

## Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?

T.E.J. Behrens,<sup>a,b,\*</sup> H. Johansen Berg,<sup>a</sup> S. Jbabdi,<sup>a</sup> M.F.S. Rushworth,<sup>a,b</sup> and M.W. Woolrich<sup>a</sup>

<sup>a</sup>Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Oxford, UK

<sup>b</sup>Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, UK

Received 26 April 2006; revised 4 September 2006; accepted 11 September 2006

Available online 27 October 2006

We present a direct extension of probabilistic diffusion tractography to the case of multiple fibre orientations. Using automatic relevance determination, we are able to perform online selection of the number of fibre orientations supported by the data at each voxel, simplifying the problem of tracking in a multi-orientation field. We then apply the identical probabilistic algorithm to tractography in the multi- and single-fibre cases in a number of example systems which have previously been tracked successfully or unsuccessfully with single-fibre tractography. We show that multi-fibre tractography offers significant advantages in sensitivity when tracking non-dominant fibre populations, but does not dramatically change tractography results for the dominant pathways.

© 2006 Elsevier Inc. All rights reserved.

Diffusion tensor tractography can reconstruct many white matter bundles that can be seen in post-mortem human dissection (Catani et al., 2002; Stieltjes et al., 2001) and inferred from ex vivo tracer studies in macaque (Behrens et al., 2003a; Lehericy et al., 2004). However, there are many examples of fibre pathways that are reported in dissection and tracer studies that are typically absent in diffusion tensor tractography studies. For example, when tracking from the cerebral peduncle or internal capsule, tractography studies tend only to find projections to the medial portion of the sensori-motor strip (e.g. Catani et al., 2002; Lazar and Alexander, 2005; Croxson et al., 2005 and many others), missing entirely the projections to lateral portions. These lateral regions, which appear unconnected to the peduncle, contain the arm representation which is known to have cortico-spinal connections in many primate species (Porter and Lemon, 1993). There are many further examples of pathways which are known to exist, but

are not reported in the tractography literature. A notable omission from the thalamic parcellation study presented in (Behrens et al., 2003a) is a projection from the medial geniculate nucleus (mgn) of thalamus to the primary auditory cortex. In Rushworth et al. (2005), the authors were able to find subcortical sites projecting to different portions of parietal lobe, but were unable to find the known parietal connections with premotor cortex and frontal eye fields that travel down the 1st and 2nd portions of the superior longitudinal fasciculus (SLF), despite several different attempted strategies. However, more lateral regions in the parietal lobe were shown to be connected to ventral premotor cortex via the 3rd portion of the SLF.

It would be tempting to suggest that, with increasing spatial resolution in our diffusion tensor images (DTI), we will see more and more of these previously invisible fibre bundles. However, these examples, and many others, exist at a spatial resolution that is available to DTI tractography. The lateral portions of the cortico-spinal and cortico-bulbar tracts, and the medial sections of the superior longitudinal fasciculus are far bigger pathways than many that we are able to see. The explanation for this apparent dichotomy can be found at a more local scale. Several recent techniques (Tuch et al., 2002; Tuch et al., 2003; Jansons and Alexander, 2003; Tournier et al., 2004; Liu et al., 2004; Alexander, 2005; Hosey et al., 2005; Van Welden et al., 2005) have been proposed that aim to describe local diffusion within an imaging voxel without imposing the Gaussian constraint central to the diffusion tensor model (Basser et al., 1994). These techniques have been able to estimate multiple dominant diffusion orientations within the same imaging voxel and suggest that they correspond to multiple fibre populations. In doing so, they have revealed large regions of white matter in which a single dominant orientation does not adequately describe the local diffusion profile. Included in these regions are all of the elusive pathways described above: the lateral motor pathways are obscured by the 3rd (lateral) branch of the SLF, the acoustic radiations from the mgn are obscured by the optic radiations, and the medial 1st portion of the SLF is obscured by the medial cortico-spinal fibres.

Attempts have been made to include such complex fibre information into both deterministic (Hagmann et al., 2004) and

\* Corresponding author. Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. Fax: +44 1865 222717.

E-mail address: [behrens@fmrib.ox.ac.uk](mailto:behrens@fmrib.ox.ac.uk) (T.E.J. Behrens).

Available online on ScienceDirect ([www.sciencedirect.com](http://www.sciencedirect.com)).

probabilistic (Hosey et al., 2005; Parker and Alexander, 2005) tractography techniques, with promising early results including pathways from cerebral peduncle to the entire motor strip. In this manuscript, we present a direct extension of a previously published probabilistic tractography routine (Behrens et al., 2003b) to the case of multiple fibre orientations in each voxel. We propose an online Bayesian method for assessing the most appropriate number of fibre orientations for the data at each voxel, and carry out probabilistic tractography through these complex orientation fields. Because this multi-fibre approach has a direct equivalent in the single fibre case, we are able to assess the advantages of increasing the local model complexity, while other factors remain constant. We find that, in general, the dominant pathway from a region remains the same between single- and multi-fibre approaches, but we show a number of examples where the multi-fibre approach is more sensitive to secondary or subordinate pathways.

## Methods

### Signal model

We use the model described in Behrens et al. (2003b) and Hosey et al. (2005). It is a partial volume model, where the diffusion-weighted MR signal is split into an infinitely anisotropic component for each fibre orientation, and a single isotropic component. Unlike Behrens et al. (2003b), here, we will infer on multiple fibre orientations.

The predicted signal for each diffusion-weighted measurement at each voxel is:

$$S_i = S_0 \left( \left( 1 - \sum_{j=1}^N f_j \right) \exp(-b_i d) + \sum_{j=1}^N f_j \exp \left( -b_i d r_i^T R_j A R_j^T r_i \right) \right) \quad (1)$$

where  $S_0$  is the non-diffusion-weighted signal value,  $d$  is the diffusivity,  $b_i$  and  $r_i$  are the  $b$ -value and gradient direction associated with the  $i$ th acquisition, and  $f_j$  and  $R_j A R_j^T$  are the fraction of signal contributed by, and anisotropic diffusion tensor along, the  $j$ th fibre orientation  $(\theta_j, \phi_j)$ , and  $N$  is the maximum number of fibres. That is,  $A$  is fixed as:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (2)$$

and  $R_j$  rotates  $A$  to  $(\theta_j, \phi_j)$ . The noise is modelled separately for each voxel as independently identically distributed (iid) Gaussian with a mean of zero and standard deviation across acquisitions of  $\sigma$ .

As in Behrens et al. (2003b), we will use Bayesian estimation to fit the parameters of this model to the signal at each voxel. However, there is a slight twist. We only want to infer multiple fibre populations when there is evidence in the data that they exist. In voxels which truly only support a single fibre orientation (such as medial callosal voxels), fitting a more complex model to the data may lead to poor estimation of the true fibre orientation and its uncertainty, and will, in any case, provide practical problems for tractography. Second fibre orientations will be followed when they do not truly exist. We solve this problem using a Bayesian trick known as

automatic relevance determination (ARD), or shrinkage priors, which were originally devised in the field of Neural Networks (MacKay, 1995), but have since been used in neuroimaging (Friston et al., 2003; Woolrich and Smith, 2001).

### Automatic relevance determination

Automatic relevance determination is different from other model selection techniques, as it does not fit different models to the data separately, and compare them on the basis of a metric reflecting data fit and model complexity. Instead, ARD fits the more complex model, but ensures that parameters that are not supported by the data, do not contribute to the likelihood (by ensuring that, in the posterior distribution, they take the value zero with very low variance). ARD can be applied to many different parameters within a model, and each ARD will act independently on its own particular parameter. A major advantage of ARD is that it only requires a single model to be fit to the data (as opposed to fitting every candidate model and comparing). Each parameter that is subject to ARD is then selected or deselected (forced to zero) depending on whether it is supported in the data.

Technically, this is achieved by placing a prior distribution on a parameter in a Bayesian model, which will force that parameter to take the value zero if, and only if, there is no evidence in the data for its existence. This prior distribution can take a number of forms. However, the most common of these is a Gaussian distribution with mean zero, but *unknown* variance. The variance is then estimated as a further parameter. If there is no evidence in the data for the existence of the original parameter, this variance term will be estimated very small, forcing the original parameter to zero. However, if the original parameter is supported by the data, the extra variance term will be estimated large, allowing the original parameter freedom to take any value.

