A Framework for the Analysis of Diffusion Compartment Imaging (DCI)

Maxime Taquet, Benoit Scherrer, Simon K. Warfield

Computational Radiology Laboratory, Boston Children's Hospital, Harvard Medical School

Abstract. The brain microstructure consists of the complex organization of cellular structures and extra-cellular space. Insights into this microstructure can be gained in vivo by means of diffusion-weighted imaging that is sensitive to the local patterns of diffusion of water molecules throughout the brain. Diffusion compartment imaging (DCI) provides a separate parameterization for the diffusion signal arising from each compartment of water molecules at each voxel. Their use in population studies and longitudinal monitoring of diseases hold promise for unraveling alterations of the brain microstructure in various disorders and conditions. Yet, to analyze multi-compartment models, high-level operations commonly used in scalar images need to be generalized. We present a framework that enables interpolation, averaging, filtering, spatial normalization and statistical analyses of multi-compartment data with a focus on multi-tensor representations. This framework is based on the generalization of linear combinations of voxel values through mixture simplification. We illustrate the impact of this framework in registration, atlas construction, tractography and population studies.

1 Introduction

The brain microstructure is the complex organization of cellular structures including the neurons, the axons, their myelin sheath, and glial cells. A key technique to probe the brain microstructure *in vivo* is to acquire diffusion-weighted images (DWI), the intensity of which depends on the local diffusion of water molecules through the microstructure. Diffusion tensor imaging (DTI) is the most widely used model to represent the brain microstructure from DWI. This model assumes that water molecules are diffusing in a single compartment, be it a single fascicle (in which the diffusion is anisotropic) or a compartment of free water (in which the diffusion is isotropic). This limits the validity of DTI to represent the brain microstructure which presents a variety of compartments in each voxel, including several crossing fascicles and partial volumes of CSF. The prevalence of voxels with crossing fascicles at typical resolutions has been shown to be around 60-90% [6]. Hence any model of the microstructure that does not represent multiple fascicles is wrong in at least 60% of the brain.

In contrast, diffusion compartment imaging (DCI) provides a separate representation for the diffusion of water molecules in each microstructural compartment in each voxel, thereby enabling a richer and more reliable characterization

of the local brain microstructure. Properties of the diffusion compartment model can be directly related to properties of the underlying microstructure. For instance, the diffusion of water molecules in a direction that is orthogonal to a bundle of axons is facilitated if the axons have a larger diameter. The radial diffusivity of the corresponding compartment is therefore larger if axonal radii are larger. Typical examples of compartments are the intra-axonal space of a particular fascicle, the extra-axonal volume in the vicinity of the fascicle, the space within glial cells, and the extra-cellular space [10]. Various DCI models spanning a range of granularity and complexity have been proposed in the literature. These include the multi-tensor model with fixed diffusivities [23], full multi-tensor models with separate representations of fascicles [14], CHARMED [2], NODDI [25] and DIAMOND [13]. By providing a more reliable representation of the brain microstructure, DCI opens new opportunities to investigate the brain in disease and injury.

Analyzing DCI models has, however, proven challenging because of the absence of a one-to-one correspondence between compartments of different voxels (neighboring voxels or voxels from different subjects). This is because different voxels may have different numbers of fascicles crossing them (Fig. 1). Therefore, corresponding compartments cannot be identified and analyzed separately and a holistic approach is required. This chapter presents a holistic mathematical framework for the analysis of DCI.

Typical parameterizations of DCI models are introduced in Section 2.1. As we shall see, many operations on DCI models boil down to computing linear combinations of these models. Section 2.2 explains why computing linear combinations of diffusion compartment models is a mixture model simplification problem. Sections 2.3 and 2.4 presents two algorithms to compute linear combinations of DCI. Section 3 provides additional details about the implementation of these methods. Finally, Section 4 illustrates the far-reaching impact of this novel analysis framework on various applications, including registration, atlas construction, smoothing, tractography, population studies and model estimation.

2 Theory

In this section, we first introduce the typical parameterization of diffusion compartment models. We then explain why typical operations of image processing boil down to computing linear combinations of voxel values and how the concept of linear combinations can be generalized to diffusion compartment models. Finally, we present two approaches to compute linear combinations of DCI. The choice between these two approaches depend on the quantity that one wants to preserve: the first approach preserves the *diffusion signal* that is generated by the model, whereas the second approach preserves the *microstructural properties* of the models. In summary, the theoretical contributions of this chapter are the following:

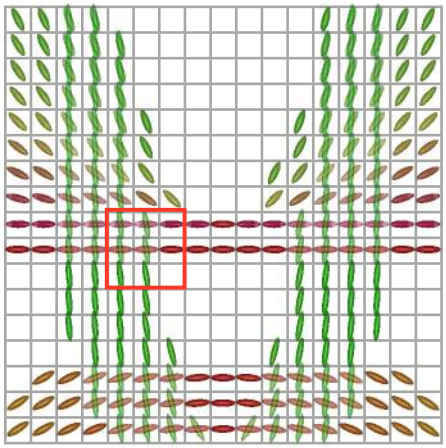

Fig. 1. Simple synthetic diffusion compartment model with fractions of occupancy encoded as the transparency of the tensors. The challenge of analyzing diffusion compartment imaging (DCI) stems from the presence of different numbers of compartments in each voxels. For example, in the red area, voxels containing one (isotropic), two and three compartments are present.

1. We propose a generic approach to generalize common operations to DCI, such as interpolation, averaging, smoothing and spatial normalization. We originally proposed this approach in [18] and applied it to registration, atlas construction and statistical analysis of DCI in [20] and [16].
2. We express two complementary approaches to computing linear combinations of DCI models: one that preserves the signal and one that preserves the microstructural properties. The former is similar (although slightly different) to the approach presented in our previous papers. The latter is a novel contribution of this chapter.

2.1 Diffusion Compartment Models

At each voxel, DCI models represent the diffusion signal arising from a set of several compartments as the weighted combination of the signals that would arise from each compartment individually. For a given set of acquisition parameters, if a compartment A is known to generate a signal S_A and a compartment B is known to generate a signal S_B, then the DCI model for the signal emanating from a voxel containing a fraction f_A of water molecules in A and a fraction $f_B = 1 - f_A$ of water molecules in B is:

$$S = f_A S_A + f_B S_B.$$

This equation can be generalized to an arbitrary number N of compartments as:

$$S = \sum_{i=1}^{N} f_i S_i, \quad \text{with } \sum_{i=1}^{N} f_i = 1.$$

The assumption underpinning this equation is that there is only slow exchange of water molecules between compartments. This is known to be an approximation since water molecules do diffuse through cell membranes. However, this approximation is reasonable insofar as the time it takes to observe a substantial exchange of water molecules between compartments is large compared to the diffusion time (which is a parameter of the acquisition).

The signal generated in each compartment, S_i, has its own parameterization. For example, the compartment containing freely diffusing water molecules (as in the CSF) can be represented by an isotropic tensor, *i.e.*, its diffusion-weighted signal for a gradient direction $\boldsymbol{g}$ and a b-value b is:

$$S_{\text{iso}} = S_0 e^{-b d_{\text{free}}},$$

where $d_{\text{free}} = 3 \times 10^{-3}\text{mm}^2/\text{s}$ is the diffusivity of free water at 37°C. In the sequel, we will focus on compartments whose signal decay can be represented by a multivariate Gaussian. This does not imply that we assume all fascicles to be represented by a single tensor. Indeed, the diffusion signal, S_j, arising from the $j-$th fascicle could be represented by two compartments, one for the intra-axonal water ($S_{j,\text{in}}$) and one for the hindered water molecules ($S_{j,\text{h}}$):

$$S_j = f_j^{\text{in}} S_j^{\text{in}} + (1 - f_j^{\text{in}}) S_j^{\text{h}} = f_j^{\text{in}} e^{-b\boldsymbol{g}^T \boldsymbol{D}_j^{\text{in}} \boldsymbol{g}} + (1 - f_j^{\text{in}}) e^{-b\boldsymbol{g}^T \boldsymbol{D}_j^{\text{h}} \boldsymbol{g}},$$

where f^{in} is the relative contribution of the intra-axonal space to the signal. In this representation, $\boldsymbol{D}^{\text{in}}$ and $\boldsymbol{D}^{\text{h}}$ have the same eigenvectors but the radial diffusivity of $\boldsymbol{D}^{\text{in}}$ is smaller than $\boldsymbol{D}^{\text{h}}$. This is similar to the CHARMED model in which the restricted diffusion would be approximated by a Gaussian diffusion tensor with small radial eigenvalues.

In summary, the class of models that we consider in the following sections is described as follows. Let N be the number of compartments including N_{iso} isotropic compartments and N_f fascicles, themselves represented by K compartments (usually $K = 1$ for a tensor representation or $K = 2$ for a bi-tensor representation). The signal decay S/S_0 is thus modeled as:

$$S/S_0 = \sum_{i=1}^{N} f_i e^{-b\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}} = \sum_{l=1}^{N_{\text{iso}}} f_l^{\text{iso}} e^{-b d_l^{\text{iso}}} + \sum_{j=1}^{N_f} \sum_{k=1}^{K} f_j^k e^{-b\boldsymbol{g}^T \boldsymbol{D}_j^k \boldsymbol{g}}.$$

2.2 Weighted Combination of DCI: a Model Simplification Problem

Weighted combinations of voxel values are ubiquitous in medical image analysis. To name a few, interpolation consists in linearly combining the values of voxels on the grid to infer the value of the image at a non-grid location. Averaging (as

used in atlas construction) consists in linearly combining the values of voxels in several subjects. Filtering consists in linearly combining the values of the voxels in an area with weights defined by a smoothing kernel.

