Reducing Model Uncertainty in Crossing Fiber Tractography

J. Gruen [ID], G. van der Voort [ID], and T. Schultz [ID]

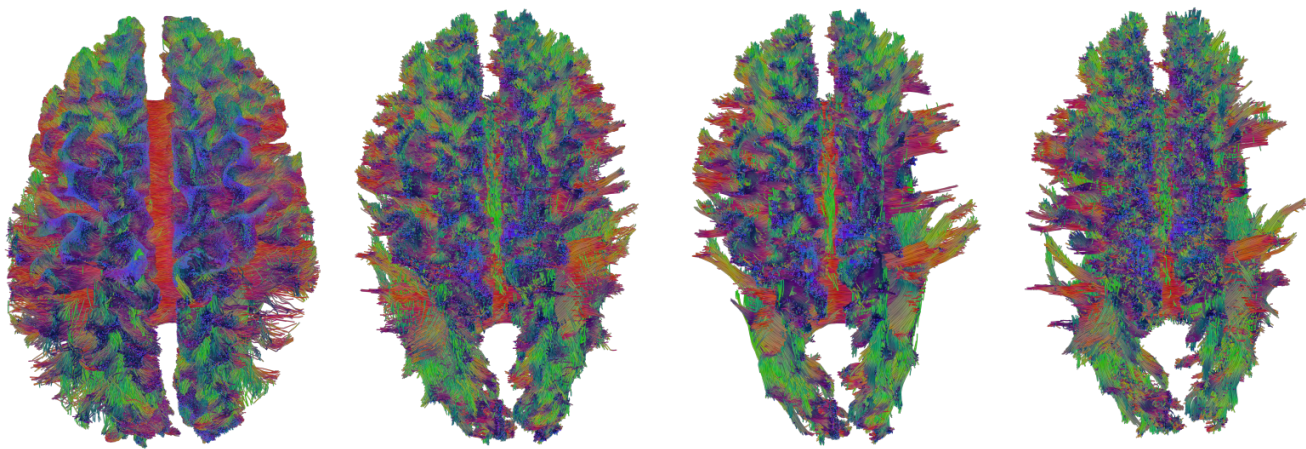University of Bonn, Bonn, Germany

Figure 1: *Left to right: A manually curated reference of the corpus callosum, a reconstruction using our proposed model averaging strategy, and two baseline reconstructions from the same seeds in the mid-sagittal plane, using tensor approximation with model selection, or standard constrained spherical deconvolution. Results from our novel approach are closest to the reference, both visually and quantitatively.*

Abstract
Diffusion MRI (dMRI) tractography permits the non-invasive reconstruction of major white matter tracts, and is therefore widely used in neurosurgical planning and in neuroscience. However, it is affected by various sources of uncertainty. In this work, we consider the model uncertainty that arises in crossing fiber tractography, from having to select between alternative mathematical models for the estimation of multiple fiber orientations in a given voxel. This type of model uncertainty is a source of instability in dMRI tractography that has not received much attention so far. We develop a mathematical framework to quantify it, based on computing posterior probabilities of competing models, given the local dMRI data. Moreover, we explore a novel strategy for crossing fiber tractography, which computes tracking directions from a consensus of multiple mathematical models, each one contributing with a weight that is proportional to its probability. Experiments on different white matter tracts in multiple subjects indicate that reducing model uncertainty in this way increases the accuracy of crossing fiber tractography.

CCS Concepts
• Applied computing → Life and medical sciences; • Mathematics of computing → Probabilistic algorithms; • Human-centered computing → Visualization techniques;

1. Introduction

Diffusion Magnetic Resonance Imaging (dMRI) [LBL*86] is a vital method for non-invasive imaging of the human brain. It is widely used in surgery planning [YYPC21] as well as in large-scale scientific studies [SJX*13; TSH*18], where it helps to learn more about the structure of healthy brains, as well as about neurological disease.

In particular, diffusion MRI is unique in its ability to non-invasively reconstruct the trajectories of major white matter tracts, using tractography algorithms [JDML19]. They are based on inferring the orientation of fiber tracts from the Brownian motion of water molecules, which is constrained by white matter microstructure such that stronger molecular motion remains possible along the tract than orthogonal to it.

Streamline-based techniques remain the most popular and widely used approach to dMRI tractography. They build on mathematical models that recover the local orientation of fiber tracts from dMRI measurements, which amounts to an ill-conditioned inverse problem [TCC07]. Diffusion tensor imaging [BML94], which estimates just one direction per voxel, is not sufficient to recover more complex fiber geometry like crossing, bending and kissing. Therefore, newer models recover multiple directions from high angular resolution diffusion imaging (HARDI). The fiber directions that are estimated by many such methods, including the ball-and-stick model [BBJ*07], or low-rank approximation of higher-order fODF tensors [SS08; ALGS17], depend on the number of distinct fibers that are assumed to be present in a voxel.

Setting the number of fibers in such a model too low will miss relevant directions, and can introduce a bias in the remaining ones. On the other hand, setting it to a number that is higher than the actual number of fibers in a voxel can lead to overfitting, i.e., the estimation of spurious fiber compartments, and an increased effect of measurement noise on the remaining ones. Unfortunately, the ideal number depends on the location within the brain and is not known a priori. Rather, it has to be estimated from the data along with the tracking directions, and it is often uncertain what is the locally optimal choice.

To our knowledge, our current work is the first one to explicitly study the effect that the uncertainty that is implied by having to make this choice has on multi-fiber tractography. Within this context, we make two main contributions. The first one is in Section 4, where we derive a mathematical framework to quantify the uncertainties in fiber rank prediction based on Bayesian model comparison. This includes a visualization of model uncertainty in crossing-fiber tractography that is, to our knowledge, the first of its kind.

Our second main contribution is introduced in Section 5, a novel approach to estimating tracking directions that reduces model uncertainty by fusing information from multiple candidate models. This new approach is compared to two previous ones that do not explicitly account for model uncertainty in Section 6. We conclude that model uncertainty has a significant impact on crossing fiber tractography that has so far not received sufficient attention, and that reducing it via model averaging can improve tractography.

To provide the required context for this main part of our work, we will start by reviewing related work on quantifying and visualizing uncertainty in dMRI tractography in Section 2, followed by some general background on the multi-fiber model that will be used in our work in Section 3.

2. Related Work

Diffusion MRI tractography is affected by various sources of uncertainty [SVBK14; SV19; GSWS21]. Among them, the propagation of measurement noise through the modeling and visualization pipeline has been studied in greatest detail. Its effect on tractography is usually quantified with probabilistic models that infer a distribution of fiber directions using Bayesian modeling [BBJ*07] or bootstrapping [CLH06]. The resulting fiber distributions have been visualized using hyperstreamlines [JTE*05; WSSS14] or illustrative confidence intervals [BPtHV13]. Bootstrapping causes considerable computational effort, but progressive visualization can reduce the resulting delay [SHV21].

A second type of uncertainty arises from the fact that tractography algorithms have several parameters that require user adjustment, including tract termination thresholds. It has been addressed by a visual tool for systematically exploring the impact of such parameters [BVPtH09].