To understand how this works, we must consider the energy functions in a Bayesian estimation scheme. The posterior energy (or negative log probability) is computed as the sum of the energies from the prior and likelihood distributions. If the original parameter does not contribute to the data, it will not reduce the likelihood energy, the prior energy will take over, the new variance will shrink to zero, and, with it, so will the original parameter. The trade-off is between the potential reduction in likelihood energy available from the better fit to the data afforded by the extra parameter, and the potential reduction in prior energy available from having low variance when the parameter is close to zero. Note that ARD cannot work in the context of point estimation schemes (for example, finding the parameters which maximise the joint posterior probability), as there is a singularity when the prior variance is 0 (generating infinite negative energy). ARD relies on the marginalisation over hyperparameters that is built in to any fully Bayesian estimation scheme. Many of the equations in the remainder of this section are expanded and explained in the Appendix.

### Toy example

To illustrate the use of ARD, we will show its behaviour in the context of fitting a mean to some data. The parameter which may or may not be included in the model is the mean itself. In this simple case, we would not need to use ARD, as we can compute the true Bayesian evidence for the two candidate models, with and without the mean. This evidence tells us the probability of the existence of the extra parameter given the data. However, we can

then use an ARD, and compare its results with the gold standard approach, which weights the two models by their Bayesian evidence.<sup>1</sup>

In our toy example, our two candidate models for the perfect Bayesian model averaging approach are:

$M0$ :

$$\begin{aligned} P(Y_i|\sigma) &\sim N(0, \sigma) \\ P(\sigma) &\sim \sigma^{-1} \end{aligned} \quad (3)$$

$M1$ :

$$\begin{aligned} P(Y_i|\mu, \sigma) &\sim N(\mu, \sigma) \\ P(\sigma) &\sim \sigma^{-1} \\ P(\mu) &\sim U(-\infty, \infty) \end{aligned} \quad (4)$$

where  $Y_i$  is the  $i$ th data point and  $\mu$  and  $\sigma$  are the mean and standard deviation to be estimated.  $U(-\infty, \infty)$  denotes a uniform distribution. Note that here, and throughout this manuscript, priors on variance parameters are reference priors  $P(\sigma) \sim \sigma^{-1}$  (see Bernardo and Smith, 2000 for a description of reference priors).

Our interest is in the posterior distribution on the mean,  $\mu$ , which we can arrive at by averaging the models, weighted by their Bayesian evidence. Note that to arrive at this equation, we have had to compute the Bayesian joint posterior distribution for each model, and then marginalise over the variance parameter (the elements in this equation are explained in the Appendix).

$$P(\mu|Y) = \frac{\delta(\mu=0)P(M0|Y) + P(\mu|Y, M1)P(M1|Y)}{P(M0|Y) + P(M1|Y)} \quad (5)$$

The ARD modelling approach can be written as follows:

$$\begin{aligned} P(Y_i|\mu, \sigma) &\sim N(\mu, \sigma) \\ P(\sigma) &\sim \sigma^{-1} \\ P(\mu|\sigma_\mu) &\sim N(0, \sigma_\mu) \\ P(\sigma_\mu) &\sim \sigma_\mu^{-1} \end{aligned} \quad (6)$$

Note that for the ARD case, we do not need a metric to select between models, nor do we need explicitly to average the models using their posterior probability given the data. The ARD is a prior on  $\mu$ , that will force  $\mu$  to zero if it is not supported by the data, hence it will not contribute to the likelihood function. If there is doubt about whether  $\mu$  is supported, then the posterior distribution on  $\mu$  will have weight at zero, and weight at the most likely value for  $\mu$ . The ARD will *implicitly* average the models with and without  $\mu$  with the appropriate weights.

In each case, we can perform analytical integration over all variance parameters leaving us with the posterior distribution of the mean,  $\mu$  given the data,  $Y$  (this integration is shown in the Appendix). Fig. 1(a) shows this distribution for data comprising 40 randomly drawn samples from a distribution with true mean, 0.5, and true variance 1. In this case, the model evidence suggests a probability of 0.93 that the more complex model is favoured. The posterior distribution for the ARD model has two peaks. One is at zero (reflecting the probability of the simpler model) and another is at a non-zero value (reflecting the probability of the more complex model). The non-zero peak accounts for 0.84 of the weight of the distribution (compared to the true probability of 0.93 for the complex model). The slight difference between the two approaches demonstrates the approximate nature of the ARD solution.

#### ARD on constrained parameters

We need to consider one further subtlety before applying an ARD prior to the volume fraction parameters: these parameters must be constrained to be between 0 and 1. Therefore, instead of

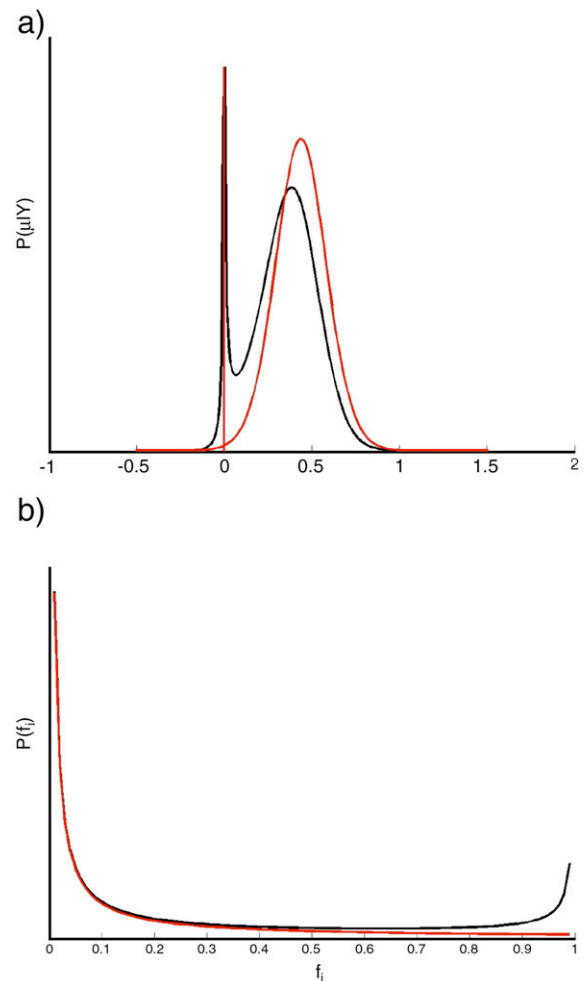


Fig. 1. (a) Marginal posterior distribution on  $\mu$  for true model averaging (red), and ARD approximation (black). (b) Effective prior distribution given by standard Gaussian ARD (red), and range-limited Beta ARD (black) in the range [0 1].

<sup>1</sup> Note that in complicated models, such as the one we will use for the diffusion data, the evidence for competing models cannot be computed analytically.

using a Gaussian distribution with an unknown variance, we use a Beta distribution with mode at zero and an unknown width (see Appendix for a description of the beta distribution).

$$P(f_i|\eta) \sim \beta(1, \eta) \quad (7)$$

$$P(\eta) \sim \eta^{-1} \quad (8)$$

where, again, we use a reference prior for  $\eta$  (Bernardo and Smith, 2000). We can integrate over  $\eta$  to give us the effective prior distribution on  $f_i$ , giving us:

$$P(f_i) = -\frac{1}{(1-f_i)\log(1-f_i)} \quad (9)$$

Similarly, we can integrate out the variance parameter from a standard Gaussian ARD prior (from Eq. (7)) giving:

$$P(\mu) = \frac{1}{|\mu|} \quad (10)$$

These two distributions can be compared in Fig. 1(b). The question of model comparison becomes relevant when  $f_i$  is small; that is when the  $i$ th fibre population is accounting for only a small proportion of the data. Note the similar behaviour in the priors in this region.

#### Model priors

We use ARD on the parameters representing the volume fractions of all but the first fibre,  $f_2^N$  (where  $\frac{N}{2}$  implies all fibres from 2 to  $N$ ). This means that if there is no evidence for a second or third fibre orientation in the data, the volume fraction attributed to these fibres will be forced to zero. We do not apply an ARD to the first fibre in order to ensure that we are directly comparable with the previously published single fibre algorithm. i.e. that any increased sensitivity we see in the results can be solely attributed to an increase in complexity in local modelling.

