values with even more different contrast. One potential method for mitigating the artefact, available to users of the MPB/F method, would be to interleave the protocol such that the acquisition of each blip-up volume is immediately followed by the acquisition of its blip-down counterpart. The dynamic field could then be estimated for every pair of DWIs, on the assumption that there was negligible motion between the

pair of DWIs. This would have the added benefit of reducing the issue of combining volumes with opposite PE direction which have different effective diffusion sensitisation, which happens when movement occurs between them. Our simulation assumed inter-volume movement, but in reality movement will occur during the acquisition of volumes. Intra-volume movement correction schemes do exist [170, 171], and would ideally be integrated into any technique that corrected for the dynamic susceptibility artefact.

There are some limitations to the work. The simulations used an off-resonance field estimated from a field-map as the input, which could introduce some circularity that favours the FMB method. It seems that this wasn't an issue in this study, as FMB techniques were outperformed by MPB methods. We also tested one available software implementation of each method; the large number available meant it was not practical to evaluate more. For FMB and MPB+MPB/F methods, we used the implementations most commonly used by the research community. We note that there are MPB methods [151] and MPB/F methods [150] that report better results than the method used in this work, TOPUP, and a comparison of these promising techniques may be the subject of future work. There are a number of available RB implementations, with no single one of them being clearly more popular than the others, each with a large number of parameters and settings that can be optimised, but we do not believe the choice of specific implementation will change our conclusions. This is because our experience with other RB methods (not shown), along with other published work comparing different RB methods to the one we tested [136, 168], support the conclusion that they are consistently outperformed by FMB and MPB+MPB/F methods.

6.6 Conclusions

I used the simulation framework to assess the three most commonly used post-processing methods for correcting the susceptibility artefact, producing practical advice on which work best. I also investigated the interaction between sus-

ceptibility and movement, and suggest that in the case where reversed PE data is acquired its impact can be reduced by a simple adjustment of the acquisition: the rearrangement of the scans so that each blip-up scan is immediately followed by its corresponding blip-down. A post-processing method that corrects for the susceptibility-movement interaction has very recently been published [172] — the simulation framework has been used to evaluate this method, and ideally the data from this can be used as a benchmark for any future methods that are developed.

Chapter 7

Application III: a tool for automated quality control

7.1 Overview

In this chapter we demonstrate the potential for using the simulation framework to develop new tools. We simulate ready-labelled data and use it to train a classifier to detect severe movement artefacts, and compare its performance to a classifier trained on manually-labelled real data.

7.1.1 Research dissemination

The work detailed here has been submitted to ISMRM.

- A supervised learning approach for diffusion MRI quality control with minimal training data. MS Graham, I Drobnjak, H Zhang. *Submitted to ISMRM 2018*

This chapter includes the extension of the framework to simulate signal dropout artefacts, and the first demonstration of the framework's ability to simulate movement during the acquisition. Datasets simulated with both these artefacts have been used to validate novel techniques for correcting signal dropout and intra-volume movement artefacts, though these works are not discussed in this thesis.

- Incorporating outlier detection and replacement into a non-parametric

framework for movement and distortion correction of diffusion MR images. JLR Andersson, MS Graham, E Zsoldos, SN Sotiropoulos. In *NeuroImage* 141, 556-572, 2016.

- Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. JLR Andersson, MS Graham, I Drobnjak, H Zhang, N Filippini, M Bastiani. In *NeuroImage* 152, 450-466, 2017.

7.2 Introduction

Quality control (QC) involves ensuring a dataset meets a certain set of standards before the dataset is given the clearance for inclusion in subsequent analyses. In MRI there are a large number of potential artefacts that need to be identified, to enable problematic images to either be excluded or accounted for in further processing and analysis. The gold standard for identification of these is visual inspection of the data.

There are a number of challenges with manual QC. For a typical study, which may involve hundreds of subjects, the process can be extremely time-consuming. This is especially true in diffusion MRI (DW-MR) where many volumes might be acquired for every subject, and there are numerous artefacts that each volume must be screened for. The current trend towards acquiring increasingly large datasets means the time required for human QC is becoming prohibitive. The HCP [173] acquired data for 1200 subjects with almost 300 DW-MR volumes per subject and the UK Biobank will eventually acquire imaging data for 100,000 subjects with over 100 volumes per subject [132]. Manual QC is also subjective. Each rater has their own sensitivity and specificity which cannot be easily altered, meaning the data is either QCed by a single rater, leading to a single standard but requiring large amounts of time, or many raters look at the data, which requires less time but means variable standards are applied across the dataset. These challenges have led to an increased interest in automated methods for QC.

Automated methods for QC fall into two classes. The first kind extracts tailored features from the datasets and applies hand-tuned cutoffs to determine whether each volume contains artefacts [174, 175]. The second kind are supervised learning approaches. These involve extracting features from the datasets and then using a training set, obtained from manual QC of a proportion of the data, to learn the mapping between these features and the classification of each volume as passing or failing QC. Recently these approaches have used support vector machines (SVMs), random forest classifiers [176] and ensembles of classifiers [177]. Whilst promising, both types of approach report performance significantly below that of a human rater.

Recently, convolution neural networks (CNNs) have been demonstrated to provide near-human levels of accuracy for identifying motion artefacts in structural [178] and DW-MR data [179]. Unlike other supervised approaches, CNNs learn features from the data during training, rather than requiring them to be hand-crafted and supplied as input. CNNs tend to have many parameters requiring optimisation — often in the millions — meaning they typically require large, labelled datasets for training.

Obtaining training datasets in medical imaging can be challenging. The acquisition of the data is time-consuming and expensive, and once acquired the data can be subject to ethical considerations or anonymisation requirements that prevent that data being shared freely. Labelling of such data is also challenging. For the case of QC, labelling requires a human rater to manually inspect each image volume and flag any that contain artefacts. The process is subjective, and the accuracy of the trained classifier will depend on the quality of the labelled dataset, so often a number of raters are used and their classifications combined in order to get a more reliable 'ground-truth' for the dataset. Furthermore, a tool trained on a specific set of training data may not generalise well to datasets acquired with different protocols or hardware, meaning new training datasets may need to be labelled for each new dataset.

One potential way to address these issues is to use simulated data. Sim-

ulation could circumvent the need for human labelling by producing realistic datasets, along with ground-truth labels, for training machine learning tools on. In the case of QC, a simulator that was capable of producing datasets containing artefacts, such as motion, could be used to produce a training set. Little research has been done to investigate the feasibility of a simulation-based approach to training supervised learning tools.

In this work, we aim to investigate the feasibility of a supervised-learning approach to QC that uses simulated data. As a first step, we focus on the problem of detecting intra-volume movement in DW-MR data. Intra-volume movement refers to both signs of head movement and the signal dropout that this gives rise to — see Fig 7.1. We focus on this artefact because they are often not tackled in QC as checking every volume in a dataset can be extremely time-consuming. Volumes containing this artefact typically need to be identified in QC so that they can either be removed, or information about them can be used as confounds in later statistical analysis [180]. Whilst post-processing techniques have been proposed to correct for intra-volume movement, [171, 181, 182] it has been reported these fail for severe cases and an initial QC needs to be performed to remove especially bad volumes before processing [179]. We compare the performance of classifiers trained on real and simulated data.

7.3 Methods

This section details the data, both real and simulated, used in this work and describes the classifier that was trained on these data.

7.3.1 Data

7.3.1.1 Real

Ten subjects were taken from the developing Human Connectome Project [133] (dHCP), which contains MRI data acquired in neonates. These were chosen because neonatal scans tend to contain large amounts of movement. The data was acquired on a 3T Philips Achieva, consisting of a spherically optimized set

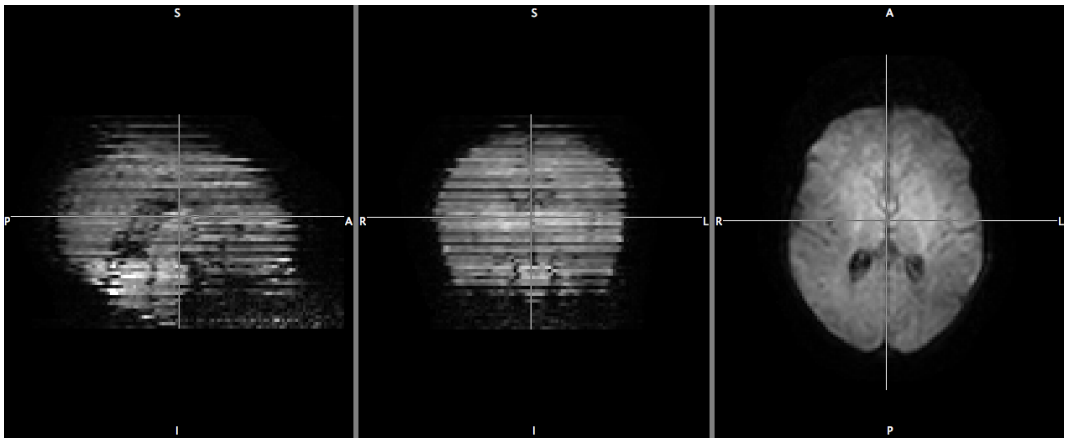


Figure 7.1: Example of the intra-volume movement artefact. For interleaved acquisitions, which are common in DW-MR, the movement causes jagged edges perpendicular to the EPI plane, here the coronal and sagittal views. This misplacement of the signal is different to signal dropout, also caused by intra-volume movement, which leads to loss of the signal — an example of dropout can be seen in one of the most inferior slices of the coronal view in this subject.

of directions on 4 shells (b0: 20, b400: 64, b1000: 88, b2600: 128) split into four subsets, each with a different phase-encoding (PE) direction. It was acquired using a multiband acceleration dactor of 4, SENSE factor 1.2 and partial fourier 0.86, TR/TE 3800/90 ms. The data has resolution 1.17x1.17x1.5 mm, matrix size 128×128, with 64 slices per volume.

For this study, the b=2600 s mm^{-2} volumes were removed as they contained very little signal, which caused even manual QC to be challenging. This left 172 volumes per subject. Manual QC was performed by visual inspection, with one rater assigning a label of either acceptable or unacceptable to each volume. The rater classified the whole dataset twice, on two separate occasions, to provide an estimate of intra-rater agreement.

7.3.1.2 Simulated

Simulated data was designed to be visually similar to the dHCP data. Data was simulated using the same *b*-values and directions as the dHCP. Voxel size and FOV were kept the same as previous simulations to minimise computation time: 2.5 mm isotropic and $72 \times 86 \times 55$ voxels. Seven subjects were simulated using input objects derived from different subjects from the HCP. Neonatal

Tissue	T_1/ms	T_2/ms	ρ
Grey matter	2200	200	0.8
White matter	2850	250	0.8
CSF	3700	280	0.8

Table 7.1: Tissue parameters used for the simulations in this chapter. Proton density ρ is in arbitrary units.

DW-MR images have different contrast to adult data, with much reduced contrast between GM, WM and CSF. MR parameter values were modified to increase the visual similarity between simulated and real data — T_1 parameters were taken from [183], T_2 parameters from [184] and then adjusted further to maximise visual similarity with the dHCP datasets — the final set of parameters used are shown in Table 7.1.

