Efficient statistical methods for inference and model selection in Diffusion-Weighted MRI models

Lisa Mott

Thesis submitted to The University of Nottingham

for the degree of Doctor of Philosophy

September 2015

Dedicated to the most hardworking and intelligent man

I know and ever will know,

Grandad Jeff!

Abstract

Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) on the brain is a revolutionary method that provides in-vivo access to tissue macrostructure non-invasively (Basser *et al.*, 1994). Recently, DW-MRI has been shown to have great potential in characterising brain microstructure, such as diameter and size distribution of neuronal fibres, features that were available so far only post-mortem or through animal studies (Zhang *et al.*, 2011). Using a process known as Tractography the existence of brain connections can be estimated using a set of DW images (Basser *et al.*, 2000).

The main aim of this thesis is to develop efficient methods for studying Tractography within a Bayesian framework. In order to characterise the white matter in the brain we focus on the widely used partial volume model (Behrens *et al.*, 2003). We describe methods that are both time and computationally efficient for estimating the parameters of the partial volume model, before reparametrising the model, so that parameter estimation is viable in some special cases. The partial volume model allows for multiple fibre orientations so we develop methodology to choose between the number of white matter fibres in a voxel. We then take into account the uncertainty in the number of fibre orientations and provide a Fully Probabilistic Tractography method as an alternative to existing Tractography algorithms. Finally we look into the Global Tractography model (Jbabdi *et al.*, 2007) and develop efficient methods for inferring connections between brain regions by investigating methods based on Thermody-

namic Integration.

Acknowledgements

First of all I would like to thank my supervisor Dr. Theo Kypraios, for his invaluable help throughout the PhD.

The biggest thank you goes out to my family. To my Dad and Mum, it is only since I went to University that I began to understand the role of social class in life opportunities in the UK. Therefore you must have done an extremely good job in bringing me up in a coal mining family during the closure of the local pit. Even though we never had much, you both always worked hard to provide the best you could for me and my brother and I feel that the stress that we went through due to constant job uncertainty, job losses and hard manual labour jobs has gave me a strong work ethic that helped me in the long run.

To my brother, Joe, thanks for putting up with having a maths geek sister!

To my special little westie Angel, thanks for always being such a cute, loving and wonderful doggy.

To my wonderful Nana and Grandad, thanks for always being there for me. Grandad you are the most intelligent and hardworking person I know and your wise advice has always helped me through life. You also helped me be less afraid to pursue Mathematics as you let me see it is not shameful to enjoy

Maths and Science. Nana, you were also always there for me and your support is also very much appreciated.

To my late Nana Agnes also, although unfortunately you never saw the end of my PhD, you always let me know how very proud you were to have a grandchild with a degree, I will always be thankful that you were not took before you saw me graduate from my bachelor's degree.

There is not enough room here to thank all my family, but I am thankful to everyone.

To Mattia, thank you very much for having to put up with me in both my bad and good days while doing this PhD.

During the PhD there have been many people who have made being in the maths department very happy. A special thanks goes out to the best office I could ever have, B50, you all made the office a pleasure to work in, and I am very grateful that I got put in the best office possible. There are many other people who have been good friends of me during the PhD, but unfortunately I cannot name all of them here but you will know who you are.

A special thanks goes to my second supervisor Chris Brignell particularly for his helpful comments on the draft of this thesis and to Paul Morgan and Stam Sotiropoulos for their invaluable help at describing the physics of MRI.

Finally I would like to thank all the people who volunteered to have a MRI scan for this project.

Contents

CHAPTER 1

Introduction

1.1 Motivation

The human brain is made up mainly of two components, grey matter and white matter. Grey matter contains the functional centres of the brain that process information, while white matter makes up about half of the brain volume and acts as the "wiring" of the brain, connecting the different functional centres (Filley, 2011). Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) enables the reconstruction of the white matter tracts in the brain, non-invasively and in-vivo by Tractography (Basser *et al.*, 2000). Currently, Tractography is the only technique that allows non-invasive reconstruction of fibre bundles in the human brain (Schultz *et al.*, 2013). The diffusion of water molecules can be quantified, and thus can provide information about the underlying structure of the brain (Sotiropoulos, 2010), this information can then be used within Tractography.

Understanding white matter tracts is crucial to understanding the brain's functions, which will allow better knowledge of how the brain works. White matter is partly or exclusively responsible for well over 100 brain disorders (Filley, 2011) and therefore by investigating the white matter tracts it can be studied how the interruption of normal white matter connections could lead to neuro-

logical disorders. Pathology induced changes could also be identified, if these tracts are well understood. Another important application of understanding white matter tracts is in neurosurgical planning. Just one example of how DW-MRI could improve the outcome in neurosurgery is in its impact on reducing postoperative motor deficits and increasing survival times in cerebral glioma surgery, as studied in Wu *et al*. (2007)

Existing Tractography methods are fairly computationally demanding and are often inaccurate (Jbabdi and Johansen-Berg, 2011). Some methods for Tractography tend to underestimate the true size of tracts (Kinoshita *et al*., 2005) and neurosurgeons are still not happy with the reproducibility of tracts using current software packages (Bürgel *et al*., 2009). The aim of this thesis is to develop a computationally feasible and time efficient statistical method for Tractography by developing novel statistical methodology for parameter estimation and model selection within the context of DW-MRI models. We then apply the developed methodology to real DW-MRI data.

1.2 White matter

In this section we briefly describe white matter; more details can be found in Filley (2011). The human brain is made up of white and grey matter, with around 50% of it being constituted of white matter. White matter was distinguished from grey matter in 1543. White matter acts as the "wiring" of the brain, thus connecting different grey matter sectors. It is therefore important to understand the connections within the brain, and how different regions interact with each other. White matter is a collection of fibres (tracts) that were classified into three different types of fibres by Meynert (1885); these different types are association fibres, projection fibres and commissural fibres. The tracts consist of myelinated axons of neurons. White matter neuroanatomy varies over the life span (Woz-

niak and Lim, 2006) so another motivation for understanding white matter is to understand how changes occur with age.

For many years the role of white matter was seen as irrelevant in comparison to that of grey matter, and more attention was given to understanding the latter. However white matter has been shown more recently to be partially or exclusively involved in over 100 disorders (Filley, 2011). All of these disorders have a serious impact on cognitive or emotional function.

Carl Wernicke (1874) put forward the idea of a disconnection paradigm being responsible for cognitive disorders in neurology and psychiatry. He believed that the breakdown of connections between different regions caused these disorders. More recently Crick and Jones (1993) stated that "to interpret the activity of the living human brains, their anatomy must be known in detail." Methods for understanding the white matter fibre tracts include dissection (Klingler, 1935), myelin stains (Weigert, 1897), tracer substances (Lanciego and Wouterlood, 2000), stereology (Schmitz and Hof, 2005) and polarized light imaging (Axer and Keyserlingk, 2000). These methods are time consuming and require very skilled workers to implement them. Due to the invasive nature of the methods, most of them are not carried out in-vivo in humans (Axer, 2011).

It wasn't until 1985 that Diffusion MRI was introduced as a tool for brain imaging (Basser *et al*., 1994). Diffusion MRI has been fundamental in providing information about white matter that was previously unknown; it has not only helped widen the knowledge of known diseases but has also helped us discover new diseases including vanishing white matter disease (van der Knapp *et al*., 2006) and adult onset leukodystrophy with neuroaxonal spheroids (Freeman *et al*., 2009).

A further development was the introduction of Diffusion Tensor Imaging (DTI) in the 1990s by Basser, Mattiello and Le Bihan, which has revised the understanding of white matter neuroanatomy (Aralasmak *et al.*, 2006). The original MRI that was introduced analyses the macrostructure of white matter, while DTI allows the examination of the microstructure of white matter (Zhang *et al.*, 2011) .

A functional connection between regions in the brain is dependent on there being an anatomical connection (Passingham *et al.*, 2002). A potential important future application of DW-MRI will be to use information from the anatomy obtained by DTI, to help in inferring information about functional connections within the brain.

1.3 Physics of diffusion

To understand how MRI works, we first need to understand the idea of diffusion. Brownian motion was discussed by Einstein (1905), who described an experiment performed by the botanist Robert Brown in 1827. In this experiment Brown observed a jiggling of pollen grains that were placed in water. Einstein suggested that the pollen grains' movement was caused by the motion of the water molecules surrounding them. Therefore the motion of the water molecules can be observed through an object that is visible by a microscope. Similarly in Magnetic Resonance Imaging the radio waves emitted from atomic nuclei act as the pollen grains allowing one to observe the diffusion of water within the brain structure (Callaghan, 2011).

In DW-MRI an assumption is made that the underlying tissue structure determines how the water diffuses (Kang *et al.*, 2005). Figure 1.1 shows typical diffusion in a barrier free medium, grey matter and white matter. In white matter

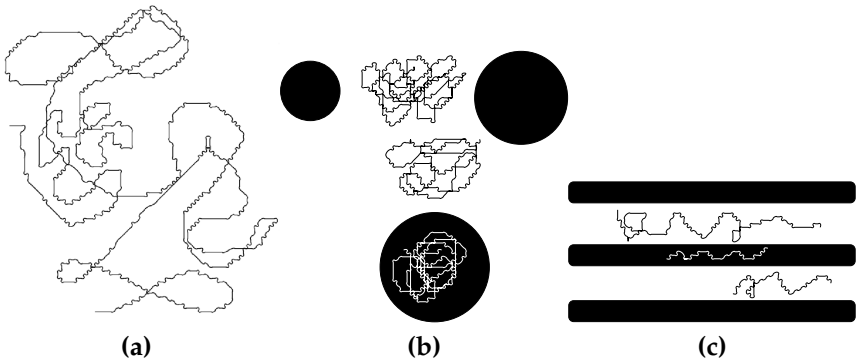


Figure 1.1: Typical diffusion in (a) a barrier free medium, (b) grey matter and (c) white matter

water will diffuse along the fibres, so it will tend to follow their direction. Although water in grey matter does not seem to diffuse in a particular direction, the diffusion is still more restricted than in a barrier-free medium.

Diffusion is the random movement of particles from a region of higher concentration to a region of lower concentration (Hobbie, 1997). Diffusion is governed by Fick's second law (Fick, 1855), also known as the diffusion equation,

$$\frac{\partial n}{\partial t} = D \bigtriangledown^2 n, \tag{1.3.1}$$

where $n(\boldsymbol{r}, t)$ is the local concentration of particles at location $\boldsymbol{r}$ and time t, $\bigtriangledown^2$ represents the Laplacian and D is the diffusion coefficient.

Einstein (1905), from Equation (1.3.1), showed that Fick's second law could be adapted so that it works also in self-diffusion. Self-diffusion is the diffusion of molecules of a medium that is caused by the medium itself. He did this by setting the local concentration of particles to be

$$n(\boldsymbol{r}', t) = \int n(\boldsymbol{r}, 0) P(\boldsymbol{r}|\boldsymbol{r}', t) d\boldsymbol{r}, \tag{1.3.2}$$

where $P(\boldsymbol{r}|\boldsymbol{r}', t)$ is the probability of a particle moving from $\boldsymbol{r}$ to $\boldsymbol{r}'$ in a time

t (Callaghan, 2011). By placing Equation (1.3.2) into Equation (1.3.1), Einstein derived the set of equations

$$\frac{\partial}{\partial t}P(\boldsymbol{r}|\boldsymbol{r}',t) = D \bigtriangledown^2 P(\boldsymbol{r}|\boldsymbol{r}',t)$$

and

$$P(\boldsymbol{r}|\boldsymbol{r}',t) = (4\pi Dt)^{-3/2}\exp\left(-\frac{(\boldsymbol{r}'-\boldsymbol{r})^2}{4Dt}\right).$$

These equations are only true in a medium where the diffusion is isotropic (Callaghan, 2011). In an anisotropic medium the above equations become

$$\frac{\partial}{\partial t}P(\boldsymbol{r}|\boldsymbol{r}',t) = \bigtriangledown \cdot [\underline{D} \bigtriangledown P(\boldsymbol{r}|\boldsymbol{r}',t)],$$

where $\underline{D}$ is the Diffusion Tensor, a 3×3 symmetric matrix that describes the diffusion process.

Hahn (1950) in a two-radiofrequency (RF) pulse experiment showed that the self-diffusing coefficients can be measured by using radiofrequency pulses. Stejskal and Tanner (1965) later on proposed the pulsed gradient spin echo sequence for calculating the self-diffusing coefficient; this is the commonly used procedure in Diffusion-Weighted MRI (DW-MRI).

1.4 Within-voxel DW-MRI models

When Diffusion-Weighted MRI data are obtained from the MRI scanner, the brain is split up into units called voxels. Then in each voxel a signal measurement is obtained for each of the Diffusion-Weighted gradient pulses that are applied to the brain. The Diffusion-Weighted gradients pulses which are applied using the pulsed gradient spin echo sequence (Stejskal and Tanner, 1965) are ap-

plied in multiple directions which are defined as the gradient directions $\mathbf{g}$, with corresponding strength b which we call the b-value (Behrens *et al.*, 2003). Thus from the MRI scanner we obtain the signal measurements and the corresponding gradient directions and b-values. In the data that we have used throughout this thesis, we have 61 measures per voxel.

Models that can be used to interpret this data and obtain useful information are available. The most commonly used models will now be described along with the metrics that can be obtained from them. We will then describe how information from these models can be used to implement Tractography.

It is often assumed that the data will contain some noise that needs to be taken into account (Behrens *et al.*, 2003). This noise amongst many other things may be due to thermal noise in the MRI electronic circuitry, electromagnetic interferences, motion of the subject that the data is being obtained from (Pajevic, 2011) and partial volume effects (Parker, 2011). Therefore a model that incorporates this noise must be used. Two of the most common models that are used to estimate the fibre orientation of each voxel are the Diffusion Tensor (DT) model (Pierpaoli and Jezzard, 1996) and the partial volume model (Behrens *et al.*, 2003). There are also some more models that can be used in place of the two models as mentioned in Jbabdi and Johansen-Berg (2011). We will use the DT and partial volume models throughout this thesis as they have been shown to be effective in modelling fibre orientations. We will focus on the partial volume model because it allows us to model multiple fibre orientations.

Both of these models model diffusion within a voxel. The observed Diffusion-Weighted signal is denoted by y_i. Furthermore it is assumed that the observed

signal y_i is a scalar and comes from a Normal distribution such that

$$y_i \sim N(\mu_i, \sigma^2), \qquad i = 1, ..., m,$$

where μ_i is the true Diffusion-Weighted signal measurement, σ is the standard deviation and m is the number of diffusion gradients.

1.4.1 Diffusion Tensor Model

The Diffusion Tensor model (Pierpaoli and Jezzard, 1996) assumes that the diffusion shape in a voxel may be modelled by a 3D Gaussian distribution with variance-covariance proportional to the diffusion tensor, **D** where

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

If the displacement of molecules within a medium can be modelled by a Gaussian distribution, then the diffusivity can be represented by the apparent diffusion coefficient (ADC) (Basser, P.J. and Özarslan, E. 2011). Tanner (1977) then proposed the following formula that relates the ADC to the measured signal in nuclear magnetic resonance (NMR),

$$\ln\left(\frac{A(b)}{A(b=0)}\right) = -b\text{ADC} \tag{1.4.1}$$

where $A(b)$ is the echo magnitude of the diffusion-weighted signal, A(b=0) is the echo magnitude of the non-diffusion weighted signal and b is the b-factor (Basser, P.J. and Özarslan, E. 2011). In grey matter the ADC is sufficient because diffusion does not depend on orientation. However in white matter where dif-

fusion is anisotropic, a three dimensional Gaussian model is used to represent the molecular displacement (Basser, P.J. and Özarslan, E. 2011). In this setting Equation 1.4.1 is generalised to

$$\ln\left(\frac{A(\mathbf{b})}{A(\mathbf{b}=\mathbf{0})}\right) = -(b_{xx}D_{xx} + 2b_{xy}D_{xy} + 2b_{xz}D_{xz} + b_{yy}D_{yy} + 2b_{yz}D_{yz} + b_{zz}D_{zz}) \tag{1.4.2}$$

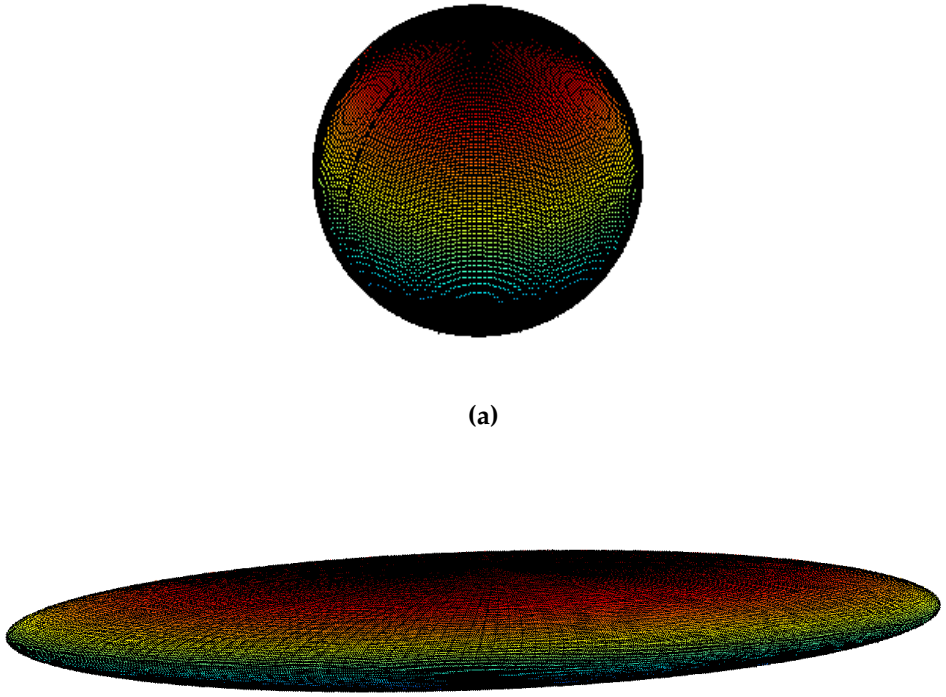
where $A(\mathbf{b})$ and $A(\mathbf{b}=\mathbf{0})$ are the echo magnitudes of the diffusion-weighted and non-diffusion weighted signals and b_{ij} is an element of a matrix $\mathbf{b}$. From Equation 1.4.2 by setting $b_{xx} = b_i g_{1i}^2$, $b_{yy} = b_i g_{2i}^2$, $b_{zz} = b_i g_{3i}^2$, $b_{xy} = b_i g_{1i} g_{2i}$, $b_{xz} = b_i g_{1i} g_{3i}$ and $b_{yz} = b_i g_{2i} g_{3i}$ we can derive the Diffusion Tensor model such that the ith predicted Diffusion-Weighted signal μ_i is

$$\mu_i = S_0 \exp(-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i), \qquad i = 1, ..., m,$$

where $\mathbf{g}_i$ is the ith gradient direction with a b-value b_i, which represents the diffusion sensitivity and is positive or zero; S_0 is the baseline signal, i.e. the signal with no diffusion gradients applied.

$\mathbf{D}$ may be diagonalised to obtain $\mathbf{D} = \mathbf{V}\mathbf{A}\mathbf{V}^T$, where $\mathbf{V}$ is a matrix of the eigenvectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ and $\mathbf{A}$ is a diagonal matrix of the corresponding eigenvalues a_1, a_2 and a_3. From the eigenvalues and eigenvectors of $\mathbf{D}$ we can obtain the diffusion ellipsoid which is a good representation of the shape of the diffusion when the model is a good fit. Such an ellipsoid has 3 axes which represent the eigenvectors obtained by diagonalising the Diffusion Tensor, each has a magnitude of $\sqrt{2\tau a_i}$, where τ is the observation time.

If $a_1 \approx a_2 \approx a_3$ then the diffusion ellipsoid will look like Figure 1.2 (a) and diffusion will be isotropic, indicating that there is no white matter in that voxel. If $a_1 \gg a_2, a_3$ then the diffusion ellipsoid will look like Figure 1.2 (b); this is an



(a)

(b)

Figure 1.2: (a) The estimated diffusion ellipsoid when $a_1 \approx a_2 \approx a_3$, i.e. diffusion is isotropic, obtained by using simulated data and (b) the estimated diffusion ellipsoid when $a_1 \gg a_2, a_3$, i.e. diffusion is anisotropic, obtained by using simulated data.

anisotropic ellipsoid, where the eigenvector with the largest eigenvalue represents the fibre orientation of white matter. This principal eigenvector can then be used in Deterministic Tractography (see Section 1.7.1) as the estimated fibre orientation.

1.5 Metrics from the Diffusion Tensor Model

The parameters in the Diffusion Tensor (DT) model can be used to calculate metrics that give information about some properties of the diffusion in a voxel. Therefore if the parameters of the DT model can be estimated, we can then obtain these metrics.

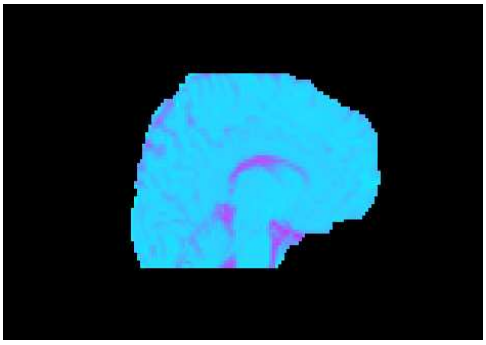


Figure 1.3: The Mean Diffusivity (MD) of the brain data: here the purple areas have a higher value of the MD than the blue areas. Higher values of the MD corresponds to more diffusivity within the voxel.

From the DT model, a measure of diffusivity in a voxel can be calculated by using the eigenvalues. This measure is called the Mean Diffusivity (MD) (Alexander, 2011) and is defined as

$$MD = \frac{D_{xx} + D_{yy} + D_{zz}}{3} = \frac{a_1 + a_2 + a_3}{3},$$

where a_1, a_2 and a_3 are the eigenvalues of the Diffusion Tensor and D_{xx}, D_{yy} and D_{zz} are elements of the Diffusion Tensor. A higher value of the MD means there is more diffusivity in that voxel. Figure 1.3 shows the MD values in a brain, using different colours to represent different values of the MD.

A measurement of the anisotropy of diffusion in a voxel is the Fractional Anisotropy (FA) (Alexander, 2011) which can be calculated as

$$FA = \sqrt{\frac{3\sum_{i=1}^{3}(a_i - \bar{a})^2}{2\sum_{i=1}^{3} a_i^2}}$$

where $\bar{a}$ denotes the mean of the eigenvalues. Alternatively we can write the

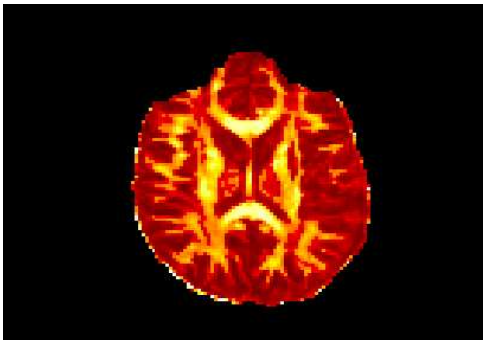


Figure 1.4: The Fractional Anisotropy (FA) of the brain data; lighter areas indicate high anisotropy and thus more white matter.

Fractional Anisotropy as

$$FA = \sqrt{\frac{\frac{1}{2}\sum_{i=1}^{3}(a_i - \bar{a})^2}{\frac{1}{3}\sum_{i=1}^{3} a_i^2}}.$$

Higher values of the FA indicates the regions where white matter is located. Figure 1.4 shows the values of the FA in a brain; a higher value of FA is indicated by a lighter colour. If the diffusion ellipsoid is perfectly anisotropic then a_2=0, a_3=0 and $\bar{a}$=$\frac{a_1}{3}$ and therefore $FA = 1$. For a perfectly isotropic ellipsoid a_1=a_2=a_3=$\bar{a}$ and $FA = 0$.

Although methods for estimating parameters of the Diffusion Tensor model can be fast, and computationally simple (see Section 2.2), in practice there may be voxels that have more than one fibre orientation. For this reason models with multiple fibre orientations were proposed (Behrens *et al.*, 2003).

1.6 Models with multiple fibre orientations

The multiple tensor model (Scherrer and Warfield, 2010) assumes that in each voxel there are N fibre orientations each of which can be modelled by the Diffusion Tensor model (Section 1.4.1). The predicted Diffusion-Weighted signal, μ_i,

is

$$\mu_i = \sum_{j=1}^{N} c_j S_0 \exp(-b_i \boldsymbol{g}_i^T \mathbf{D}_j \boldsymbol{g}_i), \qquad i = 1, ..., m,$$

such that $c_j \in (0, 1]$ and $\sum_{j=1}^{N} c_j = 1$, where c_j are the weights of the individual fibre orientations that also need to be estimated. All other parameters are the same as those defined in the DT model and m is the number of diffusion gradients in a voxel. One of the problems of the multiple tensor model is the non-identifiability of the parameters. Zhou *et al*. (2008) showed an example where $S_0 = b_i = 1$ and $N = 2$ where

$$\mu_i = 0.2 \times \exp\left(-\boldsymbol{g}_i^T \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}\right) \boldsymbol{g}_i\right) + 0.8 \times \exp\left(-\boldsymbol{g}_i^T \left(\begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) \boldsymbol{g}_i\right)$$

can also be written as

$$\mu_i = 0.1 \times \exp\left(-\boldsymbol{g}_i^T \left(\begin{bmatrix} 1 - \log(2) & 0 & 0 \\ 0 & 2 - \log(2) & 0 \\ 0 & 0 & 3 - \log(2) \end{bmatrix}\right) \boldsymbol{g}_i\right)$$

$$+0.9 \times \exp\left(-\boldsymbol{g}_i^T \left(\begin{bmatrix} 4 - \log(\frac{8}{9}) & 0 & 0 \\ 0 & 5 - \log(\frac{8}{9}) & 0 \\ 0 & 0 & 1 - \log(\frac{8}{9}) \end{bmatrix}\right) \boldsymbol{g}_i\right).$$

A solution for the non-identifiability of the parameters was found by Zhou *et al*. (2008) by reparametrising the model.

1.6.1 The partial volume model

The partial volume model is a special case of the multiple tensor model proposed by Behrens *et al*. (2003). It is derived by treating the anisotropic and

isotropic parts of the voxel separately. If the diffusion is isotropic such that the eigenvalues are $a_1 = a_2 = a_3$ then by calculating the Diffusion Tensor $\mathbf{D} = \mathbf{VAV}^T$, we obtain $\mathbf{D} = a_1\boldsymbol{I}_3$ where $\boldsymbol{I}_3$ is the 3x3 identity matrix. Similarly if the diffusion is anisotropic then $a_1 > 0$ while $a_2 = a_3 = 0$ and $\mathbf{D} = \mathbf{VAV}^T = d\mathbf{RAR}^T$.

The partial volume model with one fibre orientation has the predicted Diffusion-Weighted signal μ_i, which is

$$\mu_i = S_0((1-f)\exp(-b_i d) + f\exp(-b_i d\mathbf{g}_i^T\mathbf{RAR}^T\mathbf{g}_i)), \qquad i = 1, ..., m, \quad (1.6.1)$$

where S_0, b_i and $\mathbf{g}_i$ are the same as in the Diffusion Tensor model, d is the diffusivity, f is the fraction of the signal contributed by the fibre, with direction (θ, ϕ), $\mathbf{RAR}^T$ is the anisotropic Diffusion Tensor along that fibre direction and

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

$\mathbf{R}$ is defined to be a matrix consisting of a rotation around the y-axis by an angle θ followed by a rotation around the z-axis by an angle ϕ. Thus $\mathbf{R}$=$\mathbf{R}_Z\mathbf{R}_Y$,where $\mathbf{R}_Y$ is the rotation around the y-axis and $\mathbf{R}_Z$ is the corresponding rotation around the z-axis such that

$$\mathbf{R}_Y = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

and

$$\mathbf{R}_Z = \begin{bmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus

$$\mathbf{R} = \mathbf{R}_Z\mathbf{R}_Y = \begin{bmatrix} \cos(\phi)\cos(\theta) & -\sin(\phi) & \cos(\phi)\sin(\theta) \\ \sin(\phi)\cos(\theta) & \cos(\phi) & \sin(\phi)\sin(\theta) \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

and $\mathbf{RAR}^T$ which is in Equation 1.6.1 is

$$\mathbf{RAR}^T = \begin{bmatrix} (\cos(\phi)\cos(\theta))^2 & \cos(\phi)\sin(\phi)\cos(\theta)^2 & -\sin(\theta)\cos(\phi)\cos(\theta) \\ \sin(\phi)\cos(\phi)\cos(\theta)^2 & (\sin(\phi)\cos(\theta))^2 & -\sin(\theta)\sin(\phi)\cos(\theta) \\ -\sin(\theta)\cos(\phi)\cos(\theta) & -\cos(\theta)\sin(\theta)\sin(\phi) & \sin(\theta)^2 \end{bmatrix};$$

this last matrix can be alternatively written as

$$\mathbf{RAR}^T = \boldsymbol{v}\boldsymbol{v}^T,$$

where

$$\boldsymbol{v} = \begin{bmatrix} \cos(\phi)\cos(\theta) \\ \sin(\phi)\cos(\theta) \\ -\sin(\theta) \end{bmatrix}.$$

By introducing $\theta'=\theta+\frac{\pi}{2}$, we know that cos($\theta$)=sin($\theta'$) because cos($\theta$)=sin($\theta+\frac{\pi}{2}$)

and thus alternatively $\boldsymbol{v}$ is

$$\boldsymbol{v} = \begin{bmatrix} \cos(\phi)\sin(\theta') \\ \sin(\phi)\sin(\theta') \\ \cos(\theta') \end{bmatrix}.$$

We work with θ and ϕ rather than $\boldsymbol{v}$ because we can easily obtain uninformative priors for these parameters when working in a Bayesian framework.

Thus Equation (1.6.1) can now be rewritten as

$$\mu_i = S_0 \left((1-f)\exp(-b_i d) + f\exp(-b_i d(\boldsymbol{g}_i^T \boldsymbol{v})^2) \right), \qquad i = 1, ..., m,$$

where there are m gradient directions. We derive this by noting that

$$\boldsymbol{g}_i^T \mathbf{R}\mathbf{A}\mathbf{R}^T \boldsymbol{g}_i = \boldsymbol{g}_i^T \boldsymbol{v}\boldsymbol{v}^T \boldsymbol{g}_i = (\boldsymbol{g}_i^T \boldsymbol{v})^2.$$

The observed Diffusion-Weighted signal values of the ith acquisition are denoted by y_i, which are assumed to come from a Normal distribution with mean μ_i and standard deviation σ such that $y_i \sim \mathrm{N}(\mu_i, \sigma^2)$. The values of b_i and $\boldsymbol{g}_i$ are known in advance, whilst the values of θ, ϕ, f, d and S_0 can be estimated.

When there are N fibre orientations within a voxel the predicted Diffusion-Weighted signal μ_i is

$$\mu_i = S_0 \left(\left(1 - \sum_{j=1}^{N} f_j \right) \exp(-b_i d) + \sum_{j=1}^{N} f_j \exp(-b_i d(\boldsymbol{g}_i^T \boldsymbol{v}_j)^2) \right), \qquad i = 1, ..., m,$$

such that $\sum_{j=1}^{N} f_j < 1$, where S_0, d, b_i, $\mathbf{g}_i$ were defined earlier, f_j is the fraction of

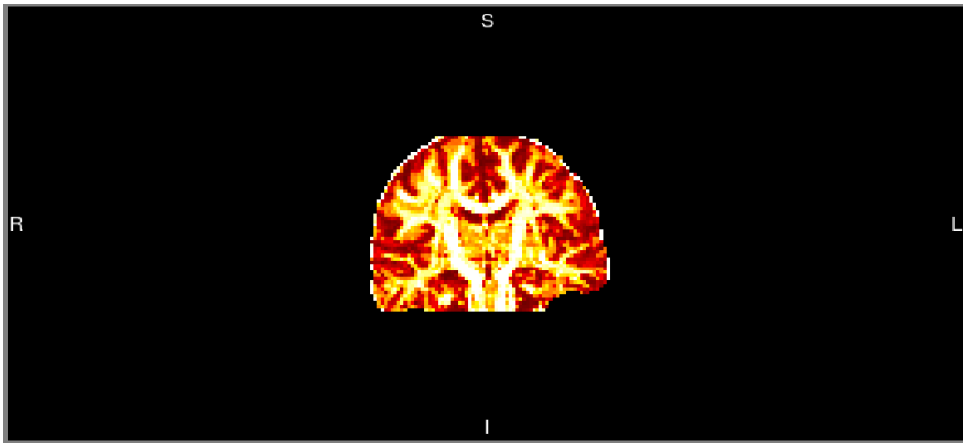


Figure 1.5: The estimated value of the parameter f in the partial volume model on real data, where a darker colour corresponds to a smaller value of f and hence less white matter.

the signal contributed by the jth fibre, with orientation (θ_j, ϕ_j) and

$$\boldsymbol{v}_j = \begin{bmatrix} \cos(\phi_j)\sin(\theta_j) \\ \sin(\phi_j)\sin(\theta_j) \\ \cos(\theta_j) \end{bmatrix}.$$

Figure 1.5 shows the estimated values of the parameter f in a brain. Methods for obtaining estimates of f will be described later. From this figure we can see that the estimated white matter tracts look a bit more detailed than the corresponding tracts when using the FA (Figure 1.4).

1.7 Tractography methods

Tractography uses the fibre orientations that are estimated from the observed data in each voxel to reconstruct tracts in the brain. By doing this it can be seen whether two brain regions, say A and B, are likely to be connected to each other or not. Currently two Tractography methods are commonly used,

namely a Probabilistic Tractography and a Deterministic Tractography (Behrens *et al.*, 2003). These two methods are both local approaches, i.e. they start from a starting point which we call the seed and produce a tract. The tracts end when either the change in fibre orientation from one voxel to another is too great or when a measure of the white matter within the current voxel is too low (Behrens *et al.*, 2003). One measure of the white matter that we could use is the Fractional Anisotropy in the DT model or the value of the f parameter in the partial volume model.

1.7.1 Deterministic Tractography

Deterministic Tractography uses the estimated fibre orientations for each voxel to trace a path from the seed. This method is described in Algorithm 1. Figure 1.6 shows the principal eigenvectors of voxels in real brain data. These eigenvectors could be used within Deterministic Tractography as the fibre orientation within a voxel. In this method the uncertainty in parameter estimates is not taken into account and thus each voxel only has one direction. Each voxel can either be connected or not connected to the seed.

Algorithm 1 Deterministic Tractography

1: Start from a point in a voxel which we define to be the seed.
2: In this voxel use the estimated fibre orientation and follow it to continue the tract until we enter a new voxel.
3: In the next voxel follow the estimated fibre orientation to continue the tract.
4: Go back to Step 3 until stopping conditions are met, such as the Fractional Anisotropy being too small.

Rather than using only the estimated orientation, i.e. a best guess, we want to introduce uncertainty when implementing Tractography. Probabilistic Tractography (Behrens *et al.*, 2003) was introduced to assign a confidence measure to the reconstructed path.

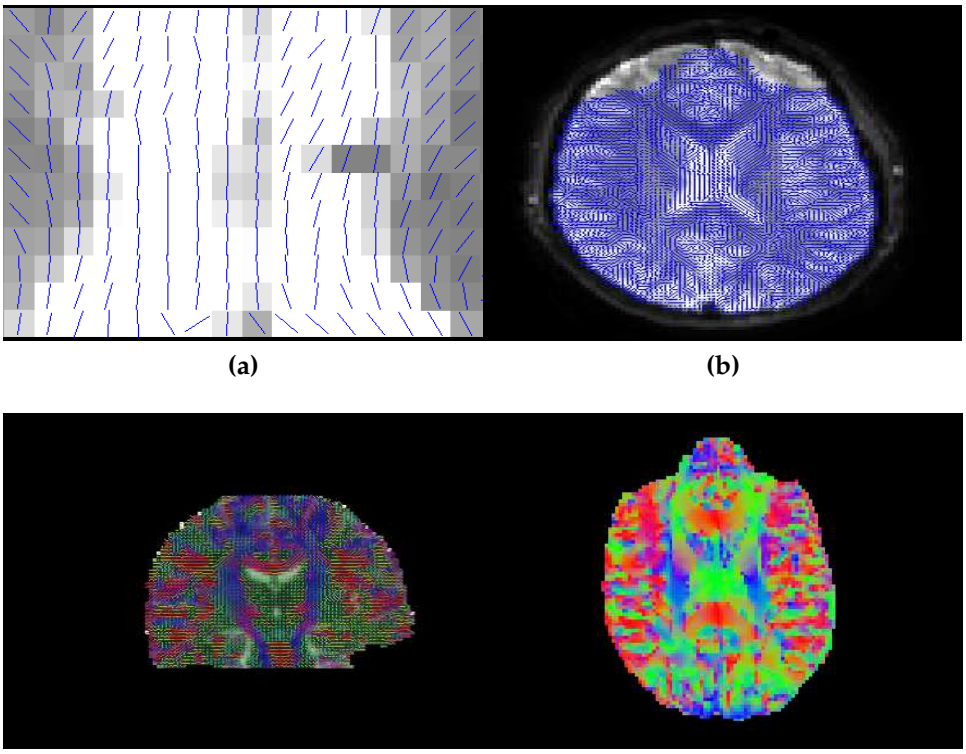


Figure 1.6: The principal eigenvector estimates in each voxel from the DT model to represent the fibre orientation (a) in 192 voxels of real data, (b) in one slice of real brain data and (c) in one slice of real brain data with the corresponding fibre direction represented by the colour. (d) The direction of the estimated fibre orientations represented by the colour as follows red=left-right, green=anterior-posterior, blue=feet-head.

1.7.2 Probabilistic Tractography

In Probabilistic Tractography (Behrens *et al.*, 2003) we may infer the fibre orientations by using Bayesian inference. For each voxel rather than having one direction for a fibre, there is a sample of most probable orientations, this way there is uncertainty in the fibre orientations. Probabilistic Tractography is summarised in Algorithm 2. Figure 1.7 shows the tracts in the brain from one voxel, using the Probabilistic Tractography algorithm in FSL (Woolrich *et al.*, 2009)[1].

Algorithm 2 Probabilistic Tractography

1: Start from a point in a voxel which we define to be the seed.
2: In this voxel choose one of the estimated fibre orientations with equal probability and follow it to continue the tract until we enter a new voxel.
3: In the next voxel choose one of the estimated fibre orientations with equal probability and follow the estimated fibre orientation to continue the tract.
4: Go back to Step 3 until stopping conditions are met, such as the Fractional Anisotropy being too small.
5: Go back to Step 1 until we have N tracts.
6: Calculate the probabilistic index of connectivity to any point in the region by counting how many tracts pass through the point which we denote M and calculating $p = \frac{M}{N}$.

Although Probabilistic Tractography is promising, uncertainty due to noise or partial volume effects within the image in small local regions can cause the pathways to deflect, and some known pathways are not reconstructed using this method (Parker, 2011). To overcome the issues with Probabilistic Tractography, a framework for Global Tractography was proposed (Jbabdi *et al.*, 2007).

[1]FSL is free software that is available from FMRIB, University of Oxford at http://www.fmrib.ox.ac.uk/fsl/. Within Diffusion MRI FSL can estimate the parameters of both the DT and partial volume models and it can implement Probabilistic Tractography.

Figure 1.7: A 2D slice of the brain showing the tracts (black) that are produced when implementing Probabilistic Tractography in 3D in FSL. The seed that we start from is the voxel that is shown in white.

1.7.3 Global Tractography

There is currently not an established method to statistically test for the existence of a connection between two brain regions. A framework for Global Tractography was proposed by Jbabdi *et al*. (2007) as a potential solution. In basic terms Global Tractography parametrises the connections between two brain regions at a global level. Its advantages include being able to acknowledge any known connections by putting this information into the algorithm and it also reduces the sensitivity to local noise by inferring on connections by introducing new parameters. We introduce novel methods for estimating the parameters within this framework in Chapter 4.

For every pair of brain regions Tractography can be done twice, once when a connection exists between the two regions and once when the connection is absent. Then it can be tested to see whether the data supports the connection or not.

The current framework for Global Tractography assumes that there are $\mathcal{N}$ regions of interest in the brain. Each of the $\mathcal{N}$ regions contains a number of voxels. Each of these voxels can be modelled using one of the most commonly used models such as the partial volume model (see Section 1.6.1). Global parameters are introduced, alongside the local parameters that we already have from the partial volume model. A connection matrix $\boldsymbol{\mathcal{C}}$ of size $\mathcal{N} \times \mathcal{N}$ is introduced. If brain Region i is connected to brain Region j then the $(i,j)^{th}$ element of $\boldsymbol{\mathcal{C}}$ is 1, otherwise it is 0. Methods for defining brain regions are discussed in Section 1.10.

We will be doing inference in a Bayesian setting so we denote by $\boldsymbol{\mathcal{F}}$ a random variable that represents the pathways connecting the regions and let $\boldsymbol{\mathcal{L}}$ be a random variable that represents the extremities of the pathways. There is an infinite number of paths that $\boldsymbol{\mathcal{F}}$ could take in the 3D space; one way to overcome the problem of modelling $\boldsymbol{\mathcal{F}}$ is by using splines. We follow Jbabdi *et al*. (2007) and choose to model $\boldsymbol{\mathcal{F}}$ using the Catmull-Rom splines (Farin, 1996) with k control points. The Catmull-Rom splines are described in Appendix A. We choose this type of spline because it assumes that the spline passes through the control points (denoted by $\boldsymbol{\mathcal{K}}$), which makes them easier to work with. The values of the control points can then be inferred. Other options of splines/models are possible but do not affect the method such as Gaussian Processes (Mackay, 1998) and Bayesian P-splines (Lang, 2001). Figure 1.8 shows the global parameters which model connections between different regions of the brain, while Figure 1.9 shows the hierarchical model where the local parameters, which are the parameters that model the fibre orientation in each voxel, are in the box and generate the data $\boldsymbol{Y}$.

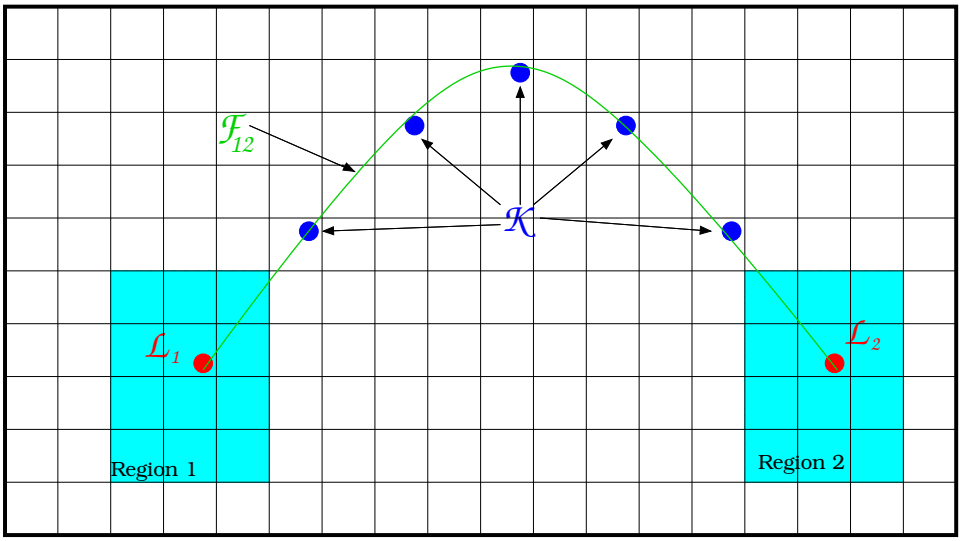


Figure 1.8: A diagram that demonstrates the global parameters from the Global Tractography framework when there is a connection between two brain regions, i.e. $c_{12} = 1$, in a 2D example. The two regions, Region 1 and Region 2, are the set of voxels in the blue areas. The connection between the two regions is the green curve that is modelled by the spline $\boldsymbol{\mathcal{F}}_{12}$. The spline is constructed from the 5 knots $\boldsymbol{\mathcal{K}}$ and the extremities of the spline, $\boldsymbol{\mathcal{L}}_1$ and $\boldsymbol{\mathcal{L}}_2$.

1.8 Bayesian Inference

In this section we will describe the fundamentals of Bayesian Inference which will be used widely throughout this thesis. A more detailed account can be found in Gilks *et al*. (1996). We denote $\boldsymbol{y}$ to be the data and $\boldsymbol{\omega}$ to be the parameters of interest. We can then derive the posterior distribution $\pi(\boldsymbol{\omega}|\boldsymbol{y})$ by combining the likelihood $\pi(\boldsymbol{y}|\boldsymbol{\omega})$ and a prior distribution of the parameters $\pi(\boldsymbol{\omega})$ using Bayes theorem.

$$\pi(\boldsymbol{\omega}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{\omega})\pi(\boldsymbol{y}|\boldsymbol{\omega})}{\int_{\omega} \pi(\boldsymbol{\omega})\pi(\boldsymbol{y}|\boldsymbol{\omega})d\boldsymbol{\omega}}. \tag{1.8.1}$$

One can obtain the value of the posterior distribution for different values of $\boldsymbol{\omega}$, by placing the values of $\boldsymbol{\omega}$ into Equation (1.8.1), however the calculation of the denominator in Equation (1.8.1), which is known as the normalising constant, can be difficult particularly in high dimensions. We will now describe Markov Chain Monte Carlo (MCMC) that allows us to sample from the posterior distri-

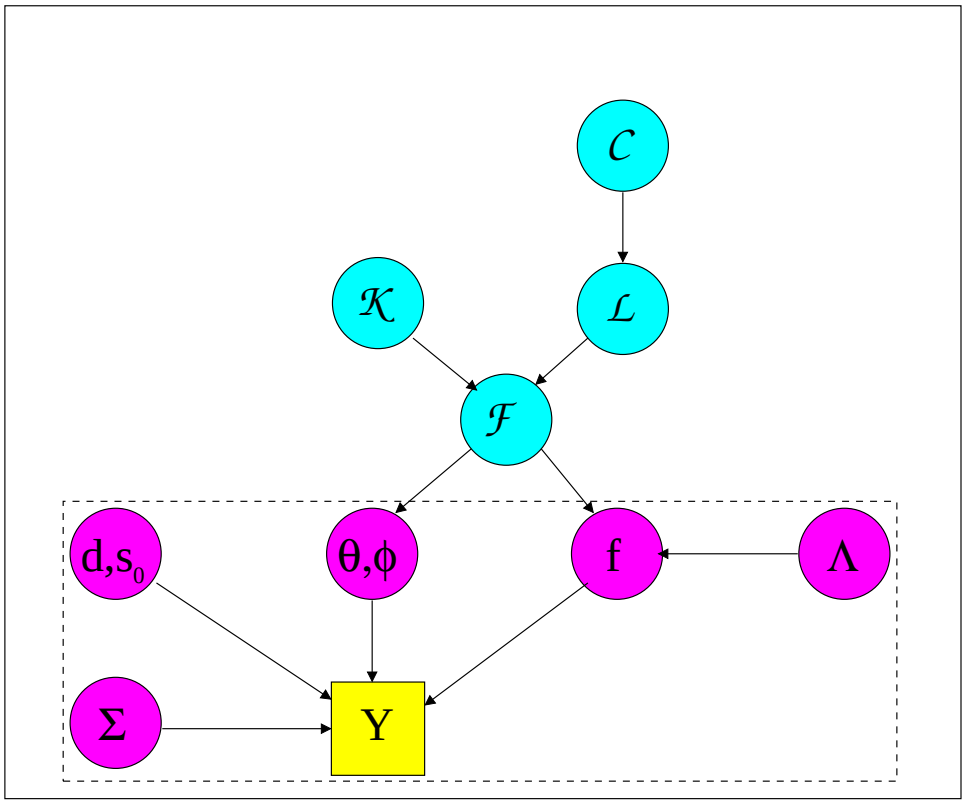


Figure 1.9: The hierarchical model which describes the parameters of the Global Tractography model. The data $\boldsymbol{Y}$ is generated by the local parameters from the partial volume model which are within the dotted box, outside the dotted box are the global parameters. The first global parameter is $\boldsymbol{\mathcal{F}}$ which is the parameter that represent the splines that connect brain regions. $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{K}}$ represent the extremities and the knots of the splines. Finally $\boldsymbol{\mathcal{C}}$ is the connection matrix. The splines $\boldsymbol{\mathcal{F}}$ are determind by the knots and extremities of the splines, whilst the local parameters that represent the fibre orientation within a voxel (ϕ, θ and f) are determined by the spline if it passes through the voxel.

bution whilst avoiding calculating the normalising constant.

1.8.1 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) (See for example Gilks *et al*. (1996)) can be used to simulate random variables from a distribution, $\pi(\cdot)$, which only has to be known up to a normalising constant, which thus avoids the problem of having to calculate complicated normalising constants.

Throughout this thesis MCMC will be used to sample from posterior distributions of the form $\pi(\boldsymbol{\omega}|\boldsymbol{y})$, when estimating parameters, which we denote $\boldsymbol{\omega}$. The two main algorithms used to implement MCMC are the Metropolis-Hastings algorithm (Metropolis *et al*., 1953, Hastings, 1970) and the Gibbs sampler (Geman and Geman, 1984) which are described in Algorithms 3 and 4. The Gibbs sampler, requires us to derive conditional posterior distributions which is not always possible. Therefore throughout this thesis we use the Metropolis-Hastings algorithm.

Algorithm 3 Metropolis-Hastings MCMC

1: Start with an initial value of the parameter that we wish to estimate, $\boldsymbol{\omega}$, and call it $\boldsymbol{\omega}_{cur}$.
2: Propose a new candidate value of $\boldsymbol{\omega}$ which is denoted by $\boldsymbol{\omega}_{can}$ from a given proposal distribution which is denoted by $q(\boldsymbol{\omega}_{cur}, \boldsymbol{\omega}_{can})$.
3: Calculate $\alpha(\boldsymbol{\omega}_{cur}, \boldsymbol{\omega}_{can}) = \min\left(1, \frac{\pi(\boldsymbol{\omega}_{can}|\boldsymbol{y})q(\boldsymbol{\omega}_{can},\boldsymbol{\omega}_{cur})}{\pi(\boldsymbol{\omega}_{cur}|\boldsymbol{y})q(\boldsymbol{\omega}_{cur},\boldsymbol{\omega}_{can})}\right)$.
4: The new value $\boldsymbol{\omega}_{can}$ is either accepted with probability $\alpha(\boldsymbol{\omega}_{cur}, \boldsymbol{\omega}_{can})$ or rejected so that

$$\boldsymbol{\omega}_{cur} = \begin{cases} \boldsymbol{\omega}_{can} & \text{with probability } \alpha(\boldsymbol{\omega}_{cur}, \boldsymbol{\omega}_{can}), \\ \boldsymbol{\omega}_{cur} & \text{with probability } 1 - \alpha(\boldsymbol{\omega}_{cur}, \boldsymbol{\omega}_{can}). \end{cases}$$

5: Start from Step 2 again until we have the required amount of samples.

Algorithm 4 Gibbs sampler MCMC

1: Choose initial values for $\boldsymbol{\omega} = (\omega^1, \omega^2, ..., \omega^N)$ and denote this as $\boldsymbol{\omega}_{cur}$. Then derive the conditional distributions $\pi(\omega^i|\boldsymbol{y}, \boldsymbol{\omega}^{-i})$, $i = 1, ..., N$ where $\boldsymbol{\omega}^{-i}$ is $\boldsymbol{\omega}$ without the i^{th} element.
2: Sample $\omega^1_{cur} \sim \pi(\omega^1|\boldsymbol{y}, \boldsymbol{\omega}^{-1}_{cur})$.
3: For $i = 2, ..., N$ sample $\omega^i_{cur} \sim \pi(\omega^i|\boldsymbol{y}, \boldsymbol{\omega}^{-i}_{cur})$
4: Go back to Step 2 until we obtain the required number of samples.