The prior distributions on model variables are therefore as follows:

$$P(S_0) \sim U(0, \infty) \quad (11)$$

$$P(f_1) \sim U(0, 1) \quad (12)$$

$$P(f_2^N) \sim \text{ARD} \quad (13)$$

$$P(\theta_1^N) \sim \sin(\theta) \quad (14)$$

$$P(\phi_1^N) \sim U(0, 2\pi) \quad (15)$$

$$P(\sigma) \sim \sigma^{-1} \quad (16)$$

For details about these priors, see Behrens et al. (2003b). We also apply the further constraint that  $\sum_{i=1}^N f_i < 1$ . This constraint is

applied easily in Markov Chain Monte Carlo, by rejecting any sample in which the constraint is not met.

#### Model estimation

We perform model estimation using Metropolis Hastings Markov Chain Monte Carlo sampling, after analytically marginalising over variance parameters, as in Behrens et al. (2003b). We burn in for 2000 jumps, then run the chain for a further 1000 jumps sampling every 20. Initialisation is by the log-linear diffusion tensor fit as in Behrens et al. (2003b) unless any preprocessed neighbouring voxel has more than one surviving fibre orientation, in which case parameters are initialised according to the mean values from this neighbouring voxel. At the end of this procedure, we have samples from the posterior probability distribution on every parameter in the model, including the orientation and volume fraction parameters from each fibre population.

#### Probabilistic tractography in a multi-fibre field

We perform tractography using the same sampling scheme that was shown to sample from the global probability of connection in Behrens et al. (2003b), and has also been used in Parker and Alexander (2003), Hagmann et al. (2003), Jones and Pierpaoli (2005), and Lazar and Alexander (2005).

This scheme amounts to streamline tractography except, at each step, instead of progressing along the most likely principal diffusion direction, we draw a sample from the posterior distribution on principal diffusion directions and progress along this same direction. This streamline becomes a sample from the connectivity distribution, and we draw a large number of samples. Two streamline samples arriving at the same point in space will choose different samples from the posterior distribution on local orientation at that voxel, and therefore leave along different directions, hence accounting for the uncertainty in local fibre orientation. After drawing a large number of samples, we are able to compute the probability of the dominant streamline passing through any single voxel or region, by counting the number of streamlines that passed through that region and dividing by the total number of samples drawn.

The only adaption we need to make this procedure amenable to the multi-fibre case is, at each step, to choose from which population we draw the orientation sample. There are a number of available options here. Particularly appealing, given the modelling approach above, would be to draw a sample from a whole set of fibre orientations, and choose each orientation,  $i$ , with a probability proportional to  $f_i$  at that sample. However, in a crossing fibre region, this scheme will preferentially sample the dominant fibre, independent of the previous orientation of the streamline. Instead, we choose a scheme that aims to maintain the orientation of the streamline, as this will allow us to track non-dominant pathways through crossing regions. Again, we draw a single sample from a set of fibre orientations, we examine each fibre population to confirm that the volume fraction has not been forced to zero by automatic relevance determination (we threshold at 0.05). We then draw a sample from *each* remaining population (i.e. those have been judged to be supported by the data), and choose that sample whose orientation is closest to parallel to the preceding orientation of the streamline.



## Data acquisition

### 60 direction data

Diffusion-weighted data were acquired on a 1.5 T Siemens Sonata scanner by using echo planar imaging ( $72 \times 2$  mm thick axial slices, matrix size  $128 \times 104$ , field of view  $256 \times 208$  mm<sup>2</sup>, giving a voxel size of  $2 \times 2 \times 2$  mm). The diffusion weighting was isotropically distributed along 60 directions by using a  $b$ -value of  $1000 \text{ s mm}^{-2}$ , allowing an echo time of 89 ms and a repetition time of 9 s. For each set of diffusion-weighted data, 5 volumes with no diffusion weighting were acquired at points throughout the acquisition. Three sets of diffusion-weighted data were acquired

for subsequent averaging to improve signal-to-noise ratio. The total scan time for the diffusion-weighted imaging (DWI) protocol was 45 min.

### 12 direction data

From these same data, we selected the 12 gradient orientations that performed best under the energy minimisation scheme proposed by Jones et al. (1999), to create a 12 direction diffusion data set with a single non-diffusion-weighted volume.

In each data set, each scan was subsequently aligned to a reference ( $b=0$ ) scan using affine registration intended to maximise mutual information (Jenkinson and Smith, 2001). Brain

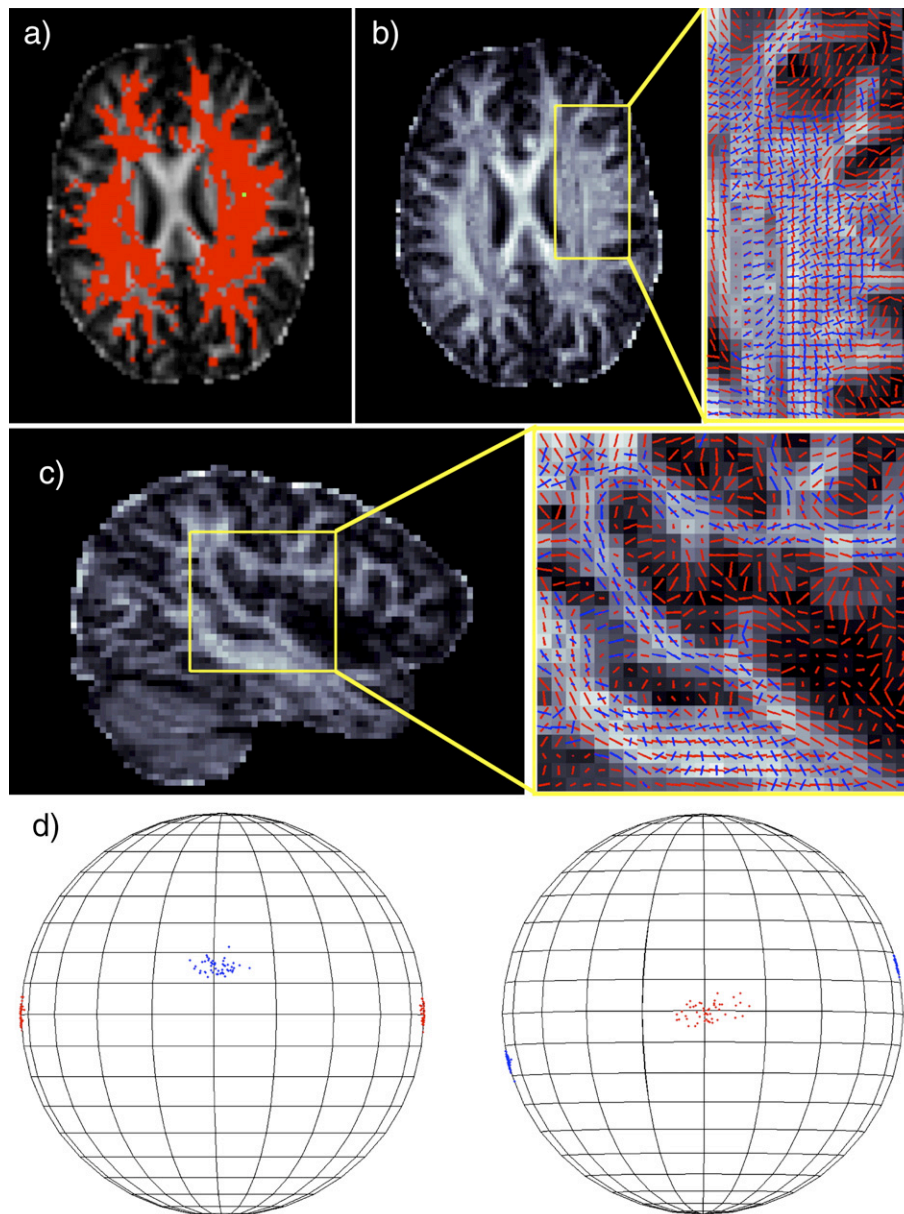


Fig. 2. Probabilistic multi-orientation fitting. (a) Axial slice showing regions where more than a single orientation were supported (thresholded at  $f_i$  0.5 after ARD-based estimation). (b and c) Axial and sagittal close ups of crossing fibre bundles with dominant fibre orientation in red and second in blue. Directions shown are the mean vectors of the posterior distribution samples. (d) Samples from the posterior distributions on the first two fibre orientation in a voxel (green dot in a) where the lateral motor projections (red) cross the longitudinal SLF projections (blue). The right hand-side is a  $90^\circ$  rotation of the left.

extraction derived from this reference scan was then applied to each volume (Smith et al., 2002).

