ABSTRACT
We demonstrate how structural parcellation can be implemented using shortest-path tractography, thereby addressing some of the shortcomings of the conventional approaches. In particular, our algorithm quantifies, via $p$-values, the confidence that a voxel in the parcellated region is connected to each cortical target region. Calculation of these statistical measures is derived from a rank-based test on the histogram of tract-based scores from all the shortest-paths found between the seed voxel and each voxel within the target region. Using data from the Human Connectome Project, we show that parcellation of the thalamus results in $p$-value maps that are spatially coherent across subjects. Whilst some agreement to a hard segmentation as in previous studies is observed, the soft segmentation exhibits better stability for voxels connected to multiple target regions.

1. INTRODUCTION
Structural parcellation [1, 2, 3, 4, 5, 6] is the data-driven segmentation of a source region defined by structural connectivity to a set of pre-defined target regions of interest (ROIs). We propose a parcellation algorithm which statistically quantifies the degree of connectivity to each target ROI for each voxel in the region to be parcellated. In contrast to classical majority voting methods [7], our method outputs a soft segmentation based on these confidence measures. We demonstrate the approach on 5 subjects from the Human Connectome Project [8].

Our approach is based on shortest-path tractography (SPT) [9] and overcomes all of the following problems, which remain unsolved by traditional approaches to structural parcellation:

a) Traditional fibre tracking methods start at a source point and randomly walk in the most likely direction as defined by a fibre orientation function (fODF). These methods may have a hard time retrieving connections from a voxel $v$ to a specific cortical region $C$ if the voxel $v$ is also connected to other regions, whose connections are easier to track. In particular, fibre tracking has a bias towards finding connections between nearby regions and will typically find more connections to large target regions than to small ones, even if this is not supported by the data [10, 11].

b) Traditional structural parcellation assumes that every voxel in the region to be segmented (in our case, the thalamus) is connected to one of the target ROIs. However, this is not always the case. Moreover, there may not be enough signal in the data to support such an assignment even if the connections exist. This is especially true for inside the thalamus, where fibres are usually hard to resolve.

c) Traditional structural parcellation performs voting to obtain a hard segmentation of the region to be segmented based on connections to the cortical ROIs. However, this is suboptimal for several reasons [2]: First, if a source voxel is not physically connected to either of the target regions, the found connections are based only on noise and may therefore be very unstable. Second, both anatomically and because of partial volume effects, many voxels are with high probability connected to multiple cortical regions. Since a hard segmentation is not able to model this, it may again, lead to unstable results.

Shortest-path tractography finds the most likely trajectory for a fibre connecting any two locations in the brain. This formulation lends itself well to structural parcellation, which precisely seeks connections between two pre-specified regions of interest: The source region to be segmented (in this paper, the thalamus) and the cortical target region, one at a time. Since SPT will always find a most likely fibre connecting any two voxels, it avoids problem a): Connections will be found both to small regions and between regions that are any distance apart.

However, this property of SPT also introduces a new problem: A connection will always be found, even if it is not there. Many SPT methods assign a score to the found paths which can be used to threshold unlikely paths, but such a threshold will reintroduce the biases from problem a). In this paper, we therefore propose an alternative approach which, assuming some thalamic voxels are physically connected to each target region, aims to define these as statistically more sig-
significant than those which are not. This is obtained through a statistical test over source region voxels.

Inspired by Kasenburg et al. [12] we quantify the confidence with which a voxel is significantly connected to each of the target regions, by assigning a p-value to each source voxel. This also solves problem b), as we allow voxels within the source region to have a p-value close to 1 for every single target region, i.e. not being strongly connected to any target region. Using the p-values as confidence maps, we obtain a soft structural parcellation of the thalamus, which allows for overlap between the segments corresponding to different target ROIs. In this way, we also solve problem c).

2. METHODOLOGY

Shortest-path tractography (SPT). In SPT [9], the diffusion weighted image (DWI) is turned into a graph where all gray- and white matter voxels from the brain are nodes. Edges are formed between pairs of voxels in a $3 \times 3 \times 3$ neighbourhood, whenever at least one of the voxels is white matter. Each edge $\vec{e}$ pointing out of a voxel $v$ is given a weight reflecting the probability of a fibre bundle tangential to $\vec{e}$. This probability is defined by integrating the fODF estimated for the voxel $v$ over the set of directions out of $v$ which is closer to $\vec{e}$ than to any other edge pointing out of $v$. The integral is estimated by sampling. To obtain an undirected graph, edge weights are averaged over their start and end nodes, obtaining probabilistic edge weights $p(e)$.

From this weighted graph, tractography between two points $v$ and $w$ is phrased as finding the most probable path from $v$ to $w$ in the graph, that is, the path maximizing $p(\pi) = \prod_{i=1}^{n} p([v_{i-1}, v_i])$, where $\pi = [v_0, v_1, \ldots, v_n]$ is a path in the brain graph from $v_0 = v$ to $v_n = w$, and $[v_{i-1}, v_i]$ is the edge connecting $v_{i-1}$ and $v_i$. Most probable paths are computed using Dijkstra’s shortest path algorithm after log-transforming the weights into edge lengths $l(e) = -\log(p(e))$.

