# MesoFT: Unifying Diffusion Modelling and Fiber Tracking

Marco Reisert[1], V.G. Kiselev[1], Bibek Dihtal[1], Elias Kellner[1], and D.S. Novikov[2]

[1] Department of Diagnostic Radiology, Medical Physics, University Medical Center Freiburg, Breisacher Street 60a, 79106 Freiburg, Germany
[2] Bernard and Irene Schwartz Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine, New York, NY 10016

**Abstract.** One overarching challenge of clinical magnetic resonance imaging (MRI) is to quantify tissue structure at the cellular scale of micrometers, based on an MRI acquisition with a millimeter resolution. Diffusion MRI (dMRI) provides the strongest sensitivity to the cellular structure. However, interpreting dMRI measurements has remained a highly ill-posed inverse problem. Here we propose a framework that resolves the above challenge for human white matter fibers, by unifying intra-voxel mesoscopic modeling with global fiber tractography. Our algorithm is based on a Simulated Annealing approach which simultaneously optimizes diffusion parameters and fiber locations. Each fiber carries its individual set of diffusion parameters which allows to link them by their structural relationships.

## 1 Introduction

Diffusion MRI (dMRI) has become an essential tool for noninvasive mapping of brain tissue [5]. A unique advantage of dMRI originates from the diffusion length, a typical displacement of water molecules, being of a few $\mu$m in a clinical scan, which is commensurate with cell dimensions. Hence, in addition to the millimeter-level anatomical MRI resolution, dMRI is sensitive to the tissue structure on the *mesoscopic scale* — an intermediate length scale between the molecular level where the NMR signal originates, and the macroscopic imaging voxel size. Mesoscopic brain tissue modeling attempts to quantify cellular-level tissue organization [2,1,4,8] in each voxel. However, determining the $\mu$m-scale parameters, such as axonal dimensions, water fraction and myelin thickness from a dMRI signal acquired at a 100–1000 times lower resolution, is a difficult problem riddled by unstable model fitting and sensitivity to noise. Fiber tractography, on the other hand, focuses on characterizing the structural connectome and inferring the interregional relationships of the human brain. It delineates white matter tracts based on the empirical anisotropy of the dMRI signal, and does not attempt to quantify the mesoscopic structure. Tractography algorithms come in many varieties, divided into deterministic streamline-based, probabilistic and global [7,9] approaches. Recently, the concept of tractometry [3] was developed, which projects mesostructural properties on top of already tracked fiber bundles.

P. Golland et al. (Eds.): MICCAI 2014, Part III, LNCS 8675, pp. 201–208, 2014.

This idea combines both fields above, however it is just a rather a retrospective 'combination' of outcomes of conceptually different approaches.

In contrast, this work tries to merge both fields into one framework such that both problems can benefit from each other. There are a few attempts that try similar approaches [6,10], but from an different perspective. The proposed algorithm performs a global tracking and local modeling simultanously. Instead of fitting the parameters independently in a voxel-by-voxel manner we treat the problem as one global optimization problem. On the one hand, we are able to link voxels by their structural relationships. On the other hand, the presence of fibers is purely driven by the fact that they correctly explain the observed signal and no further assumptions about topology are made. To make the optimization of this difficult, non-convex problem tractable we propose an efficient approximation of the data likelihood. A full brain reconstruction of the human brain takes about 12 hours on a standard Desktop PC.

## 2 The Fiber Model

The fiber model is built of small segments $X \in \mathcal{X}$. Each segment contributes to the predicted MR-signal $M(\mathcal{X})$ with a small signal contribution. Each segment carries its individual diffusion parameters that define this contribution. The segments can connect and polymerize to form long chains, called fibers. The set of edges connecting the segments is denoted by $\mathcal{E}$. The complete model $\mathcal{F} = (\mathcal{X}, \mathcal{E}, v)$ consists of the set of segments, their edges between them and the volume fractions $v$. Our mesoscopic model $M(\mathbf{r}, \mathbf{q})$ is composed of axially-symmetric Gaussian diffusion signals of the form

$$m_{\mathbf{n}}^{D_\parallel, D_\perp}(\mathbf{q}) = e^{-D_\parallel t(\mathbf{q}\cdot\mathbf{n})^2 - D_\perp t(|\mathbf{q}|^2 - (\mathbf{q}\cdot\mathbf{n})^2)}$$