## Results

### Local fitting of the complex fibre model

Fig. 2 shows typical examples of the probabilistic multi-fibre fit to the 60 direction data set. Note the wide extent of complex fibre architecture in Fig. 2(a). In this case, the diffusion data at one third of voxels with Fractional Anisotropy  $>0.1$  were able to support more than one fibre orientation, but no single voxel in the data set supported more than two orientations. In the 12 direction set, no single voxel supported more than a single fibre orientation. Note that the method is general to any number of fibre populations, given a richness of data to support them (see Appendix for relevant simulations investigating the angular resolution of the technique). (b) and (c) show close ups in lateral portions of the brain of the crossings between the longitudinal and the medio-lateral fibres. The second fibre direction (blue) is consistent between voxels, and joins into a connected path even in fine fibre bundles (c). (d) Shows samples from the posterior distributions on orientations of motor (red) and SLF (blue) orientations in a single voxel at the crossing. In human DTI data, the SLF dominates the motor pathways at this crossing, but fitting multiple orientations reveals that the fibre populations are not dissimilar in contribution to the MR signal. In this voxel (which is typical), the mean of the posterior distributions on  $f_1$  and  $f_2$  is 0.41 and 0.32, respectively. These numbers can be interpreted as the relative contributions to the MR signal of the SLF ( $f_1$ ) and motor ( $f_2$ ) projections. Note that the slight difference in volume fractions can also be seen as a slightly wider spread in the posterior distribution on the second orientation.

### Probabilistic tractography

#### Tracking dominant projections

To interrogate the effect of the multi-fibre model on dominant fibre projections, we performed a connectivity-based thalamic segmentation as described in Behrens et al. (2003a) using both single and multi-fibre approaches. We defined seven cortical regions (prefrontal cortex, premotor cortex, primary motor cortex, primary sensory cortex, posterior parietal cortex, occipital cortex, and temporal cortex) and thalamus from population probability maps (taken from Johansen-Berg et al., 2004), thresholded at 30%. We drew 5000 samples from the connectivity distribution from each seed voxel in thalamus and computed the probability of connection to each cortical mask. We then assigned an index to each seed voxel corresponding to the target cortical zone with the highest connection probability to the seed. Here, and in subsequent experiments, we used a curvature threshold of  $80^\circ$  for stopping streamline trajectories (Behrens et al., 2003b). The results with the single and multi-fibre approaches in a single typical subject can be seen in Fig. 3. The dominant projections can be seen to be similar between the two approaches.

#### Tracking non-dominant projections

To interrogate the increased sensitivity provided by the multi-fibre model, we performed tractography on pathways where known anatomical connections have previously proved difficult to trace in diffusion data.

### Motor pathways

We defined a seed mask as the internal capsule on a single slice ( $Z=6$ ) of the MNI 152 average brain (Evans et al., 1992). We then defined primary motor cortex (M1) from a population probability map (taken from Johansen-Berg et al., 2004) thresholded at 30%. In each of nine subjects, we then drew 5000 samples from the connectivity distribution from each seed voxel and retained only those samples that passed through the M1 target mask. To avoid inter-hemispheric connections, samples were excluded from further analysis if they crossed the mid-sagittal plane. We then repeated exactly the same experiment using single fibre tractography described in Behrens et al. (2003b). Fig. 4 shows maximum intensity projections of the results for the two approaches.

In each subject, the multi-fibre approach reveals more of the cortico-spinal pathways than the single-fibre approach. However, the ability of the multi-fibre approach to find the lateral motor projections is variable across subjects.

### Medial SLF

We performed the same experiment in the medial portion of the SLF. Seed and target masks were defined as rectangles covering medial parietal and medial premotor cortex on a single slice ( $Z=64$ ) of the MNI-152 brain. Samples crossing the midline or MNI slice  $Z=40$  were terminated, so as any projections running through the corpus callosum or the cingulum bundle.

Fig. 5 shows maximum intensity projections of the results for the two approaches. In each subject, the medial portion of the SLF is found by the multi-fibre approach, but is invisible to the single fibre approach as it is obscured by the medial motor projections.

### Acoustic radiations

Seeds were defined as a cuboid on three slices ( $Z=-6$  to  $-2$ ) of the MNI average brain just medial to the left lateral geniculate nucleus (e.g. Devlin et al., 2006). The target was defined as a region encompassing primary auditory cortex on a single sagittal slice ( $X=-56$ ). To exclude other thalamo-temporal pathways (see e.g. Behrens et al., 2003a) and thalamo-parietal pathways (due to misregistration between individual subject space and the MNI space target), samples crossing the following planes were terminated:  $Z=-12$ ,  $Z=46$ ,  $Y=-44$ ,  $Y=4$ .

Fig. 6 shows maximum intensity projections of the results for the two approaches. In each subject, the acoustic radiation was

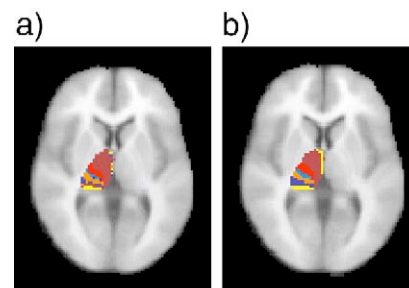


Fig. 3. Thalamic parcellation with single fibre (a) and crossing-fibre (b) tractography. Key to cortical projections is as follows. Prefrontal cortex in burgundy, premotor cortex in red, primary motor cortex in light blue, primary sensory cortex in dark blue, posterior parietal cortex in orange, occipital cortex in mid-blue, and temporal cortex in yellow.

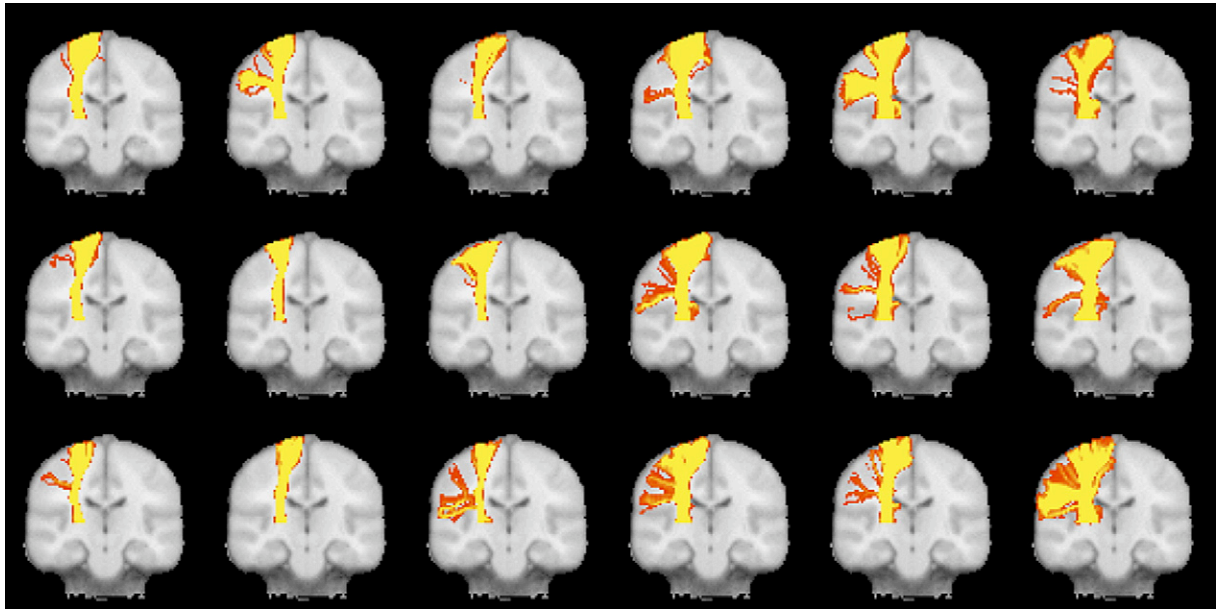


Fig. 4. Tracking the cortico-spinal tract from the internal capsule to the primary motor cortex with single fibre (left) and multi-fibre (right) tractography. A coronal maximum intensity projection is shown. Voxels are colour coded from 25 (red) to 200 (yellow) samples passing through the voxel. Single fibre results are shown as the leftmost nine ( $3 \times 3$ ) subjects, and multi-fibre results are the rightmost nine subjects. Each subject is displayed in the same location in the two grids.

found by the multi-fibre approach, but was invisible to the single fibre approach as it is obscured by the optic radiations.