Voxels in DCI contain compartment models and linearly combining these models is challenging. The challenge arises from the absence of a one-to-one correspondence between compartments of different voxels. There may be fascicles that are not present in all voxels (one-to-zero correspondence). There may be fascicles that are represented by different number of compartments in different voxels (one-to-many correspondence), for example in the case of fanning fascicles. For this reason, it is impossible to simply group compartments in pairs and compute linear combinations of single compartments in those pairs.

The goal of weighted combinations is to determine the parameters of a diffusion compartment model that can be interpreted as a weighted sum of other DCI models in a meaningful way. Let us define a virtual voxel containing all compartments from an original set of M compartment models, in proportion equal to some predefined weights $\alpha_m, m = 1, ..., M$. If S_{mi} is the signal arising from the $i-$th compartment in the $m-$th model ($i = 1, ..., N_m$) and f_{mi} is the fraction of its contribution to the total signal, then the diffusion signal arising from the microstructure in this virtual voxel would be accurately modeled as:

$$S = \sum_{m=1}^{M} \alpha_m \sum_{i=1}^{N_m} f_{mi} S_{mi} = \sum_{m=1}^{M} \sum_{i=1}^{N_m} \alpha_m f_{mi} S_{mi} \triangleq \sum_{k=1}^{N_c} w_k S_k.$$

A meaningful definition of the weighted combination of M diffusion compartment models could thus be the diffusion compartment model made of all these compartments with fractions equal to the original fractions multiplied by the weights of the combination. This *complete* model would, however, have an increased number of compartments (equal to $M\bar{N}$, where $\bar{N}$ is the average number of compartments in the original models) compared to all original models. This is not desirable for two reasons. First, the model complexity may become computationally intractable when linear combinations need to be recursively computed. Second, it may be that some compartments from the complete model actually represent the same microstructural environment and should therefore be merged in some way.

We want to simplify the *complete model* to obtain a *simplified model* that accurately represents the microstructure in the virtual voxel containing all compartments. To formalize this problem, let us introduce some notations. Let the complete model be denoted by $\mathcal{M}_c$ and let N_c be the number of its compartments. Let its k-th compartment be represented by a tensor $\boldsymbol{D}_k^c$ and generate a signal $S(\boldsymbol{D}_k^c, b, \boldsymbol{g}) = S_0 e^{-b\boldsymbol{g}^T \boldsymbol{D}_k^c \boldsymbol{g}}$ at a b-value b and for a gradient direction $\boldsymbol{g}$. Finally, let w_k be the relative contribution of the k-th compartment to the signal decay, so that the signal generated by the complete model at a b-value b and for a gradient direction $\boldsymbol{g}$ is modeled as:

$$S_{\mathcal{M}_c}(b, \boldsymbol{g}) = \sum_{k=1}^{N_c} w_k S(\boldsymbol{D}_k^c, b, \boldsymbol{g}).$$

Similarly, let $\mathcal{M}_s$ be the simplified model with $N_s \leq N_c$ compartments, with the j-th compartment described by a tensor $\boldsymbol{D}_j^s$ and generating a signal $S(\boldsymbol{D}_j^s, b, \boldsymbol{g})$ that contributes in a fraction f_j to the signal decay from the simplified model, so that the latter signal is modeled as:

$$S_{\mathcal{M}_s}(b, \boldsymbol{g}) = \sum_{j=1}^{N_s} f_j S(\boldsymbol{D}_j^s, b, \boldsymbol{g}). \tag{1}$$

The parameters of the simplified model are to minimize some discrepancy measure with respect to the complete model:

$$\mathcal{M}_s^* = \arg\min_{\mathcal{M}_s} d(\mathcal{M}_c, \mathcal{M}_s). \tag{2}$$

The discrepancy function depends on whether one is interested in preserving the signal or in preserving the microstructure. We detail the solutions to the mixture simplification problem in those two scenarios in the next two sections.

2.3 Signal-Preserving Model Simplification

If the model simplification should preserve, as much as possible, the signal that would arise from the complete model, then the discrepancy function reads:

$$\begin{aligned} d_S(\mathcal{M}_c, \mathcal{M}_s) &= \int \left(S_{\mathcal{M}_s}(b, \boldsymbol{g}) - S_{\mathcal{M}_c}(b, \boldsymbol{g})\right)^2 db\, d\boldsymbol{g} \\ &= \int \left(\sum_{j=1}^{N_s} f_j S(\boldsymbol{D}_j^s, b, \boldsymbol{g}) - \sum_{k=1}^{N_c} w_k S(\boldsymbol{D}_k^c, b, \boldsymbol{g}) \right)^2 db\, d\boldsymbol{g}. \end{aligned}$$

In other words, we want the simplified model to generate a signal that is as close as possible to the signal generated by the complete model, throughout the q-space. Minimizing this function is challenging. However, Zhang and Kwok provided an efficient algorithm to compute the following upper bound to this function [26,27]:

$$\begin{aligned} \tilde{d}_S(\mathcal{M}_c, \mathcal{M}_s) &= N_s \sum_{j=1}^{N_s} \int \left(f_j S(\boldsymbol{D}_j^s, b, \boldsymbol{g}) - \sum_{k \in \Omega_j} w_k S(\boldsymbol{D}_k^c, b, \boldsymbol{g}) \right)^2 db\, d\boldsymbol{g} \\ &\triangleq N_s \sum_{j=1}^{N_s} \tilde{d}_S^j \\ &\geq d_S(\mathcal{M}_c, \mathcal{M}_s) \end{aligned} \tag{3}$$

where $\boldsymbol{\Omega} = \{\Omega_1, \ldots, \Omega_{N_s}\}$ is a partition of the N_c components of the complete model in N_s clusters (*i.e.*, N_s groups of compartments that share some similarities) and $\tilde{d}_S^j$ is the discrepancy of the signal within each cluster. Minimizing $\tilde{d}_S$

is much easier than minimizing d_S because the optimization can be performed independently in each cluster, *i.e.*, the terms $\tilde{d}_S^j$ can be independently optimized.

Zhang and Kwok proposed an iterative algorithm in which the partition variables (Ω_j) and the parameters of the simplified model in each cluster (f_j and $\boldsymbol{D}_j^s$) are alternatively optimized [26,27]. Their method can be applied insofar as the components of the models can be expressed as kernel functions, $S_k = |\boldsymbol{H}_k|^{-1/2} K_{\boldsymbol{H}_k}(\boldsymbol{g} - \boldsymbol{g}_k)$, which is the case for Gaussian compartment models. The algorithm for the particular case of Gaussian DCI is presented in Algorithm 1 (with the step on Line 14 being detailed in Algorithm 2).

This algorithm deserves some comments regarding its interpretation and implementation.

1. The algorithm only involves simple matrix computations.
2. Both Algorithm 1 and Algorithm 2 relies on iterative approaches. These approaches were empirically shown to converge to a local minimum [27] and this was also observed in all of our experiments.
3. The equations on Lines 10 and 12 of Algorithm 2 result from decoupling the minimization of $\tilde{d}_S^j$ among the parameters f_j and $\boldsymbol{D}_j^s$. Line 12 follows from the fact that $\tilde{d}_S^j$ is a quadratic form in f_j and Line 10 arises by computing the partial derivative of $\tilde{d}_S^j$ with respect to $(\boldsymbol{D}_j^s)^{-1}$.
4. The formulation of the distance on Line 20 of Algorithm 1 is a direct reformulation of the distance (3) for multivariate Gaussian.
5. Convergence checking for the clustering variables on Line 12 of Algorithm 1 simply consists in assessing whether any element k moved from one cluster to another during the last iteration.
6. Convergence for the tensor estimation on Line 5 of Algorithm 2 is verified by computing the Frobenius norm of the difference between two consecutive estimates and dividing it by the norm of the current estimate. When this ratio is sufficiently small, then convergence is claimed.
7. For computational efficiency, the inverse of all tensors should be computed a priori, given their frequent occurrence throughout the algorithm.
8. The initialization of clusters on Line 11 of Algorithm 1 can be done in several manners. We found empirically that the spectral clustering approach of Ng *et al.* [8] with the similarity between two tensors defined as the cosine of the scalar product between their principal orientation is both efficient and accurate. This approach involves a k-means algorithm on the projection of similarity vectors onto the basis of the first N_s eigenvectors of the Laplacian matrix. The efficiency of the algorithm enables us to run it multiple times with random initializations and select the best clustering (based on the within-cluster sum of distances).

The results of this algorithm in terms of interpolation (a prototype application of weighted combinations) are depicted in Fig. 2. As expected from the minimization of the discrepancy between signals, the interpolated models tend to preserve the signal of the original tensors. This preservation should be understood in the following way. When the two original models (at the extreme left

and extreme right of Fig. 2a) generate the same signal (for a particular b-value and gradient direction), then all the interpolated models (in between original models) also generate the same signal, so that the signal is preserved along the interpolated line. When the original models generate different signals then the sequence of interpolated models should generate signals that monotonically evolve from one model to the other. This can be seen by looking at all the signals at any x-location on the graph of Fig. 2(b).

As observed in the top row of Fig. 2(a), the tensors tend to look inflated. This is confirmed by the plot of the radial and axial diffusivities and the fractional anisotropy which all present non-monotonic evolutions (Fig. 2(c)). By preserving the signal, we do not preserve the microstructure. In the next section, we will show that the opposite is also true.