from different white matter compartments (here $t$ is a fixed diffusion time and the b-value [5] $b = |\mathbf{q}|^2 t$). The parameter $\mathbf{n}$ denotes the bundle direction, $D_\parallel$ and $D_\perp$ denote axial and radial diffusivities. The signal model is composed of the sum of two such tensor models, where for one of those the perpendicular diffusion is zero, and an additional constant reflecting non-diffusing water molecules. So, the signal from the $i$th segment is:

$$M_i(\mathbf{r}, \mathbf{q}) = \; v_r(\mathbf{r}) + m_{\mathbf{n}_i}^{D^i_\parallel, 0}(\mathbf{q}) v_a(\mathbf{r}) + m_{\mathbf{n}_i}^{D^i_\parallel, D^i_\perp}(\mathbf{q})(1 - v_a(\mathbf{r}) - v_r(\mathbf{r})). \qquad (1)$$

where $v_r$ is the volume fraction of completely restricted water, $v_a$ the intra-axonal fraction and $v_e = (1 - v_a - v_r)$ the extra-axonal fraction. While empirically this model captures the signal from a straight fiber bundle quite well, it currently has a limitation of setting the intra- and extra-axonal diffusivities to be the same. In general, these parameters seem not to be equal. The total expected signal is composed of a sum over all segments: $M(\mathbf{r}, \mathbf{q}) = \sum_{X_i \in \mathcal{X}} w_i \, I(\mathbf{r}, \mathbf{r}_i) M_i(\mathbf{q})$, where $I$ is an indicator function giving contributions if $\mathbf{r}$ and $\mathbf{r}_i$ is in same voxel. Each segment carries 5 variables $X_i = (\mathbf{r}_i, \mathbf{n}_i, D^i_\parallel, D^i_\perp, w_i)$, the position, direction, the axial diffusivity along the fiber, the perpendicular diffusivity and its overall

weight. Note that the volume fractions $v_r(\mathbf{r})$, $v_a(\mathbf{r})$ and $v_e(\mathbf{r}) = 1 - v_a(\mathbf{r}) - v_r(\mathbf{r})$ are not properties of the segment but of the position. This is, on the one hand, conceptually quite natural and, on the other hand, a way to avoid the ambiguities of the model. To increase the number of segments (to get a higher number of fibers) the voxels may be divided into subvoxels which all share the same signal.

The cost functional, or energies as called in the following, consist of two parts: the data-likelihood and the prior that regularize the problem and control the connections between the segments. For optimization we use, like in [9], a simulated annealing approach. The idea is to simulate the Gibbs distribution $P(\mathcal{F}) = \frac{1}{Z}\exp(-(E_{\text{data}}(M(\mathcal{X}, v)) + E_{\text{prior}}(\mathcal{F}))/T)$ while lowering the temperature $T$. For lower temperature it gets more and more likely that we sample from minimum of the energy. The simulation principle is based on a Reversible Jump Monte Carlo Markov Chain (RJMCMC).

### 2.1 The Energy: Data Likelihood and Priors

The data term consists of a simple quadratic difference between signal and model, that is, we falsely assume a Gaussian data likelihood, which might cause a substantial bias on the parameters. However, we found in the numerical experiments that the Rician noise floor is mostly disrupting the $v_r$-fraction leaving the rest of the parameters nearly unbiased.

The priors control the number of segments, their connections, foster smoothness of the variables along fibers. Due to the freedom of the diffusion parameters we need a prior to prevent the fiber model to build unreasonable, non fiber like configurations, therefore we introduced an additional term $E_{\text{guide}}$ similar to original data-likelihood, but each segment has a fixed diffusion model. We found that very sharp diffusion models, i.e. no extra-axonal compartment and high parallel diffusion, help to resolve sharp crossings. The second prior controls the number of particles and the third the number of connections. To each particle a cost is assigned, called chemical potential $E_{\text{chem}}(\mathcal{X}) = \mu|\mathcal{X}|$ where $\mu$ is strength of the prior, or equivalently the cost of one particle. The prior controlling the connection is similar to [9], but with one important extension. Each segment $X$ has to two ports that can make connections with other segments. The location of the port is $\mathbf{r} \pm \ell\mathbf{n}$. If two segments are connected an additional potential is turned on which controls, the curvature and the similarity of the diffusion parameters. Let the segments $X_1$ and $X_2$ be connected, then we have the additional energy

$$U_{\text{con}}(X_1^{\alpha_1}, X_2^{\alpha_2}) = \lambda_d \sum_{P \in \{D_\parallel, D_\perp, v\}} (P^1 - P^2)^2 + U_{\text{bend}}(X_1^{\alpha_1}, X_2^{\alpha_2}),$$

where $\alpha_1, \alpha_2$ specify the ports. For a detailed description of the second term $U_{\text{bend}}$ consult [9]. The first term gives an additional penalty on differences between the diffusion parameters, i.e. it drives the diffusion parameters to be similar along connected segments.