#### *Limbic pathways of the subgenual white matter*

The subgenual white matter has been a successful target for deep-brain stimulation in patients with treatment resistant depression

(Mayberg et al., 2005). Beneficial affects of stimulation could be mediated by paths to remote interconnected regions of the frontal and limbic system (Mayberg et al., 2005). Connections from this region to the amygdala, a region critically implicated in depression and recovery (Drevets, 2003), are largely lacking with single fibre tractography, but connections have been found connecting the same two regions tracking in the opposite direction (Crosson et al., 2005).

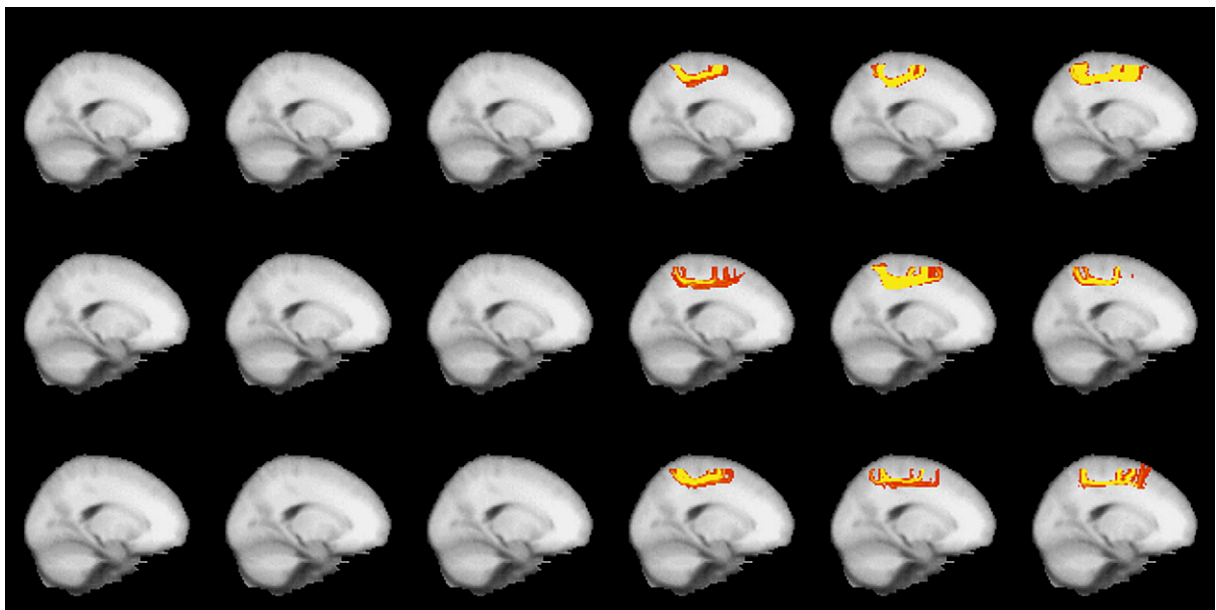


Fig. 5. Tracking of the parietal–premotor connections of the medial portion of the superior longitudinal fasciculus in nine subjects with single fibre ( $3 \times 3$  on left) and multi-fibre ( $3 \times 3$  on right) tractography. A sagittal maximum intensity projection is shown. Voxels are colour coded from 25 (red) to 200 (yellow) samples passing through the voxel. Note that the single fibre tractography was not able to find any connections that reached the target in any subject.



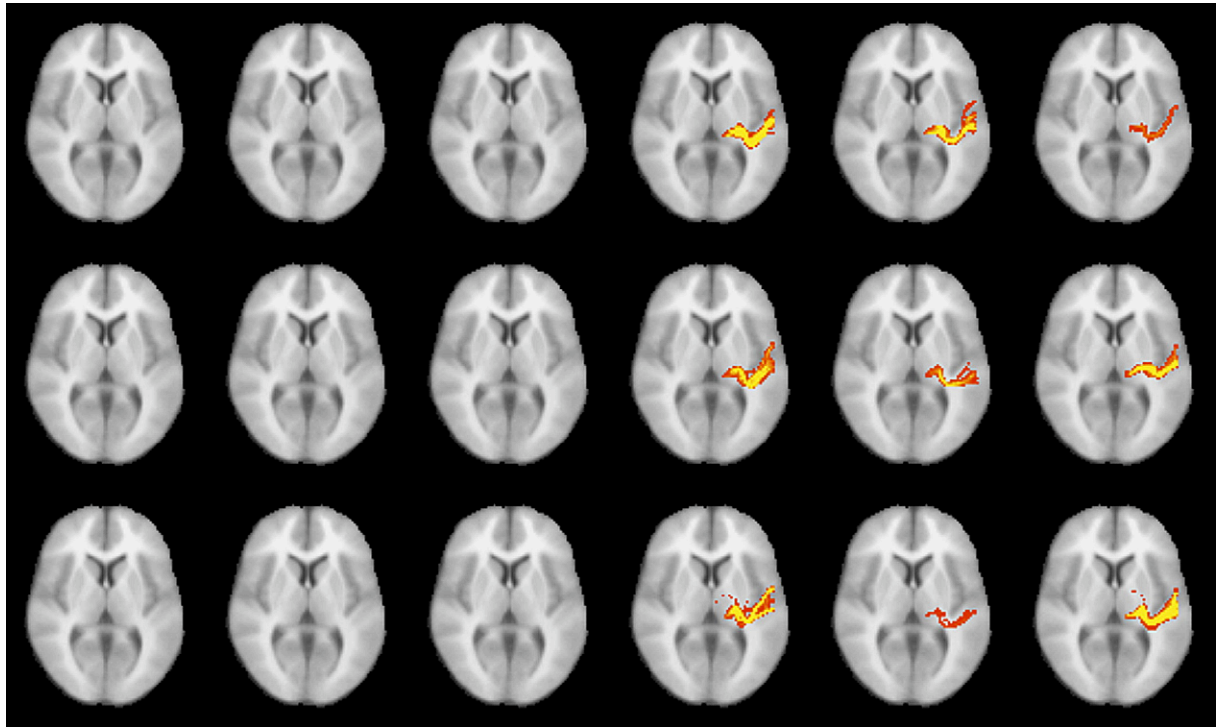


Fig. 6. Tracking of the acoustic radiations between medial geniculate nucleus of thalamus and primary auditory cortex with single fibre (left) and multi-fibre (right) tractography. An axial maximum intensity projection is shown. Voxels are colour coded from 10 (red) to 50 (yellow) samples passing through the voxel. Note that the single fibre tractography was not able to find any connections that reached the target in any subject.

We generated MNI space seed and target masks of the subgenual white matter, and the amygdala, and ran tractography using both single- and multi-fibre techniques as described above.

Fig. 7 shows maximum intensity projections of the results for the two approaches. With the single-fibre algorithm, the pathway

was visible in 4 of 9 subjects, and with high variability in connection likelihoods. With the multi-fibre algorithm, the pathway could be seen in all subjects, but again with varying degrees of connection likelihood. This pathway is the non-dominant pathway in the region of complex fibre architecture in the subgenual white