In this work, we investigate a third source of uncertainty, which has received little attention so far. It arises from the fact that a wide range of mathematical models have been proposed for estimating fiber directions from dMRI data [PSS*12], and can lead to different results. Which model is the most suitable depends on the anatomical location [BKN04; FÖK*07], and the uncertainty from having to choose between them based on the dMRI data has been referred to as model uncertainty [SVBK14].

Estimating how many distinct fibers are contained in a given voxel is a special case of model selection, and the resulting uncertainty is at the center of our work. Most algorithms for crossing fiber tractography include some mechanism for making this choice, such as ad-hoc thresholds [QRO*09], statistical tests [ABA02], Automatic Relevance Determination [BBJ*07], or data-driven approaches [Sch12].

Recently, the use of supervised machine learning has shown a strong potential to increase the accuracy of fiber tractography [PJJD19]. This is, in part, due to their ability to exploit prior knowledge about the expected position and shape of bundles. Relying on such priors is questionable when we have to expect deviations from the norm, e.g., due to displacements from brain tumors. Therefore, our work focuses on traditional streamline-based tractography, which still remains the most widely used approach in practice.

3. Background: Crossing Fiber Tractography

The earliest algorithms for dMRI tractography were based on the diffusion tensor model [MCCvZ99; BPP*00]. However, they were limited to following a single dominant fiber orientation per voxel. Meanwhile, there is a consensus that the majority of white matter voxels in dMRI contain significant contributions from multiple bundles [JLT*12], and that accounting for this improves the quality of dMRI tractography [NDH*15].

Constrained spherical deconvolution [TCC07] is a widely used mathematical model that accounts for multiple fiber orientations. It estimates an antipodally symmetric non-negative fiber orientation distribution function (fODF), i.e., a function on the sphere that captures the fraction of fibers in any given direction. Crossing fiber tractography can then be performed by tracking in the directions of local fODF maxima. However, it has been demonstrated that the interference between fODF peaks that represent different fiber orientations limits the angular resolution, and introduces a bias in crossing angles that are estimated with this approach [SS08].

Our work builds on a variation of spherical deconvolution that represents the fODF as a symmetric fourth order tensor $\mathcal{T}$, and estimates r fiber directions from it via a rank-r approximation

$$\mathcal{T}^{(r)} = \sum_{i=1}^{r} \lambda_i \mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i \otimes \mathbf{v}_i, \tag{1}$$

where the scalar λ_i represents the volume fraction of the ith fiber, the 3D unit vector $\mathbf{v}_i$ its direction, and $\otimes$ the outer product. The parameters are chosen to minimize the Frobenius norm of the residual $\|\mathcal{T} - \mathcal{T}^{(r)}\|$. It has been demonstrated that, compared to peak extraction, this optimization increases the angular resolution, and reduces the bias in estimated directions by accounting for the interference between crossing fibers [SS08; ALGS17].

However, it introduces the need to select the number r of fibers that should be reconstructed in a given voxel. Setting r too low will miss relevant directions, and can introduce a bias in the remaining ones. Setting it too high can lead to overfitting, so that measurement noise can give rise to spurious fiber estimates. It is known that r varies across the brain [JLT*12], so it has to be adapted for each voxel. The uncertainty in which choice of r is the locally most suitable one is the type of model uncertainty that is considered in our work.

4. Quantifying Uncertainty in Fiber Rank Prediction

Most existing approaches to crossing fiber tractography include some rule for selecting the local fiber rank $r = \{1, 2, 3\}$, or some corresponding parameter in their respective mathematical model for the estimation of fiber directions. Several such approaches are surveyed in Section 2.

The first contribution of our work is a probabilistic framework that, instead of merely selecting a single value of r, allows us to quantify the posterior probability with which different potential values can be considered to be suitable given the local fODF. This framework will allow us to quantify the resulting model uncertainty. It will also serve as the basis of our novel tracking approach, which determines the local direction via a weighted sum of estimates from different models, and which will be described in more detail in Section 5.

4.1. Bayesian Model Comparison

We follow a Bayesian model comparison approach, i.e., we compute the posterior probability $p(\mathcal{H}_r|\mathcal{T})$, where $\mathcal{H}_r$ denotes the hypothesis that rank r is the optimal rank for extracting tracking directions from fODF $\mathcal{T}$. According to Bayes' theorem of conditional probability, it is given as

$$p(\mathcal{H}_r|\mathcal{T}) \propto p(\mathcal{T}|\mathcal{H}_r)p(\mathcal{H}_r) \qquad (2)$$

where $p(\mathcal{H}_r)$ is our prior belief that rank r is suitable, without considering the fODF. If we had reliable prior information concerning the overall fraction of white matter voxels that contain r significant tracts, it could be reflected in this factor. However, even though there is a consensus that multi-fiber configurations are very common, there is no clear agreement among estimates of the exact fraction of two- or three-fiber voxels [BBJ*07; JLT*12; Sch12]. Using various literature values as priors, we found that their effect on our posteriors were minor. Therefore, we decided to simply use a non-informative prior that assigns equal prior probability to the values of $r \in \{1, 2, 3\}$ that are widely used in crossing fiber tractography. Note that the case $r = 0$ can be excluded since we limit tracking to a white matter mask.

The remaining term, $p(\mathcal{T}|\mathcal{H}_r)$, is the probability of the fODF $\mathcal{T}$

given an r-fiber model. In Bayesian model comparison, it is referred to as model evidence. It is derived from $p(\mathcal{T}|\mathcal{H}_r, \theta_r)$, the posterior probability of $\mathcal{T}$ given an r-fiber model with a specific parameter vector θ_r. In our context, $\theta_r = (\lambda_1, \mathbf{v}_1, ..., \lambda_r, \mathbf{v}_r)$ contains the parameters of the low-rank tensor approximation $\mathcal{T}^{(r)}$ from Eq. (1). In Section 4.2, we propose a parametric form of $p(\mathcal{T}|\mathcal{H}_r, \theta_r)$ that assigns a high probability if $\mathcal{T}$ is well-approximated by $\mathcal{T}^{(r)}$.

The overall model evidence does not depend on any particular values for the model parameters. Rather, the specific choice of model parameters θ_r is marginalized out:

$$p(\mathcal{T}|\mathcal{H}_r) = \int p(\mathcal{T}|\mathcal{H}_r, \theta_r)\, p(\theta_r|\mathcal{H}_r)\, d\theta_r \qquad (3)$$

Intuitively, $p(\mathcal{T}|\mathcal{H}_r)$ should depend on the fODF shape: An fODF $\mathcal{T}_m$ with multiple peaks is unlikely to arise from a single fiber. Eq. (3) respects this because there is no set of single fiber parameters θ_1 that yield a high value of $p(\mathcal{T}_m|\mathcal{H}_1, \theta_1)$. On the other hand, an fODF $\mathcal{T}_s$ with a single dominant peak should not support multiple fiber compartments. Eq. (3) achieves this because a greater r corresponds to a larger parameter space θ_r, which includes many crossing fiber configurations that do not approximate $\mathcal{T}_s$ well. Given that $p(\theta_r|\mathcal{H}_r)$ is normalized over this larger parameter space, the contribution of parameters that fit the data well decreases in Eq. (3). This naturally penalizes unnecessarily complex models.