Known motion was injected into both datasets during simulation. The motion traces were designed to produce data with clear signs of intra-volume movement, similar to those seen in the real data (Fig 7.1), in order to produce a suitable training set for the classifier. This was achieved by synthesising traces with large, sudden motion spikes, modelling a subject suddenly moving their head. The traces describe the object's translations along and rotations about each of the three axes, with movement occurring between the acquisition of each slice. Algorithm 1 describes more precisely how the traces were generated, and Figure 7.2 shows an example trace. Interleaved slice-ordering was simulated, without multiband, so that these intra-volume movement spikes produced the characteristic zig-zag edge pattern as seen in Figure 7.1. Signal dropout was also simulated. In the dHCP data signal dropout is often, but not always, present in volumes that show other signs of severe intra-volume movement. To reflect this, dropout was added to a volume containing significant motion with a probability of 70%. Dropout is applied by directly reducing the signal of slices in k-space, rather than simulating the effect of the interaction between movement and the diffusion gradients. This simpler approach still produces realistic-looking dropout artefacts. The full details of how dropout

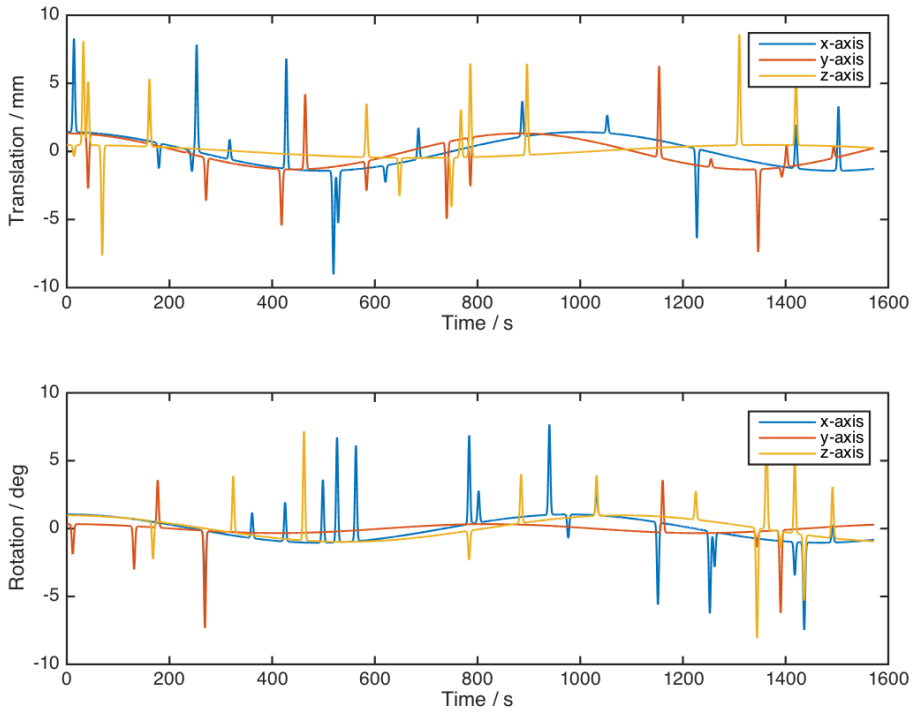


Figure 7.2: Example simulated motion trace, shown over the time required to acquire all 172 volumes in a dataset.

was added is described in Algorithm 2.

Labels were assigned to each volume using the following scheme. The amount of intra-volume movement for each volume was calculated for each of the three rotations and three translations. If all of the translations were less than 1 mm and rotations less than 1°, the volume was given a label 0, for acceptable. If any of the translations were greater than 1 mm and less than 2.5 mm, or rotations greater than 1° and less than 2.5° the volume was assigned 1, for moderate. If any translations were greater than 2.5 mm or rotations greater than 2.5°, the volume was assigned a 2. It was observed that inclusion of volumes with moderate movement (label 1) with the volumes with severe movement (label 2) made it much more difficult to train the classifier. These volumes included very subtle signs of motion that were challenging to QC when visually inspected. By contrast, we found that most volumes if the real data were much more straightforward to classify; it was usually very

Algorithm 1 Synthesising movement traces for the simulated data

for Each volume to be simulated **do**
  Set volume's movement trace for T_x,T_y,T_z,R_x,R_y,R_z to 0
  Draw number from random-number generator (RNG) between 0 and 1
  if number ≤ 0.4 **then**
    for Each movement trace T_x,T_y,T_z **do**
      Draw number from RNG
      if number $\leq 1/6$ **then**
        Generate Gaussian motion spike with height randomly selected between 0 and 10 mm, standard deviation 0.2*repetition time (T_R)
        Add Gaussian motion spike to randomly selected location in volume's trace
      end if
    end for
    for Each movement trace R_x,R_y,R_z **do**
      Draw number from RNG
      if number $\leq 1/6$ **then**
        Generate Gaussian motion spike with height randomly selected between 0 and 10°, standard deviation 0.2*T_R
        Add Gaussian to randomly selected location in volume's trace
      end if
    end for
  end if
end for

Algorithm 2 Adding signal dropout to simulated data

for Each volume to be simulated **do**
  if Volume's motion trace has any translations ≥ 2.5 mm or rotations $\geq 2.5°$ **then**
    Draw number from RNG
    if Number ≤ 0.70 **then**
      for Every slice in volume **do**
        Draw number from RNG
        if Number ≤ 0.85 **then**
          Draw number from RNG
          Multiply signal in slice by number drawn
        end if
      end for
    end if
  end if
end for

obvious whether or not they contained movement artefacts. For this reason it was decided to remove the 108 volumes with moderate motion from the simulated dataset. This left a total of 1096 volumes, 732 without movement and 364 with.

7.3.2 Classifier

We based our classifier on a type of neural network called a convolutional neural network. These networks have provided state-of-the-art performance in computer vision tasks in recent years [185], making them sensible candidates for the task of identifying motion-artefacts in scans. A drawback of such networks is that they contain millions of parameters, and so they typically require large amounts of training data and large amounts of computational power to successfully train. To circumvent this we adopted a transfer learning approach [186], which consists of taking a classifier trained to perform on a certain task and re-training a small number of parameters using a small amount of data to perform well on another, often similar, task.

Our transfer learning approach here is similar to that described in [179], where they successfully trained a classifier on real data to detect motion artefacts. We used the pre-trained InceptionV3 network as the base network, which has achieved state-of-the-art performance in the classification of natural images (i.e. photos taken on standard cameras of everyday objects such as dogs, boats, cars) [187]. To finetune InceptionV3, the top layer of the network was removed, and replaced with a fully connected layer with 16 neurons, followed by a prediction layer with 2 neurons for the two classes in our problem (motion-corrupted or normal). All parameters were fixed apart from those in the newly added layers, vastly reducing the number of parameters required for training.

The classifier was trained by passing it sagittal slices through the brain along with ground-truth labels. We trained two classifiers: one on real data, and one on simulated data, using seven subjects for training as in [179]. Five subjects were used for training, and two for validation. We chose to use three sagittal slices from each volume at training time, though in principle using

more might provide better results. One of the slices was taken from the central plane of the volume and the other two from either side of this central plane, towards the edges of the brain. For the real data, these side slices were 14 slices away from the middle slice on either side, for the simulated data, these were 16 slices away, as the simulated brains were slightly larger.

For both classifiers, images were zero-padded along the shorter dimension to make them square, resized to 299 by 299 pixels, and replicated three times for the three channels of the network (a fixed requirement of the InceptionV3 network). Each image was scaled so that its intensity lay between -1 and 1. Each classifier was trained for 30 epochs using the Adam optimizer with a learning rate of 0.001 [188] and a cross entropy loss function. The classifiers were implemented in Keras [189]. Training took less than 20 minutes on a Titan X Pascal GPU.

Testing was performed on the three reserved dHCP subjects. To assign a label to each volume, 30 saggital slices from the volume were extracted and classified; if the mean of these 30 scores was greater than a certain threshold, the volume was labelled as containing motion. For the real-trained classifier we used the natural threshold of 0.5. Due to the shift from the simulated to real domain, the threshold for the simulation-trained classifier had to be calibrated using a single subject from the real training set. We experimented with two methods for determining the optimal threshold from this subject. In the first, the threshold that maximized the F1-score between the true and predicted labels for this subject was chosen for use at test-time. The F1 score is defined as:

$$F = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{7.1}$$

with precision and recall defined as:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7.2}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7.3}$$

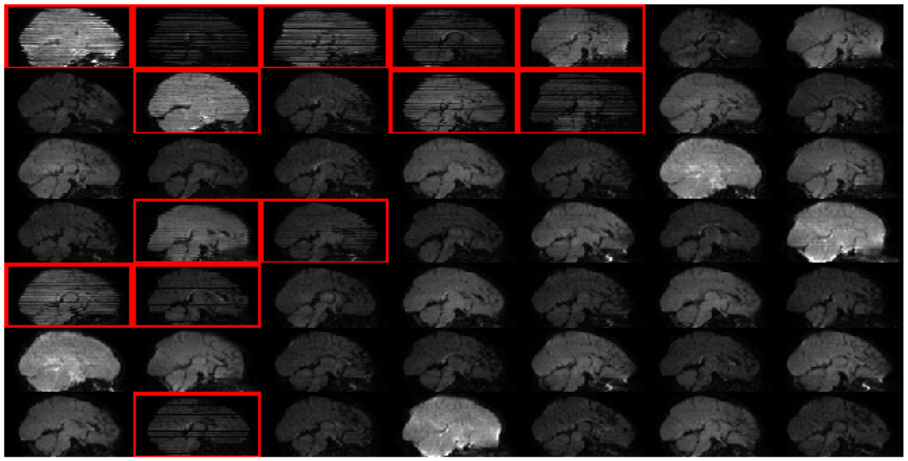
where TP are true positives, TN true negatives, FP false positives and FN false negatives. The rationale for this method was finding a threshold that balanced precision and recall. In the second, the greatest threshold that gave >95% recall for motion-corrupted volumes in this subject was chosen. The rationale for this method was that it may be preferable to ensure the majority of corrupted volumes are found, even if this means rejecting some false positives.