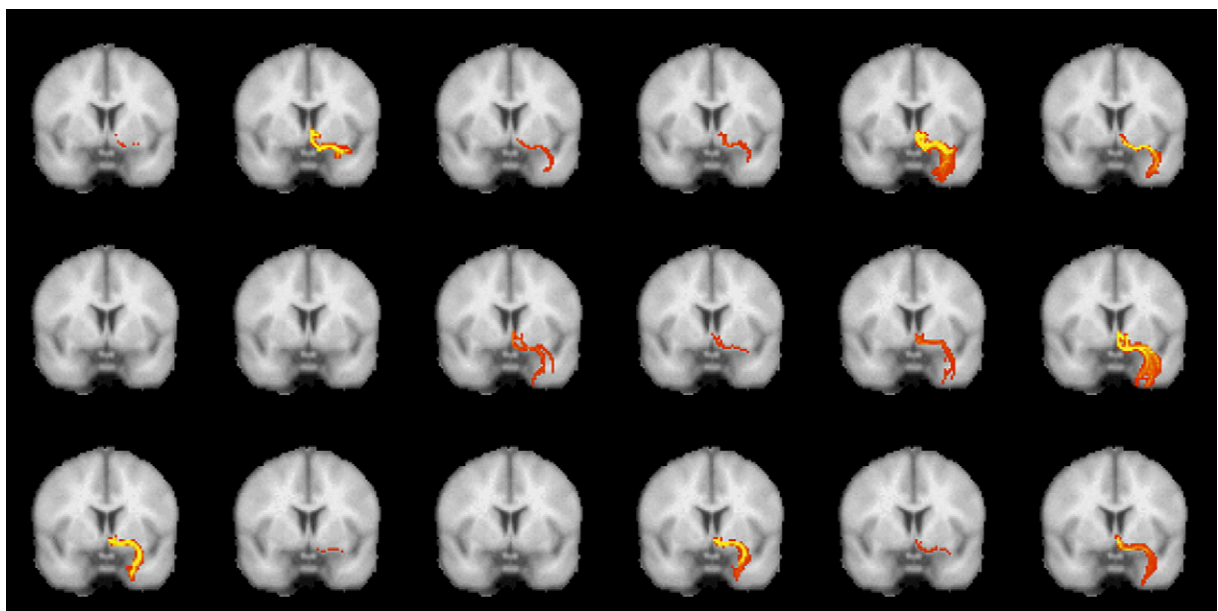


Fig. 7. Tracking the projection from the subgenual white matter to amygdala using single fibre (left) and multi-fibre (right) tractography. An coronal maximum intensity projection is shown. Voxels are colour coded from 10 (red) to 100 (yellow) samples passing through the voxel.



matter. Therefore, as expected the multi-fibre approach has higher connection likelihoods than the single-fibre approach in all subjects.

## Discussion

We have presented a direct extension of Behrens et al. (2003b) to the case of complex fibre architecture, and shown that, using this modelling approach, it is possible to detect white matter regions of complex fibre architecture that have previously been identified by model-free inversion approaches (Tuch et al., 2003; Tuch et al., 2005; Jansons and Alexander, 2003; Donald Tournier et al., 2004; Ozarslan et al., 2005). Furthermore, by inferring on a model of local diffusion, we are able to limit the regions in which multiple fibres are detected to those where the data require such a complex model. In order to perform this reduction, we use *automatic relevance determination* in a formal Bayesian estimation scheme. Using a clinically feasible acquisition scheme with 60 diffusion directions, and a  $b$ -value of 1000, we were able to detect complex fibre architecture in approximately a third of voxels with an FA greater than 0.1. As shown in Hosey et al. (2005) and Parker and Alexander (2005), such approaches dramatically reduce the problem of tractography in a multi-fibre field to an extent that it is tractable to probabilistic sampling schemes. Tractography approaches that rely on techniques that detect complex fibre architecture in every voxel, whether or not it is supported by the data, will be much more susceptible to false positive connections than approaches that infer complexity in a selective manner.

Here, by using a natural extension to a single fibre tractography technique, we are able to examine directly the potential advantages of estimating complex architecture before performing tractography.

Connectivity-based segmentation of thalamus (Behrens et al., 2003a) which relies only on identifying the target with the highest connection probability was largely unchanged between the single and multi-fibre approaches, suggesting that the same dominant pathway is found by both approaches.

Furthermore, in four separate fibre systems where non-dominant pathways have previously been hard to find, we find that multi-fibre tractography is more sensitive to non-dominant projections than single fibre tractography. For example, the medial portion of the SLF was found strongly in all subjects, whereas it is invisible to single fibre approaches. However, in some cases this non-dominant projection is found with variable success and connection likelihood from subject to subject. For example, in the cortico-spinal tract, lateral projections to primary motor cortex were not found strongly in every subject (Fig. 4). It has previously been reported (Tuch et al., 2003) that there are regions of the cortico-spinal tract where it has been possible to estimate three populations of fibres in high  $b$ -value diffusion data. The data we collected were not able to support a third orientation. It is possible that if we acquired more diffusion directions at a higher  $b$ -value, we would detect more than two orientations and therefore find the whole cortico-spinal tract in a larger proportion of subjects. Simulations provided in the Appendix of this manuscript suggest that the data we present here can resolve 2 fibre systems with an angular resolution of around  $45^\circ$ , but that in order to resolve a 3 fibre orthogonal system robustly, we would require  $b$ -values upwards of 4000 (see Appendix and Figs. 8 and 9).

In fact, the sensitivity to detect multiple fibre populations will depend not only on the data, but also on specifics of the

reconstruction technique. Techniques for inferring multiple populations of fibres from diffusion data can largely be separated into two groups (Alexander, 2005): Those that use a model of the underlying diffusion profile, and therefore try to estimate fibre orientations (e.g. Tuch et al., 2002; Hosey et al., 2005; Donald Tournier et al., 2004; Assaf et al., 2004); and those that attempt to estimate the diffusion profile itself in a model-free environment (e.g. Tuch et al., 2003; Van Welden et al., 2005; Ozarslan and Mareci, 2003). Two of these methods have particular similarities to the approach taken here. CHARMED imaging (Assaf et al., 2004) models the diffusion signal as a mixture of hindered and restricted diffusion compartments from around and within white matter axons. These compartments are modelled with a diffusion tensor (for the extra-axonal hindered compartment) and restricted cylindrical diffusion (for the intra-axonal restricted compartment). This model can be shown to estimate multiple fibre orientations with lower uncertainty (and hence a better fit to the data) than a model which ignores the restricted component (Assaf and Basser, 2005). There are strong parallels between CHARMED imaging, and the compartment model used in this manuscript. The key differences are the stronger biophysical motivation for the compartment divisions giving CHARMED the potential to ask quantitative questions about, for example, relative compartment volumes; and the simplicity of the model used in this manuscript meaning that it can be used with much sparser data (of the sort that is commonly acquired in research environments, and is easily transferred to the clinical domain). Note that the simple model used in this manuscript will not allow quantitative analysis of, for example, the relative contribution of signal from extra- and intra-cellular compartments, or of the specific biophysical parameters associated with either compartment. Such quantitative investigations may be available to a biophysically motivated model such as CHARMED. Q-ball imaging (Tuch et al., 2003; Tuch, 2004) takes a very different approach, attempting to reconstruct the entire diffusion profile, rather than fibre orientations per se. However, there are parallels with the approach taken here. The Funk Radon projection used to estimate this profile from a shell of  $q$ -space data is a projection of the data onto each infinitely anisotropic spike (much like those used in this paper). There are also key differences here, which highlight the advantages of both the model-based and model-free estimation approaches. Using a model-free approach (e.g. Q-ball imaging), the potential angular resolution of the reconstructed maxima in the diffusion function is limited by the diffusion response function of a single fibre, even in a noise-free environment. Multiple lobes in the diffusion function can merge into one, giving an incorrect maximum. When explicit assumptions are made about the single fibre response function (using a model-based approach), there is no theoretical limit to the angle that can be resolved, if the model is a good predictor of the signal. However, errors in the assumed model will lead to errors in the recovered fibre orientations (and uncertainties).

Notwithstanding differences between techniques, our experience suggests that the most likely connections tend to be common to the two approaches (e.g. the thalamus segmentation presented above), but that in cases where false negatives in single-fibre tractography are due to difficulties in tracing a non-dominant pathway, the multi-fibre approach gives strikingly different results. This is encouraging, as it means that the increased complexity is unlikely to result in many more false positives, but it is also discouraging, as it means that any false positive results present in the simpler approach are also likely to

be present when complex fibre architecture is modelled explicitly.

In summary, we believe that explicit modelling of fibre complexity will offer significant increases in sensitivity over traditional tensor, or single-fibre approaches, but that results such approaches should be interpreted with the anatomical care that has become indispensable in tractography studies.

## Acknowledgments

The authors would like to acknowledge funding from the UK Medical Research Council (TEJB, HJB), the Royal Society (MFSR), the UK Engineering and Physical Sciences Research Council (MWW), the Dr. Hadwen Trust (SJ) and the Wellcome Trust (HJB). We would also like to thank Professor Paul Matthews and the Isle of Islay for providing their combined support for this work.

## Appendix A

### A.1. Supplementary information for ARD

Referring to Eq. (5): Here all probability distributions have been marginalised over parameters of no interest, so:

$$P(M0|Y) \propto P(Y|M0) = \int_0^\infty P(Y, \sigma) d\sigma \quad (17)$$

$$P(M1|Y) \propto P(Y|M1) = \int_{-\infty}^\infty \int_0^\infty P(Y, \mu, \sigma) d\sigma d\mu \quad (18)$$

$$P(\mu|Y, M1) = \int_0^\infty P(\mu, \sigma|Y, M1) d\sigma \quad (19)$$

In this case, all of these integrals give closed form analytical expressions, given a set of data.  $Y$ , the respective evidences for each model and the overall model-averaged posterior distribution on  $\mu$  may be computed exactly (Bernardo and Smith, 2000).