To understand why we introduce microstructural artifacts by preserving the signal, let's picture the weighted combinations of many identical highly anisotropic compartments spanning a spectrum of orientations (Fig. 3). Preserving the microstructure would result in one such highly anisotropic compartment aligned with the mean direction in the spectrum. However, as a function of gradient direction, the signal decay generated by this average compartment would present a sharp peak in the orientation of the compartment. This sharp peak is not present in the signal decay of the original spectrum which presents moderate decay values for all gradient directions within the spectrum. To most accurately represent these moderate values spanning a larger area, an inflated tensor needs to be fit.

Another way to preserve the signal would be to compute weighted sums of the signal, *i.e.*, the set of original scalar diffusion-weighted images (DWI). This is sometimes performed in the literature as a work-around to the problem of processing DCI. In this view, the signal-preserving weighted combination presented above can be interpreted as an acceleration technique to carrying all the processing in the space of DWI (since by processing DCI, one avoids the computational burden associated with re-estimating the DCI at each processing step). However, the discussion above also makes it clear that if the goal is to preserve properties of the microstructure then one should not process DWI directly.

Finally, alternatives to the distance in Equation 3 can also be proposed. For instance, the original method proposed to compute weighted combinations of DCI was based on the minimization of the differential entropy between compartments [18]. The associated algorithm, presented in [16], is framed as an Expectation-Maximization clustering problem and is guaranteed to converge to a local minimum. The preserved quantity of this energy function is less clear. However, the simplified model simply consists in the weighted mean of covariance matrices in each cluster, which speaks more to the intuition than the equation on Line 10 of Algorithm 2. The two cost functions mostly differ theoretically and lead to very similar results in practice.

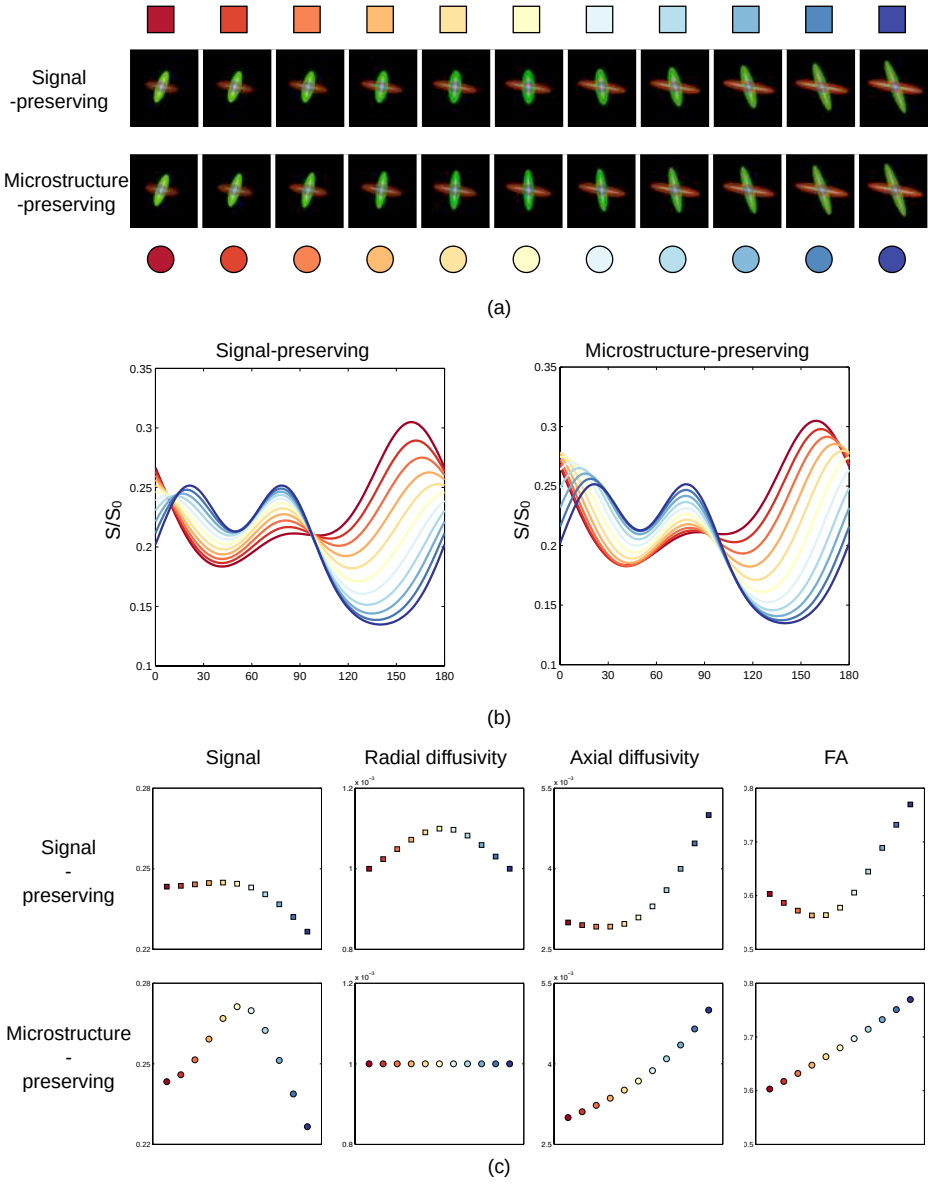

Fig. 2. Interpolation results using the signal-preserving and microstructure-preserving interpolations. Signal-preservation implies that the signal monotonically transits from one value to another, whereas microstructure-preservation implies that microstructural features monotonically transits from one value to another. When signal-preservation is achieved, microstructure is not preserved and vice versa. (a) The models at the left and right extremities are fixed and the others are interpolation results with weights increasing from 0.1 to 0.9 with steps of 0.1. The colored circles and squares match those in the graphs below. (b) Signals arising from each model for a b-value of 1000s/mm^2 and for in-plane gradients making an angle between 0 and π with the horizontal. (c) Signal generated in a direction of $\pi/25$ and various microstructural properties of the green tensor.

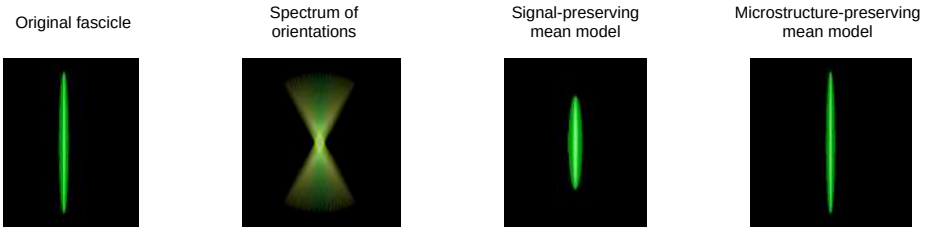

Fig. 3. Let an identical highly anisotropic original compartment be repeated many times (here 21) with various orientations. The signal-preserving mean of these compartments is an inflated tensor with higher radial diffusivity and lower axial diffusivity than the original fascicle. This inflated tensor better represents the signal generated by the spectrum of fascicles although it has microstructural features that do not correspond to original compartments. By contrast, the microstructure-preserving mean accurately represents the microstructure of the compartments in the spectrum but its generated signal departs from the signal generated by the spectrum.

2.4 Microstructure-Preserving Model Simplification

In many applications, preserving the microstructural features is arguably more important than preserving the signal. For instance, when we align DCI from various subjects in order to compare their microstructure, it is important that the interpolation does not result in inflated tensors. This may jeopardize our capability to detect group differences because the amount of inflation depends on the weights of the interpolation (tensors further away form the extremities in Fig. 2 are more inflated).

If weighted combinations must preserve the microstructure, then the discrepancy function in Equation (2) should express differences between the parameters of the DCI models rather than differences between the generated signals. The j-th compartment of the simplified model should accurately represent a subset Ω_j of the compartments from the complete model so that each compartment of the complete model is well represented by a compartment in the simplified model. The discrepancy function therefore reads:

$$d_{\mathcal{M}}(\mathcal{M}_c, \mathcal{M}_s) = \sum_{j=1}^{N_s} \left(\left| f_j - \sum_{k \in \Omega_j} w_k \right| + \sum_{k \in \Omega_j} w_k d'(\boldsymbol{D}_k^c, \boldsymbol{D}_j^s) \right).$$

The first term expresses that the fraction f_j of the j-th compartment in the simplified model should be close to the total fraction of all the compartments that the j-th compartment represents. The second term expresses that $\boldsymbol{D}_j^s$ should be close to the tensors $\boldsymbol{D}_k^c$ that it represents in a proportion that is weighted by the fraction w_k of occupancy of those compartments. This discrepancy should be simultaneously minimized for all f_j, all $\boldsymbol{D}_j^s$ and all Ω_j. Given any clustering of compartments into N_s clusters $(\Omega_j, j = 1, ..., N_s)$, one can always find a set of fractions $(f_j, j = 1, ..., N_s)$ that globally minimizes the first term. Indeed, for

$f_j = \sum_{k\in\Omega_j} w_k$, the first term is minimum and equal to zero and the constraints that $\sum_{j=1}^{N_s} f_j = 1$ with $0 \leq f_j \leq 1$ are respected. It is therefore sufficient to simultaneously optimize the tensors and the partition by minimizing the second term of the discrepancy:

$$\begin{aligned}\mathcal{M}_c^* &= \arg\min_{\mathcal{M}_c} d_{\mathcal{M}}(\mathcal{M}_c, \mathcal{M}_s) \\ &= \arg\min_{\mathcal{M}_c} \sum_{j=1}^{N_s} \sum_{k\in\Omega_j} w_k d'(\boldsymbol{D}_k^c, \boldsymbol{D}_j^s), \quad \text{with } f_j = \sum_{k\in\Omega_k} w_k.\end{aligned}$$

This is a weighted k-means clustering problem which can be solved by iterating the following two steps:

$$\boldsymbol{D}_j^s = \text{WeightedMean}\left\{(w_k, \boldsymbol{D}_k^c)\right\}_{k\in\Omega_j} \tag{4}$$

$$\Omega_j = \left\{k \,\middle|\, d'\left(\boldsymbol{D}_k^c, \boldsymbol{D}_j^s\right) \leq d'\left(\boldsymbol{D}_k^c, \boldsymbol{D}_l^s\right), \forall l = 1, ..., N_s\right\}. \tag{5}$$

In order for the k-means algorithm to converge, the definitions of the distance function d' and the weighted mean must be interrelated so that the weighted mean is the tensor that is at the smallest (weighted) cumulative distance of all the elements in the cluster. Here, we choose the recently introduced anisotropy-preserving metric of Collard *et al.* [4] for it best preserves the microstructural features while being computationally tractable.

