Characterization and Propegation of Uncertainty in Diffusion Weighted MR images.

FMRIB Technical Report TR03TB1
(A related paper is published in Magnetic Resonance in Medicine)

T.E.J. Behrens, M.W. Woolrich *, M. Jenkinson, H. Johansen-Berg, R.G. Nunes, S. Clare, P.M Matthews, J.M. Brady and S.M. Smith

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB),
Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital,
Headley Way, Headington, Oxford, UK
Corresponding author is Tim Behrens: behrens@fmrib.ox.ac.uk

1 Abstract

A fully probabilistic framework is presented for estimating local probability density functions on parameters of interest in a model of diffusion. This technique is applied to the estimation of parameters in the diffusion tensor model, and also to a simple partial volume model of diffusion. In both cases the parameters of interest include parameters defining local fiber direction. A technique is then presented for using these density functions to estimate global connectivity (i.e the probability of the existence of a connection through the data field, between any two distant points), allowing for the quantification of belief in tractography results. This technique is then applied to the estimation of the cortical connectivity of the human thalamus. The resulting connectivity distributions correspond well with predictions from invasive tracer methods in non-human primate.

2 Introduction

Uncertainty and its representation have an important role to play in any situation where the goal is to infer useful information from noisy data. In Diffusion Weighted MRI (DWMRI) scientists attempt to infer information about, for example, diffusion anisotropy or underlying fiber tract direction, by fitting models of the diffusion and measurement processes to DWMRI data (e.g. [1, 2]). In this scheme there is uncertainty caused both by the noise and artifacts present in any MR scan, but also by the incomplete modelling of the diffusion signal. That is, the true diffusion signal is more complicated than we choose to model. This additional complexity in the diffusion signal appears as residuals when we fit a simple model to the data, causing additional uncertainty in the model parameters. All of the uncertainty in these parameters may be represented in the form of probability density functions (*pdf*s). This paper is essentially divided into two parts, dealing separately with uncertainty at the local and global levels. In the first part, we describe a technique for estimating the pdfs on all parameters in any *local* model of diffusion. We will show results derived from two simple models of the diffusion process within a voxel: The *diffusion tensor* model which assumes a local 3D Gaussian diffusion profile, and a simple partial volume model of local diffusion, which assumes that a fraction of diffusion is along a single dominant direction, and that the remainder is isotropic. We will then make suggestions for the extension to more complete models of the diffusion process which are able to account for one, or more, distribution of fiber directions within the voxel. In all of these models, the use of Bayesian techniques allows for the application of prior constraints on parameters in the model where such constraints are sensible. For example, in the fitting of the diffusion tensor model, the eigenvalues of the diffusion tensor are constrained to be positive.

The distributions on parameters in a diffusion model are of great significance when making inference on the basis of these parameters. Inference may be at a group level; for example there have been studies showing reduced anisotropy in groups of Multiple Sclerosis patients, in comparison with groups of normal subjects (e.g. [3]). However, inference may also be within a single subject. There have been many recent papers (e.g. [4, 5, 6]) describing techniques for using parameters from a diffusion tensor fit to follow major white matter pathways in the brain. However, none of these techniques attempt to quantify the uncertainty in the resulting white matter connections. The output of these algorithms is a set of nodes describing the *maximum likelihood* pathway through the DTI data, with no measure of confidence on the location of this pathway. The lack of this information makes interpretation of the output pathways difficult, and also makes it hard to devise strategies for tracing reliably in uncertain areas. For both of these reasons, streamlining algorithms to date have chosen not to trace pathways through areas of low diffusion anisotropy

(e.g. [5, 7]). Diffusion anisotropy tends to be low in areas of high uncertainty in fiber direction (although the converse is not necessarily true [8]), and therefore, by tracing fibers only when anisotropy is high, streamlining algorithms have tended to generate pathways which (if they had been calculated) would have had narrow confidence bounds on them. This knowledge means that reconstructed pathways are often interpretable as major fiber tracts in the brain [9], but places limits on areas where it is possible to create them. In the second part of this manuscript, we give the mathematical formulation for deriving *spatial PDF* on connectivity between point A and every other point in the data field given the *local pdf*s. This *PDF* is an explicit representation of the confidence regions for pathways in the data. We go on to present a sampling technique to generate this *PDF* in a computationally efficient manner, and describe and discuss technical details, such as data interpolation, required in any fiber tracing algorithm. We present resulting connectivity *PDF*s from seed voxels in the thalamus, a deep gray matter structure with relatively low diffusion anisotropy. We show that connectivity distributions estimated from diffusion imaging data in human correspond well with predictions from sacrificial tracer studies in primate. Further results from this study appear with detailed discussion and interpretation in [8].

An important point to note is that, throughout this paper, the estimated probability distributions are *pdf*s on parameters in a model. This is to be contrasted with the Gaussian distribution described by the diffusion tensor fit [10], and with more recent work (e.g. [11]) which have attempted to recreate the diffusion spectrum as a probability distribution on the displacement, $r_{final} - r_0$ of a particle with initial location r_0 in the voxel after a diffusion time t_d. There are crucial differences here, both conceptually and practically.

2.1 Densities, Bayes and MCMC

When fitting a parametrized model to data, there are two general approaches which may be taken. The first is to look for the set of parameters (ω) which best fit the data. This is called a point estimate of the parameters. A special case of this is Maximum Likelihood estimation where we look for the set of parameters which maximize the probability of seeing this realization of the data given the model and its parameters.

$$\omega_{ML} = arg \max_{\Omega} \mathcal{P}(Y|\omega, M), \tag{1}$$

where Y is the data and M is the model.

The second approach is to associate a *probability density function* with the parameters. In the Bayesian framework, this distribution is called the posterior distribution on the parameters given the data

$$\mathcal{P}(\omega|Y, M) = \frac{\mathcal{P}(Y|\omega, M)\mathcal{P}(\omega)}{\mathcal{P}(Y|M)}. \tag{2}$$

This posterior density allows us to ask the question of any hypervolume $\mathcal{V}$ in parameter space Ω, "What is our belief given the measured data that the true value of ω is in $\mathcal{V}$?". In the one dimensional case, this question becomes, for any (ω_0, ω_1) "What is our belief that ω lies between ω_0 and ω_1?". These questions, and their answers, represent the uncertainty we have in the values of the parameters ω.