Even though Bayesian model evidence provides a well-founded framework for model uncertainty, the computational effort of evaluating Eq. (3) directly is unfortunately prohibitive, since it would require the numerical solution of a high-dimensional integral. Therefore, we use a popular approximation via the Bayesian Information Criterion (BIC) [Sch78]:

$$\text{BIC} = k \ln(n) - 2 \ln\left(p(\mathcal{T}|\mathcal{H}_r, \hat{\theta}_r)\right),$$

where $p(\mathcal{T}|\mathcal{H}_r, \hat{\theta}_r)$ corresponds to the likelihood of the rank-r model with the parameters $\hat{\theta}_r$ that best fit the fODF $\mathcal{T}$, k is the number of parameters in θ_r, and n denotes the number of data points to which the model was fitted. Note that, for a given number of datapoints, the BIC gets smaller if the number of parameters increases while the likelihood is not growing sufficiently. Under certain conditions, the BIC relates to $p(\mathcal{T}|\mathcal{H}_r)$ in the following way [KK08]:

$$p(\mathcal{T}|\mathcal{H}_r) \approx \exp(-\text{BIC}/2) \qquad (4)$$

This provides us with an estimate of model evidence that is simple and efficient to compute, and that we will use in our framework.

4.2. From Model Likelihood to Model Uncertainty

In order to apply the Bayesian framework, we still need to provide an equation for $p(\mathcal{T}|\mathcal{H}_r, \hat{\theta}_r)$, i.e., the likelihood of a given model $\mathcal{H}_r$ with a fitted set of parameters $\hat{\theta}_r$. We propose to compute it from the relative magnitude of the corresponding low-rank approximation residual:

$$\|\widetilde{\mathcal{R}}^{(r)}\| = \frac{\|\mathcal{T} - \mathcal{T}^{(r)}\|}{\|\mathcal{T}\|} \qquad (5)$$

This quantity is bounded within $[0, 1]$, with zero indicating a crossing fiber model that perfectly explains the observed fODF.
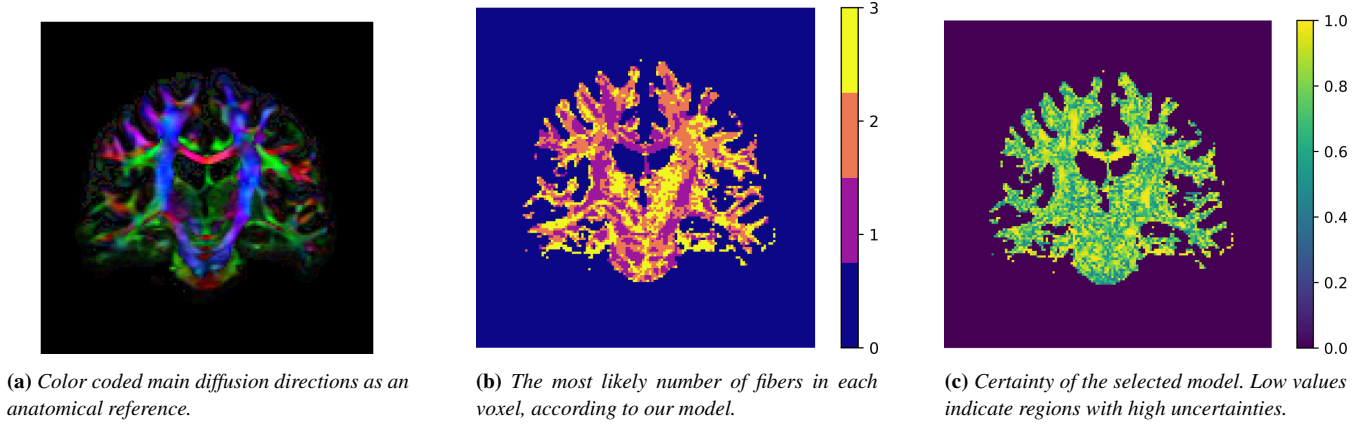
(a) *Color coded main diffusion directions as an anatomical reference.*

(b) *The most likely number of fibers in each voxel, according to our model.*

(c) *Certainty of the selected model. Low values indicate regions with high uncertainties.*

Figure 2: *Visualization of the selected model and the certainty of the model for $a = 1$ and $b = 20$.*
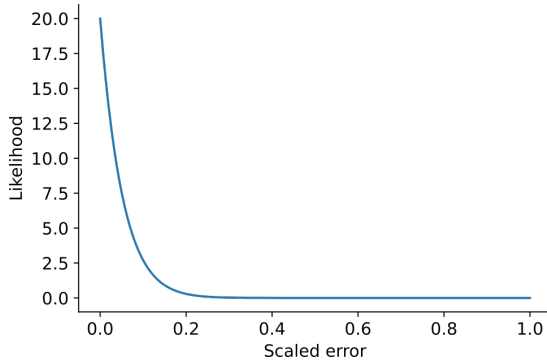
Figure 3: *We use the Kumaraswamy PDF with parameters $a = 1$ and $b = 20$ to compute the model likelihood based on the relative magnitude of the fitting residual. With these parameters, fits that result in a relative error above 20% are considered to be implausible.*

In practice, measurement noise and effects that our crossing fiber model does not account for, such as continuous fiber spread or variations in the single fiber response function, will lead to non-zero residuals. Unfortunately, it would be extremely challenging to explicitly model those effects. Therefore, we use the Kumaraswamy Probability Density Function (PDF) [Kum80] as an ansatz. It is defined as

$$f(x; a, b) = abx^{a-1} \left(1 - x^a\right)^{b-1} \text{ for } x \in (0, 1), a, b > 0 \quad (6)$$

The Kumaraswamy PDF has the benefits of being supported on the suitable interval, being computationally more efficient than the related β-distribution, and providing considerable flexibility through its parameters a and b which allows us to adapt it to our application.

It is clear that a and b should be chosen so that the probability decreases monotonically as the relative residual $\|\widetilde{\mathcal{R}}^{(r)}\|$ increases. However, it is important to select a suitable slope: Parameters that strongly penalize even small residuals would encourage overfitting, and lead to an overrepresentation of the three fiber model. Vice versa, if we assign high probabilities even to models that fit poorly,

the penalization of high model complexity in the Bayesian model comparison approach would dominate and always prefer the model with the lowest number of fibers.

In our experiments, we set $p(\mathcal{T}|\mathcal{H}_r, \hat{\theta}_r) = f(\|\widetilde{\mathcal{R}}^{(r)}\|; 1, 20)$. The corresponding shape of the Kumaraswamy PDF is shown in Fig. 3. For the data used within our current study, these settings agree well with the observed distribution of relative residual norms within a white matter mask, and resulted in anatomically plausible choices for the most likely number of fibers, as shown in Fig. 2b. In particular, these settings select a single fiber compartment within the center of the corpus callosum (CC) and parts of the corticospinal tract (CST) that are commonly used to estimate the single fiber response function in spherical deconvolution [TCC07], while permitting two or three compartments within well-known crossing regions, such as between CC and CST. These values also lead to an overall fraction of two- and three-fiber voxels that is within the range of estimates from the literature [JLT*12; Sch12].

This concludes our first important result, a framework that allows us to quantify model uncertainty. The probability of the selected model that is estimated with our approach is visualized in Fig. 2c. In a few small regions, including the central part of the CC, this value is close to one, indicating very little model uncertainty. In most other parts of the white matter, this value is clearly below one, indicating considerable model uncertainty. In the next section, we propose an approach for crossing fiber tractography that accounts for this uncertainty.