### 2.2 Approximation of Q-Space Correlations

The RJMCMC algorithm needs to compute energy differences like $E_{\text{data}}(M + M_{\text{mod}}) - E_{\text{data}}(M)$. The computation is dominated by correlations of the current model $M$ with the newly added or modified segment $M_{\text{mod}}$, and the correlation of segment $M_{\text{mod}}$ with the signal. The spatial part of these correlations is trivial, however the q-space part is quite costly as it involves the evaluation of the exponential model. To compute these correlations efficiently we found a power series approximations that can speed up the computation by an order of magnitude. The approximations are of type

$$\langle m_{\mathbf{n}}^{D_\parallel, D_\perp}, S\rangle_Q = \frac{1}{Q}\sum_{k=1}^{Q} m_{\mathbf{n}}^{D_\parallel, D_\perp}(\mathbf{q}_k)\; S(\mathbf{q}_k) \approx \sum_{l,m=1}^{M} \frac{b_{lm}(\mathbf{n})}{(\kappa + D_\parallel)^l (\kappa + D_\perp)^m} \quad (2)$$

where the $b_{lm}(\mathbf{n})$ can be found by a least squares minimization and the parameter $\kappa$ is fixed and has to be found empirically to obtain good fits. The form is reminiscent of the Laplace transformation of exponential-type functions. For the two-shell scheme (a b=1000 and b=2000 shell) considered in the experiments we found $\kappa = 4$ to work well. We found values $M > 6$ do not improve fitting accuracy.

## 3 The Algorithm

As already stated the optimization of the proposed energies is accomplished by an RJMCMC-type algorithm together with a cooling process. The idea behind the RJMCMC-algorithm is to repeatedly make random distortions to the current state $\mathcal{F}$ . The distortion, called $\mathcal{F}'$, usually depends on the previous state and follows some distribution $P_{\text{prop}}(\mathcal{F} \mapsto \mathcal{F}')$, which can be arbitrarily chosen by the algorithm designer. The only condition is that the reverse transition has to be possible, i.e. $P_{\text{prop}}(\mathcal{F}' \mapsto \mathcal{F}) > 0$. The algorithm needs usually a certain number of initial iterations such that the sequence of generated states follows the desired distribution and is in equilibrium. Once equilibrium is reached (which can be checked by statistics of the energy differences), the system is slowly cooled down. In the following we present the different proposals used in our implementation.

**Segment Birth:** A segment $X = (\mathbf{r}, \mathbf{n}, D_\parallel, D_\perp, w)$ is proposed by choosing all parameters uniformly. Then, the energy difference regarding the data-likelihood is computed according to $\Delta E_{\text{data}} = -2\langle M_X, S(\mathbf{r})\rangle_Q + 2\sum_k \langle M_X, M_k\rangle_Q + \langle M_X, M_X\rangle_Q$, where the sum over $k$ ranges over all segments that lie within the voxel containing the new segment. For the efficient computation of such correlation the approximation (2) is used. The computation $\Delta E_{\text{guide}}$ is similar. Finally, the Gibbs ratio is $R = N_0 \exp(-(\Delta E_{\text{data}} + \Delta E_{\text{guide}})/T)/(N+1)$, where $N$ is the number of segments currently present and $N_0$ the expected number of segments of the underlying Poisson process. **Segment death:** A segment $X$ is randomly chosen. The energy differences that have to be computed are just the negated differences from the birth proposal. The Gibbs ratio