Unfortunately, calculating this pdf is seldom straightforward. The denominator in equation 2 is

$$\mathcal{P}(Y|M) = \int_{\Omega} \mathcal{P}(Y|\omega, M)\mathcal{P}(\omega)d\omega, \tag{3}$$

an integral which is often not tractable analytically. To make matters worse, this *joint* posterior pdf on all parameters is often not the distribution which we are most interested in. We are often interested in the posterior pdf on a single parameter or an interesting subset of parameters. Obtaining these *marginal* distributions again involves performing large integrals,

$$\mathcal{P}(\omega_I|Y,M) = \int_{\Omega_{-I}} \mathcal{P}(\omega|Y,M)d\omega_{-I}, \tag{4}$$

where ω_I are the parameters of interest and ω_{-I} are all other parameters. Again these integrals are seldom tractable analytically.

One solution to this problem is to draw samples in parameter space from the joint posterior distribution, implicitly performing the integrals numerically. For example, we may repetitively choose random sets of parameter values and choose to accept or reject these *samples* according to a criterion based on the value of the numerator in equation 2. It can be shown (e.g [12]) that a correct choice of this criterion will result in the *accepted* samples being distributed according to the joint posterior pdf (equation 2). Schemes such as this are *rejection sampling* and *importance sampling* which generate independent samples from the posterior. Any marginal distributions may then be generated by examining the samples from only the parameters of interest. However, these kinds of sampling schemes tend to be painfully slow, particularly in high dimensional parameter spaces, as samples are proposed at random, and thus each has a very small chance of being accepted.

Markov Chain MonteCarlo (MCMC) (e.g. [12, 13]) is a sampling technique which addresses this problem by proposing samples preferentially in areas of high probability. Samples drawn from the posterior are no longer independent of one another, but the high probability of accepting samples, allows for many samples to be drawn and, in many cases, for the posterior pdf to be built in a *relatively* short period of time.

3 Local Parameter Estimation

In this section we present 3 models of the local diffusion process. The first is the familiar diffusion tensor model ([10]), which models the local diffusion as a 3 dimensional Gaussian. Then we choose two different models which attempt to model *underlying fiber structure* in a voxel and, from this, predict the diffusion weighted signal. The first of these is a simple partial volume model allowing for a single fibre direction mixed with an isotropically diffusing compartment in a voxel. The second is a parametrized model of the transfer function between a distribution of fibre orientations in a voxel and the measured diffusion weighted signal. We infer on the first two of these models, using MCMC to estimate the posterior distributions on parameters of interest. We present detailed results from a single white matter voxel showing recovered distributions from both models. We go on to present a validation study, comparing distributions throughout a slice with those recovered from empirical measurements of uncertainty ([14]).

3.1 Local Parameter Estimation: Theory

The Diffusion Tensor Model. The diffusion tensor has often been used to model local diffusion within a voxel (e.g. [10, 15, 16]). The assumption made is that local diffusion may be characterized with a 3 Dimensional Gaussian distribution ([10]), whose covariance matrix is proportional to the diffusion tensor, $\mathbf{D}$. The resulting diffusion weighted signal, μ_i along a gradient direction $\mathbf{r}_i$, with b-value b_i is modeled as:

$$\mu_i = S_0 \exp\left(-b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i\right), \tag{5}$$

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

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} \tag{6}$$

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

$$\mathbf{D} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}}, \tag{7}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \tag{8}$$

and $\mathbf{V}$ rotates $\mathbf{\Lambda}$ to (θ, ϕ, ψ), such that the tensor is first rotated so that its principal eigenvector aligns with (θ, ϕ) in spherical polar coordinates, and then rotated by ψ *around* its principal eigenvector[1].

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

$$\begin{aligned} \mathcal{P}(\mathbf{Y}|\omega, M) &= \prod_{i=1}^{n} \mathcal{P}(y_i|\omega, M) \\ \mathcal{P}(y_i|\omega, M) &\sim \mathcal{N}(\mu_i, \sigma) \end{aligned} \tag{9}$$

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

$$\mu_i = S_0 \exp -b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i. \tag{10}$$

Thus, the model at each voxel has 8 free parameters each of which is subject to a prior distribution. Priors are chosen to be non-informative, with the exception of ensuring positivity where sensible[2].

$$\begin{aligned} \mathcal{P}(\theta, \phi, \psi) &\propto \sin(\theta) \\ \mathcal{P}(S_0) &\sim \mathcal{U}(0, \infty) \\ \mathcal{P}(\lambda_1) = \mathcal{P}(\lambda_2) = \mathcal{P}(\lambda_3) &\sim \Gamma(a_\lambda, b_\lambda) \\ \mathcal{P}(\frac{1}{\sigma^2}) &\sim \Gamma(a_\sigma, b_\sigma) \end{aligned} \tag{11}$$

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

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

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

[2]A description of the Γ distribution may be found in the appendix

The predicted diffusion signal is

$$\begin{aligned}\mu_i &= S_0((1-f)\exp(-b_i d) \\ &\quad +f\exp(-b_i d\mathbf{r_i}^T\mathbf{RAR}^T\mathbf{r_i}))\end{aligned} \tag{12}$$

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

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{13}$$

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

Again noise is modeled as *iid* Gaussian:

$$\begin{aligned}\mathcal{P}(\mathbf{Y}|\omega, M) &= \prod_{i=1}^{n} \mathcal{P}(y_i|\omega, M) \\ \mathcal{P}(y_i|\omega, M) &\sim \mathcal{N}(\mu_i, \sigma),\end{aligned} \tag{14}$$

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

$$\begin{aligned}\mathcal{P}(\theta, \phi) &\propto \sin(\theta) \\ \mathcal{P}(S_0) &\sim \mathcal{U}(0, \infty) \\ \mathcal{P}(f) &\sim \mathcal{U}(0, 1) \\ \mathcal{P}(d) &\sim \Gamma(a_d, b_d) \\ \mathcal{P}(\frac{1}{\sigma^2}) &\sim \Gamma(a_\sigma, b_\sigma).\end{aligned} \tag{15}$$

