MS. TABINDA SARWAR (Orcid ID : 0000-0001-7313-5350)

Article type : Full Paper

TITLE PAGE

Mapping connectomes with diffusion MRI: deterministic or probabilistic tractography?

Tabinda Sarwar[1], Kotagiri Ramamohanarao[1] and Andrew Zalesky[2,3]

[1] Department of Computing and Information Systems, The University of Melbourne, Victoria 3010, Australia

[2] Department of Biomedical Engineering, The University of Melbourne, Victoria 3010, Australia

[3] Melbourne Neuropsychiatry Centre, The University of Melbourne, Victoria 3010, Australia

Abstract Word Count:	192
Manuscript Word Count:	5495
Number of figures:	8
Number of tables:	2
Number of supporting figures:	14
Number of supporting tables:	4

Corresponding author:

Tabinda Sarwar

Email: tsarwar@student.unimelb.edu.au

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/mrm.27471

Address: Department of Computing and Information Sytems
Doug McDonnell Building
The University of Melbourne
Victoria, 3010, AUSTRALIA

Mapping connectomes with diffusion MRI: deterministic or probabilistic tractography?

Abstract

Purpose

Human connectomics necessitates high-throughput, whole-brain reconstruction of multiple white matter fiber bundles. Scaling-up tractography to meet these high-throughput demands yields new fiber tracking challenges, such as minimizing spurious connections and controlling for gyral biases. The aim of this study is to determine which of the two broadest classes of tractography algorithms—deterministic or probabilistic—is most suited to mapping connectomes.

Methods

This study develops numerical connectome phantoms that feature realistic network topologies and that are matched to the fiber complexity of *in vivo* diffusion magnetic resonance imaging (dMRI) data. The phantoms are utilized to evaluate the performance of tensor-based and multi-fiber implementations of deterministic and probabilistic tractography.

Results

For connectome phantoms that are representative of the fiber complexity of *in vivo* dMRI, multi-fiber deterministic tractography yields the most accurate connectome reconstructions (F-measure=0.35). Probabilistic algorithms are hampered by an abundance of false positive connections, leading to lower specificity (F=0.19). While omitting connections with the fewest number of streamlines (thresholding) improves the performance of probabilistic algorithms (F=0.38), multi-fiber deterministic tractography remains optimal when it benefits from thresholding (F=0.42).

Conclusions

Multi-fiber deterministic tractography is well suited to connectome mapping, while connectome thresholding is essential when using probabilistic algorithms.

Keywords: connectome, phantom, ground truth, tractography, diffusion MRI, network

Introduction

Mapping the human connectome (1, 2) is a major goal in neuroscience. The connectome refers to a comprehensive network description of the brain, often represented as a graph, where nodes denote brain regions and edges represent white matter pathways (3-5). Tractography can be used to reconstruct white matter pathways *in vivo* using diffusion-weighted Magnetic Resonance Imaging (dMRI) (6, 7). Tractography was originally developed to reconstruct and virtually dissect individual or small groups of white matter fascicles (8, 9), whereas connectomics requires high-throughput, whole-brain reconstruction of thousands of connections. Scaling-up tractography to meet these high-throughput demands has given rise to new challenges, such as minimizing spurious connections (10), identifying fiber terminations and acceptance criteria (11, 12), and controlling for gyral biases (13, 14). Which of the many tractography algorithms available can best address these important challenges remains unclear (15, 16).

Tractography in connectomics is a process that comprises three steps: (i) estimation of local fiber orientations; (ii) linking together these local fiber orientations to generate streamlines that represent the long-range trajectories of white matter fibers; and, (iii) assigning streamlines to pairs of nodes to build a connectivity matrix. Numerous models have been developed to estimate local fiber orientations. The classical diffusion tensor (17-19) is one such model that is unable to estimate multiple fiber orientations. To overcome this limitation, a probability distribution of orientations can be fitted using any of a number of models, including deconvolution methods (20, 21), q-ball imaging (22-25) or diffusion spectrum imaging (26). Their utility has been demonstrated in tracking deep white matter (27) and the transition between gray and white matter (28). Once local fiber orientations have been estimated, tractography is used to link together these local orientations to generate streamlines that trace out the trajectories of fiber bundles.

Tractography can be broadly categorized into two classes: deterministic and probabilistic algorithms. The salient feature that distinguishes these two categories is the sampling of fiber directions for streamline propagation. With deterministic tractography, streamlines are steered according to a fixed (deterministic) direction at each voxel. Deterministic methods can be used for both single (9, 29-31) and multiple fiber orientations estimated per voxel. A criticism leveled against deterministic methods is that they are unable to account for inherent uncertainty in estimates of fiber orientations (32), are sensitive to the estimated principal direction (33) and susceptible to noise (34). In contrast, with probabilistic tractography, a

distribution of fiber orientations is estimated for each voxel (27, 35, 36), where a sample is randomly drawn from this distribution to determine the direction of streamline propagation. Probabilistic methods are usually more computationally expensive, but can better account for uncertainty in data (33, 37, 38).

Probabilistic tractography is generally considered a superior method for reconstructing and dissecting individual white matter fascicles, particularly if there is a need to quantify the confidence (probability) of each reconstructed pathway (6). However, for high-throughput applications of tractography such as connectome mapping, where spuriously reconstructed fibers can rapidly multiply due to the multiplicity of testing, it remains unclear whether deterministic or probabilistic tractography is superior. Tractography in connectomics involves seeding millions of streamlines throughout the entire white matter volume. With probabilistic methods, each of these streamlines can traverse a unique trajectory, and if a sufficient number of streamlines is generated, it is likely that streamlines will be found between virtually all pairs of nodes, resulting in a fully-connected connectome. To reduce connection density, thresholding can be performed to eliminate connections comprising the fewest numbers of streamlines, under the assumption that these connections are spurious (39, 40). While eliminating connections with a low streamline count can improve connectome specificity, not all eliminated connections are necessarily spurious, and thus any gain in specificity is inevitably traded for a loss in sensitivity (41-43). Given these difficulties, deterministic fiber tracking might be considered a viable alternative for high-throughput applications such as connectome mapping. Indeed, a recent study that compared a plethora of tractography algorithms found that deterministic methods can sometimes outperform their probabilistic counterparts (15). It is important to remark that several advanced fiber tracking methods are not amenable to classification as either deterministic or probabilistic tractography algorithms. These include global (44-51) and graph-theoretic tractography methods (52-55), machine learning approaches (56, 57), fiber tracking based on anatomical models (14) and fast marching methods (58, 59). Advanced methods are not evaluated in the present study.

The absence of a ground truth limits the ability to assess the accuracy and precision of connectomes reconstructed with tractography. The current gold standard in quantitative tractography evaluation is axonal tract tracing (13, 42, 43, 60-62), but tract tracing is highly invasive and yields a micron resolution that is orders of magnitude finer than the millimeter resolution afforded by dMRI. This mismatch in spatial scale can complicate the comparison between the two modalities. To overcome the shortcomings of using tract tracing as a benchmark for tractography, dMRI phantoms have been developed in the form of physical (63-67) and numerical (68-74) dMRI phantoms. The majority of these phantoms were developed to evaluate the accuracy with which fiber bundle geometry, spatial extent and curvature is reconstructed by tractography. However, faithful reconstruction of these geometric attributes is not a key

consideration for connectomics. When mapping a connectome, the most important consideration is whether a fiber bundle is present/absent between a pair of brain regions, and the strength of fibers that are present. Most phantom-based evaluations undertaken to date have not assessed the important step of assigning streamline endpoints to pairs of nodes to build a connectivity matrix. A notable exception is the recent study of Maier-Hein and colleagues (15), where a numerical phantom (composed of 25 manually segmented white matter bundles) was used to evaluate tractography performance with respect to the number of valid (true positives) and invalid (false positives) bundles. These two metrics explicitly quantify the accuracy and precision with which connectomes (connectivity matrices) can be reconstructed with tractography. These connectivity measures were first developed in the Tractometer study (75,76), a tractography evaluation and validation system, which evaluated valid and invalid bundles estimated by tractography algorithms under the simplified conditions of the realistic dMRI Fibercup phantom, comprising only seven fibers (63). Subsequently, Girard and colleagues (77) used a numerical phantom comprising 27 fibers (73, 75), to show that connectivity measures derived from Tractometer can be improved by seeding streamlines from the white/gray-matter interface, using partial volume as stopping and streamline filtering criteria.

Here, we aim to develop numerical connectome phantoms to evaluate which of single and multi-fiber implementations of deterministic or probabilistic tractography can reconstruct binary connectivity matrices with the greatest accuracy and precision. We develop connectome phantoms with realistic topological attributes, at the expense of a more rudimentary modeling of fiber geometry. Fiber bundles are modeled as curved tubular structures of fixed diameter, and nodes are positioned along the circumference of a circle and interconnected with realistic network topologies. We generate ensembles of ground truth connectivity matrices, simulate corresponding dMRI signals, estimate local fiber orientations in the simulated data, perform tractography and finally reconstruct connectivity matrices. We evaluate the performance of connectivity matrix reconstruction with respect to the presence or absence of connections, without considering connectivity strength.