is $R = N \exp(-(\Delta E_{\text{data}} + \Delta E_{\text{guide}})/T)/(N_0)$. **Segment move** A segment $X$ is randomly chosen. The position and orientation is distorted by normally distributed random numbers, $\mathbf{r}' := \mathbf{r} + \sigma_s \eta$ and $\mathbf{n}' := \mathbf{n} + \sigma_n \eta$. The Gibbs ratio is just $R = \exp(-(\Delta E_{\text{data}} + \Delta E_{\text{guide}} + \Delta E_{\text{con}})/T)$. **Change of segment's diffusion parameter** A segment $X$ is randomly chosen. The current diffusion parameters are distorted by normally distributed random numbers, where the variance is proportional to the current temperature. The energy difference is computed in the same way like for the move proposal. **Change of volume fraction** A random voxel is chosen. Let us call $E_M$ the data energy before the parameter change, then: $E_M = \sum_{k,j} \langle M_k, M_j \rangle_Q - 2 \sum_k \langle M_k, S(\mathbf{r}) \rangle_Q$, where the sum runs over all segments within the voxel. And correspondingly $E_{M'}$ after the change, then $\Delta E_{\text{data}} = E_{M'} - E_M$. In the same way like for the diffusion parameters the new volume fraction is proposed by distorting the old one by a normal distribution with a variance proportional to the current temperature. **Dis/Connecting segments** For the connection of segments follows the same principle as proposed in [9].

**Parameters:** The segment parameters are chosen similar to [9]. The length $\ell$ is chosen to be $2mm$ and the potential of connection is $L = 0.5$ (see [9] for notations). The chemical potential of a segment is chosen proportional to the number of measurement in q-space. We found $\mu = 0.005\ Q$ to be a good choice. That is, if a segment explains on average more than 0.005 of the variance of the signal, the segment is probably kept. For the strength of $E_{\text{guide}}$ we found $\lambda_{\text{guide}} = 15T/T_{\text{start}}$ to work well. For the strength of the connection priors we found that values of $\lambda_c = 1$ and $\lambda_d = 1$ work already quite well. The temperature schedule starts at $T_{\text{start}} = 0.3$ and cools down to $T_{\text{end}} = 0.0025$, which corresponds to a SNR level of $1/\sqrt{T_{\text{end}}} = 20$.

## 4 Experiments

We consider a 2-shell scheme at b-values of 1000 and 2000 acquired with 60 directions per shell. The in vivo diffusion measurement was acquired on a Siemens 3T TIM Trio using an SE EPI sequence, with a TE of 107 ms. A healthy male volunteer (aged 36) was scanned at an isotropic resolution of 2.5mm. Additionally, a $T_1$ data set was acquired which was segmented into white matter (WM), gray matter (GM), and CSF using SPM. White matter was thresholded at a probability of 0.5 to determine the area of reconstruction.

First, to understand the importance of our approximation we did a brute force search on a synthetic data. By sweeping through the 3-parameter space of $D_{\parallel}, D_{\perp}$ and $v_i$ we found that our approximation speeds up the likelihood computation by a factor of 20 compared to an ordinary implementation. To validate the accuracy of the approximation we simulated a simple crossing/bending configuration (see Figure 1) consisting of three bundles. The central crossing has a crossing angle of 50°. The phantom is simulated on $24 \times 24 \times 9$ grid with a isotropic voxel size of 2mm. Each of the three bundles has the same axonal volume fraction of 0.4, extra-axonal fraction of 0.6 and different diffusion parameters $(D_{\parallel}, D_{\perp})$. Bundle a) has $(1, 0.5)$, bundle b) $(1.5, 0.5)$ and bundle c) $(2, 1)$.

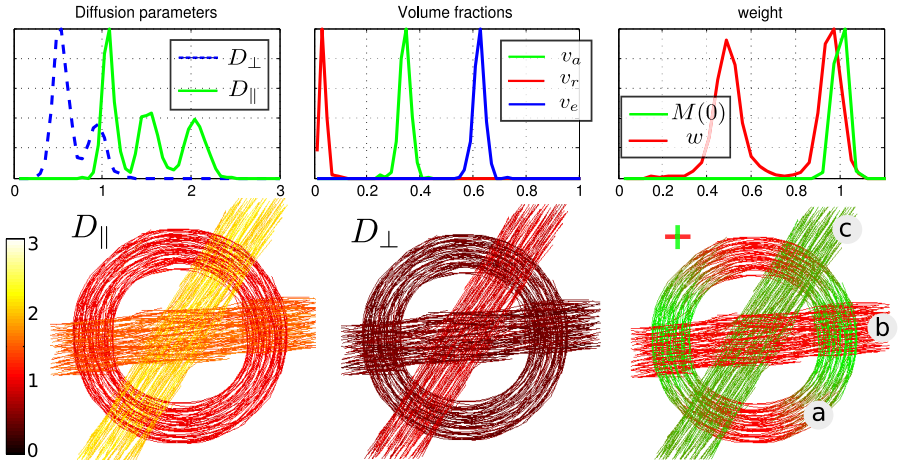


**Fig. 1.** Results for the phantom. Top: histograms of the diffusion parameters, volume fractions and signal weights/magnitudes. Bottom: Tracking result in direction coloring and coloring by their diffusion coefficients.