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

Such a model clearly requires some assumptions. We start by assuming that each subvoxel has only one fibre direction through it and that the MR signal is the sum of the signal from arbitrarily small subvoxels, and that the signal from each *subvoxel* behaves as described by Equation 12. (Note that this is a strong assumption to make, but it is *explicit* in the model. Any other model of the local diffusion characteristics of a single fibre orientation may be used as a replacement.)

$$\mu_{total} = \sum_{j \in sub-voxels} \mu_j \tag{16}$$

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

If we now consider, instead of the individual sub-voxels, the set $\Theta\Phi$ of major directions (θ, ϕ) in these subvoxels (note the discretization of$\Theta\Phi$) , then Equation 16 is identically equivalent to (see equation 12):

$$\begin{aligned}\mu_i = \sum_{(\theta,\phi)\in\Theta\Phi} \Bigg(\sum_{j \in V_{\theta\phi}} \frac{S_{0_j}}{N} & \left[(1-f_j)\exp\left(-b_i d_j\right) + \right. \\ & \left. f_j \exp\left(-b_i d_j \mathbf{r}_i^T \mathbf{R}_{\theta\phi}\mathbf{A}\mathbf{R}_{\theta\phi}^T \mathbf{r}_i\right)\right] \Bigg)\end{aligned} \tag{17}$$

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

$$\frac{\mu_i}{S_0} = (1-f)\exp(-b_i d) + \\ f\int_0^{2\pi}\int_0^{\pi} \mathcal{H}(\theta,\phi)\exp\left(-b_i d \mathbf{r}_i^T \mathbf{R}_{\theta\phi}\mathbf{A}\mathbf{R}_{\theta\phi}\mathbf{r}_i\right)\sin(\theta)d\theta d\phi. \tag{18}$$

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

$$\frac{\mu_i(\alpha_i,\beta_i)}{S_0} = (1-f)\exp(-b_i d) + \\ f\int_0^{2\pi}\int_0^{\pi} \mathcal{H}(\theta,\phi)\exp\left[-b_i d\cos^2\gamma_i\right]\sin(\theta)d\theta d\phi. \tag{19}$$

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

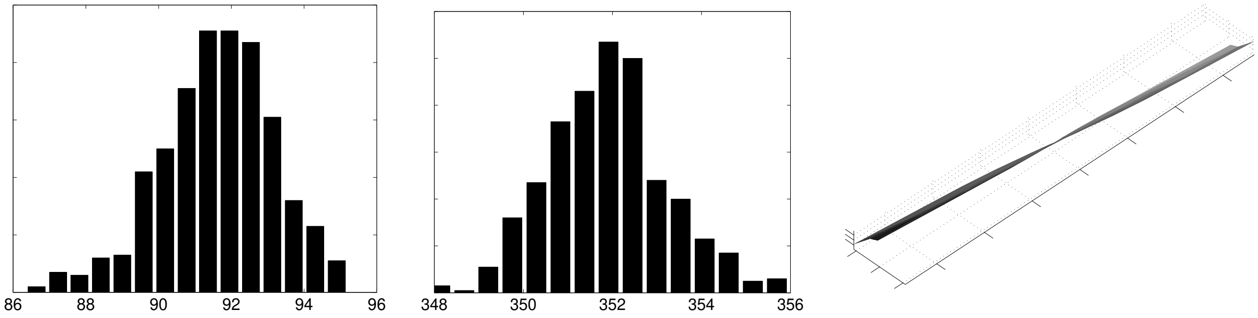
3.2 Local Parameter Estimation: Methods

Data acquisition DT-MRI datasets were acquired on a single healthy volunteer. The images were obtained on a 3.0 T Varian Inova scanner using a diffusion-weighted single-shot EPI sequence. To minimize eddy currents, a doubly-refocused spin-echo sequence was implemented [17]. A birdcage radio-frequency head coil was used for both pulse transmission and signal detection. The diffusion gradients achieved a maximum gradient strength of 22 mTm^{-1}. Each data set consisted of 3 non-diffusion-weighted and 60 diffusion-weighted images acquired with a b-value of 1000 smm^{-2}. The diffusion gradients were uniformly distributed through space using the optimized scheme proposed by Jones [18]. Each set of images contained 42 contiguous slices with a 2.5 mm thickness. A half k-space acquisition was performed with a matrix size set to 62×96 and a field of view of $240 \times 240mm^2$. The images were interpolated to achieve a matrix size of 128×128 and a final resolution of $1.875 \times 1.875 \times 2.5mm^3$. To minimize motion artifacts, peripheral gating was used such that triggering occurred on every cardiac cycle. The echo time was set to 106 ms while the effective repetition time was 14 R-R intervals. The total scan time for each dataset was approximately 15 minutes, depending on heart rate.

Estimation MCMC estimation was performed for the diffusion tensor model and for the simple partial volume model. In both cases parameters were initialized with a least squares diffusion tensor fit. The Markov Chains were then jumped 500 times without sampling as a "burnin" (see [12]), followed by 2000 times, sampling every second jump, to give 1000 samples. A single jump of the parameter set consisted of independent jumps of each parameter. In both models samples were drawn from the precision ($\frac{1}{\sigma^2}$) with a Gibbs sampler, and from all other parameters with Metropolis Hastings samplers. Proposal distributions for Metropolis Hastings parameters were zero mean Gaussians with standard deviations tuned adaptively to give a jump acceptance rate of 0.5. The full conditional distributions for the Gibbs sampling of the precision in both models are given in the appendix. Computation time for diffusion data with 63 acquisitions is approximately 0.3 seconds per voxel on a Pentium IV 2GHz. Voxels are processed independently, so computation is easily parallelized.

