

Streamline-Based Analysis: A novel framework for tractogram-driven streamline-wise statistical inference

Simone Zanoni^{1,2}, Jinglei Lv^{1,2,3}, Sharon L. Naismith^{2,4}, Robert E. Smith^{5,6}, Fernando Calamante^{1,2,7}

1. School of Biomedical Engineering, The University of Sydney, Sydney, Australia
2. Brain and Mind Centre, The University of Sydney, Sydney, Australia
3. Sydney Medical School, The University of Sydney, Sydney, Australia
4. School of Psychology, The University of Sydney, Sydney, Australia
5. The Florey Institute of Neuroscience & Mental Health, Melbourne, Australia
6. Florey Department of Neuroscience and Mental Health, The University of Melbourne, Melbourne, Australia
7. Sydney Imaging, The University of Sydney, Sydney, Australia

Supplementary Information

Supplementary note 1. Back-tracking Re-tracking Algorithm

Template streamlines that are naively projected to subject space can result in anatomically implausible streamline-tissue intersection (e.g. CerebroSpinal Fluid - CSF) due to residual local misalignment in cortical anatomy. To address this, we introduce a back-tracking/re-tracking algorithm that refines the trajectories of the template tractogram in subject space, thereby improving subject-level anatomical accuracy near the cortical and subcortical boundaries. This back-tracking/re-tracking algorithm was applied to each subject in the study cohort before computation of streamline-wise quantitative measures (see “Streamline-wise Quantification and Statistical Modelling” in Methods). The implementation consists of three main components:

- Streamline tissue profiling: a tissue segmentation computed from an anatomical T1-weighted image¹ is used to determine the underlying dominant tissue type (White Matter - WM, CSF, subcortical Grey Matter - sGM, cortical Grey Matter - cGM) for each vertex of the streamline, producing a profile of tissue composition along the streamline trajectory.
- Truncation: identification of a suitable point along the streamline where the terminal segment can be truncated. A distinction is made between streamlines terminating within sGM and those extending in cGM or CSF (see algorithms below).
- Re-tracking: if a suitable truncation point could be identified, it is used as a seed to re-initiate streamline tractography, but this time driven by subject-specific Fibre Orientation Distribution (FOD) field and tissue segmentation. The maximum re-tracking length for each streamline end is constrained to 25% of the original streamline length.

These steps are further explained in pseudo-code format as follows:

Algorithm 1.

```

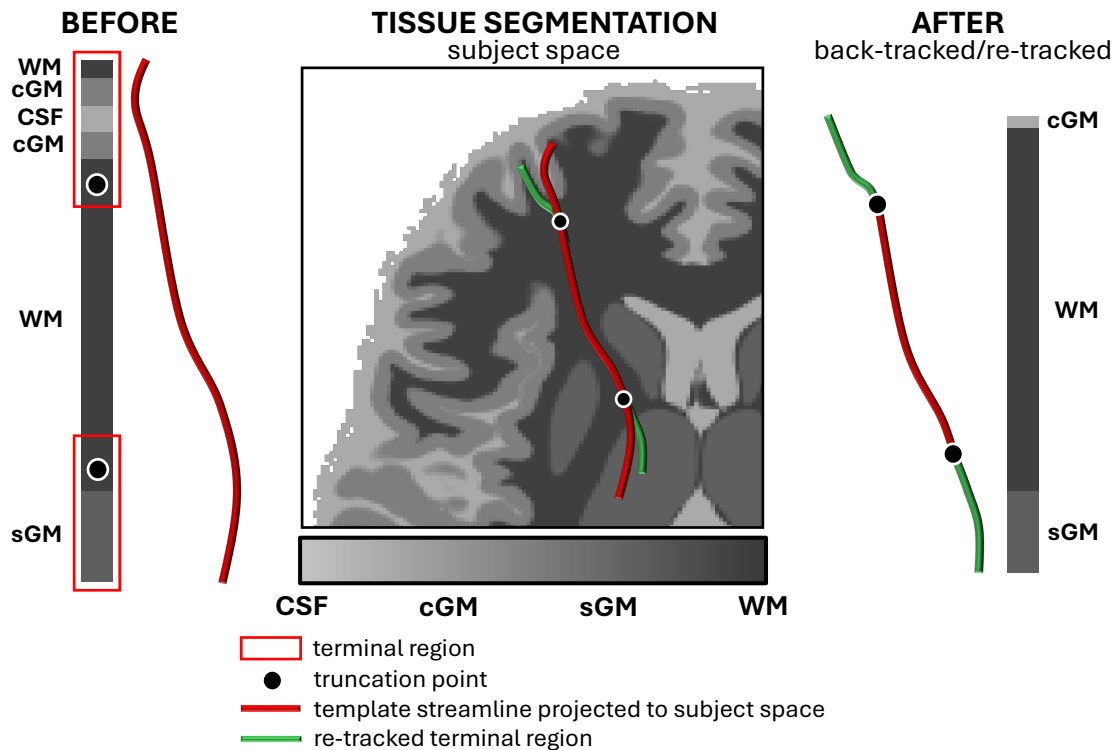
38  for each streamline in template tractogram do
39      for each vertex in streamline do
40          sample tissue type
41      end for
42      execute algorithm 2
43      for each truncation_point identified do
44          truncate streamline at truncation_point
45          re-track from truncation_point using subject-specific data
46      end for
47  end for
48
49  Algorithm 2.
50  set terminal_length = 20 mm
51  set tolerance = 5 mm
52  set min_wm_length = 5 mm
53  set min_sgm_length = 2 mm
54
55  find longest continuous WM segment in streamline
56  if longest WM segment length < min_wm_length then
57      skip process for current streamline
58  end if
59
60  for each streamline endpoint (head and tail) do
61      if streamline_length > 2 × terminal_length then
62          define terminal_region as last terminal_length
63      else
64          define terminal_region as portion outside longest WM portion
65  padded by tolerance into WM