7.4 Results

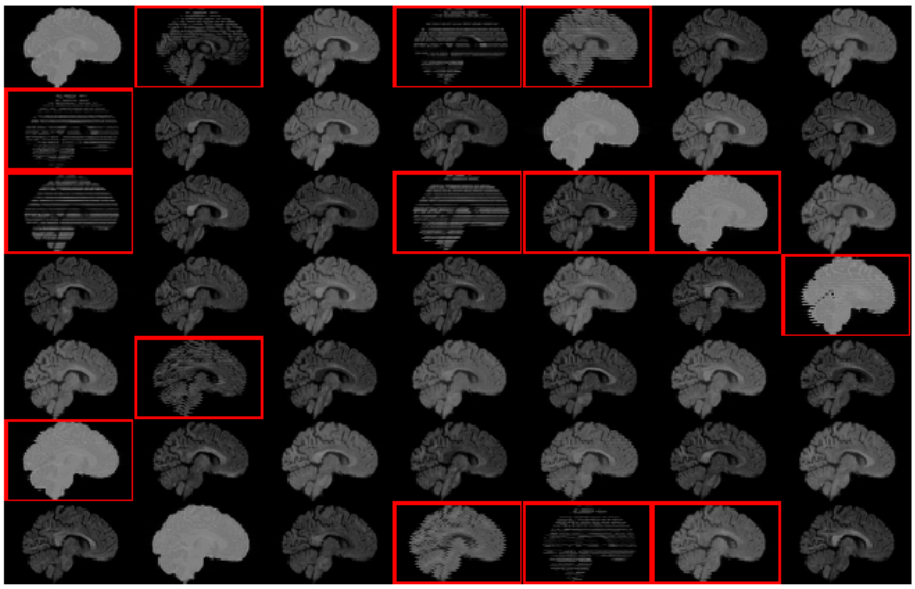
Simulated and real data is shown in Figure 7.3. Note that the saturated appearance of the b=0 volumes is expected for neonatal MR due to the longer tissue relaxation times. Both simulation-trained and real-trained classifiers fit their validation sets well — the simulation-trained achieved 95% accuracy on the simulated validation set, and the real-trained achieved 93% accuracy on the real validation set, with accuracy defined as the fraction of correct classifications (both true positives and true negatives).

We determined the optimal threshold for the simulation-trained classifier using each of the seven real subjects in the training+validation set, in order to get a sense for the extent to which the threshold depends on choice of subject. Thresholds determined from the F1-score criterion were more tightly clustered (from 0.86-0.97) then those determined from the sensitivity criterion (0.71-0.96), indicating the F1 criterion is a more reliable way of determining a threshold. We decided to use the F1 criterion thresholds for the test dataset.

Both classifiers were tested on the three held-back dHCP subjects. Figure 7.4 shows the precision-recall curve for the two classifiers. Whilst it was decided to use the F1 criterion for the results, thresholds for the sensitivity criterion are also plotted on this Figure to demonstrate the greater variance in precision/recall scores this introduces. The real-trained classifier achieved precision and recall of 97% and 98% for classification of corrupted volumes, results comparable to the state-of-the-art results reported in [179]. Intra-rater agreement on the test set was 99%, showing this classifier approaches human level



(a) Real data.



(b) Simulated data.

Figure 7.3: Real and simulated data. Red bounding boxes indicate the volume was labelled as containing intra-volume movement.

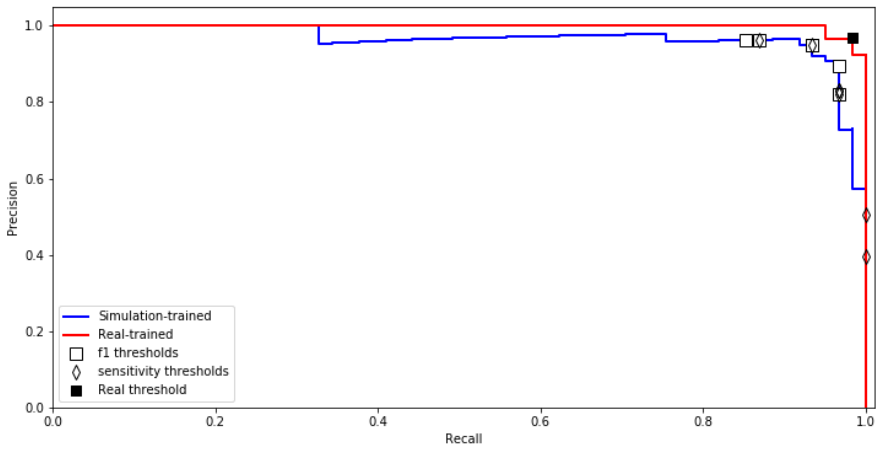


Figure 7.4: Precision-recall curve for both classifiers in the test set, consisting of 516 volumes. The threshold for the real-trained classifier of 0.5 is plotted on the curve, as are the seven thresholds determined for both the F1- and sensitivity-based criterions for the simulation-trained classifier.

performance. The simulation-trained classifier achieved precision and recall of 95% and 93% for the most common F1-determined threshold (0.94, occurred in 3/7 subjects). If the lower range threshold was used (0.87) precision and recall was 83% and 97%, and for the upper threshold (0.97) these values were 96% and 85%. Figure 7.5 shows results for both classifiers on some of the test data. Figure 7.6 shows the mean classifier score for each volume in the test set, along with the classification for each volume.

7.5 Discussion

In this work we compared the performance of a QC tool trained on simulated data to that of a tool trained on real data. The real-trained classifier achieved near-human performance, confirming the findings in [179]. The simulation-trained classifier demonstrated performance approaching that of the real-trained classifier. It was able to detect the majority of the motion corrupted volumes in the test set, though it showed slightly reduced precision and recall compared to the real-trained classifier.

The classifier presented here is a modified version of the one presented by

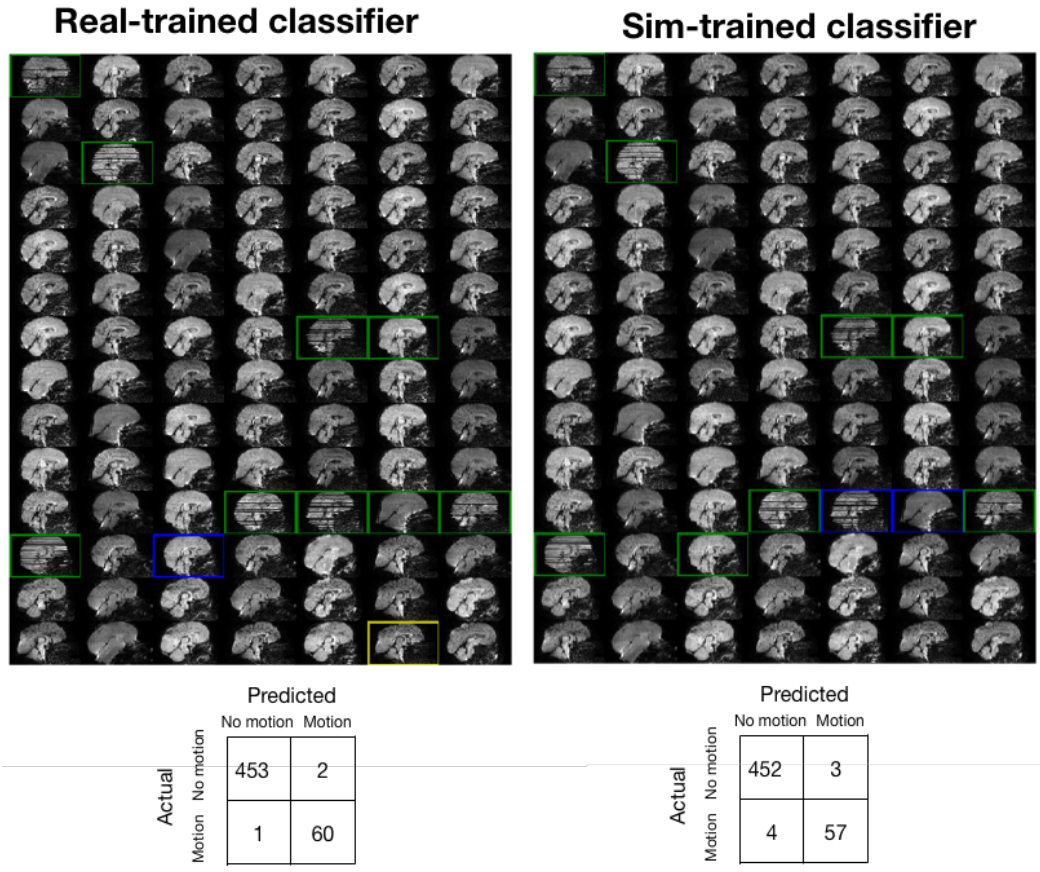


Figure 7.5: Sample classifier results for images in the test-set. Green border indicates a correct classification as containing motion, blue borders indicate false-negatives and yellow borders are false-positives. Confusion matrices for classification on all 516 volumes in the test set are shown below. Threshold of 0.94 used for the sim-trained results.

Kelly et al [179]. It offers comparable performance when trained on real data whilst offering a number of improvements. Firstly, ours involves training a single neural network, compared to the 11 trained in [179]. Our classifier only requires magnitude data, whilst the previous classifier uses both magnitude and phase data, and we don't need to distinguish between b-values for training. Our final decision is made by a simple thresholding of the classifier outputs, whilst [179] requires the additional training of a random forest classifier on the CNN outputs.

We only trained our classifier to detect movement artefacts, but there

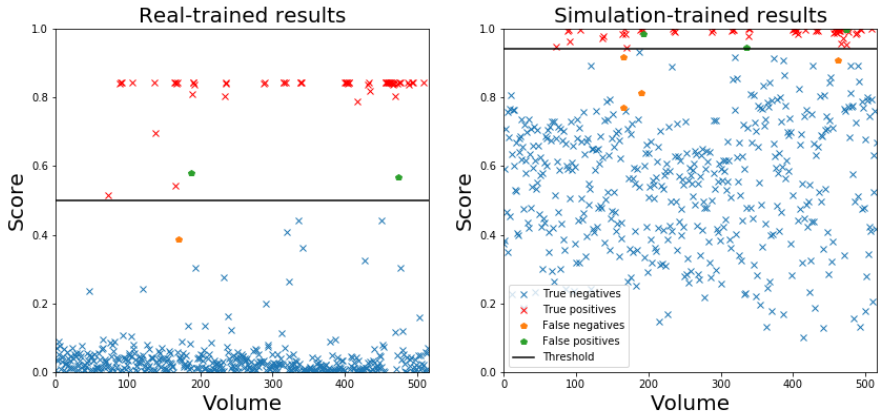


Figure 7.6: Classifier scores for each volume in the test-set. The score is produced by averaging the classifier outputs for the 30 slices classified in each volume. Threshold of 0.94 used for the simulation-trained plot.

are many more artefacts that would ideally be identified by the QC process, and future work will look into extending the classifier. One advantage of training using simulated data is that the training set can be designed to include numerous examples of artefacts that might be very rare in practice (such as RF spikes). Training to identify these rarer artefacts on real data may require labelling a very large dataset in order to find sufficient training examples.

The real data used in this study was relatively unusual, being acquired in a challenging postnatal cohort. Not only does the data has different contrast and appearance to adult data, but a specialised acquisition was adopted involving overlapping slices and interleaved b-values and phase-encode direction. Furthermore there were much higher levels of movement in the dataset than are usual, which is an advantage for the purposes of this work but could mean classifiers would show higher false-positive rates on data with less movement. In future work it would be interesting to explore the performance of classifiers on more standard, adult datasets.