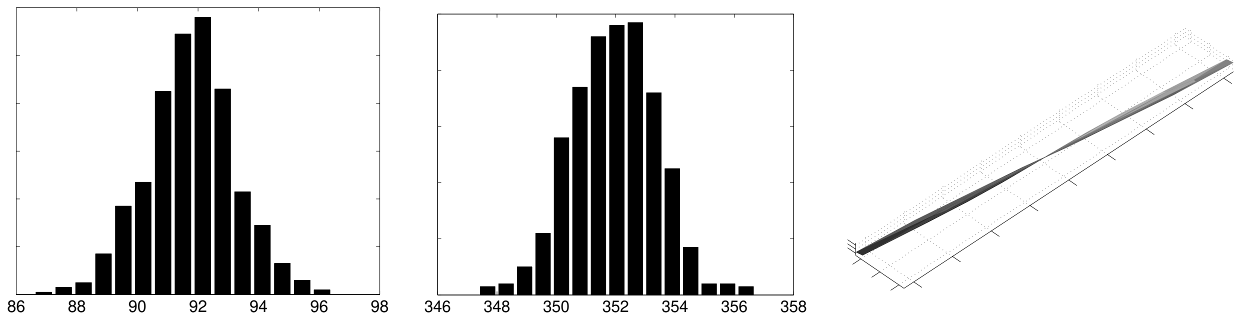
3.3 Local Parameter Estimation: Results

Example distributions from a single voxel. Figure 1 (a) and (b) show samples from the marginal posterior distributions on θ and ϕ from the *diffusion tensor* model. The voxel was chosen from the splenium of the Corpus Callosum. (c) shows (a) and (b), plotted as a joint histogram around the surface of a sphere. This is then the joint marginal posterior distribution of θ and ϕ or the marginal posterior distribution of principal diffusion direction (PDD). Note how narrow this distribution is. This represents a high confidence in our calculated PDD, which is as predicted in an area of dense white matter such as the corpus callosum. Figure 2 (a) and (b) show samples from the marginal posterior distributions on θ and ϕ from the simple *partial volume* model. The same voxel was chosen as in Figure 1. Again (c) shows (a) and (b), plotted as a joint histogram around the surface of a sphere.



(a) (b) (c)

Figure 1: Samples from marginal posterior distributions of the diffusion tensor model in a white matter voxel. (a) shows samples from the marginal posterior distribution on θ. (b) shows samples from the marginal posterior distribution on ϕ. (c) shows (a) and (b) plotted around a sphere, representing the marginal posterior distribution on principal diffusion direction.



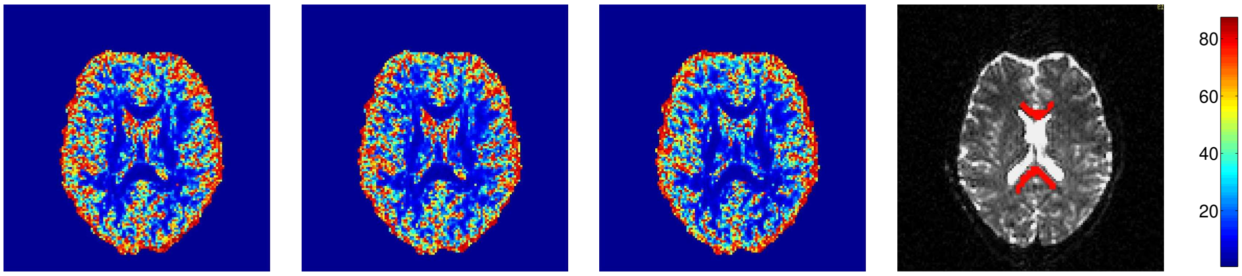
(a) (b) (c)

Figure 2: Samples from marginal posterior distributions of the partial volume model in a white matter voxel. (a) shows samples from the marginal posterior distribution on θ. (b) shows samples from the marginal posterior distribution on ϕ. (c) shows (a) and (b) plotted around a sphere, representing the marginal posterior distribution on principal diffusion direction.

Validation: Comparison with empirical measurements. The posterior *pdf*s on the parameter estimates, in either of the above models, characterize our uncertainty in these parameters. In [14], Jones proposes an empirical method for estimating this uncertainty. Following this method, we acquired 3 repeats of diffusion data with 63 gradient directions and bootstrapped,to create 1000 datasets of different combinations of these repeats. We fit a diffusion tensor at each voxel in each of these new datasets, and calculated the uncertainty between the 1000 principal eigenvectors at each voxel. This uncertainty is measured as the size of the 95% confidence angle from the mean direction.

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

Figure 3 shows these 95% angles for the diffusion tensor model in (a), and the partial volume model in (b). (c) shows the same angles predicted by Jones' method.



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

Figure 3: 95% uncertainty values (in degrees), predicted from the Diffusion Tensor model (a), the partial volume model (b), and Jones' empirical method (c). (d) shows a mask of the corpus callosum used in some calculations

There are various factors to consider when comparing these results. The first is that the empirical method in (c) is not necessarily "ground truth". It has errors associated with it due to subject motion and interpolation related effects, but also, more subtly, due to the dependence within the bootstrapped datasets. This is likely to cause an *underestimate* in the measured uncertainty. The second factor is the difference in the two models. (a) and (c) predict uncertainty levels in the principal eigenvector of a diffusion tensor model. (b) predicts the same thing for the less flexible partial volume model. In areas of complex fibre structure, the partial volume model, which has only one fibre direction available to it, is forced to represent this structure as uncertainty in the single direction (this will turn out to be extremely useful when trying to do tractography, as will be seen in later sections. In contrast, the diffusion tensor model will tend to account for complex fibre structure in a voxel not only with uncertainty in the principal fibre direction, but also with a change in the diffusion profile (i.e. a change in the relative sizes of the three eigenvectors). For this reason we would predict that, in regions of complex fibre structure, the partial volume model would show more uncertainty in principal diffusion direction than the diffusion tensor model. We would expect the two models to predict very similar uncertainties in regions of high fibre co-alignment, such as in the corpus callosum (Figure 3 (d)).

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

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

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

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

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

Predictions of uncertainty by MCMC on the two models are within 10% of each other throughout the brain, and within 5% in the callosal mask, showing, as predicted, very similar uncertainty where fibers are highly co-aligned, and slight differences in uncertainty in other areas. With the diffusion tensor model, uncertainties predicted by

MCMC are within 15% of those predicted by the empirical method when considering the whole brain, and 13% when only considering the corpus callosum. These differences are small and may be due to errors in either or both approaches.