5. Crossing Fiber Tractography With Reduced Model Uncertainty

The main novelty in this section is the model averaging approach that is introduced in Section 5.1. Section 5.2 provides further detail on the tractography algorithm in which we compared this strategy to two previous ones. Section 5.3 explains the post-processing that was required in our experiments.

5.1. Computing Local Tracking Directions

Tractography based on the low-rank tensor approximation model in Eq. (1) has considered three candidate models at each step, corresponding to the fiber counts $r \in \{1, 2, 3\}$. The strategy in previous work has been to determine an optimal rank r in each integration step, and to use the resulting set of directions $\mathbf{v}_i$ for tracking [ALGS17]. However, it ignores the model uncertainty, which arises in cases in which multiple models have a non-negligible probability, but lead to different estimates of tracking directions. We refer to this approach as model selection, and use it as a baseline in our experiments. To enable a direct comparison to our newly proposed model averaging strategy, in our current experiments, model selection uses the value of r with the largest probability $p(\mathcal{H}_r|\mathcal{T})$ according to the framework from the previous section.

An important contribution of our current work is to explore model averaging as an alternative strategy for computing local tracking directions from multi-fiber models. The idea is to reduce the model uncertainty by fusing the information from the candidate models, rather than relying on one of them alone, similar to how Bretthorst et al. [BKN04] took the expectation of scalar invariants over different models. Our implementation derives an overall estimate of directions $\mathbf{v}_i$ and volume fractions λ_i as a weighted sum of the corresponding parameters $\mathbf{v}_i^{(r)}$ and $\lambda_i^{(r)}$ from the different r-fiber models, where the weights are given by the respective probabilities $p(\mathcal{H}_r|\mathcal{T})$.

Computing the weighted sums as such is straightforward. However, implementing model averaging requires establishing a correspondence between the directions from the different r-fiber models, so that the direction $\mathbf{v}_1^{(1)}$ from the single fiber model corresponds to the first directions $\mathbf{v}_1^{(2)}$ and $\mathbf{v}_1^{(3)}$ of the two- and three-fiber models. Moreover, the second directions $\mathbf{v}_2^{(2)}$ and $\mathbf{v}_2^{(3)}$ of the two- and three-fiber models need to agree. Among the $2! \times 3! = 12$ possible assignments, we select the one that minimizes the overall sum of angles between the resulting weighted means $\mathbf{v}_i$, and their corresponding $\mathbf{v}_i^{(r)}$.

As another baseline in our experiments, we consider estimates from a standard state-of-the-art multi-shell multi-tissue constrained spherical deconvolution approach [JTD*14], which uses peak finding on order-8 spherical harmonics to estimate the $\mathbf{v}_i$ and λ_i. This amounts to an implicit model selection approach, where fiber compartments are included if they are strong enough to result in a distinct local fODF maximum.

5.2. Probabilistic Streamline-Based Tractography

Our work refines the widely used idea of probabilistic streamline-based tractography. Streamlines are started at a given seed point, and are iteratively grown in both directions using Euler integration. A set of r local tracking directions $\mathbf{v}_i$ are extracted according to one of the three methods described in the previous subsection. We handle the ambiguity in the sign of $\mathbf{v}_i$ by re-orienting it to have a non-negative inner product with the current tracking direction $\mathbf{w}$. If the crossing fiber model results in r candidate directions, we select one of them according to a probabilistic model.
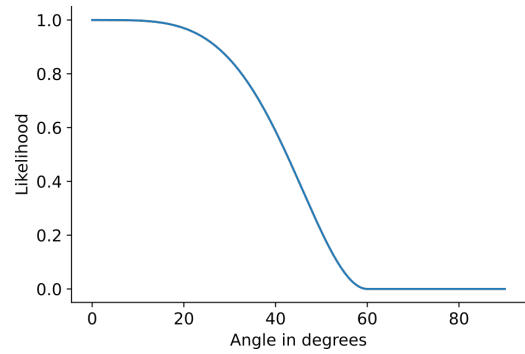

Figure 4: *Probability on the y axes of selecting a direction based on the angle in degrees between current direction and possible next direction - on the x axes - for unit length direction vectors.*

Our algorithm selects the unit direction $\mathbf{v}_i$ with volume fraction λ_i for $i \in \{1, \ldots, r\}$ with probability

$$p(\mathbf{v}_i) = \frac{\mathbb{1}_{\{\theta_i < \frac{1}{3}\pi\}} \lambda_i \cos\left(\left(\frac{3}{\sqrt{2\pi}}\theta_i\right)^2\right)^2}{\sum_j \mathbb{1}_{\{\theta_j < \frac{1}{3}\pi\}} \lambda_j \cos\left(\left(\frac{3}{\sqrt{2\pi}}\theta_j\right)^2\right)^2} \tag{7}$$

where θ_i denotes the angle between the current curve tangent direction $\mathbf{w}$ and the possible new direction $\mathbf{v}_i$. Fig. 4 illustrates the enumerator of this equation as a function of θ_i, with $\lambda_i = 1$.

To account for the limited angular resolution of spherical deconvolution [TCC07], the probabilities for angles below 30 degrees are set to be almost equal. In streamline-based tractography, it is very common to limit the maximum angular deviation that is permitted in each step. This is accounted for by assigning probability zero to all angles greater than 60 degrees. By also accounting for the volume fraction λ_i, we prefer the more significant contributions that yield more reliable tracking directions.

In our experiments, we perform Euler integration with a stepsize of 0.9 mm and terminate the streamline if no valid tracking direction is found, or if the overall white matter density drops below 0.3. For simplicity, we pre-compute the multi-vector fields at the dMRI voxel resolution and use nearest neighbor interpolation, as in FACT [MCCvZ99].

5.3. Postprocessing

Diffusion MRI tractography in general is known to be able to reconstruct large parts of many well-known white matter tracts, but is also notorious for generating false positives that, at the current state of the art, have to be removed based on anatomical knowledge [WCP*07; MNH*17]. Moreover, probabilistic tractography algorithms generate a certain fraction of outliers with low overall probability that should be removed from their output.

Therefore, we filter all tracking results in two ways. First, we use exclusion filters, which exclude streamlines if they enter certain regions, which are defined according to anatomical knowledge. For example, if a streamline from the left or right CST enters the other

hemisphere, it is excluded. In our experiments, all exclusion filters are manually calibrated for one reference subject. Then, a linear registration of all other subjects to the reference is done [JBBS02], and the linear transformation is used to map the exclusion regions to all other subjects.

Second, a density filter is used. A density map is created by counting the number of streamlines intersecting each voxel. Based on this map, the density filter cuts off streamlines when they first enter a low density region, starting at their seed point. The threshold for a low density region is set for one subject and transferred to all others by scaling it with the ratio of seed points for a tract in the reference, compared to the seed points of the tract in the target.

6. Results

6.1. Data

To evaluate the benefit of the proposed model uncertainty reduction, we compare the three above-described tractography approaches on data from the Human Connectome Project (HCP) [VSB*13]. The corresponding diffusion MR images have a resolution of 1.25 mm isotropic with $145 \times 174 \times 145$ voxels.