1.9 Directional distributions

Due to the difficulty in estimating θ and ϕ when $\theta \approx 0$ in the partial volume model which we will discuss in Chapter 2, an alternative reparameterisation is proposed. Instead of estimating θ and ϕ we estimate the fibre orientation $\boldsymbol{v}$, where

$$\boldsymbol{v} = \begin{bmatrix} \cos(\phi)\sin(\theta) & \sin(\phi)\sin(\theta) & \cos(\theta) \end{bmatrix}^T.$$

Although $\boldsymbol{v}$ has three dimensions, when we use this new parameterisation we are still estimating two parameters because $\boldsymbol{v}$ is constrained so that $\boldsymbol{v}^T\boldsymbol{v} = 1$. When using this alternative reparameterisation, parameter estimation in a Bayesian framework using Markov Chain Monte Carlo (MCMC) (Section 1.8.1) will require the use of a good proposal distribution for candidate values of $\boldsymbol{v}$ to be simulated from.

Two distributions that can be used to model antipodal symmetric data are the Bingham distribution (Bingham, 1974) and the Angular Central Gaussian distribution (Tyler, 1987). These two distributions will be used in Chapter 2.

1.9.1 The Bingham distribution

The Bingham distribution (Bingham 1974) can be derived by conditioning a multivariate Normal distribution to lie on the sphere $\boldsymbol{S}_{p-1}$ of unit radius in $\mathbb{R}^p$. We require that $p = 3$ and thus $\boldsymbol{v}$ takes values on the surface of a 3-dimensional sphere $\boldsymbol{S}_2$. $\boldsymbol{S}_2$ is of unit radius and has its centre at the origin. The probability

density function (pdf) of the Bingham distribution is

$$f_{\text{Bing}}(\boldsymbol{x}; \boldsymbol{A}) = {}_1F_1\left(\frac{1}{2}; \frac{p}{2}; \boldsymbol{A}\right)^{-1} \exp(-\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}), \quad \boldsymbol{x}^T \boldsymbol{x} = 1, \quad \boldsymbol{x} \in \mathbb{R}^p$$

such that

$$f_{\text{Bing}}(\boldsymbol{x}; \boldsymbol{A}) = {}_1F_1\left(\frac{1}{2}; \frac{p}{2}; \boldsymbol{A}\right)^{-1} f_{\text{Bing}}{}^*(\boldsymbol{x}; \boldsymbol{A}),$$

where $f_{\text{Bing}}{}^*(\boldsymbol{x}; \boldsymbol{A})$ is an unnormalised density such that

$$f_{\text{Bing}}{}^*(\boldsymbol{x}; \boldsymbol{A}) = \exp(-\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}),$$

${}_1F_1(\frac{1}{2}; \frac{p}{2}; \mathbf{A})$ is the hypergeometric function and $\mathbf{A}$ is a given matrix of size $p \times p$.

In order to use the Bingham distribution as a proposal distribution in the MCMC algorithm (Section 1.8.1), we must be able to simulate from it. There are methods in the literature for simulating from this (Marsaglia, 1972, Hoff, 2007, Kume, 2006 and Ganeiber, 2012). The most efficient of these seems to be the rejection sampling method proposed by Ganeiber (2012). Simulated efficiency rates for the method can be found in Ganeiber (2012).

Rejection sampling (Ripley, 1987) simulates samples from a probability density function $f(\boldsymbol{x}) = c_f f^*(\boldsymbol{x})$, by using a second probability distribution $g^*(\boldsymbol{x})$ that need not be normalised, that we can simulate from. $g^*(\boldsymbol{x})$ must be chosen where there is a constant M^* such that $f^*(\boldsymbol{x}) \leq M^* g^*(\boldsymbol{x})$ for all $\boldsymbol{x}$. Once such a $g^*(\boldsymbol{x})$ is found, we simulate from $f(\boldsymbol{x})$ using the method in Algorithm 5.

For the Bingham distribution, an Angular Central Gaussian distribution is used as the envelope function $g^*(\boldsymbol{x})$ within the rejection sampling algorithm. Ganeiber

Algorithm 5 Simulation from the Bingham distribution

1: Generate a value $\boldsymbol{y}$ from $g^*(\boldsymbol{x})$ and some u from $U(0,1)$.
2: Set $\boldsymbol{x} = \boldsymbol{y}$ if $u \leq \frac{f^*(\boldsymbol{y})}{M^* g^*(\boldsymbol{y})}$; otherwise go back to step 1 until a value is accepted.

(2012), derived the bound

$$M^*(p,b) = \left(\frac{p}{b}\right)^{p/2} \exp\left(-\frac{1}{2}(p-b)\right),$$

such that

$$f^*_{Bing}(\boldsymbol{x};\boldsymbol{A}) \leq M^*(p,b) g^*_{ACG}(\boldsymbol{x};\boldsymbol{\Psi}),$$

where

$$f^*_{Bing}(\boldsymbol{x};\boldsymbol{A}) = \exp(-\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}),$$

$$g^*_{ACG}(\boldsymbol{x};\boldsymbol{\Psi}) = (\boldsymbol{x}^T\boldsymbol{\Psi}^{-1}\boldsymbol{x})^{-p/2},$$

approximately

$$b = (p+2)/2$$

and

$$\Psi^{-1} = \boldsymbol{I}_p + \frac{2}{b}\boldsymbol{A}.$$

1.9.2 The Angular Central Gaussian distribution

The Angular Central Gaussian(ACG) distribution was introduced by Tyler (1987) as an alternative to the Bingham distribution due to the complicated normalising constant of the latter. The ACG distribution works by projecting a multivariate Gaussian distribution with mean zero to lie on the unit sphere $\boldsymbol{S}_{p-1}$. The probability density function (pdf) of the ACG distribution is:

$$g_{\text{ACG}}(\boldsymbol{x};\boldsymbol{\Psi}) = \boldsymbol{w}_p^{-1}|\boldsymbol{\Psi}|^{-1/2}(\boldsymbol{x}^T\boldsymbol{\Psi}^{-1}\boldsymbol{x})^{-p/2},$$

where $\mathbf{\Psi}$ is a given matrix of size $p \times p$.

If $\boldsymbol{x} \sim N_p(\mathbf{0}, \mathbf{\Psi})$ then $\|\boldsymbol{x}\|^{-1}\boldsymbol{x} \sim ACG(\mathbf{\Psi})$ (Mardia and Jupp 2000), thus it is very easy to simulate from the ACG distribution with parameter $\mathbf{\Psi}$, by first simulating $\boldsymbol{y}$ from $N_p(\mathbf{0}, \mathbf{\Psi})$ and then setting $\boldsymbol{x}$=$\|\boldsymbol{y}\|^{-1}\boldsymbol{y}$.

1.10 Defining regions

Within the thesis we will need to be able to define brain regions within real data. Within FSL there is an Atlas feature, that allows a structure of the brain to be selected, and gives the probability of a voxel within the brain being part of that region. There are many different atlases which can be selected and they were created by averaging over brain images.

To use the atlases that are available, first the images that we have must be transformed such that their coordinate space is the same as that for the atlases. This can easily be implemented by using FMRIB's Linear Image Registration Tool (FLIRT) feature in FSL (Jenkinson and Smith, 2001, Jenkinson *et al*., 2002). Once this has been implemented we can then select a structure from an atlas and obtain a new mask that shows on the brain image where the structure is. For each voxel the atlas will give a estimate that represents the probability of that voxel being within the chosen structure.

Once we have the required probability masks for the data we then transform the masks back to the original coordinate space by using the inverse of the matrix that was used to transform the original data to the standard coordinate system. We can then choose a threshold for the probability of a voxel being within a certain region such that we can assume that voxels that are less than this threshold

are not part of the defined region. Finally we can then use a binary tool to get a mask of the brain, such that the voxels take the value 1 if the voxel is above a certain threshold to be part of the structure and 0 otherwise. An example of using the FSL Atlas feature to find a mask of the left primary motor cortex using the Juelic Histological Atlas (Eickhoff *et al.*, 2005) is shown in Figure 1.10.

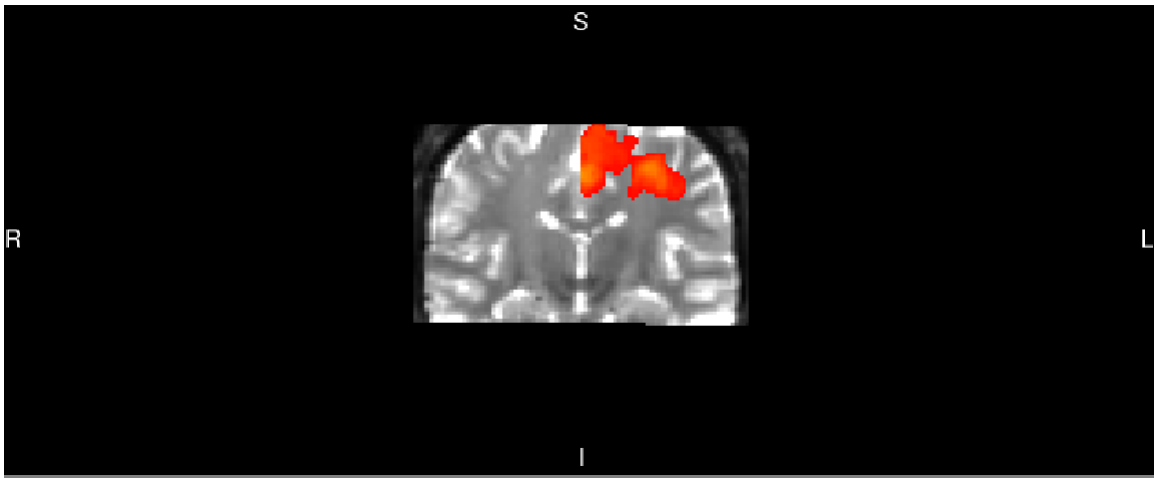
1.11 Thesis outline

The thesis is structured as follows. In Chapter 2 we investigate methods to quickly and efficiently estimate the parameters in the Diffusion Tensor and partial volume models. We then discuss how to reparameterise the partial volume model to allow successful parameter estimation in a certain case when the value of θ is close to 0.

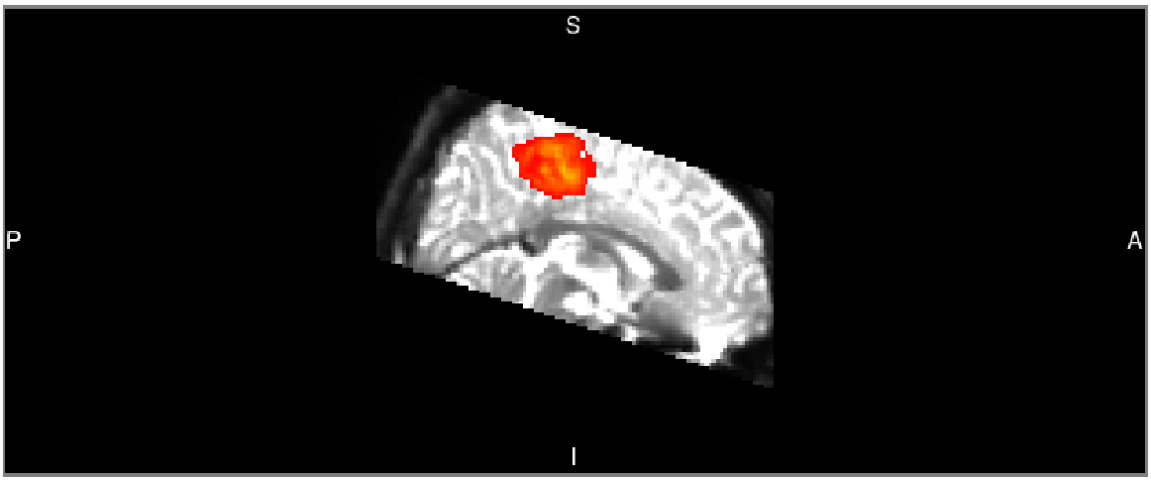
In Chapter 3 we implement model selection to choose between different models for the number of fibres in a voxel. We then use this model selection to propose a Fully Probabilistic Tractography that uses model uncertainty within Tractography.

In Chapter 4 we study Global Tractography and infer the parameters of this model. We then employ efficient model selection methods to test for the existence of a connection between two brain regions.

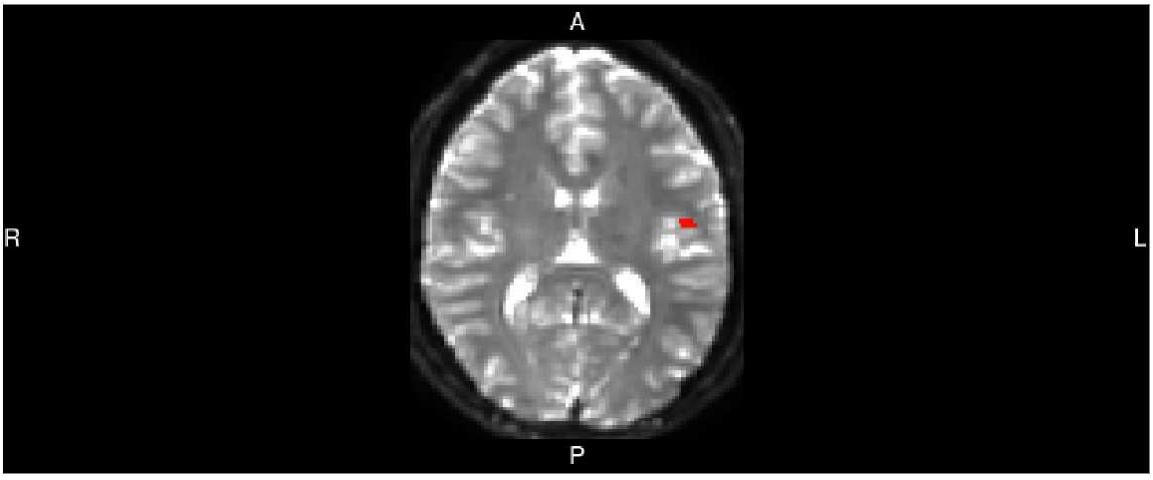
In Chapter 5 we conclude this thesis and then discuss possible future work.



(a)



(b)



(c)

Figure 1.10: The (a) coronal, (b) sagittal and (c) axial views of a brain that show the mask of the left primary motor cortex in orange that is estimated by using the Atlas feature in FSL.

CHAPTER 2

Inference within a voxel

2.1 Motivation

We would like to efficiently infer the local parameters of the models for the observed Diffusion-Weighted signal within a voxel (Section 1.4). Global Tractography (see Section 1.7.3) and Probabilistic Tractography (Section 1.7.2) require us to have estimates of the parameters of the partial volume model (see Section 1.6.1) for voxels, whilst in Deterministic Tractography (Section 1.7.1) we use the parameter estimates of the Diffusion Tensor model (see Section 1.4.1). Therefore efficient inference of the local parameters of these models will benefit these algorithms. One assumption we make is that voxels within the brain region are independent of each other such that the local parameters of the partial volume model are independent of the local parameters within other voxels. Then if we can efficiently do parameter estimation in one voxel, then we can take advantage of parallel computing methods (Hernández *et al.*, 2013) to efficiently apply parameter estimation to all the voxels.

In this chapter we show how the parameters of both the Diffusion Tensor (DT) model and the partial volume model can be estimated. A range of different methods are proposed to infer the parameters. The developed methodology

first is illustrated with simulated datasets and then is applied to a real dataset. Finally a simulation study is carried out to thoroughly assess the methods that are introduced in this chapter.

We adopt a Bayesian approach to inference for both the DT model and the partial volume model. We then develop a series of MCMC algorithms ranging from standard random-walk Metropolis-Hastings to more advanced algorithms such as Adaptive algorithms.

2.2 Inference for the Diffusion Tensor model

The Diffusion Tensor model (see Section 1.4.1) assumes that the diffusion shape in a voxel may be modelled by a 3D Gaussian distribution with variance-covariance proportional to the Diffusion Tensor, $\mathbf{D}$ where

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

In the Diffusion Tensor model, the ith predicted Diffusion-Weighted signal, μ_i is

$$\mu_i = S_0 \exp(-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i) \tag{2.2.1}$$

where $\mathbf{g}_i$ is the ith gradient direction with a b-value b_i, which is the diffusion sensitivity; S_0 is the baseline signal, i.e. the signal with no Diffusion-Weighted gradients applied. We further assume that there is additive independent and identically distributed (i.i.d) Gaussian noise:

$$y_i \sim N(\mu_i, \sigma^2), \qquad i = 1, ..., m,$$

where y_i are the observed Diffusion-Weighted signal values of the ith acquisition and σ is the standard deviation.

Denote the observed data by $\boldsymbol{y}$ and the parameters of the Diffusion Tensor model by $\boldsymbol{\omega}$=(D_{xx},D_{yy},D_{zz},D_{xy},D_{xz},D_{yz},S_0,σ). We will now illustrate how the parameter vector $\boldsymbol{\omega}$ can be estimated using a frequentist approach and a Bayesian approach.

As an illustrative example, a dataset of size 61 is simulated from the Diffusion Tensor model, such that $y_i \sim N(\mu_i, \tau^{-1})$ where $\mu_i = S_0 \exp(-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i)$ with parameters S_0=10.0, precision $\tau = \frac{1}{\sigma^2} = 100$ and

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} = \begin{bmatrix} 0.001 & 0.0002 & 0.0003 \\ 0.0002 & 0.002 & 0.0005 \\ 0.0003 & 0.0005 & 0.003 \end{bmatrix},$$

such that

$$\mathbf{D} = \mathbf{V}\mathbf{U}\mathbf{V}^{\mathbf{T}},$$

where

$$\mathbf{V} = \begin{bmatrix} 0.9848 & 0.0786 & 0.1547 \\ -0.1333 & 0.9133 & 0.3849 \\ -0.1110 & -0.3997 & 0.9099 \end{bmatrix}$$

and

$$\mathbf{U} = \begin{bmatrix} 0.0009 & 0 & 0 \\ 0 & 0.0018 & 0 \\ 0 & 0 & 0.0033 \end{bmatrix}.$$

2.2.1 Linearised DT model

In Deterministic Tractography (Section 1.7.1) it is only required to know the principal eigenvector of the Diffusion Tensor, $\mathbf{D}$ in each voxel, such that we have a point estimate of $\mathbf{D}$. Therefore any uncertainty associated with this estimate is ignored. Hence it is enough to just estimate the maximum likelihood estimate (MLE) for each of the parameters in the Diffusion Tensor model. The model that we wish to maximise is non linear but it can be transformed to a linear model if we assume that S_0 is known (Bates and Watts, 2007).

By rearranging Equation (2.2.1), note that we get the linearised Diffusion Tensor model

$$\log(\mu_i) - \log(S_0) = \mathbf{A}_i \boldsymbol{B}, \qquad i = 1, ..., m.$$

where

$$\mathbf{A}_i = -b_i \begin{pmatrix} \mathbf{g}_i(1)^2 & \mathbf{g}_i(2)^2 & \mathbf{g}_i(3)^2 & 2\mathbf{g}_i(1)\mathbf{g}_i(2) & 2\mathbf{g}_i(1)\mathbf{g}_i(3) & 2\mathbf{g}_i(2)\mathbf{g}_i(3) \end{pmatrix}$$

and

$$\boldsymbol{B} = \begin{pmatrix} D_{xx} \\ D_{yy} \\ D_{zz} \\ D_{xy} \\ D_{xz} \\ D_{yz} \end{pmatrix}.$$

The ordinary least square estimator of $\boldsymbol{B}$, $\widehat{\boldsymbol{B}}$ is $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{Y}$, where

$\mathbf{Y}$=$(y_1, y_2, ..., y_m)$, such that there are m gradient directions, and

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ . \\ . \\ . \\ \mathbf{A}_m \end{pmatrix}.$$

To implement this method we will have to use an approximate value of S_0. S_0 is defined to be the baseline signal, the signal value when no diffusion gradients are applied i.e. the signal when $b_i = 0$. Hence the value of S_0 is estimated to be the diffusion intensity y_i, that corresponds to the b-value that has the value 0.

When this method is implemented on the simulated Diffusion Tensor dataset,

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} 0.001010 & 0.001983 & 0.002970 & 0.000189 & 0.000294 & 0.000460 \end{pmatrix}^T.$$

This is very close to the true value of $\boldsymbol{B}$ which is

$$\boldsymbol{B} = \begin{pmatrix} 0.001 & 0.002 & 0.003 & 0.0002 & 0.0003 & 0.0005 \end{pmatrix}^T.$$

2.2.2 Bayesian inference for the DT model

We now investigate Bayesian inference of the parameters in the Diffusion Tensor model which we denote by $\boldsymbol{\omega}$. A Bayesian approach to estimate $\boldsymbol{\omega}$, would be to generate samples from the posterior distribution $\pi(\boldsymbol{\omega}|\mathbf{y})$. As in the previous section the value of S_0 is defined to be the y_i, that has a corresponding b-value of 0. We will follow the approach of Behrens *et al*. (2003).

The prior distributions of the parameters are

$$\pi(D_{xx}) = \pi(D_{yy}) = \pi(D_{zz}) \sim \Gamma(a_D, b_D)$$

$$\pi(D_{xy}) = \pi(D_{xz}) = \pi(D_{yz}) \propto 1$$

$$\pi(\tau) \sim \Gamma(\alpha_\sigma, \beta_\sigma)$$

where $\tau = \frac{1}{\sigma^2}$. The prior distributions are chosen such that they are uninformative, such that $\alpha_\sigma = 1$ and $\beta_\sigma = 0.001$. We further ensure that the Diffusion Tensor is positive definite by rejecting values of the elements that do not satisfy this condition. The likelihood of the data is

$$\pi(\boldsymbol{y}|\boldsymbol{\omega}) = \prod_{i=1}^{n} \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp\left(-\frac{\tau}{2}(y_i - \mu_i)^2\right)$$

$$= \left(\frac{\tau}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \mu_i)^2\right)$$

where

$$\mu_i = S_0 \exp(-b_i \boldsymbol{g}_i^T \boldsymbol{D} \boldsymbol{g}_i), \qquad i = 1, ..., m.$$

Thus the posterior distribution, $\pi(\boldsymbol{\omega}|\boldsymbol{y})$, is

$$\pi(\boldsymbol{\omega}|\boldsymbol{y}) = \pi(\boldsymbol{y}|\boldsymbol{\omega})\pi(D_{xx})\pi(D_{yy})\pi(D_{zz})\pi(D_{xy})\pi(D_{xz})\pi(D_{yz})\pi(\tau)$$

$$\propto \pi(\boldsymbol{y}|\boldsymbol{\omega})\pi(D_{xx})\pi(D_{yy})\pi(D_{zz})\pi(\tau).$$

The full conditional distributions of $\pi(\boldsymbol{\omega}|y)$ can be derived which are $\pi(D_{xx}|\ \mathbf{y}, \boldsymbol{\omega}_{-D_{xx}})$, $\pi(D_{yy}|\ \mathbf{y}, \boldsymbol{\omega}_{-D_{yy}})$, $\pi(D_{zz}|\ \mathbf{y}, \boldsymbol{\omega}_{-D_{zz}})$, $\pi(D_{xy}|\ \mathbf{y}, \boldsymbol{\omega}_{-D_{xy}})$, $\pi(D_{xz}|\ \mathbf{y}, \boldsymbol{\omega}_{-D_{xz}})$, $\pi(D_{yz}|\ \mathbf{y}, \boldsymbol{\omega}_{-D_{yz}})$ and $\pi(\tau|\ \mathbf{y}, \boldsymbol{\omega}_{-\tau})$, where the notation $\boldsymbol{\omega}_{-a}$ denotes the parameter matrix $\boldsymbol{\omega}$ without parameter a. Then single-component

Metropolis Hastings Markov Chain Monte Carlo (MCMC) (defined in Section 1.8.1) was used to generate samples from the full conditional distributions. We used a random-walk proposal distribution within MCMC, so we had to tune the proposal distribution which required trial and error.

When data were simulated from the Diffusion Tensor model, this algorithm worked very well at generating samples from the posterior distributions that were close to the true values of the parameters and had good mixing as can be seen in Figures 2.1 and 2.2.

We now have a good Bayesian method for obtaining estimates for the parameters in the Diffusion Tensor model. We also have a good frequentist method for a linearised version of the Diffusion Tensor model. If we just want a single estimate for each parameter, then the linearised DT model is ideal, otherwise if we require a distribution of values for a parameter then the Bayesian MCMC method can be implemented. From now on we focus on inference in the more complicated partial volume model (Section 1.6.1).

2.3 MCMC for estimating parameters in the partial volume model

We now focus on estimating the parameters of the partial volume model (Section 1.6.1) because it enables us to have multiple fibre orientations. Initially we will infer the partial volume model with one fibre orientation. As in Section 1.6.1 we assume that the observed data y_i has noise such that

$$y_i \sim N(\mu_i, 1/\tau), \qquad i = 1, ..., m$$

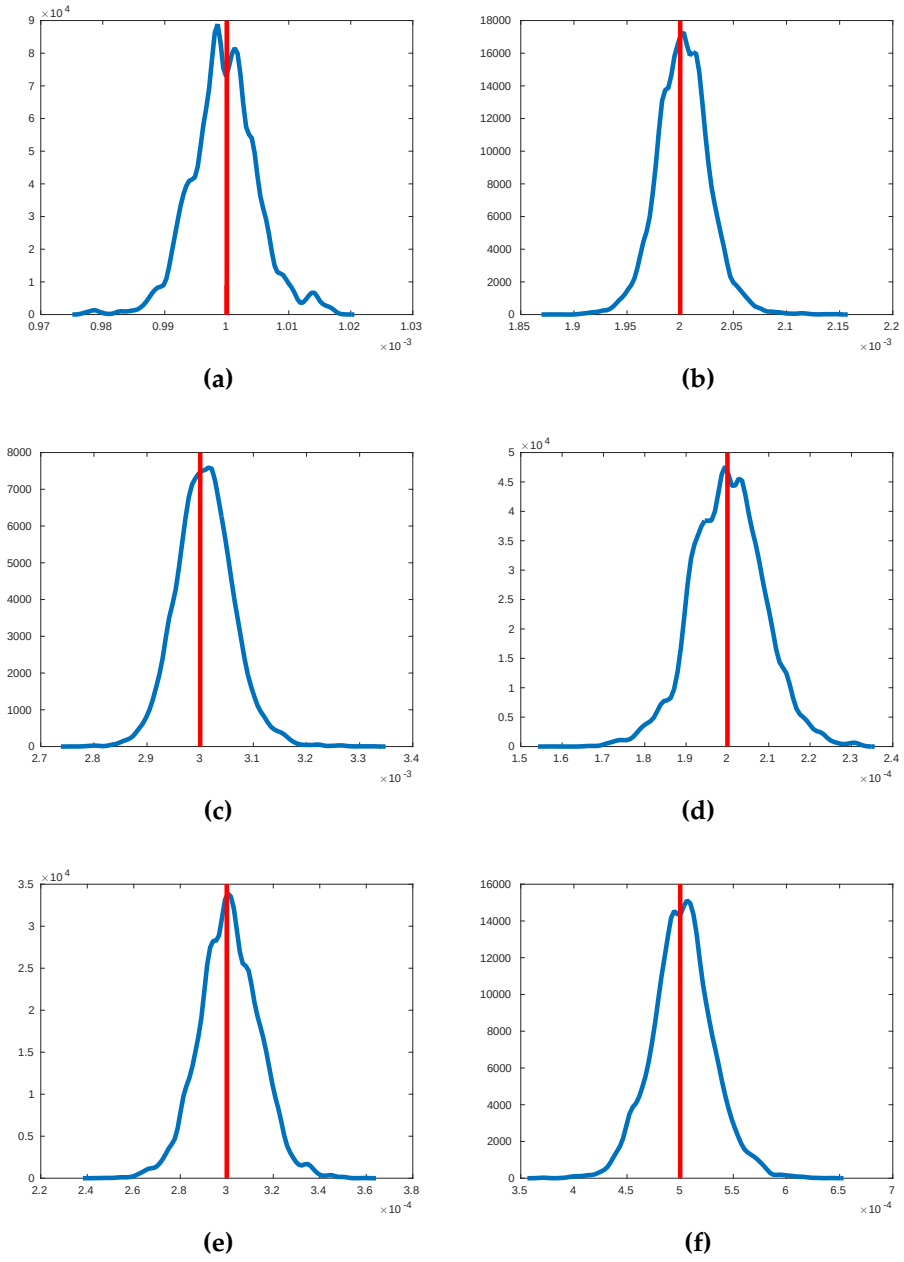


Figure 2.1: Kernel density plots of the estimated parameters of the DT model by simulating from the posterior distribution using MCMC. (a) D_{xx}, (b) D_{yy}, (c) D_{zz}, (d) D_{xy}, (e) D_{xz} and (e) D_{yz}; the red line denotes the true value of the parameter. The algorithm that was used was the random-walk Metropolis-Hastings MCMC.

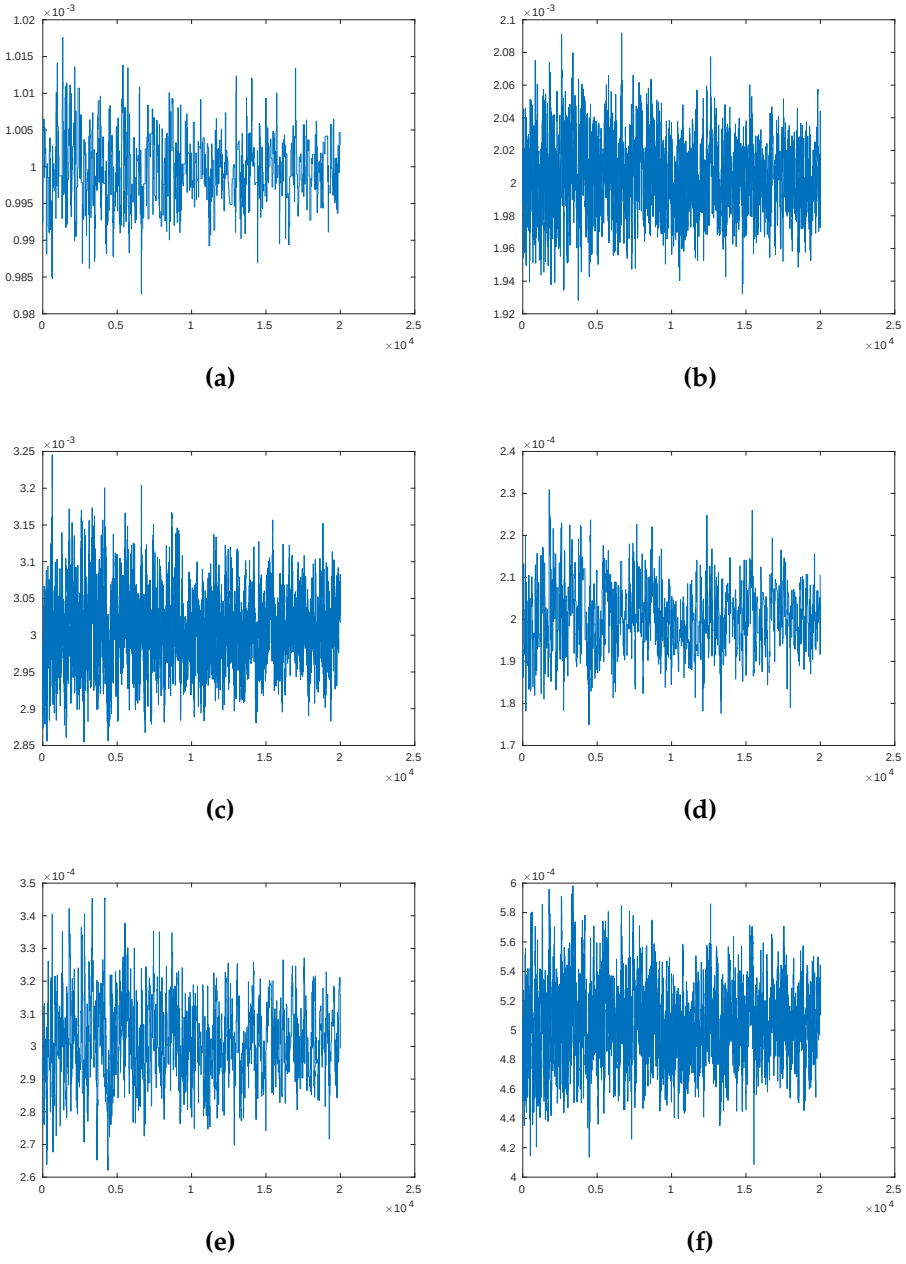


Figure 2.2: Traceplots of the estimated parameters of the DT model by simulating from the posterior distribution using MCMC. (a) D_{xx}, (b) D_{yy}, (c) D_{zz}, (d) D_{xy}, (e) D_{xz} and (f) D_{yz} when using the random-walk Metropolis Hastings MCMC

where $\tau = \frac{1}{\sigma^2}$ is the precision and

$$\mu_i = S_0((1-f)\exp(-b_i d) + f\exp(-b_i d(\mathbf{g}_i^T \boldsymbol{v})^2)), \qquad i = 1, ..., m.$$

We simulated a dataset for one voxel using the partial volume model where the parameter values are θ=0.5, ϕ=1.0, f=0.7, d=0.0015, S_0=1 and $\tau = 100$. These values were chosen as they are typical values in the partial volume model. This dataset has 61 values because there are 61 gradient directions which have 61 corresponding b-values. We used values of the gradient directions and b-values that came from a real dataset. This dataset will be used throughout this chapter.

The methods that we use throughout this chapter can easily be extended to infer on the partial volume model with more than one fibre orientation as shown in Section 2.7.2. The parameters of the partial volume model with one fibre orientation are denoted by $\boldsymbol{\omega}$=(θ,ϕ,f,d,S_0,τ). By using a Bayesian framework and suitable priors on each parameter, MCMC can be used to generate samples from the posterior distribution $\pi(\theta,\phi,f,d,S_0,\tau|\mathbf{y})$.

We follow Behrens (2003) and adopt the following priors because they are all non-informative except for where positivity is required.

$$\pi(\theta, \phi) \sim |\sin(\theta)|$$

$$\pi(f) \sim U(0, 1)$$

$$\pi(d) \sim U(0, \infty)$$

$$\pi(S_0) \sim U(0, \infty)$$

$$\pi(\tau) \sim \Gamma(\alpha_\sigma, \beta_\sigma).$$

The likelihood of the data is

$$\pi(\boldsymbol{y}|\boldsymbol{\omega}) = \prod_{i=1}^{n} \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp\left(-\frac{\tau}{2}(y_i - \mu_i)^2\right)$$

$$= \left(\frac{\sqrt{\tau}}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \mu_i)^2\right)$$

where

$$\mu_i = S_0\left((1-f)\exp(-b_i d) + f\exp(-b_i d(\boldsymbol{g}_i^T \boldsymbol{v})^2)\right).$$

The posterior distribution is then

$$\pi(\boldsymbol{\omega}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\omega})\pi(\theta,\phi)\pi(f)\pi(d)\pi(S_0)\pi(\tau)$$

$$= \left(\frac{\sqrt{\tau}}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \mu_i)^2\right) |\sin(\theta)| \frac{\beta_\sigma^{\alpha_\sigma}}{\Gamma(\alpha_\sigma)} (\tau)^{\alpha_\sigma - 1} \exp(-\beta_\sigma \tau).$$

The precision τ is integrated out of the posterior distribution as it is not a parameter of interest. The posterior distribution without τ is obtained by

$$\pi(\theta,\phi,f,d,S_0|\boldsymbol{y}) = \int \pi(\boldsymbol{\omega}|\boldsymbol{y}) d\tau$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n |\sin(\theta)| \frac{\beta_\sigma^{\alpha_\sigma}}{\Gamma(\alpha_\sigma)} \int (\tau)^{\frac{n}{2}+\alpha_\sigma - 1} \exp\left(-\tau\left(\frac{S(\boldsymbol{y},\boldsymbol{\mu})}{2} + \beta_\sigma\right)\right) d\tau$$

where $S(\boldsymbol{y},\boldsymbol{\mu}) = \sum_{i=1}^{n}(y_i - \mu_i)^2$.

By comparing the integrand to the probability density function of the Gamma distribution, the posterior distribution can then be derived as

$$\pi(\theta,\phi,f,d,S_0|\boldsymbol{y}) = \frac{\beta_\sigma^{\alpha_\sigma}}{\Gamma(\alpha_\sigma)} \left(\frac{1}{\sqrt{2\pi}}\right)^n |\sin(\theta)| \frac{\Gamma(\frac{n}{2}+\alpha_\sigma)}{\left(\frac{S(\boldsymbol{y},\boldsymbol{\mu})}{2} + \beta_\sigma\right)^{\frac{n}{2}+\alpha_\sigma}}.$$

We will now describe the MCMC algorithms to draw samples from $\pi(\widetilde{\boldsymbol{\omega}}|\boldsymbol{y})$

where $\widetilde{\boldsymbol{\omega}}$ does not contain τ such that $\widetilde{\boldsymbol{\omega}} = (\theta, \phi, f, d, S_0)$.

2.3.1 Vanilla MCMC

Initially we attempt to use the simplest MCMC algorithm, i.e. single-component Metropolis-Hastings (see Section 1.8.1) with a normal proposal distribution for each of the parameters with the mean being the current parameter value and some variance σ^2. This algorithm is called Vanilla MCMC and is described in Algorithm 6.

Algorithm 6 Vanilla MCMC

1: Start with initial values for θ, ϕ, f, d and S_0.
2: Use the Metropolis-Hastings random-walk algorithm to first propose and then accept or reject a new sample from $\pi(\theta|\phi, f, d, S_0, \boldsymbol{y})$. We then similarly propose and reject or accept values from $\pi(\phi|\theta, f, d, S_0, \boldsymbol{y})$, $\pi(f|\theta, \phi, d, S_0, \boldsymbol{y})$, $\pi(d|\theta, \phi, f, S_0, \boldsymbol{y})$ and $\pi(S_0|\theta, \phi, f, d, \boldsymbol{y})$. If values are proposed that are outside the permitted values for any of the parameters then we reject the sample.
3: Repeat Step 2 until the required number of samples are obtained.

We then implemented this algorithm on the simulated partial volume dataset. Figure 2.3 show the kernel density plots and trace plots for Vanilla MCMC on the parameters with τ integrated out. The mixing appears to be very good and the results correspond to the true parameter values.

Vanilla MCMC produces results that are accurate, however when there are a lot of voxels, we would prefer to update all the parameters as a block rather than update one at a time so that the algorithm is quicker. Therefore we will now try Block-update MCMC.

2.3.2 Block-update MCMC

Block-update MCMC updates all of the parameters together rather than updating each parameter one at a time as in Vanilla MCMC (Section 2.3.1). In this

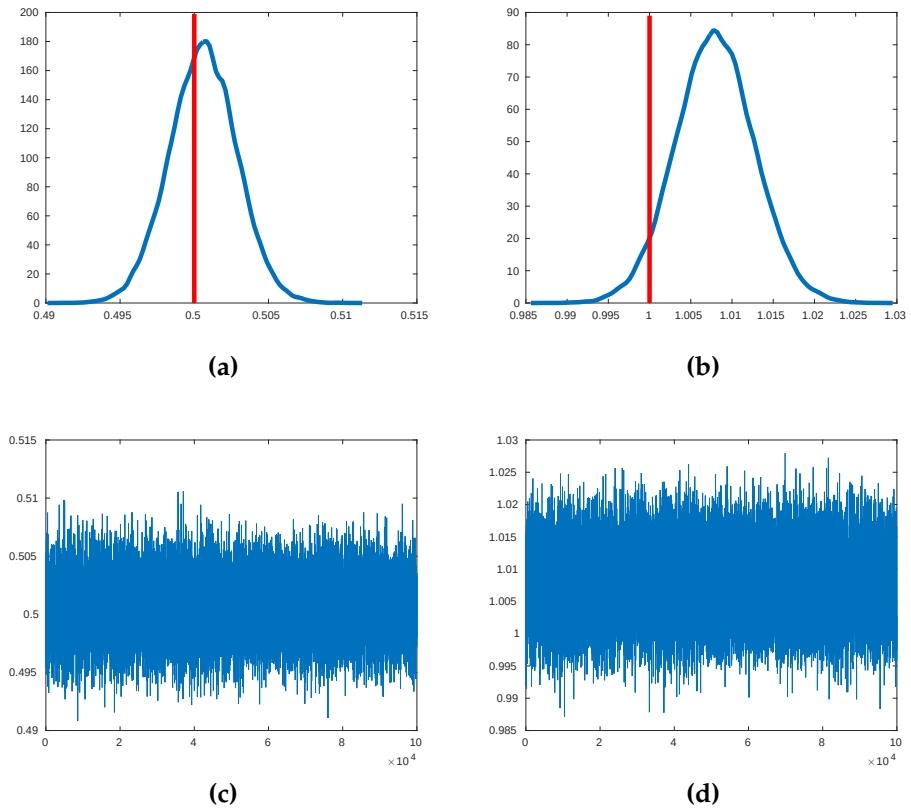


Figure 2.3: Kernel density plots for the parameter estimates of the partial volume model simulated from the posterior distribution using the Vanilla MCMC algorithm. (a) θ and (b) ϕ, where the true value of the parameters is shown by the red line. The traceplots for (c) θ and (d) ϕ.

algorithm a multivariate normal proposal distribution is used with the mean being a vector of the current parameter values. The algorithm is summarised in Algorithm 7.

Algorithm 7 Block-update MCMC

1: Start with initial values for θ, ϕ, f, d and S_0.
2: Use the Metropolis-Hastings algorithm, to propose values of $(\theta, \phi, f, d, S_0)$ from a random-walk normal proposal distribution and then accept or reject values from $\pi(\theta, \phi, f, d, S_0|\boldsymbol{y})$
3: Repeat Step 2 until the required amount of samples are obtained.

The Block-update MCMC algorithm is much faster than the Vanilla MCMC algorithm but deciding on a suitable covariance matrix for the proposal distribution is not easy. Therefore we consider the Adaptive MCMC algorithm (Section 2.3.3).

2.3.3 Adaptive MCMC

Adaptive MCMC (Haario *et al.*, 2001) automatically calculates the covariance matrix for the proposal distribution in MCMC at each iteration and will accept about 23.4% of the candidate values (Rosenthal, 2010). This is ideal as it gives an automatic way to find the covariance matrix rather than having to tune. Rosenthal (2010) discusses solutions to cases where Adaptive MCMC may not work as well.

Adaptive MCMC works by sampling the candidate value of the parameters at the nth iteration, $\boldsymbol{\omega}_{n+1}$ as

$$\boldsymbol{\omega}_{n+1} \sim N(\boldsymbol{\omega}_n, [(2.38)^2/d]\boldsymbol{\Sigma}_n)$$

where $\boldsymbol{\omega}_n$ is the current parameter value, d is the number of variables in $\boldsymbol{\omega}_n$ and

$\boldsymbol{\Sigma}_n$ is the empirical covariance matrix of $\boldsymbol{\omega}_0$, $\boldsymbol{\omega}_1$,...,$\boldsymbol{\omega}_n$. The value of 2.38 within the algorithm and acceptance rate of 23.4% are asymptotic as $d \to \infty$, however they have been shown in numerical studies to be good approximations when d is as small as five (Rosenthal, 2010).

We would like an algorithm which is as efficient as possible, therefore we will calculate the empirical covariance matrix in the following way. If we define the output $\boldsymbol{\omega}_n$=$(\omega_{n1}, \omega_{n2}, ..., \omega_{nd})$, then the (j,k)th element of the sample covariance matrix is

$$q_{jk} = \frac{1}{N-1} \sum_{i=1}^{N} (\omega_{ij} - \bar{\omega_j})(\omega_{ik} - \bar{\omega_k}).$$

Since $\boldsymbol{\Sigma}_n$ needs to be evaluated at each iteration, to increase efficiency we evaluate q_{jk} recursively as follows:

$$q_{jk} = \frac{1}{N-1} \left(\sum_{i=1}^{N} \omega_{ij}\omega_{ik} - \frac{1}{N} \sum_{i=1}^{N} \omega_{ij} \sum_{i=1}^{N} \omega_{ik} \right). \tag{2.3.1}$$

We will now find transformations for the parameters in the partial volume model so that there are no constraints on the parameters when we use Adaptive MCMC. The parameters in the partial volume model have the following constraints $0 < \theta < \pi$, $0 < \phi < 2\pi$, $0 < f < 1$, $d > 0$ and $S_0 > 0$. By transforming the variables it can be ensured that they can take any value in the real line to ease optimisation.

The following transformations were performed

$$\theta' = \log\left(\frac{\theta}{\pi - \theta}\right)$$

$$\phi' = \log\left(\frac{\phi}{2\pi - \phi}\right)$$

$$f' = \log\left(\frac{f}{1-f}\right)$$

$$d' = \log(d)$$

$$S_0' = \log(S_0).$$

The posterior distribution was then changed to include the transformed variables. The determinant of the Jacobian matrix was included in the posterior distribution because we're using transformations and it is the following.

$$|J| = \begin{vmatrix} \frac{\partial\theta}{\partial\theta'} & \frac{\partial\theta}{\partial\phi'} & \frac{\partial\theta}{\partial f'} & \frac{\partial\theta}{\partial d'} & \frac{\partial\theta}{\partial S_0'} \\ \frac{\partial\phi}{\partial\theta'} & \frac{\partial\phi}{\partial\phi'} & \frac{\partial\phi}{\partial f'} & \frac{\partial\phi}{\partial d'} & \frac{\partial\phi}{\partial S_0'} \\ \frac{\partial f}{\partial\theta'} & \frac{\partial f}{\partial\phi'} & \frac{\partial f}{\partial f'} & \frac{\partial f}{\partial d'} & \frac{\partial f}{\partial S_0'} \\ \frac{\partial d}{\partial\theta'} & \frac{\partial d}{\partial\phi'} & \frac{\partial d}{\partial f'} & \frac{\partial d}{\partial d'} & \frac{\partial d}{\partial S_0'} \\ \frac{\partial S_0}{\partial\theta'} & \frac{\partial S_0}{\partial\phi'} & \frac{\partial S_0}{\partial f'} & \frac{\partial S_0}{\partial d'} & \frac{\partial S_0}{\partial S_0'} \end{vmatrix}$$

Once the algorithm had finished the variables were then transformed back by using:

$$\theta = \frac{\pi}{1+\exp(-\theta')}$$

$$\phi = \frac{2\pi}{1+\exp(-\phi')}$$

$$f = \frac{1}{1+\exp(-f')}$$

$$d = \exp(d')$$

$$S_0 = \exp(S_0').$$

The steps of the Adaptive MCMC algorithm are summarised in Algorithm 8.

We attempted Adaptive MCMC on our simulated partial volume dataset. The

Algorithm 8 Adaptive MCMC

1: Implement Vanilla MCMC or Block-update MCMC for a pre-determined number of iterations, r, and then using these MCMC estimates calculate the empirical covariance matrix for the parameters which is denoted by Σ_n.
2: Propose $\boldsymbol{\omega}_{i+1} \sim N(\boldsymbol{\omega}_i, [2.38^2/d]\Sigma_n)$, where d is the number of dimensions.
3: Calculate the acceptance ratio which is

$$\alpha(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1}) = \min\left(1, \frac{\pi(\boldsymbol{\omega}_{i+1}|\boldsymbol{y})q(\boldsymbol{\omega}_{i+1}, \boldsymbol{\omega}_i)}{\pi(\boldsymbol{\omega}_i|\boldsymbol{y})q(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1})}\right) = \min\left(1, \frac{\pi(\boldsymbol{\omega}_{i+1}|\boldsymbol{y})}{\pi(\boldsymbol{\omega}_i|\boldsymbol{y})}\right)$$

because $\frac{q(\boldsymbol{\omega}_{i+1}, \boldsymbol{\omega}_i)}{q(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{i+1})}$=1.
4: Recalculate the empirical covariance matrix and start from step 2 again until we obtain the required number of estimates.

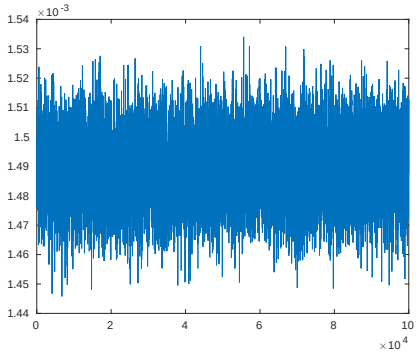


Figure 2.4: Traceplot for the parameter estimates of d from the partial volume model simulated from the posterior distribution using Adaptive MCMC.

output of the MCMC algorithm is shown in Figure 2.4. The graph demonstrates that the mixing within Adaptive MCMC is very good.

The kernel density plots of the parameter estimates are shown in Figure 2.5. These kernel density plots illustrate that Adaptive MCMC performs very well.

2.3.4 The independence sampler and the Laplace approximation

When implementing MCMC, up until now the proposal distribution has depended on the current values of the parameters. A proposal distribution that does not depend on the current value of the parameters is called an indepen-

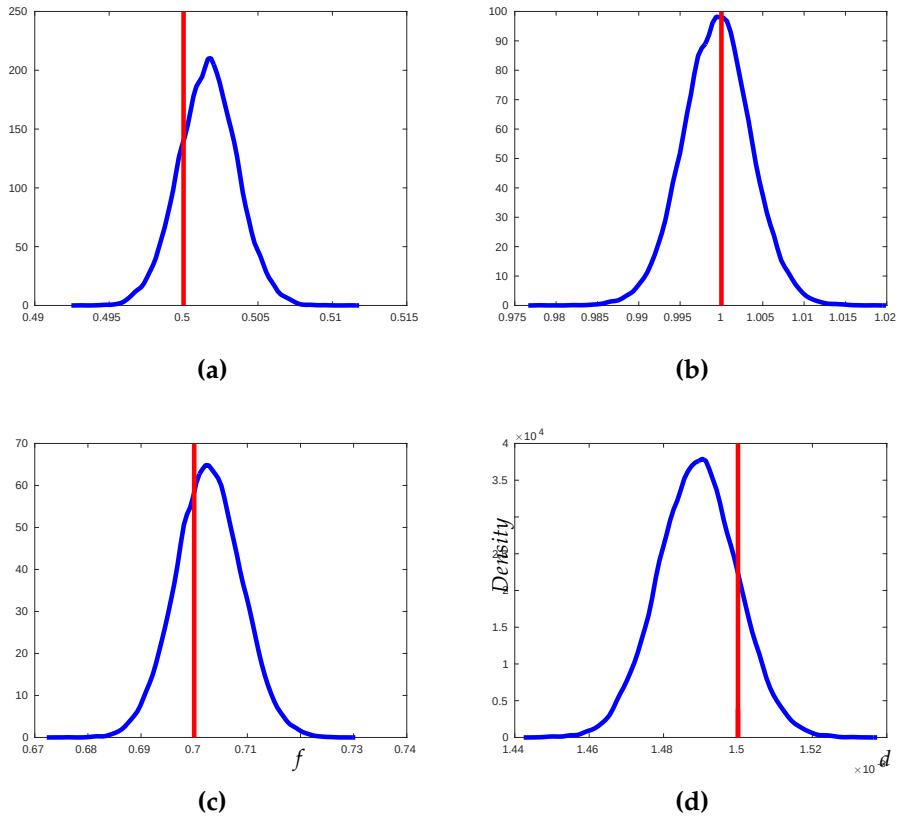


Figure 2.5: Kernel density plots of the parameter estimates of the partial volume model simulated from the posterior distribution using Adaptive MCMC (a) θ, (b) ϕ, (c) f and (d) d, with the true values indicated by the red line.

dence sampler (Gilks *et al.*, 1996). A good independence sampler will propose values for the parameters that are close to the real values, such that there is a sufficient acceptance rate in MCMC. One way of implementing this is to use a multivariate normal proposal distribution that approximates the target distribution. A suitable multivariate normal proposal distribution may be approximated by using the Laplace approximation (Tierney and Kadane, 1986).