4 Global Connectivity Estimation

4.1 Global Connectivity estimation: Theory

In the previous section we described techniques for estimating, at each voxel, probability distributions on every parameter in the chosen model of diffusion. In this section we use these local *pdf*s from the simple partial volume to infer on a model of global connectivity. The reason we choose this model is explained in detail in the previous section. We wish to maximize the chances that complex fiber structure will be represented by *uncertainty in principal direction*. We now require a model to take us from the local parameters in this model to parameters describing global connectivity. Note that, throughout the remainder of this paper, subscript $\mathbf{x}$ refers to "every voxel in the brain". Hence $(\theta, \phi)_{\mathbf{x}}$ refers to the complete set of principal diffusion directions.

Consider the case where the values of the local parameters are known with no uncertainty. What do they tell us about anatomical connectivity between voxels in the brain? In the case where our local model describes only a single fiber direction passing through the voxel, this *global* model can only take one form:

$$\mathcal{P}(\exists A \to B|(\theta,\phi)_{\mathbf{x}}) = \begin{cases} 1 & \text{If there is a connecting} \\ & \text{path through } (\theta,\phi)_{\mathbf{x}} \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

Where $\mathcal{P}(\exists A \to B|(\theta,\phi)_{\mathbf{x}})$ is the probability of a connection existing between points A and B, given knowledge of local fiber direction.

In order to solve this equation we may simply start a connected path from a *seed point*, A, and follow local fiber direction until a stopping criterion is met. If B lies on this path we may say that a connection exits between A and B. This procedure is at the heart of all "streamlining" algorithms (e.g. [6, 19, 5]), which choose $(\theta,\phi)_{\mathbf{x}}$ to be the principal eigen-direction of the estimated diffusion tensor at each voxel.

However, in the case where there is uncertainty associated with $(\theta,\phi)_{\mathbf{x}}$ we would like to compute the probability of a connection existing given the *data*, $\mathbf{Y}_{\mathbf{x}}$, which is known. That is, we would like to compute $\mathcal{P}(\exists A \to B|\mathbf{Y}_{\mathbf{x}})$. In order to calculate this *pdf* we would have to perform the following integrations:

$$\begin{aligned}\mathcal{P}(\exists A \to B|Y) = \\ \int_0^{2\pi}\int_0^{\pi}\dots\int_0^{2\pi}\int_0^{\pi} \mathcal{P}(\exists A \to B|(\theta,\phi)_x)\mathcal{P}((\theta,\phi)_{x_1}|Y)\dots \\ P((\theta,\phi)_{x_n}|Y)d\theta_{x_1}d\phi_{x_1}\dots d\theta_{x_n}d\phi_{x_n}\end{aligned} \tag{21}$$

That is, for each possible value of fiber direction at every voxel $(\theta,\phi)_{\mathbf{x}}$, we must incorporate the probability of connection given this $(\theta,\phi)_{\mathbf{x}}$, and also the probability of this $(\theta,\phi)_{\mathbf{x}}$ given the acquired MR data. This process is known as *marginalization* (see e.g. [20]).

It can be seen from equation 21, that $\mathcal{P}(\exists A \to B|Y)$ reduces to $\mathcal{P}(\exists A \to B|(\theta,\phi)_{\mathbf{x}})$ when the local pdfs on fiber direction $\mathcal{P}((\theta,\phi)_{\mathbf{x}})$ are delta functions . That is, when there is no uncertainty in the local fiber direction, equation 21 reduces to the streamlining (maximum likelihood) solution. However, when local fiber direction *is* uncertain, $\mathcal{P}(\exists A \to B|Y)$ will be non-zero for some B not on the maximum likelihood streamlines. That is, the global connectivity pattern from A, will spread to incorporate the known uncertainty in local fiber direction.

However, even in the discrete data case, equation 21 represents a v dimensional (where v is the number of voxels in the brain) integral over distributions with no analytical representation (the local pdfs, generated with MCMC), and hence clearly cannot be solved analytically.

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

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

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

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

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

Stopping criteria Algorithms which generate streamlines based on maximum likelihood fiber directions (principal eigenvector from a diffusion tensor fit) have tended to require harsh streamline stopping criteria based on fractional anisotropy and local curvature (angle between successive steps). Fractional anisotropy thresholds have tended to be in the range of 0.2-0.4 (e.g. [7]), and curvature thresholds have been as strict as requiring successive steps to be within 45° (e.g. [5]). These criteria are in place to reduce the sensitivity of the streamlining to noise in the image, partial volume effects, and other related problems. The aim is to reduce the possibility of seeing false positives in the results by only progressing when there is high confidence in fiber direction, and when the direction is anatomically plausible. The downside of these constraints is the limitations that they impose on which fiber tracts may be reconstructed and where in the brain they may occur. For example deep gray matter structures, despite displaying a high degree of order in their principal diffusion directions, tend to have low anisotropy (often below the threshold for streamlining algorithms). Streamlines will also tend to terminate well before cortex as anisotropy reduces, and uncertainty in fiber direction increases.

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

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

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

4.2 Global Connectivity estimation: Methods

Data acquisition Diffusion-weighted data were acquired with an optimized method based on echo planar imaging, implemented on a General Electric 1.5 T Signa Horizon scanner with a standard quadrature head-coil and maximum gradient strength of 22 $mTm^{-1.}$

The diffusion weighting followed an optimized scheme [21] where the diffusion weightings were isotropically distributed along 54 directions. With the diffusion parameters d and D equal to 34 and 40ms respectively, the b-value was 1150 smm^{-2}, the optimum for white matter DTI measurements. 6 diffusion-weighted volumes were acquired with b-value 300 smm^{-2}, and 6 volumes were acquired with no diffusion weighting. Each volume covered the whole brain with 60 slices of 2.3 mm slice thickness, field of view $220 \times 220mm^2$. An imaging matrix of 96 x 96 was used, giving isotropic voxels of $2.3 \times 2.3 \times 2.3mm^3$ and the images were reconstructed on a 128×128 matrix, giving a final resolution of $1.7 \times 1.7 \times 2.3mm^3$. An optimized cardiac gating scheme [21] was used to minimize artifacts arising from cerebrospinal fluid pulsatile flow. The total scan time for the DTI protocol was approximately (depending on heart rate) 20 minutes.