Rician noise was added with $\sigma = 0.05$ corresponding to a SNR of 20. Figure 1 shows histograms of our tracking results: fitted diffusion parameters, volume fractions, weight parameters and the tractogram. The reconstructed tracts are shown in three different coloring, one by directions, one by parallel diffusion $D_\parallel$ and one by perpendicular diffusion $D_\perp$. One can observe that all parameters are nearly unbiased. While the intra axonal volume fraction $v_a$ shows a small underestimation, the $v_r$ fractions and the diffusion coefficients show a small overestimation. For the in vivo dataset a voxel was subdivided into $3^3 = 27$ subvoxels to get a sufficient number of segments/fibers. For this setting the running time of the complete tracking procedure took about 10 hours on a Intel I7 (16GB) with four threads in parallel. The reconstruction contains 1.5 million particles forming about 50000 fibers longer than 10 segments.

In Figure 2 we show the results: Parametric brain maps of the diffusion parameters (Fig.2a), first and second order statistics (Fig.2b) of all parameters including tortuosity $t = D_\parallel/D_\perp$. The diagonal of the plot matrix shows ordinary histograms, the off-diagonal plots joint histograms of all parameter pairs. We also show histograms of the $w$ parameter, the predicted signal at $b = 0$ and number of segments per voxels. Further, we selected several tracts (Fig. 2c) and d)) by two ROIs, namely, Cingulum (CG), Arcquate Fascicle (AF), Cortical Spinal Tract (CST), left Optic Radiation (OR), Fronto Occipital Fascicle (IFO) and callosal fibers to the precentral gyrus (CC). Finally, Figure 2e) shows fibers sliced coronally and colored by $D_\parallel$.