In this anisotropy-preserving framework, the weighted mean of two tensors, $\boldsymbol{D}_1$ with weight w and $\boldsymbol{D}_2$ with weight $1-w$, is defined as follows (the generalization to N tensors is straightforward). First, we compute the spectral decomposition of each tensor:

$$\boldsymbol{D}_i = \boldsymbol{U}_i \boldsymbol{\Lambda}_i \boldsymbol{U}_i^T.$$

Second, the eigenvector matrices $\boldsymbol{U}_i$ are transformed to their quaternion representations $\boldsymbol{q}_i = (a_i, \boldsymbol{v}_i)$ where a_i is a scalar and $\boldsymbol{v}_i$ is a vector in $\mathbb{R}^3$ such that $\|q_i\|^2 = 1$. The quaternion representation can be understood as the rotation required to align the canonical basis of $\mathbb{R}^3$ to the eigenvectors $\boldsymbol{U}_i$. The scalar a_i is related to the angle of the rotation and $\boldsymbol{v}_i$ is proportional to the invariant axis of rotation.

Averaging the quaternions directly would give equal importance to the orientation of all tensors regardless of their anisotropy. However, tensors with low anisotropy have less relevant orientational information than highly anisotropic ones. To reflect this relative importance, Collard *et al.* weight quaternions by a function of the anisotropy of the tensors. Formally, if we let $\text{HA}_i = \log \frac{\lambda_{\max}}{\lambda_{\min}}$ be the Hilbert anisotropy of the tensors, we compute the weighted mean of quaternions as follows (for details on the derivation, see [4]):

$$\begin{aligned}\bar{\boldsymbol{q}} &= w^* \boldsymbol{q}_1 + (1 - w^*)\boldsymbol{q}_2 \\ \text{with } w^* &= \frac{w f\left(\min(\text{HA}_1, \overline{\text{HA}})\right)}{w f\left(\min(\text{HA}_1, \overline{\text{HA}})\right) + (1-w) f\left(\min(\text{HA}_2, \overline{\text{HA}})\right)},\end{aligned}$$

where f is any sigmoid function. From the representation of $\boldsymbol{D}_i$ as $(\boldsymbol{q}_i, \boldsymbol{\Lambda}_i)$, the weighted mean of two tensors is then defined as:

$$\begin{aligned}&\text{WeightedMean}\big\{(w, \boldsymbol{q}_1, \boldsymbol{\Lambda}_1), (1-w, \boldsymbol{q}_2, \boldsymbol{\Lambda}_2)\big\}\\&= \Big(w^* \boldsymbol{q}_1 + (1-w^*)\boldsymbol{q}_2, \exp\big(w \log \boldsymbol{\Lambda}_1 + (1-w)\log \boldsymbol{\Lambda}_2\big)\Big).\end{aligned} \tag{6}$$

In other words, in the anisotropy-preserving framework, the eigenvalues are independently averaged in the log-domain and the eigenvectors are averaged in their quaternion representations, with weights that depend on the anisotropy of tensors.

Since the sign of eigenvectors is undefined, there are 2^3 different representations of $\boldsymbol{U}_i$ as a basis of $\mathbb{R}^3$ and therefore 8 different equivalent quaternions. Before summing quaternions, they must therefore be aligned. Collard *et al.* proposed to perform this alignment by first selecting one arbitrary tensor from the set of tensors to be averaged, computing one of its quaternion representation, calling it the *reference* quaternion and then selecting for each other tensor the quaternion representation that maximizes the scalar product with the reference quaternion [4].

The distance between two tensors in the anisotropy-preserving framework of [4] can be defined in terms of a metric compose of one term related to the quaternions (the chordal distance) and one term related to the eigenvalues. Similarly to the weighted mean, the term related to the quaternion is weighted by a function of the Hilbert anisotropy. The distance induced by this metric has no obvious closed-form and we therefore employ the following approximation which is an upper bound of the distance [4]:

$$d'(\boldsymbol{D}_1, \boldsymbol{D}_2) = f\left(\min(\mathrm{HA}_1, \mathrm{HA}_2)\right) \|\boldsymbol{q}_1 - \boldsymbol{q}_2\|_2 + \sum_{i=1}^{3} \left|\log \frac{\Lambda_{1,ii}}{\Lambda_{2,ii}}\right|. \tag{7}$$

The weighted mean (Equation (6)) and distance (Equation (7)) in the anisotropy-preserving framework are ill-posed when the tensors are cylindrically symmetric, because each cylindrically symmetrical tensor can be decomposed in a quaternion in an infinite number of ways. Cylindrical symmetry is often enforced in DCI to reduce the number of parameters to estimate [10,22]. We thus introduce the following weighted average and distance operators that can be used for cylindrically symmetric tensors whose representations only depend on the main orientation $\boldsymbol{e}_i$ of tensors and their eigenvalues:

$$\begin{aligned}&\text{WeightedMean}_c\big\{(w, \boldsymbol{e}_1, \boldsymbol{\Lambda}_1), (1-w, \boldsymbol{e}_2, \boldsymbol{\Lambda}_2)\big\}\\&= \left(\frac{w^* \boldsymbol{e}_1 + (1-w^*)\boldsymbol{e}_2}{\|w^* \boldsymbol{e}_1 + (1-w^*)\boldsymbol{e}_2\|_2}, \exp\big(w \log \boldsymbol{\Lambda}_1 + (1-w)\log \boldsymbol{\Lambda}_2\big)\right)\end{aligned} \tag{8}$$

and

$$d'_c(\boldsymbol{D}_1, \boldsymbol{D}_2) = f\left(\min(\mathrm{HA}_1, \mathrm{HA}_2)\right) \operatorname{acos}(|\boldsymbol{e}_1 \cdot \boldsymbol{e}_2|) + \sum_{i=1}^{3} \left|\log \frac{\Lambda_{1,ii}}{\Lambda_{2,ii}}\right|. \tag{9}$$

We now have operators to compute the weighted mean and distance for non-symmetric tensors (Equations (6) and (7)) and for symmetric tensors (Equations (8) and (9)). We can plug these single-tensor operators into the k-means algorithm (Equations (4) and (5)) to obtain a framework for the analysis of DCI that preserves the microstructure.

Expressing the weighted combinations of diffusion compartment models as a k-means clustering problem enables the incorporation of useful extensions developed in the general k-means framework. One particularly interesting extension is the so-called *constrained* k-means in which prior knowledge about the clusters can be incorporated [24]. This is useful if we know that some compartments definitely represent different microstructural environments despite the possible similarity between their parameters. For instance, one may want to force the free water compartments to remain separate from other compartments; or one may want to keep all compartments that were separate in one voxel to remain so in the simplified model. Conversely, if some tensors are known to be represent the same microstructural environments in advance, this can also be incorporated in the k-means algorithm. Both types of constraints can be expressed as a constraint matrix C whose k_1k_2-entry defines the link between the k_1-th and k_2-th tensors in the complete model:

$$C(k_1, k_2) = \begin{cases} -1 & \text{if } k_1 \text{ and } k_2 \text{ must not belong to the same cluster} \\ 1 & \text{if } k_1 \text{ and } k_2 \text{ must belong to the same cluster} \\ 0 & \text{otherwise.} \end{cases}$$

The full algorithm to perform microstructure-preserving weighted combinations of DCI is presented in Algorithm 3. The results using this algorithm with constraints imposing that two tensors from a same voxel cannot be clustered together are depicted in Fig. 2. As expected, the microstructural features are preserved in this framework, whereas the signal is not preserved.

3 Details of Implementation

In this section, we provide additional details regarding the implementation of the weighted combinations of DCI.

3.1 Selecting the Number of Compartments in the Output

One aspect of the algorithm that we have not yet discussed is how the number of compartments of the output (N_s in Equation (1)) is determined. We argue that any choice for N_s that is lower than the maximum number of non-empty (*i.e.*, with a non-zero fraction) compartments in the input models is not self-consistent. In other words, we argue that if we have M input DCI models and if the m-th input model has N_m non-empty compartments, then we need:

$$N_s \geq \max_{m=1,\dots,M} N_m.$$

To understand why other choices would not be self-consistent, imagine that we want to compute the average DCI model in the following two situations. In the first situation, the first model is a three-compartment model $S_1 = f_a S_a + f_b S_b + f_c S_c$ and the second model is a two-compartment model $S_2 = (1 - \epsilon) S_A + \epsilon S_B$ with a fraction ϵ that is arbitrarily close to zero. In the second situation, the two-compartment model is replaced by a one compartment model $S_2 = S_A$. There is essentially no difference between the two situations in terms of the microstructure being represented and in terms of the diffusion signal being generated (the signal generated by the compartment with an infinitesimal fraction will itself be infinitesimal). Yet, if the number of compartments of the output depends on the number of compartments in the second model (e.g., if we set N_s to be the average number of compartments in the input), then we may end up with two different average DCI models in the two situations.