One potential source of error was the automatic threshold chosen to produce ground-truth labels for the simulated data. A volume with more than 2.5 mm translation or rotations greater than 2.5° was labelled as containing intra-volume movement. If volumes with this level of movement looked sig-

nificantly different to volumes in the real data that were manually labelled as containing movement the simulation-trained classifier's performance would be affected. We investigated this by performing manual QC on two subjects from the simulated dataset. Manual and automatic QC agreed for 95% of cases. When they disagreed it was because the automatic threshold picked up on slightly more subtle cases of movement. This could have led to a slightly more sensitive classifier than the real-trained one, but this does not seem to have been the case.

There is room for more investigation. It would be interesting to understand how similar the real and simulated data need to be in order to obtain good performance. Further work could determine how performance depends on the amount of movement simulated, signal dropouts, image contrast and choice of b-values and directions. This ties in to how well the trained classifiers will generalise to new, unseen datasets — for example a dataset acquired on adult subjects, or with a different protocol. We could also look at how much training data is required for good performance; in this work we matched the amount of simulated and real data used for training, but we could test whether performance can be improved even further by using more simulated data for training.

There are some drawbacks to the approach described in this chapter. These centre around the difficulties inherent in training on one domain (simulated data) and then classifying in a different domain (real data). This can be seen in Figure 7.6. It shows that the simulation-trained classifier was good at spotting artefacts, and assigned high scores to volumes containing movement, but was much less sure for volumes that did not contain movement, producing scores with a very large spread. This contrasts with the real-trained classifier, which was able to assign high scores to volumes with movement and low scores to volume without. This meant the simulation-trained classifier had a smaller margin for error, which caused the occasional large mistake: in Figure 7.5, it can be seen the simulation-trained classifier predicted a false-negative on a

volume that quite clearly contained movement artefacts — the errors in the real-trained classifier tend to be more straightforward 'borderline cases' that a human might find difficult to classify. The shift in scores caused by the transfer between domains also meant that the simulation-trained classifier still requires some labelled, real-data for calibration. Whilst this is still a big reduction in the amount of labelled data required when compared to the amount needed to train the classifier, it would be ideal if none was required. Furthermore, the choice of subject used for the calibration introduces variability into the final performance of the classifier. These limitation can be addressed by including recent research on domain adaptation in machine learning, getting a classifier trained to perform well in domain A (e.g. simulated data) to perform well on domain B (e.g. real data) without requiring any labelled data from domain B. In [190], they encourage the classifier to learn features which are domain-invariant by pitting it adversarially against a discriminator which attempts to predict the domain the classifier is working in by examining the classifier's activations. In [191], a neural network is used to adapt simulated data to appear more like real data; training on this adapted data gives performance equivalent to training on real data. Future work will incorporate similar techniques to bridge the simulation-trained and real-trained gap seen in our results.

7.6 Conclusion

I demonstrated that a QC classifier trained on simulated data can approach the performance of one trained on real data. The simulation classifier still requires a small amount of labelled real data for calibration, but this limitation will be addressed in future work. The use of simulated data shows some promise for machine-learning in medical image analysis where labelled training data can be expensive or difficult to acquire.

Chapter 8

Conclusions

8.1 Summary

The aim of this thesis was to develop a DW-MRI simulator to assist the testing and development of methods for the acquisition and processing of DW-MR datasets.

In chapter 3 I developed the framework by combining the POSSUM MR simulator with a realistic representation of diffusion weighting, taken from real data. I demonstrated that this approach enables the simulation of much more realistic data than was possible using existing DW-MR simulators. I also described how to produce ground-truth displacement fields for data simulated with artefacts that cause geometric distortions, and suggested it as a new, more direct way of assessing post-processing algorithms that seek to predict this ground-truth field.

In chapter 4 I applied this simulator to assess existing techniques for correcting eddy-current and motion artefacts. The results showed that the most commonly used technique, registration of each DWI to a b=0 image, can give poor results, especially for higher b-value data which is being acquired with increasing frequency. I showed that poor correction is difficult to detect using either visual inspection of the data or assessment of model fits to corrected data, two techniques that are commonly employed in the literature to assess correction quality. By using the assessment of displacement fields I was able to

demonstrate the effect clearly and quantitatively. The work also demonstrated that a new correction tool, `eddy`, provided very good correction, and enabled me to investigate the acquisition protocol required to obtain good correction using the tool.

In chapter 5 I addressed one of the main limitations of the simulator: its inability to simulate spin-echo pulse sequences. This prevents the realistic simulation of susceptibility artefacts, which in practice are present in nearly all acquired DW-MR datasets. This was used in chapter 6 to assess existing techniques for correcting the susceptibility artefact. I also investigated the interaction of the susceptibility artefact with head movement, an effect that would be difficult to study without a simulator. I highlight that this can affect analysis of DW-MR data and suggest a simple modification to the acquisition protocol that could help mitigate this effect - the adjacent acquisition of blip-up and blip-down pairs for each DWI.

In chapter 7 I demonstrated that the simulator can also be used to develop new tools. It was used to train a classifier to detect movement artefacts in DW-MR data. The classifier's performance was close to that of one trained on real data, but required significantly less manually labelled real data. An approach such as this could have applications in other areas of medical image analysis where machine learning techniques can offer good performance but labelled training data is expensive or difficult to obtain.

8.2 Future directions

8.2.1 Simulator improvements

Ideally, simulated data would be seen as an integral part of the validation of a processing pipeline, or the comparison of a newly proposed tool to the existing state-of-the-art. Whilst the simulator has been made available online, the computation time required to generate a dataset presents a current barrier to use. A single volume takes $\sim$24 hours to simulate on a single CPU and so in practice a computer cluster is necessary to produce a full dataset.

GPU acceleration of the code could provide orders of magnitude speed-up that would allow complex datasets to be simulated on a single computer, and would potentially assist the adoption of the simulator.

There are several features that could be implemented in the simulator. Simulating a denser grid of spins — assisted by GPU acceleration — would mean current model-based approaches to phenomena such as T_2^* dephasing and the spin-echo could be replaced with more faithful modelling of the physics. This would enable the simulation of pulse sequences with multiple spin-echos, as well as the reproduction of subtler effects such as imperfect spin-echo rephasing. Modelling of multiple RF coils would enable parallel imaging [192] and more recent advances such as simultaneous multislice techniques [193] to be simulated. This would make the simulator more useful for helping to understand the effect of more sophisticated image protocols on image quality, and help better understand how this is affected by interaction with other artefacts, such as movement. Finally, extending the simulator to numerically solve the Bloch equations during the RF pulse would enable the testing of non-standard RF pulses, which would combine usefully with multislice imaging techniques which tend to use more complicated pulses.

8.2.2 Simulator object improvements

As discussed in Chapter 3.3, there are some potential limitations to the use of the HCP dataset to produce simulator input objects, such as the possibility of propagating artefacts present in the data into the simulations themselves. This could be addressed by acquiring a bespoke dataset with an acquisition carefully designed to minimise artefacts. This would also enable the acquisition of a richer diffusion acquisition protocol, with higher b-values than present in the HCP data, enabling the simulation of even higher b-value data than presently possible. The limits to the b-values that can be simulated could also be addressed by replacing the current spherical-harmonic approach with a model that allows extrapolation to higher b-values, such as MAP-MRI [115].

An interesting avenue for future work might be to integrate a description

of underlying microstructure into the input object. Diffusion contrast could be generated using a model-based approach, or even a Monte-Carlo approach that explicitly models the diffusion of the spins. The advantage of such methods is that they provide a ground-truth for the underlying microstructure, which make them better suited for assessing techniques that aim to reconstruct this underlying microstructure, such as tractography. The Monte-Carlo approach would also enable the simulation of artefacts caused by flow within blood vessels.

It would be useful to have a simulator that was able to model non-rigid movement, such as that caused by CSF pulsation. This could enable the simulation of effects such as signal dropout localised to regions within a slice. This could be achieved by moving from a voxel-based to a mesh-based description of the input object, with tissue parameters specified at each vertex in the mesh. The movement of each vertex in the mesh could be specified independently.

8.2.3 Further applications

There are numerous ways the framework can be used in future work. We have seen in Chapter 7 that the simulator may be used to train a tool to detect artefacts, but we can also imagine the extension to training tools to correct these artefacts. In [194], a deep neural network was trained to predict the deformation field that registered two MR images. The training data for this was pairs of images registered by conventional means and a description of the displacement field between them. The quality of the training data is dependent on the quality of the initial registration; poor registration would lead to a poorly trained network. The simulation framework could be used to provide gold-standard training data, consisting of pairs of misaligned images and the ground-truth displacement fields between them. If the simulator could be adapted to simulate non-rigid movement then it could be used to accurately train non-rigid registration tools.

Another challenge in MRI is the harmonisation of data acquired on different scanners and with different protocols. The ability to combine data from

different scanners would much increase the statistical power and sensitivity of studies. The simulation framework could be used to understand how differences between scanners and protocols affect the data , and be used to develop either physics-based or machine-learning based tools to harmonise such data.

Appendix A

Publications

Journal papers

- JLR Andersson, **MS Graham**, I Drobnjak, H Zhang, J Campbell. Susceptibility-induced distortion that varies due to motion: Correction in diffusion MR without acquiring additional data. *NeuroImage* In press.

- **MS Graham**, I Drobnjak, M Jenkinson, H Zhang. Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI. *PLOS ONE* 12 (10), e0185647, 2017.

- JLR Andersson, **MS Graham**, I Drobnjak, H Zhang, N Filippini, M Bastiani. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *NeuroImage* 152, 450-466, 2017.

- JLR Andersson, **MS Graham**, E Zsoldos, SN Sotiropoulos. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage* 141, 556-572, 2016.

- **MS Graham**, I Drobnjak, H Zhang. Realistic simulation of artefacts in diffusion MRI for validating post-processing correction techniques. *NeuroImage* 125, 1079-1094, 2016.

Conference papers

- **MS Graham**, I Drobnjak, H Zhang. A simulation framework for quantitative validation of artefact correction in diffusion MRI. *International Conference on Information Processing in Medical Imaging*, volume 9123 of *Lecture Notes in Computer Science*, pages 638-649, 2015.

Bibliography

[1] Felix Bloch. Nuclear induction. *Physical review*, 70(7-8):460, 1946.

[2] Edward M Purcell, HC Torrey, and Robert V Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, 69(1-2):37, 1946.

[3] PC Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 1973.

[4] EM Haacke, RW Brown, MR Thompson, and R Venkatesan. *Magnetic Resonance Imaging - Physical Principles and Sequence Design*. Wiley-Blackwell, 1999.

[5] Henry C Torrey. Bloch equations with diffusion terms. *Physical Review*, 104(3):563, 1956.

[6] J P Wansapura, S K Holland, R S Dunn, and W S Ball. NMR relaxation times in the human brain at 3.0 tesla. *Journal of magnetic resonance imaging : JMRI*, 9:531–538, 1999.

[7] Zhi-Pei Liang and Paul C Lauterbur. *Principles of magnetic resonance imaging: a signal processing perspective*. SPIE Optical Engineering Press, 2000.

[8] Albert Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 322(8):549–560, 1905.