```

```

66     end if
67     if terminal_region ends with continuous sGM portion (>
68 min_sgm_length) then
69         define truncation_point as (last WM vertex before sGM run
70 padded by tolerance into WM)
71     else
72         find innermost WM portion within terminal_region
73         define truncation_point as last WM vertex before non-WM vertex
74 sequence padded by tolerance into WM
75     end if
76     yield truncation_point
77 end for
78

```



79
80 **Supplementary Fig. 1. Overview of the back-tracking/re-tracking process for one exemplar template**
81 **streamline projected to subject space.** In the middle, the streamline is overlaid on a tissue type
82 segmentation image (Cerebro Spinal Fluid – CSF, cortical Gray Matter – cGM, subcortical Gray Matter – sGM,
83 White Matter – WM). Tissue profiles are shown for both the raw subject space projected (left) and back-
84 tracked/re-tracked (right) versions of the streamline.

Supplementary note 2. Numerical simulation

Bundles of Interest for simulated pathology

Three individual tracts and one multi-bundle complex were chosen and segmented from the study-specific template tractogram using TractSeg² to represent simulated pathologies, corresponding to those previously considered in the simulations carried out to optimise the Connectivity-based Fixel Enhancement (CFE) method³. Arcuate fasciculi, corticospinal tracts, and cingulum bundles, were selected for their distinct streamline densities, morphologies, and crossing-fibre environments. Additionally, an Alzheimer's-like complex of tracts was defined as: left arcuate fasciculus, uncinate fasciculi, cingulum bundles, anterior commissure, and genu and splenium of corpus callosum⁴ (Supplementary Fig. 2).