The high resolution T1 weighted scan was obtained with a 3D inversion recovery prepared spoiled gradient echo (IR-SPGR). Parameters for the acquisition were: FOV = 310×155; matrix size = 256×128; in-plane resolution = $1.2 \times 1.2mm^2$; 156 slices of 1.2mm slice thickness; inversion time = 450ms; repetition time = 2s; echo time = 53ms.

Estimation Estimation was carried out exactly as before, except that, for reasons of computational storage, when carrying out estimation on the whole brain, as opposed to a single slice, we drew samples every 20^{th} jump instead of every 2^{nd}.

4.3 Global Connectivity estimation: Results

In this section, we present some results of applying this methodology to the estimation of connectivity distributions from voxels in human thalamus. Knowledge of thalamo-cortical connectivity is sparse in human, but rich in primate. These results, and others, are analyzed, interpreted and compared with the non-human literature in detail in [8]. Here we present them as a first step toward validation of probabilistic tractography as presented in this paper, and evidence suggesting that connectivity studies are feasible with diffusion weighted imaging, even between Gray matter structures.

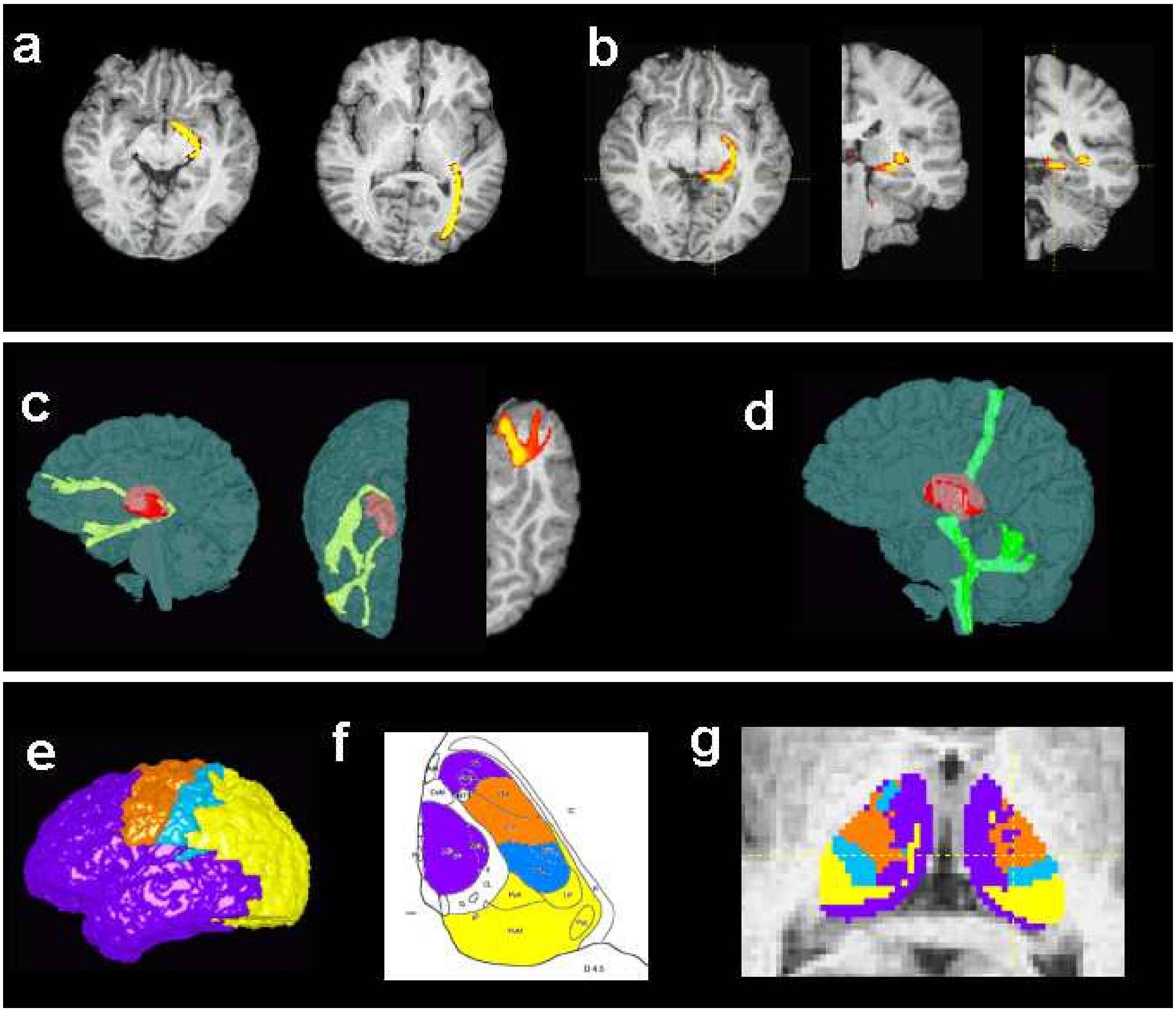


Figure 4: Connectivity distributions and connectivity based segmentation from seed points in the visual system and thalamus.

Results from the thalamus Figure 4 (a) and (b) show results from seeding different parts of the visual system.In (a) the seed point was in the lateral geniculate nucleus (LGN), a thalamic nucleus which processes visual information. The connectivity distribution heads anteriorly into optic tract, and posteriorly into visual cortex, consistent with the known connections of the lgn in non-human primate [22, 23]. However, when the seed is placed in the optic tract (b),two distinct pathways emerge. The two coronal scans in this figure show these two pathways just after they split (near the (coronal) level of lgn) and around 10mm posterior to the split. The righthand pathway follows the route of the pathway in (a). The left hand pathway (see the axial slice in Figure 4 (b)), heads inferior to lgn, round the posterior ventral edge of thalamus to the superior colliculus. These distributions correspond to the two known branches of the primate visual system ([23]); the optic radiations (via LGN) and the superior-collicular brachium.