Methods

Connectome Phantom Generation

This section provides an overview of the construction and simulation of the connectome phantom. Details can be found in Supporting Information S1 and a schematic is shown in Figure 1(a).

i. **Node Delineation:** The phantom was instantiated in three dimensions, where fiber trajectories and node locations were fixed across the third dimension for simplicity and ease of visualization. A circle was positioned at the center of a $140 \times 140 \times 5$ grid of voxels and fiber trajectories were

only permitted to traverse the interior of the circle. The circumference of the circle was uniformly subdivided into N_R sectors, where each sector represented a single node. For the current study, we used N_R=*25*.

ii. **Connectivity Matrix Generation:** The absence/presence of connections was encoded with a symmetric connectivity matrix G ($N_R \times N_R$) where element (i, j) was set to one if a connection was present between nodes i and j, otherwise this element was set to zero. The connectivity matrix was sampled from a generative model that yielded networks with topological attributes comparable to empirical brain networks (78-79). Specifically, the relative probability of connection formation between nodes i and j was given by $P(i,j) \propto E(i,j)^{\eta} \times K(i,j)^{\gamma}$ where $E(i,j)$ denotes the Euclidean distance and $K(i,j)$ represents a non-geometric relationship between nodes i and j. Betzel and colleagues (79) tested different non-geometric models for $K(i,j)$ and found that the matching index resulted in networks with the greatest topological resemblance to empirical connectomes. We thus generated networks based on the matching index model, with exponents that have previously been found to yield realistic topologies ($\eta = -0.98$ and $\gamma = 0.42$; 79). The number of connections was varied to generate phantoms with connection densities ranging from 2 to 20%.

iii. **Fiber Bundles:** Tubular fibers with varying curvature and diameter were generated between all pairs of nodes *(i,j)* for which G_{ij} = *1*. For each such pair of nodes, a curved trajectory was drawn between randomly chosen voxels in nodes i and j. Specifically, control points were positioned along the straight line between the two endpoints, each of which defined the center of a circle around which the straight line was deformed to yield a curved fiber trajectory (Figure 1(b)). For a fiber with only a single bend (Figure S.1(b)-left image), the mid-point of the line connecting the two endpoints was computed, which served as the center of curvature. The point located perpendicularly with distance r from this center acted as a control point for the fiber. Afterwards cubic spline interpolation of the end points and control point was performed to generate intermediate points defining the complete fiber. Similarly, for a fiber with multiple bends (Figure 1(b)-right image), three center of curvature points were equidistantly located on the line joining the two end points. A local tangent vector, denoted with v, was estimated for each of 1000 points parameterizing the length of each fiber trajectory. Each trajectory was then dilated cross-sectionally to endow it with volume, thus resulting in a tubular bundle (Figure 1(c)). Tangent vectors were used to simulate the dMRI signal for each voxel.

iv. **Synthetic dMRI Generation:** The compartment model was used to generate dMRI data. Each fiber that traversed a voxel contributed an anisotropic compartment and voxels that were not

traversed by any fiber comprised an isotropic compartment. We used the "ball" model (38) for the isotropic and the zeppelin model (80-82) for anisotropic compartments, with volume fraction *f* and *1-f* respectively. The dMRI signal for the *i*th voxel for gradient direction g and b-value *b* was given by (71),

$$\mu_i = fe^{-bd} + (1-f)\sum_{p\in P_i}\int_p e^{-bg^T D_{p(x)} g}\, dx \quad (1)$$

where d is the diffusivity of the isotropic compartment, P_i is the set of all fiber paths inside the *i*th voxel and $D_{p(x)}$ is the prolate tensor defining the anisotropic diffusion at point x of path $p \in P_i$, given by,

$$D_{p(x)} = \alpha vv^T + \beta I \quad (2)$$

where v denotes the local tangent vector of the fiber trajectory at point x (Step iii), I is the identity tensor, α and β are diffusivity parameters. If $d_{||}$ and $d_{\perp}$ are the apparent diffusivities parallel and perpendicular to the principal direction respectively, then $d_{||}=\alpha+\beta$ and $d_{\perp}=\beta$. The integral in Equation (1) was evaluated by sampling at most 3 tangent vectors within each voxel along the length of each fiber path. Therefore, for a given fiber path, p, multiple tangent vectors were often estimated within the same voxel.

v. Noise: Finally, Rician noise (83) was added to the dMRI signal. The signal-to-noise ratio (SNR) was computed as $SNR = S_0/\sigma$, where S_0 is the non-gradient weighted signal (b=0) averaged across voxels traversed by fibers.

Phantom Simulation

We generated phantoms with connection densities ranging from 2 to 20% in increments of 2%. For each connection density, we generated 100 phantoms, where for each phantom a new connectivity matrix was sampled using the generative model. Diffusivity parameters were chosen to match human brain tissue (71); namely, $d_{||}$= 1.5x10^{-3} mm^2/s and $d_{\perp}$= 0.2x10^{-3} mm^2/s for the zeppelin compartment, whereas d= 0.9x10^{-3} mm^2/s for the ball compartment. The choice of the volume fraction value (f) for compartment models is somewhat arbitrary (84, 85). In this study, we used a volume fraction of $f = 0.2$ for all voxels. We also evaluated a range of alternative volume fraction values and found that the relative performance ranking of the four tractography algorithms was not substantially affected by this parameter choice (Supporting Information S4).

Nominal dMRI acquisition parameters for this study were: 2x2x2 mm^3-voxel resolution, 60 diffusion gradients distributed equally over the sphere, b-value of 2000 s/mm^2 and an SNR of 10 dB for the non-

gradient weighted image. Investigation of alternative acquisition parameters (b-value = 1000, 3000 s/mm^2 and SNR = 20, 30 dB) can be found in Supporting Information S2 and S3.

Tractography

We utilized our phantoms to evaluate which of deterministic or probabilistic tractography can reconstruct binary connectivity matrices with the greatest accuracy and precision. We considered four tractography implementations: i) deterministic tensor-based tracking (*Det Tensor*) with Euler integration (29); ii) probabilistic tensor-based tracking (*Prob Tensor*) using bootstrapping (37); iii) deterministic tracking steered according to peaks in the fiber orientation distribution (FOD) estimated with constrained spherical deconvolution (*Det CSD*; 36); and, iv) probabilistic CSD tracking (*Prob CSD*; 35). We used implementations of these four algorithms provided by the MRtrix `tckgen` function (www.mrtrix.org; `tensor_det`, `tensor_prob`, `sd_stream` and `ifod2`, respectively).

A spherical harmonic order of 8 was used for CSD and 1 million streamlines were generated for each phantom. Default parameter settings recommended in the MRtrix `tckgen` function were used for the angle threshold, step size, FA (fractional anisotropy) and FOD threshold (Table 1).

Evaluation Methodology

Tractography was performed on the phantoms to infer connectivity matrices, which were then compared to the ground truth connectivity matrices. A *connection* was defined as the set of all streamlines with endpoints residing in the same pair of nodes. Streamlines with one or both endpoints terminating before reaching a node were discarded. Each connection was associated with a *streamline count*, defined as the number of streamlines that it comprised. Connections that were present in both the ground truth and reconstructed connectivity matrices were called true positives (TPs), while connections that were only present in reconstructed matrices were called false positives (FPs). False negatives (FNs) were connections that were absent in the reconstructed connectivity matrices, but present in the ground truth matrices.

Before computing TPs, FPs and FNs, the streamline count for each connection was thresholded to yield binary connectivity matrices *B*. The streamline threshold was increased incrementally from zero to the maximum streamline count of a connecting pair of nodes (reported as percentage of streamlines). If the connection between regions i and j comprised fewer streamlines than the threshold, we set $B_{ij} = 0$; otherwise $B_{ij} = 1$. The binary matrices *B* were used to compute the FP and TP rate,

$$\#TP = \sum_{ij} 1_{\{B_{ij} \cap G_{ij}\}} \; ; \; \#FP = \sum_{ij} 1_{\{B_{ij} - G_{ij}\}} \; ; \; \#FN = \sum_{ij} 1_{\{G_{ij} - B_{ij}\}} \quad (3)$$

$$TPR = \frac{\#TP}{\sum_{ij} 1_{\{G_{ij}\}}}; FPR = \frac{\#FP}{\sum_{ij}(1-G_{ij})} \quad (4)$$

The F-measure was computed for the reconstructed connectivity matrices to assess the tradeoff between TPs and FPs with a single index,

$$F = \frac{2 \times \#TP}{2 \times \#TP + \#FN + \#FP} \quad (5)$$

We used the F-measure primarily because this index was not overwhelmed by the disproportionate number of true negatives that was characteristic of the relatively sparse connectivity matrices considered here.

Phantom Complexity

To quantify the impact of variations in connection densities, fiber configurations and network topologies, we defined measures of voxel and fiber complexity. Voxel complexity, C_v, was defined as the proportion of voxels intersected by *more* than one fiber bundle out of the set of all voxels intersected by *at least* one fiber bundle. Complementarily, fiber complexity, C_F was defined as,

$$C_F = \frac{\sum_{f_p>1} f_p}{\sum_{f_p \geq 1} f_p} \quad (6)$$

where f_p denotes the number of fibers traversing the pth voxel in the ground truth. For C_v, complexity is measured categorically, with any voxel comprising more than one fiber considered complex. In contrast, C_F quantifies the percentage of multiple fibers in a phantom and together these two measures help to understand the impact of multiple fiber configurations on tractography.

Using our measures of voxel and fiber complexity, we benchmarked the complexity of each phantom to the complexity of *in vivo* dMRI data from the Human Connectome Project (HCP; 86). To measure C_v and C_F in the HCP data, the number of fiber orientations per voxel was estimated with two complementary methods: i) estimation of the FOD using multi-tissue constrained spherical deconvolution for multi-shell dMRI (MSMT;87); and, ii) the bedpostx routine provided by FSL (38). Further details are provided in Supporting Information S5.

Results

Figure 2(a-d) shows the dMRI image for a typical connectome phantom with connection density of 12%, including an axial slice of the non-diffusion weighted (b=0) image and FA image. Figure 3 shows representative streamlines generated by the four tractography algorithms for dMRI images with varying

connection density. Figure 4 shows representative streamlines generated by the four tractography algorithms for the dMRI images shown in Figure 2(c), together with the diffusion tensor and FOD estimates for a slice of voxels in a region of intersecting fibers. It can be seen that Det Tensor generated several broken streamlines that failed to traverse the intersecting fibers, whereas the CSD-based tracking algorithms navigated this region with greater success. However, Prob CSD yielded streamline trajectories that were somewhat rippled (i.e. trajectory oscillates from side-to-side). We quantify this rippling artifact below.

For brevity, we only report performance for the case of dMRI images with a b-value of 2000 s/mm^2 and SNR of 10 dB. Alternative b-values (1000, 3000 s/mm^2) and/or SNR (20, 30 db) affected performance but did not substantially affect the relative performance ranking of each tractography algorithm (Supporting Information S2 and S3). Performance was quantified using three criteria: i) receiver operating characteristic (ROC) curves, which were parameterized by the streamline threshold (Figure 5); ii) F-measure evaluated as a function of the streamline threshold (Figure 6); and iii) FN-to-FP ratio evaluated as a function of the streamline threshold (Supporting Information S6). For each of the four tractography algorithms, we quantified performance in terms of the mean and standard deviation over 100 phantom realizations for 10 cases corresponding to distinct connection densities (2, 4, …, 20%). Finally, C_v and C_F were used to evaluate the complexity of *in vivo* dMRI data from the HCP. We compared the complexity of our phantoms with the *in vivo* dMRI data to determine which phantom was most representative of the complexity of the human connectome (Figure 7).