Since there is no obvious reason to increase the number of compartments in the output DCI model compared to the input models, we set, in every applications, the number of output compartments to be equal to the maximum number of non-empty compartments in the input models, *i.e.*,

$$N_s = \max_{m=1,\dots,M} N_m.$$

3.2 Computational Time

Several factors impact the computational time required to estimate weighted combinations of DCI models. In general, weighted combinations are computed at every voxel of an image, whether it is to apply a transformation field to a DCI image or to estimate an atlas from a set of spatially aligned image (see Section 4). This process is therefore completely parallelizable, which significantly decreases computational time. Furthermore, efficient initializations as described above makes the actual algorithm converge extremely fast. For 27 input voxels each containing a three-compartment model (which is a typical situation in the tridimensional interpolation of three-compartment models), less than 10 iterations are usually required to achieve convergence of the K-means. As a typical illustration of the computational time, the registration of three-compartment DCI models at a resolution of 1mm×1mm×1mm containing 220×220×176 voxels takes approximately 40 minutes on a 10 cores Linux machine. Amongst others, this process requires tens of millions of calls to the weighted combinations of DCI models (to interpolate the model at every iteration of the algorithm). This demonstrates that the proposed method can be used in practice for the analysis of DCI in a large number of subjects.

4 Applications

In this section, we demonstrate how the proposed framework for weighted combinations of diffusion compartment models can be used in various applications. Unless otherwise mentioned, we present results for the signal-preserving approach

with the differential entropy as a cost function since it corresponds to the method used in most of our published results. When appropriate, we compare the obtained results with those of a heuristic approach described in [18]. The heuristic approach considers DCI as multi-channel images with a tensor in each channel and performs log-euclidean single-tensor processing in each channel [1]. Attribution of a tensor to a channel is based on the FA with the tensor with the highest FA occupying the first channel. When, in a particular voxel, a channel is missing a tensor (because the selection of the number of fascicles in that voxel resulted in fewer tensors than the maximum number allowed), then the channel with the highest fraction is split into two to enforce one-to-one correspondences.

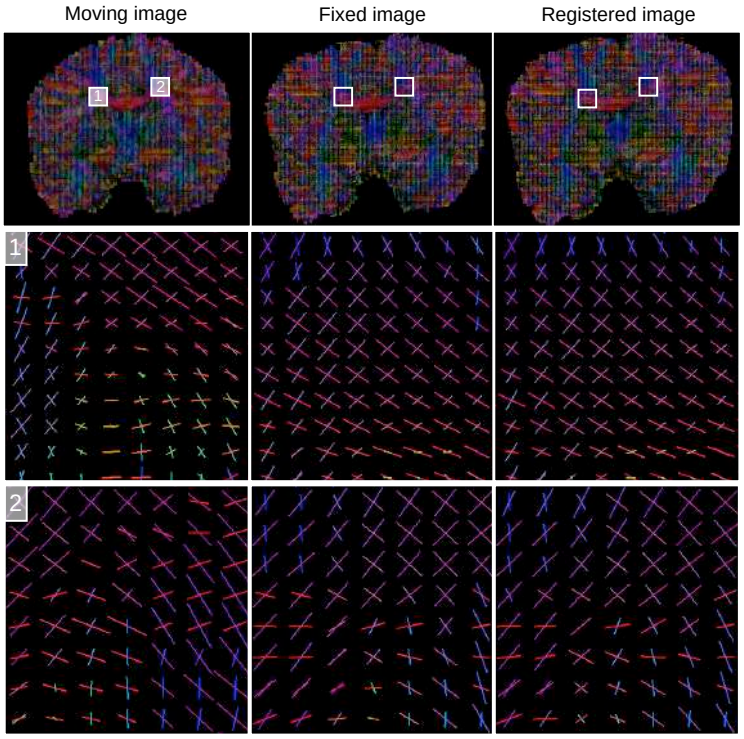

Fig. 4. The capability to compute weighted combinations of DCI combined with an appropriate similarity metric for block matching (here the generalized correlation coefficient for DCI) enables us to spatially align DCI.

4.1 Registration

Registration is pervasive in medical imaging. A subject's image can be registered to an atlas for further comparison between groups (we will discuss this appli-

cation in a following section) or to a previous image from the same subject for longitudinal analysis.

We developed a block matching approach to registration of DCI in [20,16]. This approach requires the definition of a similarity metric specifically dedicated to DCI models and a method to interpolate DCI images. Our definition of a generalized correlation coefficient as a similarity metric can be found in [17] and the interpolation of DCI was performed using the framework presented in this chapter. Examples of registered images using this approach are depicted in Fig. 4. To assess the performance of this registration approach, ten random log-Euclidean polyaffine fields were applied to the DCI of 24 subjects and noise at 6 different levels was then added to both the original and the transformed DCI. Details of the experiment can be found in [16]. Registering DCI using the presented approach is significantly more accurate than performing it with the heuristic multi-channel approach : both the magnitude and the variance of the registration errors are significantly smaller (Fig. 5).

The poor performance of the multi-channel approach can be understood by its inability to deal with the absence of one-to-one correspondence. Fig. 5(b-c) illustrates how the heuristic approach fails in those cases and results in aberrant tensors that may eventually mislead conclusions from population studies or longitudinal studies.

4.2 Construction of an Atlas of the Brain Microstructure

An atlas of the brain microstructure represents the average microstructure in a standardized anatomy. Such an atlas is an important asset to conduct population studies as we will see in a subsequent section. Constructing an atlas typically consists in iterating between the following three steps: registering all subjects to a common frame, averaging the aligned images, and applying the average inverse transform to the mean image [5]. The first and last steps require interpolation of DCI and the second step requires averaging of DCI. Both interpolation and averaging can be interpreted as a weighted combination of DCI models and we can therefore use the presented framework to build an atlas of the brain microstructure. The result, after ten iterations, is depicted in Fig. 6 with highlighted regions that are known to present crossing fascicles.

4.3 Filtering

Filtering DCI data proves very useful when the models have been estimated from noisy data. Filtering consists in replacing every voxels of the image by a weighted combination of voxels in a neighborhood. This is a simple application of the weighted combination operator for DCI. We illustrate this capability by generating a noisy version of the phantom presented in Fig. 1. To obtain a realistic noisy version of this phantom, we generated synthetic DWI in a three-shell HARDI sequence with 30 gradient directions on each shell and with b-values of 1000, 2000 and 3000s/mm^2. We added Rician noise to each DWI with a scale parameter of 40 ($= S_0/10$) and estimated the DCI model from these

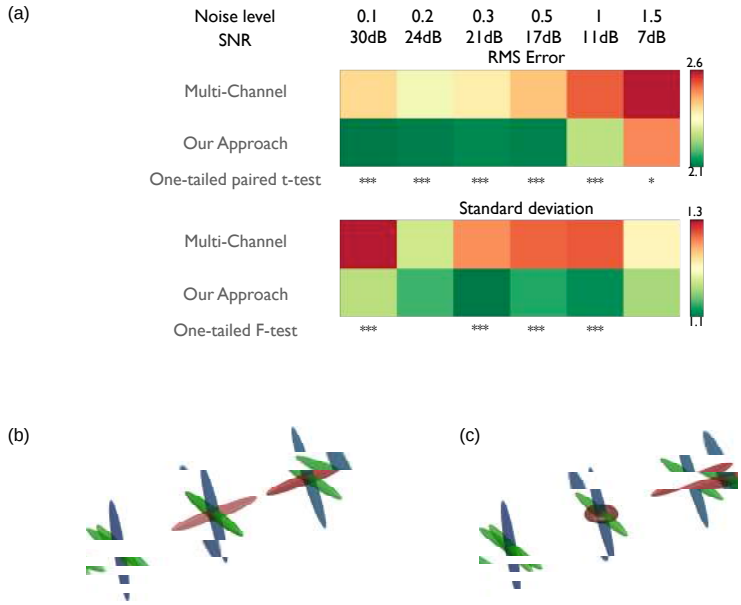

Fig. 5. (a) Registering the DCI with the proposed approach to weighted combinations is significantly more accurate than the multi-channel heuristic, with registration errors that have both smaller magnitude (top table) and smaller variance (bottom table) (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.001$). The poor performance of the heuristic approach to registration can be understood from its inability to deal with the absence of one-to-one comparisons between compartments. The model in the middle is the result of interpolating the models on the left and on the right with equal weights, (b) using the presented approach to interpolate DCI and (c) using the heuristic approach. (Figure reproduced from [16].)

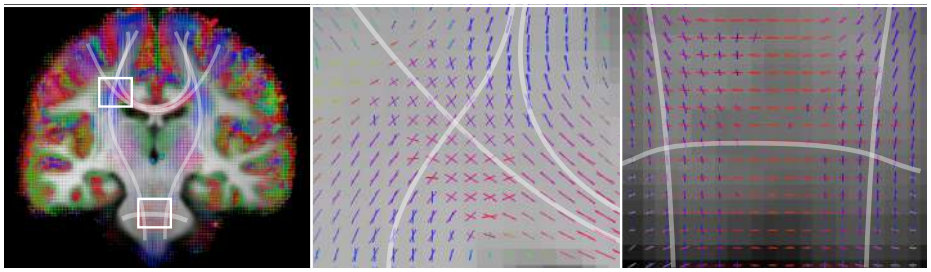

Fig. 6. From our capability to register and average DCI, we can build an atlas of the brain microstructure. This atlas represents the average microstructure in a standardized anatomy. Regions where fascicles cross are accurately represented, such as (middle) the corona radiata where projections of the corpus callosum and cortico-spinal tracts cross, and (right) a region where the pyramidal tracts (vertical lines) and the medial cerebellar peduncle (horizontal lines) cross.

noisy DWI using the technique developed in [14]. The noisy result is depicted in Fig. 7. We then applied a 3×3 Gaussian filter with a standard deviation of 0.5 and a 5×5 Gaussian filter with a standard deviation of 1 to the noised DCI using the microstructure-preserving weighted combination operator. The result, also depicted in Fig. 7, shows that aberrant tensors are adequately filtered and brought closer to the source image.