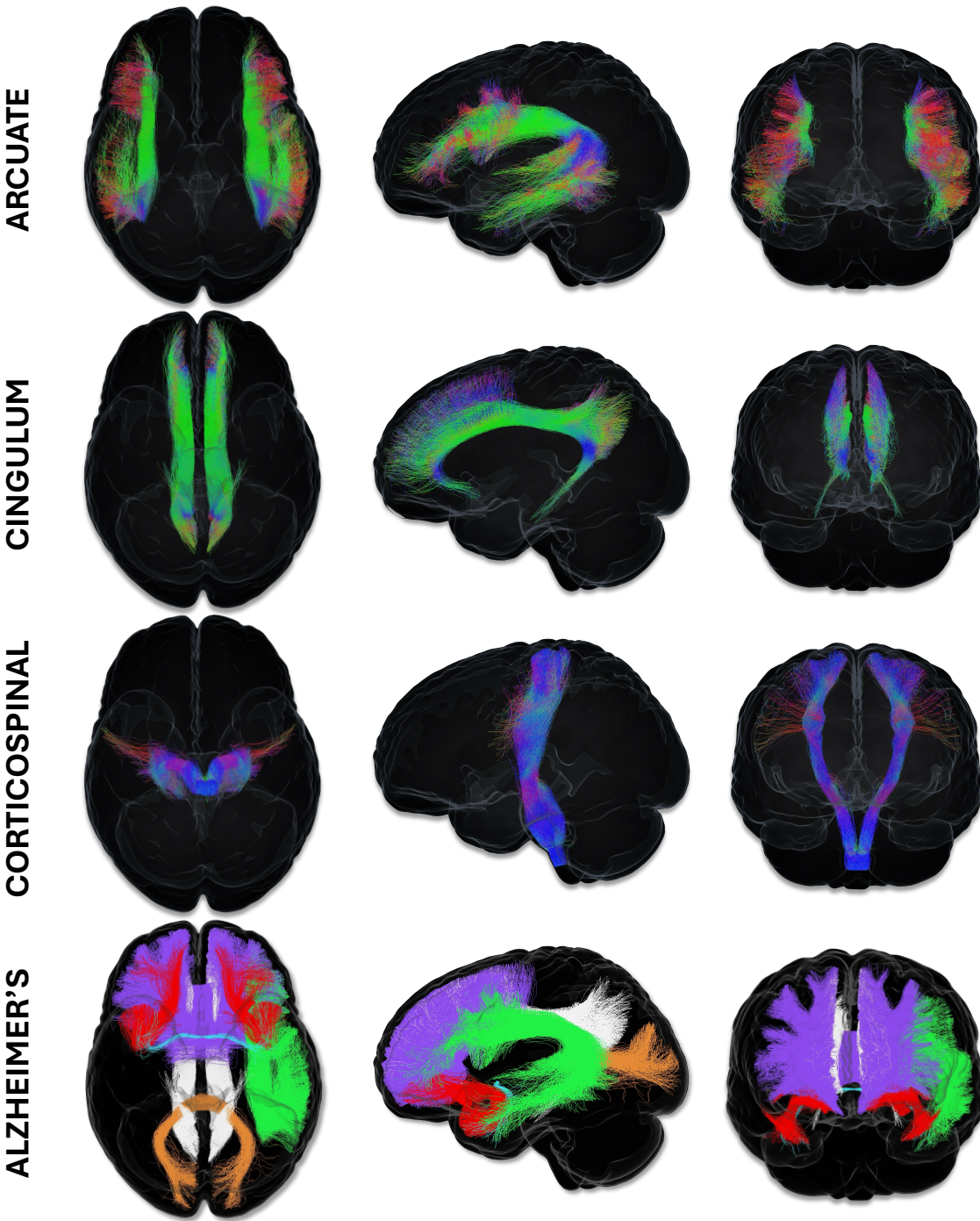
Given the basis by which Streamline-Based Analysis (SBA) operates, it was necessary to attribute each simulated pathology to a subset of streamlines within the template tractogram. This was achieved through use of the voxel masks for bundles and their terminal regions provided by TractSeg: the bundle endpoint masks and the inverse of the bundle mask were used as inclusion and exclusion regions of interest⁵.

Synthetic Data

Following an approach used in previous related studies^{3,6}, our simulation involved synthesising test-statistic data directly, rather than generating diffusion MRI metric maps and computing statistical effects explicitly.

Simulation of noisy test statistics in the absence of an effect necessitates consideration of the non-independence of streamlines within the tractogram. For instance, for a streamline-wise quantitative measure that involves sampling from underlying voxel-wise or fixel-wise image data along the streamline trajectories, each underlying value should be sampled by multiple streamlines, which would result in covariance in the subject streamline-wise measures and therefore the test statistics computed for those streamlines. The nature of this covariance however depends on the specific streamline-wise metric under investigation. For this reason, we considered in our simulations two classes of streamline-wise measures: the mean of a voxel-wise measure along the whole streamline trajectory (which was the case used for the *in vivo* ageing example included here), and the mean of a voxel-wise measure at the two streamline endpoints.

A total of 1,000 whole-brain voxel-wise maps were generated by randomly drawing values from the normal distribution $\mathcal{N}(0,1)$. Following computation of streamline-wise measures, variance normalization was performed to restore unit variance across the tractogram, thereby counteracting the inherent, length-dependent variance suppression caused by averaging independent stochastic samples. While in the endpoint averaging case an empirical normalisation was required to fully recover the target distribution due to the limited number of samples employed, the along-streamline averaging approach allowed us to instead leverage known statistics properties to analytically normalise the data.