[9] W. S. Price. Pulsed-Field Gradient Nuclear Magnetic Resonance as a Tool for Studying Translational Diffusion: Part 1. Basic Theory. *Concepts in Magnetic Resonance*, 9(5):299–336, aug 1997.

[10] Michael E Moseley, Yoram Cohen, J Kucharczyk, J Mintorovitch, HS Asgari, MF Wendland, J Tsuruda, and D Norman. Diffusion-weighted mr imaging of anisotropic water diffusion in cat central nervous system. *Radiology*, 176(2):439–445, 1990.

[11] Michael E Moseley, John Kucharczyk, Haleh S Asgari, and David Norman. Anisotropy in diffusion-weighted mri. *Magnetic Resonance in Medicine*, 19(2):321–326, 1991.

[12] E. O. Stejskal and J. E. Tanner. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*, 42(1):288, jul 1965.

[13] P Mansfield and B Chapman. Active magnetic screening of coils for static and time-dependent magnetic field generation in NMR imaging. *Journal of Physics E: Scientific Instruments*, 19:540–545, 2000.

[14] Peter J Basser, James Mattiello, and Denis LeBihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, 1994.

[15] Peter J Basser and Carlo Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. *Journal of magnetic resonance*, 111:209–219, 1996.

[16] Peter J Basser, Sinisa Pajevic, Carlo Pierpaoli, Jeffrey Duda, and Akram Aldroubi. In vivo fiber tractography using dt-mri data. *Magnetic resonance in medicine*, 44(4):625–632, 2000.

[17] Carson Ingo, Richard L Magin, Luis Colon-Perez, William Triplett, and Thomas H Mareci. On random walks and entropy in diffusion-weighted

magnetic resonance imaging studies of neural tissue. *Magnetic resonance in medicine*, 71(2):617–627, 2014.

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

[19] Jens H Jensen, Joseph A Helpern, Anita Ramani, Hanzhang Lu, and Kyle Kaczynski. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic resonance in medicine*, 53(6):1432–1440, 2005.

[20] J Tournier, Fernando Calamante, David G Gadian, Alan Connelly, et al. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23(3):1176–1185, 2004.

[21] J-Donald Tournier, Fernando Calamante, and Alan Connelly. Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage*, 35(4):1459–1472, 2007.

[22] Eleftheria Panagiotaki, Torben Schneider, Bernard Siow, Matt G Hall, Mark F Lythgoe, and Daniel C Alexander. Compartment models of the diffusion mr signal in brain white matter: a taxonomy and comparison. *Neuroimage*, 59(3):2241–2254, 2012.

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

[24] TEJ Behrens, MW Woolrich, M Jenkinson, H Johansen-Berg, RG Nunes, S Clare, PM Matthews, JM Brady, and SM Smith. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic resonance in medicine*, 50(5):1077–1088, 2003.

[25] Yaniv Assaf, Tamar Blumenfeld-Katzir, Yossi Yovel, and Peter J Basser. Axcaliber: a method for measuring axon diameter distribution from diffusion mri. *Magnetic resonance in medicine*, 59(6):1347–1354, 2008.

[26] Daniel C Alexander, Penny L Hubbard, Matt G Hall, Elizabeth A Moore, Maurice Ptito, Geoff JM Parker, and Tim B Dyrby. Orientationally invariant indices of axon diameter and density from diffusion mri. *Neuroimage*, 52(4):1374–1389, 2010.

[27] Stamatios N Sotiropoulos, Timothy EJ Behrens, and Saad Jbabdi. Ball and rackets: inferring fiber fanning from diffusion-weighted mri. *NeuroImage*, 60(2):1412–1425, 2012.

[28] Hui Zhang, Torben Schneider, Claudia A Wheeler-Kingshott, and Daniel C Alexander. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4):1000–1016, 2012.

[29] Eletheria Panagiotaki, Simon Walker-Samuel, Bernard Siow, S Peter Johnson, Vineeth Rajkumar, R Barbara Pedley, Mark F Lythgoe, and Daniel C Alexander. Noninvasive quantification of solid tumor microstructure using verdict mri. *Cancer research*, 74(7):1902–1912, 2014.

[30] Denis Le Bihan, Cyril Poupon, Alexis Amadon, and Franck Lethimonnier. Artifacts and pitfalls in diffusion MRI. *Journal of Magnetic Resonance Imaging*, 24(3):478–488, sep 2006.

[31] Derek K. Jones and Mara Cercignani. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR in Biomedicine*, 23(7):803–820, 2010.

[32] Jacques-Donald Tournier, Susumu Mori, and Alexander Leemans. Diffusion tensor imaging and beyond. *Magnetic Resonance in Medicine*, 65(6):1532–1556, 2011.

[33] Carlo Pierpaoli. Artifacts in Diffusion MRI. In Derek K Jones, editor, *Diffusion MRI: theory, methods, and applications.*, pages 303–317. Oxford University Press, 2010.

[34] Jesper L.R. Andersson and Stefan Skare. Image Distortion and Its Correction in Diffusion MRI. In Derek K. Jones, editor, *Diffusion MRI: theory, methods, and applications.*, pages 285–302. Oxford University Press, 2010.

[35] BP Poncelet, VJ Wedeen, RM Weisskoff, and MS Cohen. Brain parenchyma motion: measurement with cine echo-planar mr imaging. *Radiology*, 185(3):645–651, 1992.

[36] Jesper L R Andersson, Mark S. Graham, Enikö Zsoldos, and Stamatios N. Sotiropoulos. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*, 141:556–572, 2016.

[37] Daniel Gallichan, Jan Scholz, Andreas Bartsch, Timothy E Behrens, Matthew D Robson, and Karla L Miller. Addressing a systematic vibration artifact in diffusion-weighted mri. *Human brain mapping*, 31(2):193–202, 2010.

[38] Jesper L.R. Andersson and Stamatios N. Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125:1063–1078, 2016.

[39] Dan Xu, Joseph K Maier, Kevin F King, Bruce D Collick, Gaohong Wu, Robert D Peters, and R Scott Hinks. Prospective and retrospective high order eddy current mitigation for diffusion weighted echo planar imaging. *Magnetic resonance in medicine*, 70(5):1293–1305, 2013.

[40] Peter Jezzard and Stuart Clare. Sources of Distortions in Functional MRI Data. *Human Brain Mapping*, 8(May):80–85, 1999.

[41] Hákon Gudbjartsson and Samuel Patz. The rician distribution of noisy mri data. *Magnetic resonance in medicine*, 34(6):910–914, 1995.

[42] Santiago Aja-Fernández, Antonio Tristán-Vega, and W Scott Hoge. Statistical noise analysis in grappa using a parametrized noncentral chi approximation model. *Magnetic resonance in medicine*, 65(4):1195–1206, 2011.

[43] Derek K Jones and Peter J Basser. âĂIJSquashing peanuts and smashing pumpkinsâĂİ: How noise distorts diffusion-weighted MR data. *Magnetic Resonance in Medicine*, 52(5):979–993, nov 2004.

[44] Thies H. Jochimsen and Michael Von Mengershausen. ODIN - Object-oriented Development Interface for NMR. *Journal of Magnetic Resonance*, 170(1):67–78, 2004.

[45] Zhipeng Cao, Sukhoon Oh, Christopher T Sica, John M McGarrity, Timothy Horan, Wei Luo, and Christopher M Collins. Bloch-based mri system simulator considering realistic electromagnetic fields for calculation of signal, noise, and specific absorption rate. *Magnetic resonance in medicine*, 72(1):237–247, 2014.

[46] Nicolas Boulant, Xiaoping Wu, Gregor Adriany, Sebastian Schmitter, Kamil Uğurbil, and Pierre-François Moortele. Direct control of the temperature rise in parallel transmission by means of temperature virtual observation points: simulations at 10.5 tesla. *Magnetic resonance in medicine*, 75(1):249–256, 2016.

[47] Peter Nicholas, David Fushman, Vladislav Ruchinsky, and David Cowburn. The virtual nmr spectrometer: a computer program for efficient simulation of nmr experiments involving pulsed field gradients. *Journal of Magnetic Resonance*, 145(2):262–275, 2000.

[48] Wyndham B Blanton. Blochlib: a fast nmr c++ tool kit. *Journal of Magnetic Resonance*, 162(2):269–283, 2003.

[49] Andrew Simmons, Simon R Arridge, Gareth J Barker, and Steven CR Williams. Simulation of mri cluster plots and application to neurological segmentation. *Magnetic resonance imaging*, 14(1):73–92, 1996.

[50] Stephen J Riederer, SA Suddarth, SA Bobman, JN Lee, HZ Wang, and James R MacFall. Automated mr image synthesis: feasibility studies. *Radiology*, 153(1):203–206, 1984.

[51] JS Petersson, J-O Christoffersson, and K Golman. Mri simulation using the k-space formalism. *Magnetic Resonance Imaging*, 11(4):557–568, 1993.

[52] Cheng Guan Koay, Joelle E Sarlls, and Evren Özarslan. Three-dimensional analytical magnetic resonance imaging phantom in the fourier domain. *Magnetic Resonance in Medicine*, 58(2):430–436, 2007.

[53] H Michael Gach, Costin Tanase, and Fernando Boada. 2d & 3d shepp-logan phantom standards for mri. In *Systems Engineering, 2008. ICSENG'08. 19th International Conference on*, pages 521–526. IEEE, 2008.

[54] Lawrence A Shepp and Benjamin F Logan. The fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, 1974.

[55] Matthieu Guerquin-Kern, L Lejeune, Klaas P Pruessmann, and Michael Unser. Realistic analytical phantoms for parallel magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 31(3):626–636, 2012.

[56] Tri M Ngo, George SK Fung, Shuo Han, Min Chen, Jerry L Prince, Benjamin MW Tsui, Elliot R McVeigh, and Daniel A Herzka. Realistic analytical polyhedral mri phantoms. *Magnetic resonance in medicine*, 76(2):663–678, 2016.

[57] RK-S Kwan, Alan C Evans, and G Bruce Pike. MRI simulation-based

evaluation of image-processing and classification methods. *Medical Imaging, IEEE Transactions on*, 18(11):1085–1097, 1999.

[58] Chris A Cocosco, Vasken Kollokian, Remi K-S Kwan, G Bruce Pike, and Alan C Evans. Brainweb: Online interface to a 3d mri simulated brain database. In *NeuroImage*. Citeseer, 1997.

[59] Pavel Shkarin and Richard GS Spencer. Direct simulation of spin echoes by summation of isochromats. *Concepts in Magnetic Resonance Part A*, 8(4):253–268, 1996.

[60] J Bittoun, J Taquin, and M Sauzade. A computer algorithm for the simulation of any nuclear magnetic resonance (nmr) imaging method. *Magnetic Resonance Imaging*, 2(2):113–120, 1984.

[61] Ronald M Summers, Leon Axel, and Solomon Israel. A computer simulation of nuclear magnetic resonance imaging. *Magnetic resonance in medicine*, 3(3):363–376, 1986.

[62] Magnus BE Olsson, Ronnie Wirestam, and Bertil RR Persson. A computer simulation program for mr imaging: application to rf and static magnetic field imperfections. *Magnetic Resonance in Medicine*, 34(4):612–617, 1995.