Significance of SPT tracts. The main disadvantage of SPT is that for any two brain voxels $v$ and $w$, SPT will find a path connecting them, whether the data really supports this or not. This can be alleviated by thresholding the path-length corrected path probability score $s(\pi) = \left[ p(\pi) \right] ^{1/4}$, where $|\pi|$ is the number of nodes on $\pi$. However, this requires a choice of a threshold. A test for the statistical significance of SPT tracts would therefore be desirable. In this paper, we present a statistical test for the significance of connectivity from a source voxel $v \in R_1$ to a target region $R_2$.

Assumptions. We seek a segmentation of a given source region $R_1$ (the thalamus) defined by its connection to a set of cortical target regions. Let us focus on one target region $R_2$. Consider all paths found by SPT from a voxel $v$ to any voxel $w$ in the target region $R_2$. We assume that the majority of voxels in $R_1$ are not physically connected to the target region $R_2$ and that the scores of SPT paths found from such $v_{unc} \in R_1$ to voxels $w \in R_2$ describe noise. This noise may be specific to the regions $R_1$ and $R_2$, but we assume that it is independent of the source voxel $v_{unc}$. We also assume that voxels $v_{con} \in R_1$ that actually are physically connected to region $R_2$ show more high scoring SPT paths than expected for $v_{unc}$. Thus, their score distribution should be skewed towards the right, as the distribution is a mixture of noise from the target voxels $w$ in $R_2$ that are not connected to $v_{con}$, and high-scores from those target voxels $w$ which are connected to $v_{con}$.

Histograms and cumulative histograms. For each source voxel $v \in R_1$ we extract a histogram $H_v$ of scores corresponding to SPT paths from $v$ to any $w \in R_2$ as follows: Scores are divided into $N$ bins and the number $H_v(i)$ of scores falling into bin $i$ is counted for $i = 1, \ldots, N$. Examples of these histograms are shown in the top row of Fig. 1 together with the average histogram over all source voxels. Some histograms (top right in Fig. 1) indeed represent a distribution of scores that is more skewed to the right compared to the average histogram, suggesting that the corresponding voxel is physically connected to the target region.

Such raw histograms are, however, very jagged and noisy, therefore they are rarely used for statistical testing. Instead, one usually transforms them into cumulative normalized histograms (also known as empirical cumulative distribution functions) prior to testing without losing information. The cumulative normalized histogram of voxel $v$ is given by a vector $(C_v(1), \ldots, C_v(N))$, where

$$C_v(i) = \frac{\sum_{j=1}^{i} H_v(j)}{\text{norm}(H)}, \quad \text{norm}(H) = \sum_{i=1}^{N} H_v(i). \quad (1)$$

Assigning a p-value to histograms. To quantify if and how much the histogram of voxel $v$ deviates from the null distribution, we compare it to a simulated sample of $s$ histograms representing noise. When a histogram is skewed to the right, the corresponding cumulative histogram lies below a typical cumulative histograms representing the null distribution (see the right of Fig. 1).

Inspired by the envelope rank tests of Myllymäki et al. [13], we present a one-sided rank based test for the cumulative histograms. Our test computes a $p$-value for each source voxel $v$ given by

$$p = \frac{1}{s+1} \sum_{k=1}^{s+1} 1(\text{rank}_k \leq \text{rank}_1), \quad (2)$$

where $1(x)$ returns 1 if $x$ is true and 0 otherwise, and the rank$_k$ is the average rank from below over all bins of the $k$th sample from the null distribution. That is,

$$\text{rank}_k = \frac{1}{N} \sum_{j=1}^{N} \#\{k' = 1, \ldots, s+1 | C_{k'}(j) < C_k(j)\}, \quad (3)$$
where $k = 1$ is the index of the observed histogram. The rank thus measures how skewed the corresponding original histogram is to the right, compared to the other histograms in the sample of $s + 1$ histograms. The original test by Myllymäki et al. [13] was based on minimum ranks rather than the average rank, which led to an envelope interpretation of the test. For our application, the mean rank provides additional stability, but lacks the envelope interpretation.

**Drawing from the null distribution of cumulative histograms.** In order to define a statistical test for significantly connected source voxels, we need to be able to draw samples from the null distribution. Under the assumption that the majority of source voxels are not physically connected to the target region, this could be solved by drawing entire histograms from the population. However this could lead to an unnecessarily conservative test, since we risk to draw histograms from voxels that actually are physically connected.

Instead, we propose bootstrapping the histograms as follows: To draw a cumulative histogram $C$ from the null distribution, we first draw a histogram $H$ from the null distribution of score histograms and then obtain a cumulative sample $C$ from $H$ as in (1). The sample $H$ is drawn by randomly drawing a bin value $H(i)$ from the $i^{th}$ bin values in the entire population for every bin $i$ in $H$. This gives a histogram $H$ whose bin values $H(i)$ are $i^{th}$ bin values drawn from different source voxels $v$.