Evaluation of Connectome Reconstruction Performance

ROC curves for the four tractography algorithms are presented in Figure 5. Prob CSD yielded the highest TPR for all connection densities, whereas Det Tensor yielded the lowest. However, Prob CSD also yielded the highest FPR. Therefore, in the absence of streamline thresholding, Prob CSD provides high sensitivity but poor specificity. Conversely, Det CSD yielded improved specificity compared to Prob CSD but lower sensitivity. This emphasizes the tradeoff between connectome sensitivity and specificity. Increasing connection density, and hence the number of crossing fibers, decreased the TPR for all algorithms and Det CSD became comparable with tensor-based tractography for a connection density of 20%. At this density, Prob CSD continued to yield the highest TPR, but at the expense of a very high FPR.

To quantify the tradeoff between TPR and FPR, we evaluated the F-measure as a function of the streamline threshold for each tractography algorithm (Figure 6). Det CSD achieved the maximum F-measure for all connection densities in the range 2-16%. For phantoms with connection densities exceeding 16%, Prob CSD either rivaled or exceeded the performance of Det CSD.

We therefore conclude that the choice between Det and Prob CSD should be guided by connection density. Det CSD is well suited to mapping relatively sparse connectomes, whereas Prob CSD achieves a superior F-measure once the ground truth connection density exceeds a given threshold (20% for our phantoms). While streamline thresholding can be used to improve the performance of Prob CSD by eliminating FP connections, the choice of streamline threshold to maximize accuracy is unknown in practice. However, if an appropriate streamline threshold can be chosen, it can be seen that the connection density at which Prob CSD first rivals the performance of Det CSD is reduced to 14%. The advantage of Det CSD is that near-optimal performance can be achieved without the need for streamline thresholding. In summary, our performance evaluation indicates that: i) Det CSD is optimal for mapping sparse connectomes with connection densities below 20%; ii) Prob CSD can achieve comparable performance to Det CSD, if streamline thresholding is enforced; and iii) CSD-based methods generally outperform tensor-based tractography.

It has been suggested that connectomes should be reconstructed with an FN-to-FP ratio of approximately 2 (10). We therefore investigated the FN-to-FP ratio as a complementary measure to assess algorithm performance in addition to the F-measure which suggests that streamline threshold is mandatory to achieve a 2:1 FN-to-FP ratio (Supporting Information; Figure S8).

Complexity Estimation

Having evaluated a range of connectome phantoms with different connection densities, fiber curvatures and network topologies, we next sought to identify which of our phantoms most resembled the human connectome by comparing their complexity with *in vivo* dMRI data. Figure 7 shows C_v (blue bars) and C_F (red) for our phantoms and the HCP dMRI data. It can be seen that complexity increases monotonically with density, since the number of intersecting fibers also increases with density (Figure 7(a)). The complexity estimated under different FOD thresholds using the supervised response function are shown in Figure 7(b) whereas the estimation using unsupervised response function can be found in Supporting Information S8. For bedpostx based complexity estimation, the volume fraction of fibers was thresholded to estimate the number of multiple fibers, shown in Figure 7(c). A FOD threshold of 0.1 and a volume fraction of 0.05 has been previously recommended (38, 88). Based on these suggested thresholds, we estimated $C_v = 0.52$ and $C_F = 0.71$ with MSMT. With bedpostx, we estimated $C_v = 0.46$ and $C_F = 0.67$ (dashed boxes in Figure 7(b) and 7(c) respectively). We were able to generate this level of complexity in our phantoms when using a ground truth connection density of approximately 10% (dashed box in Figure 7(a)). This provided a rationale for centering the range of ground truth connection densities evaluated at 10%.

Table 2 shows the maximum F-measure achieved by each tractography algorithm for the set of phantoms with a ground truth connection density of 10%, with and without application of a streamline threshold. We also show the number of streamlines that need to be discarded to achieve the maximum F-measure and to obtain a 2:1 FN-to-FP ratio. From Table 2, a clear difference in thresholds for maximum F-measure and FN-to-FP suggests that there is no universal choice of threshold that can be used to study the connectome. The significance (two-sample t-test) and percentage difference in F-measure (with and without threshold) are reported in Supporting Information Table S3.

Fiber Length Estimation

Prob CSD yielded streamlines with rippled trajectories (Figure 4). We considered this rippling behavior an artifact, given that it was evident in simulated fibers with an entirely linear trajectory. We therefore aimed to quantify the magnitude of the artifact by comparing the lengths of the reconstructed and ground truth streamlines. We generated 100 phantoms with the same parameters as in Methods section except that each phantom was limited to only one fiber. This ensured the absence of interference from other fibers. Fiber length was estimated as the average length of all streamlines produced by tractography. Length bias was then defined as the difference between the ground truth and tractography estimated fiber length, shown in Figure 8. Prob CSD introduced the largest bias, whereas tensor-based tractography and Det CSD produced comparable bias. The streamlines produced by Prob CSD could be post-processed, smoothing or filtering, to potentially reduce the length bias under the constraint of minimal change in spatial position or geometry of the fiber bundle. However, these post-processing steps increase the computational burden of the tractography pipeline and necessitate somewhat arbitrary choices about the extent to which the streamline trajectory is smoothed.

Spherical Connectome Phantom

Two-dimensional connectome phantoms might be considered unrealistic given that brain networks are embedded in three-dimensional space. To investigate the impact of omitting a spatial dimension in our models, we developed a simple three-dimensional connectome phantom in which nodes were positioned on the surface of a sphere and fiber bundles traversed the sphere's interior. Fiber bundle generation and simulation of synthetic dMRI data followed the same models described above for the two-dimensional case. We generated a spherical phantom comprising 60 nodes and a connection density of 15% to achieve $C_v = 0.5$ and $C_F = 0.7$. We found that our key conclusions were consistent between the two- and three-dimensional phantoms; namely, Det CSD produced the highest F-measure and probabilistic methods were hampered by an abundance of false positive connections. Supporting Information S9 provides further details about the simulation and analysis of the spherical phantom.

Discussion

In this study, we developed numerical connectome phantoms to specifically evaluate the performance of four tractography algorithms with respect to reconstructing binary connectivity matrices. The connectome phantoms we generated were based on realistic ground truth connectivity matrices and matched to the fiber complexity of *in vivo* dMRI data. We utilized our phantoms to evaluate the performance of two popular implementations (tensor-based and multi-fiber) of the two broadest classes of tractography algorithms (deterministic and probabilistic). Many alternative implementations of these tractography algorithms have been developed, rendering an exhaustive evaluation intractable, although the recent study of Maier-Hein and colleagues (15) provides a very comprehensive assessment. Therefore, it is important to remark that our conclusions do not necessarily generalize to all available algorithms.

The primary aim of this study was to assess whether deterministic or probabilistic tractography is best suited to connectome mapping. We found that multi-fiber deterministic tractography (Det CSD) provided superior performance in mapping connectome phantoms that were matched to the fiber complexity of *in vivo* dMRI data. However, for connectome phantoms with a fiber complexity that exceeded the complexity of *in vivo* dMRI data, we found that multi-fiber probabilistic tractography (Prob CSD) outperformed the other three algorithms evaluated. The performance of Prob CSD deteriorated for the lower complexity phantoms due to reconstruction of an abundance of spurious (false positive) connections. While connectome thresholding substantially reduced this problem, selecting an optimal threshold to eliminate connections with the fewest streamlines is challenging. We thus recommend stringent connectome thresholding when using Prob CSD for connectome mapping and advocate Det CSD as a viable alternative. This deviates from conventional tractography dogma suggesting that probabilistic algorithms are generally superior to their deterministic counterparts. While this may indeed be true for classic applications of tractography that involve virtual dissection of individual fibers, our findings suggest that probabilistic algorithms are not necessarily superior for high-throughput applications such as connectomics. More specifically, we hypothesize that the fiber complexity of the human brain might not be sufficiently high to warrant probabilistic tractography algorithms. Probabilistic algorithms were only warranted when fiber complexity substantially exceed the complexity of the human brain.

We generated connectome phantoms with varying connection density and fiber complexity (Figure 7). The ground truth connectivity matrices were generated using a generative network model (Methods) to produce networks with topological attributes that are characteristic of nervous systems. While the connectome phantoms comprised only 25 nodes, we have verified that the ground truth topologies comprised hubs and evidence of small-world organization. Specifically, high clustering between nodes and the presence of short path lengths were verified by calculating the small-world index for the

generated networks. The calculated index was always greater than 1, e.g. phantom with connection density of 20% had small-world index ≈ 1.2.

We matched the fiber complexity (C_v and C_F) of our connectome phantoms to the complexity of *in vivo* dMRI data. Jeurissen and colleagues (88) used classical FA based single-fiber response function to estimate the FODs using CSD and suggested a complexity $C_v = 90\%$ for single-shell dMRI data. In this study, the response function was calculated using supervised (87) and unsupervised co-registered T1 image (89) methodologies for a HCP multi-shell dataset of five subjects. The FODs were estimated using the MSMT technique, which suggested $C_v = 52\%$. This difference in complexity is due to the overestimation of number of crossings in a single-shell dataset used by Jeurissen and colleagues (88). Using the bedpostx fiber estimation method, Jeurissen and colleagues (88) suggested $C_v = 63\%$, whereas using the same methodology and parameters the current study approximated $C_v = 46\%$. This dissimilarity is again due to the usage of a multi-shell dataset in the current study.

In the following, we specifically discuss results pertaining to the connectome phantoms that were most representative of the *in vivo* dMRI data. The connection density of these phantoms was 10%. The classical tensor model cannot resolve multiple fiber configurations, which resulted in the lowest TPR, regardless of the tractography algorithm (Figure 5). In contrast, multi-fiber models (CSD) yielded the highest TPR. The FPR of Det CSD coincides with the tensor model (Figure 5(e)), but its TPR is substantially higher. Both Det and Prob Tensor have similar TPR and FPR, whereas a substantial difference can be observed for Det and Prob CSD. Det CSD has comparatively low TPR and FPR to Prob CSD, whereas Prob CSD has the highest TPR but at the expense of high FPR.

Is connectome thresholding necessary? Zalesky and colleagues (10) recommended the 2:1 rule of thumb for a connectome i.e. FNs should be twice the number of FPs. This ratio is not achievable if the generated streamlines are not thresholded (Table 2). Like F-measure, Prob CSD required the highest streamline threshold to achieve FN/FP ratio of 2:1 (Table 2). The optimal threshold value was obvious here because the ground truth was known, but for *in vivo* data approximation of this threshold is challenging.

The tractography algorithms are dependent on angle threshold and variation in outcome is observed by changing angle threshold (39, 63, 71, 77). The investigate the impact of angle threshold, the nominal angle thresholds were swapped between the algorithms, 9^o with 45^o (Supporting Information S7) and the simulations were repeated. Det CSD with default angle threshold however produced the highest F-measure, though the outcome of Prob CSD notably improved (Supporting Information; Figure S10). The SNR and b-value were also varied but a substantial change in the performance of Det CSD could not be observed.