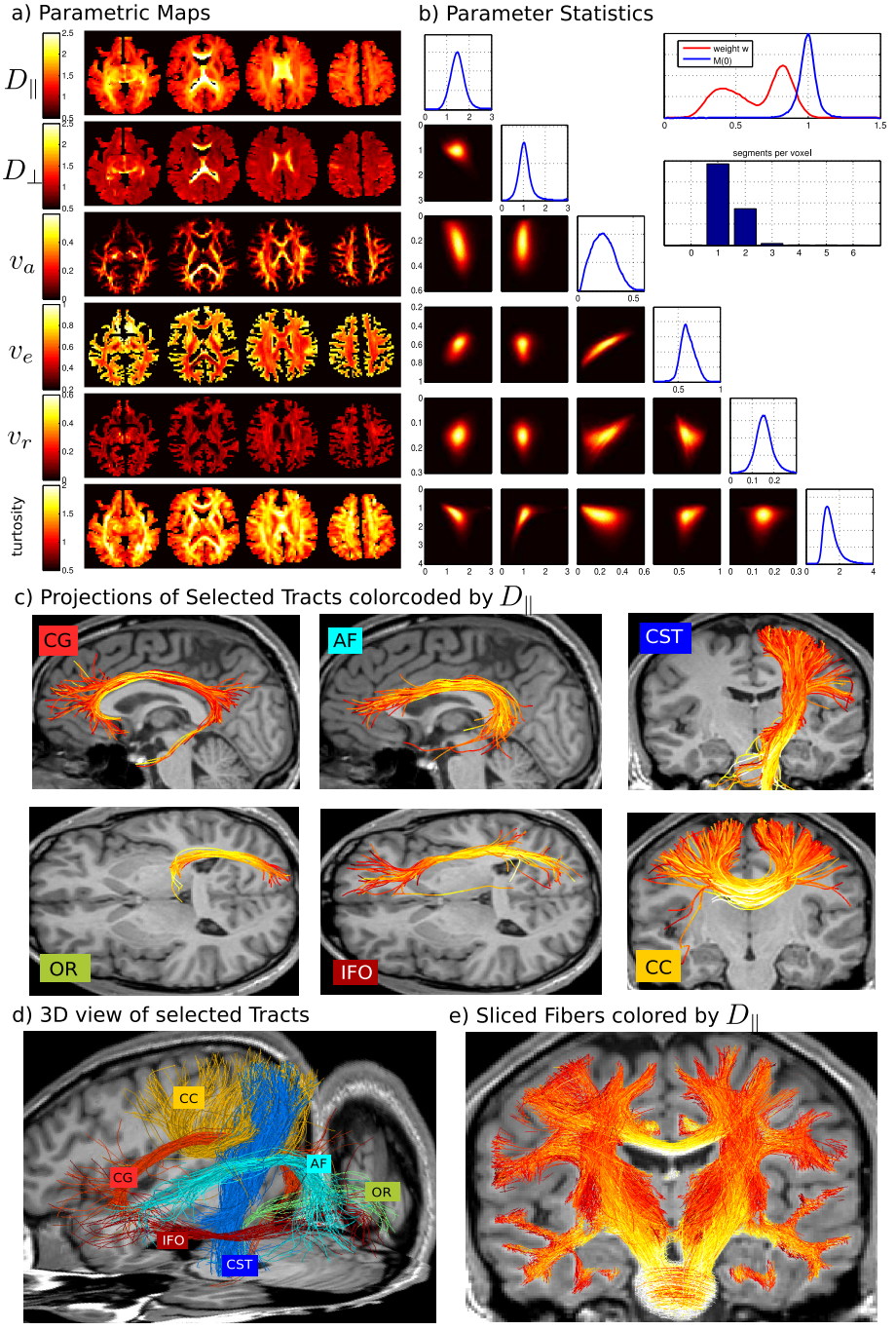


**Fig. 2.** In vivo-results for a 2-shell scheme (at b=1000,2000) at resolution of $2.5mm^3$

## 5 Discussion and Conclusion

We proposed a novel algorithm that unifies tractography and mesoscopic modeling to simultaneously reconstruct the human brain fiber bundle network and derives fiber specific diffusion parameters. The in vivo experiments show that the derived parameters go in-line with the current literature [4]. However, for the first time, we provide whole brain maps of the parameters including crossing regions. For single fiber voxel populations (like the Corpus Callosum) the putative axonal volume fraction ($v_a$) is in a range of about 40 to 50 percent, while $D_{\parallel} \approx 2$ and $D_{\perp} \approx 1$ which is similar to [4], where these parameters where derived via kurtosis imaging. The inferred parameters from multi fiber voxels differ, one can observe reduced $v_a$ and $D_{\parallel}$ while an increase in $D_{\perp}$. The source of the restricted fraction $v_r$ is not yet clear. There is definitely a certain amount caused by the Rician noise. The generated tract bundles show similar appearance like [9], but they additionally carry individual diffusion parameters.

**Acknowledgement.** The work of M. Reisert is supported by the Deutsche Forschungsgemeinschaft (DFG), grant RE 3286/2-1.

## References

1. Alexander, D.C., Hubbard, P.L., Hall, M.G., Moore, E.A., Ptito, M., Parker, G.J.M., Dyrby, T.B.: Orientationally invariant indices of axon diameter and density from diffusion mri. NeuroImage 52, 1374–1389 (2010)
2. Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y., Basser, P.J.: Axcaliber: a method for measuring axon diameter distribution from diffusion mri. Magn. Reson. Med. 59, 1347–1354 (2008)
3. Bells, S., Cercignani, M., Deoni, S., Assaf, Y., Pasternak, O., Evans, C.J., Leemans, A., Jones, D.K.: Tractometry–comprehensive multi-modal quantitative assessment of white matter along specific tracts. In: Proc. ISMRM, vol. 678 (2011)
4. Fieremans, E., Jensen, J.H., Helpern, J.A.: White matter characterization with diffusional kurtosis imaging. Neuroimage 58(1), 177–188 (2011)
5. Jones, D.K. (ed.): Diffusion MRI: Theory, Methods and Applications. Oxford University Press (2010)
6. Malcolm, J.G., Shenton, M.E., Rathi, Y.: Filtered multitensor tractography. IEEE Transactions on Medical Imaging 29(9), 1664–1675 (2010)
7. Mangin, J.: A framework based on spin glass models for the inference of anatomical connectivity from diffusion-weighted mr data - a technical review. NMR Biomed. 15(7-8), 481–492 (2002)
8. Panagiotaki, E., Schneider, T., Siow, B., Hall, M.G., Lythgoe, M.F., Alexander, D.C.: Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. Neuroimage 59(3), 2241–2254 (2012)
9. Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., Kiselev, V.: Global fiber reconstruction becomes practical. Neuroimage 54(2), 955–962 (2011)
10. Sherbondy, A.J., Rowe, M.C., Alexander, D.C.: Microtrack: an algorithm for concurrent projectome and microstructure estimation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 183–190. Springer, Heidelberg (2010)