As a reference, we used high-quality tractography results that were published along with the learning-based tractography method TractSeg. Details on how this reference data was created can be found in the corresponding publication [WNM18]. Briefly, streamlines belonging to several major white matter tracts were extracted automatically from a generously seeded whole brain tractography, followed by a careful manual cleanup in which wrongly excluded streamlines were added and falsely included ones were removed.

We ran our experiments on 12 HCP subjects for which such reference tractography results were available. For each method, we evaluated the extent to which it managed to reconstruct the full bundle from the intersection of the reference curves with a single plane. To account for cases in which seeds fall into regions of fiber crossings, the tangent direction of the reference curve was stored and used as the prior direction when starting the tracking. This should mimic a directional region of interest as it might be placed for seeding by an expert on brain anatomy [GRNM16].

6.2. Qualitative Comparison

In a first experiment, we tried to track the right Corticospinal Tract (CST) from the seed region that is indicated with a dashed black line in Fig. 5. This tract has a large lateral spread in the upper part. Visually, our novel model averaging approach manages to reconstruct that spreading part more completely compared to the model selection strategy, which results in a lower density in that region, or to constrained spherical deconvolution, which has problems to cover parts of the spread at all.

A potential explanation for this is that individual voxels in which model selection prefers a model with fewer fibers, despite considerable model uncertainty, could prevent a reconstruction of the spread at that point. Figure 6 investigates this further by showing the estimated fiber directions in the crucial part of the right CST, in which the streamlines have to split to achieve the desired lateral spread.
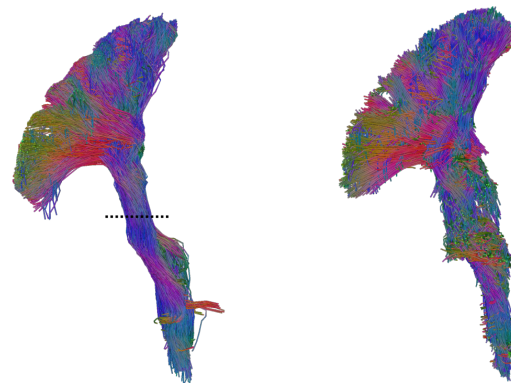
(a) *Manually curated reference.*

(b) *Low-rank approximation with proposed model averaging.*
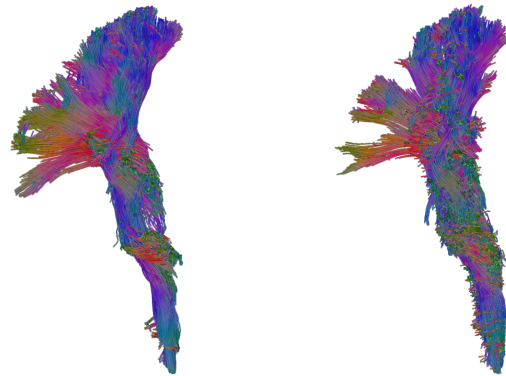
(c) *Low-rank approximation with model selection.*

(d) *Standard constrained spherical deconvolution.*

Figure 5: *Reconstructions of the right Corticospinal Tract from seeds in an axial slice whose position is indicated by the dashed line in (a).*

There are some clear differences visible between model averaging (left) and selection (right). At the first marker, we see that the selection model decided for the two fiber model with a high uncertainty. At this location, the third fiber remains available in the averaging model. We also observe that model averaging affected the direction of the main fiber: It is straight up in the selection result, but rotated in a way that is more similar to the local neighborhood after averaging.

At the second marker, model selection decided for the single fiber model. Again, model averaging permits an additional fiber compartment, even though with a small volume fraction, which is coherent with the local neighborhood and facilitates the continuous tracking of crossing fibers.

However, when focusing on the boundaries of the bundle, we also observe that it seems "fuzzier" in the model averaging and
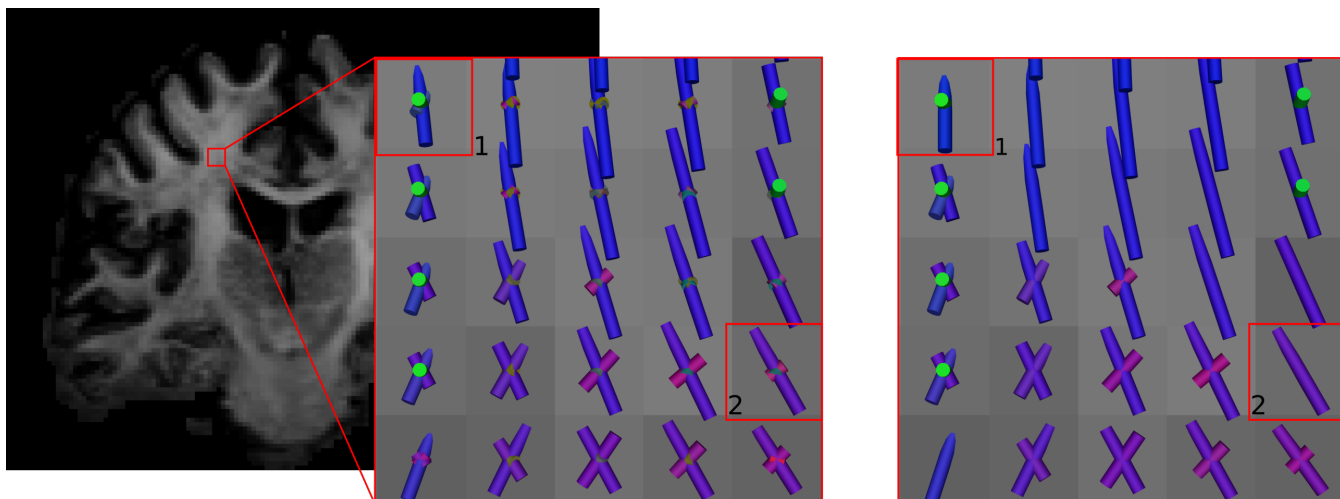
Figure 6: *Model averaging and model selection in the upper right hemisphere. The left crop shows model averaging and the right shows model selection.*

CSD approaches compared to model selection, visually indicating a larger number of streamlines that had to be cut off by the density filter. Therefore, it seems that model averaging leads to an increase in both the true and the false positive fibers.

6.3. Quantitative Comparison

To compare the different results objectively [SDM*19], we create a binary mask out of the reference data (RD) and out of the tracking results (TR). The overlap (OL) between the RD mask and the TR is calculated via

$$\text{OL} = \frac{|RD \cap TR|}{|RD|}.$$

The overreach (OR) is calculated as

$$\text{OR} = \frac{|RD^c \cap TR|}{|RD|}$$

where RD^c is the complement of RD, and the Dice score is calculated as

$$\text{Dice} = \frac{2|RD \cap TR|}{|TR| + |RD|}.$$

A high OL score indicates a high coverage of the RD, while a high OR shows that there exist many streamlines which are not within the GT, and are thus likely to be false positives. Hence, it is desirable to achieve a high OL score with a low OR score, which yields a high similarity between the tracts, and is reflected in a high Dice score.