The Laplace approximation can be obtained by finding the mode of the logarithm of the posterior distribution. Then within MCMC the proposal distribution is chosen to be $N(\widehat{\boldsymbol{\omega}}, \mathbf{H}^{-1})$, where $\mathbf{H}^{-1}$ is the inverse of the Hessian matrix and $\widehat{\boldsymbol{\omega}}$ is the value of $\boldsymbol{\omega}$ that maximises the logarithm of the posterior. The Laplace approximation is not affected by the unknown normalising constant of the posterior distibution because we work with the logarithm of the posterior distribution. Then when we calculate the Hessian matrix by finding the second derivatives of the logarithm of the posterior distribution, the logarithm of the normalising constant term will differentiate to 0. The Laplace approximation is an independence sampler on which candidate values do not depend on the current values of the parameters.

One issue with the Laplace approximation is deciding which value to choose as the initial vector of parameter values when maximising the logarithm of the posterior. Eventually the metrics obtained from the DT model as mentioned in Section 1.5 were used as the initial values for θ, ϕ, f and d. These initial values were then transformed to obtain θ', ϕ', f' and d' as in Section 2.3.3. The value of $S_0{}'$ was assigned to be the transformed value of the y_i that corresponds to the b_i that is equal to 0. It should be fairly simple to extend the Laplace approximation for partial volume models with more fibre orientations. The algorithm for MCMC that uses the independence sampler is summarised in Algorithm 9.

Algorithm 9 Independence sampler MCMC

1: Assign the initial value of S_0 to be the y_i value that corresponds to the b-value that is equal to 0 and using this value of S_0 use the linearised DT model to obtain $\widehat{\boldsymbol{B}}$, the ordinary least square estimator of $\boldsymbol{B}$ described in Section 2.2.1.
2: Use $\widehat{\boldsymbol{B}}$ to get values for the Fractional Anisotropy, the Mean Diffusivity and also the values of θ and ϕ from the principal eigenvector of the estimated Diffusion Tensor as described in Section 1.5. The values for the Fractional Anisotropy and Mean Diffusivity can be used as initial estimates for f and d.
3: Use the values obtained in Steps 1 and 2 as initial estimates for θ, ϕ, f, d and S_0 and optimise the posterior density, so that we obtain the Laplace approximation.
4: Use the Laplace approximation as the proposal distribution in Block-update MCMC as in Section 2.3.2.

When the Laplace approximation was applied to the simulated partial volume model dataset, the posterior mode was

θ=0.4966, ϕ=0.9957, f=0.6981, d=0.0015 and S_0=1.0011.

This compares to the true parameter values of

θ=0.5, ϕ=1.0, f=0.7, d=0.0015 and S_0=1.0.

The numerical covariance matrix that is obtained by calculating the inverse of the Hessian matrix is

$$\mathbf{H}^{-1} = \begin{bmatrix} 0.0001 & -0.0000 & 0.0000 & -0.0000 & -0.0000 \\ -0.0000 & 0.0001 & -0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -0.0000 & 0.2065 & -0.0259 & -0.0487 \\ -0.0000 & 0.0000 & -0.0259 & 0.0033 & 0.0061 \\ -0.0000 & 0.0000 & -0.0487 & 0.0061 & 0.0115 \end{bmatrix}.$$

Then using these values as described in Algorithm 9 as a proposal distribution, the results in Figures 2.6 and 2.7 are obtained. The results look very good when compared to the true parameter values and also the mixing is very good within MCMC. The acceptance rate of proposed samples in the MCMC appears to be

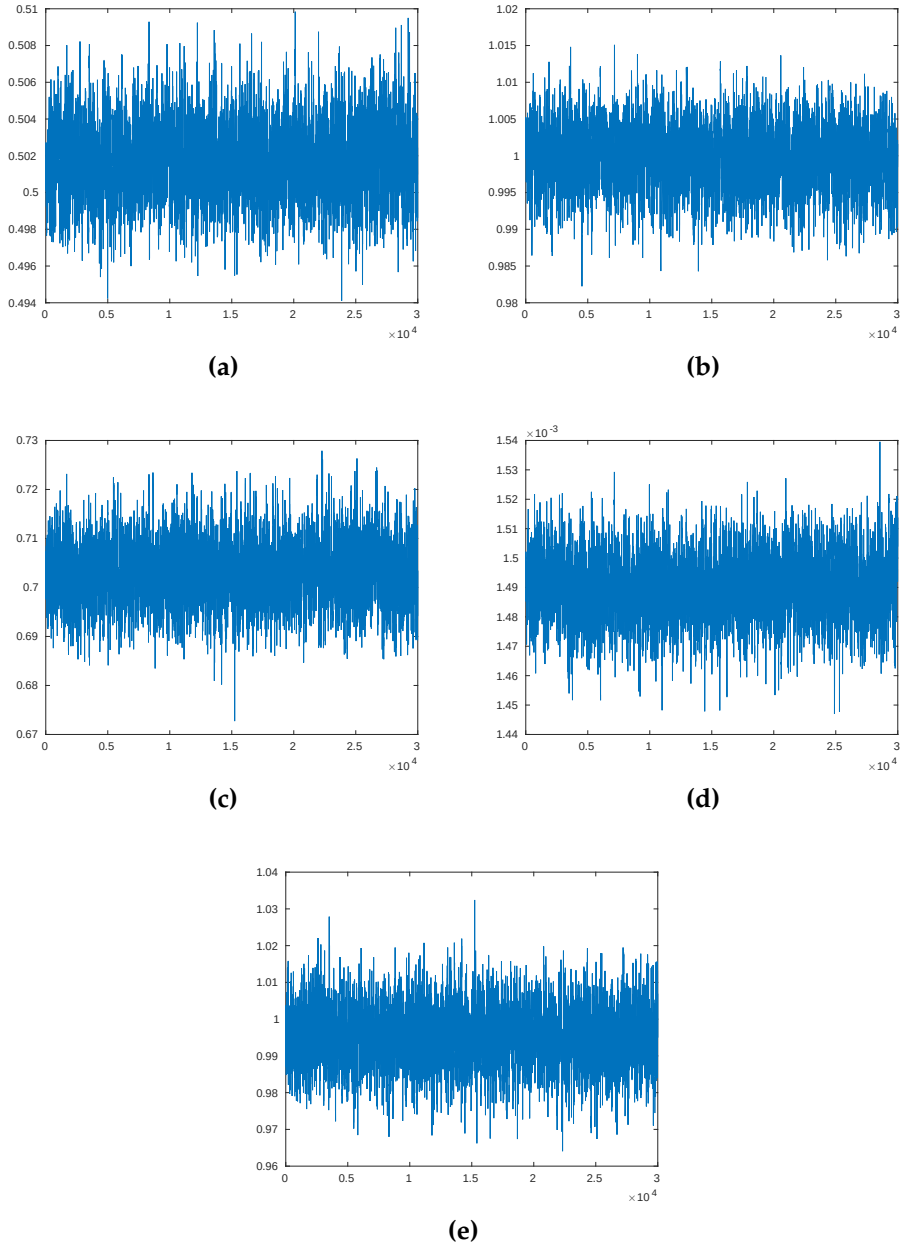


Figure 2.6: Traceplots for the parameter estimates of the partial volume model (a) θ, (b) ϕ, (c) f, (d) d and (e) S_0 simulated from the posterior distribution using the Laplace approximation as a proposal distribution in Block-update MCMC.

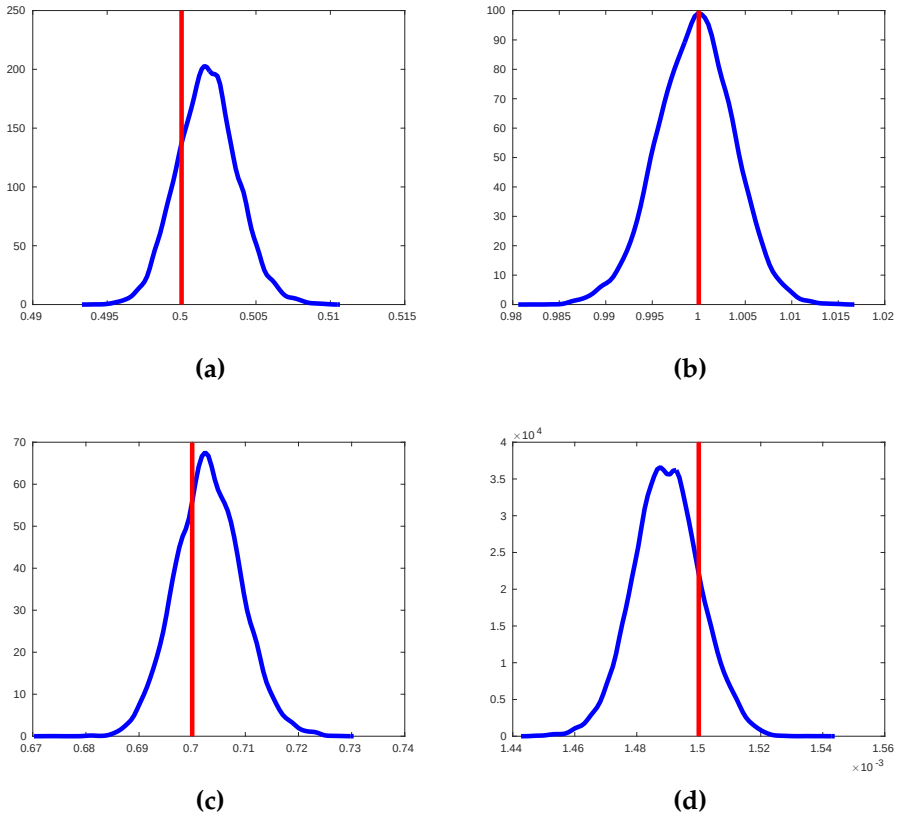


Figure 2.7: Kernel density plots for the parameter estimates of the partial volume model simulated from the posterior distribution using the Laplace approximation as a proposal distribution in Block-update MCMC.(a) θ, (b) ϕ, (c) f and (d) d. The true values of the parameters are indicated by the red line.

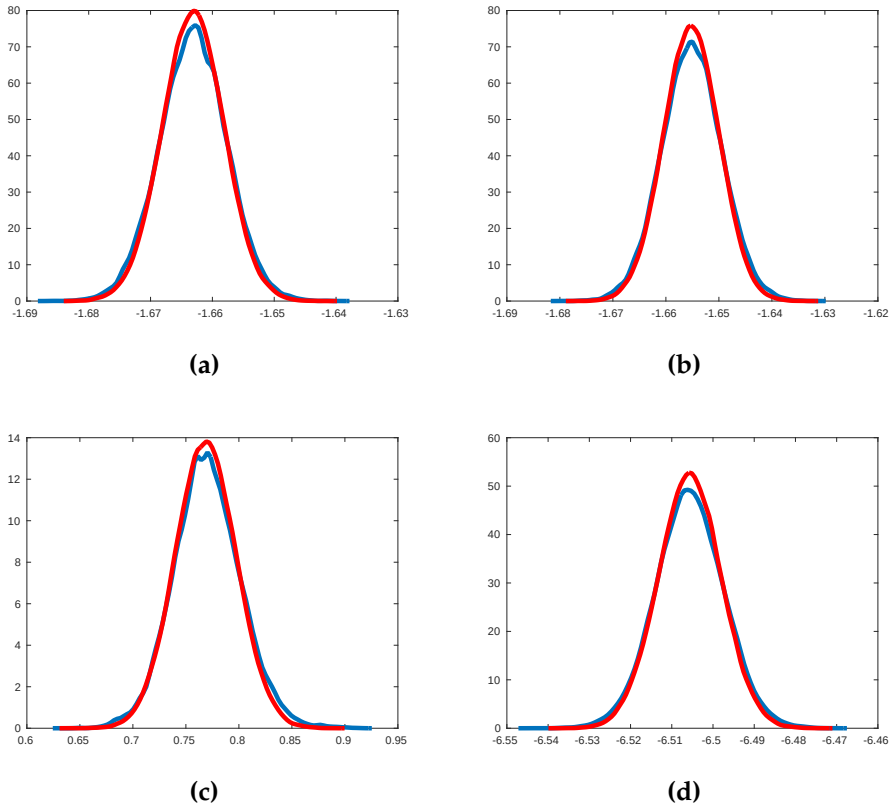


Figure 2.8: The kernel density plots from Vanilla MCMC and the Laplace approximation shown by the red curve for the transformed parameters of the partial volume model (a) θ', (b) ϕ', (c) f' and (d) d'

extremely good.

Throughout this chapter the MCMC methods that have been tried, work well at estimating the parameters of the partial volume model with one fibre orientation in a voxel. However MCMC methods in general can be very time consuming, and this will be more evident when many more voxels are considered. Instead we may use the density of the Laplace approximation as a good estimate of the posterior distribution, rather than as a proposal distribution.

The density of the Laplace approximation was then compared with the posterior density obtained from Vanilla MCMC. This is seen in Figure 2.8. From the graphs we can see that the Laplace approximation is in pretty good agreement

with the estimated posterior distribution obtained by MCMC for each parameter.

2.4 Reparameterisation

We now look at a special case of the partial volume model where the MCMC methods that we used in Section 2.3 may not work as well. Suppose that there is a voxel where the true value of θ is such that $\theta \approx 0$. In this case

$$\boldsymbol{v} = \begin{bmatrix} \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

so that the value of ϕ does not affect the value of $\boldsymbol{v}$ much and therefore many values of ϕ will be accepted when running MCMC to approximate the posterior distribution. In this case it would be better to directly infer $\boldsymbol{v}$ rather than θ and ϕ. The parameters θ and ϕ from the partial volume model can be reparameterised as in Section 1.9. Now this new reparameterisation will be used and the new parameter, $\boldsymbol{v}$ will be estimated.

MCMC will be used to estimate the posterior distribution, $\pi(\boldsymbol{v}, f, d, S_0|\boldsymbol{y})$. At each iteration of MCMC a new candidate value of $\boldsymbol{v}$ will be simulated from the proposal distribution, which will either be the Bingham distribution or the Angular Central Gaussian (ACG) distribution (see Section 1.9). Both the ACG and Bingham distributions have a parameter matrix which can be chosen such that the proposal distribution simulates values that are accepted often. Assume that the parameter matrix of both the Bingham and ACG distributions is denoted $\mathbf{A}$, then a good approximation for $\mathbf{A}$ can be found as follows.

Let

$$\mathbf{A} = \mathbf{V}\mathbf{U}\mathbf{V}^T,$$

such that this is the spectral decomposition of $\mathbf{A}$. Making the eigenvalues in $\mathbf{U}$ larger corresponds to more clustering around the eigenvector that this eigenvalue corresponds to (Mardia and Jupp, 2000). If the Laplace approximation (Section 2.3.4) is carried out such that estimates for θ and ϕ are available, then a corresponding estimate for $\boldsymbol{v}$, can be calculated. Estimates that cluster around this approximation of $\boldsymbol{v}$ would be desirable.

We will denote $\boldsymbol{v} = [a, b, c]$ and create a 3x3 symmetric matrix $\mathbf{V}$ such that the third column of $\mathbf{V}$ is $\boldsymbol{v}$, then

$$\mathbf{V} = \begin{bmatrix} d & e & a \\ e & f & b \\ a & b & c \end{bmatrix}$$

and

$$\mathbf{U} = \begin{bmatrix} u_1 & 0 & 0 \\ 0 & u_2 & 0 \\ 0 & 0 & u_3 \end{bmatrix}$$

where the value of u_3 is positive and very large in relation to u_1 and u_2 (e.g. $u_1 = u_2 \approx 0.0$, $u_3 = 900$). Values of d, e and f can be found such that $\mathbf{V}\mathbf{V}^T$=$\mathbf{I}_3$, the 3x3 identity matrix. For this to hold true

$$\begin{bmatrix} d & e & a \\ e & f & b \\ a & b & c \end{bmatrix} \begin{bmatrix} d & e & a \\ e & f & b \\ a & b & c \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

There are then six equation to solve which are

$$d^2 + e^2 + a^2 = 1$$

$$e(d + f) + ab = 0$$

$$a(d + c) + eb = 0$$

$$e^2 + f^2 + b^2 = 1$$

$$b(f + c) + ea = 0$$

$$a^2 + b^2 + c^2 = 1.$$

The last equation can be ignored because $\boldsymbol{v}\boldsymbol{v}^T = 1$ so already

$$a^2 + b^2 + c^2 = 1,$$

by solving the other five equations, the following solution is obtained:

$$e = \frac{(-\frac{cb}{a}) - \frac{b}{a}}{1 + \frac{b^2}{a^2}}$$

$$d = \frac{-eb - ac}{a}$$

$$f = \frac{-bc - ea}{b}.$$

We can then first use the Laplace approximation to get estimates for θ and ϕ and then find the corresponding value of $\boldsymbol{v} = [a, b, c]$. Then we find the values of d, e and f to obtain a good approximation for the parameter matrix $\mathbf{A}$ in both the Bingham and the ACG distributions. In Section 1.9 we described how to simulate from the ACG distribution and the Bingham distribution. Therefore we

will now use the value for the estimate of the parameter matrix $\mathbf{A}$ and compare the different proposal distributions within MCMC.

2.5 Comparing the different proposal distributions

Metropolis-Hastings MCMC (Section 1.8.1) was now implemented using a variety of proposal distributions on the partial volume dataset that we have used throughout this chapter. The proposal distributions were then compared by looking at plots of the Autocorrelation function (Box and Jenkins, 1976). The Autocorrelation function (ACF) allows us to see how the correlation between the parameter estimates from MCMC changes between consecutive samples.

First we used the reparameterisation such that the fibre direction is a vector, $\boldsymbol{v}$ and then the Bingham and ACG proposal distributions are compared. Afterwards we also compared MCMC methods that do not use the reparameterisation. We investigated MCMC methods that uses a random-walk proposal distribution and the independence sampler proposal distribution obtained from the Laplace approximation. The ACF plots when comparing the different proposal distributions are in Figures 2.9 and 2.10.

Clearly the Laplace approximation independence sampler is the best proposal distribution in this example. The ACG distribution is a better proposal distribution than the Bingham distribution, when using the vector reparameterisation. The ACG distribution is also much easier to sample from than the Bingham distribution and thus when working with the reparameterisation, the ACG proposal distribution should be used.

We will now return to the motivating example that made us propose the repa-

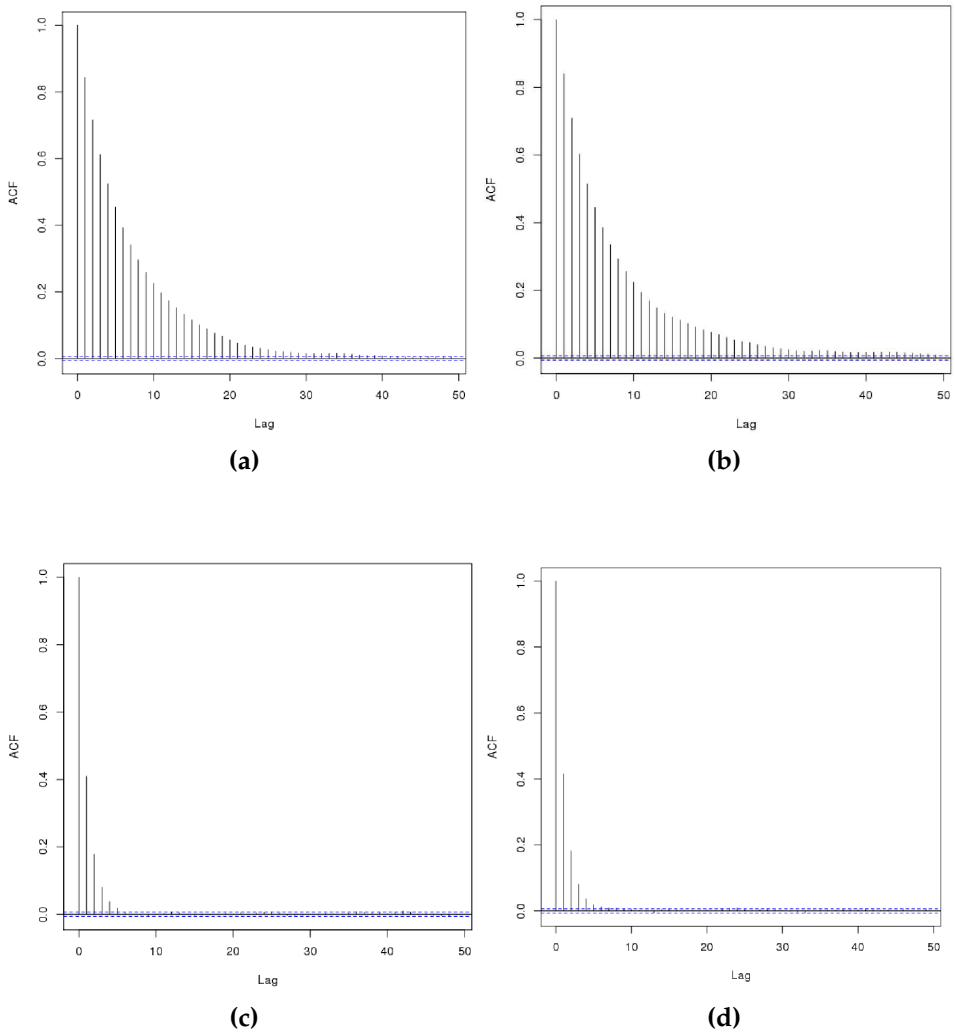


Figure 2.9: The ACF from the MCMC estimates of (a) θ and (b) ϕ in the partial volume model when using the Bingham proposal distribution, (c) θ and (d) ϕ in the partial volume model when using the ACG proposal distribution.

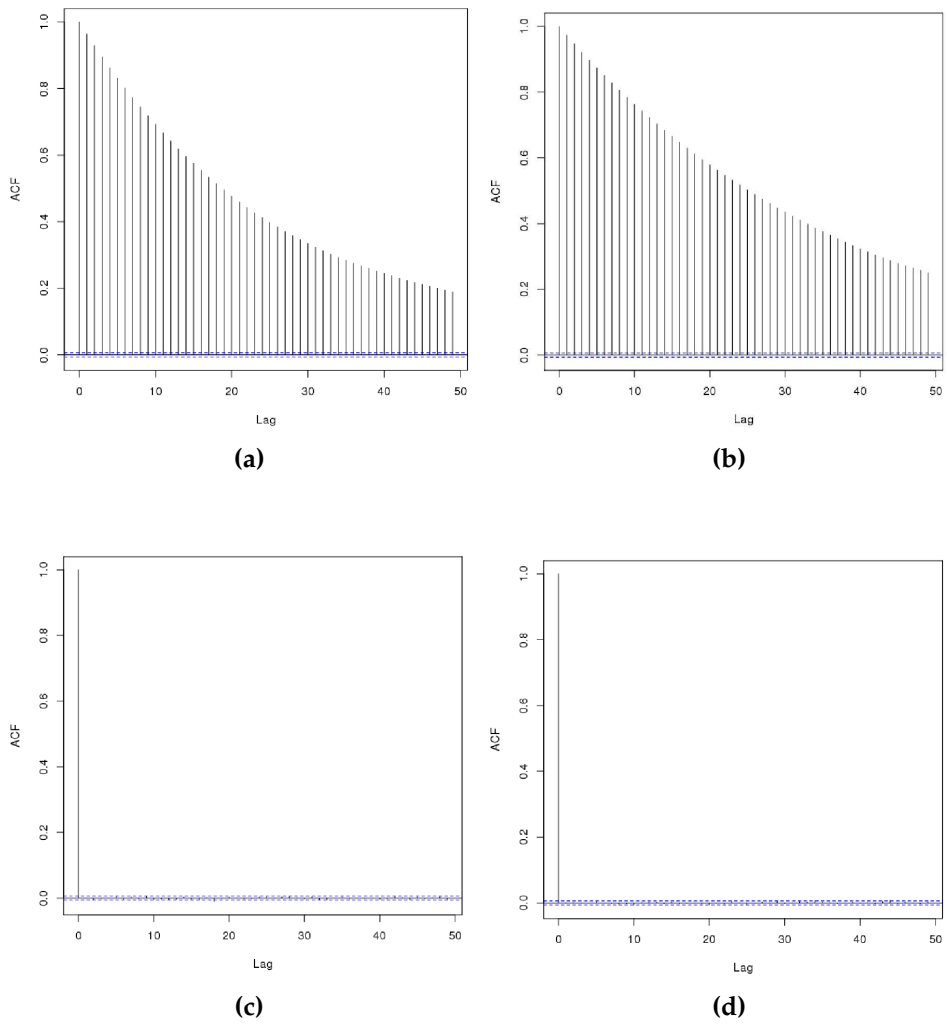


Figure 2.10: The ACF of (a) θ (b) ϕ in the partial volume model when using a random-walk proposal distribution, (c) θ and (d) ϕ in the partial volume model when using the Laplace approximation as a proposal distribution within Block-update MCMC.

rameterisation of the model. Suppose that $\theta \approx 0$, then

$$\boldsymbol{v} = \begin{bmatrix} \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

It is expected that it will be better to use the reparameterisation in this example. For this a new dataset was generated from the partial volume model with θ=0.01 and ϕ=0.1. Then we tested the four proposal distributions and both MCMC methods when using θ and ϕ as the parameters were allowing ϕ to take any value, while the MCMC methods using the reparameterisation generated good values for $\boldsymbol{v}$ as in Figure 2.11. The mixing when not using the reparameterisation seems to be bad for θ and ϕ as it seems to stop at points as in the graph in Figure 2.11 (d).

We have now demonstrated that there are some cases where it is better to use the vector $\boldsymbol{v}$ as the fibre orientation within a voxel, rather than θ and ϕ. We will keep this in consideration when doing further inference. When we are working with real datasets we may decide to use the reparameterisation if the Laplace approximation returns values of θ that are close to 0.

2.6 An application to real data

We have demonstrated that the methods that were proposed in this chapter work very well on simulated datasets. The main purpose of these methods is to be applied on real datasets. We will compare the results from our code with the results obtained by FSL (Section 1.7.2).

One voxel was chosen from a real dataset and then the code was attempted

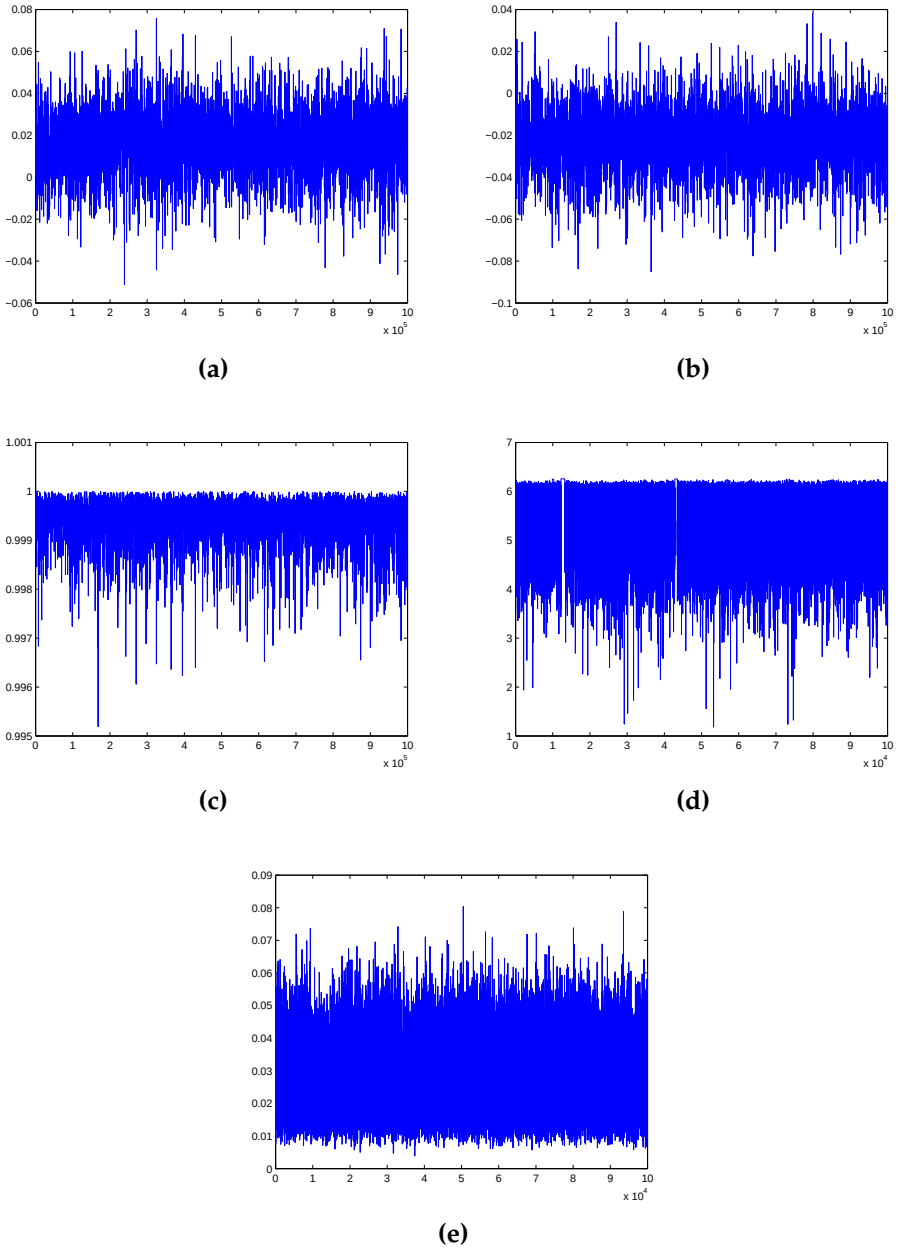


Figure 2.11: The MCMC trace plots when using a Bingham proposal distribution (a,b and c) and when using the independence sampler (d and e) for (a) the first, (b) second and (c) third element of $\boldsymbol{v}$ in the partial volume model, (d) ϕ and (e) θ in the partial volume model.

on this voxel. First the Laplace approximation was used as an independence sampler in MCMC, MCMC was then implemented using both our code and FSL to compare the results. The results are shown in Figure 2.12. From the graphs it can be seen that the estimates for the parameters in the partial volume model are similar using both our code and FSL. Note that FSL only records every 20th iteration of MCMC to account for correlations between output.

Then when we also compare these results with the Laplace approximation of the dataset, the results in Figure 2.13 are obtained, showing that the Laplace approximation of the real data is a very good approximation to the posterior distribution when compared to both our MCMC results and the FSL results. This section seems to confirm that our MCMC algorithms for inferring the values of the parameters in the partial volume model are as good as the corresponding results from FSL because the samples from both our MCMC algorithms and FSL correspond with the Laplace approximation. In FSL the proposal distribution for each parameter is a zero mean Gaussian where the standard deviation is tuned to give an acceptance rate of 0.5. Thinning is used so that every 20th iteration is sampled (Behrens *et al.*, 2003). In our algorithm we use an independence sampler obtained from the Laplace approximation which allows us to propose and then accept or reject samples for every parameter in one iteration. We will now compare the methods that we have introduced in this chapter.

2.7 Simulation study

Within this chapter we have investigated many different methods for inferring the parameters of the partial volume model within a voxel. We will now conduct a simulation study in which we will consider two different cases; a voxel with one fibre orientation and a voxel with two fibre orientations. For both

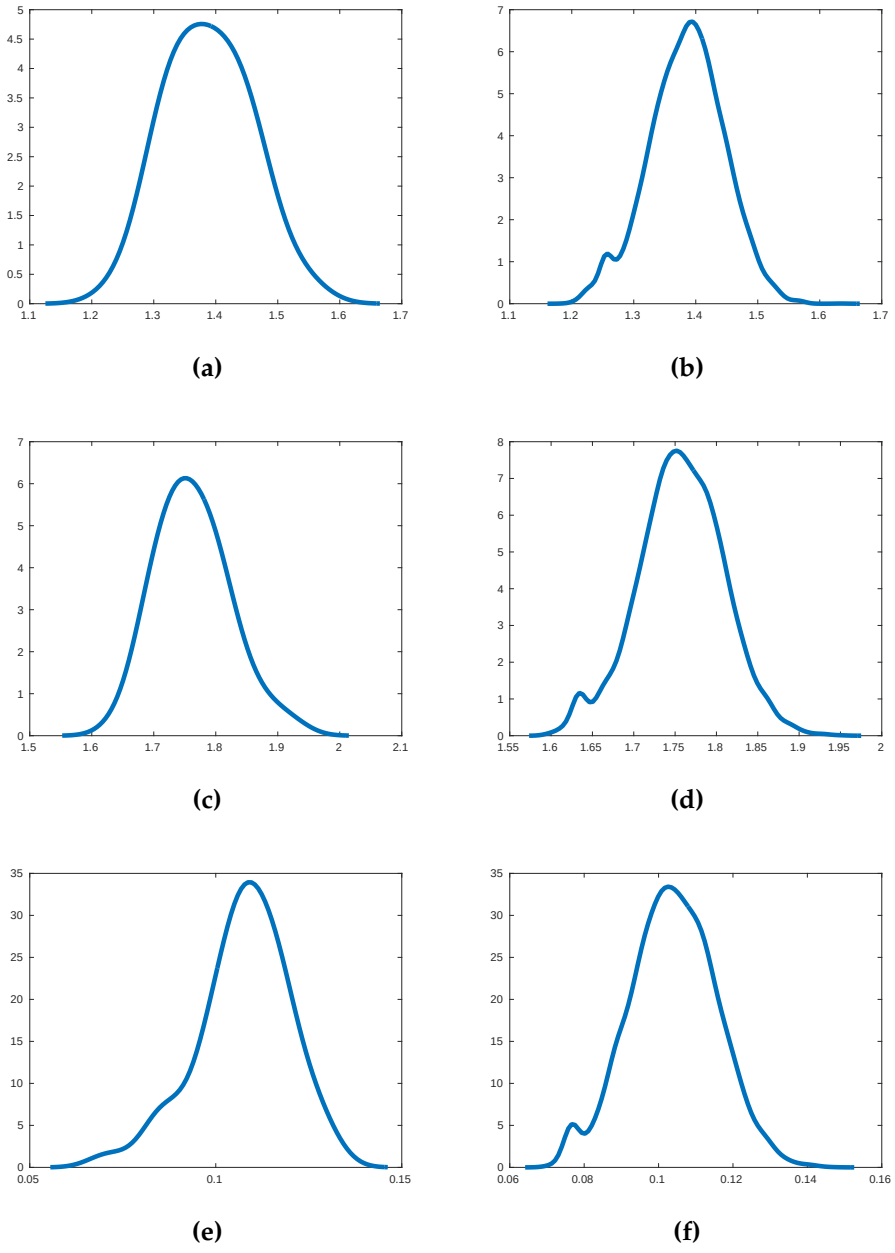


Figure 2.12: The kernel density plots of the MCMC parameter estimates from the partial volume model for (a) θ, (c) ϕ and (e) f using FSL compared with (b) θ (d) ϕ and (f) f using our code.

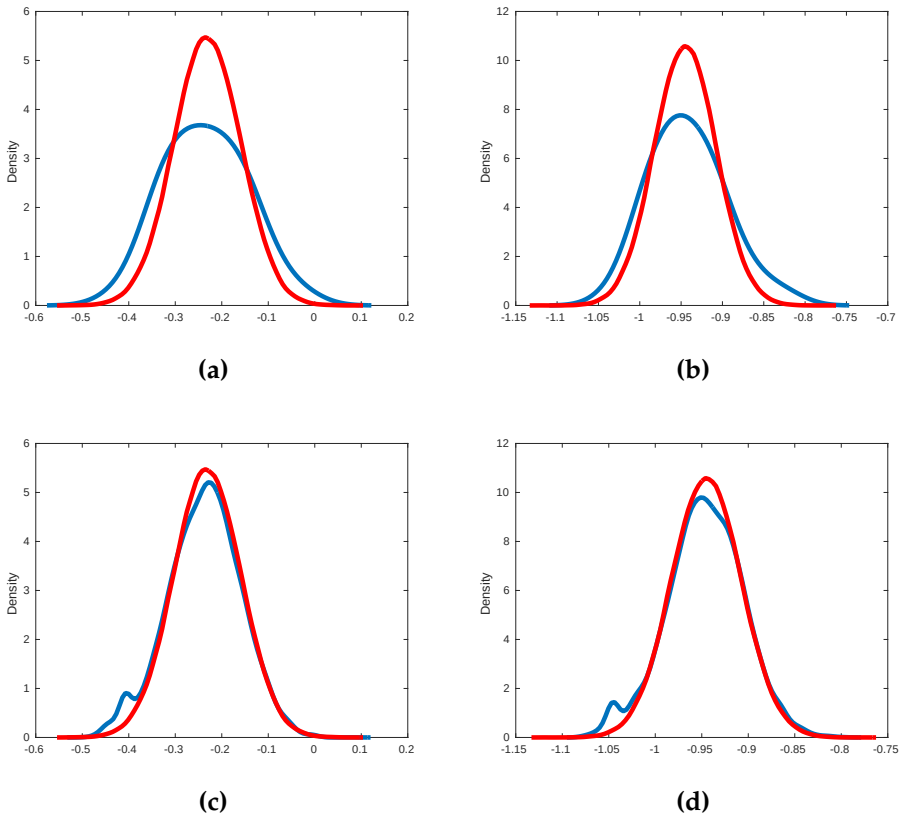
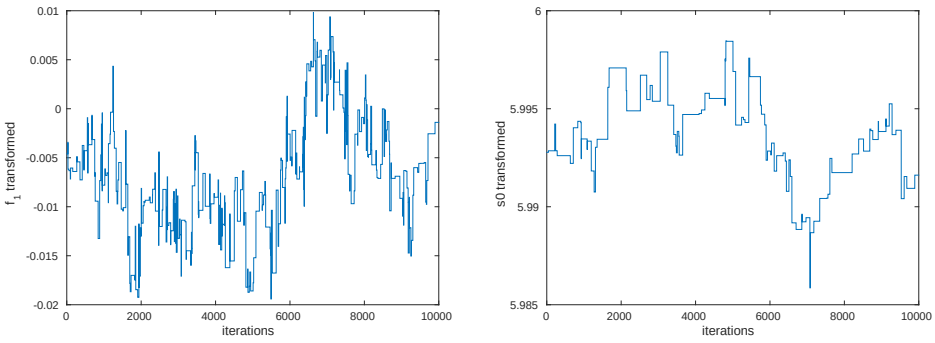


Figure 2.13: The kernel density plots of the MCMC parameter estimates from the partial volume model compared with the Laplace approximation (in red) of (a) θ and (b) ϕ using FSL, (c) θ and (d) ϕ using our code.



(a) f **(b)** S_0

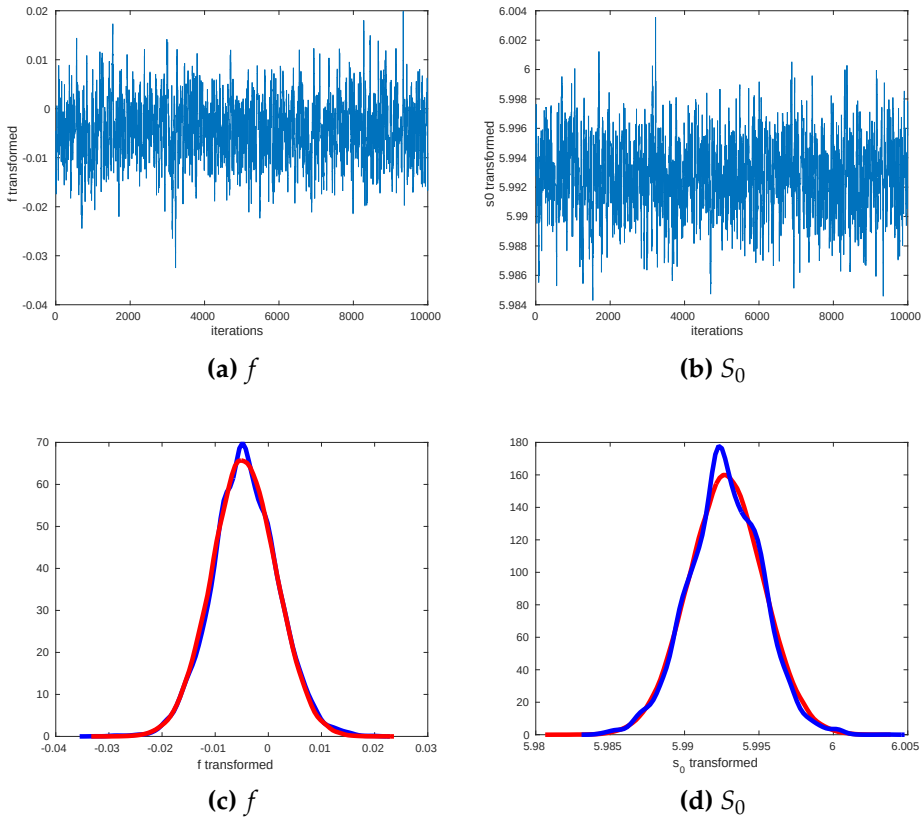
Figure 2.14: The MCMC estimates when using Vanilla MCMC for the parameters of the partial volume model (a) f and (b) S_0 when $\tau = 1$.

voxels we will look at different values of τ to see how the precision affects the results. We will compare Vanilla MCMC (Algorithm 6), Block-update MCMC (Algorithm 7), Adaptive MCMC (Algorithm 8), independence sampler MCMC (Algorithm 9) and the Laplace approximation (Section 2.3.4).

2.7.1 One fibre orientation partial volume model dataset

At first we simulated a dataset of size 61 from the partial volume model with one fibre orientation with parameter values of $\theta_1 = 1$, $\phi_1 = 1$, $f_1 = 0.5$, $d = 0.001$ and $S_0 = 400$ when $\tau = 1$. Initially Vanilla MCMC was attempted on this dataset and the results of the parameter estimates can be seen in Figure 2.14. Clearly these traceplots reveal quite bad mixing because of tuning issues and therefore we will obtain the Laplace approximation so that we can choose it as a random-walk proposal distribution that can be used in Block-update MCMC. We will use the values of the posterior mode from the Laplace approximation as the initial values of the parameters. The inverse of the Hessian matrix will be used as the covariance matrix in the proposal distribution.

The Block-update MCMC algorithm results as shown in Figure 2.15 show the



(a) f **(b)** S_0

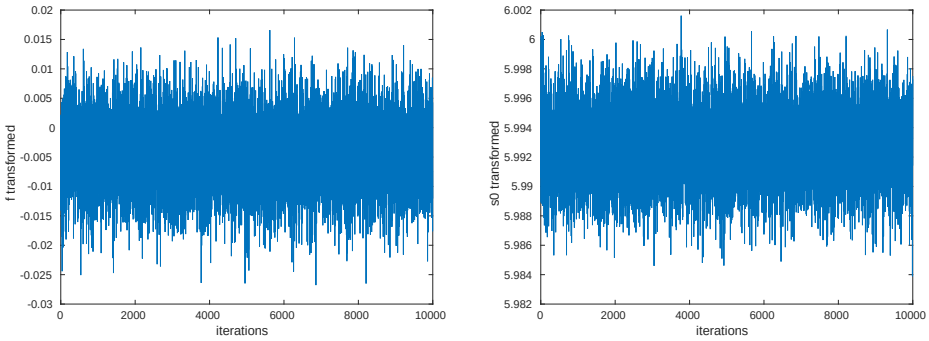
(c) f **(d)** S_0

Figure 2.15: The traceplots and kernel density plots of the MCMC estimates when using Block-update MCMC for the parameters of the partial volume model (a) (c) f and (b) (d) S_0 when $\tau = 1$.

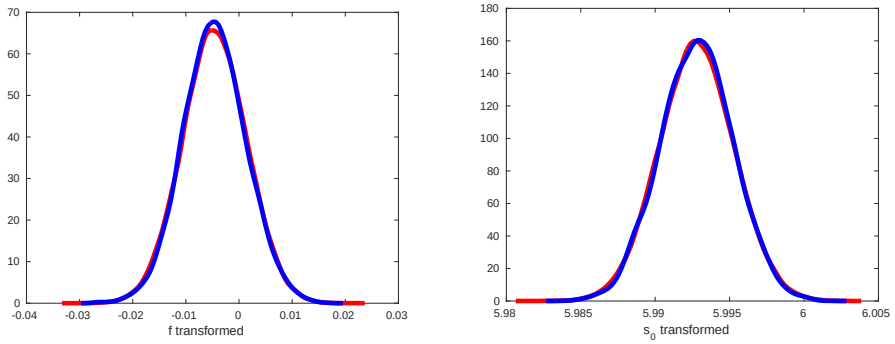
traceplots of some of the parameter estimates and the kernel density plots of the parameter estimates compared with the Laplace approximation. We observe that the mixing has greatly improved when compared with the traceplots when using Vanilla MCMC.

Afterwards independence sampler MCMC was attempted. Instead of using a random-walk to propose values, this MCMC uses the mean and Hessian from the Laplace approximation. The graphs of the results are shown in Figure 2.16. The mixing looks excellent and the kernel density plots of the parameter estimates resemble the Laplace approximation density.

Finally the MCMC is then implemented using Adaptive MCMC as in Figure

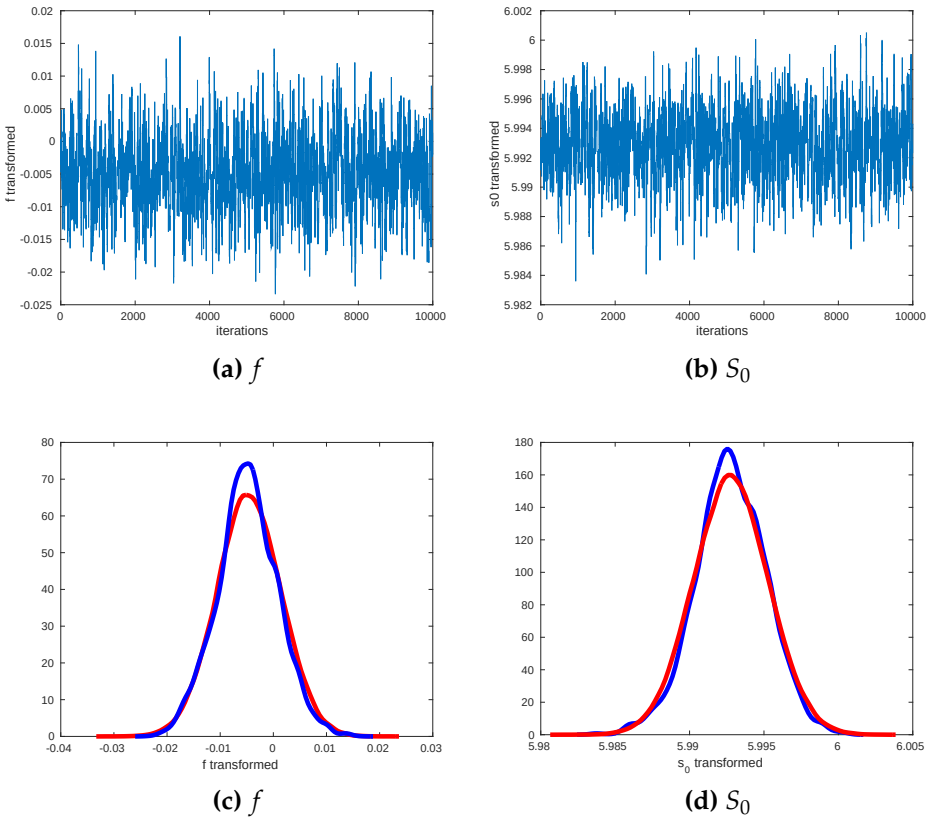


(a) f **(b)** S_0



(c) f **(d)** S_0

Figure 2.16: The traceplots and kernel density plots of the MCMC estimates when using independence sampler MCMC for the parameters of the partial volume model (a) (c) f and (b) (d) S_0 when $\tau = 1$.



(a) f **(b)** S_0

(c) f **(d)** S_0

Figure 2.17: The traceplots and kernel density plots of the MCMC estimates when using Adaptive MCMC for the parameters of the partial volume model (a) (c) f and (b) (d) S_0 when $\tau = 1$.

2.17. The results appear to be good but do not look as good as the independence sampler MCMC results because the acceptance rate is lower.

We now compare the mean and standard deviation of the MCMC results. We omit the results using Vanilla MCMC because the estimates do not appear to be good. These results are summarised in Table 2.1. The results for each of the three MCMC algorithms look similar. All of the approaches are quick, although the independence sampler MCMC is faster than the other methods because the proposal distribution is always the same.

To compare the different types of MCMC we will look at the Automatic Correlation Function (ACF) plots. These are in Figure 2.18. From these graphs it

	Block-update	Independence	Adaptive
θ_1'	-0.7646 (0.0015)	-0.7646 (0.0015)	-0.7647 (0.0014)
ϕ_1'	-1.6654 (0.0014)	-1.6654 (0.0014)	-1.6655 (0.0014)
f_1'	-0.0043 (0.0061)	-0.0046 (0.0059)	-0.0049 (0.0057)
d′	-6.9103 (0.0022)	-6.9102 (0.0022)	-6.9104 (0.0022)
S_0'	5.9927 (0.0024)	5.9928 (0.0024)	5.9928 (0.0024)

Table 2.1: The mean (and standard deviation) of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC and Adaptive MCMC on the simulated dataset when $\tau = 1$.

Samples	Block-update	Independence	Adaptive	Vanilla (f)
10000	571.6955	7993.229	550.8174	19.50942
9500	536.9074	7687.069	525.0637	18.46492

Table 2.2: The ESS of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC, Adaptive MCMC and Vanilla MCMC on the simulated dataset when $\tau = 1$. The first row is from all 10000 samples, the second row is when the first 500 samples are removed as burn-in.

is clear that the best MCMC estimates are from independence sampler MCMC and that Vanilla MCMC is bad when compared to the other methods. We will also compare the Effiecient Sample Size (ESS) (Roberts, 1996) of all the methods. This gives an estimate of the number of independent samples within the samples that we have. The ESS results can be found in Table 2.2. From the results of the ESS it is very clear that the indepedence sampler MCMC is much better than the other methods as it gives a lot more independent samples.

We repeat the investigation on a dataset with all the same parameters except for τ which is $\tau = 0.1$. The results from the MCMC algorithms were similar to obtained obtained when using $\tau = 1$. For each of the MCMC algorithms the mean and standard deviation of the parameter estimates were calculated and are summarised in Table 2.3. There does not seem much difference in the mean and standard deviation from using different MCMC methods.

