A Probabilistic Model-based Approach to Consistent White Matter Tract Segmentation

Jonathan D. Clayden, Amos J. Storkey, and Mark E. Bastin*

Abstract— Since the invention of diffusion MRI, currently the only established method for studying white matter connectivity in a clinical environment, there has been a great deal of interest in the effects of various pathologies on the connectivity of the brain. As methods for *in vivo* tractography have been developed it has become possible to track and segment specific white matter structures of interest for particular study. However, the consistency and reproducibility of tractography-based segmentation remain limited, and attempts to improve them have thus far typically involved the imposition of strong constraints on the tract reconstruction process itself. In this work we take a different approach, developing a formal probabilistic model for the relationships between comparable tracts in different scans, and then using it to choose a tract, *a posteriori*, which best matches a predefined reference tract for the structure of interest. We demonstrate that this method is able to significantly improve segmentation consistency without directly constraining the tractography algorithm.

Index Terms— magnetic resonance imaging, diffusion, brain, white matter, tractography, segmentation, model, probabilistic

I. INTRODUCTION

THE ADVENT and development of diffusion magnetic resonance imaging (dMRI), and the more recent establishment of methods for modelling water self-diffusion in the brain, starting with the tensor model [1], has made it possible for the first time to study noninvasively the white matter structures that make up the brain's circuitry [2]. Tractography— the algorithmic reconstruction of white matter pathways— has become a major application of the dMRI technique; and segmentation of these pathways, or tracts, has in turn become a significant use for tractography [3]. Unlike more established region of interest (ROI) methods, the tractographic approach to segmentation allows one to semiautomatically "select" white matter structures of arbitrarily complex shape which would be extremely difficult to identify by hand with much consistency.

An element of subjectivity typically remains in tractographic segmentations, however, due to the near-ubiquitous use of "seed points". Although many tractography algorithms and supporting models of diffusion have now been developed

(e.g. [4]–[13]), the principle is broadly common: it consists of integrating the local white matter orientation information provided by a dMRI data set across the brain [3]. However, one must choose a location from which to begin the reconstruction process—the seed point—and the final segmentation is often strongly dependent on the exact location of this point. Hence, the process of seeding at a single hand-chosen point has significant reproducibility disadvantages.

The alternative is broadly to seed at a number of points, after which one can simply combine together the output from all the individual seed points (e.g. [14]) or constrain the results to include only those tract paths that pass through two or more ROIs [15]. The latter approach has been shown to have reproducibility advantages over less constrained methods [16], but this kind of additive seeding has the converse disadvantage of losing the inherent specificity of single seed point tractography. Moreover, it represents a very strong constraint on the reconstruction process that is quite independent of the data. Another option is to seed throughout the entire brain and then segment individual structures by applying clustering techniques to the resulting set of streamlines [17]–[20]. However, such an approach would not be directly applicable to probabilistic tract representations.

We have recently demonstrated proof of concept for an approach to automated seed point placement in which a set of "candidate" seed points in some neighbourhood of dMRI space are each used to generate a tract, and the single seed point is chosen whose corresponding tract matches best to a predefined reference tract [21]. In that case, each candidate seed point is effectively treated as a hypothesis, and the hypothesis with the best evidence to support it—in terms of tract similarity— is chosen as the final segmentation for the tract of interest. In order to assess tract similarity, we presented a heuristic similarity measure which aimed to capture both shape and length similarities of the two tracts in question.

In the present work we take this approach further, developing a formal probabilistic model for the shape and length relationships between comparable tracts, which we represent using B-splines. We fit the model's parameters using supervised learning, and then use it to find the candidate tract that best matches a reference tract amongst a novel data set.

II. THEORY

In the tract matching model described below, we work on the principle that two tracts are equivalent if they have identical shape and length. Thus, given a reference tract, the topology of some other tract—rather than any of its diffusion properties— is the characteristic that determines whether it matches the

reference tract or not. We assume that the shape and length of a matching tract are well predicted by the shape and length of the reference tract, and consequently that small deviations from the properties of the reference tract are far more likely than very large ones. Our aim is to encode these relationships probabilistically.

A. Tract representation

Ideally, any model for matching tracts should not depend explicitly on the tractography algorithm being used to generate them. Since different tractography algorithms produce various different forms of output, it is necessary to reduce the inhomogeneity in tract representations as far as possible, while still capturing the topological properties that are of interest for assessing tract similarity.

Given a single seed point, some tract reconstructions consist of a single line running through that point [4], [10], while others produce a number of sample streamlines with the seed point in common [5], [8]. In either case, the process for generating streamlines is typically to choose a local tract orientation (starting at the seed point), move a short distance in the corresponding direction, and repeat until some termination criterion is met. This process has to be performed twice to reconstruct the complete streamline, since all dMRI derived tract orientation information is directionally nonspecific. As a result, each streamline can be conceptually split at the seed point into two sets of points, representing what we will refer to as the "left" and "right" substreamlines. Each streamline can thus be said to have a "left length", N_1—the number of points on its left side, excluding the seed point itself—and a "right length", N_2. Here the names left and right are used for convenience only, and have no strict significance.