In Table 1, the numerical results for the right CST are shown, averaged over all 12 subjects. It confirms our visual impression. The average model has by far the highest OL score, followed by the selection and the CSD models. The fuzzy boundaries correspond to a high OR score for the averaging and CSD models, while the OR score of the selection model is greatly reduced. The resulting Dice

Model	OL	OR	Dice
Average	**0.71 ± 0.034**	0.5 ± 0.092	**0.65 ± 0.025**
Selection	0.66 ± 0.029	**0.42 ± 0.063**	0.63 ± 0.016
CSD	0.63 ± 0.025	0.51 ± 0.077	0.59 ± 0.19

Table 1: *Average of scores of all twelve patients for the right CST. The best model of each score is marked bold.*

score is largest for the averaging model, closely followed by the selection model.

For a more representative experiment, we selected the corpus callosum (CC), the cingulum (CG), the corticospinal tract (CST), the inferior fronto-occipital (IFO) and the inferior longitudinal fasciculus (ILF), the optic radiation (OR), and the superior longitudinal fasciculus (SLF). In the reference tractography, the CC and SLF tracts were divided into subtracts, which we joined into a single tract (CC) or one per hemisphere (SLF) for our experiment.

The Dice scores of all tracts, averaged over all 12 subjects, are shown in Fig. 7. In all cases, our novel model averaging strategy achieved the best result. The margin towards the next best model was large in some cases (such as the CC, shown in Fig. 1), but less so in some others (such as the right OR). In most cases, low-rank tensor approximation with model selection achieved better results than CSD with peak finding, confirming previously published results [ALGS17]. An exception are the SLF tracts, which have been reconstructed by that particular approach less completely. This can be seen from Figs. 8 and 9, which report the overlap and overreach, respectively.

In all case, model averaging leads to the highest OL score. In the CC and both IFO tracts, the score is around 0.1 higher than with the next best model. Low-rank approximation with model selection performs slightly better than the CSD model: It has a higher OL score in 10 out of 13 tracts.
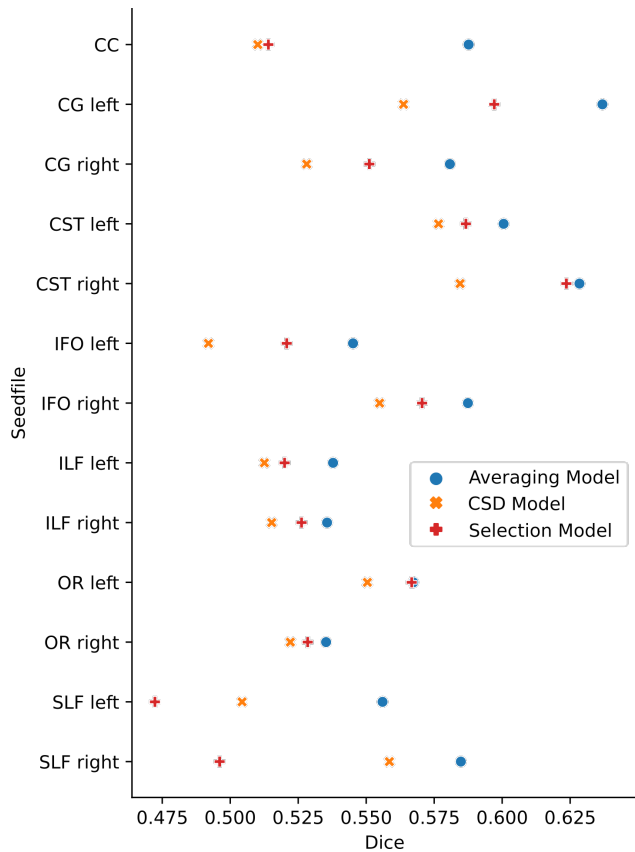
Figure 7: *Averaged Dice score over all patients tract wise for each model.*

Figure 8: *Averaged OL score over all patients tract wise for each model.*

The observation that model averaging reconstructs the bundles most completely can be explained by that fact that, when it doubt, it preserves additional fibers, even though with a reduced weight. However, Fig. 9 shows that this also increases the overreach. In 11 out of 13 tracts, model averaging leads to the highest OR score, while model selection and CSD split the second place: Both have the highest OR score once, and the lowest one almost equally often. Fig. 7 shows that, in terms of the Dice score, the benefits of model averaging with respect to a more complete reconstruction outweigh the slightly larger overreach.

6.4. Computational Effort

Experiments were computed on an Intel i9 with 3.3 GHz and 64 GB RAM. All durations below are in min:s. The code is available at `https://github.com/MedVisBonn/bonndit`.

Multi-threaded fODF estimation took 4:33 for the fourth-order tensors, 16:09 for the order-8 spherical harmonics. Single-threaded computation of directions via low-rank approximation took 1:30 with model averaging, 1:10 with model selection. Peak extraction from standard CSD took 1:30. The tracking itself is mostly independent from the pre-processing and took approximately 2:30 for a bundle such as it is shown in Figure 5, again on a single CPU core, and including post-processing.

In summary, when pre-computing the multi-vector fields for tractography, model averaging only leads to a modest increase in computational effort compared to an equivalent model selection strategy. Compared to the widely used standard CSD, low-rank approximation of fourth-order tensor fODFs reduces the overall effort [ALGS17]. This remains true when using model averaging.

7. Conclusion

While other sources of uncertainty in diffusion MRI tractography have been studied widely, the aspect of model uncertainty has received little attention so far. In this work, we proposed a novel mathematical framework to quantify this type of uncertainty, based on Bayesian model comparison. We applied it to the uncertainty that results from having to choose the number of fibers in multifiber models. However, this framework is more widely applicable. We intend to use it to study other cases of model uncertainty in the future, for example, the uncertainty resulting from deviations in the single fiber response function that can arise in cases of neurodegenerative disease [SG13].

Our second main contribution is a strategy that reduces this type of model uncertainty when estimating fiber directions for tractography, by fusing information from multiple candidate models. We implemented this idea in a probabilistic streamline-based tractography
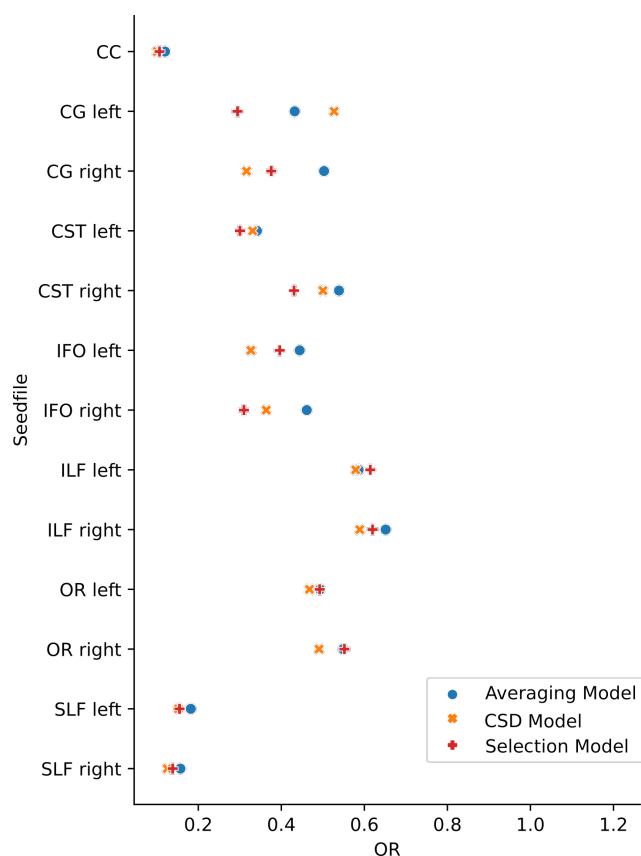
Figure 9: *Averaged OR score over all patients tract wise for each model.*

algorithm and demonstrated that, on a range of different tracts and in twelve different subjects, it led to a more complete reconstruction compared to a comparable algorithm that uses model selection, and to state-of-the-art constrained spherical deconvolution.