We compare the different MCMC methods by looking at the ACF plots which

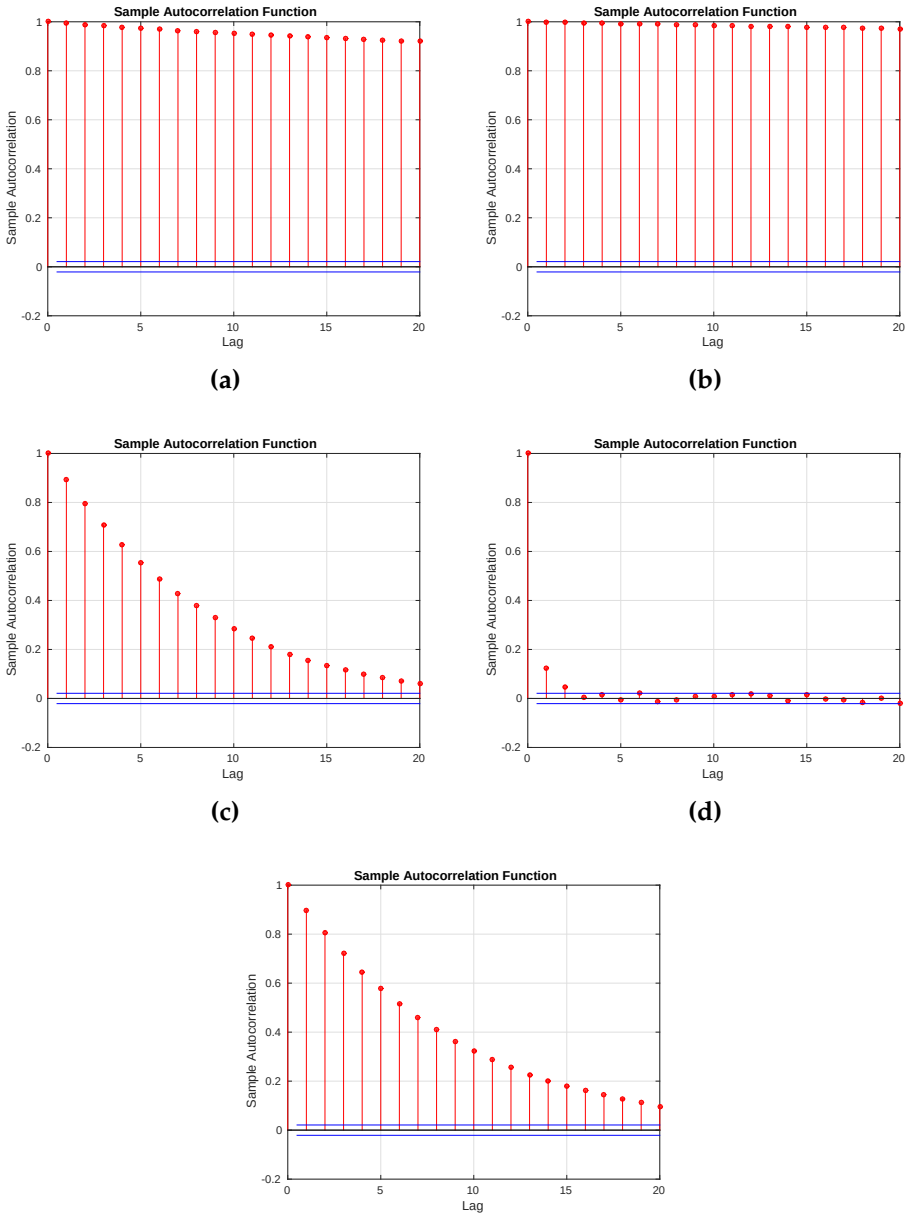


Figure 2.18: The plots of the ACF for the parameter estimates when using Vanilla MCMC for the partial volume model (a) f and (b) S_0 and for f using (c) Block-update MCMC, (d) independence sampler MCMC and (e) Adaptive MCMC when $\tau = 1$.

	Block-update	Independence	Adaptive
θ_1'	-0.7506 (0.0132)	-0.7511 (0.0127)	-0.7509 (0.0134)
ϕ_1'	-1.6576 (0.0120)	-1.6574 (0.0122)	-1.6574 (0.0128)
f_1'	0.1140 (0.0569)	0.1175 (0.0592)	0.1165 (0.0615)
d'	-6.8901 (0.0210)	-6.8910 (0.0202)	-6.8921 (0.0199)
S_0'	5.9426 (0.0226)	5.9406 (0.0233)	5.9405 (0.0233)

Table 2.3: The mean (and standard deviation) of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC and Adaptive MCMC on the simulated dataset when $\tau = 0.1$.

Samples	Block-update	Independence	Adaptive
10000	543.0111	7210.751	621.8843
9500	521.7032	6852.391	603.0329

Table 2.4: The ESS of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC and Adaptive MCMC on the simulated dataset when $\tau = 0.1$. The first row is from all 10000 samples, the second row is when the first 500 samples are removed as burn-in.

are shown in Figure 2.19 and by comparing the ESS which can be found in Table 2.4. From these again we see that the independence sampler MCMC produces the best estimates. Vanilla MCMC produces very correlated results.

Finally for the partial volume data with one fibre orientation we simulate data with $\tau = 2$. After implementing the MCMC algorithms we observed that the results of the mixing are very similar to the mixing when $\tau = 1$. For each of the MCMC algorithms the mean and standard deviation of the parameter estimates were calculated and can be summarised in Table 2.5. These seem to show that there is not really any difference in the mean and standard deviation of the estimates obtained using the different MCMC methods.

The ACF plots of the different MCMC methods are shown in Figure 2.20. Once again from these we can see that the independence sampler MCMC produces results with less correlation between samples, while the Vanilla MCMC produces results that are very correlated.

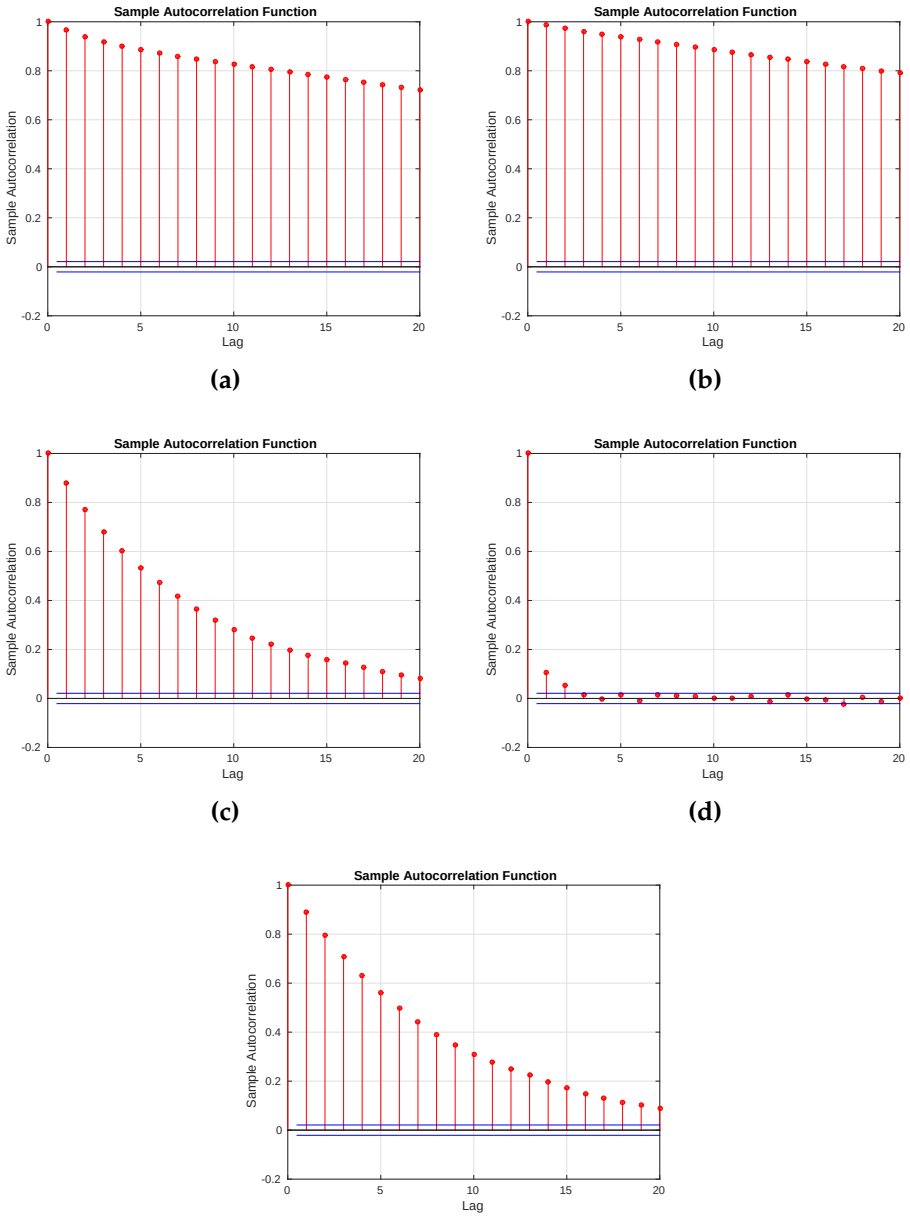


Figure 2.19: The plots of the ACF for the parameter estimates when using Vanilla MCMC for the partial volume model (a) f and (b) S_0 and (c) Block-update MCMC, (d) independence sampler MCMC and (e) Adaptive MCMC when $\tau = 0.1$.

	Block-update	Independence	Adaptive
θ_1'	-0.7613 (0.0009)	-0.7612 (0.0008)	-0.7612 (0.0009)
ϕ_1'	-1.6641 (0.0008)	-1.6641 (0.0008)	-1.6641 (0.0008)
f_1'	0.0012 (0.0035)	0.0015 (0.0035)	0.0015 (0.0035)
d′	-6.9070 (0.0013)	-6.9071 (0.0013)	-6.9071 (0.0013)
S_0'	5.9913 (0.0014)	5.9912 (0.0014)	5.9911 (0.0014)

Table 2.5: The mean (and standard deviation) of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC and Adaptive MCMC on the simulated dataset when $\tau = 2$.

Samples	Block-update	Independence	Adaptive	Vanilla (f)
10000	571.6245	7163.868	563.9729	7.614286
9500	602.8724	7086.818	527.0038	6.784524

Table 2.6: The ESS of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC, Adaptive MCMC and Vanilla MCMC on the simulated dataset when $\tau = 2$. The first row is from all 10000 samples, the second row is when the first 500 samples are removed as burn-in.

The same conclusions are also reached when looking at the ESS in Table 2.6.

In summary after looking at all the results it can be seen that Vanilla MCMC is not very good due to the fact that it is very much dependent on choosing a good proposal distribution. However because we now have good initial approximations for the parameters of the partial volume model due to the Laplace approximation, there is not much difference in the other three MCMC methods and all of them perform equally well. Changing the value of τ also does not affect which of the methods is best in terms of the mean. From the ACF plots it is clear that the independence sampler MCMC is the most efficient algorithm as it produces results that are almost uncorrelated samples.

2.7.2 Two fibre orientations partial volume model dataset

Until now we have considered datasets from the partial volume model where there is only one fibre orientation in a voxel. Therefore we will now implement

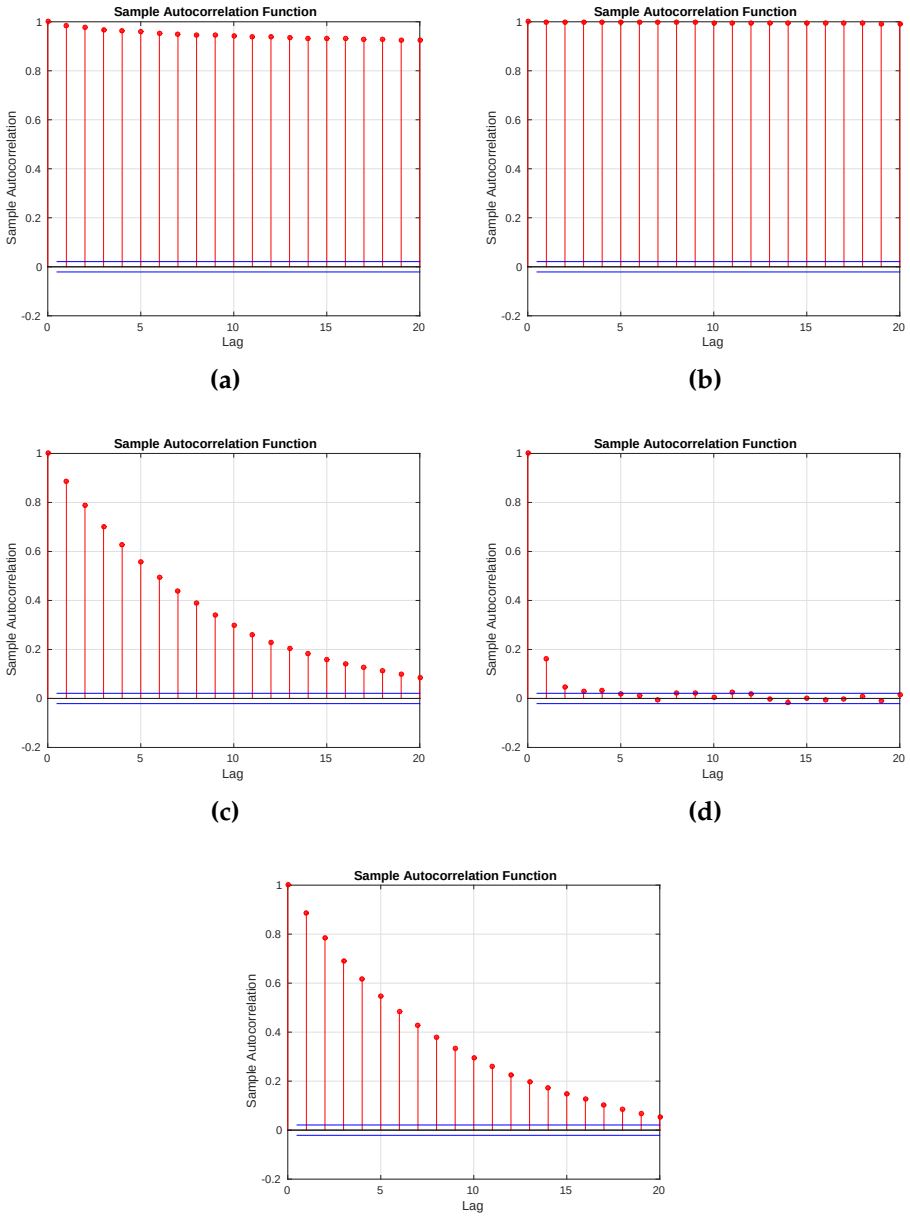
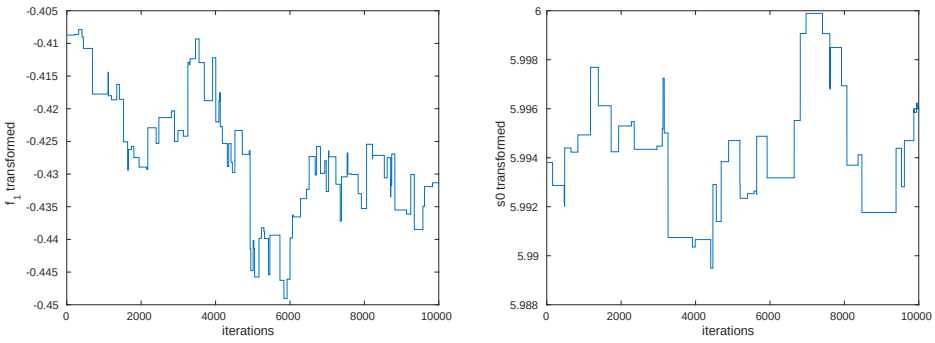


Figure 2.20: The plots of the ACF for the parameter estimates when using Vanilla MCMC for the partial volume model (a) f and (b) S_0 and (c) Block-update MCMC, (d) independence sampler MCMC and (e) Adaptive MCMC when $\tau = 2$.



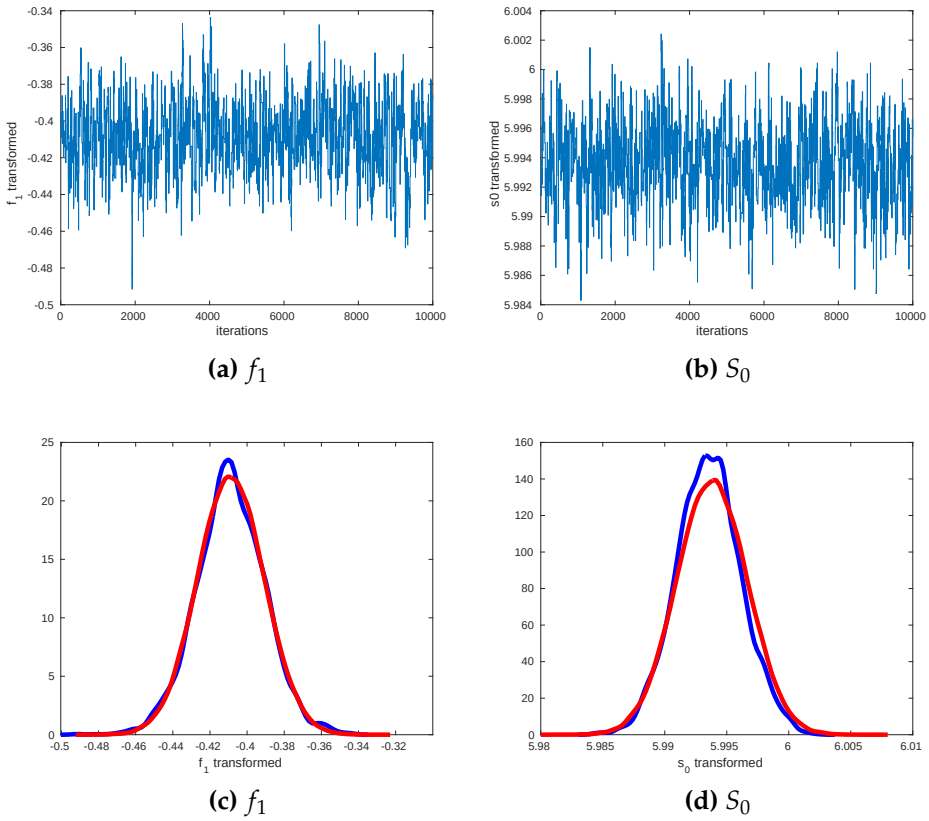
(a) f_1 **(b)** S_0

Figure 2.21: The MCMC estimates using Vanilla MCMC for the parameters of the partial volume model with two fibre orientations (a) f_1 and (b) S_0 when $\tau = 1$.

Vanilla MCMC (Algorithm 6), Block-update MCMC (Algorithm 7), independence sampler MCMC (Algorithm 9) and Adaptive MCMC (Algorithm 8) on partial volume datasets with two fibre orientations and compare these results with the Laplace approximation. This will allow us to observe if our methods can be easily extended to data with more than one fibre orientation. We vary the value of τ so that we can investigate the effect it has on the results. We use values of τ which are 1, 2 and 0.5.

First we simulated a dataset of size 61 from the partial volume model with two fibre orientations that has parameter values $\theta_1 = 1$, $\phi_1 = 1$, $f_1 = 0.2$, $\theta_2 = 0.5$, $\phi_2 = 1.5$, $f_2 = 0.4$, $d = 0.001$ and $S_0 = 400$ when $\tau = 1$ such that $\boldsymbol{v}_1 = \begin{bmatrix} 0.4546 & 0.7081 & 0.5403 \end{bmatrix}^T$ and $\boldsymbol{v}_2 = \begin{bmatrix} 0.0339 & 0.4782 & 0.8776 \end{bmatrix}^T$ We chose these values because they are typical values in real data and the two fibre orientations are very distinct. We implemented Vanilla MCMC and obtained the graphs in Figure 2.21. Similarly to the results when investigating datasets with one fibre orientation we see that the estimates from Vanilla MCMC are bad due to the mixing in the traceplots.

We then used Block-update MCMC with a random-walk proposal distribution



(a) f_1 **(b)** S_0

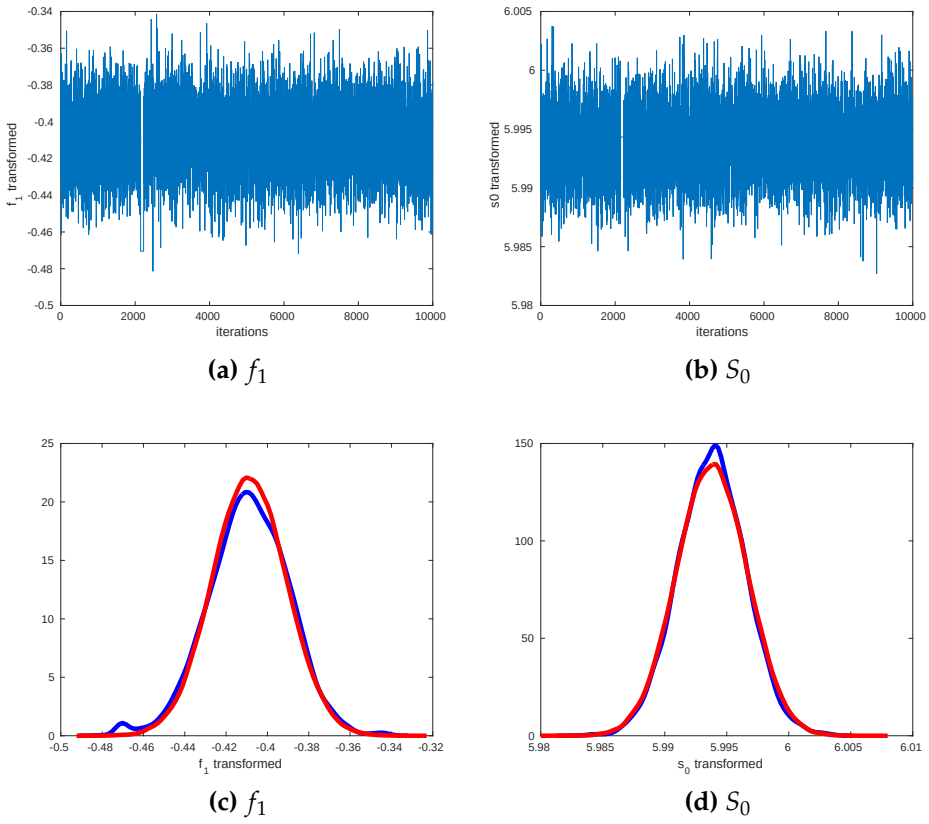
(c) f_1 **(d)** S_0

Figure 2.22: The traceplots and kernel density plots of the MCMC estimates using Block-update MCMC for the parameters of the partial volume model with two fibre orientations (a) (c) f_1 and (b) (d) S_0 when $\tau = 1$.

estimated by the Laplace approximation. The traceplots and kernel density plots of the parameter estimates are in Figure 2.22. We immediately notice that the mixing of the Block-update MCMC algorithm is an improvement on Vanilla MCMC. The estimates resemble the Laplace approximation density.

Afterwards independence sampler MCMC was attempted. The corresponding graphs of the parameter estimates are shown in Figure 2.23. The mixing in the traceplots is excellent and the estimates seem to come from the same density as the Laplace approximation.

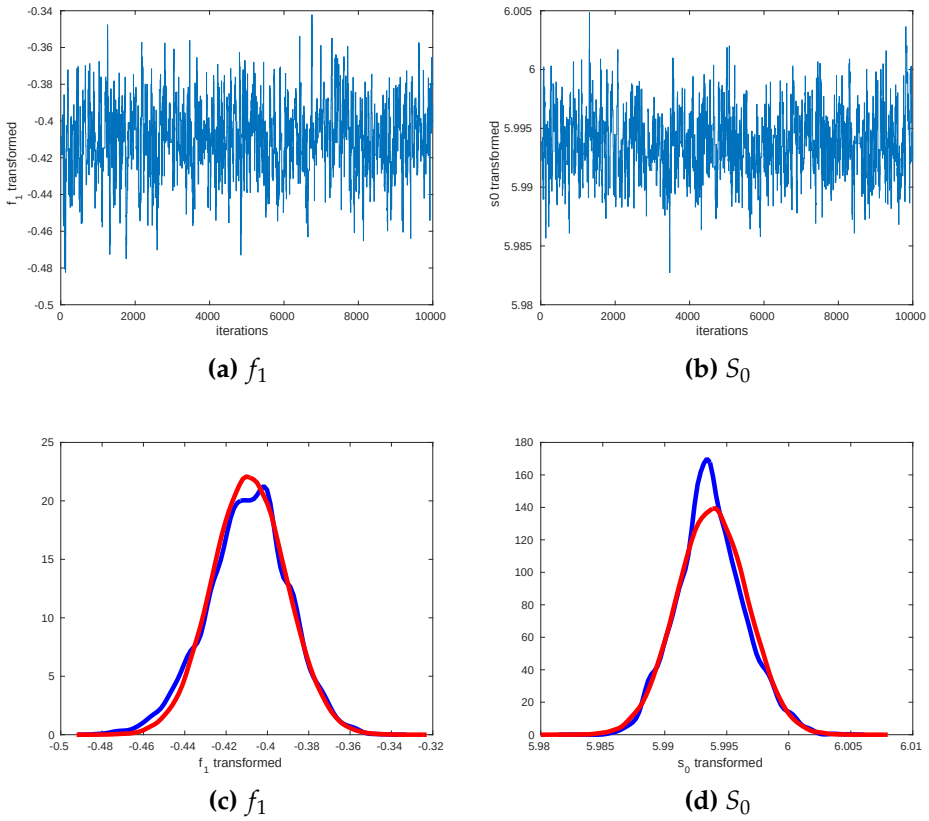
Finally Adaptive MCMC was also implemented. The kernel density plots of the posterior distributions are in Figure 2.24. Although the traceplots do not



(a) f_1 **(b)** S_0

(c) f_1 **(d)** S_0

Figure 2.23: The traceplots and kernel density plots of the MCMC estimates when using the independence sampler MCMC for the parameters of the partial volume model with two fibre orientations (a) f_1 and (b) S_0 when $\tau = 1$.



(a) f_1 **(b)** S_0

(c) f_1 **(d)** S_0

Figure 2.24: The traceplots and kernel density plots of the MCMC estimates using Adaptive MCMC for the partial volume model with two fibre orientations (a) (c) f_1 and (b) (d) S_0 when $\tau = 1$.

seem as good as the results from the independence sampler MCMC they are still satisfactory. Once again the kernel density plots seem to suggest that the Laplace approximation is a good approximation of the posterior distribution of the parameter estimates.

For each of the MCMC algorithms the mean and standard deviation of the parameter estimates can be calculated and are summarised in Table 2.7. The values of the mean are very similar in all the methods, whilst the standard deviation seems to have slightly smaller values in the Block-update MCMC results.

We then compared the MCMC methods by looking at the ACF plots and the values of the ESS. These are in Figure 2.25 and Table 2.8. From these it is clear

	Block-update	Independence	Adaptive
θ_1'	-1.6629 (0.0074)	-1.6624 (0.0079)	-1.6626 (0.0081)
ϕ_1'	-1.1600 (0.0059)	-1.1604 (0.0062)	-1.1602 (0.0065)
f_1'	-0.4088 (0.0185)	-0.4092 (0.0197)	-0.4093 (0.194)
θ_2'	-0.7596 (0.0112)	-0.7596 (0.0116)	-0.7597 (0.0116)
ϕ_2'	-1.6632 (0.0065)	-1.6630 (0.0070)	-1.6633 (0.0068)
f_2'	-1.3982 (0.0275)	-1.3996 (0.0295)	-1.3983 (0.0301)
d'	-6.9088 (0.0033)	-6.9089 (0.0034)	-6.0986 (0.0034)
S_0'	5.9935 (0.0026)	5.9938 (0.0027)	5.9937 (0.0027)

Table 2.7: The mean (and standard deviation) of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC and Adaptive MCMC on the simulated dataset with two fibre orientations when $\tau = 1$.

Samples	Block-update	Independence	Adaptive
10000	392.5688	5400.187	376.6503
9500	373.3775	5441.109	361.7738

Table 2.8: The ESS of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC and Adaptive MCMC on the simulated dataset when $\tau = 1$. The first row is from all 10000 samples, the second row is when the first 500 samples are removed as burn-in.

that the best estimates are obtained using independence sampler MCMC, whilst the estimates obtained using Vanilla MCMC are much worse.

We then simulated a dataset with the same parameters as previously apart from τ which is now $\tau = 0.5$. For each of the three MCMC algorithms the mean and standard deviation of the parameter estimates can be calculated and are summarised in Table 2.9. The means of the estimates appear to be very similar whilst the standard deviation results suggests that the estimates obtained from independence sampler MCMC have a slightly smaller standard deviation.

The ACF results are in Figure 2.26 and the values of the ESS are in Table 2.10. Again the estimates using independence sampler MCMC are the least correlated, whilst the Vanilla MCMC results are the worst.

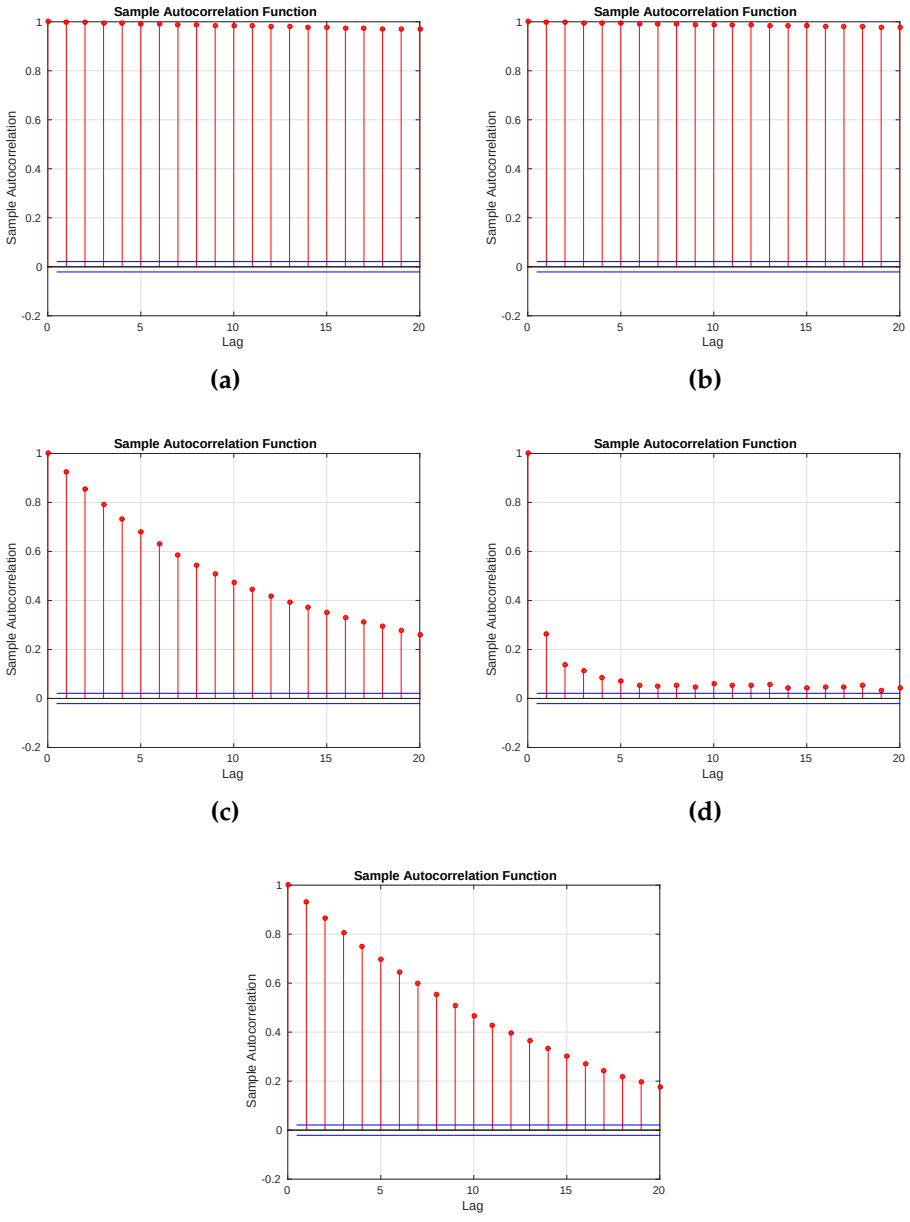


Figure 2.25: The plots of the ACF for the parameter estimates when using Vanilla MCMC for the partial volume model with two fibre orientations (a) f_1 and (b) S_0 and (c) Block-update MCMC, (d) independence sampler MCMC and (e) Adaptive MCMC when $\tau = 1$.

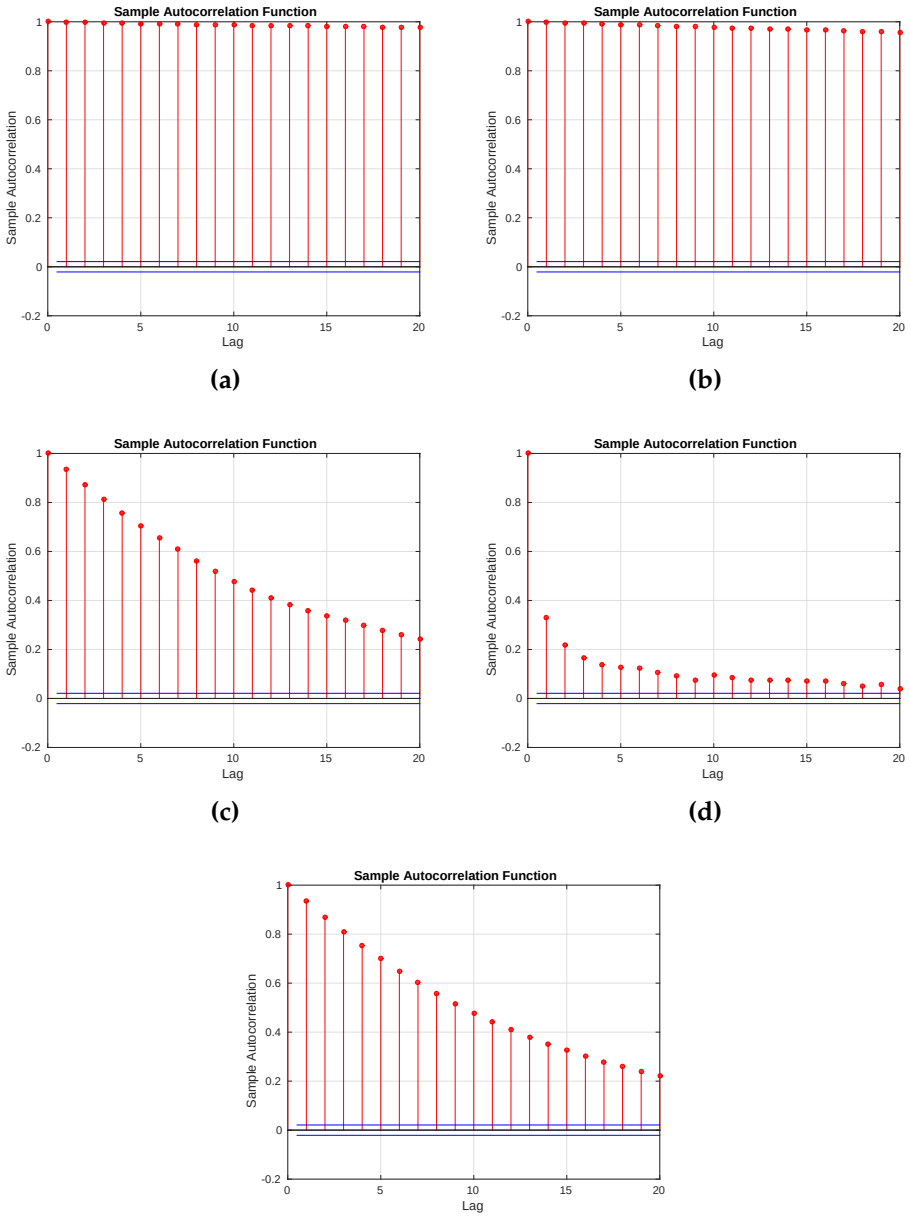


Figure 2.26: The plots of the ACF for the parameter estimates when using Vanilla MCMC for the partial volume model with two fibre orientations (a) f_1 and (b) S_0 (c) Block-update MCMC, (d) independence sampler MCMC and (e) Adaptive MCMC when $\tau = 0.5$.

	Block-update	Independence	Adaptive
θ_1'	-0.7833 (0.0215)	-0.7834 (0.0206)	-0.7804 (0.0213)
ϕ_1'	-1.6531 (0.0127)	-1.6536 (0.0121)	-1.6530 (0.0123)
f_1'	-1.3190 (0.0553)	-1.3185 (0.0526)	-1.3222 (0.0546)
θ_2'	-1.6883 (0.0155)	-1.6889 (0.0151)	-1.6885 (0.0156)
ϕ_2'	-1.1353 (0.0130)	-1.1352 (0.0122)	-1.1359 (0.0125)
f_2'	-0.4516 (0.0383)	-0.4522 (0.0366)	-0.4472 (0.0374)
d'	-6.9022 (0.0066)	-6.9023 (0.0064)	-6.9022 (0.0065)
S_0'	5.9917 (0.0051)	5.9918 (0.0051)	5.9911 (0.0049)

Table 2.9: The mean (and standard deviation) of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC and Adaptive MCMC on the simulated dataset with two fibre orientations when $\tau = 0.5$.

Samples	Block-update	Independence	Adaptive	Vanilla (f)
10000	366.9532	3698.739	362.8368	4.737862
9500	353.0739	3492.721	341.9449	5.538274

Table 2.10: The ESS of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC, Adaptive MCMC and Vanilla MCMC on the simulated dataset when $\tau = 0.5$. The first row is from all 10000 samples, the second row is when the first 500 samples are removed as burn-in.

We then simulated the data with $\tau = 2$. For each of the three MCMC methods the mean and standard deviation of the parameter estimates were calculated and can be summarised in Tables 2.11. There does not seem to be any difference in the means or standard deviations of the estimates from the different methods.

We then compared the different MCMC methods by using the ACF which are shown in Figure 2.27 and the ESS whose values are shown in Table 2.12. From the ACF plots independence sampler MCMC obtains the least correlated estimates whilst Vanilla MCMC obtains the worst results.

In conclusion it seems that independence sampler MCMC gives the best results as its estimates are the least correlated and the mixing is good. The Vanilla MCMC results are very bad, but this makes sense due to us using trial and error to get a good proposal covariance matrix. We have now shown that the MCMC

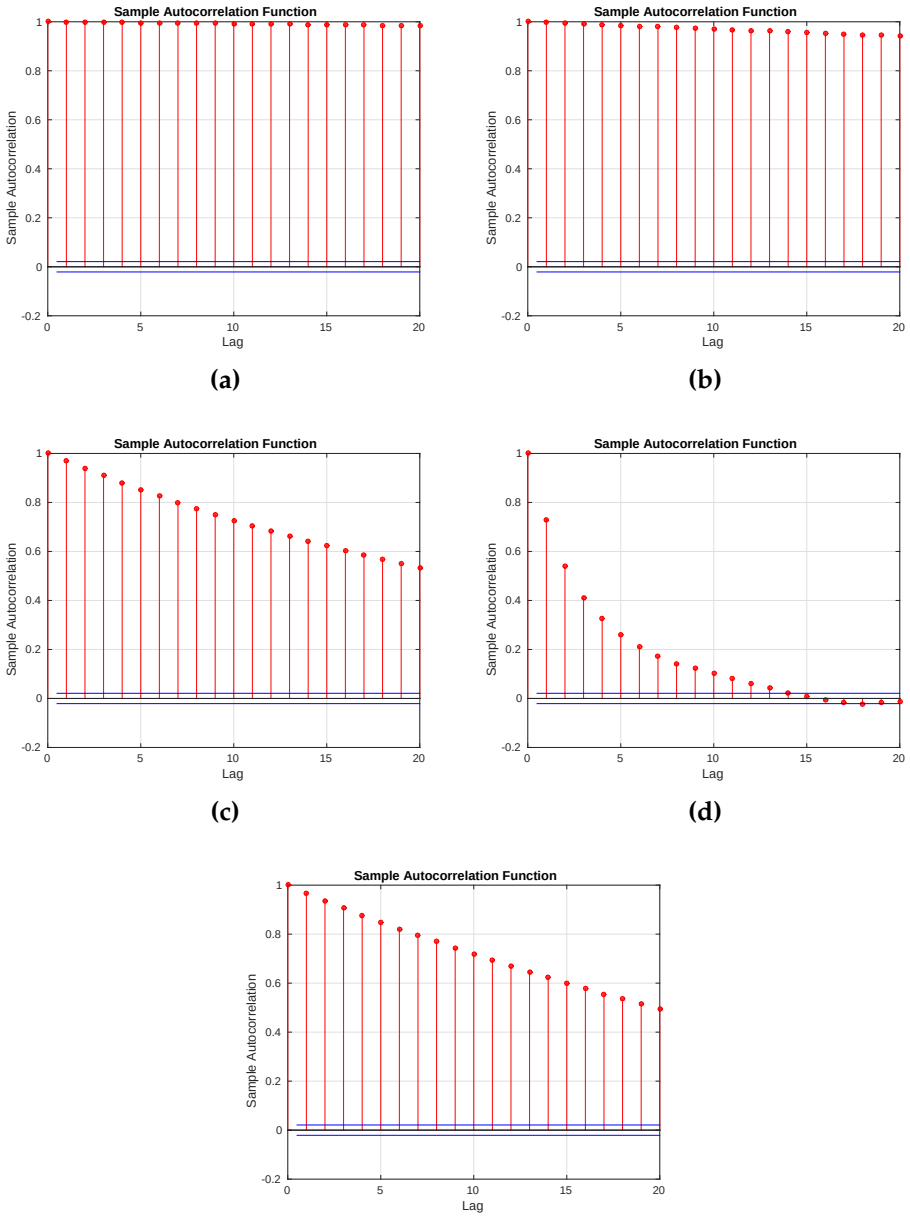


Figure 2.27: The plots of the ACF for the parameter estimates when using Vanilla MCMC for the partial volume model with two fibre orientations (a) f_1 and (b) S_0 (c) Block-update MCMC, (d) independence sampler MCMC and (e) Adaptive MCMC when $\tau = 2$.

	Block-update	Independence	Adaptive
θ_1'	-0.7662 (0.0042)	-0.7658 (0.0045)	-0.7658 (0.0042)
ϕ_1'	-1.6626 (0.0028)	-1.6627 (0.0028)	-1.6626 (0.0026)
f_1'	-1.3821 (0.0110)	-1.3834 (0.0115)	-1.3831 (0.0108)
θ_2'	-1.6650 (0.0031)	-1.6647 (0.0032)	-1.6645 (0.0030)
ϕ_2'	-1.1565 (0.0024)	-1.1568 (0.0025)	-1.1567 (0.0024)
f_2'	-0.4133 (0.0072)	-0.4125 (0.0078)	-0.4127 (0.0074)
d′	-6.9093 (0.0014)	-6.9094 (0.0013)	-6.9095 (0.0013)
S_0'	5.9928 (0.0010)	5.9927 (0.0011)	5.9927 (0.0011)

Table 2.11: The mean (and standard deviation) of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC and Adaptive MCMC on the simulated dataset with two fibre orientations when $\tau = 2$.

Samples	Block-update	Independence	Adaptive	Vanilla (f)
10000	169.1363	1728.894	183.0949	3.442952
9500	157.4462	1654.076	177.3746	3.698462

Table 2.12: The ESS of the transformed parameter estimates of the partial volume model using Block-update MCMC, independence sampler MCMC, Adaptive MCMC and Vanilla MCMC on the simulated dataset when $\tau = 2$. The first row is from all 10000 samples, the second row is when the first 500 samples are removed as burn-in.

methods can be successfully and easily extended to the partial volume model when there is more than one fibre orientation. We revisit the Laplace approximation for the partial volume model with two fibre orientations in Chapter 3.

2.8 Conclusions

We have introduced efficient methods for implementing parameter estimation of both the Diffusion Tensor (DT) model and the more useful partial volume model. First we showed that a good approximation of the parameters in the DT model in the case when we assume that S_0 is known, can be obtained by linearising the model. We can also use MCMC to obtain parameter estimates for this model. Both of these methods produce very good parameter estimates.

The partial volume model will be used more throughout this thesis, therefore

we then tried to infer the parameters in this model when there is one fibre orientation. First Vanilla MCMC was implemented which works but is not very fast. Afterwards to try and make the inference faster Block-update MCMC was used, which is difficult because of choosing a good covariance matrix for the proposal distribution, this problem was solved by using Adaptive MCMC. Adaptive MCMC uses the previous results in MCMC to calculate a proposal covariance matrix at each iteration of MCMC.

A good approximation of the parameters in the partial volume model is found by using the Laplace approximation. This approximation can then be used as either an estimate for the posterior distribution of the parameters or as a good independence sampler in MCMC. The estimates of the parameters from the DT model are used to initialise the Laplace approximation.

Due to the problems that can sometimes occur when using θ and ϕ in the partial volume model, a novel parameterisation which uses a vector for the fibre orientation within a voxel was introduced. Then the Angular Central Gaussian and Bingham distributions can be used as proposal distributions for this vector within MCMC. We showed that both are effective but that the Angular Central Gaussian proposal distribution is better.

All of this was first implemented on simulated datasets where we know the real values of the parameters, then the methods were attempted on real datasets, where the answers were comparable to those obtained by the software package FSL and the mixing appeared to be better in our algorithms.

We then started a simulation study and compared our different MCMC methods, which showed that in general Vanilla MCMC is not very good, however

the other methods are good and in particular independence sampler MCMC obtained very good parameter estimates that are less correlated than the other methods. We then extended the simulation study to start looking at data with more than one fibre orientation which was successfully implemented. All of these methods will help us when inferring the parameters of the Global Tractography model.

CHAPTER 3

Model selection within voxel

3.1 Motivation

In Chapter 2 efficient methods were proposed to estimate the values of the local parameters of the partial volume model in a voxel. Another problem that remains is to infer how many fibre orientations there are within one voxel. If we incorrectly guess the number of fibre orientations within a voxel then the local parameter estimates of the partial volume model may be very inaccurate. This could then lead to very bad consequences in the Global Tractography model (see Section 1.7.3). The aim of this chapter is to develop methods for selecting the number of fibre orientations within a voxel, so that once the parameters of the partial volume model in a voxel are estimated, we can be confident that the estimates are reflective of the truth.

When Tractography methods (Section 1.7) were first introduced, it was assumed that there was only one dominant fibre orientation in each voxel (Behrens *et al.*, 2003). Behrens *et al.* (2007), estimated that approximately one-third of all brain voxels with a Fractional Anisotropy value (Section 1.5) greater than 0.1 contain a crossing-fibre configuration. Attempts have been made to extend Tractography such that multiple fibre orientations are permitted within a voxel

(Behrens *et al.*, 2007). This was implemented using both Deterministic Tractography (Section 1.7.1) within Diffusion Spectrum Imaging (Hagmann *et al.*, 2004) and in Probabilistic Tractography (Section 1.7.2) using High Angular Resolution Diffusion Imaging (Hosey *et al.*, 2005) and the angular structure of cerebral tissue (Parker and Alexander, 2005). However the most widely used current Tractography method uses the partial volume model (see Section 1.6.1) and was introduced by Behrens *et al.* (2007).

The partial volume model allows us to have multiple fibre orientations in a voxel. The methods from Chapter 2 can help us in inferring the parameters of the partial volume model. The local parameter estimates that are obtained can then be used within a Tractography framework to construct tracts between different brain regions (see Section 1.7). If we do not consider models with multiple fibre orientations then this could cause us to produce tracts that are not true due to the model not being representative of the true fibre orientation. By recognising the existence of multiple fibre orientations, known connections in the brain that were previously not reconstructed using Tractography methods have been identified (Behrens *et al.*, 2007).

In existing Tractography methods it is assumed that the number of fibres within a voxel is known. Then at each stage of Tractography the fibre is selected by considering the fraction of the signal contributed by each of the fibres in a voxel. Therefore we consider the estimated values of the ith fibre, f_i. The existing methods do not take into account the uncertainty in the number of fibres. A Tractography method which we term as *Fully Probabilistic Tractography* is introduced in this chapter. It differs from Probabilistic Tractography by also selecting the number of fibres to model in the partial volume model at each step based on the probability of choosing the model with that many fibre orientations. This has the advantage that it may reconstruct tracts that would otherwise be missed

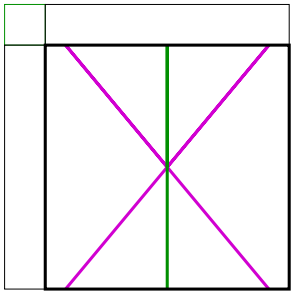


Figure 3.1: An example of two equally weighted fibres (pink) in a voxel, and the resulting direction that may be estimated in the partial volume model with one fibre orientation (green).

by existing Tractography methods.

For example it is possible that there are two fibres that cross each other and are both equally weighted as shown in Figure 3.1, so that when implementing parameter estimation, due to partial volume effects (Parker, 2011), one fibre may fit the model well, but in practice will not be the true fibre orientation. Thus when Tractography is performed then the wrong model is fitted and the Tractography results will be misleading. By taking into account model uncertainty, tracts involving the two true fibres will also be reconstructed. Schultz *et al*. (2013), discuss sources of error in Tractography and the fact that model selection has not received much attention with respect to its effect on Tractography results.

3.2 Bayesian model choice and Bayes factor

This section is a review of Bayesian solutions for implementing model selection. In Section 3.3 we will use Bayesian methods to choose between partial volume models with different numbers of fibre orientations. From these model selection estimates, the probability of choosing a certain model can be calculated (see Section 3.6), such that model uncertainty can also be taken into account when implementing Tractography.

Bayes factors are used to obtain a measure to decide which of two proposed models is a better fit for some data. In particular suppose that there are two models denoted M_0 and M_1. The Bayes factor, here denoted by K, for model M_0 versus M_1 is defined as

$$K = \frac{\pi(\boldsymbol{y}|M_0)}{\pi(\boldsymbol{y}|M_1)}.$$

$\pi(\boldsymbol{y}|M_0)$ and $\pi(\boldsymbol{y}|M_1)$, represent the evidence (or in other words the marginal likelihood) in favour of M_0 and M_1 respectively. If K>1, then the model M_0 is more likely than model M_1, while if K<1 then model M_1 is more likely than M_0. Alternatively the Bayes factor can be written in terms of the posterior odds and prior odds such that

$$K = \frac{\pi(M_0|\mathbf{y})}{\pi(M_1|\mathbf{y})} \frac{\pi(M_1)}{\pi(M_0)}$$

where $\pi(M_0|\mathbf{y})$ and $\pi(M_1|\mathbf{y})$ are the posterior probabilities of M_0 and M_1. The Bayes factor is often difficult to calculate because it requires the marginal likelihoods $\pi(\boldsymbol{y}|M_0)$ and $\pi(\boldsymbol{y}|M_1)$. The marginal likelihood for model M_k and observed data $\boldsymbol{y}$ is given as

$$\pi(\boldsymbol{y}|M_k) = \int_{\theta_k} \pi(\boldsymbol{y}|\boldsymbol{\theta}_k, M_k)\pi(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k$$

where $\boldsymbol{\theta}_k$ are the parameters that are in model M_k. In many cases $\boldsymbol{\theta}_k$ can be high-dimensional and therefore the calculation of such an integral can be difficult.

In order to take the model uncertainty into account a method that provides estimates of $\pi(M_k|\boldsymbol{y})$ is required. By calculating $\pi(M_k|\boldsymbol{y})$ and using this to choose which model to look at, we can take into account model uncertainty within Tractography. One disadvantage of Bayes factors are that they can be very sensitive to the choice of prior distribution (Kass and Raftery, 1995).

3.2.1 Reversible Jump Markov Chain Monte Carlo

One method that is commonly used in the literature to estimate $\pi(M_k|\boldsymbol{y})$ is Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995). This can be used to sample from the posterior distribution

$$\pi(\boldsymbol{\theta}_k, k|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta}_k, k)\pi(\boldsymbol{\theta}_k|k)\pi(k)$$

where k is some parameter that indicates which model we are in so that we are sampling from the joint posterior distribution of model indicators and parameters. In more detail $\boldsymbol{\theta}_k$ are the parameters that are specific to model k whilst $\boldsymbol{\theta}$ is the collection of all model parameters. $\pi(\boldsymbol{\theta}_k|k)$ is the prior distribution for parameters within model k, $\pi(k)$ is the prior for the model indicator and $\pi(\boldsymbol{y}|\boldsymbol{\theta}_k, k)$ is the likelihood for the data.