Streamline weighting and filtering methods such as COMMIT (90) and SIFT (91) can potentially improve the accuracy of connectome reconstruction. However, SIFT can potentially yield spurious between-group differences when comparing streamline counts across all pairs of nodes with methods such as the network-based statistic (92). An example of this phenomenon is shown in Supplementary Figure S14.

Several important limitations require consideration. First, conclusions derived from numerical phantoms do not necessarily generalize to *in vivo* dMRI data. In this respect, our connectome phantoms are relatively simplistic with respect to fiber geometry and spatial embedding of the connectome. For example, our phantoms are effectively two-dimensional with nodes positioned along a circular circumference, whereas the human brain is three-dimensional with a convoluted cortical geometry. However, we ensured that that our two-dimensional phantoms were matched to the fiber complexity of three-dimensional *in vivo* dMRI data. Moreover, we developed a three-dimensional spherical phantom and found that inclusion of a third spatial dimension is consistent with the conclusions of our performance evaluation. Future work will focus on developing three-dimensional connectome phantoms with fiber configurations that resemble the human brain more closely. Second, for computational tractability, our connectome phantoms comprised only 25 nodes. In practice, connectome mapping is undertaken with a greater number of nodes. Despite this, we found that the topology of our ground truth connectivity matrices exhibited topological attributes that are characteristic of nervous systems. Thirdly, although our connectome phantoms were produced using generative models (77), which aim to recapitulate brain network topologies, the geometrical configuration of the simulated fiber bundles does not map to the true geometry of the brain fiber bundles. The impact of fiber bundle geometry has previously been well studied (61, 62, 64-66, 69, 72, 74). Fourth, we studied binary connectivity matrices (absence/presence of a connection), whereas connectome mapping also focuses on the strength of connections (i.e. number of streamlines per connection) in practice. But the connectivity strength is dependent on data quality, tractography algorithm and its inherent limitations (6). Finally, we evaluated only four tractography algorithms, which are not necessarily representative of alternative algorithms. Therefore, our conclusions do not necessarily generalize to all implementations of deterministic and probabilistic tractography.

In conclusion, we developed numerical connectome phantoms to evaluate whether deterministic and probabilistic tractography algorithms are suited to connectome mapping. We identified which of our phantoms were most closely matched to the fiber complexity of *in vivo* dMRI data and found that multi-fiber deterministic tractography was best able to recover the ground truth connectivity matrices for these phantoms. Single-fiber methods based on the diffusion tensor generally performed poorly in most cases. We showed that the performance of probabilistic tractography is burdened by a high FPR. In the future, the use of multi-shell multi-tissue methods (85) can potentially reduce the FPR of probabilistic

tractography. In the meantime, we recommend stringent connectome thresholding if probabilistic tractography is used to map connectomes.

Acknowledgements

The authors do not have any conflicts of interest to declare.

Data were provided (in part) by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

1. Sporns O, Tononi G, Kötter R. The human connectome: a structural description of the human brain. PLoS Comput Biol 2005; 1: 245-251. *doi: 10.1371/journal.pcbi.0010042*

2. Hagmann P, Kurant M, Gigandet X, Thiran P, Wedeen V, Meuli R, Thiran, J. Mapping human whole-brain structural networks with diffusion MRI. PLoS ONE 2007; 2: e597. *doi*: *10.1371/journal.pone.0000597*

3. Bullmore E, Bassett D. Brain Graphs: Graphical Models of the Human Brain Connectome. Annu Rev Clin Psychol 2011; 7: 113-140. *doi*: 10.1146/annurev-clinpsy-040510-143934

4. Behrens T, Sporns O. Human connectomics. Curr Opin in Neurobiol 2012; 22: 144-153. *doi*: *10.1016/j.conb.2011.08.005*

5. Fornito A, Zalesky A, Breakspear, M. Graph analysis of the human connectome: Promise, progress, and pitfalls. NeuroImage 2013; 80: 426-444. *doi*: *10.1016/j.neuroimage.2013.04.087*

6. Sotiropoulos SN, Zalesky A. Building connectomes using diffusion MRI: why, how and but. NMR Biomed 2017. *doi*: *10.1002/nbm.3752*

7. Jeurissen B, Descoteaux M, Mori S, Leemans A. Diffusion MRI fiber tractography of the brain. NMR Biomed 2017; e3785. *doi*: *10.1002/nbm.3785*

8. Catani M, Howard R, Pajevic S, Jones D. Virtual in Vivo Interactive Dissection of White Matter Fasciculi in the Human Brain. NeuroImage 2002; 17: 77-94. *doi*: *10.1006/nimg.2002.1136*

9. Mori S, Crain B, Chacko V, Van Zijl P. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Ann Neurol 1999; 45: 265-269. *doi*: *10.1002/1531-8249(199902)45:2<265::AID-ANA21>3.0.CO;2-3*

10. Zalesky A, Fornito A, Cocchi L, Gollo LL, Heuvel MP, Breakspear M. Connectome sensitivity or specificity: which is more important? NeuroImage 2016; 142: 407-420. *doi*: *10.1016/j.neuroimage.2016.06.035*

11. Jbabdi S, Johansen-Berg H. Tractography: Where do we go from here? Brain Connect 2011; 1: 169-183. *doi: 10.1089/brain.2011.0033*

12. Smith R, Tournier J, Calamante F, Connelly A. Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. NeuroImage 2012; 62: 1924-1938. *doi: 10.1016/j.neuroimage.2012.06.005*

13. Reveley C, Seth A, Pierpaoli C, Silva A, Yu D, Saunders R, Leopold D, Ye F. Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion MR tractography. Proc Natl Acad Sci U S A 2015; 112: 2820-2828. *doi: 10.1073/pnas.1418198112*

14. St-Onge, E., Daducci, A., Girard, G., Descoteaux, M. Surface-enhanced tractography (SET). NeuroImage 2018; 169: 524–539. *doi: /10.1016/j.neuroimage.2017.12.036*

15. Maier-Hein KH, Neher PF, Houde JC, et al. The challenge of mapping the human connectome based on diffusion tractography. Nature Commun 2017; 8: e1349. *doi: 10.1038/s41467-017-01285-x*

16. Petrov D, Ivanov A, Faskowitz J, Gutman B, Moyer D, Villalon J, Jahanshad N, Thompson P. Evaluating 35 methods to generate structural connectomes using pairwise classification. Med Image Comput Comput Assist Interv (MICCAI) 2017:515-522

17. Alexander AL, Lee JE, Lazar M, Field AS. Diffusion tensor imaging of the brain. Neurotherapeutics 2007; 4: 316-329. *doi: 10.1016/j.nurt.2007.05.011*

18. Koerte IK, Muehlmann M. Diffusion Tensor Imaging. MRI Psychiatry 2014: 77-86

19.O'Donnell LJ, Westin C. An introduction to diffusion tensor image analysis. Neurosurg Clin N Am 2011; 22: 185-196. *doi*: *10.1016/j.nec.2010.12.004*

20. Tournier JD, Calamante F, Gadian DG, Connelly A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage 2004; 23: 1176-1185. *doi*: *10.1016/j.neuroimage.2004.07.037*

21. Tournier J, Calamante F, Connelly A. Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. NeuroImage 2007; 35: 1459-1472. *doi*: *10.1016/j.neuroimage.2007.02.016*

22. Canales-Rodríguez EJ, Melie-García L, Iturria-Medina Y. Mathematical description of q-space in spherical coordinates: Exact q-ball imaging. Magn Reson Med 2009; 61: 1350-1367. *doi*: *10.1002/mrm.21917*

23. Descoteaux M, Angelino E, Fitzgibbons S, Deriche R. Regularized, fast, and robust analytical Q-ball imaging. Magn Reson Med 2007; 58(3): 497-510. *doi*: *10.1002/mrm.21277*

24. Hess CP, Mukherjee P, Han ET, Xu D, Vigneron DB. Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. Magn Reson Med 2006; 56: 104-117. *doi*: *10.1002/mrm.20931*

25. Tuch DS. Q-ball Imaging, Magn Reson Med 2004; 52: 1358-1372. *doi*: *10.1002/mrm.20279*

26. Wedeen VJ, Hagmann P, Tseng WI, Reese TG, Weisskoff RM. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn Reson Med 2005; 54(6): 1377-1386. *doi*: *10.1002/mrm.20642*

27. Behrens T, Berg H, Jbabdi S, Rushworth M, Woolrich M. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? NeuroImage 2007; 34: 144-155. *doi*: *10.1016/j.neuroimage.2006.09.018*

28. Schilling K, Janve V, Gao Y, Stepniewska I, Landman B, Anderson A. Comparing diffusion MRI with the fiber architecture and tract density of gyral blades. Proceedings of the ISMRM Meeting 2016; 924.

29. Basser P, Pajevic S, Pierpaoli C, Duda J, Aldroubi A. In vivo fiber tractography using DT-MRI data. Magn Reson Med 2000; 44: 625-632. *doi*: *10.1002/1522-2594(200010)44:4<625::AID-MRM17>3.0.CO;2-O*

30. Conturo T, Lori N, Cull T, Akbudak E, Snyder A, Shimony J, McKinstry R, Burton H, Raichle M. Tracking neuronal fiber pathways in the living human brain. Proc Natl Acad Sci U S A 1999; 96: 10422-10427

31. Yeh F, Verstynen, TD, Wang Y, Fernández-Miranda JC, Tseng WI. Deterministic diffusion fiber tracking improved by quantitative anisotropy. PLoS ONE 2013; 8. *doi*: 10.1371/journal.pone.0080713

32. Jones DK. Studying connections in the living human brain with diffusion MRI. Cortex 2008; 44: 936-952.*doi: 10.1016/j.cortex.2008.05.002*

33. Descoteaux M, Deriche R, Knosche T, Anwander A. Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Trans Med Imaging 2009; 28: 269-286. *doi*: *10.1109/TMI.2008.2004424*

34. Jones DK. Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI. Imaging Med 2010; 2: 341–355. *doi*: *10.2217/iim.10.21*

35. Tournier JD, Calamante F, Connelly A. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. In Proceedings of the 18th Annual Meeting of ISMRM 2010. p. 1670.