123

124

125

126

127

128

129

Supplementary Fig. 2. Bundles of interest used to synthesise test statistics tractograms across different simulated scenarios. The individual fibre tracts (row 1-3) are coloured by local streamline direction (red: left-right; green: anterior-posterior; blue: inferior-superior). Conversely, the complex of tracts commonly affected by the Alzheimer's disease depicted on the bottom row are coloured by tract: left arcuate fasciculus (green), uncinate fasciculi (red), cingulum bundles (white), anterior commissure (blue), genu of the corpus callosum (purple), splenium of the corpus callosum (orange).

130 The uncorrected along-streamline average for streamline j is:

$$\hat{t}_{noise,j} = \frac{\sum_{i \in I} l_{i,j} x_i}{\sum_{i \in I} l_{i,j}} \quad (S1)$$

131 where $l_{i,j}$ is the length of the segment of streamline j intersecting voxel i in image I with value
 132 $x_i \sim \mathcal{N}(0,1)$ ⁷. Therefore, to analytically achieve unit variance, $\hat{t}_{noise,j}$ was divided by:

$$\sigma(\hat{t}_{noise,j}) = \sigma\left(\frac{\sum_{i \in I} l_{i,j} x_i}{\sum_{i \in I} l_{i,j}}\right) = \frac{\sqrt{\sum_{i \in I} l_{i,j}^2}}{\sum_{i \in I} l_{i,j}} \quad (S2)$$

133 This approach ensures that the simulated test statistics in absence of effects t_{noise} can be
 134 generated with the expected level of inter-streamline correlation and the desired distribution.

135 Streamline-wise test statistics for evaluation inclusive of effect were produced by adding to t_{noise} a
 136 t_{signal} value of 1, 2 or 3 (corresponding to Signal-to-Noise Ratio - SNR) to those streamlines
 137 attributed to the respective simulated pathology.

138 This procedure resulted in 1,000 synthetic test-statistic tractograms $t_{signal} + t_{noise}$ for each
 139 combination of simulated pathology, SNR, and class of streamline-wise metric, alongside a set of
 140 1,000 noise-only instance to simulate the null-effect scenario (Supplementary Fig. 3).