RJMCMC is an extension of the Metropolis-Hastings algorithm which moves within models whilst allowing us to jump from one model to another such that we have samples of $(\boldsymbol{\theta}_k, k)$. RJMCMC works by introducing some distribution, g that we can simulate random numbers $\boldsymbol{u}$ from and another distribution g^* that we can simulate random numbers $\boldsymbol{u}^*$ from. Then a deterministic function that allows us to move from $(\boldsymbol{\theta}_k, k)$ to $(\boldsymbol{\theta}_l, l)$ is denoted as

$$(\boldsymbol{\theta}_l, l) = f_{kl}(\boldsymbol{\theta}_k, \boldsymbol{u}, k).$$

Similarly

$$(\boldsymbol{\theta}_k, k) = f_{lk}(\boldsymbol{\theta}_l, \boldsymbol{u}^*, l),$$

f_{kl} must be a bijection and its derivative must be invertible (Friel and Pettitt, 2008). This only holds if

$$\dim(\boldsymbol{\theta}_k) + \dim(\boldsymbol{u}) = \dim(\boldsymbol{\theta}_l) + \dim(\boldsymbol{u}^*).$$

Then by denoting the probability of moving from model l to model k as $\pi(l \rightarrow k)$ the probability of accepting a proposed move from $(\boldsymbol{\theta}_k, k)$ to $(\boldsymbol{\theta}_l, l)$ is

$$\min \left\{ 1, \frac{\pi(\boldsymbol{\theta}_l, l|\boldsymbol{y})\pi(l \rightarrow k)g^*(u^*)}{\pi(\boldsymbol{\theta}_k, k|\boldsymbol{y})\pi(k \rightarrow l)g(u)}|J| \right\}$$

where J is the Jacobian that is required because of the transformation from (θ_k, u, k) to (θ_l, u^*, l).

RJMCMC can be an efficient algorithm with good mixing properties. However the model mixing across dimensions which requires choosing efficient proposal distributions and mappings (Friel and Pettitt, 2008) can cause problems. In our partial volume model problem we have many voxels and therefore many parameters to infer. Therefore it will be even more difficult to choose efficient proposal distributions and mappings in RJMCMC. Thus we focus on calculating the Bayes factor by investigating approximations to the marginal likelihood.

3.2.2 Importance sampling estimators

Kass and Raftery (1995) proposed importance sampling estimators as a way of estimating the marginal likelihood. This approach assumes that we have access

to an unnormalised density $g(\boldsymbol{\theta}_k)$ that we can sample from. Then we can derive

$$\begin{aligned}
\pi(\boldsymbol{y}|M_k) &= \int_{\boldsymbol{\theta}_k} \pi(\boldsymbol{y}|\boldsymbol{\theta}_k, M_k)\pi(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k \\
&= \frac{\int_{\boldsymbol{\theta}_k} \pi(\boldsymbol{y}|\boldsymbol{\theta}_k, M_k)\pi(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k}{\int_{\boldsymbol{\theta}_k} \pi(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k} \\
&= \frac{\int_{\boldsymbol{\theta}_k} \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}_k,M_k)\pi(\boldsymbol{\theta}_k|M_k)}{g(\boldsymbol{\theta}_k)} g(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k}{\int_{\boldsymbol{\theta}_k} \frac{\pi(\boldsymbol{\theta}_k|M_k)}{g(\boldsymbol{\theta}_k)} g(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k} \\
&= \frac{E_g\left[\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}_k,M_k)\pi(\boldsymbol{\theta}_k|M_k)}{g(\boldsymbol{\theta}_k)}\right]}{E_g\left[\frac{\pi(\boldsymbol{\theta}_k|M_k)}{g(\boldsymbol{\theta}_k)}\right]}.
\end{aligned}$$

By obtaining a sample of size J, which we denote $\boldsymbol{\theta}_k^1, \boldsymbol{\theta}_k^2, ..., \boldsymbol{\theta}_k^{J-1}, \boldsymbol{\theta}_k^J$, from $g(\boldsymbol{\theta}_k)$ the expectations of the denominator and numerator can be approximated using Monte Carlo methods such that

$$\pi(\boldsymbol{y}|M_k) \simeq \frac{\sum_{j=1}^{J} \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}_k^j,M_k)\pi(\boldsymbol{\theta}_k^j|M_k)}{g(\boldsymbol{\theta}_k^j)}}{\sum_{j=1}^{J} \frac{\pi(\boldsymbol{\theta}_k^j|M_k)}{g(\boldsymbol{\theta}_k^j)}}. \tag{3.2.1}$$

Then the only thing that must be chosen in advance for this method is the unnormalised density $g(\boldsymbol{\theta}_k)$. The simplest case would be to choose it to be the prior distribution such that

$$g(\boldsymbol{\theta}_k^j) = \pi(\boldsymbol{\theta}_k^j|M_k).$$

Then Equation (3.2.1) becomes

$$\pi(\boldsymbol{y}|M_k) \simeq \frac{1}{J}\sum_{j=1}^{J} \pi(\boldsymbol{y}|\boldsymbol{\theta}_k^j, M_k).$$

This approximation is called the *prior arithmetic mean estimator*. However if the high-likelihood region is small then this estimator is not good as then the prior

will not produce many samples from this region. Therefore to attempt to overcome this problem the *posterior harmonic mean estimator* was proposed. Newton and Raftery (1994), suggested drawing samples from the posterior such that

$$g(\boldsymbol{\theta}_k^j) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta}_k^j, M_k)\pi(\boldsymbol{\theta}_k^j|M_k).$$

If we put this $g(\boldsymbol{\theta}_k^j)$ into Equation (3.2.1), the following estimate is obtained.

$$\begin{aligned}\pi(\boldsymbol{y}|M_k) &\simeq \frac{\sum_{j=1}^{J} 1}{\sum_{j=1}^{J} \frac{\pi(\boldsymbol{\theta}_k^j|M_k)}{\pi(\boldsymbol{\theta}_k^j|M_k)\pi(\boldsymbol{y}|\boldsymbol{\theta}_k^j,M_k)}} \\ &= \frac{J}{\sum_{j=1}^{J} \frac{1}{\pi(\boldsymbol{y}|\boldsymbol{\theta}_k^j,M_k)}}.\end{aligned}$$

Then

$$\frac{1}{\pi(\boldsymbol{y}|M_k)} \simeq \frac{1}{J}\sum_{j=1}^{J} \frac{1}{\pi(\boldsymbol{y}|\boldsymbol{\theta}_k^j, M_k)}. \tag{3.2.2}$$

If we use the Laplace approximation as an approximate density for $g(\boldsymbol{\theta}_k^j)$, then we may take samples from this approximate density and use these to calculate an estimate for Equation 3.2.2.

It has been shown that in some situations the posterior harmonic mean estimator has a variance which is infinite (Kass and Raftery, 1995) and therefore it is not used often in practice. We propose an alternative way of estimating the marginal likelihood in Section 3.4.

In this section we have done a review of the various Bayesian solutions for implementing model selection. We will focus on approximating the Bayes factor in Section 3.4 for the number of fibre orientations in the partial volume model.

3.3 Model selection methods for the number of fibre orientations

We will now extend the partial volume model so that it allows for more than one fibre orientation. Then an existing method that enables us to choose between the number of fibre orientations in a voxel will be discussed.

3.3.1 The partial volume model with multiple fibre orientations

Parameter estimation within a voxel assuming that there is only one fibre orientation can be done efficiently (i.e. quickly and accurately). We extend the methods for parameter estimation that were developed in Chapter 2 in the case where there is more than one fibre orientation, which we briefly visited in Section 2.7.2. In the partial volume model with more than one fibre orientation, the value of the predicted Diffusion-Weighted signal, μ_i, changes and is

$$\mu_i = S_0 \left(\left(1 - \sum_{j=1}^{N} f_j \right) \exp(-b_i d) + \sum_{j=1}^{N} f_j \exp\left(-b_i d (\boldsymbol{g}_i^T \boldsymbol{v}_j)^2 \right) \right), \quad i = 1, ..., m,$$

where $\boldsymbol{v}_j$ is the jth fibre orientation, such that the parameters that represent the fibre orientations within a voxel are θ_1 ,.., θ_N ,ϕ_1 ,.., ϕ_N. The likelihood of the observed data is similar to the likelihood in the case of one fibre orientation (Section 2.3) with the only difference being that the predicted Diffusion-Weighted signal, μ_i, is as above. By the definition of the partial volume model in Section 1.6.1, f_j is the fraction of the signal contributed by the jth fibre. We denote

$$f_0 = 1 - \sum_{j=1}^{N} f_j$$

so that

$$\sum_{j=0}^{n} f_j = 1.$$

Each of the f_j must be positive but the sum of the f_j can not exceed 1. Therefore

$$0 < \sum_{j=1}^{n} f_j < 1.$$

If there are two fibre orientations, then

$$0 < f_1 + f_2 < 1.$$

If we assume without loss of generality that f_1 is greater than or equal to f_2, then we can write

$$0 < f_2 \leq f_1 \leq 1 - f_2 < 1.$$

If the likelihood of the observed data was maximised then transformations of f_1 and f_2 would have to be used to ensure that

$$0 < f_2 \leq f_1 < 1.$$

The following transformations were used

$$f_1' = \log\left(\frac{f_1}{1 - f_1}\right)$$

and

$$f_2' = \log\left(\frac{f_2}{f_1 - f_2}\right).$$

Samples where $f_1 + f_2 > 1$ are rejected within the MCMC algorithm.

Similar transformations were found in the case of $N \geq 3$ so if the fibres are ordered such that

$$f_N < f_{N-1} < .... < f_2 < f_1$$

then

$$f_1' = \log\left(\frac{f_1}{1-f_1}\right)$$

and

$$f_i' = \log\left(\frac{f_i}{f_{i-1}-f_i}\right), \quad i = 1, ..., N-1.$$

3.3.2 Estimating the fibres in the partial volume model

When looking at real data, a method is needed to decide how many fibre orientations to have in each voxel. One way would be to model each voxel using a varying amount of fibre orientations, then decide how many fibre orientations to have, by using the Bayesian information criterion (BIC) (Schwarz, 1978) or Akaike information criterion (AIC) (Akaike, 1983). These are both asymptotic approximations that rely on sufficiently large sample sizes, where the interpretation of sufficiently large is not defined (Kass and Raftery, 1995). The AIC has been shown to overestimate the number of parameters needed (Shibata, 1976, Katz, 1981), while the BIC favours simpler models. However Findley (1991) shows examples where the AIC works but not the BIC. For these reasons the AIC and BIC are not very reliable estimates for inferring the number of fibre orientations in a voxel; given that we only have 61 data points in general (Section 1.4).

A solution that was proposed to implement model selection within the partial volume model is Automatic Relevance Determination (ARD) (Behrens *et al.*, 2007). This is the method implemented in FSL (see Section 1.7.2).

3.3.3 Automatic Relevance Determination prior

Automatic Relevance Determination (ARD) works by first fitting the most complex model, and then forces any parameters that the data does not support to zero. It does this by forcing the parameter's value to zero in the posterior distribution with a very low variance, by using sparsity induced priors (MacKay, 1995). In the partial volume model these sparsity induced priors could be used on the parameter f_i to determine whether the ith fibre contributes to the signal. Each parameter that is being considered, can have a prior distribution, whose variance is unknown and whose mean is zero. If the variance is inferred to be very low it will ensure that the parameter is forced to zero. If the variance is large then it will allow the parameter to take any value. We will now investigate ARD when it is applied to the partial volume model.

3.3.4 ARD applied to the partial volume model

We now illustrate how to impose an ARD prior for the parameters of the partial volume model. In particular we need to assign a prior for $f_i, \quad i \geq 2$ because we assume that at least one fibre orientation exists. We will also ensure that the sum of the f_is is less than one by rejecting samples of f_i where this constraint doesn't hold true. The prior for $f_i, \quad i \geq 2$ is a Beta distribution with parameters Beta$(1, \eta)$. This distribution has its mode at zero. Then

$$f_i|\eta \sim \text{Beta}(1, \eta)$$
$$\pi(\eta) \propto \eta^{-1}$$

where η follows an improper prior distribution with density proportional to $1/\eta$. The imposed prior on f_i is then derived by integrating η out.

$$\begin{aligned}
\pi(f_i) &= \int_0^\infty \pi(f_i|\eta)\pi(\eta)d\eta \\
&= \int_0^\infty \frac{f_i^{1-1}(1-f_i)^{\eta-1}}{\eta\beta(1,\eta)}d\eta \\
&= \int_0^\infty (1-f_i)^{\eta-1}d\eta \\
&= \left[\frac{(1-f_i)^{\eta-1}}{\log(1-f_i)}\right]_0^\infty \\
&= \frac{-1}{(1-f_i)\log(1-f_i)}
\end{aligned}$$

where $\beta(1,\eta)$ is the beta function. Then

$$\pi(f_i) = -\frac{1}{(1-f_i)\log(1-f_i)}.$$

When we plot the prior density for different values of f_i, it can be seen that the prior takes very high values when f_i gets closer to 0 and slightly high values when f_i gets closer to 1, as in Figure 3.2.

The single-component Metropolis-Hastings MCMC algorithm can then be implemented as in Section 1.8.1 to infer on the local parameters of the new posterior distribution that takes into account the ARD priors for f_i when $i \geq 2$. The other priors are the same as them in the partial volume model (see Section 2.3).

Behrens *et al*. (2007) suggests to threshold the f_i at 0.05 within the Probabilistic Tractography algorithm, such that if f_i is less than 0.05 then we assume that the corresponding fibre is not there. The value of 0.05 is arbitrary and this is the de-

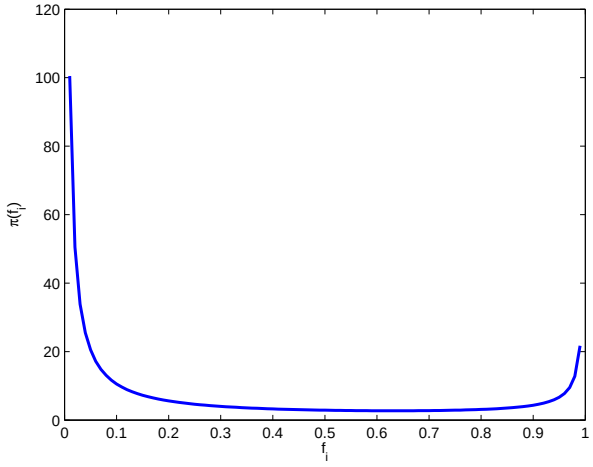


Figure 3.2: The implied ARD prior density for f_i, $\pi(f_i)$ in the partial volume model.

fault value in the FSL software (Section 1.7.2). Then all other fibres that remain are considered and within a voxel we choose the fibre orientation that is most parallel to the previous voxel's fibre orientation to continue the tract. Thus the weights of the fibres are ignored and also model uncertainty is ignored.

3.3.5 Examples of ARD applied to the partial volume model

Data were simulated from the partial volume model with one fibre orientation, i.e. $f_2 = 0$. We fitted a model with two fibre orientations and imposed an ARD prior on f_2. The graph in Figure 3.3 is the traceplot of the MCMC estimated values of f_2. The MCMC mixing for f_2 appears to be particularly bad, although at some iterations it does seem to be suggesting that $f_2 = 0$ because the accepted samples of f_2 are close to 0. The conclusions of whether the second fibre orientation exists is very much dependent on the arbitrary threshold that is chosen for f_2.

Data were simulated with two fibre orientations such that θ_1=0.5, θ_2=0.5, ϕ_1=1, ϕ_2=5, f_1=0.5, f_2=0.3, d=0.0015, S_0=1.0 and τ=1000. Then the ARD prior was used

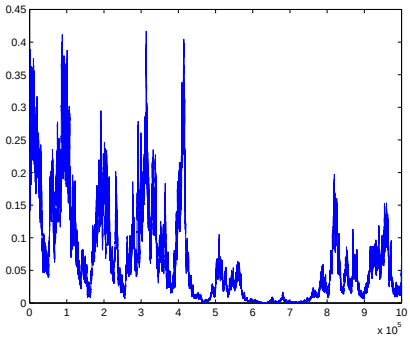
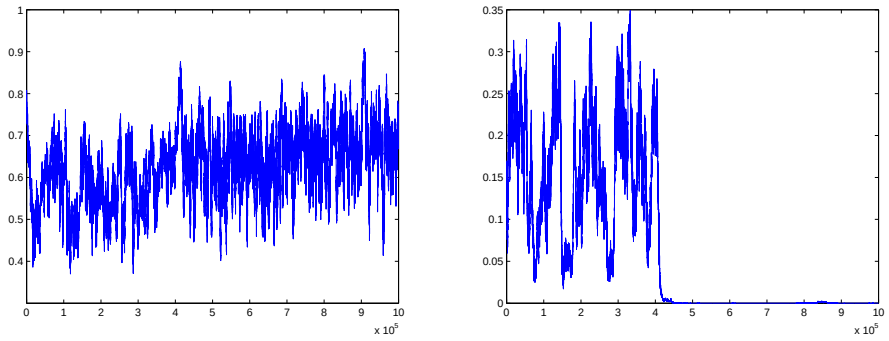


Figure 3.3: Traceplot of the parameter estimates from the partial volume model f_2 simulated from the posterior distribution when using ARD priors within MCMC on a simulated dataset where $f_2 = 0$.



(a) **(b)**

Figure 3.4: Traceplots of the parameter estimates from the partial volume model (a) f_1 and (b) f_2 simulated from the posterior distribution when using ARD priors within MCMC on a simulated dataset where $f_2 \neq 0$.

on the parameter f_2. The ARD method did not seem to work as demonstrated in Figure 3.4. As we can see in the graphs the posterior implied by the ARD prior is very bad at estimating the value of f_2 in this case.

Although ARD is the method that is widely used to determine the number of fibre orientations within a voxel, it has many issues. Firstly a value has to be proposed where we decide whether f_i is large enough to be included as a fibre or not, the choice of the threshold is arbitrary. Furthermore the mixing in random-walk MCMC can often be problematic when using the ARD prior. Finally the parameter that we use the ARD prior on is constrained to be within

(0,1). The ARD is therefore not ideal and this was demonstrated by the examples in this section. Therefore we consider alternative methods in the section below.

3.4 Marginal likelihood estimation using Thermodynamic Integration

In the DW-MRI literature there have been attempts to use model selection methods such as in Freidlin *et al*. (2007) using the Bayesian Information Criteria for comparing Diffusion Tensor models with simpler models and also Bretthorst *et al*. (2004) to choose between models in baboon brains, but there does not seem to be widely known attempts to do Bayesian model selection between partial volume models with differing numbers of fibre orientations. Therefore in this section we will investigate methods based on Thermodynamic Integration as a way of estimating the marginal likelihood.

Recent work by Lartillot and Philippe (2006) and Friel and Pettit (2008), showed that Thermodynamic Integration which originated from the physics community to compute the free energy difference between two molecular-dynamic systems (Gelman and Meng, 1998) is one of the most promising methods for model selection when compared with other available methods. Despite being promising they have not been used yet in the context of the DW-MRI methods. The approaches are first illustrated with a basic example that has a known analytical solution. They are then extended to compare the partial volume models with varying numbers of fibres.

3.4.1 Annealing-Melting Integration

A method that is based on ideas from Thermodynamic Integration was introduced by both Lartillot and Philippe (2006) and Friel and Pettit (2008) independently. Following the presentation of Friel and Pettit (2008), by introducing a temperature parameter which we denote t, where $t \in [0, 1]$, we define the power posterior as

$$\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}) \tag{3.4.1}$$

$$\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \pi(\boldsymbol{y}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

where $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood, $\pi(\boldsymbol{\theta})$ is the prior distribution and the normalising constant of $\pi_t(\boldsymbol{\theta}|\boldsymbol{y})$ is

$$z(\boldsymbol{y}|t) = \int_{\boldsymbol{\theta}} \pi(\boldsymbol{y}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{3.4.2}$$

From Equation (3.4.2) we derive

$$z(\boldsymbol{y}|t = 0) = \int_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1$$

and

$$z(\boldsymbol{y}|t = 1) = \int_{\boldsymbol{\theta}} \pi(y|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

which is the marginal likelihood of the data.

Then the following identity can be derived

$$\log(\pi(\boldsymbol{y})) = \log\left(\frac{z(\boldsymbol{y}|t = 1)}{z(\boldsymbol{y}|t = 0)}\right) = \int_0^1 E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))] dt \tag{3.4.3}$$

as follows

$$\begin{aligned}
\frac{d}{dt}\log(z(\boldsymbol{y}|t)) &= \frac{1}{z(\boldsymbol{y}|t)}\frac{d}{dt}z(\boldsymbol{y}|t) \\
&= \frac{1}{z(\boldsymbol{y}|t)}\frac{d}{dt}\int_{\boldsymbol{\theta}} \pi(\boldsymbol{y}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \frac{1}{z(\boldsymbol{y}|t)}\int_{\boldsymbol{\theta}} \pi(\boldsymbol{y}|\boldsymbol{\theta})^t \log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))\pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\theta}} \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})^t \pi(\boldsymbol{\theta})}{z(\boldsymbol{y}|t)} \log(\pi(\boldsymbol{y}|\boldsymbol{\theta})) d\boldsymbol{\theta} \\
&= E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))]
\end{aligned}$$

such that

$$\frac{d}{dt}\log(z(\boldsymbol{y}|t)) = E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))]. \tag{3.4.4}$$

Then by integrating Equation (3.4.4) with respect to t we obtain the identity in Equation (3.4.3). This identity can be used to approximate the logarithm of the marginal likelihood.

Friel and Pettit (2008) suggested two methods for approximating the identity in Equation (3.4.3). One approach involves obtaining estimates of both $\boldsymbol{\theta}$ and t from $\pi(\boldsymbol{\theta}, t|\boldsymbol{y})$ by treating them both as random variables. An alternative method first runs separate chains for different values of t. Values of $\boldsymbol{\theta}$ are drawn from the power posterior in Equation 3.4.1. Then these estimates are used along with the trapezoidal rule over t to approximate $\log(\pi(\boldsymbol{y}))$. If t is discretised such that $0 = t_0 < t_1 < t_2 < ... < t_{n-1} < t_n = 1$ then the approximation is

$$\log(\pi(\boldsymbol{y})) \approx \sum_{i=0}^{n-1}(t_{i+1} - t_i)\frac{E_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))] + E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))]}{2} \tag{3.4.5}$$

where

$$E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))] = \frac{1}{p-k+1}\sum_{j=k}^{p}\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}_j^i)) \tag{3.4.6}$$

and $\boldsymbol{\theta}_j^i$ is the jth MCMC estimate of the power posterior from the ith temperature. The MCMC will run for p iterations for $n+1$ different temperatures.

Friel *et al*. (2013) recently proposed a method for calculating a more accurate estimate to the logarithm of the marginal likelihood when using Annealing-Melting Integration. In Equation (3.4.5) the trapezoidal rule is used to approximate the logarithm of the marginal likelihood. Instead the corrected trapezium rule (Atkinson and Han, 2004), can be used to obtain a more accurate approximation with hardly any extra computational cost. The corrected trapezium rule calculates an approximation to the integral of some function f between points a and b as follows

$$\int_a^b f(x)dx = (b-a)\left[\frac{f(b)+f(a)}{2}\right] - \frac{(b-a)^3}{12}f''(c),$$

where c is some point in the interval $[a,b]$. The first part of the approximation is the same as the normal trapezium rule. We can use the fact that

$$f''(c) \approx \frac{f'(b)-f'(a)}{b-a}$$

to obtain the corrected trapezium rule which is

$$\int_a^b f(x)dx \approx (b-a)\left[\frac{f(b)+f(a)}{2}\right] - \frac{(b-a)^2}{12}[f'(b)-f'(a)]. \tag{3.4.7}$$

Previously we needed to calculate Equation (3.4.3) to obtain the approximation of the logarithm of the marginal likelihood. Instead if we wanted to approximate Equation (3.4.3) using Equation (3.4.7) then $f(x) = E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))]$, $a = 0$ and $b = 1$. By discretising the temperatures t we obtained Equation

(3.4.5) for calculating the logarithm of the marginal likelihood. Similarly for the corrected trapezium rule

$$\log(\pi(\boldsymbol{y})) \approx \sum_{i=0}^{n-1}(t_{i+1}-t_i)\left[\frac{E_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}\left[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))\right]+E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))]}{2}\right] - \sum_{i=0}^{n-1}\frac{(t_{i+1}-t_i)^2}{12}\left[\frac{dE_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}\left[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))\right]}{dt}-\frac{dE_{\boldsymbol{\theta}|\boldsymbol{y},t_i}\left[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))\right]}{dt}\right].$$

Friel *et al*. (2013) show that

$$\frac{d}{dt}E_{\boldsymbol{\theta}|\boldsymbol{y},t}\log(\pi(\boldsymbol{y}|\boldsymbol{\theta})) = V_{\boldsymbol{\theta}|\boldsymbol{y},t}(\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))),$$

where $V_{\theta|\boldsymbol{y},t}(\log(\pi(\boldsymbol{y}|\theta)))$ is the variance of the logarithm of the marginal likelihood with respect to the power posterior. This variance can be estimated at minimal computational cost from the MCMC output.

Annealing-Melting Integration can be used to approximate the marginal likelihood, additionally we can use the corrected trapezium rule to obtain a more accurate approximation. We will now describe another method based on Thermodynamic Intregration for approximating the marginal likelihood.

3.4.2 Importance Power Posterior

In the power posterior of the Annealing-Melting Integration method (Section 3.4.1) we could see that as we change temperatures we are slowly moving from the prior when $t = 0$ to the posterior when $t = 1$. However the prior may be very diffuse, therefore we introduce the power posterior in this section such that we are moving from the proposal distribution $q(\boldsymbol{\theta})$ when $t = 0$ to the posterior when $t = 1$. $q(\boldsymbol{\theta})$ needs to be a good approximation of the posterior distribution so we choose it to be some multivariate normal distribution where the parameters are estimated using the Laplace approximation (Section 2.3.4). We now

introduce a novel modification to the power posterior method such that we are defining a new power posterior. We define

$$\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) \propto (\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))^t (q(\boldsymbol{\theta}))^{1-t}.$$

where $q(\boldsymbol{\theta})$ is some distribution.

We derive the estimate to the logarithm of the marginal likelihood by first noting that

$$z(\boldsymbol{y}|t) = \int_{\boldsymbol{\theta}} (\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))^t q(\boldsymbol{\theta})^{1-t} d\boldsymbol{\theta}$$

and then

$$
\begin{aligned}
& \frac{d}{dt}\log(z(\boldsymbol{y}|t)) \\
= \; & \frac{1}{z(\boldsymbol{y}|t)}\frac{d}{dt}z(\boldsymbol{y}|t) \\
= \; & \frac{1}{z(\boldsymbol{y}|t)}\frac{d}{dt}\int_{\boldsymbol{\theta}}(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))^t q(\boldsymbol{\theta})^{1-t}d\theta \\
= \; & \frac{1}{z(\boldsymbol{y}|t)}(\int_{\boldsymbol{\theta}}(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))^t \log(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))q(\boldsymbol{\theta})^{1-t}d\boldsymbol{\theta} \\
& + \int_{\boldsymbol{\theta}}(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))^t(-q(\boldsymbol{\theta}))^{1-t}\log(q(\boldsymbol{\theta}))d\boldsymbol{\theta}) \\
= \; & \int_{\boldsymbol{\theta}}\frac{(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))^t q(\boldsymbol{\theta})^{1-t}}{z(\boldsymbol{y}|t)}\log(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))d\boldsymbol{\theta} \\
& - \int_{\boldsymbol{\theta}}\frac{(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))^t q(\boldsymbol{\theta})^{1-t}}{z(\boldsymbol{y}|t)}\log(q(\boldsymbol{\theta}))d\boldsymbol{\theta} \\
= \; & E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))] - E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(q(\boldsymbol{\theta}))] \\
= \; & E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})) - \log(q(\boldsymbol{\theta}))] \\
= \; & E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right].
\end{aligned}
$$

Then by integrating both sides with respect to t, we obtain the identity

$$\log(\pi(\boldsymbol{y})) = \log\left(\frac{z(\boldsymbol{y}|t=1)}{z(\boldsymbol{y}|t=0)}\right) = \int_0^1 E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right]dt.$$

Then using the same approach as Annealing-Melting Integration (Section 3.4.1) we obtain

$$\log(\pi(\boldsymbol{y})) \approx \sum_{i=0}^{n-1}(t_{i+1}-t_i)\frac{E_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right] + E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right]}{2}$$

where

$$E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right] \approx \frac{1}{p-k+1}\sum_{j=k}^{p}\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}_j^i)\pi(\boldsymbol{\theta}_j^i)}{q(\boldsymbol{\theta}_j^i)}\right)$$

where $\boldsymbol{\theta}_j^i$ is the jth MCMC estimate of the power posterior from the ith temperature. The MCMC will run for p iterations.

We would like to use the corrected trapezium rule (Section 3.4.1) in the Importance Power Posterior to get a potentially better estimate. By using a similar method for calculating the corresponding estimate in Section 3.4.1, the extra term that is required for the corrected trapezium rule is calculated by

$$\frac{d}{dt}E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right) = \int_{\boldsymbol{\theta}}\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\frac{d}{dt}\pi_t(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}.$$

By using the quotient rule for differentiation.

$$\frac{d}{dt}\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) = \pi_t(\boldsymbol{\theta}|\boldsymbol{y})\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right) - E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right)\right).$$

Then putting this back into the equation for $\frac{d}{dt}E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right)$;

$$\begin{aligned}\frac{d}{dt}E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right) &= E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)^2\right)\\ &\quad - E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right)^2\\ &= V_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right)\right).\end{aligned}$$

The estimate with the added variance term is then

$$\log(\pi(\boldsymbol{y})) \approx 0.5 \sum_{i=0}^{n-1} (t_{i+1} - t_i) \left(E_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}} \left[\log \left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \right) \right] \right.$$

$$\left. + E_{\boldsymbol{\theta}|\boldsymbol{y},t_i} \left[\log \left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \right) \right] \right) - \sum_{i=0}^{n-1} \frac{(t_{i+1} - t_i)^2}{12} \left[V_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}} \left[\log \left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \right) \right] \right.$$

$$\left. - V_{\boldsymbol{\theta}|\boldsymbol{y},t_i} \left[\log \left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \right) \right] \right].$$

We now have an alternative method to Annealing-Melting Integration for approximating the marginal likelihood which may be better in cases where the prior is very diffuse. We will now investigate a method that instead directly estimates the logarithm of the Bayes factor.

3.4.3 Model-Switch Integration

When the Annealing-Melting Integration method (Section 3.4.1) is used to calculate the logarithm of the Bayes Factor in favour of one model over another we first approximate each model's marginal likelihood and then use these to calculate the Bayes factor,

$$K = \frac{\pi(\boldsymbol{y}|M_0)}{\pi(\boldsymbol{y}|M_1)}.$$

It is possible that the difference between the logarithm of the marginal likelihoods may be small in comparison to the values of the logarithm of the marginal likelihoods. For this reason a method was proposed by Lartillot and Philippe (2006) which directly calculates the logarithm of the Bayes factor instead of first approximating the two marginal likelihoods separately. This method is also more computationally efficient as only one MCMC is run instead of two.

Suppose that there are two models that we wish to compare denoted M_0 and

M_1. Also let's denote the parameters that appear across both models as

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_0, \boldsymbol{\theta}_1]$$

where $\boldsymbol{\theta}_0$ are the parameters in M_0 and $\boldsymbol{\theta}_1$ are the parameters in M_1. The likelihood for M_0 is $\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0)$ with prior $\pi(\boldsymbol{\theta}|M_0)$. The corresponding likelihood and prior for model M_1 are $\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1)$ and $\pi(\boldsymbol{\theta}|M_1)$. A path which goes from model M_0 to model M_1 is

$$\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) \propto [\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0)\pi(\boldsymbol{\theta}|M_0)]^{1-t}[\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1)\pi(\boldsymbol{\theta}|M_1)]^t$$

such that

$$\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{[\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0)\pi(\boldsymbol{\theta}|M_0)]^{1-t}[\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1)\pi(\boldsymbol{\theta}|M_1)]^t}{z(\boldsymbol{y}|t)}$$

where

$$z(\boldsymbol{y}|t) = \int_{\boldsymbol{\theta}} [\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0)\pi(\boldsymbol{\theta}|M_0)]^{1-t}[\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1)\pi(\boldsymbol{\theta}|M_1)]^t d\boldsymbol{\theta}.$$

Then

$$\begin{aligned} z(\boldsymbol{y}|t=0) &= \int_{\boldsymbol{\theta}} \pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0)\pi(\boldsymbol{\theta}|M_0)d\boldsymbol{\theta} \\ &= \pi(\boldsymbol{y}|M_0) \end{aligned}$$

and

$$\begin{aligned} z(\boldsymbol{y}|t=1) &= \int_{\boldsymbol{\theta}} \pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1)\pi(\boldsymbol{\theta}|M_1)d\boldsymbol{\theta} \\ &= \pi(\boldsymbol{y}|M_1). \end{aligned}$$

Then the following identity can be derived

$$\log\left(\frac{\pi(\boldsymbol{y}|M_1)}{\pi(\boldsymbol{y}|M_0)}\right) = \log\left(\frac{z(\boldsymbol{y}|t=1)}{z(\boldsymbol{y}|t=0)}\right)$$

$$= \int_0^1 E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right] dt$$

by using

$$\begin{aligned}
& \frac{d}{dt}\log(z(\boldsymbol{y}|t)) \\
= \; & \frac{1}{z(\boldsymbol{y}|t)}\frac{d}{dt}z(\boldsymbol{y}|t) \\
= \; & \frac{1}{z(\boldsymbol{y}|t)}\frac{d}{dt}\int_{\boldsymbol{\theta}}[\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)]^{1-t}[\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)]^t d\boldsymbol{\theta} \\
= \; & \frac{1}{z(\boldsymbol{y}|t)}\Big(\int_{\boldsymbol{\theta}}(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0))^{1-t} \\
& (\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1))^t\log(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1))d\boldsymbol{\theta} \\
& -\int_{\boldsymbol{\theta}}(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0))^{1-t}(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1))^t \\
& \log(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0))d\boldsymbol{\theta}\Big) \\
= \; & E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)] - E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)] \\
= \; & E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right].
\end{aligned}$$

Similarly to the Annealing-Melting Integration method we can use the trapezoidal rule to find an approximation to the logarithm of the Bayes factor, which

is

$$\log\left(\frac{\pi(\boldsymbol{y}|M_1)}{\pi(\boldsymbol{y}|M_0)}\right) \approx \quad 0.5\sum_{i=0}^{n-1}(t_{i+1}-t_i)\Bigg(E_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right] + E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right]\Bigg) \tag{3.4.8}$$

where

$$E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right]$$

$$\approx \frac{1}{p-k+1}\sum_{j=k}^{p}\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}_j^i,M_1)\pi(\boldsymbol{\theta}_j^i|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta}_j^i,M_0)\pi(\boldsymbol{\theta}_j^i|M_0)}\right). \tag{3.4.9}$$

We would like to use the corrected trapezium rule (Section 3.4.1) to get a better approximation when using Model-Switch Integration. By following the approach taken by Friel *et al*. (2013) to use the corrected trapezium rule within Annealing-Melting Integration, we wish to find

$$\frac{d}{dt}E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right]$$

$$= \int \log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\frac{d}{dt}\pi_t(\boldsymbol{\theta}|y)d\boldsymbol{\theta}.$$

We first must calculate

$$\frac{d}{dt}\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{d}{dt}\frac{(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0))^{1-t}(\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1))^t}{z(\boldsymbol{y}|t)}.$$

Then by using the quotient rule we calculate

$$\frac{d}{dt}\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) = \pi_t(\boldsymbol{\theta}|\boldsymbol{y})\Bigg(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right) - E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right)\Bigg).$$

If we put this into the equation to calculate $\frac{d}{dt}E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right]$ we obtain the following approximation

$$\begin{aligned}\frac{d}{dt}E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right] &= E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)^2\right)\\ &\quad - \left(E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right)\right)^2\\ &= V_{\boldsymbol{\theta}|\boldsymbol{y},t}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right).\end{aligned}$$

Then the new approximation for the logarithm of the Bayes factor in favour of M_0 over M_1 is

$$\begin{aligned}&\log\left(\frac{\pi(\boldsymbol{y}|M_0)}{\pi(\boldsymbol{y}|M_1)}\right)\\ &\approx \sum_{i=0}^{n-1}(t_{i+1}-t_i)\frac{E_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right] + E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0}\right)\right]}{2}\\ &\quad - \sum_{i=0}^{n-1}\frac{(t_{i+1}-t_i)^2}{12}\Bigg[V_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right)\\ &\quad - V_{\boldsymbol{\theta}|\boldsymbol{y},t_i}\left(\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right)\Bigg].\end{aligned}$$

We have reviewed three methods based on Thermodynamic Integration for estimating the logarithm of the Bayes factor. The first Annealing-Melting Integration (Lartillot and Philippe, 2006) estimates the logarithm of the marginal likelihood. Annealing-Melting Integration may not work well if the prior is very dif-

fuse, therefore we introduced Importance Power Posterior, which moves from the proposal distribution when $t = 0$ to the posterior distribution when $t = 1$. The final method that we reviewed is Model-Switch Integration (Lartillot and Phillippe, 2006) which directly calculates the logarithm of the Bayes factor. It was introduced because it is possible that the differece between the logarithm of the marginal likelihoods may be small in comparison to the values of the logarithm of the marginal likelihoods. We then looked at a suggestion by Friel *et al*. (2013) to apply the corrected trapezium rule when approximating the logarithm of the marginal likelihood using Annealing-Melting Integration. We finally applied the corrected trapezium rule to Importance Power Posterior and Model-Switch Integration. We will now investigate these methods on a toy example before using them in the setting of partial volume models.

3.4.4 Simulation study on a toy example

We will now introduce an example where we can obtain an analytical solution for the logarithm of the marginal likelihood. Then we will use this exact solution to investigate how good the estimates using the proposed methods based on Thermodynamic Integration are.

Suppose that we have data and we want to answer the question whether they come from a Gamma distribution or an Exponential distribution. In addition we assume that the Gamma distribution's shape parameter α is known and equal to 2, while its rate parameter β is assumed to be unknown. The rate parameter of the Exponential distribution λ is also unknown. We can then execute model selection using Annealing-Melting Integration (Section 3.4.1), Importance Power Posterior (Section 3.4.2) and Model-Switch Integration (Section 3.4.3) to calculate estimates of the Bayes factor, by estimating the marginal likelihoods. In this example we will denote M_0 to be the Exponential distribution while M_1 will be

the Gamma distribution. We can obtain analytically the marginal likelihoods because we have chosen conjugate priors, which will allow us to see how well the methods for approximating the marginal likelihoods work before applying them on the partial volume models.

In more detail we denote

$$\boldsymbol{\theta} = [\lambda, \beta]$$

$$\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0) = \lambda^n \exp\left(-\lambda \sum y_i\right)$$

$$\pi(\boldsymbol{\theta}|M_0) = \lambda_\theta \exp(-\lambda_\theta \lambda)$$

$$\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1) = \frac{\beta^{2n}}{\Gamma(2)^n} \exp(-\sum y_i \beta) \prod_{i=1}^{n} y_i$$

$$\pi(\boldsymbol{\theta}|M_1) = \lambda_\beta \exp(-\beta \lambda_\beta).$$

To calculate the analytical solution to the marginal likelihood in the case of the Gamma distribution,

$$\pi(\boldsymbol{y}|M_1) = \int_\beta \pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1)\pi(\boldsymbol{\theta}|M_1) d\beta$$

$$= \int_\beta \lambda_\beta \exp(-\lambda_\beta \beta) \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} \exp\left(-\sum_{i=1}^{n} y_i \beta\right) \prod_{i=1}^{n} y_i^{\alpha-1} d\beta$$

$$= \frac{\lambda_\beta}{\Gamma(\alpha)^n} \prod_{i=1}^{n} y_i^{\alpha-1} \int_\beta \exp\left(-\beta(\lambda_\beta + \sum_{i=1}^{n} y_i)\right) \beta^{n\alpha} d\beta.$$

Then

$$\pi(\boldsymbol{y}|M_1) = \frac{\lambda_\beta}{\Gamma(\alpha)^n} (\prod_{i=1}^{n} y_i^{\alpha-1}) \frac{\Gamma(n\alpha + 1)}{(\lambda_\beta + \sum_{i=1}^{n} y_i)^{n\alpha+1}}.$$

Similarly for the marginal likelihood of the Exponential distribution

$$\pi(\boldsymbol{y}|M_0) = \int \pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0)\pi(\boldsymbol{\theta}|M_0)d\lambda$$

$$= \lambda_\theta \int \lambda^n \exp(-\lambda(\sum y_i + \lambda_\theta))d\lambda$$

$$= \lambda_\theta \frac{\Gamma(n+1)}{(\sum y_i + \lambda_\theta)^{n+1}}.$$

When we investigated a few examples using the Model-Switch Integration estimate we observed that the estimate of Equation (3.4.9) behaves very strangely when $t = 0$ and $t = 1$. The Expected deviance contributes to our overall Model-Switch Integration estimate of the logarithm of the Bayes factor. After further investigation when $t = 0$ and $t = 1$, the power posterior is

$$\pi_{t=0}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0)\pi(\boldsymbol{\theta}|M_0)$$

and

$$\pi_{t=1}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1)\pi(\boldsymbol{\theta}|M_1).$$

Denote $\boldsymbol{\theta}_0$ as the parameters that are only in M_0 and not in M_1 and similarly $\boldsymbol{\theta}_1$ as the parameters that are only in M_1 and not in M_0. Then within the MCMC, for the temperatures $t = 0$ and $t = 1$, the value of $\pi_{t=0}(\boldsymbol{\theta}|\boldsymbol{y})$ does not depend on $\boldsymbol{\theta}_1$ and the value of $\pi_{t=1}(\boldsymbol{\theta}|\boldsymbol{y})$ does not depend on $\boldsymbol{\theta}_0$, but the approximation to $\log\left(\frac{\pi(\boldsymbol{y}|M_0)}{\pi(\boldsymbol{y}|M_1)}\right)$ depends on $\boldsymbol{\theta}_0$ when $t = 1$ and depends on $\boldsymbol{\theta}_1$ when $t = 0$.

However within MCMC when $t = 0$ any value of $\boldsymbol{\theta}_1$ will be accepted and when $t = 1$ any value of $\boldsymbol{\theta}_0$ will be accepted. This could then greatly affect the value of the approximation to the logarithm of the Bayes Factor in a bad way. Therefore in the examples that follow we will use the approximation we had before and a

modified approximation of

$$\log\left(\frac{\pi(\boldsymbol{y}|M_0)}{\pi(\boldsymbol{y}|M_1)}\right)$$

$$\approx 0.5\sum_{i=1}^{n-2}(t_{i+1}-t_i)\Bigg(E_{\boldsymbol{\theta}|\boldsymbol{y},t_{i+1}}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right]$$

$$+E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right]\Bigg).$$

From this approximation we will see how the $t = 1$ and $t = 0$ terms affect the original approximation (Equation 3.4.8). In the alternative approximation we can choose the lower value of t to be close to 0 and the higher value of t to be close to 1, e.g. 0.01 and 0.99, such that we are approximating the logarithm of the Bayes factor by

$$\log\left(\frac{\pi(\boldsymbol{y}|M_1)}{\pi(\boldsymbol{y}|M_0)}\right)=\log\left(\frac{z(\boldsymbol{y}|t=1)}{z(\boldsymbol{y}|t=0)}\right)$$

$$=\int_0^1 E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right]dt$$

$$\approx\int_{0.01}^{0.99} E_{\boldsymbol{\theta}|\boldsymbol{y},t}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_1)\pi(\boldsymbol{\theta}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\theta},M_0)\pi(\boldsymbol{\theta}|M_0)}\right)\right]dt.$$

One thing that we can modify in the algorithm is the spacing between the temperatures, t. Each temperature has the form $t_i = (\frac{i}{n})^c$, for $i = 0 : n$, where n is the number of points available. The value of c can be adjusted, as c increases the approximation using Annealing-Melting Integration should get better for the following reason. As shown in Friel and Pettitt (2008), the choice of prior distribution and its parameters can greatly affect the value of $E_{\boldsymbol{\theta}|\boldsymbol{y},t}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))]$ for values of t close to 0.

The Monte Carlo standard error for $E_{\boldsymbol{\theta}|\boldsymbol{y},t_i}[\log(\pi(\boldsymbol{y}|\boldsymbol{\theta}))]$ is denoted by s_i. The overall Monte Carlo standard error for the approximation of the logarithm of the marginal likelihood in Equation (3.4.5) is

$$\sqrt{\frac{(t_2-t_1)^2}{2}s_1^2+\sum_{i=2}^{n-1}\frac{(t_{i+1}-t_{i-1})^2}{2}s_i^2+\frac{(t_n-t_{n-1})^2}{2}s_n^2}. \tag{3.4.10}$$

If the spacing between temperatures which are closer to 0 is made smaller then it should help to decrease the overall Monte Carlo standard error.

3.4.5 Results

We will simulate 50 datasets of size 100 from the Gamma$(2,\beta)$ distribution in which we showed there is an analytical answer to the logarithm of the marginal likelihood in Section 3.4.4. Then by using Annealing-Melting Integration, Importance Power Posterior and Model-Switch Integration we will compare the estimates of the logarithm of the Bayes factor in favour of the Gamma distribution over the Exponential distribution with the analytical solution. We will alter the spacing in the temperatures and also change the prior parameters such that we can see how these affect the estimation. We will also see how the corrected trapezium rule affects our three methods. After an investigation we found that temperature spacing of $c = 5$ appears to work well when using Annealing-Melting Integration. The choice of temperature spacing c will be dependent on the problem, however Friel and Pettitt (2008) found that in general c works well when its value is between 3 and 5. A plot of t against the Expected deviance (Equation 3.4.6) is a good way of spotting whether the choice of temperature spacing is good by looking at how smooth the plotted curve is. In Model-Switch Integration the estimate without the extreme temperature values and Importance Power Posterior worked well when $c = 1$.

To assess the efficiency of the estimators we compare their mean squared error (MSE). If we denote the analytical logarithm of the Bayes factor to be $\pi(\boldsymbol{y})$, the estimate without the added variance term $\widehat{\pi(\boldsymbol{y})}$ and the other estimate with the added variance term to be $\widetilde{\pi(\boldsymbol{y})}$, then the first squared error (SE) which we will call SE_1 is

$$SE_1 = (\widehat{\pi(\boldsymbol{y})} - \pi(\boldsymbol{y}))^2$$

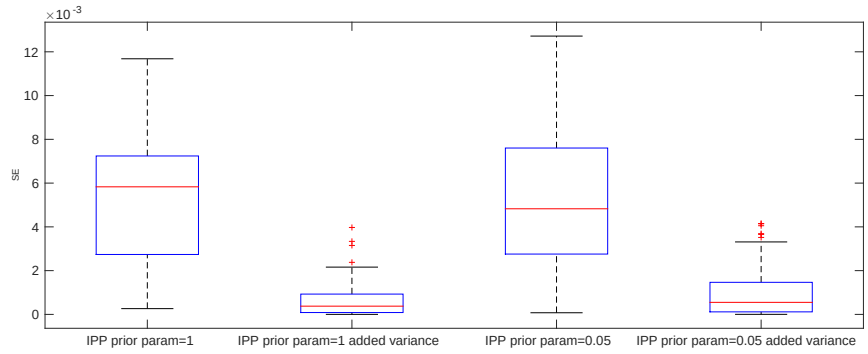
while the second SE, SE_2 is

$$SE_2 = (\widetilde{\pi(\boldsymbol{y})} - \pi(\boldsymbol{y}))^2.$$

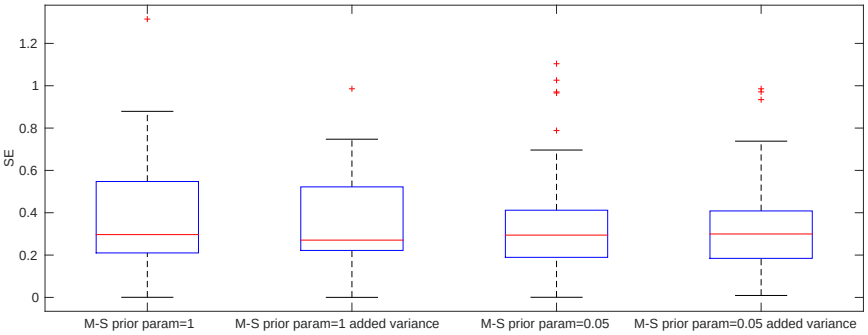
We then use these definitions of the squared error to obtain the MSE by calculating the mean of the SEs.

We then estimated the logarithm of the Bayes factor in favour of the Gamma distribution over the Exponential distribution using the Annealing-Melting Integration, Model-Switch Integration and Importance Power Posterior estimates and compared these to the true value obtained analytically by calculating the MSE, which is shown in Table 3.1 for different spacings and different methods. The boxplots of the SE are in Figure 3.5. It can be seen that in all cases the approximation with the added variance term is better than the original approximation.

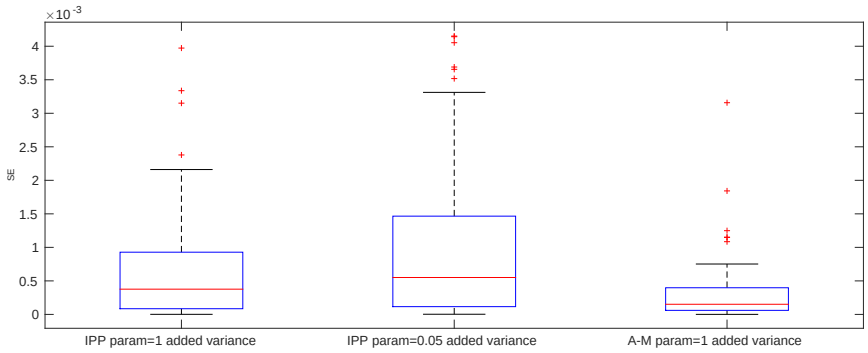
From the boxplots of the SE we can see that the best estimate is the Annealing-Melting Integration estimate when $\lambda_\theta = \lambda_\beta = 1$. Both of the estimates when using the Importance Power Posterior are extremely good. The Annealing-Melting Integration estimate is not as good when $\lambda_\theta = \lambda_\beta = 0.05$, although the MSE is still relatively small. The Model-Switch Integration estimates are worse than the other two methods. However they are still not bad.



(a)



(b)



(c)

Figure 3.5: The SE for the Gamma distribution toy example (Section 3.4.5) in (a) Importance Power Posterior, (b) Model-Switch Integration and (c) comparing Importance Power Posterior and Annealing-Melting Integration.

	$\lambda_\theta = \lambda_\beta = 1$	$\lambda_\theta = \lambda_\beta = 0.05$
A-MI c=5	11.0518 (0.0004)	11.5110(0.0347)
A-MI added variance term c=5	11.0521 (0.0004)	11.5096 (0.0342)
IPP c=1	10.9811 (0.0054)	11.4592 (0.0051)
IPP added variance term c=1	11.0690 (0.0007)	11.5483 (0.0011)
M-S c=1	10.5083 (0.3515)	10.9738 (0.3555)
M-S c=1 added variance term	10.5121 (0.3306)	10.9780 (0.3380)

Table 3.1: The estimate of the logarithm of the Bayes factor in favour of the Gamma distribution over the Exponential distribution and (the MSE) when using different prior parameters and when using the different power posterior methods with and without the added variance term to estimate the logarithm of the Bayes factor in favour of the Gamma distribution over the Exponential distribution. A-MI denotes Annealing-Melting Integration, IPP denotes Importance Power Posterior and M-S denotes Model-Switch Integration. Here 50 datasets of size 100 were simulated from the Gamma distribution distribution with $\beta = 1$. The analytical solution of the logarithm of the Bayes factor in favour of the Gamma distribution is 11.0510 when $\lambda_\theta = \lambda_\beta = 1$ and 11.5266 when $\lambda_\theta = \lambda_\beta = 0.05$.

