

# Probabilistic combination of tractography data from multiple seed points for white matter segmentation

J. D. Clayden<sup>1</sup>, A. J. Storkey<sup>2</sup>, and M. E. Bastin<sup>3</sup>

<sup>1</sup>Institute of Child Health, University College London, London, United Kingdom, <sup>2</sup>School of Informatics, University of Edinburgh, Edinburgh, United Kingdom, <sup>3</sup>Medical and Radiological Sciences (Medical Physics), University of Edinburgh, Edinburgh, United Kingdom

## Introduction

Segmentation of specific tissue regions from magnetic resonance images for comparative analysis necessitates a tradeoff between maximizing coverage of the structure of interest and minimizing the inclusion of other tissues. For the case of white matter tracts, a reasonable approach is to focus on a “core” of the tract where partial volume effects are likely to be minimal. This is the approach taken by, for example, the tract-based spatial statistics (TBSS) method [1]. On the other hand, it is sometimes desirable to produce a more inclusive segmentation, to counter noise problems.

Neighborhood tractography (NT) [2] selects a segmentation for a white matter tract of interest from a number of candidates generated by a tractography algorithm. The single tract thus selected is generated from a single seed voxel and is thus highly specific, tending to produce quite a narrow segmentation of the structure, as with TBSS. When probability distributions are used to find a suitable segmentation from among the candidates [3], this tract corresponds to the maximum likelihood (ML) seed voxel. In this work we describe a natural alternative use for the probability distribution over candidate seeding voxels, to produce a weighted average composite tract which results in a more inclusive segmentation than with the ML method.

## Methods

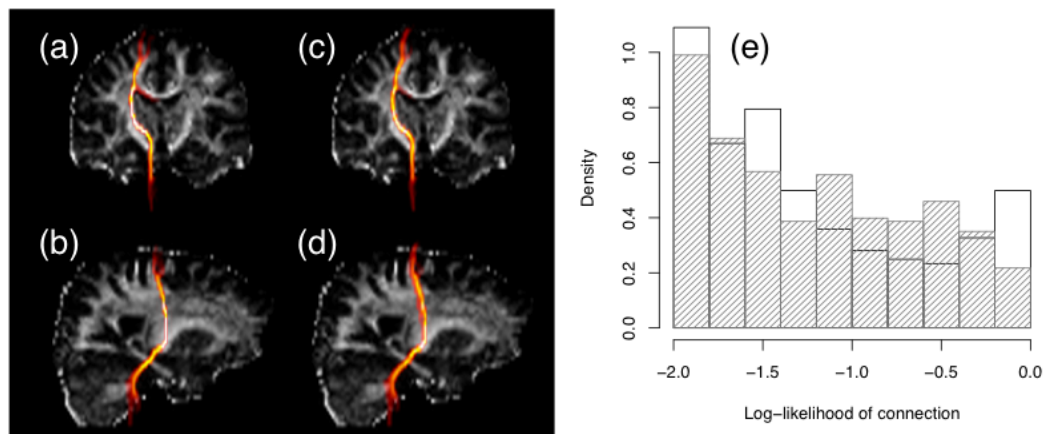
The data used for this study were acquired from a healthy 31 year old female volunteer, using a spin-echo echo-planar diffusion MRI protocol consisting of 51 noncollinear diffusion sensitizing gradient directions at 1.5 T, using a  $b$ -value of  $1000 \text{ s mm}^{-2}$ . Three T2-weighted volumes were also acquired. In-plane resolution was  $1.72 \times 1.72 \text{ mm}$ , and slice thickness was  $2.8 \text{ mm}$ . The data were processed to correct for eddy current induced distortions and extract the brain, using FSL tools (FMRIB, Oxford, UK). The BEDPOST/ProbTrack algorithm [4] was used for tractography, which produces quantitative output in the form of the likelihoods of a path connecting the seed voxel with each voxel in the brain. We represent this mathematically with a three-dimensional scalar field,  $\phi(\mathbf{x})$ , where  $\mathbf{x}$  denotes location.

Using the probabilistic NT method described in [3], we established a distribution  $P(m = i | D)$ , which encapsulates the probability that each candidate tract (indexed by  $i$ ) represents the best match,  $m$ , to a reference representing the corticospinal tract, given the data,  $D$ . The ML tract is then identical to the field  $\phi_i(\mathbf{x})$ , for the value of  $i$  where  $P(m = i)$  is maximal. By contrast, we can create a composite tract that is formed by calculating

$$\hat{\phi}(\mathbf{x}) = \sum_{i=1}^N P(m = i | D) \phi_i(\mathbf{x}), \quad (1)$$

which represents the statistical expectation of the voxel data with respect to the matching distribution.  $N$  is the number of candidate tracts. In practice, only seed voxels with matching probabilities of greater than 0.01 were incorporated into the results, in order to reduce computation time.

**Fig. 1:** Maximum likelihood corticospinal tract images in coronal (a) and sagittal (b) projections, and equivalent images for the weighted average tract data (c,d), all thresholded at the 1% level. Histograms of connection log-likelihoods for the maximum likelihood (plain) and weighted average (shaded) images are also shown (e).



## Results

Fig. 1 shows the the maximum likelihood and weighted average tract data for the subject's corticospinal tract. The general appearances and trajectories of the two segmentations appear very similar, but the composite tract is more inclusive, and therefore spatially broader. After thresholding at 1% there are 414 nonzero voxels in the composite tract, as opposed to 321 in the ML version. Moreover, the distribution of connection likelihoods is markedly altered in the weighted average tract, as shown by Fig. 1(e).

## Discussion

Since the composite tract shown above is wider than the ML one, it gives more complete coverage of the voxels that are likely to represent the physical corticospinal tract, lending greater power to any downstream statistical comparisons of tract-averaged anisotropy. In addition, the composite tract's use of multiple seed points reduces the tendency for very high connection probabilities to occur near the location of the seed. It can be seen in Fig. 1(e) that the general trend in both tracts is for the larger connection likelihoods to occur less frequently, but in the ML tract—shown with unshaded bars—there is a significant upturn at the very top end of the range, representing an uncharacteristically large number of voxels that are very likely to be connected to the seed point. These are a direct result of the high directional certainty of streamlines near to the seed point, and are heavily seed point dependent. The average tract, on the other hand, incorporates data from several seed points and is therefore less affected by this issue. It can be seen from the histograms that the downward trend continues over the whole range of voxel values in this case. (Only 7% of suprathreshold connection likelihood values are greater than 0.5 in the average tract, against 13% in the ML tract.)

In conclusion, we have shown that probabilistic neighborhood tractography can be used to create both relatively exclusive (maximum likelihood) and inclusive (weighted average) segmentations of white matter structures, improving its flexibility for applications. The latter approach is more similar to commonly applied region-of-interest seeding strategies, but it uses the NT matching distribution to maintain high confidence in the segmentation.

**References:** [1] Smith et al., *NeuroImage* 31(4):1487–1505; [2] Clayden et al., *NeuroImage* 33(2):482–492; [3] Clayden et al., *IEEE Transactions on Medical Imaging* 26(11):1555–1561; [4] Behrens et al., *Magnetic Resonance in Medicine* 50(5):1077–1088.