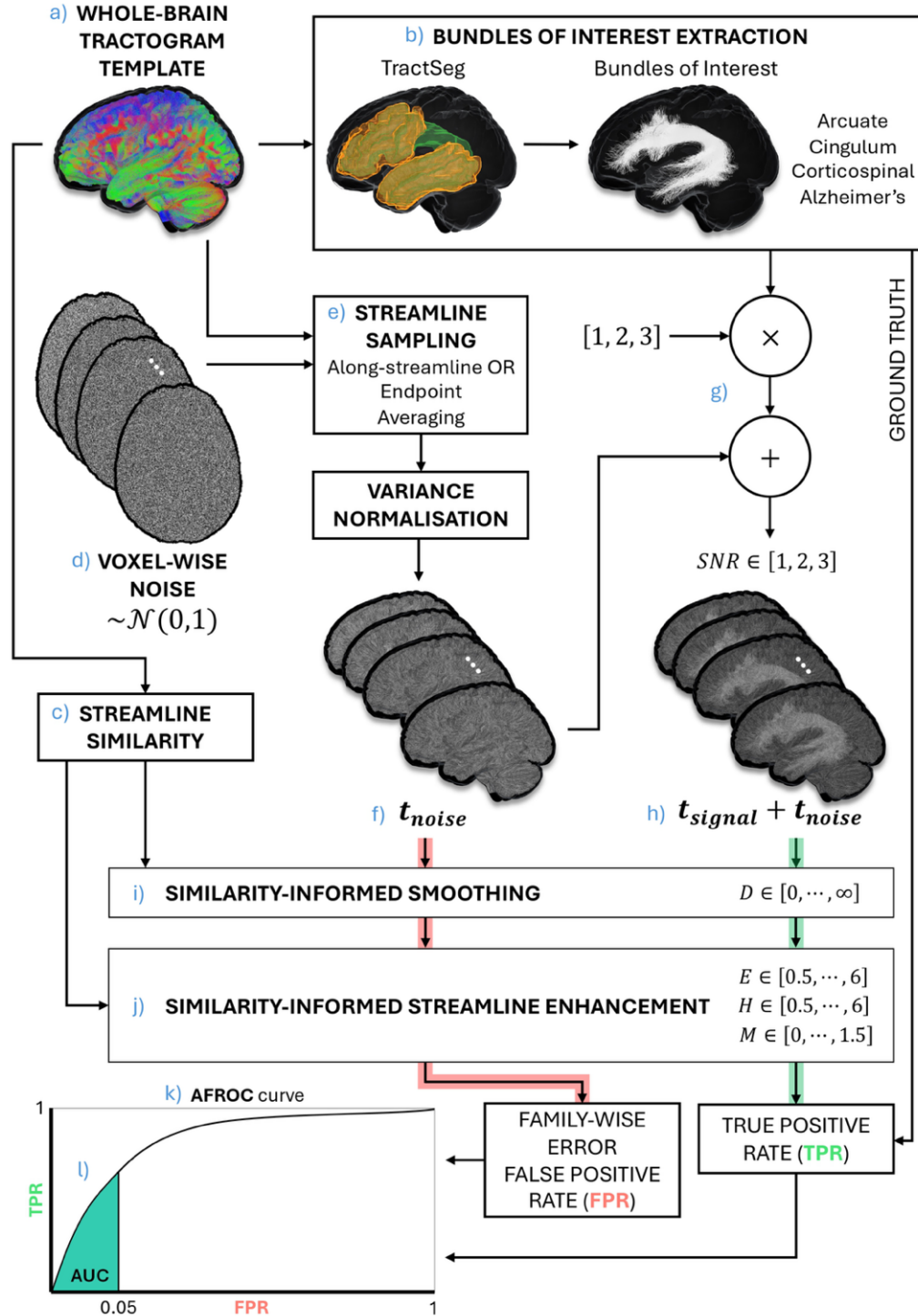
141 **ROC-Based Evaluation**

142 An ROC curve illustrates the relationship between True Positive Rate (TPR) and False Positive Rate
 143 (FPR) of a classifier at various threshold settings. The Area Under the Curve (AUC) provides a
 144 quantitative measure of overall classifier performance. For all the combinations tested in this
 145 simulation, we used the Alternative Free-response ROC (AFROC)⁸ approach, as per precedent set in
 146 the evaluation of Threshold Free Cluster Enhancement (TFCE) and CFE, which bases the FPR on the
 147 reporting of a false positive anywhere for data in the absence of a true effect, thereby seeking Family-
 148 Wise Error control. Both simulated “true effect” and null-effect test-statistic tractograms underwent
 149 streamline data smoothing and Similarity-informed Statistical Enhancement (SSE). Unlike how SBA
 150 would be applied in practise, where smoothing is applied to per-subject data, here smoothing was
 151 performed directly on the streamline-wise test-statistic data; this approach is considered
 152 comparable to computing the test statistic from smoothed input data provided that variance
 153 normalisation is applied post-smoothing⁶. The enhanced noise-only tractograms were used to
 154 compute the FPR, defined as the fraction of the 1,000 noise-only instances in which at least one
 155 streamline test statistic exceeded the given threshold. The TPR was defined as the average fraction
 156 of streamlines attributed to the simulated pathology that survived thresholding across the 1,000
 157 realisations inclusive of that pathology. We computed the AUC solely over the [0, 0.05] FPR range
 158 (and rescaled appropriately) to capture the TPR-FPR trade-off of interest.

159 For each combination of enhancement parameters E , H , M , and smoothing degree D , a cost function
 160 was computed as a composite score of reward and cost terms for performance comparison. First,

the mean performance $\overline{AUC}$ across simulated pathologies and SNRs was used as a reward term. The performance inconsistency was also quantified as the standard deviation σ of the AUCs across simulated pathologies and SNRs and penalised. The resulting cost function follows:

$$score = \overline{AUC} - \sigma \quad (S3)$$



Supplementary Fig. 3. Pipeline of the AFROC quantitative evaluation of the similarity-informed smoothing and enhancement of SBA using simulated test-statistic tractograms. Tractogram segmentation with extraction of the bundles of interest (b) and streamline similarities (c) are obtained from the whole-brain tractogram template (a). The 1,000 realisations of the simulated per-streamline test statistics t_{noise} (f) are obtained via streamline sampling (e) on normally distributed voxel-wise noise images (d). The t_{noise} is then added to the modulated (g) bundles of interest to achieve $t_{signal} + t_{noise}$ with a target SNR (h). Both the t_{noise} and the $t_{signal} + t_{noise}$ undergo different degrees of smoothing (i) and statistical enhancement with different internal parameters (j). The enhanced t_{noise} and $t_{signal} + t_{noise}$ are then used to compute FPR and TPR, respectively, using the ground truth (b). The AFROC curve (k) is finally used to compute the AUC (l) where $FPR < 0.05$.