For comparison purposes we looked at the estimate for the logarithm of the Bayes factor when using the Laplace approximation of the posterior to produce samples to be used in the posterior harmonic mean estimator (Section 3.2.2) which is

$$\pi(\boldsymbol{y}|M_k) = \frac{J}{\sum_{j=1}^{J} \frac{1}{\pi(\boldsymbol{y}|\boldsymbol{\theta}_k^j, M_k)}}.$$

The analytical logarithm of the Bayes factor in favour of the Gamma distribution is 15.4987 when $\lambda_\theta = \lambda_\beta = 1$ and 15.9508 when $\lambda_\theta = \lambda_\beta = 0.05$. We produced 10000 samples from the Laplace approximation of a dataset of size 100 from the Gamma distribution, and we obtained the estimates of the Bayes factor to be 17.7661 when $\lambda_\theta = \lambda_\beta = 1$ and 16.2298 when $\lambda_\theta = \lambda_\beta = 0.05$. Thus when $\lambda_\theta = \lambda_\beta = 0.05$ the estimate appears to be very good, whilst when $\lambda_\theta = \lambda_\beta = 1$ the estimate is not very good.

As a second example we simulated 50 datasets of size 100 from the Exp(1) distribution. Again we can obtain an analytical answer to the logarithm of the

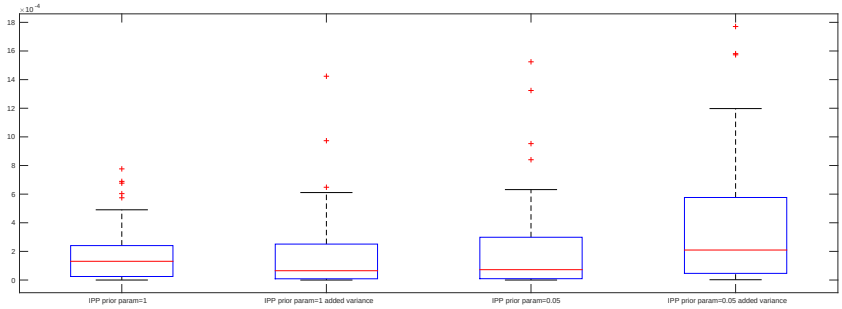
	$\lambda_\theta = \lambda_\beta = 1$	$\lambda_\theta = \lambda_\beta = 0.05$
A-MI c=5	-18.7218 (0.0002)	-17.7427(0.0201)
A-MI added variance term c=5	-18.7205 (0.0002)	-17.7438 (0.0197)
IPP c=1	-18.7196 (0.0002)	-17.7573 (0.0002)
IPP added variance term c=1	-18.7108 (0.0002)	-17.7481 (0.0004)
M-S c=1	-18.6735 (0.0725)	-17.7428 (0.0921)
M-S c=1 added variance term	-18.6649 (0.0625)	-17.7432 (0.0787)

Table 3.2: The estimate of the logarithm of the Bayes factor in favour of the Gamma distribution over the Exponential distribution and (the MSE) when using different prior parameters and when using the different power posterior methods with and without the added variance term to estimate the logarithm of the Bayes factor in favour of the Gamma distribution over the Exponential distribution. A-MI denotes Annealing-Melting Integration, IPP denotes Importance Power Posterior and M-S denotes Model-Switch Integration. Here 50 datasets of size 100 were simulated from the Exponential distribution distribution with $\lambda = 1$. The analytical solution of the logarithm of the Bayes factor in favour of the Gamma distribution is -18.7169 when $\lambda_\theta = \lambda_\beta = 1$ and -17.7634 when $\lambda_\theta = \lambda_\beta = 0.05$.

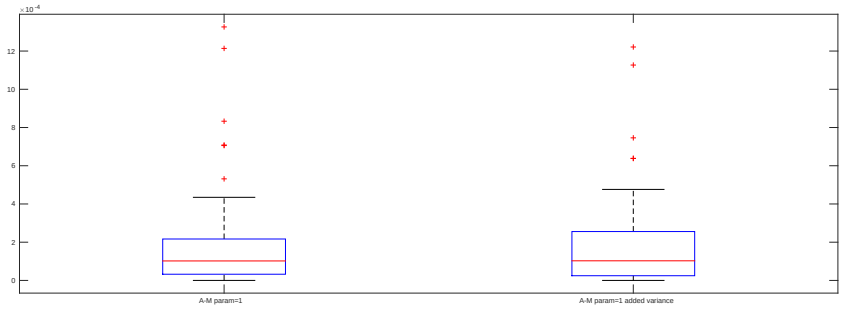
marginal likelihood in Section 3.4.4. We will compare the analytical solution of the logartihm of the Bayes factor in favour of the Gamma distribution over the Exponential distribution with the estimates obtained by using Annealing-Melting Integration, Importance Power Posterior and Model-Switch Intregration. We will again look at how the prior parameters and the added variance term in the corrected trapezium affect our results. We will then calculate the mean squared error to test how close the estimates are to the analytical solution.

The values of the estimated logarithm of the Bayes factor in favour of the Gamma distribution over the Exponential distribution and the MSE can be found in Table 3.2 whilst the boxplots of the SE are in Figure 3.6. We can see that the methods work as well on the Exponential distribution datasets as they did in the case of the Gamma distribution.

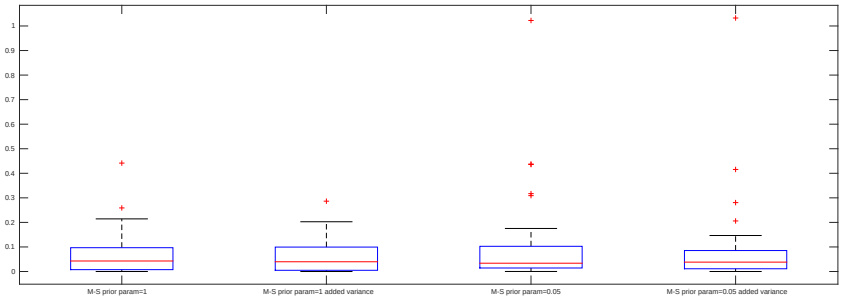
In this section we have showed that all three of our methods for estimating the logarithm of the marginal likelihood and the logarithm of the Bayes factor are



(a)



(b)



(c)

Figure 3.6: The SE for the Exponential distribution toy example (Section 3.4.5) in (a) Importance Power Posterior, (b) Annealing-Melting Integration and (c) Model-Switch Integration.

good in the toy example where we have an analytical solution to compare with. We have also found that the added variance term in the corrected Trapezium rule improves the approximation. In the next section we will use these methods to estimate the Bayes factor when comparing partial volume models, where we will not have an analytical solution.

3.5 Thermodynamic Integration applied to the partial volume model

In Section 3.4 we showed that in a simple Exponential vs Gamma distribution example, Annealing-Melting Integration, Importance Power Posterior and Model-Switch Integration are very good at approximating the logarithm of the Bayes factor. Therefore this is now extended to the problem of model selection between the number of fibres in a voxel in the partial volume model. Initially we discuss how to get initial parameter estimates for the partial volume model with two fibre orientations as until now we have mainly focused on the partial volume model with one fibre orientation. We can also consider having more than two fibre orientations in a voxel which is discussed further in Section 5.3. Then the methods based on Themodynamic Integration will be described for the partial volume model.

We will now be working with the partial volume model with two fibre orientations. Therefore we want to find methods for obtaining good initial parameter estimates in this model which will help us in implementing an efficient MCMC algorithm. Previously when assuming that there is only one fibre orientation (e.g. $f_k = 0$ for $k \geq 2$), the values of the parameters in the partial volume model are initialised by first using metrics obtained from the estimate of the Diffusion Tensor (Section 1.5). These metrics are then used to initialise

the Laplace approximation for the parameters. In particular the value of the Fractional Anisotropy (FA) (Section 1.5) which is a measure of the anisotropy of diffusion in a voxel, is used as the initial value of f_1 as it gives an approximate measure of the white matter in a voxel. From the model with one fibre orientation to the model with two fibre orientations the additional parameters are θ_2, ϕ_2 and f_2. To obtain an initial estimate for f_2, three solutions are proposed:

Method 1 :

Split the FA equally by setting

$$f_1 = FA/2$$

and

$$f_2 = FA/2.$$

Method 2 :

The FA can be separated into linear, planar and spherical measures denoted c_l, c_p and c_s which are taken from the DT literature (Cortez-Conradis *et al.*, 2013) and sum to one. In more detail by using the eigenvalues a_1, a_2 and a_3 that we derive from the Diffusion Tensor (see Section 1.5)

$$c_l = \frac{a_1 - a_2}{a_1 + a_2 + a_3}, \quad a_1 \gg a_2 \approx a_3,$$

$$c_p = \frac{2(a_2 - a_3)}{a_1 + a_2 + a_3}, \quad a_1 \approx a_2 \geq a_3$$

and

$$c_s = \frac{3a_3}{a_1 + a_2 + a_3}, \quad a_1 \approx a_2 \approx a_3.$$

Different combinations of these measures are then used to assign values for f_1

and f_2. As an example we can estimate f_1 to be the linear measure of the FA such that the estimate is

$$\frac{c_l}{c_l + c_p + c_s} FA.$$

Method 3 :

Split the FA by using the values of the eigenvalues from the Diffusion Tensor so that

$$f_1 = \frac{a_1}{a_1 + a_2 + a_3} \text{FA}$$

and

$$f_2 = \frac{a_2}{a_1 + a_2 + a_3} \text{FA}.$$

After simulating different datasets from the partial volume model and investigating the three methods, Method 3 seemed to give consistently good estimates. Therefore this is used to obtain estimates for f_1 and f_2 throughout the rest of this chapter. Then for $\theta_1, \theta_2, \phi_1$ and ϕ_2, we just use the eigenvectors which correspond to a_1 and a_2, so that the two eigenvectors are $\boldsymbol{v}_1 = [v_1(x)\ v_1(y)\ v_1(z)]^T$ and $\boldsymbol{v}_2 = [v_2(x)\ v_2(y)\ v_2(z)]^T$. Then by using the definition of the direction vector $\boldsymbol{v}$ in Section 1.6.1 the initial estimates are

$$\begin{aligned} \theta_1 &= \cos^{-1}(v_1(z)), \\ \theta_2 &= \cos^{-1}(v_2(z)), \\ \phi_1 &= \tan^{-1}\left(\frac{v_1(y)}{v_1(x)}\right), \end{aligned}$$

and

$$\phi_2 \quad = \quad \tan^{-1}\left(\frac{v_2(y)}{v_2(x)}\right).$$

Alongside Method 3 we use Algorithm 10 to obtain a good Laplace approximation.

Algorithm 10 Calculating the Hessian

1: Calculate the Laplace approximation for the local parameters of the partial volume model in each voxel by starting with the eigenvectors that correspond to the two highest eigenvalues from the DT approximation to estimate the values of θ_1, ϕ_1, f_1, θ_2, ϕ_2 and f_2.
2: For each voxel test if the Hessian matrix is positive semi-definite. If it isn't then for that voxel initialise the Laplace approximation using the estimates of θ_1, ϕ_1, f_1, θ_2, ϕ_2 and f_2 that are obtained from the eigenvectors that correspond to the highest eigenvalue and lowest eigenvalue.
3: Retest the remaining voxels to see if the Hessian matrix is positive semi-definite, if not continue by using the DT approximation corresponding to the two lowest eigenvalues to obtain estimates for θ_1, ϕ_1, f_1, θ_2, ϕ_2 and f_2.
4: If we have still not found an appropriate Hessian matrix then continue the algorithm using various different starting values for θ_1, ϕ_1, f_1, θ_2, ϕ_2 and f_2.

3.5.1 Implementing the methods in the partial volume model

A difference between the MCMC algorithms that we will use to calculate the estimate of the Bayes factor and the MCMC algorithms that we used in Chapter 2 is that we must use the full likelihood and priors when estimating the Bayes factor as this is a requirement of the methods based on Thermodynamic Integration. Therefore we must use the posterior before τ is integrated out that is $\pi(\widetilde{\omega}, \tau|\boldsymbol{y})$ described in Section 2.3. We then need to obtain good initial parameter estimates for all of the parameters including τ by using the Laplace approximation. We obtain an approximation for $\widetilde{\omega}$, which denotes all the parameters except τ, by using the Laplace approximation as in Section 2.3.4 and we can note that the distribution of $\tau|\widetilde{\omega}, \boldsymbol{y}$ is Gamma with parameters α and β (see Ap-

pendix D). Therefore we can approximate the initial value of τ by obtaining the mean of the Gamma distribution. The Hessian matrix is then calculated numerically by using the obtained values of τ and $\widetilde{\omega}$.

To apply this all to the partial volume model, first we denote M_0 as the partial volume model with one fibre orientation, while M_1 is the partial volume model with two fibre orientations. The parameter vector is defined to be

$$\boldsymbol{\theta} = (\theta_1, \phi_1, f_1, \theta_2, \phi_2, f_2, d, S_0, \tau).$$

Further the prior distributions are

$$\pi(\boldsymbol{\theta}|M_0) \propto |\sin(\theta_1)|\tau^{\alpha_0 - 1}\exp(-\beta_0\tau)$$

and

$$\pi(\boldsymbol{\theta}|M_1) \propto |\sin(\theta_1)||\sin(\theta_2)|\tau^{\alpha_0 - 1}\exp(-\beta_0\tau).$$

$\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_0)$ and $\pi(\boldsymbol{y}|\boldsymbol{\theta}, M_1)$ are the likelihoods for the partial volume model that were defined in Section 2.3. The Annealing-Melting Integration (Section 3.4.1), Model-Switch Integration (Section 3.4.3) and Importance Power Posterior (Section 3.4.2) methods can then be implemented. We will now attempt these methods on simulated datasets to determine the difference in the estimates for the logarithm of the Bayes factor that are obtained.

3.5.2 An example using data with one fibre orientation

First we simulate data from the partial volume model with one fibre orientation with parameter values of $\theta = 1, \phi = 1, f = 0.5, d = 0.00008, S_0 = 100$ and $\tau = 1$ so that the true model is M_0. We use 101 temperatures in all the methods. The estimates of the logarithm of the Bayes factor in favour of the partial volume model with two fibre orientations over the partial volume model with one fibre

	log(BF)	log(BF) added variance term
IPP	1.3552	1.3396
M-SI	-2.4340	-2.7032
A-MI	-1.8163	-2.0233

Table 3.3: The estimates of the logarithm of the Bayes factor if favour of the partial volume model with two fibre orientations (M_1) over the partial volume model with one fibre orientation (M_0). The estimates are obtained using the Annealing-Melting Integration (A-MI) method with spacing $c = 5$, the Model-Switch Integration (M-SI) method with spacing $c = 1$ and the Importance Power Posterior (IPP) estimate with spacing $c = 1$. The estimates with the extra variance term are also included. The data is simulated from the partial volume model with one fibre orientation.

orientation are in Table 3.3.

The estimate when using Importance Power Posterior is in favour of the partial volume model with two fibre orientations. Annealing-Melting Integration and Model-Switch Integration are in favour of the partial volume model with one fibre orientation. Therefore we tend to believe that the Model-Switch Integration and Annealing-Melting Integration estimates are better in this case.

We then see if the posterior harmonic mean estimator that is defined in Section 3.2.2 is a good estimator of the logarithm of the Bayes factor in favour of the partial volume model with two fibre orientations over the partial volume model with one fibre orientation. We use the Laplace approximations of the partial volume models with one and two fibre orientation to generate samples. We then use these samples within the estimate. The estimate of the logarithm of the marginal likelihood for the partial volume model with one fibre orientation is -95.9547. The estimate of the logarithm of the marginal likelihood for the partial volume model with two fibre orientations is -165.3694. Therefore the estimate of the logarithm of the Bayes factor in favour of the partial volume model with two fibre orientations is -69.4147. We clearly see that the estimate using the harmonic mean estimator is completely different from the Thermo-

	log(BF)	log(BF) added variance
IPP c=1	1.5165	2.3474
M-SI	1.4075	1.6049
A-MI	0.7895	1.404

Table 3.4: The estimates of the logarithm of the Bayes factor when using the Importance Power Posterior method with spacing $c = 1$, the Annealing-Melting Integration (A-MI) method with spacing $c = 5$ and the Model-Switch Integration (M-SI) method with spacing $c = 1$. The estimates with the extra variance term are also included. The data is simulated from the partial volume model with two fibre orientations.

dynamic Integration estimates. The estimates using the Thermodynamic Integration methods seem more realistic, thus from now we concentrate on these methods.

3.5.3 An example using data with two fibre orientations

We now simulate data from the partial volume model with two fibre orientations such that M_1 is the true model. The values of the parameters are $f_1 = 0.4$, $f_2 = 0.2$, $\theta_1 = 1$, $\theta_2 = 1.5$, $\phi_1 = 1$, $\phi_2 = 0.4$, $d = 0.00008$, $S_0 = 100$ and $\tau = 1$. All the estimates of the logarithm of the Bayes factor when using 101 temperatures are compared in Table 3.4. We see that all three methods are in support of the model with two fibre orientations and that the estimates using Model-Switch Integration and Annealing-Melting Integration are quite similar when we include the added variance term.

3.5.4 A second example with two fibre orientations

We finally simulate data from the partial volume model with two fibre orientations and a fairly large value of d (Section 1.6.1) to see how this affects the estimates. The values of the parameters are $f_1 = 0.2$, $f_2 = 0.2$, $\theta_1 = 1$, $\theta_2 = 1.5$, $\phi_1 = 1$, $\phi_2 = 0.4$, $d = 0.001$, $S_0 = 100$ and $\tau = 1$. A summary of the estimates for the logarithm of the Bayes factor in favour of the partial volume model with

	log(BF)	log(BF) added variance
IPP	50.1390	50.1390
A-MI	54.0208	51.8062
M-SI	51.7417	51.9671

Table 3.5: The estimates of the logarithm of the Bayes factor when using the Importance Power Posterior with spacing $c = 1$, Annealing-Melting Integration (A-MI) method with spacing $c = 5$ and the Model-Switch Integration (M-SI) method with spacing $c = 1$. The estimates with the extra variance term are also included. The data is simulated from the partial volume model with two fibre orientations.

two fibre orientations when using 101 temperatures are in Table 3.5.

The obtained estimates are also very close to each other showing that there is not much difference in the three methods. The obtained estimates are strongly in support of the model with two fibre orientations as expected.

3.5.5 A simulation study using Model-Switch Integration

After investigating a few simulated datasets, it appears that if d is small then the Importance Power Posterior estimate is not very good due to the Laplace approximation not being very good in this case. We decided that the best Thermodynamic Integration method to use is Model-Switch Integration as it is just as good at Annealing-Melting Integration but was only introduced because of problems that can sometimes occur in Annealing-Melting Integration. We will now try Model-Switch Integration on various datasets with different numbers of temperatures and different values of τ to see how they affect the results. In DTI the signal to noise ratio of the signal corresponding to $b = 0$ should be at least 20 to obtain good estimates (Lagana *et al.*, 2010). Therefore if S_0 is small we will only obtain good estimates if the precision τ is large.

	$\tau = 0.5$	$\tau = 1$	$\tau = 2$
51 temperatures	-2.2843	-3.1889	-2.0742
51 temperatures added variance	-2.3079	-3.2782	-2.1477
101 temperatures	-1.8396	-2.6907	-2.0311
101 temperatures added variance	-1.8226	-2.7287	-2.0253
201 temperatures	-1.6827	-2.9290	-1.0878
201 temperatures added variance	-1.6672	-2.9377	-1.0847

Table 3.6: The estimates of the logarithm of the Bayes factor in favour of the partial volume model with two fibre orientation over the partial volume model with one fibre orientation using Model-Switch Integration. The datasets are simulated from the partial volume model with one fibre orientation for various values of τ. We use different numbers of temperatures in the estimation and compare the effect of the extra variance term in the estimate.

3.5.5.1 Model-Switch Integration on datasets with one fibre orientation

The dataset from the partial volume model that we investigate has the following parameters, $\theta_1 = 1, \phi_1 = 1, f_1 = 0.7, d = \frac{1}{12000}$ and $S_0 = 100$. Three datasets will be simulated, with varying values of τ which will be 0.5, 1 and 2. The spacing of the temperatures will be equal and we will look at the cases where there are 51, 101 and 201 temperatures. The results are in Table 3.6. We see that all of the estimates are in favour of the model with one fibre orientation. The noise and number of temperatures varies the estimates, however they are quite similar to each other. The number of temperatures affects the approximation greatly in the case when $\tau = 2$.

3.5.5.2 Model-switch Integration on a dataset with two crossing fibres

In this example a dataset from the partial volume model with two fibre orientations is looked at that could potentially have the problem of crossing fibres due to f_1 being equal to f_2 (Section 3.1). The parameter values are $\theta_1 = 1, \theta_2 = 1.5$, $f_1 = 0.2, \phi_1 = 1, \phi_2 = 0.4, f_2 = 0.2, d = \frac{1}{12000}$ and $S_0 = 100$. The values of τ will again be 0.5, 1 and 2. The results are in Table 3.7.

	$\tau = 0.5$	$\tau = 1$	$\tau = 2$
51 temperatures original	-2.0263	-0.6872	2.9097
51 temperatures new	-2.0223	-0.6328	3.2338
101 temperatures original	-2.3150	0.6237	2.7692
101 temperatures new	-2.3550	0.6416	2.8336
201 temperatures original	-2.3398	0.3616	2.6745
201 temperatures new	-2.3245	0.1314	2.7212

Table 3.7: The estimates of the logarithm of the Bayes factor in favour of the partial volume model with two fibre orientation over the partial volume model with one fibre orientation using Model-Switch Integration. The datasets are simulated from the partial volume model with two fibre orientations for various values of τ. We use different numbers of temperatures in the estimation and compare the effect of the extra variance term in the estimate.

In the results we see that when $\tau = 2$ all of the estimates are in favour of the partial volume model with two fibre orientations. However when $\tau = 1$ the estimates when using 51 temperatures are in favour of the partial volume model with one fibre orientation, while when $\tau = 0.5$ all estimates are in favour of the partial volume model with one fibre orientation. Therefore we can see the effect of noise and the number of temperatures on the estimated logarithm of the Bayes factor using Model-Switch Integration. It is expected that when there is more noise and when there are less temperatures the estimation will not be as good which we have demonstrated in this example. This example motivates the use of FPT in Section 3.6 as we have shown that the Bayes factor may be in favour of the incorrect model in cases when we have crossing fibres within a voxel. FPT provides a solution by including model uncertainty within Tractography.

3.5.5.3 Model-switch Integration on a dataset with two fibre orientations

In this example a dataset from the partial volume model with two fibre orientations was simulated where $f_1 \neq f_2$. The true values of the parameter are $\theta_1 = 1$, $\phi_1 = 1$, $f_1 = 0.4$, $\theta_2 = 1.4$, $\phi_2 = 2$, $f_2 = 0.3$, $d = \frac{1}{12000}$ and $S_0 = 100$. The values of τ are 0.5, 1 and 2. The results of the Model-Switch Integration estimates are

	$\tau = 0.5$	$\tau = 1$	$\tau = 2$
51 temperatures original	-0.7614	12.5076	39.0341
51 temperatures new	-0.7507	12.6190	39.7734
101 temperatures original	-0.6290	13.4871	38.3061
101 temperatures new	-0.6280	14.1909	38.4760
201 temperatures original	2.5329	13.3712	39.0494
201 temperatures new	2.5260	13.4160	39.1979

Table 3.8: The estimates of the logarithm of the Bayes factor in favour of the partial volume model with two fibre orientation over the partial volume model with one fibre orientation using Model-Switch Integration. The datasets are simulated from the partial volume model with two fibre orientations for various values of τ. We use different numbers of temperatures in the estimation and compare the effect of the extra variance term in the estimate.

in Table 3.8 The results are mainly strongly in favour of the model with two fibre orientations. However when $\tau = 0.5$, the estimates with 51 and 101 temperatures are in favour of the model with one fibre orientation. However the estimates that uses more temperatures are in favour of the correct model.

We have seen the importance of noise and the number of temperatures when using Model-Switch Integration. In real data we will not be able to control the value of τ but we can control the number of temperatures. When choosing the number of temperatures we will have to weigh the benefit of having more temperatures with the extra computational cost. The computational cost increases proportionally with the number of temperatures. Now that we have got methods for calculating the estimate of the logarithm of the Bayes factor, we will now apply this to implement Fully Probabilistic Tractography. We would like to do this so that model uncertainty can be taken into account in Tractography.

3.6 Fully Probabilistic Tractography

Now that we can obtain a good estimate for the Bayes factor it is possible to calculate the probability of choosing a certain model given the data (Kass and

Raftery, 1995). The method is very simple; suppose that there are $k+1$ proposed models which are denoted M_0, M_1, M_2,..., M_k. Then the Bayes factor can be calculated in favour of M_j against M_0, such that we have β_{00}=1, β_{10}, β_{20},...,β_{k0}. Further let α_j=$\pi(M_j)/\pi(M_0)$ be the prior odds for Model j against Model 0. Then

$$\pi(M_j|\boldsymbol{y}) = \frac{\alpha_j \beta_{j0}}{\sum_{r=0}^{K} \alpha_r \beta_{r0}}$$

is derived by comparing M_j with M_0 and weighting this against all the other models compared with M_0. α_j allows us to add in information about the plausability of competing models. Now using these probabilities the Tractography algorithm when there are two models to choose from, M_0 and M_1, is defined in Algorithm 11. We focus on the case when $K = 1$ because in this chapter we will be looking at model selection when there are two models.

Algorithm 11 Fully Probabilistic Tractography

1: Choose a voxel in our starting region that we will start in. We will need the values of $\pi(M_0|\boldsymbol{y})$ and $\pi(M_1|\boldsymbol{y})$ within this voxel so we will have to estimate the Bayes factor in favour of M_1 over M_0.
2: Sample a value, u_1 such that $u_1 \sim U(0,1)$. If $u_1 < \pi(M_0|\boldsymbol{y})$ choose the partial volume model with one fibre orientation otherwise choose the partial volume model with two fibre orientations.
3: From the model we have chosen, M_k, get a sample from $\pi(\boldsymbol{\theta}_k|\boldsymbol{y}, M_k)$.
4: From $\boldsymbol{\theta}_k$ look at all the f_i values and by generating some u_2 from $U(0,1)$ choose the fibre orientation based on the weights of the f_i.
5: Go back to Step 1, and treat the current voxel we are in as the voxel to start. We will continue these steps until the stopping conditions are reached.

If we assume that all the models are equally likely such that $\pi(M_0) = \pi(M_1) = ... = \pi(M_K)$, then $\alpha_j = 1$. It is very easy to extend this algorithm to when there are more than two models to choose from. In this example it is a requisite that we have samples from both posterior distributions in advance.

3.6.1 FPT Example 1

A 3D dataset was simulated, where the data in each voxel is simulated from the partial volume model with one fibre orientation. The dataset had 11 units in the x direction, 3 units in the y direction and 3 units in the z direction, such that in total there are 99 voxels. There is a tract that has knots at $(1.5, 1.5, 1.5)$, $(3.0, 2.7, 1.3)$, $(5.0, 1.5, 0.7)$, $(7.0, 2.1, 0.5)$ and $(8.0, 1.5, 0.9)$. We will try to reconstruct this tract as close as possible by using Fully Probabilistic Tractography. For each of the 99 voxels we assumed $S_0 = 100$ and $d = \frac{1}{12000}$ which was consistent with earlier values of parameters. In a voxel where the tract passes through the value of f_1 in that voxel is set to be 0.7. Otherwise the f_1 value is set to 0.2. The precision in all voxels is $\tau = 1$. If the tract passes through a voxel the value of (θ, ϕ) for that voxel is chosen such that it corresponds to the fibre orientation, otherwise the value of (θ, ϕ) is chosen randomly.

Once the dataset was simulated, a separate Laplace approximation was obtained for each of the 99 voxels, this helps when MCMC is implemented to calculate the Bayes factor using the Model-Switch Integration method (Section 3.4.3). Also when implementing MCMC, for each voxel, the parameter estimates from $\pi_{t=0}(\boldsymbol{\theta}|\boldsymbol{y})$ and $\pi_{t=1}(\boldsymbol{\theta}|\boldsymbol{y})$ must be saved as these correspond to the parameter estimates for the posterior distribution of the two models.

Once Model-Switch Integration had been implemented the following were calculated

$$\begin{aligned} \pi(M_0|\boldsymbol{y}) &= \frac{a_0\beta_{00}}{\alpha_0\beta_{00} + \alpha_1\beta_{10}} \\ &= \frac{1}{1+\beta_{10}} \end{aligned}$$

because we assume that $\beta_{00} = \alpha_0 = \alpha_1 = 1$. Also

$$\begin{aligned}
\pi(M_1|\boldsymbol{y}) &= 1 - \pi(M_0|\boldsymbol{y}) \\
&= 1 - \frac{1}{1+\beta_{10}} \\
&= \frac{1+\beta_{10}-1}{1+\beta_{10}} \\
&= \frac{\beta_{10}}{1+\beta_{10}}
\end{aligned}$$

where M_0 is the partial volume model with one fibre orientation and M_1 is the partial volume model with two fibre orientations. Therefore the value of β_{10} within each voxel lets us calculate the probability of choosing a certain model at each step of the Tractography. In this example it is expected that β_{10} will be less than 1 as each voxel only has one fibre orientation. We then choose one of the models at each voxel depending on the value of $\pi(M_j|\boldsymbol{y})$ and then the values of the parameters that represent the fibre orientation are chosen depending on which model we have chosen. For this simple example we will start at the known value of the start of the tract $(1.5, 1.5, 1.5)$ and we'll finish when (x, y, z) is such that $x \geq 8$.

After estimating the logarithm of the Bayes factor using Model-Switch Integration on the dataset, in the majority of voxels, the Bayes factor was strongly in favour of the partial volume model with one fibre orientation. The lowest value of $\pi(M_0|\boldsymbol{y})$ from the 99 voxels was 2.8623x10^{-8} whilst the highest value was 0.9999. The mean of the values of $\pi(M_0|\boldsymbol{y})$ for all voxels was 0.8127. 88 of the voxels has a value of $\pi(M_0|\boldsymbol{y})$ which was above 0.5. Then Tractography was run to produce 10 tracts, which are shown in Figure 3.7.

From the graph, we see that 9 of the 10 reconstructed tracts correspond well

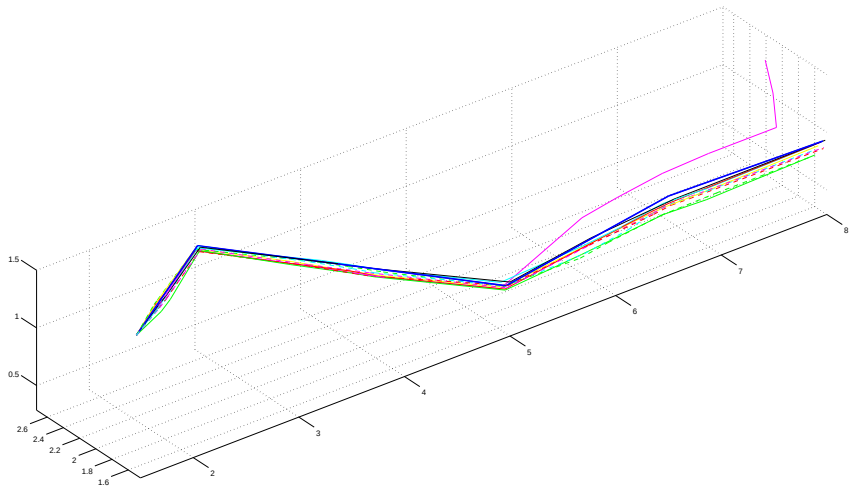


Figure 3.7: The reconstructed tracts compared to the true fibre orientation (bold blue) when using Fully Probabilistic Tractography on FPT Example 1. There is more uncertainty as we go along the tract because as we enter more voxels the probability that we will have chosen the model with two fibre orientations in one of the voxels will increase.

with the true fibre orientation, while one of the tracts at some point, diverts from the true fibre orientation. This is probably due to model uncertainty. When the Tractography is ran many times the majority of the tracts will follow the true fibre orientation, while some will take a different path, but this may be an advantage in the situations such as the one described in Figure 3.1.

3.6.2 FPT Example 2

As in Example 1, the same dataset will be simulated, with one exception. In one of the voxels that the spline passes through, there will now be two fibre orientations, such that $f_1 = 0.3$ and $f_2 = 0.3$, the values of d, S_0, θ_1 and ϕ_1 stay the same in this voxel, whilst the values of θ_2 and ϕ_2 are $\theta_2 = \phi_2 = 1$. The data was then simulated and once again the Model-Switch Integration approximation to the logarithm of the Bayes factor was calculated for each voxel in favour of the partial volume model with two fibre orientations over the partial volume model with one fibre orientation. Once this has finished then the Fully Proba-

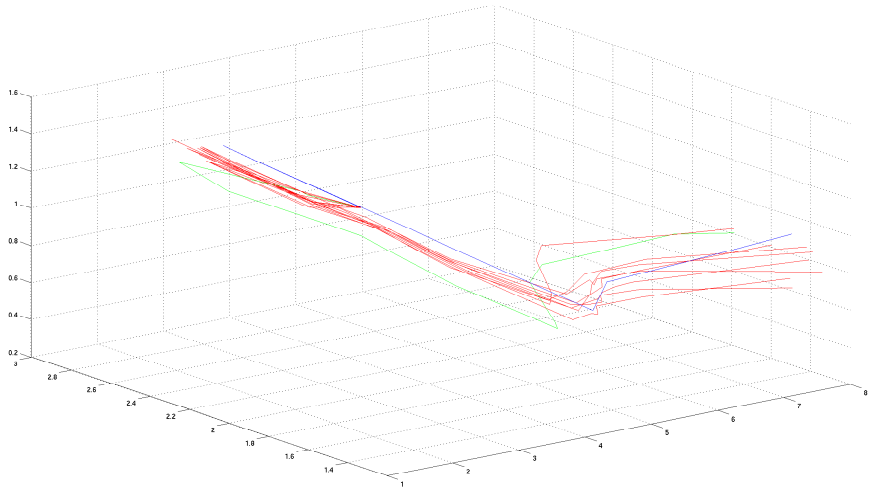


Figure 3.8: The reconstructed splines compared to the true spline (bold blue) when using Fully Probabilistic Tractography on FPT Example 2.

bilistic Tractography algorithm was implemented. When the approximation to the logarithm of the Bayes factor was calculated in the voxel of interest using Model-Switch Integration the answer was 2.2159 in favour of the model with one fibre orientation. Therefore in the voxel of interest,

$$\pi(M_0|\boldsymbol{y}) = \frac{1}{1+\exp(\text{-}2.2159)} = 0.9017$$

$$\pi(M_1|\boldsymbol{y}) = \frac{\exp(-2.2159)}{1+\exp(-2.2159)} = 0.0983.$$

Therefore there is a small amount of uncertainty in the model selection in this example. If we had not taken into account model uncertainty, then only the partial volume model with one fibre orientation would have been used in Tractography due to the Bayes factor being in favour of this. Fully Probabilistic Tractography was then used on the dataset a number of times and the tracts in Figure 3.8 were obtained, where the true orientation is in blue.

As can be seen from the graph, some of the reconstructed tracts are very close to the true fibre orientation. Whilst some other tracts start moving away from the

true fibre orientation, in particular the tract that is green. When implementing Fully Probabilistic Tractography many of the tracts left the brain region so were rejected. We observed that most of the tracts that we rejected were in the case when the model with one fibre orientation is selected in our voxel of interest.

This example demonstrates the benefit of including model uncertainty in Tractography. If model uncertainty has not been considered then the voxel that has two fibre orientations would have only had one fibre orientation because the Bayes factor was in favour of the model with one fibre orientation. When implementing Tractography a lot of tracts would have been rejected because when choosing the model with one fibre orientation in this voxel a lot of the tracts left the brain region before reaching Region 2. By introducing model uncertainty into Tractography, the model with two fibre orientations is also chosen in this voxel and then tracts are produced that represent the connection between the two regions. Now that we have shown the benefits of Fully Probabilistic Tractography, we will try to implement all of the methods that we have used so far in this chapter on real data.

3.7 An application to a real dataset

In this section the methods that were introduced in this chapter will now be applied to be used on real data. An added difficulty in analysing real data, is deciding on how to define the regions of the brain. We will use the Atlas feature in FSL (Section 1.7.2) to define regions.

3.7.1 Bayes factor estimation on real data example

First we estimate the Bayes factor to determine the number of fibre orientations within a voxel as in Section 3.5.1. We will investigate a voxel with the

	log(BF)	log(BF) with added variance
M-SI	-3.8973	-3.9936
A-MI	-1.1434	-4.094

Table 3.9: The estimates of the logarithm of the Bayes factor when using the Annealing-Melting Integration (A-MI) method with spacing $c = 5$ and the Model-Switch Integration (M-SI) method with spacing $c = 1$. The estimates with the extra variance term are also included. The data are the measured Diffusion-Weighted signals in one voxel of a real brain dataset.

following Laplace approximation for the partial volume model with two fibre orientations. $\theta_1 = 1.4307$, $\phi_1 = 1.2249$, $f_1 = 0.2650$, $\theta_2 = 0.9805$, $\phi_2 = 3.2106$, $f_2 = 0.1508$, $d = 0.0003$, $S_0 = 64.3138$ and $\tau = 0.0294$. The corresponding Laplace approximation for the partial volume model with one fibre orientation is $\theta_1 = 1.5494$, $\phi_1 = 1.0560$, $f_1 = 0.2462$, $d = 0.0002$, $S_0 = 63.9683$ and $\tau = 0.0278$. The summary of the estimates of the logarithm of the Bayes factors using the Annealing-Melting Integration and Model-Switch Integration methods are in Table 3.9.

We see that the Model-Switch Integration estimate and the estimate using Annealing-Melting Integration are very similar to each other. Both estimates are in favour of the model with one fibre orientation. We can see that the added variance term seems to greatly affect the Annealing-Melting Integration estimate.

3.7.2 Results of Fully Probabilistic Tractography

We now implement Fully Probabilistic Tractography (Section 3.6) on real datasets. Due to the large number of voxels that will be in our examples we use parallel computing for each voxel such that we are estimating the logarithm of the Bayes factor in voxels simultaneously.

3.7.2.1 FPT on real data Example 1

We will first look at data from subject 1 and use the atlases (Section 1.10) to find the Thalamus and the Amygdala centromedial group on the right hand side of the brain. Then we will investigate the voxels that contain the regions and surround the regions. In total there are 1820 voxels in the data we investigate. First we estimate the logarithm of the Bayes factor in favour of the partial volume model with two fibre orientations for each of these voxels by using Model-Switch Integration. Once this is done we then attempted Fully Probabilistic Tractography. By implementing FPT we find a few tracts that connect Region 1 and Region 2, but most of the time no tract is found which leads us to believe that no connection is available in the few voxels we have. After looking at the literature about the Amygdala (Bzdok *et al.*, 2012), it seems that the Amygdala and Thalamus are indirectly connected by the lateral nucleus and the Basolateral Amygdala. Due to the relatively low number of voxels that we are investigating, this may be the reason why we do not regularly find a connection between these two regions.

3.7.2.2 FPT on real data Example 2

We will now look at data from subject 2 with 3780 voxels. The two regions that we will investigate are the right primary auditory region and the right broca area. The broca area is concerned with speech production in the brain, so a guess would be that auditory and speech are connected somehow. When implementing Fully Probabilistic Tractography, we have shown that when starting at some voxels within the right primary auditory region we create a tract that goes to the right broca area. In the majority of times that we run FPT we find that there is a connection between the two regions, therefore it seems in this case there is evidence for a connection between the two regions. A graph of the connecting tracts can be found in Figure 3.9.

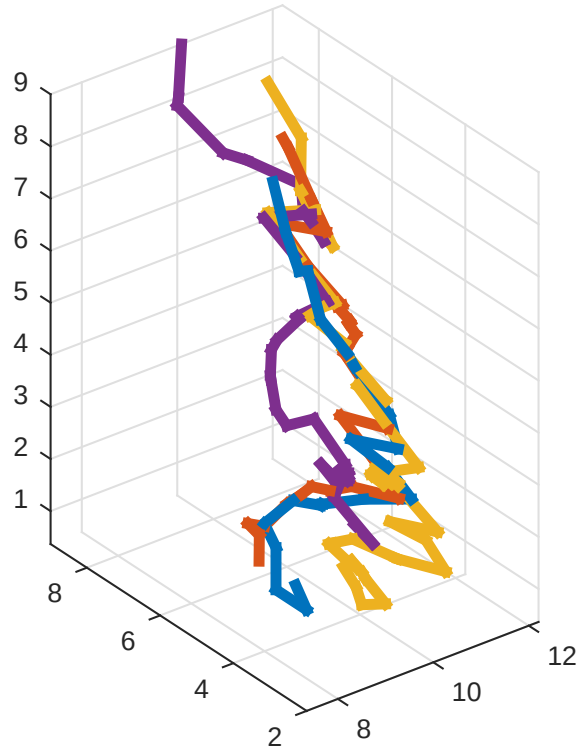


Figure 3.9: The tracts that are constructured that connect the right primary auditory region and the right broca area when using Fully Probabilistic Tractography.

3.8 Conclusions

In this chapter we have found a method to choose how many fibre orientations to model in a voxel when using the partial volume model. This is of vital importance as using the wrong number of fibre orientations could get fibre orientations that are completely wrong from the true fibre orientations.

First we investigated the ARD prior, the commonly used method to choose between the number of fibre orientations in the partial volume model. We summarised the disadvantages of this method and thus started to consider calculating the Bayes factor to compare models.

Calculating an exact analytical solution to the Bayes factor is difficult and proposed solutions to approximate this have disadvantages, thus an approximation is calculated using ideas from Thermodynamic Integration. Annealing-Melting Integration, Importance Power Posterior and Model-Switch Integration were introduced. We decided to use Model-Switch Integration due to it only being proposed due to problems that sometimes occur with Annealing-Melting Integration. We showed that sometimes Importance Power Posterior has problems, therefore we chose to reject this as a possible solution.

Now that it was possible to implement model selection to choose between the number of fibre orientations in the partial volume model, we decided to use uncertainty in model selection in Tractography. There are some cases where the model that is chosen may not be the model with the true number of fibre orientations, therefore we would like to include this model selection uncertainty such that some tracts are reproduced that are close to the truth. Now in our Tractography algorithm at each voxel we also choose the model, based on probabilities that can be calculated using the approximations from the Bayes factors. This

Tractography that we term "Fully Probabilistic Tractography" has been shown in both simulated and real data to be effective.

We then investigated the methods that were introduced in this chapter on real data. The estimates for the logarithm of the Bayes factor appeared to work well. Finally we implemented Fully Probabilistic Tractography on two real datasets. Implementing parallel computing by running the algorithm for different temperatures and different voxels simultaneously greatly improves the computational speed of the algorithm.

CHAPTER 4

Global Tractography

4.1 Motivation

In the previous chapters we have looked for potential connections between brain regions voxel-wise by inferring fibre orientations within a voxel to be used within Tractography algorithms. We would instead like to statistically test for the existence of a connection between two brain regions at a global level. The Global Tractography framework (see Section 1.7.3) was first proposed by Jbabdi *et al*. (2007) and parametrises the connections between two brain regions at a global level to reduce the sensitivity to local noise. If we know that there is a connection between two brain regions then we can include this information within this framework by using a prior density for a connection matrix. For any two brain regions we can then use model selection to choose between the model where a connection exists between two brain regions and the model that says there is no such connection. We propose a new method for inferring the global parameters of this framework which uses model selection techniques based on Thermodynamic Integration.

In this chapter we will attempt to infer the values of the parameters of the Global Tractography model which are the local parameters of the partial vol-

ume model (Section 1.6.1) that represent the fibre orientation within a voxel and the global parameters. In more detail if there are N voxels then the local parameters are $\boldsymbol{d} = [d_1, ..., d_N]$ which is the diffusivity in each voxel, the baseline signal $\boldsymbol{S_0} = [S_{01}, ..., S_{0N}]$, $\boldsymbol{\Theta} = [\theta_1, ..., \theta_N]$, $\boldsymbol{\Phi} = [\phi_1, ..., \phi_N]$ and $\boldsymbol{f} = [f_1, ..., f_N]$ which represent the fibre's directions and the proportion of the signal contributed by the fibre in each voxel and finally $\boldsymbol{\Sigma} = [\sigma_1, ..., \sigma_N]$ which represents the noise in the observed measurements of the signals in each voxel. If it is assumed that we have $\mathcal{N}$ brain regions of interest in the brain image that we are looking at then the global parameters are $\boldsymbol{\mathcal{C}}$ which is an $\mathcal{N}x\mathcal{N}$ matrix whose $(i, j)_{th}$ element is 1 if regions i and j are connected and 0 otherwise. Furthermore we have $\boldsymbol{\mathcal{F}}$ which represents the pathways that connect the regions, here we choose these pathways to be the Catmull-Rom splines (see Appendix A) as in Jbabdi *et al*. (2007), with control points $\boldsymbol{\mathcal{K}}$ and extremities $\boldsymbol{\mathcal{L}}$.

Inferring the parameters of the Global Tractography model is an extension to inferring the values of the local parameters of the partial volume model for each voxel of the brain as in Chapter 2. By inferring the parameters of the Global Tractography model we will be able to calculate the evidence for the existence of a connection between any two regions by using model selection techniques such as those that were introduced in Chapter 3 to infer on the connection matrix $\boldsymbol{\mathcal{C}}$.

Throughout this chapter due to the complexity of the problem that we wish to solve, we will assume that within each voxel there is one fibre orientation. The work within this chapter can be extended to voxels where it is assumed there is more than one fibre orientation. We discuss this in Section 5.3.

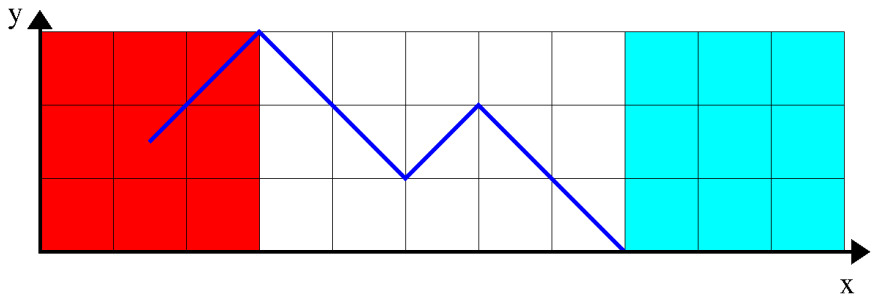


Figure 4.1: The 2D simulated global dataset in which two brain regions of interest in red and blue are connected to each other by a fibre that is modelled using knots from the Global Tractography framework. Each voxel is modelled using the partial volume model with one fibre orientation.

4.2 Simulating data from the Global Tractography model

A 2D dataset was simulated for illustration. A diagram of the dataset is shown in Figure 4.1. In this diagram Region 1, in red, is connected to Region 2, in blue, by a tract with knots at (1.5,1.5), (3,3), (5,1), (6,2) and (8,0). There are 11 voxels in the x axis direction and 3 voxels in the y axis direction. Region 1 is all the voxels such that $x \leq 3$, while Region 2 is all the voxels where $x \geq 8$

For every voxel, data is simulated from the partial volume model (Section 1.6.1) with local parameters having the values $d = \frac{1}{12000}$, $S_0 = 100$ and $\theta = \frac{\pi}{2}$. In voxels which the tract passes through $f = 0.85$, while in the voxels that it doesn't pass through $f = 0.1$. The values of ϕ in voxels which the tract passes through are chosen to be either $\frac{\pi}{4}$ or $\frac{3\pi}{4}$ depending on the orientation of the tract as in the diagram. In the voxels which the tract does not pass through the value of ϕ is simulated from a Uniform distribution, such that the values are random.

The method for simulating data from the Global Tractography framework is summarised in Algorithm 12.

Algorithm 12 Simulating data from the Global Tractography framework

1: For the voxels that the tract that connects the two regions passes though, calculate the fibre direction and assign point estimate values for θ and ϕ.
2: For all other voxels assign random values for the local parameters by using uniform random number generators such as $\theta \sim U(0, \pi)$.
3: For each gradient direction that we have, calculate the true Diffusion-Weighted signal μ_i by using the local parameter values where

$$\mu_i = S_0((1-f)\exp(-b_i d) + f\exp(-b_i d(\mathbf{g}_i^T \boldsymbol{v}^2))),\ i = 1, ..., m.$$

4: Obtain the observed Diffusion-Weighted signal, y_i for each voxel and each gradient direction by simulating it from a normal distribution with mean μ_i and variance σ^2

4.3 Framework for Bayesian inference

We will now introduce the framework that will allow us to infer the parameters of the Global Tractography model. The parameters will be inferred by adopting a Bayesian approach. First the likelihood of the model is derived and the prior distributions for each of the parameters must be defined. The posterior distribution is then derived:

$$\pi(\boldsymbol{\Omega}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\Sigma}, \boldsymbol{d}, \boldsymbol{S_0}, \boldsymbol{f}, \boldsymbol{\Theta}, \boldsymbol{\Phi})\pi(\boldsymbol{\Omega})$$

$$\text{where } \pi(\boldsymbol{\Omega}) = \pi(\boldsymbol{\Sigma})\pi(\boldsymbol{d})\pi(\boldsymbol{S_0})\pi(\boldsymbol{f})\pi(\boldsymbol{\Theta}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{F}})\pi(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{C}})\pi(\boldsymbol{\mathcal{C}})$$

and $\boldsymbol{\Omega}$=($\boldsymbol{\Sigma}$, $\boldsymbol{d}$, $\boldsymbol{S_0}$, $\boldsymbol{f}$, $\boldsymbol{\Theta}$, $\boldsymbol{\Phi}$, $\boldsymbol{\mathcal{C}}$, $\boldsymbol{\mathcal{F}}$). The local parameters are $\boldsymbol{\Sigma}$, $\boldsymbol{d}$, $\boldsymbol{S_0}$, $\boldsymbol{f}$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$, that we are familiar with from the previous chapters. The new global parameters that we will work with for the first time are $\boldsymbol{\mathcal{C}}$, the connection matrix, and $\boldsymbol{\mathcal{F}}$ that represents the splines that connect regions. We should note that ($\boldsymbol{\mathcal{K}}$,$\boldsymbol{\mathcal{L}}$), which are the knots and extreme points of the splines, are not in $\boldsymbol{\Omega}$ because there is a deterministic relationship between ($\boldsymbol{\mathcal{K}}$,$\boldsymbol{\mathcal{L}}$) and $\boldsymbol{\mathcal{F}}$.

The likelihood $\pi(\boldsymbol{y}|\boldsymbol{\Sigma}, \boldsymbol{d}, \boldsymbol{S_0}, \boldsymbol{f}, \boldsymbol{\Theta}, \boldsymbol{\Phi})$ is the same as the likelihood for the partial volume model as in Section 2.3 because the data $\boldsymbol{y}$ is only generated by the local parameters as shown in Algorithm 12. The prior distributions for $\boldsymbol{d}$, $\boldsymbol{f}$ and $\boldsymbol{S_0}$

are assumed to be the same as in the partial volume model which are

$$\pi(\boldsymbol{d}) = \prod_j \pi(d_j),$$

$$\pi(\boldsymbol{f}) = \prod_j \pi(f_j),$$

$$\pi(\boldsymbol{S}_0) = \prod_j \pi(S_{0j}),$$

where j is an index over voxels and

$$\pi(d_j) \sim U(0, \infty),$$

$$\pi(f_j) \sim U(0, 1),$$

$$\pi(S_{0j}) \sim U(0, \infty).$$

We take a different approach from Jbabdi *et al*. (2007) who choose the prior on $\boldsymbol{f}$ to be the ARD prior in voxels in which the tract does not pass through. We dismiss the ARD prior due to it not being effective (as shown in Chapter 3) in this setting.