A limitation of our current approach is that the exact parameters of the Kumaraswamy distribution that were used in our experiments might not be universally applicable. Even though setting them based on one patient and transferring them to eleven others from the same cohort worked well, they may have to be changed when using a different dMRI acquisition scheme, or when the spatial resolution or level of noise in the data change drastically. In the future, this might be addressed by a data-driven approach to estimating model probabilities [Sch12].

Acknowledgments

References

[ABA02] ALEXANDER, D. C., BARKER, G. J., and ARRIDGE, S. R. "Detection and Modeling of Non-Gaussian Apparent Diffusion Coefficient Profiles in Human Brain Data". *Magnetic Resonance in Medicine* 48 (2002), 331–340. DOI: 10.1002/mrm.10209 2.

[ALGS17] ANKELE, M., LIM, L.-H., GROESCHEL, S., and SCHULTZ, T. "Versatile, Robust, and Efficient Tractography With Constrained Higher-Order Tensor fODFs". *Int'l J. of Computer Assisted Radiology and Surgery* 12.8 (2017), 1257–1270. DOI: 10.1007/s11548-017-1593-6 2, 3, 5, 7, 8.

[BBJ*07] BEHRENS, T.E.J., BERG, H. JOHANSEN, JBABDI, S., et al. "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?": *NeuroImage* 34.1 (2007), 144–155. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2006.09.018 2, 3.

[BKN04] BRETTHORST, G. LARRY, KROENKE, CHRISTOPHER D., and NEIL, JEFFREY J. "Characterizing Water Diffusion In Fixed Baboon Brain". *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Ed. by FISCHER, RAINER, PREUSS, ROLAND, and von TOUSSAINT, UDO. 2004, 3–15. DOI: 10.1063/1.1835192 2, 5.

[BML94] BASSER, P.J., MATTIELLO, J., and LEBIHAN, D. "Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo". *Journal of Magnetic Resonance, Series B* 103.3 (1994), 247–254. ISSN: 1064-1866. DOI: 10.1006/jmrb.1994.1037 2.

[BPP*00] BASSER, PETER J., PAJEVIC, SINISA, PIERPAOLI, CARLO, et al. "In Vivo Fiber Tractography Using DT-MRI Data". *Magnetic Resonance in Medicine* 44 (2000), 625–632. DOI: 10.1002/1522-2594(200010)44:4<625::aid-mrm17>3.0.co;2-o 2.

[BPtHV13] BRECHEISEN, RALPH, PLATEL, BRAM, ter HAAR ROMENIJ, B. M., and VILANOVA, ANNA. "Illustrative uncertainty visualization of DTI fiber pathways". *The Visual Computer* 29.4 (2013), 297–309. DOI: 10.1007/s00371-012-0733-9 2.

[BVPtH09] BRECHEISEN, RALPH, VILANOVA, ANNA, PLATEL, BRAM, and ter HAAR ROMENIJ, B. M. "Parameter Sensitivity Visualization for DTI Fiber Tracking". *IEEE Trans. on Visualization and Computer Graphics* 15.6 (2009), 1441–1448. DOI: 10.1109/TVCG.2009.170 2.

[CLH06] CHUNG, SUNGWON, LU, YING, and HENRY, ROLAND G. "Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters". *NeuroImage* 33.2 (2006), 531–541. DOI: 10.1016/j.neuroimage.2006.07.001 2.

[FÖK*07] FREIDLIN, RAISA Z., ÖZARSLAN, EVREN, KOMLOSH, MICHAEL E., et al. "Parsimonious Model Selection for Tissue Segmentation and Classification Applications: A Study Using Simulated and Experimental DTI Data". *IEEE Trans. on Medical Imaging* 26.11 (2007), 1576–1584. DOI: 10.1109/TMI.2007.907294 2.

[GRNM16] GRAUMANN, ANDREAS, RICHTER, MIRCO, NIMSKY, CHRISTOPHER, and MERHOF, DORIT. "Exploring Crossing Fibers of the Brain's White Matter Using Directional Regions of Interest". *Visualization in Medicine and Life Sciences III*. Springer, 2016, 179–194. DOI: 10.1007/978-3-319-24523-2_8 6.

[GSWS21] GILLMANN, CHRISTINA, SAUR, DOROTHEE, WISCHGOLL, THOMAS, and SCHEUERMANN, GERIK. "Uncertainty-aware Visualization in Medical Imaging - A Survey". *Computer Graphics Forum* 40.3 (2021), 665–689. DOI: 10.1111/cgf.14333 2.

[JBBS02] JENKINSON, MARK, BANNISTER, PETER, BRADY, MICHAEL, and SMITH, STEPHEN. "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images". *NeuroImage* 17 (2002), 825–841. DOI: 10.1016/s1053-8119(02)91132-8 6.

[JDML19] JEURISSEN, BEN, DESCOTEAUX, MAXIME, MORI, SUSUMU, and LEEMANS, ALEXANDER. "Diffusion MRI fiber tractography of the brain". *NMR in Biomedicine* 32.4 (2019), e3785. DOI: 10.1002/nbm.3785 1.

[JLT*12] JEURISSEN, BEN, LEEMANS, ALEXANDER, TOURNIER, JACQUES-DONALD, et al. "Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging". *Human Brain Mapping* 34.11 (2012), 2747–2766. DOI: 10.1002/hbm.22099 2–4.

[JTD*14] JEURISSEN, BEN, TOURNIER, JACQUES-DONALD, DHOLLANDER, THIJS, et al. "Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion {MRI} data". *NeuroImage* 103 (2014), 411–426. DOI: 10.1016/j.neuroimage.2014.07.061 5.

[JTE*05] JONES, D. K., TRAVIS, A. R., EDEN, G., et al. "PASTA: Pointwise assessment of streamline tractography attributes". *Magnetic Resonance in Medicine* 53.6 (2005), 1462–1467. DOI: 10.1002/mrm.20484 2.

[KK08] KONISHI, SADANORI and KITAGAWA, GENSHIRO. *Information Criteria and Statistical Modeling*. Springer Science & Business Media, 2008. ISBN: 978-0-387-71886-6. DOI: 10.1007/978-0-387-71887-3 3.

[Kum80] KUMARASWAMY, P. "A generalized probability density function for double-bounded random processes". *Journal of Hydrology* 46.1-2 (1980), 79–88. ISSN: 00221694. DOI: 10.1016/0022-1694(80)90036-0 4.

[LBL*86] LE BIHAN, DENIS, BRETON, ERIC, LALLEMAND, DENIS, et al. "MR Imaging of Intravoxel Incoherent Motions: Application to Diffusion and Perfusion in Neurologic Disorders". *Radiology* 161.2 (1986), 401–407. DOI: 10.1148/radiology.161.2.3763909 1.