36. Tournier J, Calamante F, Connelly A. MRtrix: Diffusion tractography in crossing fiber regions. Int J Imaging Sys Technol 2012; 22: 53-66. http://www.mrtrix.org/ *doi*: *10.1002/ima.22005*

37. Jones D. Tractography gone wild: probabilistic fibre tracking using the wild bootstrap with diffusion tensor MRI. IEEE Trans Med Imaging 2008; 27: 1268-1274. *doi*: *10.1109/TMI.2008.922191*

38. Behrens TEJ, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 2003; 50 (5): 1077 – 1088. *doi*: *10.1002/mrm.10609*

39. van den Heuvel M, de Lange S, Zalesky A, Seguin C, Yeo B, Schmidt R. Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. NeuroImage 2017; 152: 437-449. *doi*: *10.1016/j.neuroimage.2017.02.005*

40. Roberts J, Perry A, Roberts G, Mitchell P, Breakspear M. Consistency-based thresholding of the human connectome. NeuroImage 2017; 145: 118-129. *doi*: *10.1016/j.neuroimage.2016.09.053*

41. Azadbakht H, Parkes L, Haroon H, Augath M, Logothetis N, de Crespigny A, D'Arceuil H, Parker G. Validation of high-resolution tractography against in vivo tracing in the macaque visual cortex. Cereb Cortex 2015; 25: 4299-4309. *doi*: *10.1093/cercor/bhu326*

42. Knösche TR, Anwander A, Liptrot M, Dyrby TB. Validation of Tractography: Comparison with manganese tracing. Hum Brain Map 2015; 36: 4116- 4134. *doi*: *10.1002/hbm.22902*

43. Thomas C, Ye F, Irfanoglu M, Modi P, Saleem K, Leopold D, Pierpaoli C. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. Proc Natl Acad Sci U S A 2014; 111: 16574-16579. *doi*: *10.1073/pnas.1405672111*

44. Aganj C, Lenglet N, Jahanshad E, Yacoub N, Harel PM, Thompson G, Sapiro A. Hough transform global probabilistic approach to multiple-subject diffusion MRI tractography Med Image Anal; 2014, 15: 414-425. *doi: 10.1016/j.media.2011.01.003*

45. Fillard P, Poupon C, Mangin JF. A novel global tractography algorithm based on an adaptive spin glass model. In Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention 2009: 927-934

46. Jbabdi S, Woolrich MW, Andersson JLR, Behrens TEJ. A Bayesian framework for global tractography. Neuroimage 2007; 37: 116-129. *doi: 10.1016/j.neuroimage.2007.04.039*

47. Lemkaddem A, Skiöldebrand D, A. Dal Palú, Thiran FP, Daducci A. Global tractography with embedded anatomical priors for quantitative connectivity analysis. Front Neurol 2014; 5: 232. *doi: 10.3389/fneur.2014.00232*

48. Mangin JF, Fillard P, Cointepas Y, Bihan, DL, Frouin V, Poupon C. Toward global tractography. Neuroimage 2013; 80: 290-296. *doi: 10.1016/j.neuroimage.2013.04.009*

49. Reisert M, Kellner E, Dhital B, Hennig J, Kiselev VG. Disentangling micro from mesostructure by diffusion MRI: a Bayesian approach. NeuroImage 2017; 147: 964-975. *doi: 10.1016/j.neuroimage.2016.09.058*

50. Konopleva L, Il'yasov KA, Skibbe H, Kiselev VG, Kellner E, Dhital B, Reisert M. Modelfree global tractography. NeuroImage 2018. *doi: 10.1016/j.neuroimage.2018.03.058*

51. Reisert M, Mader I, Anastasopoulos C, Weigel M, Schnell S, Kiselev V. Global fiber reconstruction becomes practical. NeuroImage 2011; 54: 955-962. *doi: 10.1016/j.neuroimage.2010.09.016*

52. Iturria-Medina Y, Canales-Rodríguez EJ, , Melie-García L , Valdés-Hernándeza PA, Alemán Gómez Y, Sánchez-Bornot JM. Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. NeuroImage 2007; 63: 645-660. *doi: 10.1016/j.neuroimage.2007.02.012*

53. Iturria-Medina Y, Sotero RC, Canales-Rodríguez EJ, Alemán Gómez, Y, Melie-García L. Studying the human brain anatomical network via diffusion-weighted MRI and Graph Theory. NeuroImage 2008; 40: 1064–1076. *doi :10.1016/j.neuroimage.2007.10.060*

54. Zalesky A, Fornito A. A DTI-derived measure of cortico-cortical connectivity. IEEE Trans Med Imaging 2009; 28: 1023-1036. *doi*: *10.1109/TMI.2008.2012113*

55. Zalesky A. DT-MRI fiber tracking: a shortest paths approach. IEEE Trans Med Imaging 2008: 27; 1458-1471. *doi:* 10.1109/TMI.2008.923644

56. Neher PF, Côté M,Houde JC, Descoteaux, Maier-Hein KH. Fiber tractography using machine learning. NeuroImage 2017; 158: 417-429. *doi: 10.1016/j.neuroimage.2017.07.028*

57. Poulin P, Côté M,Houde JC, Petit L, Neher PF, Maier-Hein KH, Larochelle H, Descoteaux Maxime. Learn to track: Deep Learning for tractography. Med Image Comput Comput Assist Interv (MICCAI) 2017: 540-547

58. Staempfli P, Jaermann T, Crelier GR, Kollias S, Valavanis A,Boesiger P. Resolving fiber crossing using advanced fast marching tractography based on diffusion tensor imaging. NeuroImage 2006; 30: 110-120. *doi: 10.1016/j.neuroimage.2005.09.027*

59. Parker GJM, Wheeler-Kingshott CAM, Barker GJ. Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging. IEEE Trans Med Imaging 2002; 21: 505-512. *doi: 10.1109/TMI.2002.1009386*

60. Jbabdi S, Sotiropoulos SN, Haber SN, Essen DC, Behrens TE. Measuring macroscopic brain connections in vivo. Nat Neurosci 2015; 18: 1546-1555. *doi*: *10.1038/nn.4134*

61. Calabrese E, Badea A, Cofer G, Qi Y, Johnson GA. A Diffusion MRI Tractography Connectome of the Mouse Brain and Comparison with Neuronal Tracer Data. Cereb Cortex 2015; 25: 4628-4637. *doi*: *10.1093/cercor/bhv121*

62. Donahue CJ, Sotiropoulos SN, Jbabdi S, et al. Using Diffusion Tractography to Predict Cortical Connection Strength and Distance: A quantitative comparison with tracers in the monkey. J Neurosci 2016; 36: 6758-6770. *doi*: *10.1523/JNEUROSCI.0493-16.2016*

63. Fillard P, Descoteaux M, Goh A, et al. Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. NeuroImage 2011; 56: 220-234. *doi*: *10.1016/j.neuroimage.2011.01.032*

64. Poupon C, Rieul B, Kezele I, Perrin M, Poupon F, Mangin J. New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging (HARDI) models. Magn Reson Med 2008; 60(6): 1276-1283. *doi: 10.1002/mrm.21789*

65. Cho K, Yeh C, Tournier J, Chao Y, Chen J, Lin C. Evaluation of the accuracy and angular resolution of q-ball imaging. NeuroImage 2008; 42: 262-271. *doi: 10.1016/j.neuroimage.2008.03.053*

66. Moussavi-Biugui A, Stieltjes B, Fritzsche K, Semmler W, Laun FB. Novel spherical phantoms for Q-ball imaging under in vivo conditions. Magn Reson Med 2010; 65(1): 190-194. *doi*: *10.1002/mrm.22602*

67. Tournier JD, Yeh C, Calamante F, Cho K, Connelly A, Lin C. Resolving crossing fibres using constrained spherical deconvolution: Validation using diffusion-weighted imaging phantom data. NeuroImage 2008; 42: 617-625. *doi*: *10.1016/j.neuroimage.2008.05.002*

68. Daducci A, Canales-Rodriguez EJ, Descoteaux M, et al. Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI. IEEE Trans Med Imaging 2014; 33: 384-399. *doi*: *10.1109/TMI.2013.2285500*

69. Panagiotaki E, Schneider T, Siow B, Hall MG, Lythgoe MF, Alexander DC. Compartment models of the diffusion MR signal in brain white matter: A taxonomy and comparison. NeuroImage 2012; 59: 2241-2254. *doi*: *10.1016/j.neuroimage.2011.09.081*

70. Tournier J, Calamante F, Connelly A. Determination of the appropriate bvalue and number of gradient directions for high-angular-resolution diffusion-weighted imaging. NMR Biomed 2013; 26: 1775-1786. *doi*: *10.1002/nbm.3017*

71. Neher PF, Laun FB, Stieltjes B, Maier-Hein KH. Fiberfox: Facilitating the creation of realistic white matter software phantoms. Magn Reson Med 2014; 72(5): 1460-1470. *doi*: *10.1002/mrm.25045*

72. Close TG, Tournier J, Calamante F, Johnston LA, Mareels I, Connelly A. A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. NeuroImage 2009; 47: 1288-1300. *doi*: *10.1016/j.neuroimage.2009.03.077*

73. Caruyer E, Daducci A, Descoteaux M, Houde JC, Thiran JP, Verma R. Phantomas: a flexible software library to simulate diffusion MR phantoms. In Proceedings of International on Magnetic Resonance in Medicine (ISMRM) 2014.

74. Daducci A, Caruyer E, Descoteaux M, Houde JC, Thiran JP. IEEE International Symposium on Biomedical Imaging (ISBI) Reconstruction Challenge 2013. http://hardi.epfl.ch/static/events/2013_ISBI/

75. Côté M, Girard G, Boré A, Garyfallidis E, Houde J, Descoteaux M. Tractometer: Towards validation of tractography pipelines. Med Image Anal 2013; 17: 844-857. *doi*: *10.1016/j.media.2013.03.009*

76. Neher PF, Descoteaux, M, Houde J, Stieltjes B, Maier-Hein KH. Strengths and weaknesses of state of the art fiber tractography pipelines – A comprehensive in-vivo and phantom evaluation study using Tractometer. Med Image Anal 2015; 26: 287-305. *doi*: *10.1016/j.media.2015.10.011*

77. Girard G, Whittingstall K, Deriche R, Descoteaux M. Towards quantitative connectivity analysis: reducing tractography biases. NeuroImage 2014; 98: 266-278. *doi*: *10.1016/j.neuroimage.2014.04.074*

78. Betzel R, Bassett DS. Generative Models for Network Neuroscience: Prospects and Promise. J. Royal Society of Interface 2017; 14. *doi: 10.1098/rsif.2017.0623*

79. Betzel R, Avena-Koenigsberger A, Goñi J, et al. Generative models of the human connectome. NeuroImage 2016; 124: 1054-1064. *doi*: *10.1016/j.neuroimage.2015.09.041*

80. Alexander DC. A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features. Magn Reson Med 2008; 60(2): 439-448. *doi*: *10.1002/mrm.21646*

81. Wu X, Zhou J, Xie M, Luo D. A Bayesian constraint stochastic framework for DT-MRI white matter Fiber Tractography. J. Electronics and Information Technology 2010; 32(8): 1786-1791. *doi*: *10.3724/SP.J.1146.2009.01120*

82. Friman O, Westin C. Uncertainty in White Matter Fiber Tractography. Med Image Comput Comput Assist Interv (MICCAI) 2005: 107-114.