The conditional prior distribution on ($\boldsymbol{\Theta}$, $\boldsymbol{\Phi}$) is obtained by looking separately at the voxels which $\boldsymbol{\mathcal{F}}$ passes through and the voxels which it doesn't. For notation purposes we will label each voxel as j and we will denote the voxels that the spline passes through as $j \in \mathcal{F}$ and the voxels that the spline doesn't

pass through as $j \notin \mathcal{F}$. Then

$$\pi(\boldsymbol{\Theta}, \boldsymbol{\Phi}|\boldsymbol{\mathcal{F}}) = \prod_{j \notin \boldsymbol{\mathcal{F}}} \pi(\theta_j, \phi_j|\boldsymbol{\mathcal{F}}) \prod_{j \in \boldsymbol{\mathcal{F}}} \pi(\theta_j, \phi_j|\boldsymbol{\mathcal{F}}),$$

$$= \prod_{j \notin \boldsymbol{\mathcal{F}}} \pi(\theta_j, \phi_j) \prod_{j \in \boldsymbol{\mathcal{F}}} \delta(\theta_j - \theta'_j, \phi_j - \phi'_j|\boldsymbol{\mathcal{F}}), \quad j = 1, ..., N,$$

where δ is the Dirac delta distribution (Dirac, 1930), θ'_j and ϕ'_j are the orientations of the spline passing through voxel j and N denotes the total number of voxels in a brain image. In voxels which the spline doesn't pass through, the prior distribution is the same as that in the partial volume model (see Section 2.3). In voxels which the spline passes through the prior distribution is the Dirac delta distribution. Within the MCMC we will have two 3x3 matrices which have an element for each voxel. In each voxel the element will either be 0 or 1 depending on whether the current or candidate spline passes through that voxel.

Finally to find $\pi(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{C}})$, define γ_l as a random variable that describes one connection between two regions and let $\boldsymbol{\mathcal{C}}_l$ be a vector of the upper diagonal elements of the connection matrix $\boldsymbol{\mathcal{C}}$, then $\boldsymbol{\mathcal{F}}$=(γ_1,γ_2,...,$\gamma_{N(N-1)/2}$) and

$$\pi(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{C}}) = \prod_{l=1}^{N(N-1)/2} \pi(\gamma_l|\boldsymbol{\mathcal{C}}_l).$$

If $\boldsymbol{S}_k$ is the set of splines with k knots and c is the element of the connection matrix that represents the connection between the two regions of interest then:

$\pi(\gamma \in \boldsymbol{S}_k, \gamma \text{ connects the two regions}|c = 1) = 1$

$\pi(\gamma \in \boldsymbol{S}_k, \ \gamma \text{ connects the two regions}|c = 0) = 0$

$\pi(\gamma \in \boldsymbol{S}_k, \ \gamma \text{ does not connect the two regions}|c = 1) = 0$

$\pi(\gamma \in \boldsymbol{S}_k, \ \gamma \text{ does not connect the two regions}|c = 0) = 1.$

We now describe how to infer the parameters of the Global Tractography model.

4.4 Inferring the whole set of parameters

We now wish to estimate the values of all the knots as well as all the local parameters in the Global Tractography model. We will denote $\boldsymbol{\omega} = (\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{f}, \boldsymbol{d}, \boldsymbol{S_0})$ and $\boldsymbol{r} = (\boldsymbol{P1}, \boldsymbol{P2}, \boldsymbol{P3}, \boldsymbol{P4}, \boldsymbol{P5})$ where $\boldsymbol{P1} = (P1(x), P1(y))$, $\boldsymbol{P2} = (P2(x), P2(y))$, $\boldsymbol{P3} = (P3(x), P3(y))$, $\boldsymbol{P4} = (P4(x), P4(y))$ and $\boldsymbol{P5} = (P5(x), P5(y))$ are the knots that represent the tract that connects two regions. We choose to model the spline using five knots because Jbabdi *et al*. (2007) show that there is not much difference in the splines that are constructed using more than five knots. We wish to generate samples from $\pi(\boldsymbol{r},\boldsymbol{\omega}|\boldsymbol{y})$. Three methods for implementing this are now described. These methods will be compared and also challenges attributed to these methods and their solutions will be addressed. All of these methods work better when a good initial estimate for the parameters is provided. One such method will be described later in Section 4.5.

4.4.1 Deterministic Scan MCMC

Deterministic Scan MCMC works by first updating $\pi(\boldsymbol{r}|\boldsymbol{y},\boldsymbol{\omega})$ and then updating $\pi(\boldsymbol{\omega}|\boldsymbol{y},\boldsymbol{r})$. When generating samples from $\pi(\boldsymbol{r}|\boldsymbol{y},\boldsymbol{\omega})$,we found that the mixing is bad if we propose the values of all the knots at the same time. Therefore we use separate proposal distributions for each knot. The method is described in Algorithm 13.

Within the MCMC algorithm, the proposal distribution for each of the conditional distributions is a multivariate Normal distribution where the initial mean is the knot values estimated using the method described later in Section 4.5, and some covariance matrix. We tried different covariance matrices in the proposal

Algorithm 13 Deterministic Scan MCMC

1: Initialise the values of the knots and local parameters.
2: Generate new values from $\pi(P1(x), P1(y)|\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{P2}, \boldsymbol{P3}, \boldsymbol{P4}, \boldsymbol{P5})$ using the Metropolis-Hastings algorithm with a random-walk proposal distribution.
3: Continue by using the Metropolis-Hastings algorithm with a random-walk proposal distribution to sample from $\pi(\boldsymbol{P2}|\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{P1}, \boldsymbol{P3}, \boldsymbol{P4}, \boldsymbol{P5})$, $\pi(\boldsymbol{P3}|\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{P1}, \boldsymbol{P2}, \boldsymbol{P4}, \boldsymbol{P5})$, $\pi(\boldsymbol{P4}|\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{P1}, \boldsymbol{P2}, \boldsymbol{P3}, \boldsymbol{P5})$ and $\pi(\boldsymbol{P5}|\boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{P1}, \boldsymbol{P2}, \boldsymbol{P3}, \boldsymbol{P4})$.
4: For each voxel, i, use the Metropolis-Hastings algorithm with an independence sampler Normal proposal distribution that is estimated using the Laplace approximation for each voxel on $\pi(\boldsymbol{\omega}_i|\boldsymbol{y}, \boldsymbol{r})$.
5: Repeat all the steps until the required amount of samples are acquired.

distribution for the knots to obtain a proposal distribution that allows satisfactory mixing within MCMC. The Laplace approximation is used to obtain a good proposal distribution for the local parameters.

4.4.2 Block-update MCMC

In this method we will use the proposed values of the local parameters $\boldsymbol{\omega}$, which are acquired using the Laplace approximation, to obtain proposed values of the knots $\boldsymbol{r}$ that correspond to the proposed values of the local parameters. Therefore we are inferring the parameters by using $\pi(\boldsymbol{r}, \boldsymbol{\omega}|\boldsymbol{y}) = \pi(\boldsymbol{\omega}|\boldsymbol{y})\pi(\boldsymbol{r}|\boldsymbol{\omega}, \boldsymbol{y})$. We then either accept or reject all of these proposed values. This method is described in Algorithm 14.

Algorithm 14 Block-update MCMC

1: Initialise the values of the local parameters by using the Laplace approximation. Then initialise the values of the knots such that they correspond to the initial values of the local parameters.
2: First propose candidate values from $\pi(\boldsymbol{\omega}|\boldsymbol{y}, \boldsymbol{r})$ separately for each voxel by using the Laplace approximation.
3: Using the proposed values of $\boldsymbol{\omega}$, the candidate knot values are proposed using the Deterministic Tractography method that is described in Section 4.5.
4: All new parameter values $(\boldsymbol{\omega}, \boldsymbol{r})$ are then either rejected or accepted, using the Metropolis-Hastings algorithm on $\pi(\boldsymbol{\omega}, \boldsymbol{r}|\boldsymbol{y})$.
5: Repeat Steps 2-4 until the required amount of samples are acquired.

This method only accepts values of r and ω that correspond to each other, i.e. the candidate values of the local parameters determines the candidate knots. However the computational time and performance of Block-update MCMC is about the same as Deterministic Scan MCMC.

4.4.3 Partially Deterministic Scan MCMC

From the previous two methods we will focus on Deterministic Scan MCMC because the proposed knot values do not depend on the the local parameter estimates so it allows a wider range of knot values to be proposed. However when running Deterministic Scan MCMC the acceptance rate of new parameter values was quite low. We will now propose a method that fixes the problems associated with Deterministic Scan MCMC.

To solve the problem with the low acceptance rate we realised that if the spline does not pass through a voxel then we can obtain parameter estimates for this voxel, using the methods that were introduced in Chapter 2 as these voxels do not depend on any of the global parameters. In Deterministic Scan MCMC at the moment every voxel is updated at each step. Alternatively we could just treat voxels that do not depend on the global parameters separately later and use parallel computing on each of these voxels to speed up the computation time. Therefore we decided to introduce Partially Deterministic Scan MCMC.

In Partially Deterministic Scan MCMC, at each step of the MCMC algorithm only the voxels where either the current spline or the candidate spline pass through are updated. Although this seems to help MCMC it does not seem as optimal as we would like it to be. We found that the independence sampler proposal distribution within the MCMC did not work because candidate samples

were not accepted often. Therefore it was decided that instead of an independence sampler we will use a random-walk proposal distribution in the MCMC algorithm. This greatly improves the acceptance rate of the algorithm.

Partially Deterministic Scan MCMC is summarised in Algorithm 15.

Algorithm 15 Partially Deterministic Scan MCMC

1: Simulate from $\pi(\boldsymbol{P1}|\boldsymbol{P2}, \boldsymbol{P3}, \boldsymbol{P4}, \boldsymbol{P5}, \boldsymbol{\omega}, \boldsymbol{y})$ and then $\pi(\boldsymbol{\omega}|\boldsymbol{r}, \boldsymbol{y})$ for all the voxels that the candidate and current splines pass through using the Metropolis-Hastings algorithm.
2: Similarly to Step 1 update the values of $\boldsymbol{P2}$, $\boldsymbol{P3}$, $\boldsymbol{P4}$ and $\boldsymbol{P5}$ and also all the local parameters for the voxels that the current and candidate spline pass through using the Metropolis-Hastings algorithm.
3: Update all the local parameters that were not involved in the previous two steps using the algorithms for updating local parameters such as those in Chapter 2.
4: Repeat this until the required amount of samples are acquired.

4.5 Initialisation of the knots

One consideration when implementing the various MCMC methods is deciding how to initialise the values of the knots. In Chapter 2 we proposed methods for initialising the local parameters so we can continue to use these method in the Global Tractography model. We propose to initialise the knot values using a method based on Deterministic Tractography (see Section 1.7.1) which is described in Algorithm 16.

Algorithm 16 Initialisation of knots

1: Run Deterministic Tractography starting from a voxel within Region 1 which we call $\boldsymbol{P1}$.
2: Finish Deterministic Tractography when we hit Region 2. Choose the final point to be $\boldsymbol{P5}$.
3: Find the points that are a quarter, a half and three quarters of the way through the tract. Denote these points as $\boldsymbol{P2}$, $\boldsymbol{P3}$ and $\boldsymbol{P4}$.

The method described in Algorithm 16 can be used to obtain the initial values of

$\boldsymbol{P2}$, $\boldsymbol{P3}$ and $\boldsymbol{P4}$ when implementing Partially Deterministic Scan MCMC. One of the problems with the method that is proposed to initialise the splines is that it chooses the knots evenly spaced on some line. If the true knots are such that they are approximately equally spaced on the tract, then this method appears to work well. However in practice it is not very likely that we will always have a true tract with approximately equally spaced knots because there are many different tracts that could represent the connection between two regions. For example we may have some spline where the distance between $\boldsymbol{P1}$ and $\boldsymbol{P2}$ is very large when compared to the total distance between $\boldsymbol{P1}$ and $\boldsymbol{P5}$.

A second problem is deciding whether to use more knots to represent the spline. In the literature it is suggested that 5 knots should be used (Jbabdi *et al.*, 2007). Using more knots was quickly discounted as it was attempted and although the estimates were better, it induced a very large computational cost. For one dataset we attempted to look at the tracts that are reconstructed using 5, 10 and 20 uniformly spaced knots. These are plotted in Figure 4.2

To solve the first problem, rather than choosing the knots to be the points that are a quarter, one half and three quarters of the way through the tract that is constructed using Deterministic Tractography, we could instead simulate three uniform random numbers from $U(0,1)$. We then use these simulated numbers as the proportion of the way we go through the distance between $\boldsymbol{P1}$ and $\boldsymbol{P5}$, to allocate values for $\boldsymbol{P2}$, $\boldsymbol{P3}$ and $\boldsymbol{P4}$. This was then attempted in Partially Deterministic Scan MCMC. Unfortunately the mixing in MCMC does not appear to be very good, because at some iterations candidate knot values will be accepted, but this is not a frequent occurrence. This can be explained by the fact that most of the time the MCMC algorithm will propose knots that are completely far away from the truth. As an example a dataset was simulated and then different uniform numbers were generated and the corresponding tracts

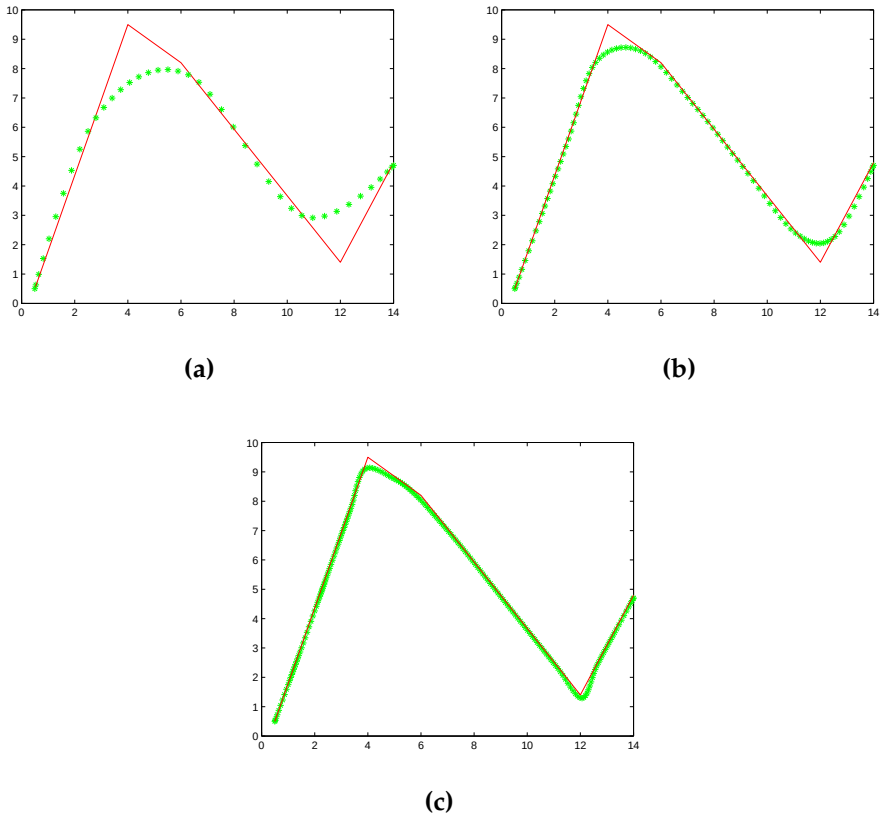


Figure 4.2: The true fibre orientation (red) compared to tracts that are reconstructed by using a varying number of uniformly spaced knots (green) in Algorithm 16 on a simulated dataset. (a) 5 knots, (b) 10 knots and (c) 20 knots. We see that as the number of knots increases the estimated spline gets closer to the true fibre orientation.

using these numbers were plotted. These can be seen in Figure 4.3. From these graphs we can see that some of the tracts are quite close to the true tract while others are very bad.

To overcome the problem with the mixing, another idea that we proposed is described in Algorithm 17. We generated one hundred sets of random num-

Algorithm 17 Improved initialisation of knots

1: Simulate many random numbers from $U(0,1)$.
2: For each set of three numbers, obtain the corresponding knot values using the method described in Algorithm 16.
3: Calculate the value of the posterior distribution and find the set of numbers which gives the maximum.
4: In Partially Deterministic Scan MCMC, start at these maximum knot values and propose new values of the knots using a random-walk proposal distribution

bers from $U(0,1)$ and calculated the posterior density using the knots values that corresponded to these random sets of numbers using Algorithm 17. The four sets of knots that produced the highest posterior distribution values were plotted in Figure 4.4. The splines that are plotted look very close to the true fibre orientation. The methods that we use to infer the knots and to initialise the knots are different from the methods used by Jbabdi *et al*. (2007).

4.6 2D example

We will now attempt to estimate the local and global parameters of a 2D simulated dataset by using Partially Deterministic Scan MCMC. In the dataset, shown in Figure 4.5, which we denote dataset 3, the whole area was covered by 11×3 voxels, so that in total there are 33 voxels. The first region of interest, shown in red, is all the voxels such that $x \in [0,3]$ and $y \in [0,3]$ which we denote as Region A, the second region, shown in blue, is all voxels such that $x \in [8,11]$ and $y \in [0,3]$ which we define to be Region B. The knots of

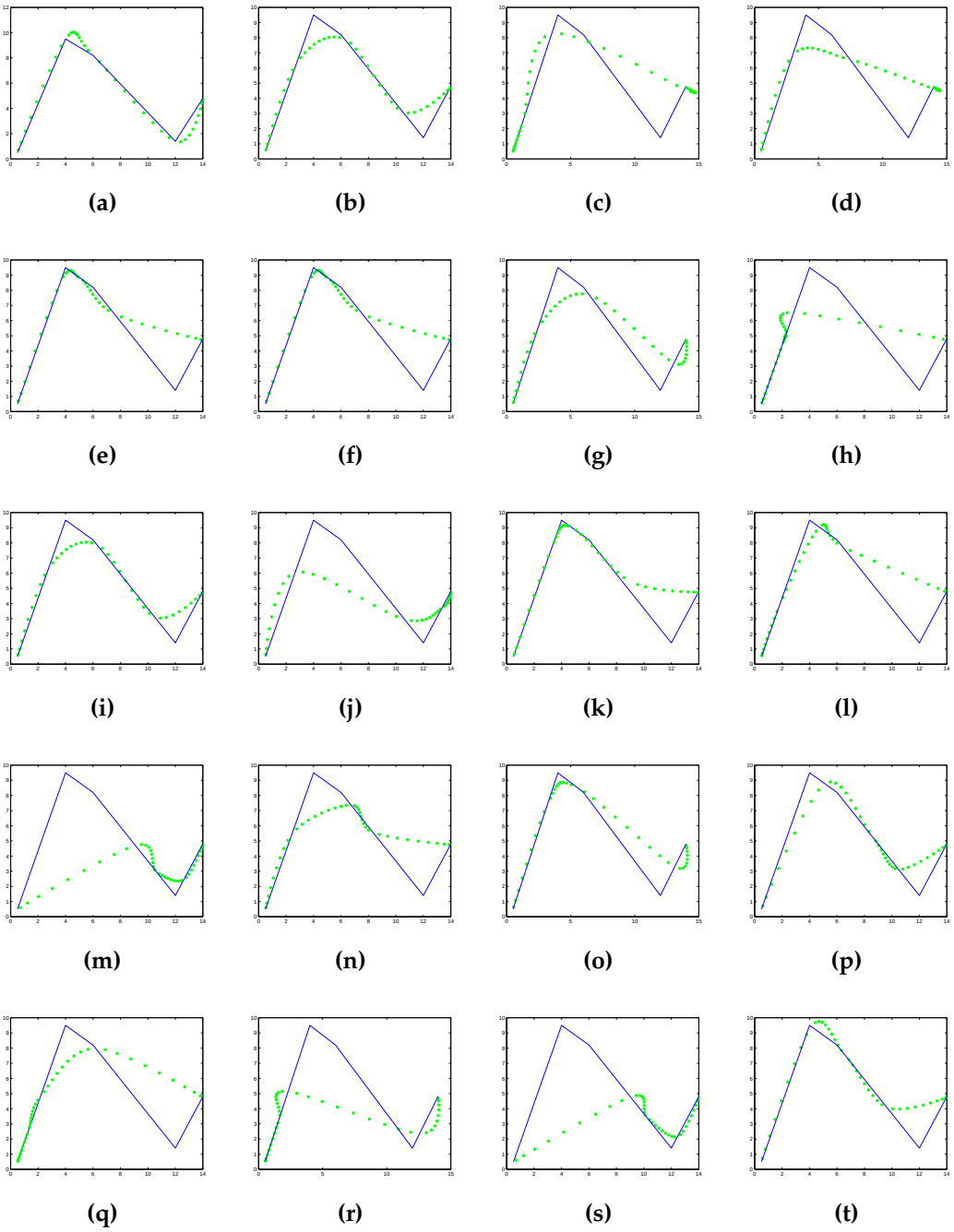


Figure 4.3: The true fibre orientation (blue) compared to tracts that are reconstructed by using different spacing within Algorithm 16 to obtain 5 knots (green), using (a) the true knots, (b) the knots exactly one quarter, one half and three quarters of the way through and (c)-(t) randomly selected knots from $U(0,1)$

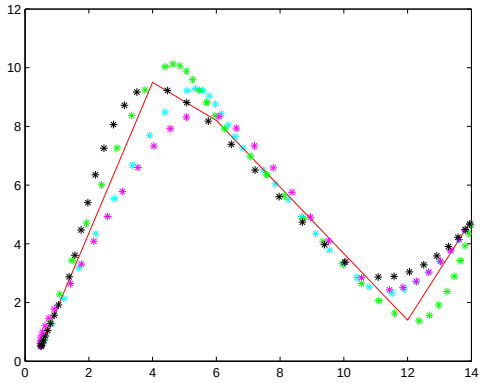


Figure 4.4: The four sets of knots out of one hundred sets of simulated knots that are reconstructed using various spacing in Algorithm 16 that have the highest posterior distribution. These are plotted against the true orientation (solid line)

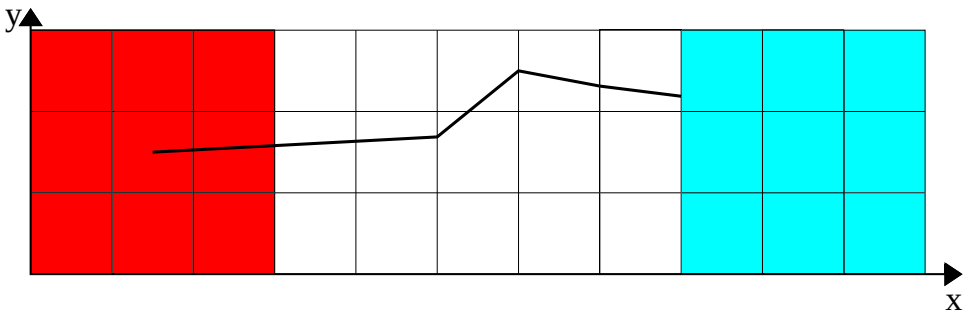
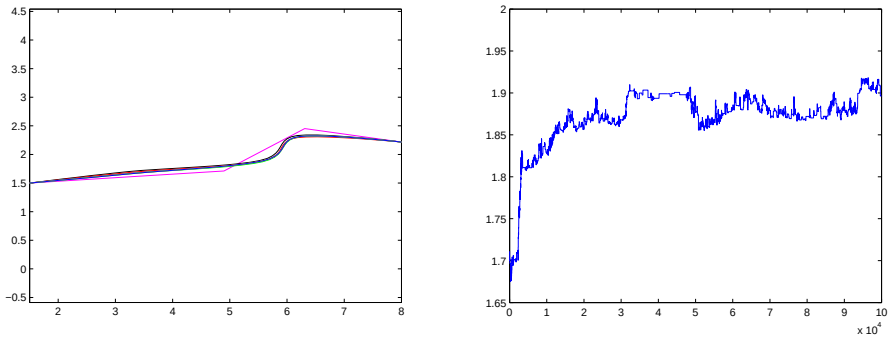


Figure 4.5: Dataset 3 that connects the two regions with a spline that uses the parameters within the Global Tractography model. Each voxel is modelled using the partial volume model with one fibre orientation.

the tract are $\boldsymbol{P1} = (1.5, 1.5)$, $\boldsymbol{P2} = (5, 1.7)$, $\boldsymbol{P3} = (6, 2.5)$, $\boldsymbol{P4} = (7, 2.3)$ and $\boldsymbol{P5} = (8, 2.2)$.

We now investigate the results when implementing Partially Deterministic Scan MCMC on dataset 3 to determine how well our algorithm works. Partially Deterministic Scan MCMC was initiated using a random-walk proposal distribution. First Deterministic Tractography was applied to initialise the knot values, also the Laplace approximation was used to get local parameter estimates and a proposal covariance matrix for each voxel. For simplicity, initially, $\boldsymbol{P1}$ and $\boldsymbol{P5}$ had their values fixed to the true values of the knots The plot of the splines



(a) **(b)**

Figure 4.6: (a) The random splines from the Partially Deterministic Scan MCMC estimates compared with the true spline (pink) when updating all knots except for the end knots in the Global Tractography model and (b) the mixing within MCMC for one coordinate of a knot.

using random samples of knots from the MCMC results are plotted alongside the true fibre orientation in Figure 4.6. The mixing in MCMC is not as good as we would like it to be. We will discuss a solution to this in Section 4.8.

4.7 Moving to 3D

We now attempt to estimate the global and local parameters of a 3D dataset. Extending the problem to 3D should be fairly easy as there is only one more coordinate for each knot plus θ for each voxel to estimate. A dataset was simulated with $\boldsymbol{P1} = (1.5, 1.5, 1.5)$, $\boldsymbol{P2} = (3, 2.7, 1.2)$, $\boldsymbol{P3} = (5, 1.5, 0.7)$, $\boldsymbol{P4} = (7, 2.1, 0.5)$ and $\boldsymbol{P5} = (8, 1.5, 0.9)$ where the true fibre orientation between these knots is a straight line. In the x direction there are 11 voxels while in the y and z directions there are 3 voxels. Thus in total there are 99 voxels. Region 1 was defined to be the voxels where the x coordinate is less than or equal to 3, while Region 2 was the voxels where the x coordinate is greater than or equal to 8.

As in the 2D example the values of $\boldsymbol{P2}$, $\boldsymbol{P3}$ and $\boldsymbol{P4}$ were inferred by using Partially Deterministic Scan MCMC while the values of the two end knots were

held to make the problem slightly easier. Then all the knots including $\boldsymbol{P1}$ and $\boldsymbol{P5}$ and all the local parameters were inferred using Partially Deterministic Scan MCMC.

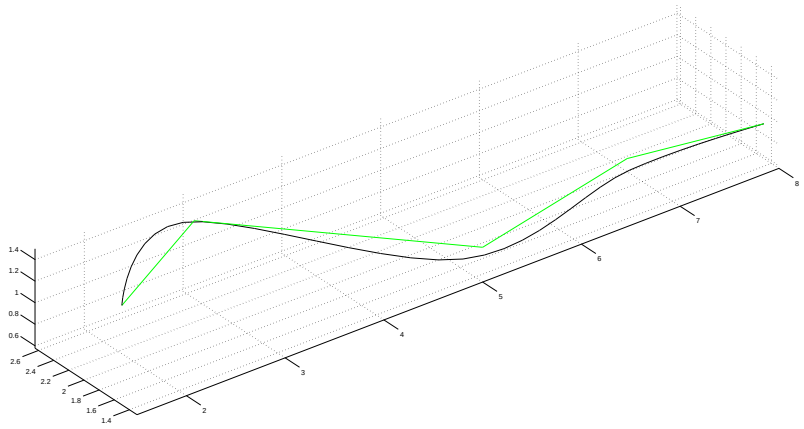
The results that show the approximate spline alongside the true fibre orientation in both the cases when $\boldsymbol{P1}$ and $\boldsymbol{P5}$ are held and when they aren't are in Figure 4.7. Also in Figure 4.7, there is a traceplot of the mixing of a coordinate of a knot in Partially Deterministic Scan MCMC. This is in the case when we are inferring the values of all 5 knots. The only problem with Partially Deterministic Scan MCMC seems to be tuning issues, but this could be solved by using Adaptive MCMC (Section 2.3.3). We will now look at implementing Adaptive MCMC within Partially Deterministic Scan MCMC.

4.8 Using Adaptive MCMC within the Global Tractography model

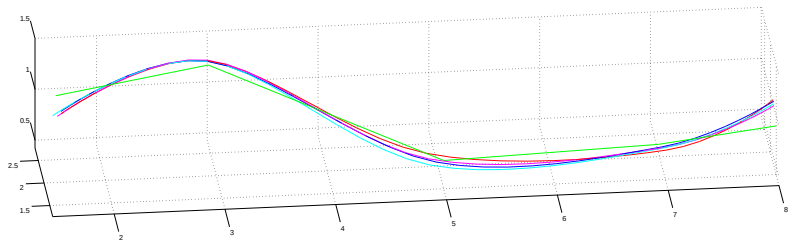
When implementing Partially Deterministic Scan MCMC, one problem is determining a good covariance matrix for the proposal distribution of the knots that allows sufficient mixing. In Chapter 2 when inferring the local parameters in a voxel using MCMC, we successfully used Adaptive MCMC to calculate a covariance matrix for the proposal distribution at each stage of the MCMC (see Section 2.3.3) that permitted good mixing. The algorithm for inferring the local and global parameters of a 3D dataset by including Adaptive MCMC in Partially Deterministic Scan MCMC is described in Algorithm 18.

4.8.1 Example

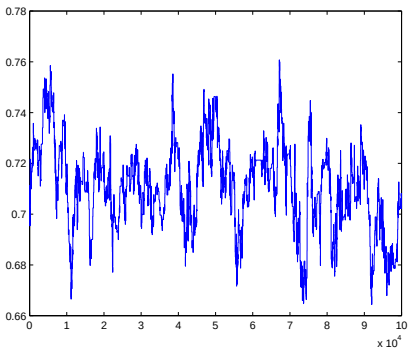
We now infer the values of the parameters of the 3D dataset using Partially Deterministic Scan MCMC with the Adaptive MCMC extension, to determine if



(a)



(b)



(c)

Figure 4.7: The splines that are constructed by estimating the knots of the Global Tractography model using Partially Deterministic Scan MCMC against the true orientation (green) when (a) updating the three non-end knots, (b) when updating all knots. (c) The typical mixing of one knot.

Algorithm 18 Adaptive MCMC (Global Tractography)

1: Run the Partially Deterministic Scan MCMC algorithm for a given number of iterations.
2: Calculate the empirical covariance matrix as described in Section 2.3.3 for all the coordinates in each knot, such that we will have a 3×3 matrix for each knot. Then using this matrix calculate the proposal distribution covariance matrix as described in Section 2.3.3.
3: In MCMC use the calculated covariance matrix in the proposal distribution for the knots and propose new values of the local parameters from a random-walk Normal distribution from the Laplace approximation. Then accept or reject these values using Partially Deterministic Scan MCMC.
4: Go back to Step 2 and continue until we have the required number of estimates.

this extension improves the results. We will first infer the three non-end knots before inferring the values of all five knots. In both the MCMC where we infer three knots and all five knots, first the knots are initialised by using the uniform knots of a tract that is obtained using Deterministic Tractography (Section 4.5). Once this has been done then Partially Deterministic Scan MCMC is implemented for 10000 iterations, before Adaptive MCMC is run for another 90000 iterations.

4.8.2 Updating 3 knots only

At first we decided to just update the values of $\boldsymbol{P2}$, $\boldsymbol{P3}$ and $\boldsymbol{P4}$, whilst holding the values of $\boldsymbol{P1}$ and $\boldsymbol{P5}$ so that the problem was easier. The plots of the traceplots of the MCMC estimates for some of the knots are shown in Figure 4.8 which reveal fairly good mixing. Further in Figure 4.9 we see some of the randomly selected splines from the MCMC results in comparison to the true spline. These figures show that the splines obtained from Partially Deterministic Scan MCMC correspond to the true spline. We have seen that Adaptive MCMC greatly improves the mixing, so we attempt to infer the values of all five knots.

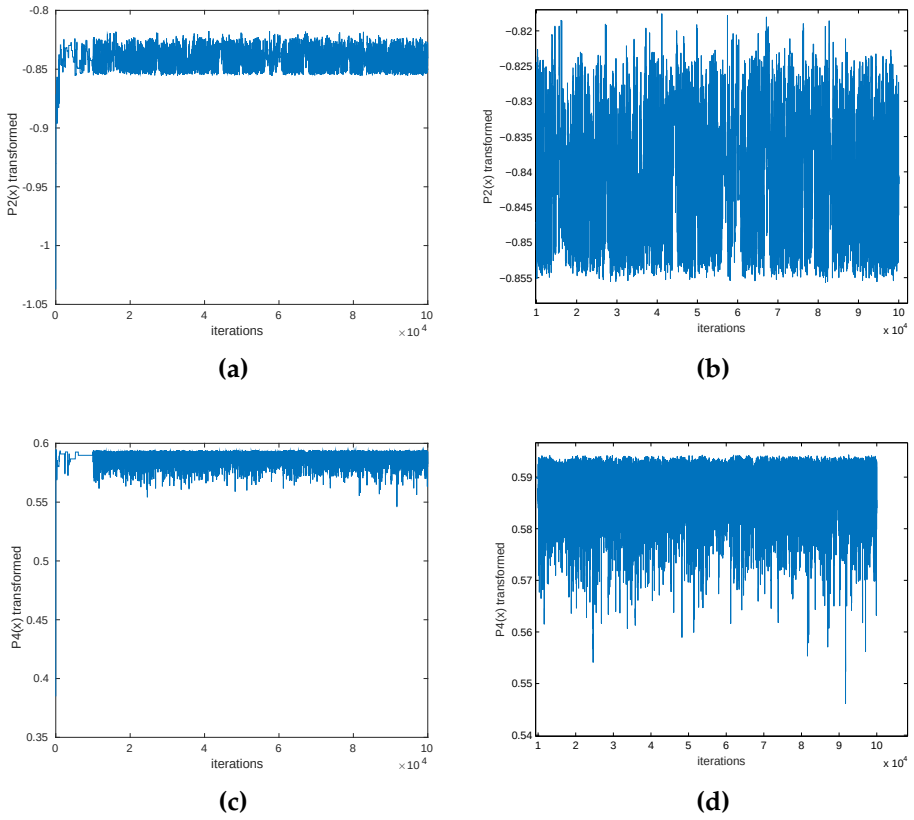


Figure 4.8: The MCMC traceplots when using the Adaptive MCMC extension of Partially Deterministic Scan MCMC to simulate the estimated values of the three non-end knots and the local parameters from the posterior distribution of the Global Tractography model (a) $P2(x)$, (b) $P2(x)$ when the burn in iterations are removed, (c) $P4(x)$ and (d) $P4(x)$ when the burn in period is removed. The dataset that is investigated is a 3D simulated dataset.

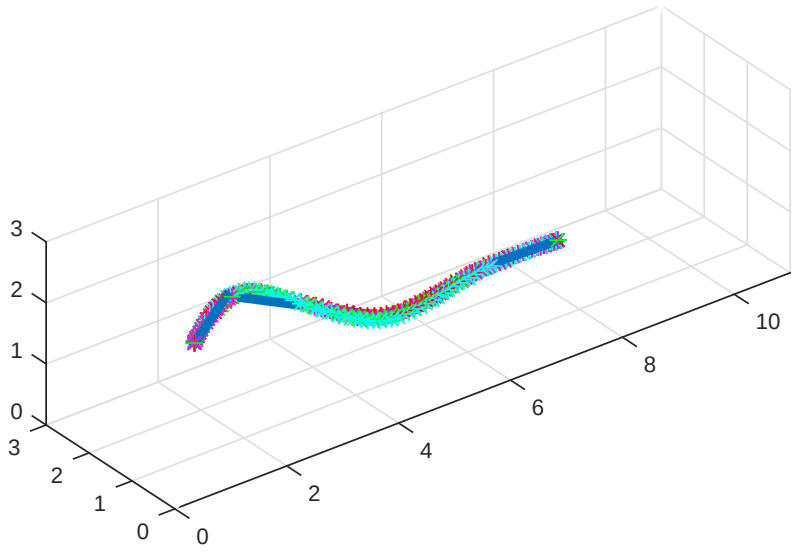


Figure 4.9: The true spline (blue) compared to four randomly selected splines when simulating the estimated values of the three non-end knots and the local parameters from the posterior distribution of the Global Tractography model using the Adaptive MCMC extension of Partially Deterministic Scan MCMC.

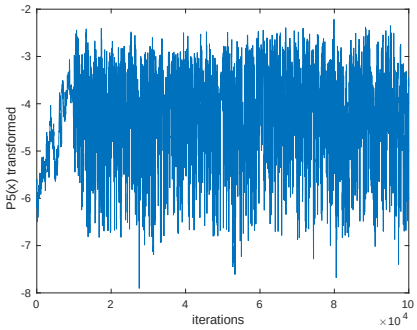


Figure 4.10: The MCMC traceplot of one knot coordinate when using the Adaptive MCMC extension of Partially Deterministic Scan MCMC to simulate the estimated values all knots and the local parameters from the posterior distribution of the Global Tractography model for $P5(x)$. The dataset that is investigated is a 3D simulated dataset.

4.8.3 Updating 5 knots

We now extend the algorithm to estimate the values of all 5 knots and all the local parameters. The traceplot of the MCMC estimates for one knot coordinate is shown in Figure 4.10. The mixing in this traceplot looks good once we start to use the Adaptive MCMC extension. We selected four samples from these estimates and compared them with the true fibre orientation, as in Figure 4.11. The splines from the MCMC estimates match the true fibre orientation.

We are now able to obtain estimates of the knots and local parameters with sufficient mixing by using Adaptive MCMC within Partially Deterministic Scan MCMC. Therefore we will now execute model selection to determine whether a connection between two brain regions exists.

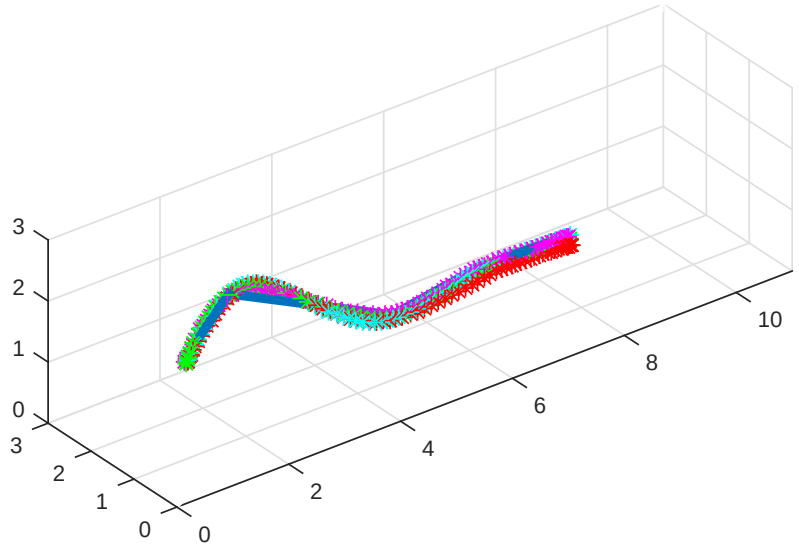


Figure 4.11: The true spline (blue) compared to four randomly selected splines when simulating the estimated values of all knots and the local parameters from the posterior distribution of the Global Tractography model using the Adaptive MCMC extension of Partially Deterministic Scan MCMC.

4.9 Model selection for the existence of a priori known connection

We now want to test whether there is evidence for the existence of an anatomical connections between brain regions of interest. A functional connection between regions in the brain is dependent on there being an anatomical connection (Passingham *et al*., 2002), so if we can judge whether there is an anatomical connection between two brain regions then it can help in determining potential functional connections in fMRI experiments.

Until now we have managed to estimate the values of the knots and the local parameters within the Global Tractography framework in Section 4.3. We now want to infer the connectivity matrix $\boldsymbol{\mathcal{C}}$. The elements within $\boldsymbol{\mathcal{C}}$ can either take the value 1 or 0. As an example if the (i, j)th element of $\boldsymbol{\mathcal{C}}$ is equal to 1, this means there is a connection between regions i and j. Otherwise if the element is 0, it means there is no connection between these regions. Jbabdi *et al*. (2007) suggested using model selection to see if there is a connection between any two regions. We assume that there are only two regions of interest Region 1 and Region 2, such that $\boldsymbol{\mathcal{C}}$ is a 2×2 matrix. The two models we will select from are defined as,

M_1: The constrained model such that $\pi(c_{12} = 1) = 1$.

M_0: The unconstrained model.

Jbabdi *et al*. (2007) use the Bayes factor (see Section 3.2) to choose between the two models which they approximate with the harmonic mean approximation (see Section 3.2.2). Despite being easy to implement the harmonic mean approximation can be a very bad estimate (Raftery *et al*., 2007). Therefore it is proposed that we will use the Bayes factor, but we will approximate it using the much more robust methods based on Thermodynamic Integration that are

described in Section 3.4

4.9.1 Estimation of Bayes factor

We now use methods based on Thermodynamic Integration to approximate the Bayes factor in favour of the model with a connection between two brain regions against the model with no such connection. To implement this we will make use of the likelihood and prior distributions introduced in the Global Tractography framework in Section 4.3. The only difference from when we were inferring the local parameters and the knots is that now we also have to include $\pi(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{C}})\pi(\boldsymbol{\mathcal{C}})$ in the posterior distribution $\pi(\boldsymbol{\Omega}|\boldsymbol{Y})$ in Section 4.3. We assume that there are only two regions of interest Region 1 and Region 2, such that $\boldsymbol{\mathcal{C}}$ is a 2×2 matrix. If we first investigate the model where we assume that $c_{12} = 1$, which we define to be model M_1, then for any knots such that Region 1 and Region 2 are connected $\pi(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{C}}) = 1$, otherwise if knots are proposed that do not connect the two regions then $\pi(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{C}}) = 0$. Similarly for model M_0 where there is no such connection, if knots are proposed such that there is a connection between the regions then $\pi(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{C}}) = 0$, otherwise $\pi(\boldsymbol{\mathcal{F}}|\boldsymbol{\mathcal{C}}) = 1$.

We can approximate the Bayes factor by using either Annealing-Melting Integration (Section 3.4.1) or Model-Switch Integration (Section 3.4.3). By looking more closely at the Model-Switch Integration approximation that is

$$\begin{aligned}
&\log\left(\frac{\pi(\boldsymbol{y}|M_0)}{\pi(\boldsymbol{y}|M_1)}\right) \\
&\approx 0.5\sum_{i=0}^{n-1}(t_{i+1}-t_i)\left(E_{\boldsymbol{\Omega}|\boldsymbol{y},t_{i+1}}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\Omega},M_1)\pi(\boldsymbol{\Omega}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\Omega},M_0)\pi(\boldsymbol{\Omega}|M_0)}\right)\right]\right. \\
&\left.+\, E_{\boldsymbol{\Omega}|\boldsymbol{y},t_i}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\Omega},M_1)\pi(\boldsymbol{\Omega}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\Omega},M_0)\pi(\boldsymbol{\Omega}|M_0)}\right)\right]\right)
\end{aligned}$$

where

$$E_{\boldsymbol{\Omega}|\boldsymbol{y},t_i}\left[\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\Omega},M_1)\pi(\boldsymbol{\Omega}|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\Omega},M_0)\pi(\boldsymbol{\Omega}|M_0)}\right)\right]$$

$$\approx \frac{1}{p-k+1}\sum_{j=k}^{p}\log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\Omega}_j^i,M_1)\pi(\boldsymbol{\Omega}_j^i|M_1)}{\pi(\boldsymbol{y}|\boldsymbol{\Omega}_j^i,M_0)\pi(\boldsymbol{\Omega}_j^i|M_0)}\right)$$

such that $\boldsymbol{\Omega}_j^i$ is the jth MCMC estimate of the power posterior from the ith temperature. The MCMC is run for p iterations where the first k iterations are defined to be the burn in period. It can be observed that problems will occur in this approximation because either $\pi(\boldsymbol{\Omega}|M_1)$ or $\pi(\boldsymbol{\Omega}|M_0)$ will be equal to 0, which means at some point in the approximation we will have log(0). Therefore this problem is avoided completely by instead using Annealing-Melting Integration (Section 3.4.1).

We must include the precision τ in the local parameters when calculating the Annealing-Melting Integration estimate because the likelihood must be the full likelihood.

4.9.2 Global Tractography when there is no connection

Previously when we have been inferring the values of the parameters in the Global Tractography framework, we have assumed that a connection exists between the two regions, as in M_1. To calculate the estimate of the logarithm of the Bayes factor we will also have to infer the model where there is no such connection.

When running Annealing-Melting integration the following conditions on the knots must be met for a candidate set of knots to be accepted.

M_1

Within the MCMC $\boldsymbol{P1}$ is constrained to be within Region 1 and $\boldsymbol{P5}$ is constrained to be in Region 2 such that

- $\boldsymbol{P1}$ is in Region 1
- $\boldsymbol{P5}$ is in Region 2

M_0
Either

- $\boldsymbol{P1}$ is in Region 1 and $\boldsymbol{P2}$, $\boldsymbol{P3}$, $\boldsymbol{P4}$ and $\boldsymbol{P5}$ are not in Region 2.
  or
- $\boldsymbol{P1}$, $\boldsymbol{P2}$, $\boldsymbol{P3}$ and $\boldsymbol{P4}$ are not in Region 1 and $\boldsymbol{P5}$ is in Region 2
  or
- $\boldsymbol{P5}$ is in Region 1 and $\boldsymbol{P1}$, $\boldsymbol{P2}$, $\boldsymbol{P3}$ and $\boldsymbol{P4}$ are not in Region 2
  or
- $\boldsymbol{P2}$, $\boldsymbol{P3}$, $\boldsymbol{P4}$ and $\boldsymbol{P5}$ are not in Region 1 and $\boldsymbol{P1}$ is in Region 2
  or
- $\boldsymbol{P1}$, $\boldsymbol{P2}$, $\boldsymbol{P3}$, $\boldsymbol{P4}$ and $\boldsymbol{P5}$ are not in Region 1 and $\boldsymbol{P1}$, $\boldsymbol{P2}$, $\boldsymbol{P3}$, $\boldsymbol{P4}$ and $\boldsymbol{P5}$ are not in Region 2.

Before calculating the Annealing-Melting Integration estimate of the logarithm of the marginal likelihood for the model where there is no connection, we first infer the knots and the local parameters in the case when there is no connection between the two regions. This will help us decide on a MCMC method for implementing Annealing-Melting Integration. We have proposed two potential methods which we will now describe.

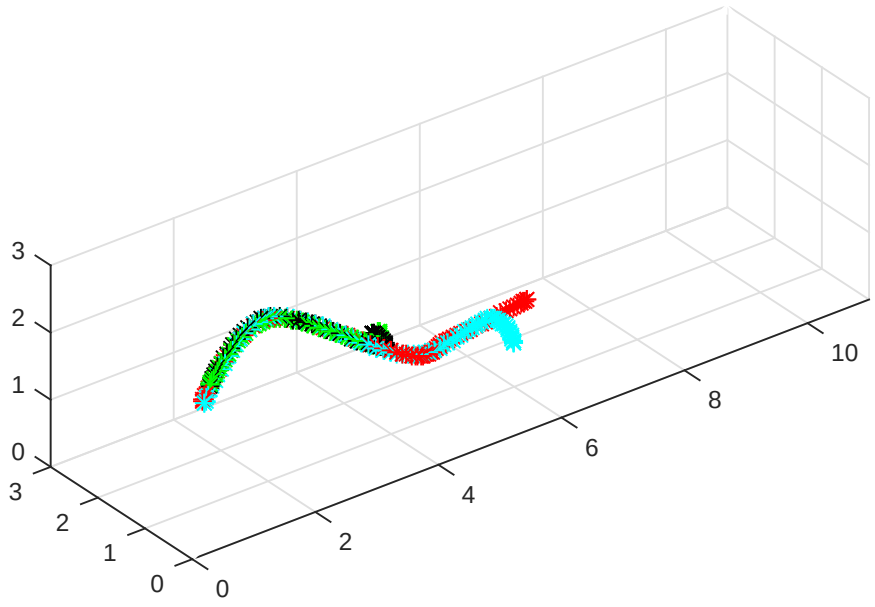


Figure 4.12: The splines constructed using the knot estimates obtained by simulating knot and local parameter estimates from the posterior distribution of the Global Tractography model that is constrained so that there is no connection between the two brain regions of interest by using BNC MCMC.

4.9.3 Basic No Connection MCMC

The first method to infer the parameters in model M_0 ensures that the only condition within MCMC is that there is no spline that connects the two brain regions such as the conditions for M_0 in the previous section. This condition is enforced by rejecting proposed knots in the MCMC that connect regions. This is the simplest case of MCMC when inferring the parameters when there is no connection and will be the easiest to run. We will name this MCMC algorithm Basic No Connection MCMC (BNC MCMC). Splines that are reconstructed by using samples from the MCMC results are shown in Figure 4.12.

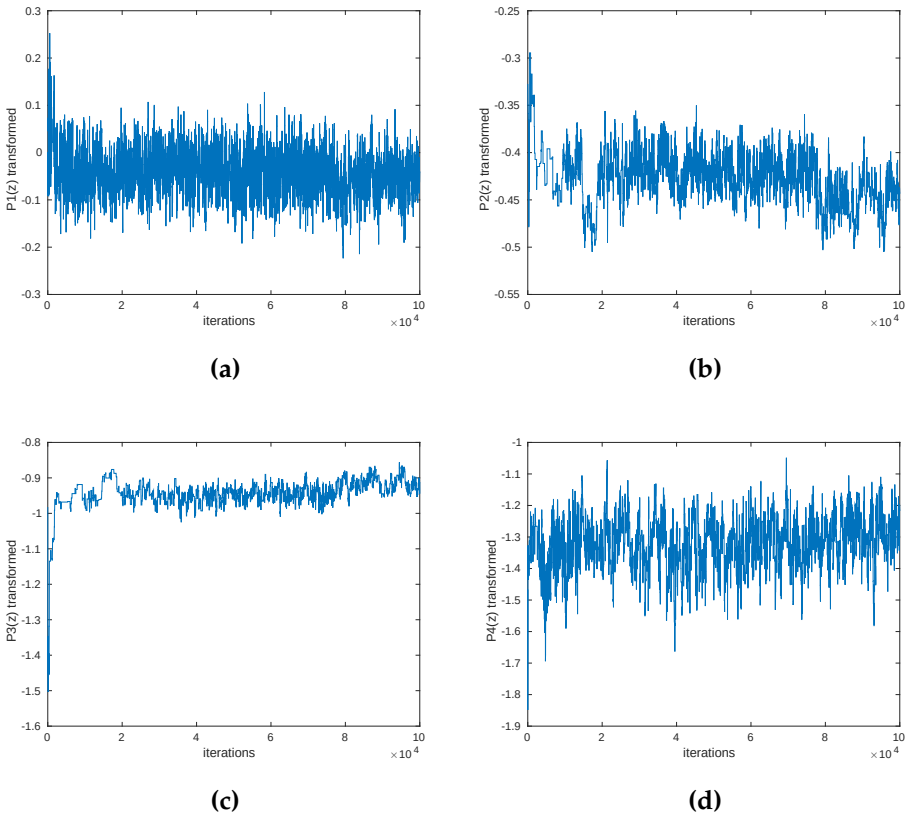


Figure 4.13: The traceplots for the transformed values of the knots (a) $P1(z)$, (b) $P2(z)$, (c) $P3(z)$, and (d) $P4(z)$. We simulate knot and local parameter estimates from the posterior distribution of the Global Tractography model that is constrained so that there is no connection between the two brain regions of interest by using CNC MCMC.