[63] H. Benoit-Cattin, G. Collewet, B. Belaroussi, H. Saint-Jalmes, and C. Odet. The SIMRI project: A versatile and interactive MRI simulator. *Journal of Magnetic Resonance*, 173(1):97–115, 2005.

[64] Duane A Yoder, Yansong Zhao, Cynthia B Paschal, and J Michael Fitzpatrick. MRI simulator with object-specific field map calculations. *Magnetic resonance imaging*, 22(3):315–328, 2004.

[65] Ivana Drobnjak, David Gavaghan, Endre Süli, Joe Pitt-Francis, and Mark Jenkinson. Development of a functional magnetic resonance imag-

ing simulator for modeling realistic rigid-body motion artifacts. *Magnetic resonance in medicine*, 56(2):364–380, 2006.

[66] Ivana Drobnjak, Gaby S. Pell, and Mark Jenkinson. Simulating the effects of time-varying magnetic fields with a realistic simulated scanner. *Magnetic Resonance Imaging*, 28(7):1014–1021, sep 2010.

[67] Tony Stöcker, Kaveh Vahedipour, Daniel Pflugfelder, and N Jon Shah. High-performance computing mri simulations. *Magnetic resonance in medicine*, 64(1):186–193, 2010.

[68] Christos G Xanthis, Ioannis E Venetis, AV Chalkias, and Anthony H Aletras. Mrisimul: a gpu-based parallel approach to mri simulations. *Medical Imaging, IEEE Transactions on*, 33(3):607–617, 2014.

[69] Fang Liu, Richard Kijowski, and Wally Block. Performance of Multiple Types of Numerical MR Simulation using MRiLab. *ISMRM 2014*, page 5244, 2013.

[70] Fang Liu, Julia V Velikina, Walter F Block, Richard Kijowski, and Alexey A Samsonov. Fast realistic mri simulations based on generalized multi-pool exchange tissue model. *IEEE transactions on medical imaging*, 36(2):527–537, 2017.

[71] John G Sled and G Bruce Pike. Quantitative imaging of magnetization transfer exchange and relaxation properties in vivo using mri. *Magnetic resonance in medicine*, 46(5):923–931, 2001.

[72] Chih-Liang Chin, Felix W Wehrli, Scott N Hwang, Masaya Takahashi, and David B Hackney. Biexponential diffusion attenuation in the rat spinal cord: computer simulations based on anatomic images of axonal architecture. *Magnetic resonance in medicine*, 47(3):455–460, 2002.

[73] J-D Tournier, F Calamante, MD King, DG Gadian, and A Connelly. Limitations and requirements of diffusion tensor fiber tracking: an as-

sessment using simulations. *Magnetic resonance in medicine*, 47(4):701–708, 2002.

[74] Alessandro Daducci, Erick Jorge Canales-Rodrı, Maxime Descoteaux, Eleftherios Garyfallidis, Yaniv Gur, Ying-Chia Lin, Merry Mani, Sylvain Merlet, Michael Paquette, Alonso Ramirez-Manzanares, et al. Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion mri. *IEEE transactions on medical imaging*, 33(2):384–399, 2014.

[75] Paul T Callaghan. A simple matrix formalism for spin echo analysis of restricted diffusion under generalized gradient waveforms. *Journal of Magnetic Resonance*, 129(1):74–84, 1997.

[76] Ivana Drobnjak, Hui Zhang, Matt G Hall, and Daniel C Alexander. The matrix formalism for generalised gradients with time-varying orientation in diffusion nmr. *Journal of Magnetic Resonance*, 210(1):151–157, 2011.

[77] Matt G Hall and Daniel C Alexander. Convergence and parameter choice for monte-carlo simulations of diffusion mri. *IEEE transactions on medical imaging*, 28(9):1354–1364, 2009.

[78] Lihui Wang, Yuemin Zhu, Hongying Li, Wanyu Liu, and Isabelle E Magnin. Multiscale modeling and simulation of the cardiac fiber architecture for dmri. *IEEE Transactions on Biomedical Engineering*, 59(1):16–19, 2012.

[79] J Chetley Ford and David B Hackney. Numerical model for calculation of apparent diffusion coefficients (adc) in permeable cylinder - comparison with measured adc in spinal cord white matter. *Magnetic resonance in medicine*, 37(3):387–394, 1997.

[80] M E Bastin. Correction of eddy current-induced artefacts in diffusion tensor imaging using iterative cross-correlation. *Magnetic Resonance Imaging*, 17(7):1011–1024, 1999.

[81] Mark E. Bastin. On the use of the FLAIR technique to improve the correction of eddy current induced artefacts in MR diffusion tensor imaging. *Magnetic Resonance Imaging*, 19(7):937–950, sep 2001.

[82] Rita G. Nunes, Ivana Drobnjak, Stuart Clare, Peter Jezzard, and Mark Jenkinson. Performance of single spin-echo and doubly refocused diffusion-weighted sequences in the presence of eddy current fields with multiple components. *Magnetic Resonance Imaging*, 29(5):659–667, jun 2011.

[83] Klaus H Maier-Hein, Peter F Neher, Jean-Christophe Houde, Marc-Alexandre Côté, Eleftherios Garyfallidis, Jidan Zhong, Maxime Chamberland, Fang-Cheng Yeh, Ying-Chia Lin, Qing Ji, et al. The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, 8(1):1349, 2017.

[84] A Leemans, J Sijbers, M Verhoye, A Van der Linden, and D Van Dyck. Mathematical framework for simulating diffusion tensor mr neural fiber bundles. *Magnetic resonance in medicine*, 53(4):944–953, 2005.

[85] Thomas G. Close, Jacques Donald Tournier, Fernando Calamante, Leigh a. Johnston, Iven Mareels, and Alan Connelly. A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. *NeuroImage*, 47(4):1288–1300, 2009.

[86] Emmanuel Caruyer, Alessandro Daducci, Maxime Descoteaux, Jean-christophe Houde, Jean-philippe Thiran, and Ragini Verma. Phantomas: a flexible software library to simulate diffusion MR phantoms. In *Proc. Intl. Soc. Mag. Reson. Med.*, 2014.

[87] Bryce Wilkins, Namgyun Lee, Niharika Gajawelli, Meng Law, and Natasha Leporé. Fiber estimation and tractography in diffusion mri: development of simulated brain images and comparison of multi-fiber analysis methods at clinical b-values. *Neuroimage*, 109:341–356, 2015.

[88] Peter F. Neher, Frederik B. Laun, Bram Stieltjes, and Klaus H. Maier-Hein. Fiberfox: Facilitating the creation of realistic white matter software phantoms. *Magnetic Resonance in Medicine*, 1470:1460–1470, 2013.

[89] Daniele Perrone, Ben Jeurissen, Jan Aelterman, Timo Roine, Jan Sijbers, Aleksandra Pizurica, Alexander Leemans, and Wilfried Philips. D-BRAIN: Anatomically Accurate Simulated Diffusion MRI Brain Data. *PloS one*, 11(3):e0149778, 2016.

[90] Hong-bo Du, Li-hui Wang, Wan-yu Liu, Feng Yang, Zhi Li, Yue-min Zhu, Cnrs Umr, Insa Lyon, and Université De Lyon. Diffusion Mri Simulation for Human Brain Based on the Atlas. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium*, pages 898–902, 2016.

[91] Chun-Hung Yeh, Benoît Schmitt, Denis Le Bihan, Jing-Rebecca Li-Schlittgen, Ching-Po Lin, and Cyril Poupon. Diffusion Microscopist Simulator: A General Monte Carlo Simulation System for Diffusion Magnetic Resonance Imaging. *PLoS ONE*, 8(10):e76626, 2013.

[92] Jesper L R Andersson and Stefan Skare. A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI. *NeuroImage*, 16(1):177–199, may 2002.

[93] Markus Nilsson, Filip Szczepankiewicz, Danielle van Westen, and Oskar Hansson. Extrapolation-based references improve motion and eddy-current correction of high b-value dwi data: Application in parkinson's disease dementia. *PloS one*, 10(11):e0141825, 2015.

[94] Timothy G. Reese, O. Heid, R. M. Weisskoff, and V. J. Wedeen. Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. *Magnetic Resonance in Medicine*, 49(1):177–182, jan 2003.

[95] P Jezzard and R S Balaban. Correction for geometric distortion in echo

planar images from B0 field variations. *Magnetic resonance in medicine*, 34(1):65–73, jul 1995.

[96] Jesper L R Andersson, Stefan Skare, and John Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20:870–888, 2003.

[97] J C Haselgrove and J R Moore. Correction for distortion of echo-planar images used to calculate the apparent diffusion coefficient. *Magnetic Resonance in Medicine*, 36(6):960–964, 1996.

[98] Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2):143–156, 2001.

[99] Ipek Oguz, Mahshid Farzinfar, Joy Matsui, Francois Budin, Zhexing Liu, Guido Gerig, Hans J Johnson, and Martin Styner. DTIPrep: quality control of diffusion-weighted images. *Frontiers in neuroinformatics*, 8, 2014.

[100] J-F Mangin, Cyril Poupon, Chris Clark, Denis Le Bihan, and Isabelle Bloch. Distortion correction and robust tensor estimation for MR diffusion imaging. *Medical image analysis*, 6(3):191–198, 2002.

[101] M a Horsfield. Mapping eddy current induced fields for the correction of diffusion-weighted echo planar images. *Magnetic resonance imaging*, 17(9):1335–1345, nov 1999.

[102] Jiancheng Zhuang, Zhong-Lin Lu, Christine Bouteiller Vidal, and Hanna Damasio. Correction of eddy current distortions in high angular resolution diffusion imaging. *Journal of Magnetic Resonance Imaging*, 37(6):1460–7, jun 2013.

[103] Y Bai and D Alexander. Model-based registration to correct for motion

between acqui- sitions in diffusion MR imaging. *5th IEEE Int Symp Biom Imaging*, (1):947–950, 2008.

[104] Shani Ben-Amitay, Derek K. Jones, and Yaniv Assaf. Motion correction and registration of high b-value diffusion weighted images. *Magnetic Resonance in Medicine*, 67(6):1694–1702, jun 2012.

[105] Ivana Drobnjak, David Gavaghan, Endre Süli, Joe Pitt-Francis, and Mark Jenkinson. Development of a functional magnetic resonance imaging simulator for modeling realistic rigid-body motion artifacts. *Magnetic Resonance in Medicine*, 56(2):364–380, aug 2006.

[106] David C Van Essen, Kamil Ugurbil, E Auerbach, D Barch, TEJ Behrens, R Bucholz, A Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

[107] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *Medical Imaging, IEEE Transactions on*, 20(1):45–57, 2001.

[108] D. C. Alexander, G. J. Barker, and S. R. Arridge. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magnetic Resonance in Medicine*, 48:331–340, 2002.

[109] J. Donald Tournier, Fernando Calamante, and Alan Connelly. Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. *NMR in Biomedicine*, 26(August):1775–1786, 2013.