83. Andersen AH. On the Rician distribution of noisy MRI data. Magn Reson Med 1996; 36(2): 331-332. *doi*: *10.1002/mrm.1910360222*

84. Ferizi U, Schneider T, Panagiotaki E, Nedjati-Gilani G, Zhang H, Wheeler-Kingshott CAM, Alexander DC. A ranking of diffusion MRI compartment models with in vivo human brain data. Magn Reson Med 2014; 72: 1785-1792. *doi: 10.1002/mrm.25080*

85. Nedjati-Gilani GL, Schneider T, Hall MG, et al. Machine learning based compartment models with permeability for white matter microstructure imaging. NeuroImage 2017; 150: 119-135. *doi: 10.1016/j.neuroimage.2017.02.013*

86. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: An overview. NeuroImage 2013; 80: 62-79. *doi*: *10.1016/j.neuroimage.2013.05.041*

87. Jeurissen B, Tournier JD, Dhollander T, Connelly A, Sijbers J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. NeuroImage 2014; 103: 411-426. *doi*: *10.1016/j.neuroimage.2014.07.061*

88. Jeurissen B, Leemans A, Tournier J, Jones D, Sijbers J. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. Hum Brain Map 2012; 34: 2747-2766. *doi*: *10.1002/hbm.22099*

89. Dhollander T, Raffelt D, Connelly A. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image. ISMRM Workshop on Breaking the Barriers of Diffusion MRI 2016: 5

90. Daducci A, Dal Palu A, Lemkaddem A, Thiran, JP. COMMIT: Convex Optimization Modeling for Micro-structure Informed Tractography. IEEE Trans Med Imaging 2015: 34: 246-257. *doi*: *10.1109/TMI.2014.2352414*

91. Smith RE, Tournier JD, Calamante F, Connelly A. The effects of SIFT on the reproducibility and biological accuracy of the structural connectome. NeuroImage 2015: 104; 253–265. *doi: 10.1016/j.neuroimage.2014.10.004*

92. Zalesky A, Fornito A, Bullmore ET. Network-based statistic: identifying differences in brain networks. NeuroImage 2010: 53; 1197-1207. *doi*: *10.1016/j.neuroimage.2010.06.041*

Main Figures

Figure 1. Schematic of phantom generation for a 10-node connectome. Nodes are delineated on the circumference of a circle (colored segments) and interconnected via fiber bundles. A ground truth connectivity matrix dictates whether a fiber bundle is present/absent between each pair of nodes. Fiber bundles are modeled as curved tubular structures with fixed diameter. A multi-compartment dMRI model is used to generate dMRI signals for each voxel, such that each fiber bundle traversing a given voxel contributes an independent anisotropic compartment to the signal. Finally, Rician noise is added. **(b)** Example fiber bundle segments comprising one (left) and three (right) control points. Fiber trajectories are denoted with solid black lines. Black circles denote endpoints, yellow circles represent control points situated at perpendicular distances *r* from the center of curvatures represented by green solid circles and red solid circles are the intermediate points generated using cubic-spline interpolation. To generate a tube, the fiber trajectory is extended perpendicularly to yield a prescribed bundle diameter. Gray shading denotes the tubular bundle **(c)** Example fiber bundle architecture for 10-node phantoms comprising 10 distinct fibers. Fiber color is determined by diameter. Fibers are tubular (three-dimensional) and fiber curvature is confined to the two-dimensions on which the circle resides.

Figure 2. Simulated dMRI images of a representative connectome phantom with connection density of 12% **(a-d)**, $d_{||}=1.5x10^{-3}$ mm^2/s and $d_{\perp}=0.2x10^{-3}$ mm^2/s for the zeppelin compartment and a diffusivity $D=0.9x10^{-3}$ mm^2/s for the ball compartment. Acquisition parameters listed in "**Methods**". **(a)** Non-diffusion weighted signal without noise, **(b)** FA image without noise, **(c)** non-diffusion weighted volume with noise (SNR=10db), **(d)** FA image with noise.

Figure 3. Tractograms of ground truth model **(a)** with connection density **(i)** 4% **(j)** 8% **(k)**12% **(l)** 16% **(m)** 20%. The ground truth tractography is generated using tensor deterministic tractography on individual fibers and merging the individual fibers to produce tractogram. Streamlines in the region encapsulated by the blue rectangular box are shown for **(b)** Det Tensor **(c)** Prob Tensor **(d)** Det CSD **(e)** Prob CSD. For visualization purposes, a ground truth reconstruction was determined by generating a separate dMRI image for each fiber. Each of these dMRI images only included the signal from a single fiber. Deterministic tensor-based tractography was performed independently on each dMRI image and the streamlines reconstructed for each fiber were merged to form a single set of streamlines **(a).**

Figure 4. Streamline trajectories **(a-e)**, diffusion tensors **(f)** and FODs **(g)** for a representative connectome phantom with connection density of 12%. Streamlines in the region encapsulated by

the blue rectangular box are shown for **(b)** Det Tensor **(c)** Prob Tensor **(d)** Det CSD **(e)** Prob CSD tractography algorithms.

Figure 5. Receiver operating characteristic (ROC) curves for connectome phantoms with connection density 2% - 20% **(a-j; increments of 2%).** Curves are parameterized by the streamline threshold. The rightmost point on each curve corresponds to a threshold of zero, while the most severe threshold corresponds to the leftmost point. Insets characterize behavior for false positives rates below 0.05. Note that Prob CSD typically generates a diffuse streamline distribution, resulting in TPR and FPR combinations that cannot be achieved with the other algorithms. This is evidenced by the red curve terminating at a higher position than the others. For the connection density of 10% **(e)**, the CSD-based algorithms (Det CSD, Prob CSD) yield better performance than their tensor-based counterparts (Det Tensor, Prob Tensor). If streamline thresholding is used, Prob CSD can achieve the same performance at Det CSD. Without thresholding, the FPR of Prob CSD is approximately 1 for a connection density of 20%.

Figure 6. F-measure as a function of the streamline threshold for connectome phantoms with connection density 2% - 20% **(a-j; increments of 2%)**. The leftmost point on each curve corresponds to the absence of a streamline threshold, where a connection is assumed to be present for any pair of regions interconnected by at least one streamline. For the connection density of 10% **(e)**, Det CSD outperforms the other three algorithms by a factor of approximately 2 if streamline thresholding is not used (leftmost data points). While the performance of Prob CSD can be substantially improved with thresholding, Det CSD yields the maximum F-measure across all threshold values for connection densities 2%-12% **(a-f)**.

Figure 7. Complexity of simulated and in-vivo dMRI data. Complexity was quantified with voxel (red bars: $\boldsymbol{C_v}$) and fiber (blue bars: $\boldsymbol{C_F}$). **(a)** Complexity of the connectome phantoms (simulated dMRI data) for varying connection density. Complexity of in-vivo dataset estimated using **(b)** MSMT **(c)** Bedpostx. Dashed boxes in **(b)** and **(c)** indicate the recommended FOD and volume fraction thresholds, respectively. Dashed box in **(a)** indicates the connection density that achieves a complexity that is best matched to the complexity of the *in vivo* data.

Figure 8. Distribution of bias in reconstructed fiber lengths. Bias was computed as the ground truth fiber length subtracted from the reconstructed fiber length. Reconstructed fiber length was estimated based on the average streamline length of all streamline comprising a fiber. Distribution of this bias across multiple fibers is shown for four tractography algorithms.

Main Table

Table 1: Tractography Parameters

Table 2: Performance of deterministic and probabilistic tractography algorithms for a connectome phantom with connection density of 10%. Performance quantified as the maximum F-measure achievable with and without streamline thresholding. The streamline threshold required to maximize the F-measure and achieve an FN/FP ratio of 2 are also shown.

Supporting Figures

Figure S1. (a) Example fiber with tangent vectors *v* represented in blue. To obtain an effective diameter of 5 voxels (*m*=4), the vectors *v* are replicated to neighboring voxels in 2 iterations **(b)** Iteration 1: The vectors *v* represented by blue are replicated to their neighboring voxels (orange) **(c)** Iteration 2: The vectors *v* in orange are replicated again to their neighboring voxels (green), resulting in a fiber with effective diameter of 5 voxels. Vectors are not replicated to voxels that already comprise a fiber. The length of the fiber is not affected as the end points of the fibers reside on the circular connectome and any replication outside the circumference is discarded.

Figure S2. Receiver operating characteristic (ROC) curves for ground truth model of connection density 10%, with b-value (s/mm^2) **(a)** 1000 **(b)** 2000 **(c)** 3000. Curves are parameterized by the streamline threshold

Figure S3. F-measure as a function of the streamline threshold for connectome phantoms of connection density 10%, with b-value (s/mm^2) **(a)** 1000 **(b)** 2000 **(c)** 3000

Figure S4. Receiver operating characteristic (ROC) curves for ground truth model of connection density 10%, with SNR (db) **(a)** 10 **(b)** 20 **(c)** 30. Curves are parameterized by the streamline threshold.

Figure S5. F-measure as a function of the streamline threshold for connectome phantoms of connection density 10%, with SNR (db) **(a)** 10 **(b)** 20 **(c)** 30

Figure S6. Receiver operating characteristic (ROC) curves for ground truth model of connection density 10%, with volume fraction f **(a)** 0 **(b)** 0.3 **(c)** 0.7

Figure S7. F-measure as a function of the streamline threshold for connectome phantoms of connection density 10%, with volume fraction f **(a)** 0 **(b)** 0.3 **(c)** 0.7

Figure S8. FN/FP ratio as a function of streamline threshold for connectome phantoms with connection density 2% - 20% **(a-j; increments of 2%)**. To achieve a prescribed FN/FP ratio, more streamlines must be thresholded as the connection density of the phantom is increased. Prob CSD mandates the most severe streamline threshold to achieve an FN/FP ratio of 2, whereas the other three algorithms achieve this recommended value for approximately the same streamline threshold.

Figure S9. Receiver operating characteristic (ROC) curves for ground truth model of connection density 10%. Curves are parameterized by the streamline threshold. Tractography with angle threshold **(a)** Tensor (Det and Prob) and Det CSD=9^o and Prob CSD=45^o **(b)** Tensor (Det and Prob) and Det CSD =45^o and Prob CSD=9^o.