Referring to Eq. (7) and Fig. 1, as above, all variance parameters in this model may be integrated out analytically, giving the following closed form analytical expression for the posterior distribution on  $\mu$  in the ARD model:

$$P(\mu|Y) = \int_0^\infty \int_0^\infty P(\mu, \sigma, \sigma_\mu) d\sigma d\sigma_\mu \quad (20)$$

$$P(\mu|Y) \propto \int_0^\infty \int_0^\infty P(Y|\mu, \sigma) P(\mu|\sigma_\mu) P(\sigma_\mu) P(\sigma) d\sigma d\sigma_\mu \quad (21)$$

$$P(\mu|Y) = \frac{\left[ \sum_{i=1}^{NP} (Y_i - \mu)^2 \right]^{\frac{NP-1}{2}}}{|\mu|} \quad (22)$$

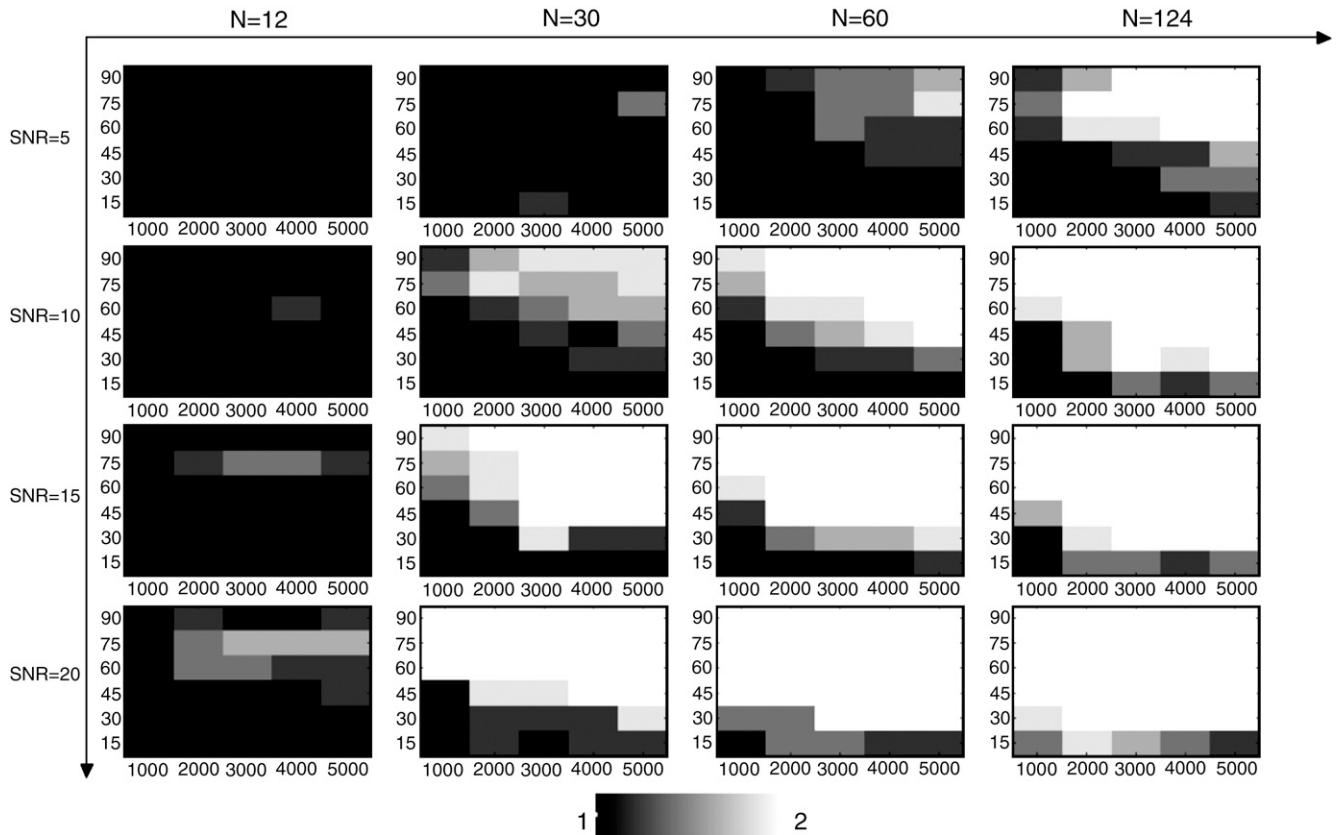


Fig. 8. Simulations of two crossing fibres with different number of diffusion encoding orientations (main x-axis), SNR (main y-axis),  $b$ -values (individual x-axes), and separation angle (individual y-axes). Simulations were repeated 5 times. Greyscale shows the average number of fibres recovered by the estimation technique.

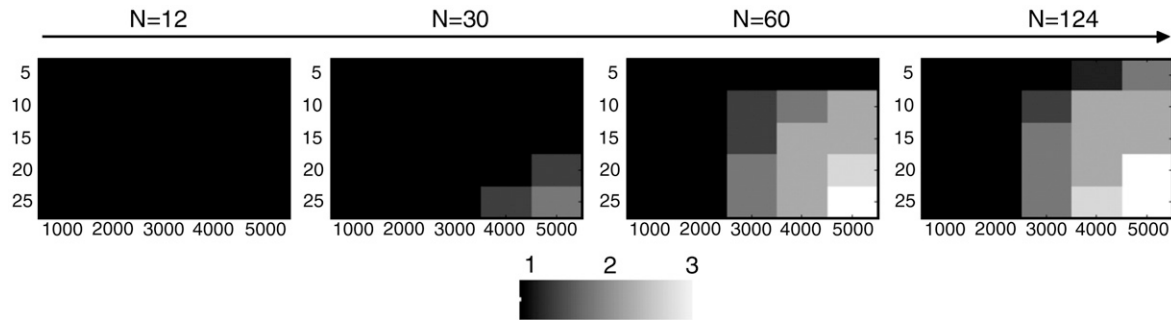


Fig. 9. Simulations of three fibres crossing at 90° with different number of diffusion encoding orientations (main x-axis),  $b$ -values (individual x-axes), and SNR (individual y-axes). Simulations were repeated 5 times. Greyscale shows the average number of fibres recovered by the estimation technique.

where NP is the number of data points. This is the function that is plotted in black in Fig. 1(a).

Referring to Eq. (7), the Beta distribution is defined as follows:

$$\beta(x; \zeta, \eta) = \frac{\Gamma(\zeta + \eta)(1 - x)^{\eta-1} x^{\zeta-1}}{\Gamma(\zeta)\Gamma(\eta)} \quad (23)$$

implying that

$$\beta(x; 1, \eta) = \eta(1 - x)^{\eta-1} \quad (24)$$

## Appendix B. Simulations of angular resolution of the technique

Here we show results of estimating multiple populations of fibres on simulated data, where the true number of fibre populations is known. It should be carefully noted that the results we show here are the best possible case for this (or any model-based) technique, as the simulated data are generated from the same model that is being used in the estimation phase. We should not assume that fibre populations that are separable in this simulated data, are also separable in real data.

### B.1. Simulations

Data were simulated from the generative model described in Eq. (1). In Fig. 8, there were 2 fibres in the generative (ground-truth) model, each of which contributed 0.35 of the signal ( $f_1=f_2=0.35$ ). Other constant parameters were:  $d=0.0015 \text{ mm}^2 \text{ s}^{-1}$ ,  $S_0=100$ . The angle between the two fibres was incremented in steps of 15°, the  $b$ -value was incremented in steps of 1000  $\text{s mm}^{-1}$  from 1000–5000  $\text{s mm}^{-1}$ . We chose 4 different sets of diffusion-encoding orientations ( $N=12, 30, 60, 124$ ), all sampled isotropically on a sphere. Finally, we added Gaussian noise at 4 different values (in the non-diffusion-weighted images  $SNR=5, 10, 15, 20$ ). This set of simulations was performed on 5 occasions. Fig. 8 shows the average number of fibres estimated on each occasion (with a fibre threshold of 0.05). Note that there was no event of more than 2 fibres being estimated by the technique, despite the model being run with a maximum of 3 fibres. In each case, the 3rd fibre was suppressed by the ARD.