Importantly, this filtering approach leverages information from both single-tensor and multi-tensor areas in an adaptive manner: the green fibers in the highlighted region of Fig. 7 are regularized in part by the single-tensor region located above the two-tensor region whereas the orange fibers are regularized by the single-tensor region located to the right.

4.4 Application to Multi-Fiber Tractography

Single-tensor tractography typically consists in (i) shooting tracts from seeding voxels in the direction of the tensor, (ii) making one step in that direction and (iii) interpolating the DTI field at the new location and reiterating the process from there. The capability to interpolate DCI fields thus enables us to perform multi-fiber tractography that can disentangle crossing fibers. However, if interpolation leads to aberrant tensors, as with the heuristic approach, then spurious tracts may appear (false positives) and expected tracts may go missing (false negatives). In [18], this effect was demonstrated on a simple synthetic phantom consisting of two crossing fascicles (Fig. 8). With the proposed approach, however, tracts can reliably be drawn even in the region of crossing fibers and under the influence of noise.

4.5 Population Studies of the Microstructure

One of the most important applications of diffusion compartment imaging is to learn how diseases affect the brain microstructure. This can be achieved by con-

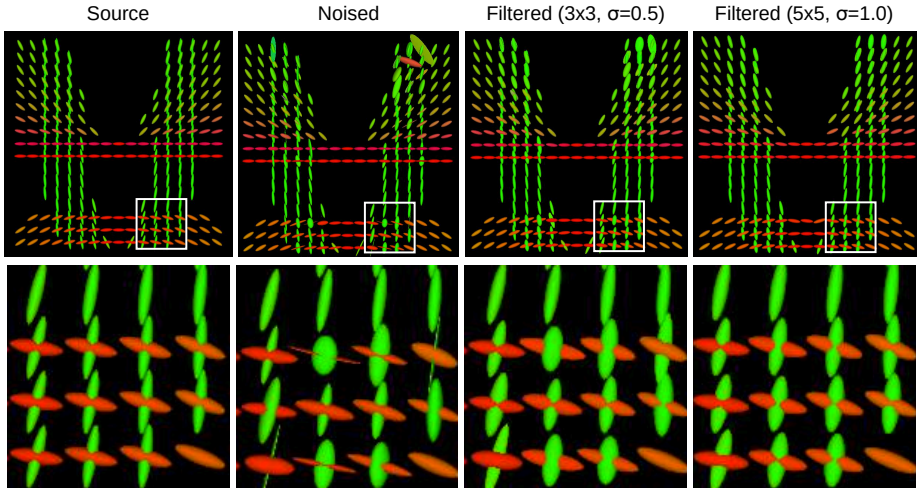

Fig. 7. Gaussian smoothing of DCI based on the weighted combination operator. The highlighted area presents voxels with very noisy tensor estimates that are being regularized by the filter.

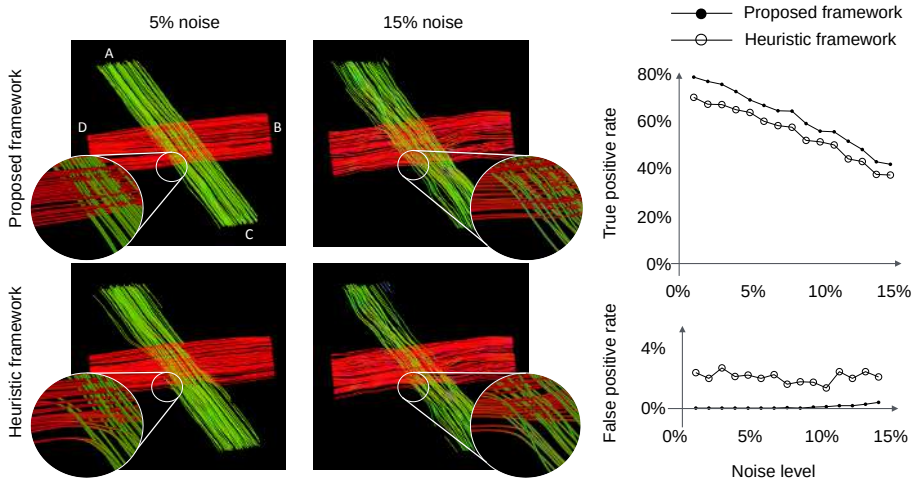

Fig. 8. The capability to compute weighted combinations enables us to perform multi-fiber tractography by interpolating the DCI field at each step. Using the heuristic multi-channel approach to interpolate DCI field results in more spurious tracts (higher false positive rate) and more missing tracts (lower true positive rate), mostly when the noise level increases.

ducting DCI-based population studies. Conducting population studies requires to construct an atlas of the brain microstructure and to register individual DCI to it. These techniques have been described in the above sections. Once all subjects have been aligned to a common coordinate frame, the parameters of the DCI models can be statistically analyzed. In [20,16], two different statistical tools for DCI analysis were introduced: fascicle-based spatial statistics and isotropic diffusion analysis. The idea behind these two tools is that some properties of DCI models pertain to individual fascicles whereas some other properties pertain to the surrounding extra-axonal volume. These tools therefore contrast with traditional DTI-based tools, such as tract-based spatial statistics (TBSS [15]) that investigate group differences in microstructure on a per-voxel basis.

Fascicle-based spatial statistics (FBSS) consists in comparing some property of the fascicle (e.g., FA, MD, radial diffusivity, axial diffusivity, etc.) along a particular fascicle of interest drawn on the atlas. Specifically, FBSS proceeds in the following steps:

1. Build a DCI atlas of the microstructure.
2. Perform multi-fiber tractography on the atlas to extract the fiber bundle of interest (tractography is only computed on the atlas).
3. Select a representative fiber tract from the bundle of interest.
4. Spatially align the DCI of all subjects to the atlas.
5. Interpolate the DCI field from each subject along the fascicle of interest.
6. Select, for each subject and at every location along the fascicle, the tensor most aligned with the fascicle.
7. Perform statistical analysis of the resulting vectors of microstructural features.

Many of the steps above (atlas construction, registration and multi-fiber tractography) require the presented framework. The last step was originally computed using cluster-based statistics [20,16]. Recently, however, a Bayesian approach to FBSS was introduced to circumvent the caveats of p-values in population studies [22]. In the latter approach, a local model of the microstructure at the population level is estimated and a Markov random field prior is assigned to its latent variables to express spatial coherence. As an illustration, FBSS was conducted in a population studies of Tuberous Sclerosis Complex (TSC) to compare the dorsal language circuit (Fig. 9) between patients with TSC and controls and, within patients with TSC, between those with autism spectrum disorder (ASD) and those without the disorder [12]. The results depicted in Fig. 10 demonstrate that FBSS reveals group differences that single-tensor DTI analysis fails to detect. FBSS also reveals specific clusters that are missing when the multi-channel heuristic processing of DCI is used. In particular, a cluster with significantly lower FA in TSC patients with autism (TSC+ASD) in the vicinity of Broca's area (Region 1 in Fig. 9 and Fig. 10) was only revealed with FBSS.

Isotropic diffusion analysis (IDA) is used to compare properties that pertain to the extra-axonal space. It was shown that a higher fraction of water molecules diffusing in this environment may be a surrogate to the presence of neuroinflammation [11]. IDA proceeds in the following steps:

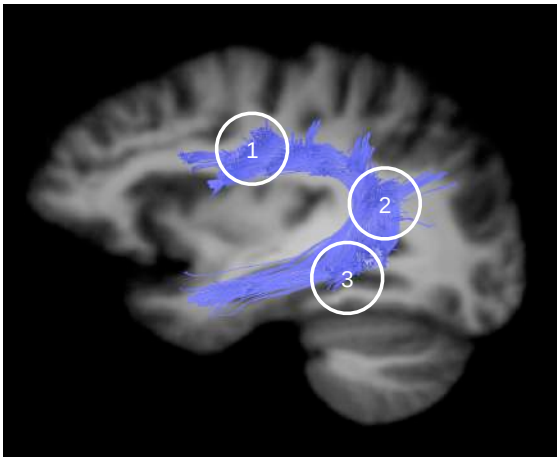

Fig. 9. The dorsal language circuit is a set of fascicles thought to connect Broca's area in the frontal lobe (Region 1), Geschwind's territory in the parietal lobe (Region 2), and Wernicke's area in the temporal lobe (Region 3).

1. Extract the scalar map of the fraction of isotropic diffusion (f_{iso}) from the aligned DCI of all subjects.
2. Transform the f_{iso} maps in l_{iso} maps with $l_{\text{iso}} = \text{logit}(f_{\text{iso}})$ to bring the distribution of the statistics close to a Gaussian.
3. Perform cluster-based statistics on the resulting l_{iso} maps as in [9].

Fig. 11 presents the results of IDA in a population study comparing children with TSC with a comorbid diagnosis of autism (TSC+ASD) and children with TSC but without autism (TSC-ASD).