Figure 4 (c,d) show connectivity distributions seeded in different areas in thalamus.(c) shows a distribution seeded in a medial dorsal area in thalamus. In primate, nuclei in the medial dorsal nuclear cluster of thalamus receive projections from anterior temporal lobe [24, 25, 26] and maintain reciprocal projections with prefrontal cortex [27, 28]. The connectivity distribution in (c) progresses anteriorly into prefontal cortex, and initially posteriorly round the posterior edge of thalamus followed by anteriorly into anterior temporal lobe. (d) shows a distribution seeded in a ventral lateral area in thalamus. In primate thalamus, the ventral lateral nuclear group processes motor information, and maintains strong connections with other motor zones [29, 30], such as primary motor cortex, and cerebellar cortex. The connectivity distribution in (d) progresses superiorly to primary motor cortex and inferiorly to cerebellar cortex and brainstem.

The interpretation issues discussed in the previous section are particularly relevant to the distributions shown in Figure 4 (a,b,c,d). (a) and (b) and (d), show pathways which mainly exist in large white matter pathways, with correspondingly low uncertainty in fiber direction. Hence the distributions seen in these figures are narrow. This should not be interpreted to mean that true connections from the seed voxel are necessarily correspondingly focused, but rather that the uncertainty on the pathway defined by the principal diffusion directions is low. The pathway seen in (c) spreads as it passes through a region of uncertainty while approaching temporal lobe, and also encounters uncertainty before entering prefrontal cortex. Again, this should be interpreted as uncertainty in the connection defined by the principal diffusion directions. To reiterate the point previously: In order to *infer on* diffuse connections from a single seed, the model of diffusion within a voxel must allow for multiple fibers passing through the voxel. However, as can be seen in (b) and (d), presence of local fiber divergence may well be reflected in the local *pdf* at, for example, branching points in the pathways. In these two examples, branches which are known to occur in primate brain, are found by accounting for uncertainty in *principal* diffusion direction.

To test the consistency of the results throughout the thalamus, we seeded every voxel in thalamus (manually outlined on the T1 weighted image), and classified the results by the cortical area with the highest probability of connection to the seed. Four cortical areas were manually outlined on the T1 image to correspond with the principal projection and reception sites of thalamic nuclear clusters in primate brain [23] (Figure 4 (e)):

Cortical Zone	Connected Nuclear Cluster in Primate	Color in Figure 4 (e)
Prefrontal-Temporal	medio-dorsal, ventral anterior	purple
Motor-Premotor	ventral lateral	orange
Sensory	ventral posterior	sky blue
Occipital	lateral geniculate, pulvinar (partly)	yellow

Figure 4 (f) shows the nuclei in human thalamus, as defined by histological staining [31] with, overlaid, a color map showing predictions derived from primate data of the strongest cortical connection sites.

We skull-stripped the diffusion weighted image [32], and performed affine registration between the diffusion weighted and T1 images [33],[34], taking care never to resample the diffusion image. We then ran probabilistic tractography seeded from every voxel in the structural scan, classifying the results as above. The results can be seen in figure 4 (g). The classification of thalamic seed voxels by their connectivity distributions, reveals a segmentation of thalamic nuclear clusters broadly consistent with the histological prediction (underlaid in figure 4 (g)), and most probable connected cortical zones consistent with predictions from primate data (overlaid in color in figure 4 (f)). Furthermore, the results show approximate bilateral symmetry in thalamic seed voxels.

These results are examined and extended in detail in [8] including a finer segmentation of the thalamic nuclei resulting from an increased number of cortical zones, and a detailed look at the information available in the probability values themselves.

5 Discussion

In general, analysis of diffusion weighted data has involved the fitting of a model of local diffusion to the diffusion weighted data at each voxel. This model may assume that local diffusion is Gaussian in profile (the *Diffusion Tensor* model [10]) or may allow a more complex structure for local diffusion (e.g. a spherical harmonic decomposition [35, 36]). However, in all cases, the assumed model is of the *diffusion profile* and not of the underlying fiber structure, and any analysis which has occurred after the fitting of this local model has made the assumption that the parameters in this model are known absolutely.

There are two important, but separate issues here. The first, is that the parameters of real interest to the scientist are ones which relate directly to the underlying fiber structure, and not to the diffusion profile. These underlying parameters may have convincing markers within the fitted diffusion profile (for example anisotropy measures [37, 2] from the diffusion tensor fit have been shown to be a marker for collinearity of fibers within a voxel), but any attempt to recreate the fiber structure from these profiles is essentially an educated guess. There has been no model proposed to predict how a specific structure or distribution of fiber directions within a voxel will reflect itself in the measured diffusion weighted NMR signal. The second issue is that, even when fitting a model of local diffusion, the resulting parameters have uncertainty associated with them. Factors such as noise in the NMR signal (both physical and physiological) and, crucially, the inadequacy of the proposed model, lead to this uncertainty which should be incorporated in any further processing (such as tractography schemes).

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

We then consider uncertainty at a global level. We outline the theory behind moving from the *pdf*s on local PDD to an estimate on the probability distribution on global connectivity. When estimating global connectivity, we first have to choose between the available *local* models of diffusion. We have chosen to use a simple partial volume model. The reason for this choice is that, by choosing a model which allows for only a single fiber direction within a voxel, we maximize the chance that the effect of diverging or splitting fibers will be seen as *uncertainty* in the principal diffusion direction, and not as a change in the diffusion profile, as might be the case if the Diffusion Tensor model were chosen. However the similarity in uncertainty between the two models that we find in the empirical validation suggests that this decision is made largely for conceptual completeness, and that results would have been similar if the Diffusion Tensor model had been chosen.

The next stage is to define a model of global connectivity. The model we choose is identical to that used in streamlining algorithms (e.g. [4, 19, 6, 5]). That is, given absolute knowledge of local fiber directions, connectivity is assumed between two points if, and only if, there exits a connected path between them through the data (see equation 20). The crucial difference between the probabilistic tractography proposed here, and the streamlining algorithms referenced above above can be seen in equation 21. Put simply, the result of this equation incorporates every possible fiber orientation at every voxel and the probability of each of these fiber directions given the acquired MR data. We simply allow for uncertainty in fiber direction when computing streamlines. The practicality of solving this equation is an algorithm similar in nature to others presented, along with this method, at ISMRM 2002 [38, 39, 40], effectively repeatedly sampling local *pdf*s to create streamlines, and regarding these streamlines as samples from a global *pdf*. A crucial difference between these methods and our method, is that we choose to compute the local *pdf*s in a rigorous fashion given the MR data . The methods referenced above all use heuristic experience-based relationships between the shape of the fitted diffusion tensor and the assumed *pdf* on local fiber orientation.