[MCCvZ99] MORI, SUSUMU, CRAIN, BARBARA J., CHACKO, V. P., and van ZIJL, PETER C. M. "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging". *Annals of Neurology* 45.2 (1999), 265–269. DOI: 10.1002/1531-8249(199902)45:2<265::aid-ana21>3.0.co;2-3 2, 5.

[MNH*17] MAIER-HEIN, KLAUS H., NEHER, PETER F., HOUDE, JEAN-CHRISTOPHE, et al. "The challenge of mapping the human connectome based on diffusion tractography". *Nature Communications* 8.1 (2017). DOI: 10.1038/s41467-017-01285-x 5.

[NDH*15] NEHER, PETER F., DESCOTEAUX, MAXIME, HOUDE, JEAN-CHRISTOPHE, et al. "Strengths and weaknesses of state of the art fiber tractography pipelines. A comprehensive in-vivo and phantom evaluation study using Tractometer". *Medical Image Analysis* 26.1 (2015), 287–305. DOI: 10.1016/j.media.2015.10.011 2.

[PJJD19] POULIN, PHILIPPE, JÖRGENS, DANIEL, JODOIN, PIERRE-MARC, and DESCOTEAUX, MAXIME. "Tractography and machine learning: Current state and open challenges". *Magnetic Resonance Imaging* 64 (2019), 37–48. DOI: 10.1016/j.mri.2019.04.013 2.

[PSS*12] PANAGIOTAKI, ELEFTHERIA, SCHNEIDER, TORBEN, SIOW, BERNARD, et al. "Compartment models of the diffusion MR signal in brain white matter: A taxonomy and comparison". *NeuroImage* 59 (2012), 2241–54. DOI: 10.1016/j.neuroimage.2011.09.081 2.

[QRO*09] QAZI, ARISH A., RADMANESH, ALIREZA, O'DONNELL, LAUREN, et al. "Resolving Crossings in the Corticospinal Tract by Two-Tensor Streamline Tractography: Method and clinical assessment using fMRI". *NeuroImage* 47.2 (2009), T98–T106. DOI: 10.1016/j.neuroimage.2008.06.034 2.

[Sch12] SCHULTZ, THOMAS. "Learning a Reliable Estimate of the Number of Fiber Directions in Diffusion MRI". *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI) Part III*. Ed. by AYACHE ET AL., N. Vol. 7512. LNCS. 2012, 493–500. DOI: 10.1007/978-3-642-33454-2_61 2–4, 9.

[Sch78] SCHWARZ, GIDEON. "Estimating the Dimension of a Model". *The Annals of Statistics* 6.2 (Mar. 1978), 461–464. ISSN: 0090-5364. DOI: 10.1214/aos/1176344136 3.

[SDM*19] SCHILLING, KURT G., DADUCCI, ALESSANDRO, MAIER-HEIN, KLAUS, et al. "Challenges in diffusion MRI tractography Lessons learned from international benchmark competitions". *Magnetic Resonance Imaging* 57 (2019), 194–209. ISSN: 0730-725X. DOI: 10.1016/j.mri.2018.11.014 7.

[SG13] SCHULTZ, THOMAS and GROESCHEL, SAMUEL. "Auto-calibrating Spherical Deconvolution Based on ODF Sparsity". *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI) Part I*. Ed. by MORI ET AL., K. Vol. 8149. LNCS. Springer, 2013, 663–670. DOI: 10.1007/978-3-642-40811-3_83 8.

[SHV21] SIDDIQUI, FAIZAN, HÖLLT, THOMAS, and VILANOVA, ANNA. "A Progressive Approach for Uncertainty Visualization in Diffusion Tensor Imaging". *Computer Graphics Forum* 40.3 (2021), 411–422. DOI: 10.1111/cgf.14317 2.

[SJX*13] SOTIROPOULOS, STAMATIOS N., JBABDI, SAAD, XU, JUNQIAN, et al. "Advances in diffusion MRI acquisition and processing in the Human Connectome Project". *NeuroImage* 80 (2013), 125–143. DOI: 10.1016/j.neuroimage.2013.05.057 1.

[SS08] SCHULTZ, THOMAS and SEIDEL, HANS-PETER. "Estimating Crossing Fibers: A Tensor Decomposition Approach". *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), 1635–1642. DOI: 10.1109/TVCG.2008.128 2, 3.

[SV19] SCHULTZ, THOMAS and VILANOVA, ANNA. "Diffusion MRI Visualization". *NMR in Biomedicine* 32.4 (2019), e3902. DOI: 10.1002/nbm.3902 2.

[SVBK14] SCHULTZ, THOMAS, VILANOVA, ANNA, BRECHEISEN, RALPH, and KINDLMANN, GORDON. "Fuzzy Fibers: Uncertainty in dMRI Tractography". *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*. Ed. by HANSEN, C., CHEN, M., JOHNSON, C., et al. Springer, 2014, 79–92. DOI: 10.1007/978-1-4471-6497-5_8 2.

[TCC07] TOURNIER, J-DONALD, CALAMANTE, FERNANDO, and CONNELLY, ALAN. "Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution". *NeuroImage* 35.4 (2007), 1459–1472. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2007.02.016 2, 4, 5.

[TSH*18] TOBISCH, ALEXANDRA, STIRNBERG, RÜDIGER, HARMS, ROBBERT L., et al. "Compressed Sensing Diffusion Spectrum Imaging for Accelerated Diffusion Microstructure MRI in Long-Term Population Imaging". *Frontiers in Neuroscience* 12 (2018), 650. DOI: 10.3389/fnins.2018.00650 1.

[VSB*13] VAN ESSEN, DAVID C., SMITH, STEPHEN M., BARCH, DEANNA M., et al. "The WU-Minn Human Connectome Project: An overview". *NeuroImage* 80 (2013). Mapping the Connectome, 62–79. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2013.05.041 6.

[WCP*07] WAKANA, SETSU, CAPRIHAN, ARVIND, PANZENBOECK, MARTINA M., et al. "Reproducibility of quantitative tractography methods applied to cerebral white matter". *NeuroImage* 36 (2007), 630–644. DOI: 10.1016/j.neuroimage.2007.02.049 5.

[WNM18] WASSERTHAL, JAKOB, NEHER, PETER, and MAIER-HEIN, KLAUS H. "TractSeg - Fast and accurate white matter tract segmentation". *NeuroImage* 183 (2018), 239–253. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2018.07.070 6.

[WSSS14] WIENS, VITALIS, SCHLAFFKE, LARA, SCHMIDT-WILCKE, TOBIAS, and SCHULTZ, THOMAS. "Visualizing Uncertainty in HARDI Tractography Using Superquadric Streamtubes". *Proc. EG Conf. on Visualization (EuroVis) Short Papers*. 2014, 37–41. DOI: 10.2312/eurovisshort.20141154 2.

[YYPC21] YANG, JOSEPH YUAN-MOU, YEH, CHUN-HUNG, POUPON, CYRIL, and CALAMANTE, FERNANDO. "Diffusion MRI tractography for neurosurgery: the basics, current state, technical reliability and challenges". *Physics in Medicine & Biology* (2021). DOI: 10.1088/1361-6560/ac0d90 1.