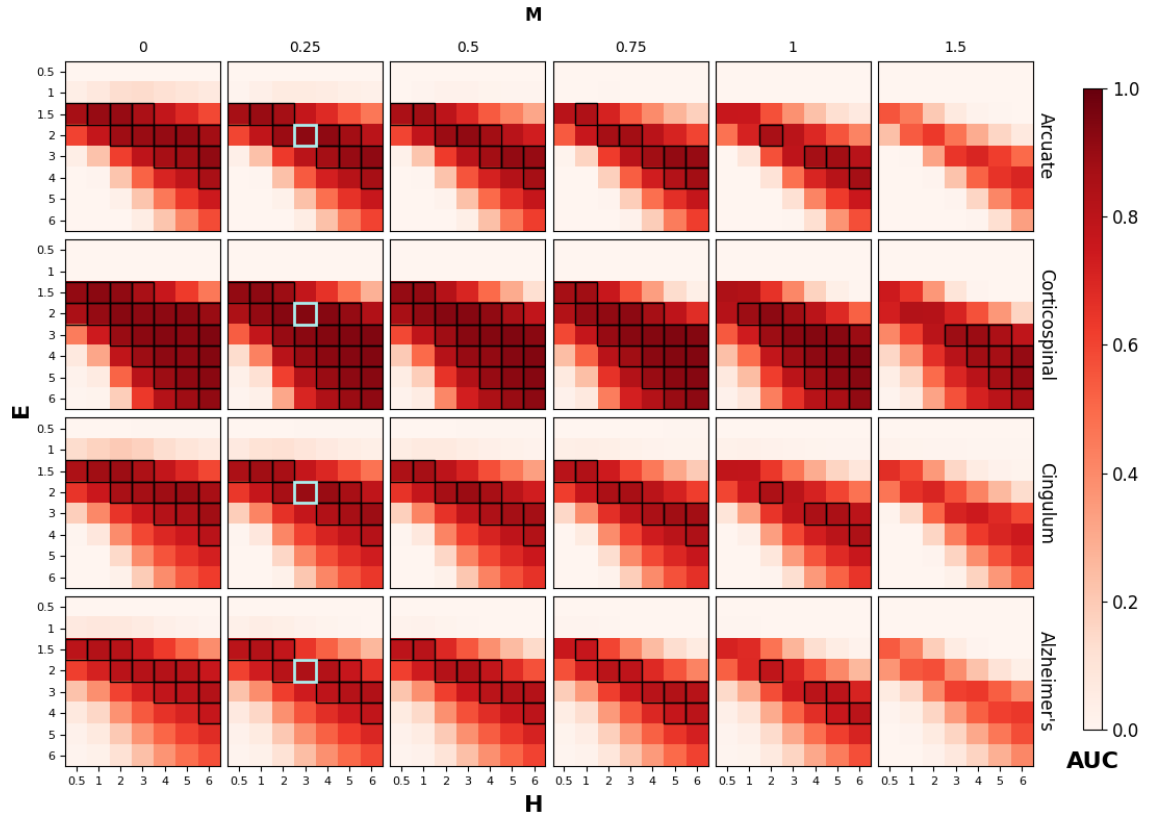
Simulation Results

Performance varied across bundles of interest, SSE parameters, smoothing degrees, SNRs, and streamline-wise measures, as shown in the AUC heat maps (Supplementary Fig. 4-6). Supplementary Figure 4 reveals that the adoption of the recommended parameters (light blue square: $E = 2$, $H = 3$, $M = 0.25$, and $D = 5$) led to a comparable performance across the 4 simulated pathologies for the along-streamline averaging. Moreover, the presence of multiple high-performing parameter combinations (black squares) suggests that the algorithm exhibits overall low sensitivity to SSE parameter selection around the optimal set, with the degree of sensitivity varying across bundles.

Supplementary Figure 5 indicates the optimal parameter set for the endpoint-averaging case (green square), corresponding to $E = 3$, $H = 5$, $M = 0.25$, and $D = 5$. Although this optimal parameter set is slightly different from that of the along-streamline averaging case, the latter parameter set achieves an AUC value within 2% of that obtained with that numerically optimal for endpoint averaging, making it a robust choice across sampling strategies.