In order to be able to model single streamlines and distributions of probabilistic streamlines in the same way, we must first find a single line, in the latter case, which epitomises the shape of the whole set of lines. We do this by calculating a median streamline whose left and right lengths, $\tilde{N}_1$ and $\tilde{N}_2$, are the ξ-quantiles of the individual streamline lengths, where ξ is a parameter to be chosen. (For $\xi = 0.9$, for example, distal spatial information would be discarded from the longest 10% of streamlines.) Then, beginning at the seed point and moving outwards in each direction in turn, the x, y and z components of the median point location are calculated at each step from all unterminated streamlines. The resultant set of median points is a single line tract representation $r = \{\mathbf{x}_i\}$, where $i \in \{-\tilde{N}_1, -\tilde{N}_1 + 1, \ldots, \tilde{N}_2 - 1, \tilde{N}_2\}$ and the point $\mathbf{x}_0$ is the seed point.

Unlike in the individual streamlines, where each step is of a fixed distance in the native space of the subject, the median line, as a composite streamline, is not in general made up of equally spaced points. However, the real world length of this piecewise linear median line can be easily calculated by summing the actual point spacings.

Finally, the path of the median tract is represented in terms of a three dimensional cubic B-spline curve [22], parameterised by the distance along the median line, t. For any uniform cubic B-spline with K knots in total, there are

Fig. 1. Graphical representation of a full set of probabilistic streamlines (grey), the median line and B-spline knot points (black), here projected into a plane normal to the superior–inferior (z) axis. $\xi = 0.99$. The seed point is indicated with an arrow.

$\kappa = K - 8$ equally spaced internal knots; and in this case they are arranged so that one of them falls on the seed point. The final tract parameterisation then becomes

$$\mathbf{r}(t) = \sum_{j=0}^{K-3} \mathbf{p}_j B_j(t) \, , \tag{1}$$

where B_j are the cubic B-spline basis functions and $\mathbf{p}_j$ are the corresponding control point vectors.

The free parameter, K, is not chosen directly. Instead, the control point coefficients are calculated for the reference tract data using a model with one internal knot (i.e. $\kappa = 1$, $K = 9$), and the residuals, $\boldsymbol{\rho}_i$, at each point, i, on the median line are used to calculate the residual standard error, according to

$$\mathbf{E}_\kappa = \sqrt{\frac{\sum_i \boldsymbol{\rho}_i{}^2}{\tilde{N}_1 + \tilde{N}_2 - \kappa - 3}} \, . \tag{2}$$

(The denominator of Eq. (2) represents the residual degrees of freedom, which is affected by the number of points on the median line and the number of internal knots.) The number of knots is then incremented and the residual standard error recalculated until the mean of the three components of $\mathbf{E}_\kappa$ is less than some threshold value, η. The knot separation distance for this fit is then fixed for each candidate tract, so the number of knot points in each case depends on—and is uniquely determined by—the length of each median line.

Fig. 1 demonstrates the process described above. A set of 5000 probabilistic streamlines is shown in grey: these represent all of the information about the connectivity distribution provided by the tractography algorithm for a single seed point. The black line represents the median, and the black filled circles represent the B-spline knot points in the final tract parameterisation. Note that, although we favour methods that produce a distribution of streamlines due to the greater amount of information they provide about spatial uncertainty, if a tractography algorithm had been used that generates only a single

streamline for each seed point, then calculating the median line would be unnecessary, but the B-spline parameterisation would still be valid. This parameterisation is used in order to reduce the dimensionality of the data and emphasise topological tract features at a scale that is not determined by voxel dimensions.

B. Matching model

With the reference and candidate tracts represented as a series of B-spline knot points as described in the previous section, we can now define a model for the relationships between tracts. We consider a finite set of candidate tracts, among which there is assumed to be a single tract that best matches the reference tract, which has been chosen in advance. We introduce a variable, m, which can take any value in $\{1, 2, \ldots, N\}$, where N is the number of candidate tracts in the set, to indicate that the corresponding tract is the best match. Given a set of data, D, describing the group of candidate tracts, we wish to establish a model for the distribution $P(m \,|\, D)$; and hence to find the most likely value of m.

For a tract, i, which has L_1 internal knot points on its left side and L_2 knots on its right side (excluding the seed point in each case), consider the vectors that link successive knots together such that they are always directed away from the seed point. We denote these vectors $\mathbf{v}_u^i$, where u indexes location in the tract such that it is negative on the left side and positive on the right side. The cosine of the angle between a contiguous pair of these vectors is given by

$$c_u^i = \cos\theta_u^i = \frac{\mathbf{v}_u^i \cdot \mathbf{w}_u^i}{\|\mathbf{v}_u^i\|\,\|\mathbf{w}_u^i\|} \qquad (3)$$

where $\|\cdot\|$ is the usual Euclidean norm and

$$\mathbf{w}_u^i = \begin{cases} \mathbf{v}_{u+1}^i & \text{if } u < -1 \\ -\mathbf{v}_{-u}^i & \text{if } u = \pm 1 \\ \mathbf{v}_{u-1}^i & \text{if } u > 1. \end{cases} \qquad (4)$$

These *continuity* angles give an indication of the local curvature of the tract.