4.9.4 Constrained No Connection MCMC

When implementing MCMC within Annealing-Melting Integration, we could use the method that Jbabdi *et al*. (2007) use to infer knots in the model with no connection. In this method first MCMC is used to obtain samples of the knots in the model M_1 where a connection is enforced between the two regions. Then splines that are constrained such that they have the same length as the splines in model M_1 but are otherwise unconstrained are simulated for model M_0. We will refer to the MCMC algorithm of Jbabdi *et al*. (2007) as Constrained No Connection MCMC (CNC MCMC). By using CNC MCMC to get estimates of the knots and local parameters we obtain the traceplots in Figure 4.13. The

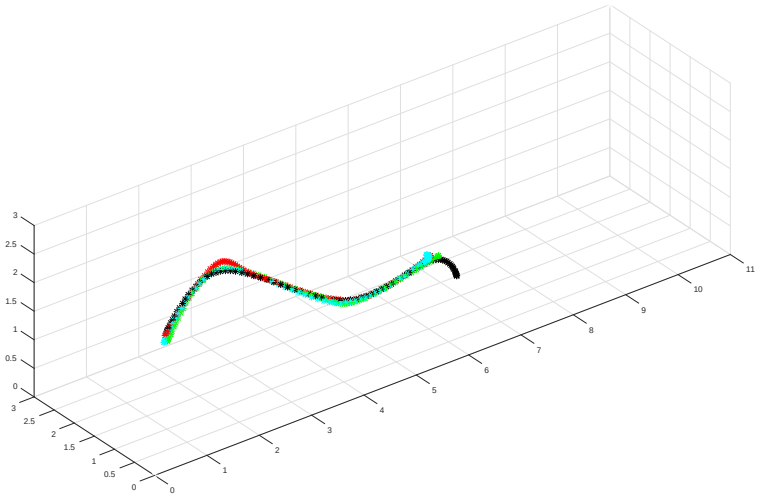


Figure 4.14: The splines constructed using the knot estimates obtained by simulating knot and local parameter estimates from the posterior distribution of the Global Tractography model that is constrained so that there is no connection between the two brain regions of interest by using CNC MCMC.

graph of the splines obtained from the CNC MCMC estimates is in Figure 4.14.

4.10 Examples for Annealing-Melting Integration

We now investigate two simulated datasets by calculating the Annealing-Melting Integration estimate for the logarithm of the marginal likelihood. We will calculate this estimate for both the model with no connection between two brain regions and the model with a connection between two brain regions which we denote M_0 and M_1. These estimates will then be used to calculate an estimate for the logarithm of the Bayes factor in favour of model M_1 over model M_0. We will investigate one dataset where there is a connection between two brain regions and one dataset where there is no such connection. For simplicity in both datasets each of the voxels only has one fibre orientation. In the example with a connection we will use the 3D dataset that we have been using throughout this chapter (Section 4.7). In the example with no connection we will simulate a

dataset with completely random local parameter values such that there is overwhelming evidence of no connection.

4.10.1 Example 1 - a dataset with a connection

By investigating the 3D dataset that we have worked with throughout this chapter (Section 4.7) we will calculate three approximations for the logarithm of the marginal likelihood. The first will be the approximation when there is the constraint of there being a connection (i.e. model M_1). To implement this we will use a slightly altered version of Partially Deterministic Scan MCMC, such that we infer the power posterior by using Annealing-Melting Integration. Then we will calculate the two approximations when there is no such connection (i.e. model M_0) by using Annealing-Melting Integration based on BNC MCMC (Section 4.9.3) and CNC MCMC (Section 4.9.4).

Initially in the Annealing-Melting Integration method we choose the temperatures to be equally spaced where we have 11 temperatures, so that the algorithm can be run quickly. Each of the three methods were run for 20000 iterations for each temperature, 10000 of which were implemented before Adaptive MCMC. The results when excluding the Expected deviance value that corresponds to $t = 0$ are shown in Table 4.1. The Expected deviance when $t = 0$, has a big influence on the approximation of the logarithm of the marginal likelihood in all three methods. The reason for this is that when $t = 0$ we are just sampling from the prior distribution. Then the knots do not have any influence on the power posterior, and therefore any values are accepted. This then causes the high magnitude values of the Expected deviance (Equation 3.4.6).

From these results we see that the logarithm of the Bayes factor is 609.6 in favour of the no connection model when using the CNC MCMC approximation and

Spacing	CNC	BNC	Partial
$c = 1$	$-8.6300\text{x}10^3$	$-9.1654\text{x}10^3$	$-9.2396\text{x}10^3$
$c = 5$	$-1.0943\text{x}10^4$	$-9.0757\text{x}10^3$	$-1.7978\text{x}10^4$

Table 4.1: The estimates of the logarithm of the marginal likelihood by using Annealing-Melting Integration; for the CNC MCMC method when there is no connection, the BNC MCMC method and finally the method when there is a connection by using Partially Deterministic Scan MCMC. Here we use 11 temperatures and temperature spacings of $c = 1$ and $c = 5$.

74.2 in favour of the no connection model when using the BNC MCMC approximation. We now try to execute Annealing-Melting Integration with spacing $c = 5$ because Friel and Pettitt (2008) found that in general c works well when its value is between 3 and 5.

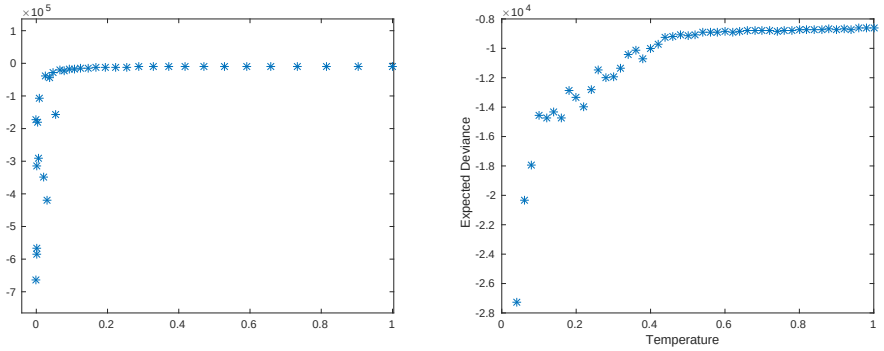
After running the algorithm with 11 temperatures such that the spacing is $c = 5$, we obtain the results in Table 4.1. From the results we estimate that the logarithm of the Bayes factor in favour of the no connection model is 7035 when using the CNC MCMC approximation and 10874 when using the BNC MCMC approximation.

We observed in the results that the approximations of the logarithm of the no connection marginal likelihood do not change too much when using spacing $c = 1$ and $c = 5$. However the approximation of the logarithm of the marginal likelihood when there is a connection changes quite significantly when using spacing $c = 1$ and $c = 5$. We will now use 51 temperatures with spacing $c = 5$ to attempt to obtain better estimates. The results can be seen in Table 4.2.

The estimate of the logarithm of the marginal likelihood in the model where there is a connection appears to be bad when using spacing $c = 5$ which we can observe by looking at the plot of the Expected deviance which is in Figure 4.15 (a). Therefore we now run Annealing-Melting Integration with 51 temperatures

CNC	BNC	Partial
$-1.0789\text{x}10^4$	$-1.0829\text{x}10^4$	$-3.6914\text{x}10^4$

Table 4.2: The estimates of the logarithm of the marginal likelihood by using Annealing-Melting integration; for the CNC MCMC method when there is no connection, the BNC MCMC method and finally the method when there is a connection based on Partially Deterministic Scan MCMC. Here we use the 51 temperatures and spacing of $c = 5$.



(a) **(b)**

Figure 4.15: The temperatures against the Expected deviance when using spacing (a) c=5 and (b) c=1 to estimate the marginal likelihood for model M_1.

with spacing $c = 1$. When we calculate this new estimate for the logarithm of the marginal likelihood in favour of the model with a connection, the approximation is $-1.0259\text{x}10^4$. Immediately we can see that this result looks more sensible than the approximation using $c = 5$ by observing the plot of the Expected deviance in Figure 4.15 (b). We can see that the plot of the Expected deviance is much more smooth and hence the trapezium rule estimate will be better.

We then use the new estimate for the logarithm of the marginal likelihood in favour of the model with a connection and compare it with our best approximations for the model with no connection. With the CNC MCMC approximation the approximation of the logarithm of the Bayes factor in favour of the model with a connection is 530, whilst the corresponding approximation when using BNC MCMC is 570. Thus we have strong evidence in favour of the model with a connection.

4.10.2 Example 2 - a dataset with no connection

We now attempt model selection on a dataset where there is no connection between the two brain regions of interest. In this example the dataset that we simulate has 11 voxels in the x direction and 3 voxels in both the y and z directions. We denote region 1 to be all the voxels such that $x \leq 3$ and region 2 to be all the voxels such that $x \geq 8$. We choose the local parameter values in each voxel to be completely random so that we expect there is no anatomical connection between the two regions.

To initialise the knot values we first attempt to implement the method based on Deterministic Tractography (Section 4.5) starting from every possible voxel in region 1. No tract could be produced that connects region 1 and region 2 which makes sense because there is no connection between the regions. Therefore in this method we start with the spline that we used in Example 1 as the initial spline.

As in Example 1 initially we implement Annealing-Melting Integration using 11 equally spaced temperatures. There seems to be problems with the acceptance rate of samples when using the CNC MCMC method for implementing Annealing-Melting Integration in this example. Therefore we dismiss CNC MCMC and just use BNC MCMC for the model where there is no connection. The approximation of the logarithm of the marginal likelihoods are shown in Table 4.3

From the results we see that the logarithm of the Bayes factor is 1758 in favour of the model with a connection. We will now use temperature spacing of $c = 5$ instead of $c = 1$ when calculating the Annealing-Melting Integration estimates for the marginal likelihood, as we would prefer to focus more on the higher

Spacing	No connection	Connection
$c = 1$	$-1.3114\text{x}10^4$	$-1.1356\text{x}10^4$
$c = 5$	$-1.1858\text{x}10^4$	$-2.3446\text{x}10^4$

Table 4.3: The estimates of the logarithm of the marginal likelihood by using Annealing-Melting Integration; for the BNC MCMC method and the method when there is a connection based on Partially Deterministic Scan MCMC. Here we use 11 temperatures and temperature spacings of $c = 1$ and $c = 5$.

No connection	Connection
$-1.2732\text{x}10^4$	$-2.4798\text{x}10^4$

Table 4.4: The estimates of the logarithm of the marginal likelihood by using Annealing-Melting Integration; for the BNC MCMC method and finally the method when there is a connection based on Partially Deterministic Scan MCMC. Here we use the 51 temperatures and temperature spacing $c = 5$.

temperatures due to the problems when $t = 0$. The results are in Table 4.3.

From these results we see that the logarithm of the Bayes factor is 11588 in favour of the no connection model. As in Example 1 we can observe that the approximation for the model with a connection when using $c = 1$ changes drastically when using $c = 5$. Finally we use more temperatures when the spacing is $c = 5$ to attempt to obtain a better approximation. The results are in Table 4.4.

As in Example 1 we can observe that the Annealing-Melting integration approximation with spacing $c = 5$ when there is a connection is not very good. We verify this by looking at the plot of the Expected deviance in Figure 4.16. Therefore for this model we attempt Annealing-Melting Integration again for the model with a connection using 51 equally spaced temperatures. The approximation of the logarithm of the marginal likelihood in favour of a connection is then $-1.6378\text{x}10^4$. When using the approximation of the logarithm of the marginal likelihood for the model where there is no connection which is $-1.2732\text{x}10^4$ we can calculate the logarithm of the Bayes factor in favour of the model with a connection to be -3646. Therefore in both Example 1 and Example 2 the result

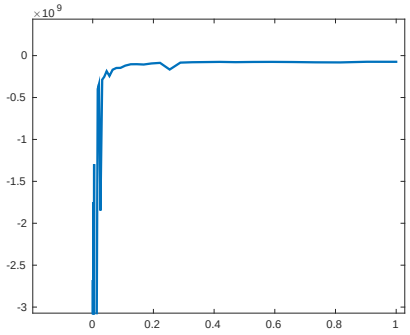


Figure 4.16: The temperatures against the Expected deviance when using spacing c=5 to estimate the marginal likelihood for model M_1.

of the logarithm of the Bayes factor corresponds to the true model.

4.11 Conclusions

The main aim of this chapter was to find some way of determining whether regions of the brain are connected by incorporating prior knowledge and using model selection methods. First we introduced the Global Tractography framework that was proposed by Jbabdi *et al*. (2007). We altered this framework slightly to take into account the conclusions in Chapter 3 that the ARD prior was inadequate.

Within the framework our first goal was to infer the global parameters. Already in Chapter 2 we had found ways to infer the local parameters in the Global Tractography model, thus we just needed to extend the work in Chapter 2 to also infer the global parameters. To aid in the inference we proposed a method based on Deterministic Tractography for initialising the values of the knots in the Global Tractography model. Then three methods were proposed for running MCMC to infer the knots in the Global Tractography model when there is a connection between two brain regions of interest. Partially Deterministic Scan MCMC was found to be the best method from the proposed methods.

Although the mixing in Partially Deterministic Scan MCMC was adequate, it was not as good as we would wish for it to be, therefore we applied Adaptive MCMC to Partially Deterministic Scan MCMC to automatically calculate a proposal covariance matrix for the knot values after a given number of iterations. This significantly improved the mixing in the Partially Deterministic Scan MCMC algorithm. Due to the introduction of Adaptive MCMC in the Partially Deterministic Scan MCMC algorithm we now have a robust algorithm for inferring the values of both the knots and the local parameters in the Global Tractography framework.

The next step was to then execute model selection between the model with a connection between two brain regions and the model with no such connection. From this we can determine whether or not it is likely that there is an anatomical connection between any two brain regions. We used Annealing-Melting Integration to aid us in approximating the logarithm of the Bayes factor in favour of a model.

For the model where there is a connection, model selection was just a slight extension on the work we had already implemented to infer the values of the knots. However it was not as easy to calculate the Annealing-Melting Integration approximation for the marginal likelihood in the model where there is no connection. Therefore we proposed two potential methods to first infer the values of the knots and the local parameters in this model. Eventually we chose one of the methods for approximating the logarithm of the marginal likelihood when there is no connection which is Basic No Connection MCMC.

We looked at two simulated examples to see how well the approximations work. The first example was a 3D dataset with a connection between two regions. The approximation to the logarithm of the Bayes factor was strongly in

favour of the model where there is a connection. Afterwards we worked with a second example where there was no connection between the two regions. The approximation of the logarithm of the Bayes factor was strongly in favour of the model with no connection. Thus from our simulated datasets Annealing-Melting Integration offers a good method to implement model selection.

CHAPTER 5

Conclusions

5.1 Synopsis

The main aim of this thesis was to develop efficient methods for studying the Global Tractography model within a Bayesian framework. This broader aim was split into three smaller goals that correspond to our three main chapters. In Chapter 2 we discussed how to provide efficient methods for estimating the parameters of the partial volume model. In Chapter 3 we reviewed existing methods and developed novel model selection techniques to choose between the number of fibre orientations in the partial volume model.

The methodology developed in Chapters 2 and 3 was employed in Chapter 4 in fitting the Global Tractography model. In the Global Tractography model the diffusion within a voxel is modelled using the partial volume model, thus estimating the parameters of the model quicker as in Chapter 2 and selecting the model with the correct number of fibre orientations as in Chapter 3, helps us when working with the Global Tractography model. Finally in Chapter 4 the Global Tractography model parameters are inferred and robust methods for model selection are used to decide whether there is evidence for a connection between two brain regions.

We have also introduced a new type of Tractography that we termed Fully Probabilistic Tractography. This new method unlike existing Tractography methods, considers model uncertainty for the number of fibre orientations within a voxel. Furthermore we have introduced a reparameterisation of the partial volume model that involves the use of directional distributions; the reason for this is that such distributions may be more convenient to use in certain special cases. We will now give an overview of the results of each chapter of this thesis.

5.2 Overview of the results

We first introduced the basics of Diffusion-Weighted MRI in Chapter 1 and then in Chapter 2 we explored ways of quickly and efficiently estimating parameters in the partial volume model. As a basis for this we first introduced fast methods for estimating the parameters of the Diffusion Tensor model. This was implemented using both the Linearised Diffusion Tensor model and Markov Chain Monte Carlo within a Bayesian framework. The estimates obtained from the Linearised Diffusion Tensor model were used to obtain initial estimates for the parameters of the partial volume model by using metrics that were defined in Chapter 1.

We used Bayesian inference to infer the parameters of the partial volume model. Initially Vanilla MCMC was used to estimate the parameters in the partial volume model but it was very slow and thus impractical in the presence of many voxels; for this reason we then considered Block-update MCMC. However the choice of covariance matrix for the proposal distribution determines how efficiently MCMC works. To overcome the problem of mixing we introduced Adaptive MCMC as a solution for obtaining a good covariance matrix for propos-

ing candidate values of the parameters within MCMC.

MCMC can potentially be very time consuming in the presence of many voxels, thus we introduced the Laplace approximation as a good approximation of the posterior distribution. We showed that the Laplace approximation density was comparable to the density obtained from MCMC results. Alternatively the Laplace approximation was found to be a good proposal distribution for the independence sampler within the MCMC.

When we inferred the parameters of the partial volume models we showed that there are problems when the true value of θ, which represents the fibre orientation is close to 0. Thus we investigated a new reparameterisation that utilises directional distributions. We investigated both the Bingham and the Angular Central Gaussian distributions as proposal distributions within MCMC. The reparameterisation was shown to be better than using the original parameters in cases where the true value of $\theta \approx 0$. The Angular Central Gaussian distribution was shown to be better than the Bingham distribution as a proposal distribution by comparing the Autocorrelation function of the samples drawn using each algorithm.

We used our methods on real datasets to determine if the proposed methods work as well as they do on simulated datasets. To determine how well the methods works we compared estimates of the parameters in the partial volume model with the corresponding results obtained by FSL software. Our algorithms appeared to produce samples with better mixing.

Finally we conducted a simulation study to compare the different methods for inferring the parameters of the partial volume model. The methods that were

compared were Vanilla MCMC, Block-update MCMC, independence sampler MCMC, Adaptive MCMC and the Laplace approximation. Of all the methods only Vanilla MCMC produced bad results due to the difficulty in the mixing. This simulation study suggested that from all the MCMC methods the independence sampler MCMC produced the least correlated samples whilst also being very quick. We then investigated simulated datasets with two fibre orientations in a voxel. The methods worked equally well in voxels with two fibre orientations as they did in voxels with only one fibre orientation.

In Chapter 3 our goal was to find a method for determining how many fibre orientations to model in the partial volume model. The existing method in the literature, Automatic Relevance Determination was discussed and its problems when implemented were demonstrated. We thus explored Bayesian model choice and Bayes factors. Due to the difficulties in calculating the marginal likelihoods for the Bayes factor we must estimate the value of the marginal likelihoods. This is not a straightforward task and a commonly used method, namely Reversible Jump MCMC, is too costly; it was therefore decided to attempt methods based on Thermodynamic Integration.

The two main methods based on Thermodynamic Integration are Annealing-Melting Integration and Model-Switch Integration. Furthermore we introduced an alternative version of Annealing-Melting Integration which involved the independence sampler, which we called Importance Power Posterior. We further looked at a suggested improvement to the approximations using the corrected Trapezium rule. We demonstrated that this suggestion improves the approximations in toy examples. We then applied all of this to simulated datasets from the partial volume model, and we determined that the best method was Model-Switch integration.

Our model selection techniques were then used within Fully Probabilistic Tractography, which is Probabilistic Tractography with added model uncertainty. This was shown to be effective in voxels with crossing fibres. All methods that were introduced in this chapter were then applied to real datasets, where results appeared to be similarly good.

In Chapter 4 the aim was to infer the parameters of the Global Tractography model which then allowed us to determine where there is evidence for connections between different brain regions of interest. Initially we just focused on obtaining estimates for the knots of the splines that represent the connections between regions. After investigating different possible methods for inferring the values of the knots we found that the proposed Partially Deterministic Scan MCMC was the best method. Initially the Partially Deterministic Scan MCMC algorithm was applied to 2D simulated datasets and this produced favourable results.

In 3D datasets the Partially Deterministic Scan MCMC mixing was not as good as we would expect, so Adaptive MCMC was used within the Partially Deterministic Scan MCMC algorithm which vastly improved the mixing within MCMC. To aid in the performance of the MCMC algorithm, initialisation of the knots was also investigated. We used a method based on Deterministic Tractography to initialise the knots; this method appeared to be better than existing ones.

The final aim of Chapter 4 was to infer the connection matrix $\boldsymbol{C}$, which is the parameter from the Global Tractography model that determines whether there is a connection between any two brain regions. We again used model selection methods based on Thermodynamic Integration to approximate the Bayes factor in favour of the model with a connection. We used Annealing-Melting Integra-

tion as we showed that Model-Switch integration is not suitable in this case. We investigated simulated datasets and found that the approximation of the Bayes factor in these simulated datasets supported the correct model.

5.3 Future work

There are still many possible extensions to the work in this thesis. The first extension would be to investigate datasets that are simulated from the partial volume model with three or more fibre orientations. In the literature there are many conflicting views on whether voxels with more than two white matter fibre orientations are actually regularly encountered (Jeurissen *et al.*, 2013). First one could determine if our methods for both initialising and estimating the parameters of the partial volume models work as well when there are more fibre orientations. It would probably be relatively simple to initialise the parameter values in the partial volume model with three fibre orientations. Similarly to the partial volume model with two fibre orientation we can obtain three eigenvalues in the Diffusion Tensor estimate and then use these eigenvalues to split the Fractional Anisotropy into three. However it will be more challenging to determine a way of initialising the parameters in the partial volume model with four or more fibre orientations. We could also investigate how having models with more fibre orientations affects the methods for model selection in Chapter 3 and Fully Probabilistic Tractography.

In this thesis we have only introduced Fully Probabilistic Tractography and investigated it briefly. This Tractography method could be used to investigate many more connections in the brain.

The methods for inferring the global parameters and connections between re-

gions in the Global Tractography model have thus far only been used on simulated datasets with one fibre orientation in each voxel. A large investigation that explores how these methods work on real data with more fibre orientations needs to be implemented to confirm that these methods work on real data. It would be interesting to see how the connections found by Global Tractography compare with functional connections found from fMRI. We would then like to use the results obtained in Global Tractography to help to determine which possible functional connections to look for in fMRI.

A final step is to use the model uncertainty for the number of fibre orientations within a voxel that we used throughout Chapter 3 within the Global Tractography model when determining the existence of a connection as this could potentially affect the results greatly as it does in Fully Probabilistic Tractography.

CHAPTER 6

Appendices

Appendix A - Catmull Rom splines

Suppose that there are four knots, labelled $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{P}_3$ and $\boldsymbol{P}_4$. These knots are such that the order of the index determines in which order they are connected, such that there is a curve from $\boldsymbol{P}_1$ to $\boldsymbol{P}_2$, a curve from $\boldsymbol{P}_2$ to $\boldsymbol{P}_3$ and a curve from $\boldsymbol{P}_3$ to $\boldsymbol{P}_4$. Then a curve, $\boldsymbol{P}$(t), can be drawn from $\boldsymbol{P}_2$ to $\boldsymbol{P}_3$, such that $\boldsymbol{P}(0)=\boldsymbol{P}_2$ and $\boldsymbol{P}(1)=\boldsymbol{P}_3$. This curve can be calculated for a parameter α, which represents the spline's tension, at any t $\in$(0,1) by:

$$\boldsymbol{P}(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} -\alpha & 2-\alpha & \alpha-2 & \alpha \\ 2\alpha & \alpha-3 & 3-2\alpha & -\alpha \\ -\alpha & 0 & \alpha & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_1 \\ \boldsymbol{P}_2 \\ \boldsymbol{P}_3 \\ \boldsymbol{P}_4 \end{bmatrix}.$$

Thus if two end knots are defined a spline that passes through all of the other knots may be constructed. When these splines are used to connect the knots in the global framework, it is required for the curve to go through all the knots including the two end knots. A way to resolve this problem is to create two auxiliary control points, which are placed at each end of the curve. Once these control points are defined, the spline will then pass through the two end knots.

A common method used to create the control points is reflection. $\boldsymbol{P}_0$, the first control point is obtained by reflecting $\boldsymbol{P}_2 - \boldsymbol{P}_1$ about $\boldsymbol{P}_1$, similarly $\boldsymbol{P}_5$, the second control point is obtained by reflecting, $\boldsymbol{P}_4 - \boldsymbol{P}_3$ about $\boldsymbol{P}_4$. Usually the default value of α is 0.5.

If we choose the default value of α, then

$$\boldsymbol{P}(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} -0.5 & 1.5 & -1.5 & 0.5 \\ 1 & -2.5 & 2 & -0.5 \\ -0.5 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_1 \\ \boldsymbol{P}_2 \\ \boldsymbol{P}_3 \\ \boldsymbol{P}_4 \end{bmatrix}$$

$$= \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \begin{bmatrix} -0.5\boldsymbol{P}_1 + 1.5\boldsymbol{P}_2 - 1.5\boldsymbol{P}_3 + 0.5\boldsymbol{P}_4 \\ \boldsymbol{P}_1 - 2.5\boldsymbol{P}_2 + 2\boldsymbol{P}_3 - 0.5\boldsymbol{P}_4 \\ -0.5\boldsymbol{P}_1 + 0.5\boldsymbol{P}_3 \\ \boldsymbol{P}_2 \end{bmatrix}$$

$$\begin{aligned} =&(-0.5\boldsymbol{P}_1 + 1.5\boldsymbol{P}_2 - 1.5\boldsymbol{P}_3 + 0.5\boldsymbol{P}_4)t^3 \\ &+ (\boldsymbol{P}_1 - 2.5\boldsymbol{P}_2 + 2\boldsymbol{P}_3 - 0.5\boldsymbol{P}_4)t^2 \\ &+ (-0.5\boldsymbol{P}_1 + 0.5\boldsymbol{P}_3)t + \boldsymbol{P}_2. \end{aligned}$$

Appendix B - The calculation of the distribution in the ARD toy example

We know that

$$\pi(y_i|\mu,\sigma) \sim N(\mu,\sigma)$$

so that

$$\pi(y_i|\mu,\sigma) = \frac{1}{\sqrt{\sigma 2\pi}}\exp\left(-\frac{(y_i-\mu)^2}{2\sigma}\right)$$

and

$$\pi(\boldsymbol{Y}|\mu,\sigma) = \left(\frac{1}{\sqrt{\sigma 2\pi}}\right)^N \exp\left(-\frac{1}{2\sigma}\sum_{i=1}^{N}(y_i-\mu)^2\right).$$

Also

$$\pi(\mu|\sigma_\mu) \sim N(0,\sigma_\mu),$$

$$\pi(\sigma) \sim \sigma^{-1}$$

and

$$\pi(\sigma_\mu) \sim \sigma_\mu^{-1}$$

so that

$$\pi(\mu|\sigma_\mu) = \left(\frac{1}{\sqrt{\sigma_\mu 2\pi}}\right)\exp\left(\frac{1}{2\sigma_\mu}\mu^2\right).$$

We calculate

$$\pi(\mu|\boldsymbol{Y}) \propto \int_0^\infty \int_0^\infty \pi(\boldsymbol{Y}|\mu,\sigma)\pi(\mu|\sigma_\mu)\pi(\sigma_\mu)\pi(\sigma)d\sigma d\sigma_\mu$$

by the following

$$\pi(\mu|\boldsymbol{Y}) \propto \int_0^\infty \int_0^\infty \left(\frac{1}{\sqrt{2\pi}}\right)^N (\sigma)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma}\sum_{i=1}^N (y_i-\mu)^2\right)$$

$$\frac{1}{\sqrt{\sigma_\mu 2\pi}} \exp\left(-\frac{1}{2\sigma_\mu}\mu^2\right) \sigma^{-1}\sigma_\mu^{-1} d\sigma d\sigma_\mu$$

$$= \int_0^\infty \left(\frac{1}{\sqrt{2\pi}}\right)^N \frac{1}{\sqrt{\sigma_\mu 2\pi}} \exp(-\frac{1}{2\sigma_\mu}\mu^2)\sigma_\mu^{-1} \int_0^\infty \sigma^{-\frac{N}{2}}$$

$$\exp(-\frac{1}{2\sigma}\sum_{i=1}^N (y_i-\mu)^2)\sigma^{-1} d\sigma d\sigma_\mu.$$

We can derive

$$\int_0^\infty \sigma^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma}\sum_{i=1}^N (y_i-\mu)^2\right)\sigma^{-1} d\sigma = \frac{\Gamma(\frac{N}{2})}{(0.5\sum_{i=1}^N (y_i-\mu)^2)^{\frac{N}{2}}}.$$

Therefore

$$\pi(\mu|\boldsymbol{Y}) \propto \frac{\Gamma(\frac{N}{2})}{(0.5\sum_{i=1}^N (y_i-\mu)^2)^{\frac{N}{2}}} \left(\frac{1}{\sqrt{2\pi}}\right)^{N+1} \int_0^\infty \sigma_\mu^{-\frac{1}{2}} \exp\left(-\frac{1}{\sigma_\mu}0.5\mu^2\right)\sigma_\mu^{-1} d\sigma_\mu$$

and

$$\int_0^\infty \sigma_\mu^{-\frac{1}{2}} \exp\left(-\frac{1}{\sigma_\mu}0.5\mu^2\right)\sigma_\mu^{-1} d\sigma_\mu = \frac{\Gamma(\frac{1}{2})}{\sqrt{0.5}|\mu|}.$$

Then

$$\pi(\mu|\boldsymbol{Y}) \propto \frac{\Gamma(\frac{N}{2})}{(0.5\sum_{i=1}^N (y_i-\mu)^2)\frac{N}{2}} \frac{\Gamma(\frac{1}{2})}{\sqrt{0.5}|\mu|} \left(\frac{1}{\sqrt{2\pi}}\right)^{N+1}$$

$$\propto \frac{(\sum_{i=1}^N (y_i-\mu)^2)^{-\frac{N}{2}}}{|\mu|}.$$

Appendix C - Transformations used in the knots

The knots in the Global Tractography model may be constrained to be within a certain interval of values. For example the first coordinate of the knot $\boldsymbol{P}_1$, $\boldsymbol{P}_1(x)$, may be constrained to take a value in the interval (a_1, b_1). It may be easier to look at the transformed value $\boldsymbol{P}_1'(x)$ which can take a value in the real line. Therefore we wish to find a way of obtaining the value of $\boldsymbol{P}_1(x)$ from $\boldsymbol{P}_1'(x)$, and vice versa. First it is easy to see that the first relation is

$$\boldsymbol{P}_1(x) = a_1 + ((b_1 - a_1)/(1 + \exp(-\boldsymbol{P}_1'(x)))).$$

Then we wish to obtain some function that gets the value of $\boldsymbol{P}_1'(x)$ given a value of $\boldsymbol{P}_1(x)$. This is done as follows

$$\begin{aligned}
\boldsymbol{P}_1(x) - a_1 &= ((b_1 - a_1)/(1 + \exp(-\boldsymbol{P}_1'(x)))) \\
(\boldsymbol{P}_1(x) - a_1)(1 + \exp(-\boldsymbol{P}_1'(x))) &= (b_1 - a_1) \\
(1 + \exp(-\boldsymbol{P}_1'(x))) &= \frac{(b_1 - a_1)}{\boldsymbol{P}_1(x) - a_1} \\
\exp(-\boldsymbol{P}_1'(x)) &= \frac{(b_1 - a_1)}{\boldsymbol{P}_1(x) - a_1} - 1 \\
\exp(-\boldsymbol{P}_1'(x)) &= \frac{(b_1 - a_1) - \boldsymbol{P}_1(x) + a_1}{\boldsymbol{P}_1(x) - a_1} \\
\exp(-\boldsymbol{P}_1'(x) &= \frac{(b_1 - \boldsymbol{P}_1(x))}{\boldsymbol{P}_1(x) - a_1} \\
-\boldsymbol{P}_1'(x) &= \log\left(\frac{(b_1 - \boldsymbol{P}_1(x))}{\boldsymbol{P}_1(x) - a_1}\right) \\
\boldsymbol{P}_1'(x) &= \log\left(\frac{\boldsymbol{P}_1(x) - a_1}{b_1 - \boldsymbol{P}_1(x)}\right).
\end{aligned}$$

Appendix D - The calculation of the posterior distribution of τ

First note that as in Section 2.3

$$\pi(\boldsymbol{\omega}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{\omega})\pi(\theta,\phi)\pi(f)\pi(d)\pi(S_0)\pi(\tau)$$

$$= \left(\frac{\sqrt{\tau}}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(y_i-\mu_i)^2\right)|\sin(\theta)|\frac{\beta_\sigma^{\alpha_\sigma}}{\Gamma(\alpha_\sigma)}(\tau)^{\alpha_\sigma-1}\exp(-\beta_\sigma\tau).$$

Then we can derive the posterior distribution of τ which is

$$\pi(\tau|\boldsymbol{y}) \propto \tau^{\frac{n}{2}+\alpha_\sigma-1}\exp(-\tau(\beta_\sigma+0.5\sum_{i=1}^{n}(y_i-\mu_i)^2))$$

which is proportional to a Gamma distribution with parameters α and β where

$$\alpha = \frac{n}{2}+\alpha_\sigma$$

and

$$\beta = \beta_\sigma + 0.5\sum_{i=1}^{n}(y_i-\mu_i)^2.$$

Bibliography

Akaike, H. 1983. Information Measures and Model Selection. *Bulletin of the International Statistical Institute* 50, pp. 277-290.

Alexander, A.L. 2011. Deterministic White Matter Tractography. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 383-395.

Aralasmak, A., Ulmer, J.L., Kocak, M., Salvan C.V., Hillis, A.E. and Yousem, D.M. 2006. Association, commissural, and projection pathways and their functional deficit reported in literature. *J Comp Assist Tomogr* 30, pp. 695-716.

Atkinson, K. and Han, W. 2004 *Elementary numerical analysis.* 3rd Edition. John Wiley and sons.

Axer, H. 2011. Invasive Methods for Tracing White Matter Architecture. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 31-42.

Axer, H. and Keyserlingk, D.G. 2000. Mapping of fiber orientation in human internal capsule by means of polarized light and confocal scanning laser microscopy. *J Neurosci Methods* 94, pp. 165-175.

Basser, P.J., Mattiello, J. and LeBihan, D. 1994. MR diffusion tensor spectoscopy and imaging. *Biophys J* 66(1), pp. 259-267.

Basser, P.J. and Özarslan, E. 2011. Anisotropic Diffusion: From the Apparent Diffusion Coefficient to the Apparent Diffusion Tensor. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 331-353.

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

Bates, D.M. and Watts, D.G. 2007 *Nonlinear Regression Analysis and Its Applications* 1st Edition. Wiley-Interscience.

Behrens, T.E.J., Johansen-Berg, H., Jbabdi, S., Rushworth, M.F.S. and Woolrich, M.W. 2007. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* 34, pp. 144-155.

Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M. and Smith, S.M. 2003. Characterization and Propagation of Uncertainty in Diffusion-Weighted MR Imaging. *Magnetic Resonance in Medicine* 50, pp. 1077-1088.

Bingham, C. 1974. An Antipodally Symmetric Distribution on the Sphere. *Ann. Statist.* 2 (6), 1201-1225.

Box, G. E. P. and Jenkins, G. 1976. *Time Series Analysis: Forecasting and Control*, Holden-Day.

Bretthorstt, G.L., Kroenke, C.D. and Neil, J.J. 2004. Characterizing water diffusion in fixed baboon brain. *24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP Conference Proceedings, Volume 735*, Garching. 25-30 July, 2004. pp. 3-15

Bürgel, U., Mädler, B., Honey, C.R., Thron., A., Gilsbach, J. and Coenen, V.A. 2009. Fiber tracking with distinct software tools results in a clear diversity in anatomical fiber tract portrayal. *Cen. Eur. Neurosurg.* 70(1), pp. 27-35.

Bzdok, D., Laird A., Zilles, K., Fox, P.T. and Eickhoff, S. 2012 An investigation of the structural, connectional and functional sub-specialization in the human amygdala. *Human Brain Mapping*

Callaghan, P.T. 2011. Physics of Diffusion. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 19-30.

Catani, M. 2011. The Functional Anatomy of White Matter: From Postmortem Dissections to In Vivo Virtual Tractography. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 5-18.

Cortez-Conradis, D., Favilla, R., Isaac-Olive, K., Martinez-Lopez, M., Rios, C. and Roldan-Valadez E. 2013. Diagnostic performance of regional DTI-derived tensor metrics in glioblastoma multiforme: simultaneour evaluation of p, q, L, Cl, Cp, Cs, RA, RD, AD, mean diffusivity and fractional anisotropy. *Eur Radiol.* 23(4) pp. 1112-21.

Crick, F. and Jones, E. 1993. Backwardness of human neuroanatomy. *Nature* 361(6408), pp. 109-110.

Dirac, P. 1958. *The Principles of Quantum Mechanics*. 4th Edition. Oxford at the Clarendon Press.

Earl, D.J. and Deem, M.W. 2005. Parallel tempering: Theory, applications, and new perspectives *Phys. Chem. Chem. Phys.* 7, pp. 3910-3916.

Eickhoff S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K. and Zilles K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25(4) pp. 1325-35.

Einstein, A. 1905. On the motion required by the molecular kinetic theory of heat of small particles suspended in a stationary liquid. *Annalen der Physik* 17, p. 549.

Farin, G. 1996. *Curves and Surfaces for Computer Aided Geometric Design*. Academic Press, London.

Fick, A. 1855. Über diffusion. *Phil Mag* 10, p. 30.

Filley, C.M. 2001. emphThe Behavioral Neurology of White Matter. New York: Oxford University Press.

Filley, C.M. 2011. Neurobiology of White Matter Disorders. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 19-30.

Findley,D.F. 1991. Counterexamples to Parsimony and BIC. *Annals of the Institue of Statistical Mathematics* 43, pp. 505-514.

Fisher, R.A. 1922. On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London. Ser. A* 222 pp. 309-368.

Freeman, S.H., Hyman, B.T. Sims, K.B., Hedley-Whyte, E.T., Vossough, A. Frosch, M.P. and Schmahmann, J.D. 2009. Adult onset leukodystrophy with neuroaxonal spheroids: clinical, neuroimaging and neuropathologic observations. *Brain Pathol* 19 pp. 39-47.

Freidlin, R.Z., Özarslan, E., Komlosh, M.E., Chang, L., Koay, C.G. and Jones, D.K. 2007. Parsimonious model selection for tissue segmentation and classification applications: A study using simulated and experimental DTI data. *IEEE Trans. on Medical Imaging* 26(11), pp. 1576-1584.

Friel, N. and Pettitt, A.N. 2008. Marginal likelihood estimation via power posteriors *J.R. Statist. Soc. B* Vol. 70, Part 3, pp. 589-607.

Friel, N., Hurn, M. and Wyse, J. 2013. Improving power posterior estimation of statistical evidence *arXiv*:1209.3198v3 [stat.C0]

Friel, N. and Wyse, J. 2012. Estimating the statistical evidence - a review. *Stat. Neerl.* 66, pp. 288-308.

Ganeiber, A.M. 2012. *Estimation and Simulation in Directional and Statistical Shape Models* PhD thesis, School of Mathematics, The University of Leeds, pp. 86-90.

Gelman, A. and Meng, X.L. 1998. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Stat. Sci.* 13(2), pp. 163-185.

Geman, S. and Geman, D. 1984 Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images *IEEE Trans PAMI* 6, pp. 721-741

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. 1996. Introducing Markov Chain Monte Carlo. In: Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. eds. *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC Press, London/Florida, 1st Edition, pp. 1-19.

Green, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4), pp. 711-732.

Haario, H., Saksman, E. and Tamminen, J. 2001. An adaptive Metropolis algorithm. *Bernoulli* 7, 223-242.

Hagmann,P., Reese, T.G., Tseng, W.-Y.I., Meuli, R., Thiran,J.P. and Wedeen V.J. 2004. Diffusion spectrum imaging tractography in complex cerebral white matter: an investigation of the centrum semiovale. *International Society for Magnetic Resonance in Medicine, ISMRM TWELFTH SCIENTIFIC MEETING*, Kyoto, Japan, 15-21 May 2004, vol. 12, p. 623.

Hahn, E.L. 1950. Spin echoes. *Phys Rev* 80, p. 580.

Hastie, D.I. and Green, P.J. 2012. Model Choice using Reversible Jump Markov Chain Monte Carlo. *Statistic Neerlandica* Vol. 66, nr. 3, pp. 309-338.

Hasings, W.K. 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, Vol. 57, No. 1, pp. 97-109.

Hernández, M., Guerrero, G., Cecilia, J., García, J., Inuggi, A. and Sotiropoulos S. 2013. Accelerating Fibre Orientation Estimation from Diffusion Weighted Magnetic Resonance Imaging using GPUs *PLoS ONE* 8(4): e61892. doi:10.1371/journal.pone.0061892

Hinne, M., Heskes, T. and Van Gerven, M.A.J. 2012. Bayesian Inference of Whole-Brain Networks *arXiv*:1202.1696v1 [q-bio.NC]

Hobbie, R.K. 1997. *Intermediate Physics for Medicine and Biology.* Springer-Verlag, New York.

Hoff, P.D. 2007. Simulation of the Matrix Bingham-Von Mises-Fisher Distribution, With Application to Multivariate and Relational Data. *Journal of Computational Statistics* Vol. 18, No. 2, pp. 438-456.

Hosey, T., Williams, G. and Ansorge, R. 2005. Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magn. Reson. Med.* 54, pp. 1480-1489.

Jbabdi, S., Woolrich, M.W., Andersson, J.L.R. and Behrens, T.E.J. 2007. A Bayesian framework for global tractography. *NeuroImage* 37, pp. 116-129.

Jbabdi, S. and Johansen-Berg, H. 2011. Tractography: Where Do We Go from Here? *Brain Connectivity* Vol. 1, No. 3, pp. 169-183.

Jenkinson, M. and Smith, S.M. 2001. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* 5(2):143-156.

Jenkinson, M., Bannister, P.R., Brady, J.M. and Smith S.M. 2002. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17(2): 825-841.

Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K. and Sijbers J. 2013. Investigating the Prevalence of Complex Fiber Configurations in White Matter Tissue with Diffusion Magnetic Resonance Imaging. *Human Brain Mapping* 34: 2747-2766.

Kang, N., Zhang, J., Carlson, E.S. and Gembris D. 2005. White matter fiber tractography via anisotropic diffusion simulation in the human brain *IEEE Transactions on Medical Imaging* 24(9):1127-1137.

Kass, R.E. and Raftery, A.E. 1995. Bayes Factors *Journal of the American Statistical Association* Vol. 90, No. 430, pp. 773-795.

Katz, R.W. 1981. On Some Criteria for Estimating the Order of a Markov Chain. *Technometrics* 23, pp. 243-249.

Kent, J.T. and Ganeiber, A.M. 2012. Simulation of random rotation matrices *46th scientific meeting of the Italian Statistical Society*, Rome, 20-22 June, 2012.

Kinoshita, M., Yamada, K., Hashimoto, N., Kato, A., Izumoto, S., Baba, T., Maruno, M., Nishimura, T. and Yoshimine, T. 2005. Fiber-tracking does not accurately estimate size of fiber bundle in pathological condition: initial neurosurgical experience using neuronavigation and subcortical white matter stimulation. *NeuroImage* 25(2), pp. 424-429.

Klingler, J. 1935. Erleichterung der makroskopischen Präparation des Gehirns durch den Gerfrierprozess. *Schwiez Arch Neurol Psychiatrie* 36 pp. 247-256.

Kume, A. and Walker, S.G. 2006. Sampling from compositional and directional distributions. *Statistics and Computing* Vol. 16, No. 3, pp. 261-265.

Lagana, M., Rovaris, M., Ceccarelli, A., Venturelli, C., Marini, S. and Baselli, G. 2010. DTI Parameter Optimisation for Acquistion at 1.5T: SNR Analysis and Clinical Application. *Computational Intelligence and Neuroscience*

Lanciego, J.L. and Wouterlood, F.G. 2000. Neuroanatomical tract-tracing methods beyond 2000: what's now and next. *J. Neurosci Methods* 103, pp. 1-2.

Lang, B. 2001. Bayesian P-Splines. *Sonderforschungsbereich* 386, Paper 236.

Lartillot, N. and Philippe, H. 2006. Computing Bayes Factors Using Thermodynamic Integration. *Syst. Biol*, 55(2), pp. 195-207.

Lawes, I.N.C. and Clark, C.A. 2011. Anatomical Validation of DTI and Tractography. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 19-30.

MacKay, D.J.C. 1995. Probable networks and plausible predictions -A review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems* 6, pp. 469-505.

Mackay, D.J.C. 1998. Introduction to Gaussian Processes. In: Bishop, C.M. ed. *Neural Networks and Machine Learning*. Springer-Verlag, Berlin, pp. 133-165

Mangin, J.F., Fillard, P., Cointepas, Y., Le Bihan, D., Frouin, V. and Poupon C. 2013. Toward global tractography. *Neuroimage* 80, pp. 290-296.

Mardia, K.V. 1975. Statistics of Directional Data. *Journal of the Royal Statistical Society. Series B*, Vol. 37, No. 3, pp. 349-393.

Mardia, K.V. and Jupp, P.E. 2000. *Directional Statistics*, Wiley, England, 2nd edition, pp. 181-182.

Marsaglia, G. 1972. Choosing a Point from the Surface of a Sphere. *Ann. Math. Stat.* 43, pp. 645-646.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. 1953. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics.* 21 (6), pp. 1087-1092.

Meynert, T. 1885. *A Clinical Treatise on Diseases of the Fore-brain Based Upon a Study of Its Structure, Functions, and Nutrition.* Trans. Bernard Sachs. New York: G.P. Putnam's Sons.

Newton, M.A. and Raftery, A.E. 1994. Approximating Bayesian inference with the weighted likelihood bootstrap. *J.R. Stat. Soc. B* 56, pp. 3-48.

Oates, C.J., Papamarkou, T. and Girolami, M. 2014. The controlled thermodynamic integral for Bayesian model comparison. *arXiv*:1404.5053v1 [stat.ME]

Ogata, Y. 1989. A Monte Carlo method for high dimensional integration. *Numer. Math* 55, pp. 137-157.

Pajevic, S. 2011. Statistical Issues in Diffusion Tensor MRI. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 331-353.

Parker, G.J.M. 2011. Probabilistic Fiber Tracking. In: Jones, D.K. ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 396-408.

Parker, G.J.M. and Alexander, D.C. 2005. Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Philos. Trans. T. Soc. London, Ser. B Biol. Sci.* 360, pp. 893-902.

Passingham, R.E., Stephan, K.E. and Kotter, R. 2001. The anatomical basis of functional localization in the cortex. *Nat. Rev., Neuroscience.* 3, pp. 606-616.

Pierpaoli, P. and Jezzard, P. 1996. Diffusion tensor imaging of the human brain. *Radiology* 201:637-648.

Raftery, A.E., Newton, M.A., Satagopan, J.M. and Krivitsky, P.N. 2007. Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity. In: Bernando, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M. and West M. eds. *Bayesian Statistics 8*. 1st Edition. Oxford University Press, Inc., New York, pp. 1-45.

Ripley,B.D. 1987. *Stochastic simulation* Wiley Series in Probability in Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York.

Roberts, G.O. 1996. Markov chain concepts related to sampling algorithms. In: Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. eds. *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC Press, London/Florida, 1st Edition, pp. 45-58.

Rosenthal, J.S. 2010. Optimal Proposal Distributions and Adaptive MCMC. In: Brooks, S., Gelman, A., Jones, G.L. and Meng, X.L. eds. *Handbook of Markov Chain Monte Carlo*. 1st edition. Taylor & Francis, US, pp. 93-112.

Scherrer, B. and Warfield S.K. 2010 Why multiple b-values are requiored for multi-tensor models: evaluation with a constrained log-Euclidean model in *ISBI*, Rotterdamm Netherlands, pp. 1389-1392, IEEE Press.

Schmitz, C. and Hof, P.R. 2005. Design-based stereology in neuroscience. *Neuroscience* 230, pp. 813-831.

Schultz, T., Vilanova, A., Brecheisen, R. and Kindlmann, G. 2013. Fuzzy Fibers: Uncertainty in dMRI Tractography. *arXiv*:1307.3271v1 [cs.CV]

Schwarz, G. 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6, pp. 461-464.

Shibata, R. 1976. Selection of the Order of an Autoregressive Model by Akaike's Information Criterion. *Biometrika* 63, pp. 117-126.

Sotiropoulos, S.N. 2010. *Processing of Diffusion MR Images of the Brain: From Crossing Fibres to Distributed Tractography*, PhD thesis, Division of Clinical Neurology, University of Nottingham.

Stejskal, E.0. and Tanner J.E. 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J Chem Phys* 42, p. 288.

Tabelow, K., Voss, H.U. and Polzehl, J. 2012. Modeling the orientation distribution function by mixtures of angular central Gaussian distributions. *Journal of Neuroscience Methods* 203 pp. 200-211.

Tanner, J.E. 1977. Self-diffusion in cells and tissues. In *Annual Report*. Crane, Ind Naval Weapons Support Center.

Tierney, L. and Kadane, J.B. 1986. Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, Vol. 81, No. 393, pp. 82-86.

Tyler, D. 1987. Statistical analysis for the Angular Central Gaussian distribution on the sphere. *Biometrika*, Vol. 74, pp. 579-589.

van der Knapp, M.S., Pronk, J.C. and Scheper G.C. 2006. Vanishing white matter disease. *Lancet Neurol* 5, pp. 413-423.

Vyshemirsky, V. and Girolami, M.A. 2008. Bayesian ranking of biochemical system models. *Bioinformatics* 24(6): pp. 833-839.

Weigert, C. 1897. Die Markscheidenfärbung. *Ergeb Anat Entwickl Gesch* 3, pp. 1-23.

Wernicke, C. 1874. *Der Apshasische Symptomencomplex, Eine psychologische Studie auf anatomischer Basis.* Trans. G. Eggert. Breslau: Cohn & Weigert.

Wood, J.H. 1980 Physiology, Pharmacology, and Dynamics of Cerebrospinal Fluid. In:Wood, J.H. ed. *Neurobiology of Cerebrospinal Fluid 1* Springer, ISBN: 978-1-4684-1041-9 pp. 1-16

Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M. and Smith, S.M. 2009 Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45 pp. 173-186.

Wozniak, J.R. and Lim, K.O. 2006. Advances in white matter imaging: a review of in vivo magnetic resonance imaging methodologies and their applicability to the study of development and ageing. *Neurosci Biobehav Rev* 30, pp. 762-774.

Wu, J.S., Mao, Y., Zhou, L.F., Tang, W.J., Hu, J., Song, Y.Y., Hong, X.N. and Du, G.H. 2007. Clinical evaluation and follow-up outcome of diffusion tensor imaging-based functional neuronavigation: A prospective, controlled study in patients with gliomas involving pyramidal tracts. *Neurosurgery* 61(5), pp. 935-949.

Zhang, J., Huang, H., Aggarwal, M. and Mori, S. 2011. Diffusion Tensor Microimaging and Its Applications. In: Jones, D.K.ed. *Diffusion MRI: Theory, Methods and Applications*. 1st Edition. Oxford University Press, Inc., New York, pp. 383-395.

Zhou, D. 2010. *Statistical Analysis of Diffusion Tensor Imaging*, PhD thesis, School of Mathematical Sciences, University of Nottingham.

Zhou, D., Dryden, I.L., Koloydenko, A. and Bai, L. 2008. A Bayesian method with reparameterisation for diffusion tensor imaging. *Proceedings of SPIE Medical Imaging 2008: Image Processing*, page 69142J.