4.6 Estimation of Diffusion Compartment Models

Estimating a DCI model is an ill-posed problem when data at only a single b-value are available [14,19]. In that case, additional information from an external population of subjects can be incorporated in the estimation to regularize it [21]. This prior information is encoded in a probabilistic atlas of the brain microstructure which contains, in every voxel, a distribution over the model parameters. This distribution is used as a prior in the estimation of the DCI models in a new subject [19]. The distribution is spatially aligned to the subject's space using the proposed registration method. As a result, population studies of the brain microstructure can be conducted with single b-value data that are clinically widely available.

5 Conclusion and Discussion

This chapter introduced a framework for the analysis of diffusion compartment imaging data. At the heart of this framework is the capability to compute

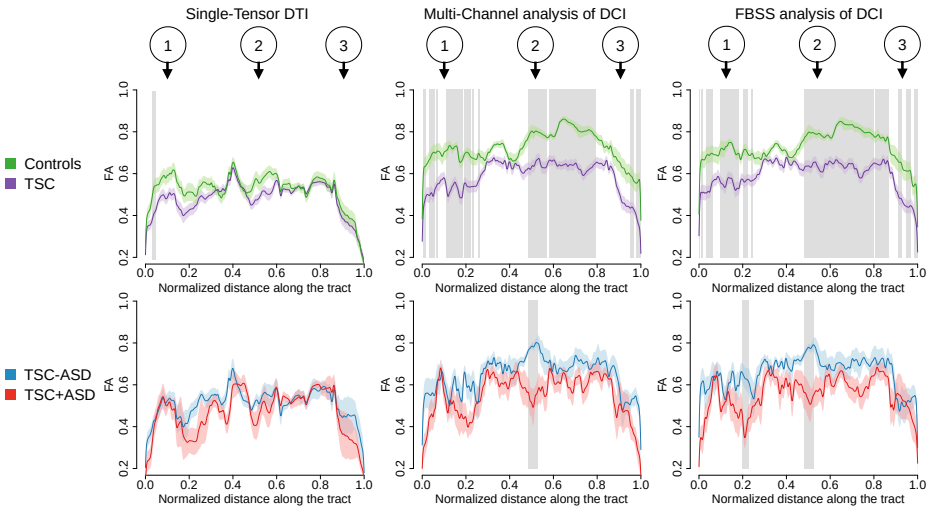

Fig. 10. Results of fascicle-based spatial statistics (FBSS) in a population studies comparing the dorsal language circuit in patients with tuberous sclerosis complex (TSC) and controls and comparing, within patients with TSC, those with autism spectrum disorder (TSC+ASD) and those without the disorder (TSC-ASD). The dark curves are the mean fractional anisotropy in the group and shaded areas around the curve represent two standard errors. Grey rectangles are locations where the FA is significantly different between the groups. FBSS reveals microstructural differences that the analysis of single-tensor DTI fails to detect. If a heuristic multi-channel approach is used to compute weighted combinations of DCI models, then some significant clusters go missing. The landmarks 1, 2 and 3 correspond to those in Fig. 9. (Figure adapted from [16]).

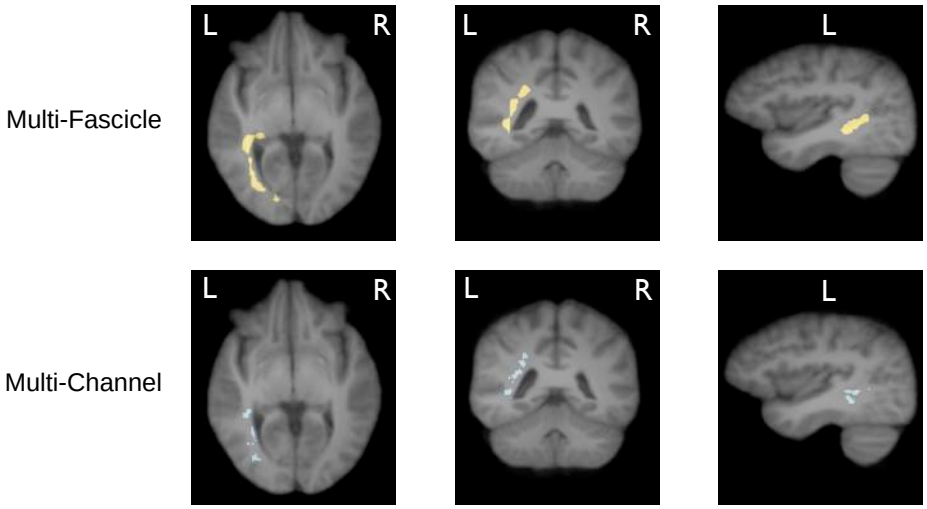

Fig. 11. IDA reveals clusters of significantly higher fraction of isotropic diffusion in children with TSC and a comorbid diagnosis of autism spectrum disorder than in children with TSC but without autism. The significant clusters are represented in the atlas space on an axial (left), coronal (center) and sagittal (right) slice. When the multi-channel heuristic approach is used (bottom line), significant clusters are located in the same areas but are less spatially coherent. (Figure adapted from [16])

weighted combinations of DCI models. This is a challenging problem because of the absence of one-to-one correspondences between the compartments of different DCI voxels (such as adjacent voxels or voxels from different subjects).

By combining all the compartments from all the DCI models of the weighted combinations, one obtain a new DCI model. This complete model is, however, not practical due to its large number of compartments. For this reason, we want to estimate a simplified model that approaches the complete model. Two approaches were introduced to simplify the complete model: a *signal-preserving* approach and a *microstructure-preserving* approach. The former leads to a simplified model whose generated signal is close to that generated by the complete model, whereas the latter has microstructural properties that are close to the original models. Importantly, the microstructure-preserving framework preserves all eigenvalues and can work equally well for tensors that have a cylindrical symmetry and for tensors that have all three eigenvalues different. In particular, averaging, smoothing or interpolating ball-and-sticks models lead to a ball-and-sticks models with the microstructure-preserving operator. This is not the case with the signal-preserving operator for which combining ball-and-sticks models may lead to multi-tensor models. In that sense, the microstructure-preserving operator generalizes the approach of Cabeen *et al.* [3] for clustering ball-and-sticks, and defines a unique framework that can be used whether the radial diffusivities are assumed null or finite.

The proposed weighted combinations operator for DCI models has far-reaching applications in microstructure imaging. We have shown in the last section how it can be used to perform registration, tractography, smoothing, atlas construction, population studies and how it can help in estimating DCI. Importantly, the operator is general enough to be incorporated in various implementations of these applications. For instance, one may be interested in developing a more advanced filtering technique for DCI based on bilateral smoothing. Such an implementation would simply require to adapt the weights of a smoothing kernel and the proposed operator could still be used. Similarly, one may want to use another algorithm for tractography, such as multi-tensor filtered tractography [7]. There again, the weighted combination operator can be incorporated to better represent the DCI field at a non grid location.

Incorporating the novel operator in various applications enables us to fully leverage DCI models in population studies of the brain microstructure, from the estimation of the models to the statistical analysis of DCI models aligned to an atlas. This opens new opportunities for the *in vivo* analysis of the brain microstructure in the normal development and in diseases and injury.

A Algorithms for the weighted combinations of DCI

The detailed algorithms for the combination of DCI are presented in Algorithm 1 the signal-preserving version and in Algorithm 3 for the microstructure-preserving version.

Algorithm 1 Signal-preserving weighted combinations of DCI

Input: (1) A set of M models $\{\mathcal{M}^m\}_{m=1,\ldots,M}$ so that $\mathcal{M}_m$ is described by fractions f_{mi} and tensors $\boldsymbol{D}_{mi}$ $(i = 1, \ldots, N_m)$, (2) The weights α_m associated with each model, (3) The number N_s of compartments in the output.

Output: A multi-fascicle model: $\mathcal{M}_s$ such that $S_{\mathcal{M}_s}(b, \boldsymbol{g}) = \sum_{j=1}^{N_s} f_j e^{-b\boldsymbol{g}^T \boldsymbol{D}_j^s \boldsymbol{g}}$

$k \leftarrow 0$

for m in 1 to M **do** ▷ Construct the complete model $\mathcal{M}^c$

  for i in 1 to N_m **do**

    $k \leftarrow k + 1$

    $w_k \leftarrow \alpha_m f_{mi}$

    $\boldsymbol{D}_k^c \leftarrow \boldsymbol{D}_{mi}$

  end for

end for

$\boldsymbol{\Omega} \leftarrow$ Initialization$(\{D_k^c, w_k\}, N_s)$ ▷ Initialize clustering

while $\boldsymbol{\Omega}$ has not converged **do**

  for j in 1 to N_s **do** ▷ Representation

    $\boldsymbol{D}_j^s, f_j \leftarrow$ Update$(\{\boldsymbol{D}_k^c, w_k\}_{k \in \Omega_j})$ (see **Algorithm 2**)

    $Z_j \leftarrow \sum_{k \in \Omega_j} w_k$

  end for

  Reset all Ω_j to empty sets

  for k in 1 to N_c **do** ▷ Clustering

    for j in 1 to N_s **do**

      $\text{Distance}(j,k) \leftarrow \left|2\boldsymbol{D}_k^c\right|^{1/2} + \frac{f_j^2}{Z_j^2}\left|2\boldsymbol{D}_j^s\right|^{1/2} - \frac{2f_j}{Z_j}\left|(\boldsymbol{D}_j^s)^{-1} + (\boldsymbol{D}_k^c)^{-1}\right|^{-1/2}$

    end for

    $c(k) \leftarrow \arg\min_x \text{ Distance}(x,k)$

    $\Omega_{c(k)} \leftarrow \Omega_{c(k)} \cup \{k\}$

  end for

end while

Algorithm 2 Update step in signal-preserving weighted combinations of DCI

Input: The tensors and fractions that are in Cluster Ω_j : $\{\boldsymbol{D}_k^c, w_k\}_{k\in\Omega_j}$