Similarly, by introducing the notation $\mathbf{v}_u^*$ for the uth vector in the reference tract, we can describe another cosine value,

$$s_u^i = \cos\phi_u^i = \frac{\mathbf{v}_u^i \cdot \mathbf{v}_u^*}{\|\mathbf{v}_u^i\|\,\|\mathbf{v}_u^*\|} \,, \qquad (5)$$

which indicates the local directional *similarity* between the reference and candidate tracts.

Fig. 2 illustrates, in two dimensions, the continuity angles, θ_u, and the similarity angles, ϕ_u. Since the cosine function is *a priori* uniform for random 3D vectors, we model the cosines of these angles—as described by Eqs (3) and (5)—rather than directly modelling the angles themselves.

The tract data that are relevant to our matching model are its continuity and similarity cosines and its left and right lengths: $\mathbf{d}^i = (L_1^i, L_2^i, \mathbf{c}^i, \mathbf{s}^i)$, where $\mathbf{c}^i = (c_u^i)$ and $\mathbf{s}^i = (s_u^i)$. The full data set, D, then contains all the $\mathbf{d}^i$ plus the left and right lengths of the reference tract, L_1^* and L_2^*. The principle of the model is that in regions where there is directionality information available from the reference tract, that information should provide the best predictor for the
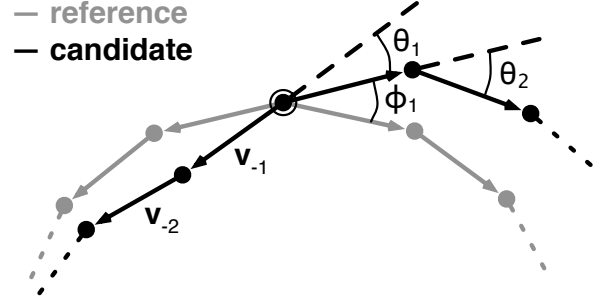

Fig. 2. Illustration of the different angles relevant to our model. Filled circles here represent successive knot points in the reference and candidate tracts. The ringed knot is the seed point, which is common to the two tracts.

direction of a matching candidate tract. If the candidate tract is longer than the reference tract, however, then in the region beyond the end of the reference, the only predictor of the tract's direction at any given step is its direction at the previous step. Hence, the full matching model is given by

$$\begin{aligned} P(m=i \,|\, D) \propto\; & P(L_1^i \,|\, L_1^*)\, P(L_2^i \,|\, L_2^*) \\ & \times \prod_{u=1}^{\check{L}_1^i} P(s_{-u}^i) \prod_{u=1}^{\check{L}_2^i} P(s_u^i) \\ & \times \prod_{u=\check{L}_1^i+1}^{L_1^i} P(c_{-u}^i) \prod_{u=\check{L}_2^i+1}^{L_2^i} P(c_u^i)\,, \quad (6) \end{aligned}$$

where $\check{L}_1^i = \min\{L_1^i, L_1^*\}$, and equivalently for $\check{L}_2^i$. The inclusion of the continuity cosine distributions expresses a preference for candidates that are not atypical in their curvature in regions unconstrained by the reference tract; it thus provides some assurance of "tract quality". It is implicitly assumed here that all unmatched tracts are equiprobable. The constant of proportionality in Eq. (6) is given by normalising over all values of i.

There are some constraints that can be applied to this model in order to reduce the number of parameters that need to be estimated. To this end, we assume that the curvature properties of tracts do not vary along their length, implying that all continuity cosines are drawn from a single distribution. We cannot, however, assume the same for the similarity cosines: Fig. 1 demonstrates that there is generally far more spatial uncertainty—as shown by the spread of the streamline set— near the ends of tracts than there is near the middle, so considerable local deviation from the reference tract can be expected near the ends of even well-matched candidate tracts. Hence, we make the weaker assumption that there is no inherent difference between the left and right sides of the tract, with distributions over similarity cosines varying only with distance from the seed point. That is,

$$\begin{aligned} P(c_u^i) = P(c_v^i) = P(c) & \qquad \forall u, v, i \\ P(s_u^i) = P(s_{-u}^i) = P(s_u) & \qquad \forall u > 0, i. \end{aligned} \qquad (7)$$

We must finally give specific forms for the distributions in Eq. (6). The length distributions are modelled as regularised

multinomial distributions, subject to a maximum length cutoff, while the cosine distributions are modelled as single-parameter beta distributions with a uniform regularisation component (see Appendix). That is,

$$P(c) = \varepsilon + (1 - \varepsilon)\,\alpha\,c^{\alpha-1} \,, \qquad (8)$$

and equivalently for each $P(s_u)$ distribution. The parameter ε is expected to be small, but since the beta distribution component will assign extremely tiny probabilities to large angles, it is needed to ensure that Eq. (6) does not grossly underestimate matching probabilities when larger angles do occur.