[110] T Niendorf, R M Dijkhuizen, D G Norris, M van Lookeren Campagne, and K Nicolay. Biexponential diffusion attenuation in various states of brain tissue: implications for diffusion-weighted imaging. *Magnetic Resonance in Medicine*, 36(6):847–857, 1996.

[111] Peter Jezzard, Alan S. Barnett, and Carlo Pierpaoli. Characterization of and correction for eddy current artifacts in echo planar diffusion imaging. *Magnetic Resonance in Medicine*, 39(5):801–812, may 1998.

[112] M. Zaitsev, J. Hennig, and O. Speck. Point spread function mapping with parallel imaging techniques and high acceleration factors: Fast, robust, and flexible method for echo-planar imaging distortion correction. *Magnetic Resonance in Medicine*, 52(5):1156–1166, nov 2004.

[113] Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, Peter Boesiger, et al. Sense: sensitivity encoding for fast mri. *Magnetic resonance in medicine*, 42(5):952–962, 1999.

[114] Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (grappa). *Magnetic resonance in medicine*, 47(6):1202–1210, 2002.

[115] Evren Özarslan, Cheng Guan Koay, Timothy M. Shepherd, Michal E. Komlosh, M. Okan Irfanoglu, Carlo Pierpaoli, and Peter J. Basser. Mean apparent propagator (MAP) MRI: A novel diffusion imaging method for mapping tissue microstructure. *NeuroImage*, 78:16–32, 2013.

[116] Julian Maclaren, Michael Herbst, Oliver Speck, and Maxim Zaitsev. Prospective motion correction in brain imaging: a review. *Magnetic resonance in medicine*, 69(3):621–636, 2013.

[117] Gustavo Kunde Rohde, a. S. Barnett, P. J. Basser, S. Marenco, and C. Pierpaoli. Comprehensive Approach for Correction of Motion and Distortion in Diffusion-Weighted MRI. *Magnetic Resonance in Medicine*, 51(1):103–114, jan 2004.

[118] P a Cook, Y Bai, K K Seunarine, M G Hall, G J Parker, and D C Alexander. Camino: Open-Source Diffusion-MRI Reconstruction and

Processing. *14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, 14:2759, 2006.

[119] D. K. Jones, M. a. Horsfield, and a. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine*, 42(3):515–525, 1999.

[120] S. M. Maniega, Mark E. Bastin, and Paul a. Armitage. A quantitative comparison of two methods to correct eddy current-induced distortions in DT-MRI. *Magnetic Resonance Imaging*, 25(3):341–349, apr 2007.

[121] Haruyasu Yamada, Osamu Abe, Takashi Shizukuishi, Junko Kikuta, Takahiro Shinozaki, Ko Dezawa, Akira Nagano, Masayuki Matsuda, Hiroki Haradome, and Yoshiki Imamura. Efficacy of Distortion Correction on Diffusion Imaging: Comparison of FSL Eddy and Eddy_Correct Using 30 and 60 Directions Diffusion Encoding. *PloS one*, 9(11):e112411, jan 2014.

[122] Siawoosh Mohammadi, Harald E. Möller, Harald Kugel, Dirk K. Müller, and Michael Deppe. Correcting eddy current and motion effects by affine whole-brain registrations: Evaluation of three-dimensional distortions and comparison with slicewise correction. *Magnetic Resonance in Medicine*, 64(4):1047–1056, oct 2010.

[123] J Hennig, A Nauerth, and H Friedburg. Rare imaging: a fast imaging method for clinical mr. *Magnetic resonance in medicine*, 3(6):823–833, 1986.

[124] Robert Turner and Denis Le Bihan. Single-shot diffusion imaging at 2.0 tesla. *Journal of Magnetic Resonance (1969)*, 86(3):445–452, 1990.

[125] C Hutton, A Bork, O Josephs, R Deichmann, J Ashburner, and R Turner. Image distortion correction in fMRI: A quantitative evaluation. *Neuroimage*, 16(1):217–240, 2002.

[126] J L Andersson, M Richter, W Richter, S Skare, R G Nunes, M D Robson, and T E Behrens. Effects of susceptibility distortions on tractography. In *Proceedings of International Society of Magnetic Resonance in Medicine*, volume 11, page 87, 2004.

[127] Karl V Embleton, Hamied A Haroon, David M Morris, Matthew A Lambon Ralph, and Geoff JM Parker. Distortion correction for diffusion-weighted mri tractography and fmri in the temporal lobes. *Human brain mapping*, 31(10):1570–1587, 2010.

[128] M Okan Irfanoglu, Lindsay Walker, Joelle Sarlls, Stefano Marenco, and Carlo Pierpaoli. Effects of image distortions originating from susceptibility variations and concomitant fields on diffusion MRI tractography results. *NeuroImage*, 61(1):275–88, 2012.

[129] Steen Moeller, Essa Yacoub, Cheryl A Olman, Edward Auerbach, John Strupp, Noam Harel, and Kâmil Uğurbil. Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fmri. *Magnetic Resonance in Medicine*, 63(5):1144–1153, 2010.

[130] Kawin Setsompop, Borjan A Gagoski, Jonathan R Polimeni, Thomas Witzel, Van J Wedeen, and Lawrence L Wald. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine*, 67(5):1210–1224, 2012.

[131] Stamatios N. Sotiropoulos, Saad Jbabdi, Junqian Xu, Jesper L. Andersson, Steen Moeller, Edward J. Auerbach, Matthew F. Glasser, Moises Hernandez, Guillermo Sapiro, Mark Jenkinson, David a. Feinberg, Essa Yacoub, Christophe Lenglet, David C. Van Essen, Kamil Ugurbil, and Timothy E J Behrens. Advances in diffusion MRI acquisition and pro-

cessing in the Human Connectome Project. *NeuroImage*, 80:125–143, oct 2013.

[132] Karla L Miller, Fidel Alfaro-Almagro, Neal K Bangerter, David L Thomas, Essa Yacoub, Junqian Xu, Andreas J Bartsch, Saad Jbabdi, Stamatios N Sotiropoulos, Jesper LR Andersson, et al. Multimodal population brain imaging in the uk biobank prospective epidemiological study. *Nature neuroscience*, 19(11):1523–1536, 2016.

[133] Emer J Hughes, Tobias Winchman, Francesco Padormo, Rui Teixeira, Julia Wurie, Maryanne Sharma, Matthew Fox, Jana Hutter, Lucilio Cordero-Grande, Anthony N Price, et al. A dedicated neonatal brain imaging system. *Magnetic Resonance in Medicine*, 2016.

[134] Colin Studholme, R. Todd Constable, and James S. Duncan. Accurate alignment of functional EPI data to anatomical MRI using a physics-based distortion model. *IEEE Transactions on Medical Imaging*, 19(11):1115–1127, 2000.

[135] Jan Kybic, Philippe Thévenaz, Arto Nirkko, and Michael Unser. Unwarping of unidirectionally distorted EPI images. *IEEE Transactions on Medical Imaging*, 19(2):80–93, 2000.

[136] M Wu, L C Chang, L Walker, H Lemaitre, a S Barnett, S Marenco, and C Pierpaoli. Comparison of EPI distortion correction methods in diffusion tensor MRI using a novel framework. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 321–329. Springer, 2008.

[137] Siamak Ardekani and Usha Sinha. Geometric distortion correction of high-resolution 3 T diffusion tensor brain images. *Magnetic Resonance in Medicine*, 54(5):1163–1171, 2005.

[138] Ran Tao, P Thomas Fletcher, Samuel Gerber, and Ross T Whitaker. A variational image-based approach to the correction of susceptibility

artifacts in the alignment of diffusion weighted and structural mri. In *International Conference on Information Processing in Medical Imaging*, pages 664–675. Springer, 2009.

[139] Ali Gholipour, Nasser Kehtarnavaz, Benoit Scherrer, and Simon K Warfield. On the accuracy of unwarping techniques for the correction of susceptibility-induced geometric distortion in magnetic resonance echo-planar images. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 6997–7000. IEEE, 2011.

[140] M. O. Irfanoglu, L. Walker, S. Sammet, C. Pierpaoli, and R. Machiraju. Susceptibility distortion correction for echo planar images with non-uniform B-spline grid sampling: A diffusion tensor image study. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 174–181, 2011.

[141] Chitresh Bhushan, Soyoung Choi, Anand a Joshi, Justin P Haldar, David W Shattuck, and Richard M Leahy. Co-registration and distortion correction of diffusion and anatomical images based on inverse contrast normalization. *NeuroImage*, pages 30–33, 2014.

[142] P. J. Reber, E. C. Wong, R. B. Buxton, and L. R. Frank. Correction of off resonance related distortion in echo planar imaging using EPI based field maps. *Magn Reson Med*, 39:328–330, 1998.

[143] Desmond T B Yeo, Thomas L. Chenevert, Jeffrey A. Fessler, and Boklye Kim. Zero and first-order phase shift correction for field map estimation with dual-echo GRE using bipolar gradients. *Magnetic Resonance Imaging*, 25(9):1263–1271, 2007.

[144] Amanda K. Funai, Jeffrey A. Fessler, Desmond T B Yeo, Douglas C. Noll, and Valur T. Olafsson. Regularized field map estimation in MRI. *IEEE Transactions on Medical Imaging*, 27(10):1484–1494, 2008.

[145] A Matakos, J Balter, and Y Cao. Estimation of geometrically undistorted B0 inhomogeneity maps. *Physics in Medicine and Biology*, 59(17):4945–4959, 2014.

[146] Hsuan Chang and J Michael Fitzpatrick. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE Transactions on medical imaging*, 11(3):319–329, 1992.

[147] Paul S Morgan, Richard W Bowtell, Dominick JO McIntyre, and Brian 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.

[148] D. Holland, J.M. Kuperman, and A.M. Dale. Efficient Correction of Inhomogeneous Static Magnetic Field- Induced Distortion in Echo Planar Imaging. *NeuroImage*, 50(1):1–18, 2011.

[149] Lars Ruthotto, Siawoosh Mohammadi, Constantin Heck, Jan Modersitzki, and Nikolaus Weiskopf. Hyperelastic susceptibility artifact correction of DTI in SPM. *Informatik aktuell*, pages 344–349, 2013.

[150] M. Okan Irfanoglu, Pooja Modi, Amritha Nayak, Elizabeth B. Hutchinson, Joelle Sarlls, and Carlo Pierpaoli. DR-BUDDI (Diffeomorphic Registration for Blip-Up blip-Down Diffusion Imaging) method for correcting echo planar imaging distortions. *NeuroImage*, 106:284–299, 2015.

[151] Renaud Hedouin, Olivier Commowick, Elise Bannier, Benoit Scherrer, Maxime Taquet, Simon Warfield, and Christian Barillot. Block-Matching Distortion Correction of Echo-Planar Images With Opposite Phase Encoding Directions. *IEEE Transactions on Medical Imaging*, 36(5):1106–1115, 2017.

[152] M D Robson, J C Gore, and R T Constable. Measurement of the point spread function in MRI using constant time imaging. *Magnetic resonance in medicine*, 38(12):733–740, 1997.