An important result of our procedure is that the recovered "connectivity distributions" are strictly probability distributions on the connected pathway through *dominant* fiber directions. That is, there is no explicit representation of splitting or diverging fibers in either the local or global model. We are strictly inferring on a single pathway leading from the seed point, and therefore in order to find, for example, splitting pathways, the effect of fiber divergence within a voxel must reveal itself as *uncertainty* in the PDD. It can be seen from the local results section that, at least in the cases presented here, this effect can be seen. Figure 4 (b) shows sensitivity to the splitting of fibers from the optic tract, into the superior collicular brachium, and the direct fibers of the optic radiations. Figure (d) also shows sensitivity to branching fibers. Descending fibers from the ventral lateral (motor) nucleus of the thalamus split into two distinct branches, as is to be expected from primate studies. The first heads down to brainstem, and the second into superior cerebellar cortex. However, because fiber divergence within a voxel is treated as uncertainty in principal diffusion direction, this sensitivity to diverging and branching fibers will be dependent on the experimental design; in general, the more information in the MR measurements, the lower the uncertainty in principal diffusion direction. Taking this effect to its logical extreme, if we were to gather an infinite number of MR measurements, there would be *no* uncertainty in principal fiber direction, and the marginal probability distribution on the dominant streamline would be infinitely narrow, i.e. the simple streamlining solution. Ideally we would like to infer, not on connectivity via a single connection, but on an *anatomical* distribution of connectivity. In order to do this we must allow for divergence, branching and crossing of fibers in our *local* model of diffusion. We propose one such model which will allow for inference on an underlying *distribution* of fiber orientations.

Probably the most important result in this paper is in Figure 4 (e,f,g). Here we seed *every* voxel in thalamus, and compute the respective connectivity distributions, recording the probability of connectivity to each of four cortical masks. There are two striking features in this figure. The correspondence of the connectivity-based thalamic segmentation between the left and right thalami (g) provides strong evidence for the robustness of the technique, even when seeding from deep gray matter areas. This is backed up by the marked similarity between the predicted cortical zones from primate data (f) and the connectivity based segmentation (g). This second feature also provides strong, albeit indirect, validation for the use of diffusion based tractography in any guise.

In summary, we have presented a technique for characterizing the uncertainty associated with parameter estimates in diffusion weighted MRI, and for propagating this uncertainty through the diffusion weighted data. This allows us to compute the probability distribution on the location of the dominant fiber pathway so that we may quantify our belief in the tractography results.

5.1 Acknowledgments

The authors thank Claudia Wheeler-Kingshott, Phil Boulby and Gareth Barker from the Institute of Neurology, Queen's Square, London and the UK Multiple Sclerosis Society for providing data for this work. We would also like to thank the following institutions for their generous support: UK Medical Research Council (PMM, SMS), UK Engineering and Physical Science Research Council (TEJB, SMS), Wellcome Trust (HJB), the EPSRC-MRC IRC "From medical images and signals to clinical information" (JMB, MWW), the Portuguese Foundation for Science and Technology (RGN).

A Appendix

A.1 Gamma Distribution

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

$$\mathcal{P}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \tag{22}$$

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

A.2 Full conditional distibution for precision parameters

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

$$\mathcal{P}(\frac{1}{\sigma^2}|\mathbf{Y}, \Omega_-) = \Gamma(a + \frac{n}{2}, b + \frac{1}{2}\sum_{i=1}^{n}(Y_i - \mu_i)^2), \tag{23}$$

where, $\mathbf{Y}$ is the data, Ω_- is the set of all parameters except σ, n is the number of acquisitions, Y_i is the value of the data at the i^{th} acquisition, a and b are the parameters in the Gamma prior on the precision, and μ_i is the value for the i^{th} acquisition predicted by the model. μ_i, for the diffusion tensor model is given by equation 10, and for the simple partial volume model, by equation 12.

References

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

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

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

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

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

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

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

[8] T.E.J. Behrens, H. Johansen-Berg, M.W. Woolrich, S.M. Smith, C.A.M. Wheeler-Kingshott, P.A. Boulby, G.J. Barker, E.L. Sillery, K. Sheehan, O. Ciccarelli, A.J. Thompson, J.M. Brady, and P.M. Matthews. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature Neuroscience. Under Revision, (2003).

[9] Marco. Catani, Robert. J. Howard, Sinisa. Pajevic, and Derek. K. Jones. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage*, **17**, 77–94. Sep 2002. Clinical Trial.

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

[11] Tuch D.S. Wieggell M.R. Reese T.G. Belliveau J.W. Van J. Wedeen. Measuring corico-cortical connectivity with diffusio spectrum imaging. In *Annual Meeting of the ISMRM*, page 502, (2001).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[32] Stephen. M. Smith, Yongyue. Zhang, Mark. Jenkinson, Jacqueline. Chen, P. M. Matthews, Antonio. Federico, and Nicola. De Stefano. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage*, **17**, 479–489. Sep 2002. Evaluation Studies.

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

[34] Mark. Jenkinson, Peter. Bannister, Michael. Brady, and Stephen. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, **17**, 825–841. Oct 2002.

[35] D. C. Alexander, G. J. Barker, and S. R. Arridge. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magn Reson Med*, **48**, 331–340. Aug 2002.

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

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

[38] Anders Brun, Mats Bjornemo, Ron Kikinis, Carl-Frerdic Westin. White matter tractography using sequential importance sampling. In *Annual Meeting of the ISMRM*, page 1131, (2002).

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

[40] Geoffrey J.M. Parker, Gareth Barker, Neil Thacker, Alan Jackson. A framework for a streamline-based probabilistic index of connectivity (pico) using a structural interpretation of anisotropic diffusion. In *Annual Meeting of the ISMRM*, page 1165, (2002).