III. PARAMETER FITTING

Diffusion MRI data from six healthy volunteers were used for this study. The dMRI acquisition protocol has been described previously [21]. Briefly, it uses 51 non-collinear diffusion weighting gradient directions at a b-value of 1000 s mm^{-2}, plus 3 T$_2$-weighted volumes. In-plane resolution was 1.72×1.72 mm, and the slice thickness was 2.8 mm. Subjects were scanned up to three times. Each volume was preprocessed to remove skull data and eddy current induced artefacts using FMRIB Software Library tools (FMRIB, Oxford, UK; [23]).

For the purposes of this study, the white matter structures of interest were the corpus callosum splenium and corticospinal tract. All tracts were generated using the BED-POST/ProbTrack algorithm [5] with its default parameters. The result was a set of 5000 probabilistic streamlines for each tract, with a fixed separation distance of 0.5 mm between successive points. Median lines were then calculated using $\xi = 0.99$, and transformed into the space of the reference tract by using the FLIRT algorithm [24] to register together T$_2$-weighted ($b = 0$) volumes from each scan. Using a residual error threshold, η, of 0.1 mm, the B-spline parameterisation was calculated for the splenium reference tract, and all candidate tract splines were fitted using the resulting knot separation distance of 6.1 mm. If any pair of successive median line points were more than this distance apart, the median line was truncated to avoid creating multiple knots (which would result in discontinuities in the spline).

In addition to the reference, nine other splenium tracts were chosen by hand from different brain volumes to form a training set of matching tracts, and the parameters of the model pertaining to the length and similarity cosine distributions were fitted using maximum likelihood (see Appendix). Specifically, three splenium tracts were taken from subject 1, two from subject 2, two from subject 3, and one each from subjects 4 and 5. The reference tract was taken from a third scan of subject 2. No more than one training tract was taken from any given scan. The continuity cosine distribution, $P(c)$, was fitted from 50 tracts generated by seeding randomly in a single brain volume, subject to an anisotropy threshold used to ensure that each seed point was in white matter. This policy is appropriate given the assumption that the continuity properties of all tracts are broadly similar, and it has the significant advantage of increasing the quantity of training data available.
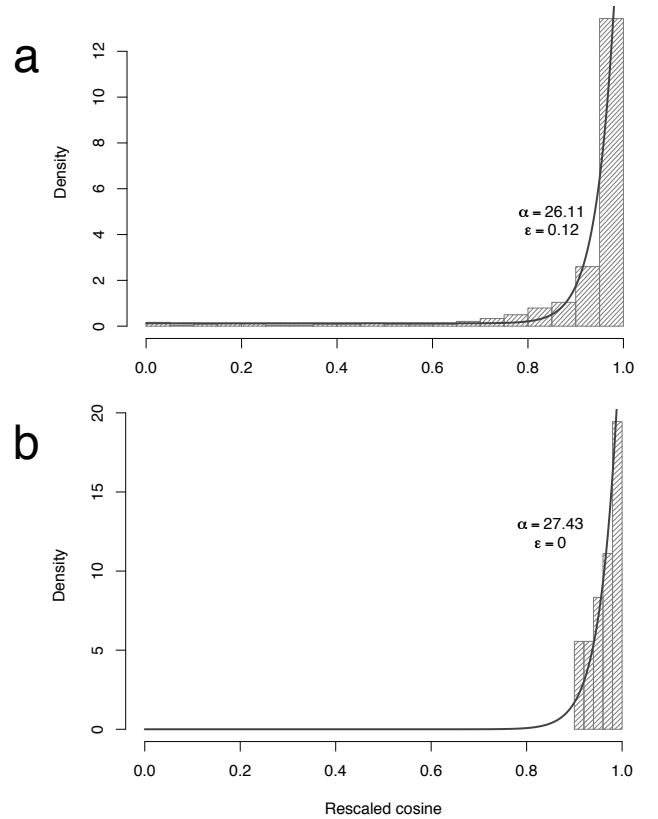

Fig. 3. Histograms of rescaled continuity cosines (a; $n = 962$) and similarity cosines for $u = 7$ (b; $n = 18$) from the splenium training data. The appropriate density functions from the model are overlaid.

Fig. 3 shows histograms of the cosine distributions, $P(c)$ and $P(s_u)$—the latter for a sample value of u. In (a), there are data from the full domain of cosine values, and the final estimate for ε reflects this. In (b), however, there are no cosine data below 0.9, and so the ε parameter has shrunk to zero. In fact, all of the similarity cosine distributions had $\varepsilon = 0$, although the α parameter—which affects the steepness of the right hand sides of the distributions—varied considerably, being 112.6 for $u = 1$ and only 6.1 for $u = 14$, the largest value of u for which a distribution was defined.

The whole process was applied in the same manner for the corticospinal tract, using an appropriate reference. The model parameters were retrained for this case, using a training set of five tracts.