Figure S10. F-measure as a function of the streamline threshold for connectome phantoms of connection density 10% Tractography with angle threshold **(a)** Tensor (Det and Prob) and Det CSD=9^{o} and Prob CSD=45^{o} **(b)** Tensor (Det and Prob) and Det CSD =45^{o} and Prob CSD=9^{o}.

Figure S11. Complexity of HCP dMRI dataset of five subjects estimated from MSMT using response function **(a)** supervised **(b)** unsupervised

Figure S12. (a) Division of a spherical surface into cells using 20 longitudinal and latitudinal bands. Nodes are defined by contiguous clusters of cells, with the red outline delineating an example node. **(b)** Tractogram of ground truth spherical phantom with connection density of 15%. The ground truth tractography was generated using tensor-based deterministic tractography on individual fibers and merging the individual fibers to produce the tractogram shown.

Figure S13. Performance evaluation based on the spherical connectome phantom with connection density 15% **(a)** Receiver operating characteristic (ROC) curve parameterized by the streamline threshold **(b)** F-measure as a function of the streamline threshold.

Figure S14. Toy example demonstrating the emergence of a spurious between-group difference after processing with SIFT (19). Each ellipse denotes the fiber orientation distribution (FOD). For Bundle 1, one of the FODs (colored red) encapsulates a substantially smaller volume than the others due to localized pathology. Streamlines encounter difficulty traversing this pathology, and thus Bundle 1 is traversed by half as many streamlines as Bundle 2. The total streamline density of Bundle 1 is therefore half that of Bundle 2. To reconcile the discrepancy between FOD volume and total streamline density between the two bundles, SIFT eliminates streamlines from Bundle 2. In particular, half of the streamlines traversing Bundle 2 are eliminated to minimize the cost function. However, Bundle 2 comprises no pathology. Therefore, if the SIFT-adjusted streamline count for Bundle 2 is compared between groups with and without pathology in Bundle 1, a spurious reduction in the streamline count will be found in Bundle 2 when in fact the true pathology is circumscribed to Bundle 1.

Supporting Tables

Table S1: F-measure for ground truth model of connection density 10%. The listed streamline threshold gives the maximum achievable F-Measure

Table S2: F-measure for ground truth model of connection density 10%. The listed streamline threshold gives the maximum achievable F-Measure

Table S3: Testing the null hypothesis of equality in F-measures between tractography algorithms. Leftmost tables show *p*-values for all pairs of tractography algorithms, where 0 indicates a *p*-value smaller than 0.00001. Rightmost tables show percentage difference in performance, as measured by the F-measure. Positive percentages indicate the method in the row outperformed the method in the column, and conversely for negative percentages. Tables S3.1 and S3.2 quantify performance without connectome thresholding, while Tables S3.3 and S3.4 quantify performance with application of the streamline threshold that maximized performance. All results pertain to connectome phantoms that were best matched to the *in vivo* dMRI data with respect to fiber complexity.

Table S3.1: P-values for F-measure in the absence of threshold

Table S3.2: Percentage difference in F-measure in the absence of threshold

Table S3.3: P-values for maximum F-measure computed using threshold

Table S3.4: Percentage difference in maximum F-measure computed using threshold

Table S4: Fraction of streamlines connecting TP's (valid connection), FP's (invalid connection) and terminating before reaching a node (no connection). Tractography generated 1,000,000 streamlines for a phantom

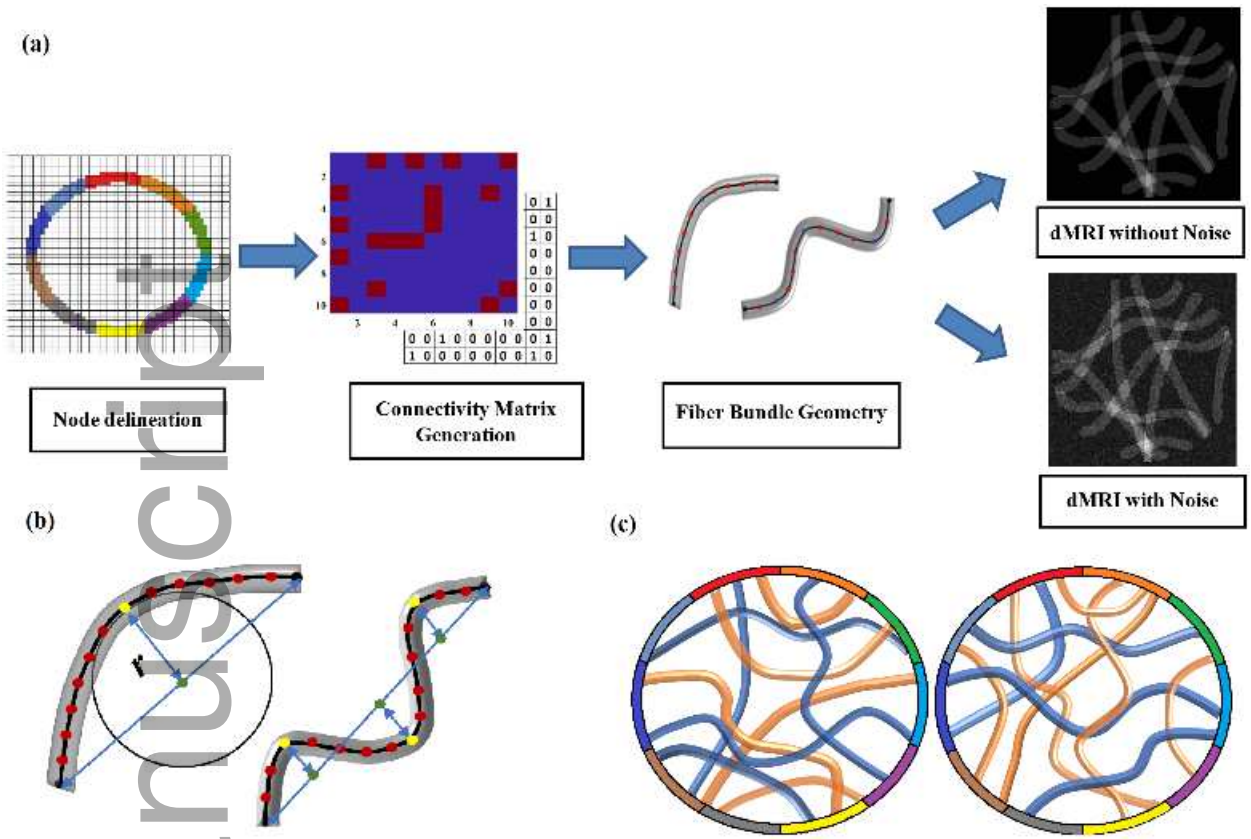
Tables

Table 1: Tractography Parameters

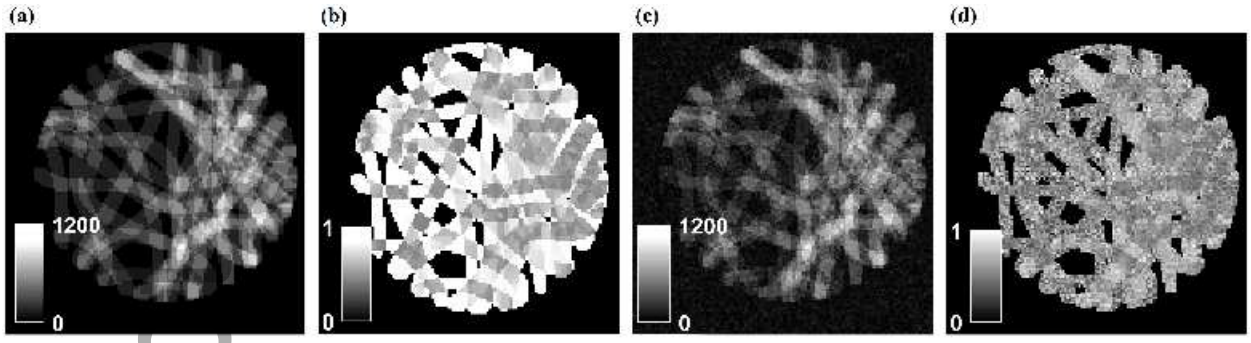
Tractography Algorithm and Fiber Reconstruction Model	Parameters	Values
Det Tensor Prob Tensor Det CSD	Step Size (0.1 x voxel size)	0.2
	Angle threshold (90^o x step size / voxel size)	9^o
	FA threshold	0.1
Prob CSD	Step Size (0.5 x voxel size)	1
	Angle threshold (90^o x step size / voxel size)	45^o
	FOD threshold	0.1

Table 2: Performance of deterministic and probabilistic tractography algorithms for a connectome phantom with connection density of 10%. Performance quantified as the maximum F-measure achievable with and without streamline thresholding. The streamline threshold required to maximize the F-measure and achieve an FN/FP ratio of 2 are also shown.