The comparison across SNRs and smoothing degrees (Supplementary Figure 6) highlighted two patterns. First, SBA barely discriminates between noise-only and true effect at $SNR=1$, but can reliably detect true effects as SNR increases, reaching a 0.94 AUC peak performance when using the recommended parameter set at $SNR=3$. Second, the analysis revealed that at a fixed SNR level, the smoothing degree has only a mild impact on SBA performance, with best performance at weak or no smoothing, contrasting conventional neuroimaging pipelines; this is likely due to those streamlines proximal enough to influence one another through smoothing being already correlated due to having sampled underlying data from the same image locations. What is shown in Supplementary Figure 6 for the cingulum bundle is representative of what was found across all synthetised bundles and sampling strategies.

ALONG-STREAMLINE AVERAGING

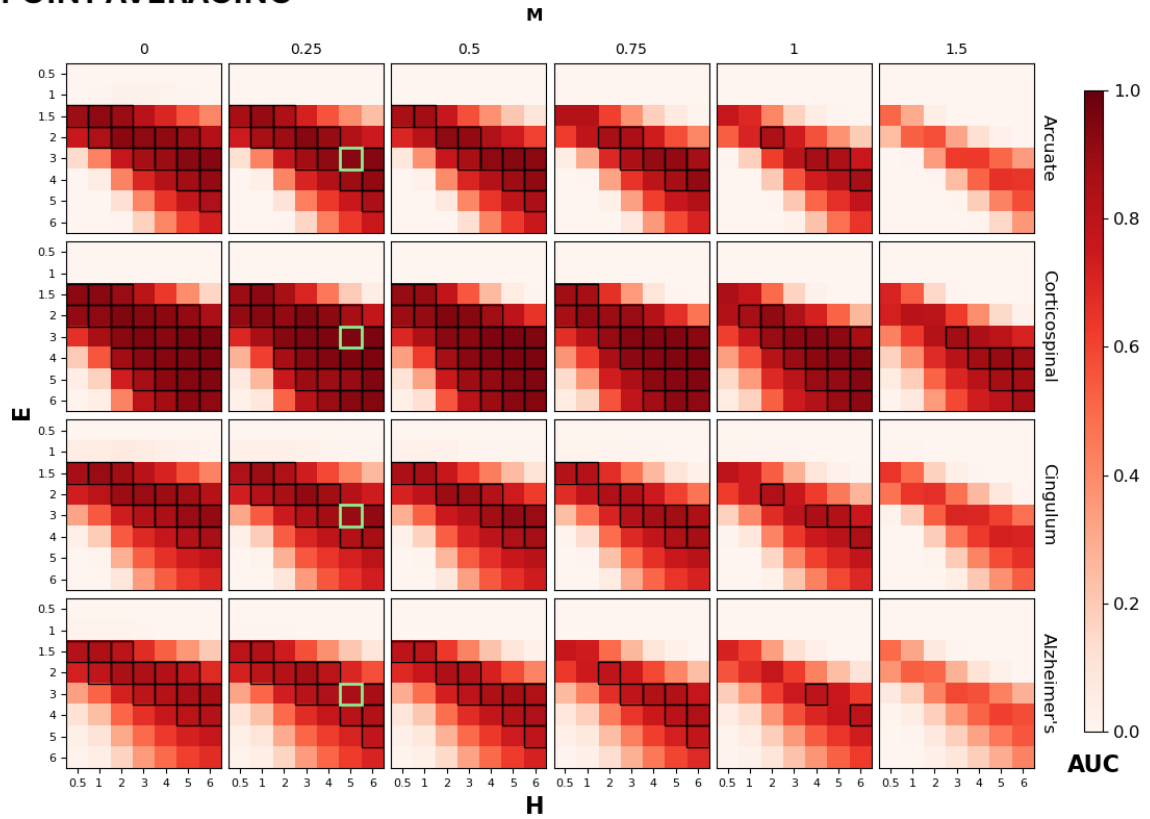


200

201 **Supplementary Fig. 4. Performance quantification in the along-streamline averaging scenario.**