Output: The updated tensor and fraction of the j-th compartment in the simplified model : $\boldsymbol{D}_j^s$ and f_j

$(\boldsymbol{D}_j^s)^{-1} \leftarrow \dfrac{1}{\sum\limits_{k\in\Omega_j} w_k} \sum\limits_{k\in\Omega_j} w_k (\boldsymbol{D}_k^c)^{-1}$

$\delta \leftarrow \infty$

while $\boldsymbol{D}_j^s$ has not converged **do**

  for k in Ω_k **do**

    $\boldsymbol{W}_k \leftarrow \dfrac{\left((\boldsymbol{D}_j^s)^{-1} + (\boldsymbol{D}_k^c)^{-1}\right)^{-1}}{\left|(\boldsymbol{D}_j^s)^{-1} + (\boldsymbol{D}_k^c)^{-1}\right|^{1/2}} w_k$

  end for

  $\boldsymbol{P}_j \leftarrow \sum\limits_{k\in\Omega_j} \boldsymbol{W}_k$

  $(\boldsymbol{D}_j^s)^{-1} \leftarrow \boldsymbol{P}_j^{-1} \sum\limits_{k\in\Omega_j} \boldsymbol{W}_k (\boldsymbol{D}_k^c)^{-1}$

end while

$f_j \leftarrow \left|\frac{1}{2}\boldsymbol{D}_j^s\right|^{-1/2} \sum\limits_{k\in\Omega_j} \dfrac{w_k}{\left|(\boldsymbol{D}_j^s)^{-1} + (\boldsymbol{D}_k^c)^{-1}\right|^{1/2}}$

Algorithm 3 Microstructure-preserving weighted combinations of DCI

Input: (1) A set of M models $\{\mathcal{M}^m\}_{m=1,\dots,M}$ so that $\mathcal{M}_m$ is described by fractions f_{mi} and tensors $\boldsymbol{D}_{mi}$ $(i = 1, \dots, N_m)$, (2) The weights α_m associated with each model, (3) The number N_s of compartments in the output, (4) a constraint matrix C.
Output: A multi-fascicle model: $\mathcal{M}_s$ such that $S_{\mathcal{M}_s}(b, \boldsymbol{g}) = \sum_{j=1}^{N_s} f_j e^{-b\boldsymbol{g}^T \boldsymbol{D}_j^s \boldsymbol{g}}$
$k \leftarrow 0$
for m in 1 to M **do** ▷ Construct the complete model $\mathcal{M}^c$
  for i in 1 to N_m **do**
    $k \leftarrow k + 1$
    $w_k \leftarrow \alpha_m f_{mi}$
    $\boldsymbol{D}_k^c \leftarrow \boldsymbol{D}_{mi}$
  end for
end for
$\boldsymbol{\Omega} \leftarrow$ Initialization$(\{D_k^c, w_k\}, , N_s, C)$ ▷ Initialize clustering
while $\boldsymbol{\Omega}$ has not converged **do**
  for j in 1 to N_s **do** ▷ Representation
    $\boldsymbol{D}_j^s \leftarrow \text{WeightedMean}\big\{(w_k, \boldsymbol{D}_k^c)\big\}_{k \in \Omega_j}$ ▷ given by Equation (6) or (8)
    $f_j \leftarrow \sum_{k \in \Omega_j} w_k$
  end for
  Reset all Ω_j to empty sets
  for k in 1 to N_c **do** ▷ Clustering
    if $\exists\, k' < k : C(k, k') = 1$ **then** ▷ Check if k must be forced to some cluster
      $c(k) \leftarrow c(k')$
      $\Omega_{c(k)} \leftarrow \Omega_{c(k)} \cup \{k\}$
    else
      ForbiddenClusters $\leftarrow \{j | \Omega_j \supset k' < k$ and $C(k, k') = -1\}$
      for j in 1 to N_s **do**
        if $j \in$ ForbiddenClusters **then**
          Distance$(j, k) \leftarrow \infty$
        else
          Distance$(j, k) \leftarrow d'(\boldsymbol{D}_j^s, \boldsymbol{D}_k^c)$ ▷ d' given by Equation (7) or (9)
        end if
      end for
      $c(k) \leftarrow \arg\min_x$ Distance(x, k)
      $\Omega_{c(k)} \leftarrow \Omega_{c(k)} \cup \{k\}$
    end if
  end for
end while

References

1. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic resonance in medicine 56(2), 411–421 (2006)
2. Assaf, Y., Basser, P.J.: Composite hindered and restricted model of diffusion (charmed) mr imaging of the human brain. Neuroimage 27(1), 48–58 (2005)
3. Cabeen, R.P., Bastin, M.E., Laidlaw, D.H.: Estimating constrained multi-fiber diffusion mr volumes by orientation clustering. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pp. 82–89. Springer (2013)
4. Collard, A., Bonnabel, S., Phillips, C., Sepulchre, R.: Anisotropy preserving DTI processing. International Journal of Computer Vision 107(1), 58–74 (2014)
5. Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: A convergence study. Computer vision and image understanding 77(2), 192–210 (2000)
6. Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K., Sijbers, J.: Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. Human brain mapping 34(11), 2747–2766 (2013)
7. Malcolm, J.G., Shenton, M.E., Rathi, Y.: Filtered multitensor tractography. Medical Imaging, IEEE Transactions on 29(9), 1664–1675 (2010)
8. Ng, A.Y., Jordan, M.I., Weiss, Y., et al.: On spectral clustering: Analysis and an algorithm. Advances in neural information processing systems 2, 849–856 (2002)
9. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human brain mapping 15(1), 1–25 (2002)
10. 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)
11. Pasternak, O., Westin, C.F., Bouix, S., Seidman, L.J., Goldstein, J.M., Woo, T.U.W., Petryshen, T.L., Mesholam-Gately, R.I., McCarley, R.W., Kikinis, R., et al.: Excessive extracellular volume reveals a neurodegenerative pattern in schizophrenia onset. The Journal of Neuroscience 32(48), 17365–17372 (2012)
12. Peters, J.M., Taquet, M., Prohl, A.K., Scherrer, B., van Eeghen, A.M., Prabhu, S.P., Sahin, M., Warfield, S.K.: Diffusion tensor imaging and related techniques in tuberous sclerosis complex: review and future directions. Future neurology 8(5), 583–597 (2013)
13. Scherrer, B., Schwartzman, A., Taquet, M., Prabhu, S.P., Sahin, M., Akhondi-Asl, A., Warfield, S.K.: Characterizing the distribution of anisotropic micro-structural environments with diffusion-weighted imaging (diamond). In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pp. 518–526. Springer (2013)
14. Scherrer, B., Warfield, S.K.: Parametric representation of multiple white matter fascicles from cube and sphere diffusion mri. PLoS one 7(11), e48232 (2012)
15. Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., et al.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31(4), 1487–1505 (2006)
16. Taquet, M., Scherrer, B., Commowick, O., Peters, J., Sahin, M., Macq, B., Warfield, S.: A mathematical framework for the registration and analysis of multi-fascicle models for population studies of the brain microstructure. Medical Imaging, IEEE Transactions on 33(2), 504–517 (Feb 2014)

17. Taquet, M., Macq, B., Warfield, S.K.: A generalized correlation coefficient: Application to dti and multi-fiber dti. In: Mathematical Methods in Biomedical Image Analysis (MMBIA), 2012 IEEE Workshop on. pp. 9–14. IEEE (2012)
18. Taquet, M., Scherrer, B., Benjamin, C., Prabhu, S., Macq, B., Warfield, S.K.: Interpolating multi-fiber models by gaussian mixture simplification. In: Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on. pp. 928–931. IEEE (2012)
19. Taquet, M., Scherrer, B., Boumal, N., Macq, B., Warfield, S.K.: Estimation of a multi-fascicle model from single b-value data with a population-informed prior. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pp. 695–702. Springer (2013)
20. Taquet, M., Scherrer, B., Commowick, O., Peters, J., Sahin, M., Macq, B., Warfield, S.K.: Registration and analysis of white matter group differences with a multi-fiber model. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012, pp. 313–320. Springer (2012)
21. Taquet, M., Scherrer, B., Macq, B., Warfield, S.K., et al.: Multi-fascicle model reconstruction from acquisitions at a single b-value with a population-informed prior. In: Proc. the 21st International Symposium on Magnetic Resonance in Medicine (ISMRM). vol. 30 (2013)
22. Taquet, M., Scherrer, B., Peters, J.M., Prabhu, S.P., Warfield, S.K.: A fully bayesian inference framework for population studies of the brain microstructure. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014. Springer (2014)
23. Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Wedeen, V.J.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine 48(4), 577–582 (2002)
24. Wagstaff, K., Cardie, C., Rogers, S., Schrödl, S., et al.: Constrained k-means clustering with background knowledge. In: ICML. vol. 1, pp. 577–584 (2001)
25. Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C.: NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 61(4), 1000–1016 (2012)
26. Zhang, K., Kwok, J.T.: Simplifying mixture models through function approximation. In: Advances in Neural Information Processing Systems. pp. 1577–1584 (2006)
27. Zhang, K., Kwok, J.T.: Simplifying mixture models through function approximation. Neural Networks, IEEE Transactions on 21(4), 644–658 (2010)