[153] Huairen Zeng and R. Todd Constable. Image distortion correction in EPI: Comparison of field mapping with point spread function mapping. *Magnetic Resonance in Medicine*, 48(1):137–146, jul 2002.

[154] Heidi A. Ward, Stephen J. Riederer, and Clifford R. Jack. Real-time autoshimming for echo planar timecourse imaging. *Magnetic Resonance in Medicine*, 48(5):771–780, 2002.

[155] A. Alhamud, Paul A. Taylor, Andre J W van der Kouwe, and Ernesta M. Meintjes. Real-time measurement and correction of both B0 changes and subject motion in diffusion tensor imaging using a double volumetric navigated (DvNav) sequence. *NeuroImage*, 126:60–71, 2016.

[156] J L Andersson, C Hutton, J Ashburner, R Turner, and K Friston. Modeling geometric deformations in EPI time series. *NeuroImage*, 13(5):903–919, 2001.

[157] Chloe Hutton, Jesper Andersson, Ralf Deichmann, and Nikolaus Weiskopf. Phase informed model for motion and susceptibility. *Human Brain Mapping*, 34(11):3086–3100, 2013.

[158] Paul A. Taylor, A. Alhamud, Andre van der Kouwe, Muhammad G. Saleh, Barbara Laughton, and Ernesta Meintjes. Assessing the performance of different DTI motion correction strategies in the presence of EPI distortion correction. *Human Brain Mapping*, 37(12):4405–4424, 2016.

[159] Jesper L R Andersson. Geometric Distortions in Diffusion MRI. *Diffusion MRI: From Quantitative Measurement to In vivo Neuroanatomy: Second Edition*, pages 63–85, 2013.

[160] Marc Modat, Gerard R Ridgway, Zeike A Taylor, Manja Lehmann, Josephine Barnes, David J Hawkes, Nick C Fox, and Sébastien Ourselin. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–284, 2010.

[161] Chitresh Bhushan, Anand A. Joshi, Richard M. Leahy, and Justin P. Haldar. Improved B0-distortion correction in diffusion MRI using interlaced q-space sampling and constrained reconstruction. *Magnetic Resonance in Medicine*, 72(5):1218–1232, 2014.

[162] Mark Jenkinson. Fast, automated, N-dimensional phase-unwrapping algorithm. *Magnetic Resonance in Medicine*, 49(1):193–197, 2003.

[163] Stephen M. Smith, Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E J Behrens, Heidi Johansen-Berg, Peter R. Bannister, Marilena De Luca, Ivana Drobnjak, David E. Flitney, Rami K. Niazy, James Saunders, John Vickers, Yongyue Zhang, Nicola De Stefano, J. Michael Brady, and Paul M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. In *NeuroImage*, volume 23, 2004.

[164] Emer Hughes, Lucilio Cordero-Grande, Maria Murgasova, Jana Hutter, Anthony Price, Ana Dos Santos Gomes, Joanna Allsop, Johannes Steinweg, Nora Tusor, Julia Wurie, et al. The developing human connectome: announcing the first release of open access neonatal brain imaging. In *23rd Annual Meeting of the Organization for Human Brain Mapping*, 2017.

[165] Mark Jenkinson, James L. Wilson, and Peter Jezzard. Perturbation method for magnetic field calculations of nonconductive objects. *Magnetic Resonance in Medicine*, 52(3):471–477, 2004.

[166] Xiaohong Joe Zhou, Yiping P. Du, Matt A. Bernstein, H. Glenn Reynolds, Joseph K. Maier, and Jason A. Polzin. Concomitant magnetic-field-induced artifacts in axial echo planar imaging. *Magnetic Resonance in Medicine*, 39(4):596–605, 1998.

[167] Sijia Wang, Daniel J. Peterson, J. C. Gatenby, Wenbin Li, Thomas J. Grabowski, and Tara M. Madhyastha. Evaluation of field map and non-

linear registration methods for correction of susceptibility artifacts in diffusion mri. *Frontiers in Neuroinformatics*, 11:17, 2017.

[168] O. Esteban, A. Daducci, E. Caruyer, K. O'Brien, M.J. Ledesma-Carbayo, M. Bach-Cuadra, and A. Santos. Simulation-based evaluation of susceptibility distortion correction methods in diffusion mri for connectivity analysis. *2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014*, pages 738–741, 2014.

[169] Vinai Roopchansingh, Robert W. Cox, Andrzej Jesmanowicz, B. Douglas Ward, and James S. Hyde. Single-shot magnetic field mapping embedded in echo-planar time-course imaging. *Magnetic Resonance in Medicine*, 50(4):839–843, 2003.

[170] Bahram Marami, Seyed Sadegh Mohseni Salehi, Onur Afacan, Benoit Scherrer, Caitlin K Rollins, Edward Yang, Judy A Estroff, Simon K Warfield, and Ali Gholipour. Temporal slice registration and robust diffusion-tensor reconstruction for improved fetal brain structural connectivity analysis. *NeuroImage*, 2017.

[171] Jesper L.R. Andersson, Mark S. Graham, Ivana Drobnjak, Hui Zhang, Nicola Filippini, and Matteo Bastiani. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *NeuroImage*, 152:450–466, 2017.

[172] Jesper L.R. Andersson, Mark S Graham, Ivana Drobnjak, Hui Zhang, and Jon Campbell. Susceptibility-induced distortion that varies due to motion: Correction in diffusion mr without acquiring additional data. *NeuroImage*, In press.

[173] D C Van Essen, K Ugurbil, E Auerbach, D Barch, T E J Behrens, R Bucholz, A Chang, L Chen, M Corbetta, S W Curtiss, S Della Penna, D Feinberg, M F Glasser, N Harel, A C Heath, L Larson-prior, D Marcus, G Michalareas, S Moeller, R Oostenveld, S E Petersen, F Prior, B L

Schlaggar, S M Smith, A Z Snyder, J Xu, E Yacoub, Wu-minn H C P Consortium, and M E G Eeg. The Human Connectome Project : A data acquisition perspective. *NeuroImage*, 62(4):2222–2231, 2012.

[174] Zhexing Liu, Yi Wang, Guido Gerig, Sylvain Gouttard, Ran Tao, Thomas Fletcher, and Martin Styner. Quality control of diffusion weighted images. In *Proceedings of SPIE–the International Society for Optical Engineering*, 2010.

[175] Ipek Oguz, Mahshid Farzinfar, Joy Matsui, Francois Budin, Zhexing Liu, Guido Gerig, Hans J Johnson, and Martin Styner. DTIPrep: quality control of diffusion-weighted images. *Frontiers in neuroinformatics*, 8(January):4, jan 2014.

[176] Oscar Esteban, Daniel Birman, Marie Schaer, Oluwasanmi O Koyejo, Russell A Poldrack, and Krzysztof J Gorgolewski. MRIQC: Predicting Quality in Manual MRI Assessment Protocols Using No-Reference Image Quality Measures. *bioRxiv*, pages 1–18, 2017.

[177] Fidel Alfaro-Almagro, Mark Jenkinson, Neal K Bangerter, Jesper LR Andersson, Ludovica Griffanti, Gwenaelle Douaud, Stamatios Sotiropoulos, Saad Jbabdi, Moises Hernandez Fernandez, Emmanuel Vallee, Diego Vidaurre, Matthew Webster, Paul D McCarthy, Chris Rorden, Alessandro Daducci, Daniel Alexander, Hui Zhang, Iulius Dragonu, Paul Matthews, Karla L Miller, and Stephen M Smith. Image Processing and Quality Control for the first 10,000 Brain Imaging Datasets from UK Biobank. *bioRxiv*, 2017.

[178] Juan Eugenio Iglesias, Garikoitz Lerma-usabiaga, Luis C Garcia-Peraza-Herrera, Sara Martinez, and Pedro M Paz-alonso. Retrospective head motion estimation in structural brain MRI with 3D CNNs. *Medical image computing and computer-assisted intervention : MICCAI ... Interna-*

tional Conference on Medical Image Computing and Computer-Assisted Intervention, pages 314–322, 2017.

[179] Christopher Kelly, Max Pietsch, Serena Counsell, and J-donald Tournier. Transfer learning and convolutional neural net fusion for motion artefact detection. *Proc. Intl. Soc. Mag. Reson. Med.*, pages 1–2, 2016.

[180] Anastasia Yendiki, Kami Koldewyn, Sita Kakunoori, Nancy Kanwisher, and Bruce Fischl. Spurious group differences due to head motion in a diffusion mri study. *Neuroimage*, 88:79–90, 2014.

[181] Estanislao Oubel, Mériam Koob, Colin Studholme, Jean Louis Dietemann, and François Rousseau. Reconstruction of scattered data in fetal diffusion MRI. *Medical Image Analysis*, 16(1):28–37, 2012.

[182] Bahram Marami, Seyed Sadegh Mohseni Salehi, Onur Afacan, Benoit Scherrer, Caitlin K. Rollins, Edward Yang, Judy A. Estroff, Simon K. Warfield, and Ali Gholipour. Temporal slice registration and robust diffusion-tensor reconstruction for improved fetal brain structural connectivity analysis. *NeuroImage*, 156(April):475–488, 2017.

[183] Lori-Anne Williams, Neil Gelman, Paul A Picot, David S Lee, James R Ewing, Victor K Han, and R Terry Thompson. Neonatal brain: regional variability of in vivo mr imaging relaxation rates at 3.0 tâĂŤinitial experience. *Radiology*, 235(2):595–603, 2005.

[184] Ilana R Leppert, C Robert Almli, Robert C McKinstry, Robert V Mulkern, Carlo Pierpaoli, Michael J Rivkin, and G Bruce Pike. T2 relaxometry of normal pediatric brain development. *Journal of Magnetic Resonance Imaging*, 29(2):258–267, 2009.

[185] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large

Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[186] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[187] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. *arXiv preprint*, 2015.

[188] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[189] François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

[190] Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, Virginia Newcombe, Joanna Simpson, Andrew Kane, David Menon, Aditya Nori, Antonio Criminisi, Daniel Rueckert, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *International Conference on Information Processing in Medical Imaging*, pages 597–609. Springer, 2017.

[191] Konstantinos Bousmalis, Alex Irpan, Paul Wohlhart, Yunfei Bai, Matthew Kelcey, Mrinal Kalakrishnan, Laura Downs, Julian Ibarz, Peter Pastor, Kurt Konolige, et al. Using simulation and domain adaptation to improve efficiency of deep robotic grasping. *arXiv preprint arXiv:1709.07857*, 2017.

[192] Anagha Deshmane, Vikas Gulani, Mark A Griswold, and Nicole Seiberlich. Parallel mr imaging. *Journal of Magnetic Resonance Imaging*, 36(1):55–72, 2012.

[193] Markus Barth, Felix Breuer, Peter J Koopmans, David G Norris, and Benedikt A Poser. Simultaneous multislice (sms) imaging techniques. *Magnetic resonance in medicine*, 75(1):63–81, 2016.

[194] Xiao Yang, Roland Kwitt, and Marc Niethammer. Quicksilver: Fast predictive image registration-a deep learning approach. *arXiv preprint arXiv:1703.10908*, 2017.