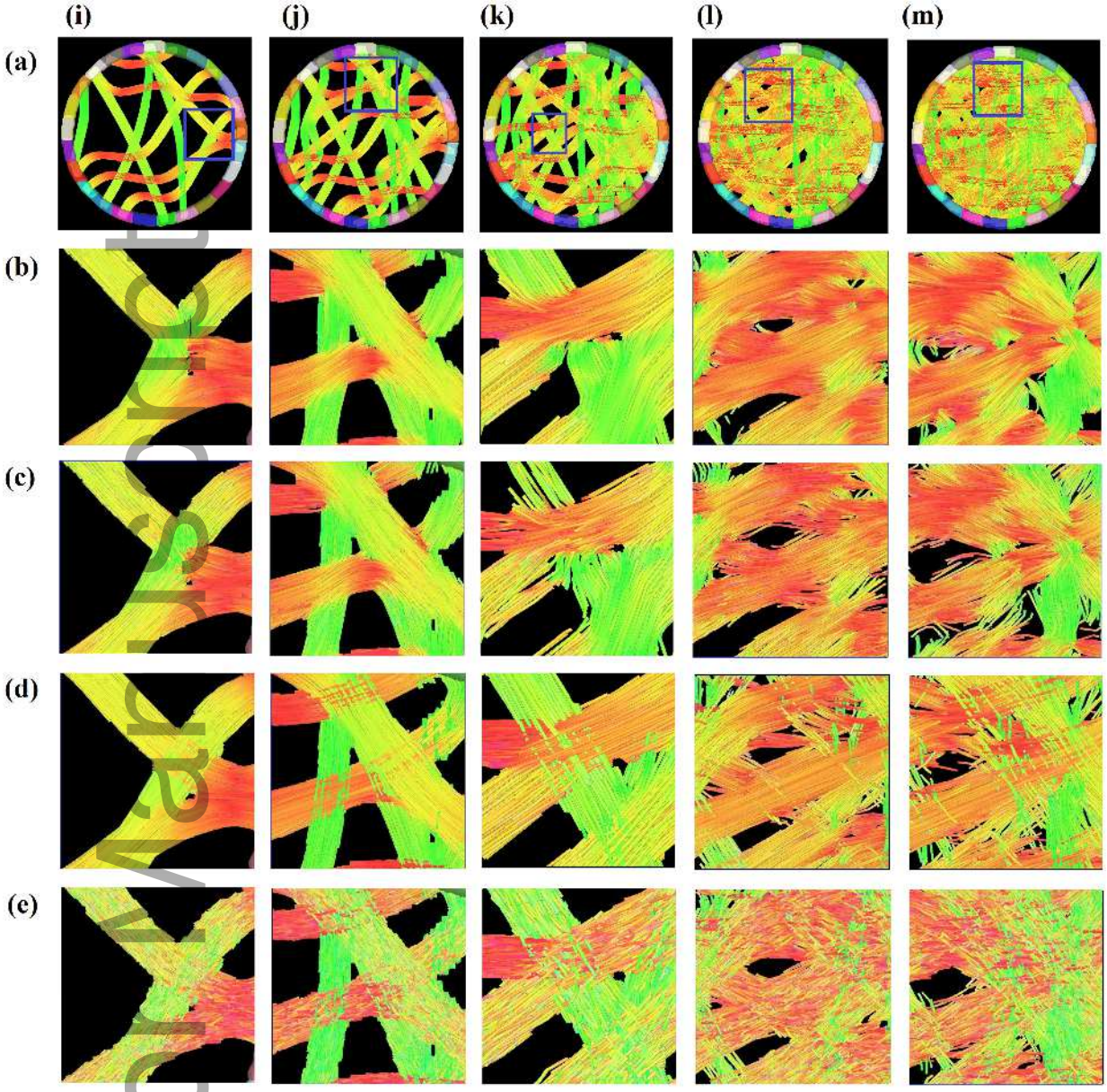
Algorithm	Maximum F-measure		Streamline Threshold for F-measure (%)	Streamline Threshold Required for FN/FP Ratio of 2:1 (%)
	Without Threshold	With Threshold		
Det Tensor	0.185	0.196	1	12
Prob Tensor	0.185	0.198	3	9
Det CSD	0.345	0.415	3	8
Prob CSD	0.192	0.382	14	30



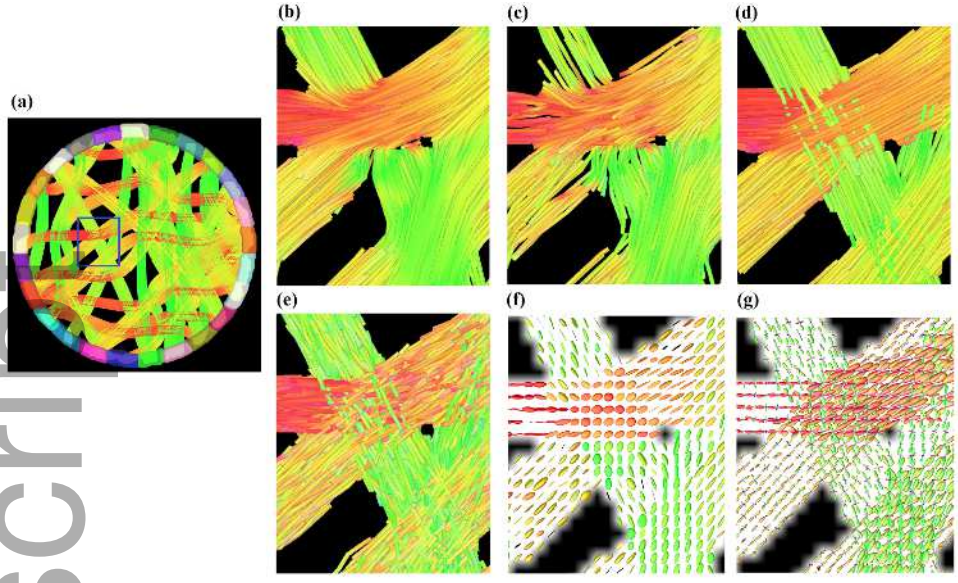
mrm_27471_f1.tif



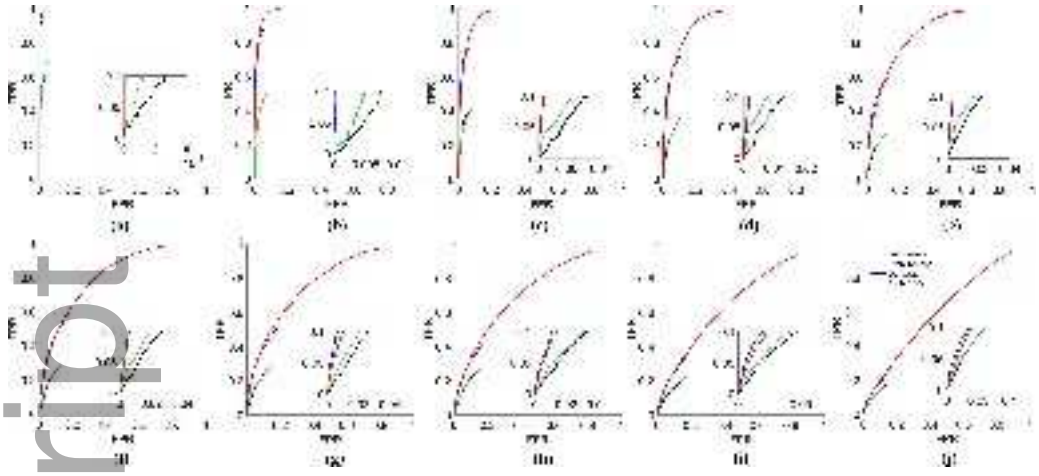
mrm_27471_f2.tif



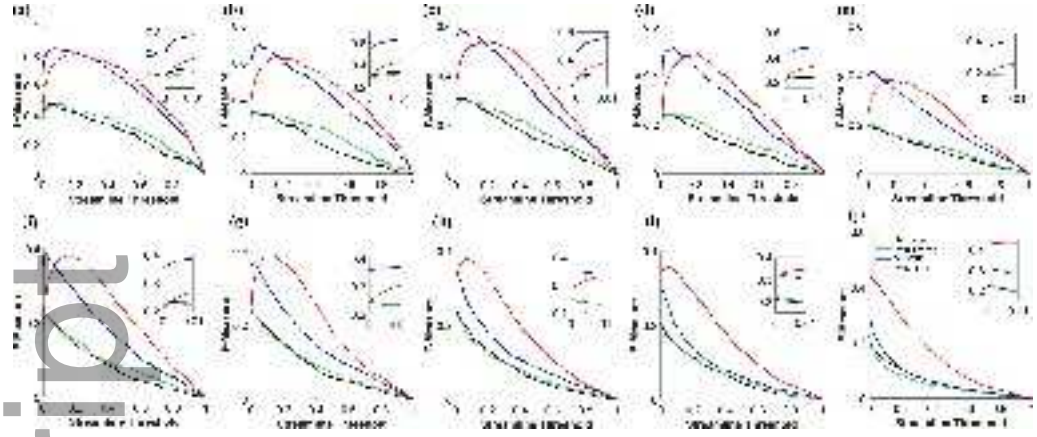
mrm_27471_f3.tif



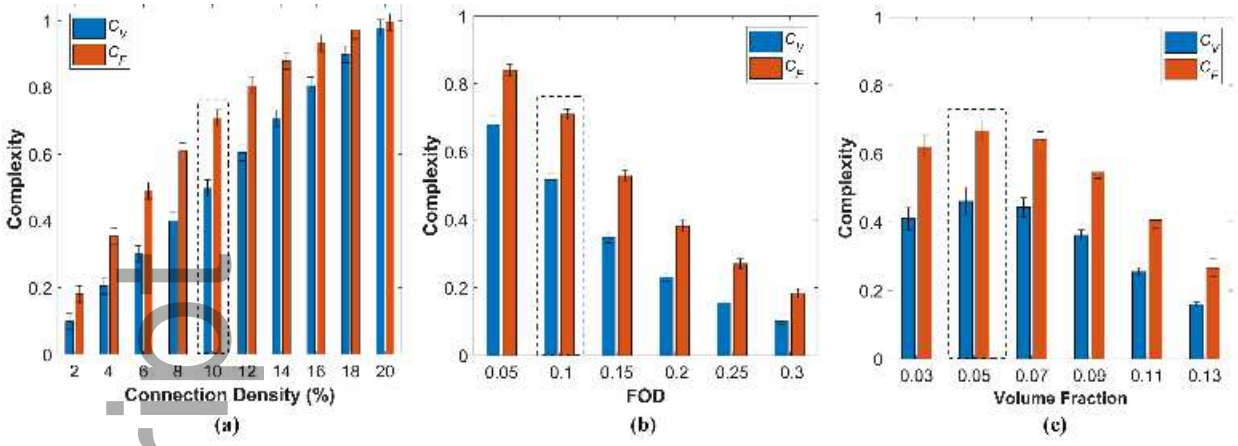
mrm_27471_f4.tif



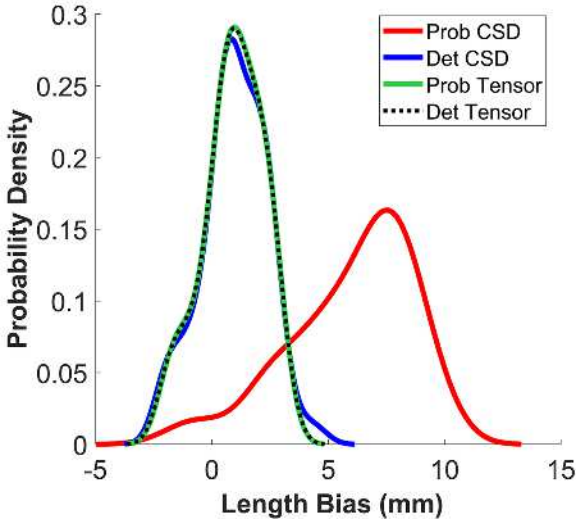
mrm_27471_f5.tif



mrm_27471_f6.tif



mrm_27471_f7.tif



mrm_27471_f8.tif