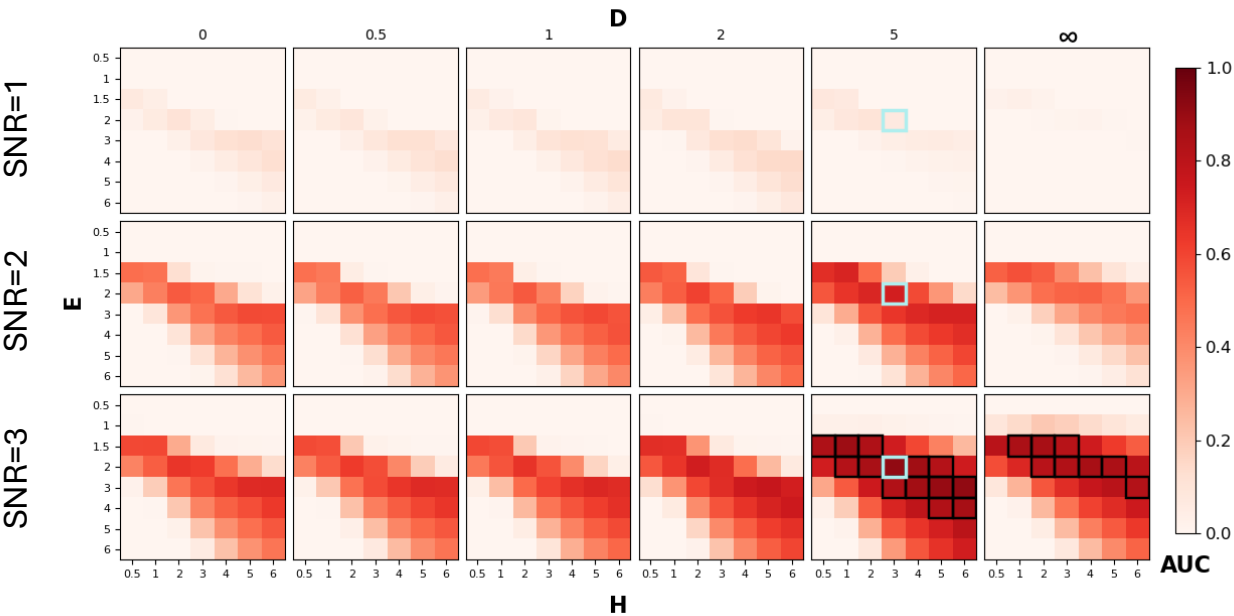
202 Evaluation across bundles of interest (rows) and SSE parameters (E, H, and M) at SNR=3 and smoothing
 203 power coefficient D=5. The light blue square highlights the best combination of SSE parameters (i.e. E=2,
 204 H=3, M=0.25), as determined by Equation (S3). Black squares indicate other parameter combinations
 205 achieving an AUC of at least 90% of the AUC of the best parameter combination (light blue square).

ENDPOINT AVERAGING



Supplementary Fig. 5. Performance quantification in the endpoint averaging scenario. Evaluation across bundles of interest (rows) and SSE parameters (E, H, and M) at SNR=3 and smoothing power coefficient D=5. The green square corresponds to best parameters of choice for the endpoint averaging case (i.e. E=3, H=5, M=0.25). Black squares indicate other parameter combinations achieving an AUC of at least 90% of the AUC of the best parameter combination (green square).

ALONG-STREAMLINE AVERAGING - Cingulum



Supplementary Fig. 6. Performance quantification across smoothing degrees and SNRs. Evaluation across smoothing degrees ($D = [0, 0.5, 1, 2, 5, \infty]$) and SNRs (rows) for a representative bundle (Cingulum) and along-streamline averaging. The light blue squares correspond to the recommended SSE parameters; black squares indicate all the other combinations achieving at least 90% of that achieved using recommended parameters.

References

1. Smith, R., Skoch, A., Bajada, C. J., Caspers, S. & Connelly, A. Hybrid surface-volume segmentation for improved anatomically-constrained tractography. in *Proceedings of the 26th Annual Meeting of the Organization for Human Brain Mapping (OHBM)* (2020).
2. Wasserthal, J., Neher, P. & Maier-Hein, K. H. TractSeg-Fast and accurate white matter tract segmentation. *NeuroImage* **183**, 239–253 (2018).
3. Raffelt, D. A. *et al.* Connectivity-based fixel enhancement: Whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres. *Neuroimage* **117**, 40–55 (2015).
4. Mito, R. *et al.* Fibre-specific white matter reductions in Alzheimer’s disease and mild cognitive impairment. *Brain* **141**, 888–902 (2018).
5. Tournier, J.-D. *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* **202**, 116137 (2019).
6. Smith, S. M. & Nichols, T. E. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* **44**, 83–98 (2009).
7. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. SIFT: Spherical-deconvolution informed filtering of tractograms. *Neuroimage* **67**, 298–312 (2013).
8. Chakraborty, D. P. & Winter, L. H. Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology* **174**, 873–881 (1990).