In Fig. 9, data were generated using 3 populations of fibres crossing at 90° ( $f_1=f_2=f_3=0.25$ ). All other settings were the same as for Fig. 8. Three fibres were never estimated with  $b$ -values lower than 3000. Note that at low  $b$ -values, the three fibre crossing at 90° is a particularly difficult problem for any

technique, as the data appear approximately isotropic. Often only one fibre is estimated. If the ARD is placed on all three fibres (instead of being left off the first), in some low  $b$ -value cases, there is no evidence for any fibre.

## References

- Alexander, Daniel C., 2005. Multiple-fiber reconstruction algorithms for diffusion MRI. *Ann. N. Y. Acad. Sci.* 1064, 113–133 (Dec).
- Assaf, Yaniv, Basser, Peter J., 2005. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage* 27, 48–58 (Aug).
- Assaf, Yaniv, Freidlin, Raisa Z., Rohde, Gustavo K., Basser, Peter J., 2004. New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magn. Reson. Med.* 52, 965–978 (Nov).
- Basser, P.J., Mattiello, J., Le Bihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson., Ser. B* 103, 247–254.
- Behrens, T.E.J., Johansen-Berg, H., Woolrich, M.W., Smith, S.M., Wheeler-Kingshott, C.A.M., Boulby, P.A., Barker, G.J., Sillery, E.L., Sheehan, K., Ciccarelli, O., Thompson, A.J., Brady, J.M., Matthews, P.M., 2003a. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci.* 6 (7), 750–757 (Sep).
- Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003b. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088 (Nov).
- Bernardo, J.M., Smith, A.F.M., 2000. *Bayesian Theory*. Wiley.
- Catani, M., Howard, R.J., Pajevic, S., Jones, D.K., 2002. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage* 17, 77–94 (Sep).
- Croxson, Paula L., Johansen-Berg, Heidi, Behrens, Timothy E.J., Robson, Matthew D., Pinski, Mark A., Gross, Charles G., Richter, Wolfgang, Richter, Marlene C., Kastner, Sabine, Rushworth, Matthew F.S., 2005. Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *J. Neurosci.* 25, 8854–8866 (Sep).
- Devlin, J.T., Sillery, E.L., Hall, D.A., Hobden, P., Behrens, T.E.J., Nunes, R.G., Clare, S., Matthews, P.M., Moore, D.R., Johansen-Berg, H., 2006. Reliable identification of the auditory thalamus using multi-modal structural analyses. *NeuroImage* 30, 1112–1120 (May).
- Drevets, Wayne C., 2003. Neuroimaging abnormalities in the amygdala in mood disorders. *Ann. N. Y. Acad. Sci.* 985, 420–444 (Apr).
- Evans, A.C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W., Milot, S., Meyer, E., Bub, D., 1992. Anatomical mapping of functional activation in stereotactic coordinate space. *NeuroImage* 1, 43–53 (Aug).
- Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *NeuroImage* 19, 1273–1302 (Aug).

- Hagmann, P., Thiran, J.-P., Jonasson, L., Vandergheynst, P., Clarke, S., Maeder, P., Meuli, R., 2003. DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection. *NeuroImage* 19, 545–554 (Jul).
- Hagmann, P., Reese, T.G., Tseng, W.-Y.I., Meuli, R., Thiran, J.-P., Wedeen, V.J., 2004. Diffusion spectrum imaging tractography in complex cerebral white matter: an investigation of the centrum semiovale. *ISMRM*, p. 623.
- Hosey, Tim, Williams, Guy, Ansoorge, Richard, 2005. Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magn. Reson. Med.* 54, 1480–1489 (Dec).
- Jansons, Kalvis M., Alexander, Daniel C., 2003. Persistent angular structure: new insights from diffusion MRI data. Dummy version. *Inf. Process. Med. Imag.* 18, 672–683 (Jul).
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156 (Jun).
- Johansen-Berg, H., Behrens, T.E., Sillery, E., Ciccarelli, O., Thompson, A.J., Smith, S.M., Matthews, P.M., 2004. Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. *Cereb. Cortex* (Jul).
- Jones, Derek K., Pierpaoli, Carlo, 2005. Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. *Magn. Reson. Med.* 53, 1143–1149 (May).
- Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn. Reson. Med.* 42, 515–525 (Sep).
- Lazar, Mariana, Alexander, Andrew L., 2005. Bootstrap white matter tractography (BOOTTRAC). *NeuroImage* 24, 524–532 (Jan).
- Lehericy, Stephane, Ducros, Mathieu, Krainik, Alexandre, Francois, Chantal, Van de Moortele, Pierre-Francois, Ugurbil, Kamil, Kim, Dae-Shik, 2004. 3-D diffusion tensor axonal tracking shows distinct SMA and Pre-SMA projections to the human striatum. *Cereb. Cortex* 14, 1302–1309 (Dec).
- Liu, Chunlei, Bammer, Roland, Acar, Burak, Moseley, Michael E., 2004. Characterizing non-Gaussian diffusion by using generalized diffusion tensors. *Magn. Reson. Med.* 51, 924–937 (May).
- 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, 469–505.
- Mayberg, Helen S., Lozano, Andres M., Voon, Valerie, McNeely, Heather E., Seminowicz, David, Hamani, Clement, Schwab, Jason M., Kennedy, Sidney H., 2005. Deep brain stimulation for treatment-resistant depression. *Neuron* 45, 651–660 (Mar).
- Ozarslan, Evren, Mareci, Thomas H., 2003. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magn. Reson. Med.* 50, 955–965 (Nov).
- Ozarslan, Evren, Shepherd, Timothy M., Vemuri, Baba C., Blackband, Stephen J., Mareci, Thomas H., 2005. Fast orientation mapping from HARDI. *Med. Image Comput. Comput. Assist. Interv. Int. Conf. Med. Image Comput. Comput. Assist. Interv.* 8, 156–163.
- Parker, Geoff J.M., Alexander, Daniel C., 2003. Probabilistic Monte Carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. *Inf. Process. Med. Imag.* 18, 684–695 (Jul).
- Parker, Geoffrey J.M., Alexander, Daniel C., 2005. Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Philos. Trans. R. Soc. London, Ser. B Biol. Sci.* 360, 893–902 (May).
- Porter, R., Lemon, R.N., 1993. *Corticospinal Function and Voluntary Movement*. Oxford Univ. Press, Oxford, UK.
- Rushworth, M.F., Behrens, T.E., Johansen-Berg, H., 2005. Connection patterns distinguish 3 regions of human parietal cortex. *Cereb. Cortex* (Nov).
- Smith, Stephen M., Zhang, Yongyue, Jenkinson, Mark, Chen, Jacqueline, Matthews, P.M., Federico, Antonio, DeStefano, Nicola, 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage* 17, 479–489 (Sep.).
- Stieltjes, B., Kaufmann, W.E., van Zijl, P.C., Fredericksen, K., Pearlson, G.D., Solaiyappan, M., Mori, S., 2001. Diffusion tensor imaging and axonal tracking in the human brainstem. *NeuroImage* 14, 723–735.
- Tournier, J.-Donald, Calamante, Fernando, Gadian, David G., Connelly, Alan, 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23, 1176–1185 (Nov).
- Tuch, David S., 2004. Q-ball imaging. *Magn. Reson. Med.* 52, 1358–1372 (Dec).
- Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Wedeen, V.J., 2002. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magn. Reson. Med.* 48, 577–582.
- Tuch, David S., Reese, Timothy G., Wiegell, Mette R., Van Wedeen, J., 2003. Diffusion MRI of complex neural architecture. *Neuron* 40, 885–895 (Dec).
- Tuch, David S., Wisco, Jonathan J., Khachaturian, Mark H., Ekstrom, Leeland B., Kotter, Rolf, Vanduffel, Wim, 2005. Q-ball imaging of macaque white matter architecture. *Philos. Trans. R. Soc. London, Ser. B Biol. Sci.* 360, 869–879 (May).
- Van Wedeen, J., Hagmann, Patric, Tseng, Wen-Yih Isaac, Reese, Timothy G., Weisskoff, Robert M., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magn. Reson. Med.* 54, 1377–1386 (Dec.).
- Woolrich, M., Smith, S., 2001. Hierarchical fully bayesian spatiotemporal analysis of fmri data. *Proceedings of the 7th Annual Conference on Functional Mapping of the Human Brain*. Wiley Interscience, Brighton, UK.