### 3. EXPERIMENTS

We used the preprocessed DWIs [14, 15] from 5 subjects (Q3 release) of the Human Connectome Project (HCP) data [8, 16]. Fibre orientation distribution functions (fODFs) were computed for every voxel using constrained spherical deconvolution [17] with order 8 implemented in the DiPy [18] package. The graph required for SPT was constructed from the given fODFs. The voxelwise diffusion parameters necessary for probabilistic tractography were generated using FSL’s BedpostX [19] with the zeppelin model and 3 fibres per voxel.

The thalamic source region was extracted from the MNI atlas [20] provided in FSL [21]. Target regions were extracted both from the MNI atlas and the Juelich atlas [22] similar to Behrens et al. [7]: prefrontal/temporal zone (frontal and temporal lobe from the MNI atlas), motor zone (primary motor cortex and premotor cortex from the Juelich atlas), somatosensory zone (primary somatosensory cortex from the Juelich atlas) and parieto-occipital zone (occipital and parietal lobe from the MNI atlas). The atlas in MNI space was warped into the respective subject-specific spaces using the warps provided in the HCP data.

For every source voxel, SPT was performed to obtain the most likely paths to all voxels in the given target region. Path scores were binned into a histogram with 1000 bins in the range from 0 to 1. Histograms for all source voxels were analysed for each target region separately as described in Section 2. The corresponding $p$-value for each voxel is shown on the left of Fig. 2. In addition, each voxel was classified as being connected to one of the target regions following the hard segmentation of Behrens et al. [7]. In brief, after probabilistic tractography using FSL’s protrackx [23] (5000 samples, 0.5 mm step length, maximum inter-step curvature $80^\circ$) to each of the target regions, the largest of the four resultant connectivity measures determined each voxel’s label (see right side of Fig. 2).

### 4. DISCUSSION AND CONCLUSION

We present a new method for structural parcellation based on structural connectivity as an alternative to hard segmentation. Our method addresses the problems mentioned in Sec. 1:

a) We choose shortest-path tractography (SPT) to avoid problems like path length dependency common in fibre tracking methods. We overcome the problem that SPT always finds a path by assigning a confidence $p$-value to every voxel in the parcellated region.

b) Most voxels in the thalamus are assigned with a low confidence (high $p$-value) for all target regions by our method (see left side of Fig. 2). This occurs mostly inside the thalamus and implies that the diffusion signal for those voxels is not strong or clear enough to assign them to a region with high confidence. This is supported by the low fractional anisotropy (FA) inside the thalamus, suggesting that the resolution is simply not high enough to resolve the location of fibres inside the thalamus.
Fig. 2. Soft (left) and hard [7] (right) segmentations of the thalamus for all target regions for all 5 subjects. Left: Low confidence $p$-values (blue) reflect regions that are more likely connected to the respective target region than high $p$-values (red). Right: As in [7] the thalamus is segmented based on which voxel have most streamlines obtained from probabilistic tractography to the respective target region (white), while the rest of the thalamus is coloured black.

c) Standard hard segmentation of the thalamus [7], as shown in Fig. 2 (right) assumes that each voxel is connected to only one target region. However, our results indicate that high confidence regions overlap strongly between the motor and somatosensory zone, as well as between the prefrontal/temporal and parieto-occipital zone. This makes it especially difficult for the hard segmentation to pick up any voxels connected to the somatosensory zone (see Fig. 2). In contrast, our soft segmentation allows for a voxel to be connected to multiple target regions.

While the soft segmentations overlap strongly with the hard segmentations (see Fig. 2) for the motor, somatosensory and parieto-occipital zone, there is only a small overlap for the prefrontal/temporal zone. One possible reason is that a large part of the hard segmentation lies in the region of high uncertainty, where the resolution of the data is not good enough to resolve the fibre directions. Additionally, the distribution of histograms can be a mixture of multiple underlying distributions corresponding to several major tract that connect the thalamus to the cortex. Here, this may affect the $p$-values of the projection of the prefrontal and temporal cortex onto the thalamus, indicating that the target regions should be sufficiently spatial coherent and reasonably small. A division of the prefrontal/temporal zone into two separate regions as in Behrens et al. [7] will potentially resolve this problem.

Here we are not testing which source voxels are significantly connected to the target region based on the confidence $p$-values, as this would require to address multiple hypothesis testing. Although this could be done via Bonferroni correction, structural connectivity of neighbouring source voxels is strongly correlated. Bonferroni correction would therefore be far too conservative and would affect the segmentation in an unknown way. Future work therefore includes multiple hypothesis testing that respects correlation between source voxels to find those significantly connected. Furthermore, it would be interesting to apply the soft segmentation to other brain structures like the corpus callosum and striatum or to extend the method to be able to parcellate the whole cortex.
based on whole brain tractography.

5. REFERENCES