IV. APPLICATION

Having used the training data to learn its parameters, the model described by Eq. (6) represents a way of assessing a set of novel tracts for their respective similarities to the reference tract. In order to create such a set, the seed point used to generate the reference tract was transferred to a new brain volume, from which no training data had been taken. Tractography was then performed for all points within a $7 \times 7 \times 7$ voxel region centred at this location, subject to

an anisotropy threshold, and each candidate seed point was processed as follows.

1) Run the tractography algorithm and recover a set of probabilistic streamlines.
2) Calculate the median line and transform it into the space of the reference tract as described above.
3) Using the knot spacing chosen for the reference tract, fit a cubic B-spline along the median line.
4) Calculate continuity and similarity angles for the inter-knot vectors, as depicted in Fig. 2.
5) Evaluate the right hand side of Eq. (6) using the length and angle distributions fitted from the training data.

This allows us to select the "best" seed point *a posteriori* by finding the starting location which generates the best-matching tract.

Using this size of neighbourhood, the runtime of the algorithm to choose a single tract is around four hours on a single-core Intel Xeon 3.2 GHz based workstation, the vast majority of which is spent running the tractography algorithm and collating the results. Fitting the B-spline model and evaluating the likelihood takes less than a second per candidate seed point. However, the process is presently implemented in R, an interpreted statistics and data analysis language [25], and so we expect that reimplementing the method in a compiled language such as C would provide significant speed benefits. Moreover, candidate tracts can be generated in parallel since they have no interdependence.

In order to test the robustness of the method to small differences in the reference tract, the corpus callosum reference was substituted for its equivalent taken from a different scan of the same subject (see Fig. 4). These two tracts do, of course, represent the same physical fasciculus, imaged in two consecutive scans. The model parameters were then recalculated for this alternative reference tract, and the experiment was repeated.

Since there is no normalisation or standardisation of matching probabilities between different sets of candidate tracts, these values are not directly comparable between data sets or reference tracts. They simply represent the probability of each candidate tract matching the given reference *relative* to the other candidates. There is no guarantee that the most likely match is a good match in any absolute sense. In order to provide an indication of absolute goodness-of-match, the log-ratio between the matching likelihood (the right hand side of Eq. (6)) of the best match and the matching likelihood of the reference tract to itself was calculated.

V. RESULTS

Fig. 4 shows the results of applying the model to tract—and hence seed point—selection. In this figure, all tracts are shown as maximum intensity projections; splenium tracts in a plane normal to the superior–inferior (z) axis, and corticospinal tracts normal to the left–right (x) axis. These perspectives are used because they show the two axes of greatest spatial variation and highlight the most common gross reconstruction inconsistencies in each case. Each tract is shown colour-coded according to the proportion of probabilistic streamlines that pass through each image voxel, thresholded at the 1% level.
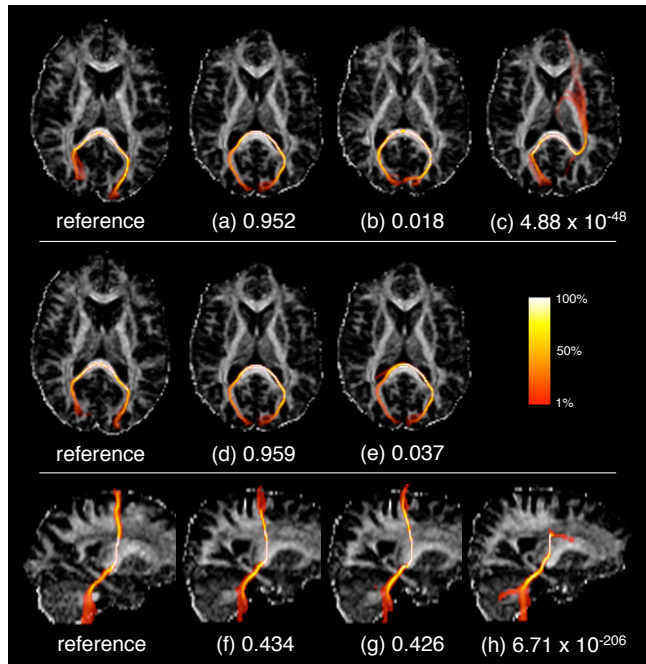

Fig. 4. The two most likely matches to the original (top row) and the alternative splenium reference tract (middle row), shown in axial projection with their associated matching probabilities. The tract generated from the neighbourhood centre point is shown with its matching probability (c), for comparison. Results for the corticospinal tract, in sagittal projection, are shown in the bottom row. It should be remembered that tracts (a)–(h) are taken from different subjects to the reference tracts. Colours represent the proportion of probabilistic streamlines passing through each voxel, as indicated by the colour bar.

(This threshold is approximately equivalent to the use of $\xi = 0.99$ above in calculating the median line.) The underlying greyscale image in each case is the anisotropy map slice in-plane with the seed point.

According to the model, tracts (a) and (b) are the two most likely matches to the reference tract adjacent to them. The point at the centre of the seeding neighbourhood generated tract (c), which is visually far less similar to the reference tract. Its matching probability is commensurately smaller, by many orders of magnitude, than those for (a) and (b). The candidate set contained 220 tracts in total, after thresholding on anisotropy.

For comparison, tracts (d) and (e) are the two best-matching tracts from the same neighbourhood, using the alternative reference tract. In this case the model parameters were re-learned, but the knot separation distance under this very similar reference tract was only slightly smaller than the old one, at 6.0 mm. Tracts (a) and (d) are in fact the same tract, so the most likely match is the same with both reference tracts.

Similarly, tracts (f) and (g) are much better matches to the corticospinal reference tract than the tract generated from the centre seed, (h). Once again the matching probabilities reflect this.

Fig. 5 shows the results of calculating log-ratios using the original reference tract for the splenium. The more negative this log-ratio, the less good a fit is compared to the "bench-mark" of the reference tract itself.
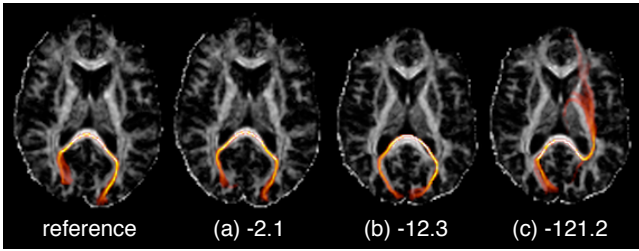
Fig. 5. Log-ratios between matching likelihoods of the tracts shown and the matching likelihood of the reference tract. The reference tract has a log-ratio of zero by definition; (a) is the alternative reference tract; (b) is the best match in the novel candidate set; and (c) is the tract generated from the neighbourhood centre point.

VI. DISCUSSION

In this work, we have presented a probabilistic model-based approach to white matter tract segmentation from dMRI data sets. We have described a tract similarity model, based on shape and length, which has no specific dependence on any particular tractography algorithm. We have demonstrated the use of the algorithm to select the best matching tract from a set of candidates, and described how a log-ratio can be used to quantify goodness-of-match in a manner that is comparable between data sets. The method has been shown to be apt for selecting matches for two example fasciculi of very different shapes. Taken together, this work represents a novel way of automatically selecting single seed points for robust segmentation of tracts of interest in groups.

The success of our approach does of course depend on it being *possible* to consistently segment a tract of interest with single seed points; and there is no guarantee that this will be possible for all scans or all tracts. However, with the recent development of tractography algorithms that can model more than one fibre orientation at each voxel (e.g. [7], [12], [13], [26]) it should become less likely for there to be no suitable match; and, ultimately, the log-ratio demonstrated in Fig. 5 can be used to reject tracts that match very badly.

We note that the training set for the splenium tract included tracts from different scans of the test subject. We do not regard this as problematic since interscan variability is as much of interest as intersubject variability, and since the reference tract itself was taken from a different subject. Nevertheless, in case this be seen as a source of bias, we avoided taking any training data for the corticospinal tract from the test subject.

Compared to the simpler methods of placing single seed points by hand or using image-registration-based transformation, our method offers advantages with respect to consistency and reproducibility. Figs 4(c) and 4(h) were generated by using registration to transfer the seed point from the reference image, and it can be seen that the results are highly dissimilar to the reference tracts, and so this approach is insufficient to achieve consistency in segmentation. By contrast, the best matches under our model—shown in Figs 4(a), 4(d) and 4(f) for the three test cases—much more closely resemble their respective reference tracts. Even using anatomical landmarks for guidance, placing seed points by hand is subjective and hence has limited reproducibility [27]. On the other hand,

reference tracts can be directly transferred between studies without modification; and since there is no need for observer interaction, presentation of an identical data set to the method described above will always yield an identical result.

The use of ROIs to constrain the paths that probabilistic streamlines may take [15], [16] is not precluded by our method. Indeed, a two-ROI constraint could be applied to ameliorate one of the limitations of our model: the median line cannot represent branches in the streamline set and therefore the model does not take them into account. Nevertheless, the similarity "preference" that our model imposes is a far weaker constraint than the two-ROI constraint, and since both use *a priori* knowledge about the tract of interest to affect a reconstruction process that should be primarily driven by the actual data, we would argue that the weaker constraint is preferable where possible.

The present model-based approach to assessing tract similarity also has advantages over the heuristic method described in [21]. The first benefit is a general matter of principle: explicitly describing a tract matching model and its assumptions makes the method more scrutable than otherwise. Secondly, and more substantially, the median line representation of a tract can undergo affine transformation without complications; whereas the previously used field representation of a tract cannot be transformed without creating interpolation issues. This is helpful because it allows us to easily correct for gross head size or rotation differences between the reference and candidate tracts using standard affine image registration—as we have done above. Thirdly, the results from our previous approach to tract matching were quite strongly affected by the particular nature of the reference tract, and had a very narrow dynamic range. By contrast, Fig. 4 demonstrates that two very similar reference tracts do produce comparable—although not identical—results under the current model, while the matching probabilities assigned to dissimilar candidate tracts vary by orders of magnitude. Tracts (a), (b), (d) and (e) all represent appropriate matches to either splenium reference tract, and the fact that the best match under the original reference tract was also the best match under the alternative reference, out of a set of more than 200 candidates, (whilst not conclusive evidence) does suggest a beneficial lack of sensitivity to small alterations in the reference tract.

Our model does have limitations, however. It has already been mentioned that the median line representation ignores branches in the original set of streamlines; and as a result, the model cannot discriminate against such tracts, which may be considered desirable. This, of course, will not be an issue in cases where the tractography algorithm produces a single streamline representation of a tract. Secondly, the nature of Eq. (6) is such that the reference tract itself does not have the highest possible matching likelihood, and so the log-ratio calculated in Fig. 5 could be positive for some tracts. Moreover, since there is very little training data available for the length distributions, and so they are heavily affected by their regularisation terms, they do not fully compensate for the likelihood-increasing effects of the continuity cosines in very long tracts. Additionally, of course, any limitations and sensitivities to data quality that the chosen tractography

algorithm may have will apply in turn to our method.

Consistent and reproducible segmentation of tracts is a prerequisite for clinical and neurological studies that use dMRI data sets and need to focus their attentions on specific white matter structures (e.g. [14]). As a result, consistency is our primary concern in this work, and we focus on this aim rather than attempting to segment the entire width of the tract in each brain. Tract segmentation may have a significant part to play in future developments in the theory of the connective architecture of the brain. A suitably rich model of the relationships between the topologies of comparable tracts in different subjects has the potential to be informative in its own right, as well as leading naturally, as we have described, to a method for choosing seed points for segmentation. In future work, alternative tract representations could be considered in order to handle branches in streamline sets; and the limitations on the current method due to the small size of the training set could be alleviated by using an unsupervised learning approach. There are also other possibilities for this kind of topological modelling: for example, probabilistic model selection could perhaps be used to classify tracts into "normal" and "distorted" groups in pathological scenarios where the topology of white matter structures is adversely affected.

APPENDIX

The distributions over cosines are modelled as a mixture of a beta distribution with a uniform regularisation component. The beta distribution has the general p.d.f.

$$P(x \,|\, \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \, x^{\alpha-1}(1-x)^{\beta-1} \,, \qquad (9)$$

where $\Gamma(\cdot)$ is the gamma function. However, since small angles are always assumed to be the most common, we can fix $\beta = 1$ in each case, leaving α and the mixture coefficient, ε, as the only unknown parameters. The distribution is then

$$P(x \,|\, \alpha, \varepsilon) = \varepsilon + (1 - \varepsilon) \, \alpha \, x^{\alpha-1} \,. \qquad (10)$$

It is defined only for $x \in [0, 1]$ (being zero for all other values of x), so we rescale our cosine values into this interval.

In order to find maximum likelihood estimates for α and ε given some data vector of rescaled cosine values, $\mathbf{x}$, we use a simple Expectation–Maximisation algorithm. Associated with each data value, x_j, is a latent variable, z_j, indicating whether the value came from the uniform distribution ($z_j = 0$) or the beta distribution ($z_j = 1$).

Given some starting estimates for the distribution parameters, $\hat{\alpha}$ and $\hat{\varepsilon}$, the E-step of the algorithm calculates

$$P(z_j = 0 \,|\, x_j) = \frac{\hat{\varepsilon}}{\hat{\varepsilon} + (1 - \hat{\varepsilon}) \, \hat{\alpha} \, x_j^{\hat{\alpha}-1}} \qquad (11)$$

and

$$P(z_j = 1 \,|\, x_j) = \frac{(1 - \hat{\varepsilon}) \, \hat{\alpha} \, x_j^{\hat{\alpha}-1}}{\hat{\varepsilon} + (1 - \hat{\varepsilon}) \, \hat{\alpha} \, x_j^{\hat{\alpha}-1}} \qquad (12)$$

for each value of j. The M-step then updates the parameter estimates according to

$$\hat{\alpha} = \frac{-\sum_j P(z_j = 1 \,|\, x_j)}{\sum_j P(z_j = 1 \,|\, x_j) \ln x_j} \qquad (13)$$

and

$$\hat{\varepsilon} = \frac{P(z_j = 0 \,|\, x_j)}{P(z_j = 0 \,|\, x_j) + P(z_j = 1 \,|\, x_j)} \,, \qquad (14)$$

and the algorithm repeats until convergence.

REFERENCES

[1] P. Basser, J. Mattiello, and D. Le Bihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *Journal of Magnetic Resonance, Series B*, vol. 103, no. 3, pp. 247–254, March 1994.

[2] D. Le Bihan, "Looking into the functional architecture of the brain with diffusion MRI," *Nature Reviews Neuroscience*, vol. 4, no. 6, pp. 469–480, June 2003.

[3] S. Mori and P. van Zijl, "Fiber tracking: principles and strategies - a technical review," *NMR in Biomedicine*, vol. 15, pp. 468–480, December 2002.

[4] P. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi, "In vivo fiber tractography using DT-MRI data," *Magnetic Resonance in Medicine*, vol. 44, no. 4, pp. 625–632, October 2000.

[5] T. Behrens, M. Woolrich, M. Jenkinson, H. Johansen-Berg, R. Nunes, S. Clare, P. Matthews, J. Brady, and S. Smith, "Characterization and propagation of uncertainty in diffusion-weighted MR imaging," *Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 1077–1088, November 2003.

[6] O. Friman, G. Farnebäck, and C.-F. Westin, "A Bayesian approach for stochastic white matter tractography," *IEEE Transactions on Medical Imaging*, vol. 25, no. 8, pp. 965–978, August 2006.

[7] K. Jansons and D. Alexander, "Persistent angular structure: new insights from diffusion magnetic resonance imaging data," *Inverse Problems*, vol. 19, no. 5, pp. 1031–1046, October 2003.

[8] D. Jones and C. Pierpaoli, "Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach," *Magnetic Resonance in Medicine*, vol. 53, no. 5, pp. 1143–1149, April 2005.

[9] M. Lazar and A. Alexander, "Bootstrap white matter tractography (BOOT-TRAC)," *NeuroImage*, vol. 24, no. 2, pp. 524–532, January 2005.

[10] S. Mori, B. Crain, V. Chacko, and P. van Zijl, "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging," *Annals of Neurology*, vol. 45, no. 2, pp. 265–269, February 1999.

[11] G. Parker, K. Stephan, G. Barker, J. Rowe, D. MacManus, C. Wheeler-Kingshott, O. Ciccarelli, R. Passingham, R. Spinks, R. Lemon, and R. Turner, "Initial demonstration of in vivo tracing of axonal projections in the macaque brain and comparison with the human brain using diffusion tensor imaging and fast marching tractography," *NeuroImage*, vol. 15, no. 4, pp. 797–809, April 2002.

[12] J.-D. Tournier, F. Calamante, D. Gadian, and A. Connelly, "Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution," *NeuroImage*, vol. 23, no. 3, pp. 1176–1185, November 2004.

[13] D. Tuch, T. Reese, M. Wiegell, N. Makris, J. Belliveau, and V. Wedeen, "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," *Magnetic Resonance in Medicine*, vol. 48, no. 4, pp. 577–582, October 2002.

[14] R. Kanaan, S. Shergill, G. Barker, M. Catani, V. Ng, R. Howard, P. McGuire, and D. Jones, "Tract-specific anisotropy measurements in diffusion tensor imaging," *Psychiatry Research: Neuroimaging*, vol. 146, no. 1, pp. 73–82, January 2006.

[15] T. Conturo, N. Lori, T. Cull, E. Akbudak, A. Snyder, J. Shimony, R. McKinstry, H. Burton, and M. Raichle, "Tracking neuronal fiber pathways in the living human brain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 18, pp. 10 422–10 427, August 1999.

[16] E. Heiervang, T. Behrens, C. Mackay, M. Robson, and H. Johansen-Berg, "Between session reproducibility and between subject variability of diffusion MR and tractography measures," *NeuroImage*, vol. 33, no. 3, pp. 867–877, November 2006.

[17] A. Brun, H. Knutsson, H.-J. Park, M. Shenton, and C.-F. Westin, "Clustering fiber traces using normalized cuts," in *Proceedings of the 7th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2004, pp. 368–375.

[18] I. Corouge, P. Fletcher, S. Joshi, S. Gouttard, and G. Gerig, "Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis," *Medical Image Analysis*, vol. 10, no. 5, pp. 786–798, October 2006.

[19] M. Maddah, A. Mewes, S. Haker, W. Grimson, and S. Warfield, "Automated atlas-based clustering of white matter fiber tracts from DTMRI," in *Proceedings of the 8th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2005, pp. 188–195.

[20] L. O'Donnell and C.-F. Westin, "White matter tract clustering and correspondence in populations," in *Proceedings of the 8th International Conference on Medical ImageComputing and Computer Assisted Intervention (MICCAI)*, 2005, pp. 140–147.

[21] J. Clayden, M. Bastin, and A. Storkey, "Improved segmentation reproducibility in group tractography using a quantitative tract similarity measure," *NeuroImage*, vol. 33, no. 2, pp. 482–492, November 2006.

[22] C. de Boor, *A practical guide to splines*, ser. Applied Mathematical Sciences Series. Springer-Verlag, 1978, no. 27.

[23] S. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, November 2002.

[24] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis*, vol. 5, no. 2, pp. 143–156, June 2001.

[25] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, 2007.

[26] T. Behrens, H. Johansen-Berg, S. Jbabdi, M. Rushworth, and M. Woolrich, "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?" *NeuroImage*, vol. 34, no. 1, pp. 144–155, January 2007.

[27] O. Ciccarelli, G. Parker, A. Toosy, C. Wheeler-Kingshott, G. Barker, P. Boulby, D. Miller, and A. Thompson, "From diffusion tractography to quantitative white matter tract measures: a reproducibility study," *NeuroImage*, vol. 18, no. 2, pp. 348–359, February 2003.