HHS Public Access

Author manuscript

Med Image Anal. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Med Image Anal. 2015 December ; 26(1): 268–286. doi:10.1016/j.media.2015.10.004.

Improved Fidelity of Brain Microstructure Mapping from Single-Shell Diffusion MRI

Maxime Taquet[*,a,b], **Benoit Scherrer**[a], **Nicolas Boumal**[b,c], **Jurriaan M. Peters**[a], **Benoit Macq**[b], and **Simon K. Warfield**[a]

[a]Computational Radiology Laboratory, Boston Children's Hospital, Harvard Medical School, Wolbach 215, 300 Longwood Avenue, Boston MA 02115, USA

[b]ICTEAM Institute, Université catholique de Louvain, Avenue Georges Lemaitre, 4, B-1348 Louvain-la-Neuve, Belgium

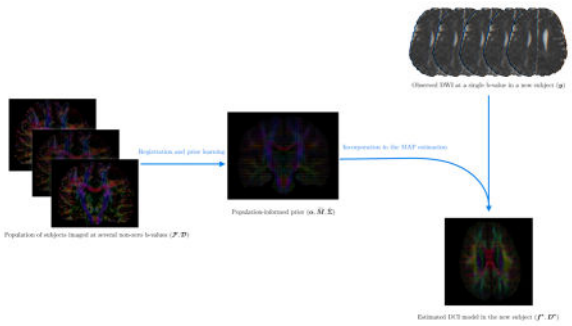
[c]Sierra Team, Computer Science Department, Ecole Normale Supérieure, 23 avenue d'Italie, CS 81321, 75214 PARIS Cedex 13

Abstract

Diffusion Weighted Imaging (DWI) is sensitive to alterations in the diffusion of water molecules caused by microstructural barriers. Different microstructural compartments are characterized by differences in DWI signal. Diffusion tensor imaging conflates the signal from these compartments into a single tensor, which poorly represents multiple white matter fascicles and extra-axonal space. Diffusion compartment imaging (DCI) models overcome this limitation by providing parametric representations for the signal contribution of each compartment, thereby improving the fidelity of brain microstructure mapping. However, current approaches fail to identify DCI model parameters from conventional single-shell DWI with the desired accuracy. It has been demonstrated that part of this inaccuracy is due to the ill-posedness of the estimation of DCI model parameters from conventional single-shell acquisitions. In this paper, we propose to regularize the estimation problem for single-shell DWI by learning a prior distribution of DCI model parameters from DWI acquired at multiple b-values in an external population of subjects. We demonstrate that this population-informed prior enables, for the first time, accurate estimation of DCI models from single-shell DWI typically acquired in clinical practice. We validated our approach on synthetic and *in vivo* data of healthy subjects and patients with autism spectrum disorder. We applied the approach to population studies of brain microstructure in autism and found that introducing a population-informed prior lead to reliable detection of group differences. Our algorithm enables novel investigation from large existing DWI datasets in normal development and in disease and injury.

Graphical Abstract

[*]Corresponding author: Maxime Taquet, maxime.taquet@childrens.harvard.edu, Computational Radiology Laboratory, Wolbach 215, 300 Longwood Avenue, Boston 02115, USA. Phone: +1 857-218-5172.



Keywords

Diffusion Compartment Imaging; HARDI; Diffusion-Weighted Imaging; Microstructure; Estimation; Population Studies

1. Introduction

Diffusion tensor imaging (DTI) provides a single tensor model to describe all of the signal arising from the water molecules diffusing within a single voxel. In voxels with multiple compartments, such as arises in regions of crossing fascicles, or with alterations in extra-axonal space, the mixing of the separate signals from these compartments into a single tensor fails to preserve the microstructural properties of the individual compartments. Furthermore, the interpretation of scalar measures from DTI, such as mean diffusivity (MD), fractional anisotropy (FA) may be misleading (Vos et al., 2012) when signal arising from multiple compartments is interpreted as if it arose from one compartment. Since 60–90% of voxels contain more than one fascicle at typical resolutions (Jeurissen et al., 2012), this limitation makes DTI inadequate in the vast majority of the white and gray matter. Various models have been proposed to overcome this limitation. Among them, diffusion compartment imaging (DCI) models such as multi-tensor models (Pasternak et al., 2009; Scherrer and Warfield, 2012; Tuch et al., 2002), NODDI (Zhang et al., 2012), CHARMED (Assaf and Basser, 2005) and DIAMOND (Scherrer et al., 2013a) seek to represent the contribution to the overall signal at a voxel arising from different populations of water molecules, and recent research has improved the tractability of their estimation (Daducci et al., 2015a,b). Since these models may improve the sensitivity and specificity of interpretation of the DWI signal, they are of great interest to characterize and compare brain tissue properties. However, the estimation of the parameters of these models from conventional data acquired at a single shell (diffusion weighting with a single non-zero b-value (Jones, 2004)) has proven ineffective.

Recent work has suggested that part of this inaccuracy is explained by the ill-posedness of the estimation problem and is not solely related to imaging noise or artifacts (Kreher et al., 2005; Scherrer and Warfield, 2010, 2012; Schultz et al., 2010; Taquet et al., 2013). Ill-posedness occurs when several different parameter values for the object being imaged can equally explain the acquired measurements, even in the absence of noise. In such a situation, it is impossible to identify the right parameter values. One possibility to circumvent this problem is to restrict the class of models by fixing *a priori* some of the model parameters. A

review of these constrained models is included in (Mishra et al., 2014). In (Kreher et al., 2005), pseudo tensors with undefined eigenvalues are estimated. This model does not allow measurements of diffusion properties (such as the fractional anisotropy and the mean diffusivity) for each fascicle, limiting its use in population studies of the brain microstructure. In (Tuch et al., 2002) and (Ramirez-Manzanares et al., 2007), the tensor eigenvalues are completely fixed a priori so that only the directions can be inferred. In (Anderson, 2005), the axial eigenvalue is fixed a priori, the radial eigenvalue is estimated from the mean signal and assumed equal for all fascicles in the voxel, and the orientation distribution is then estimated. In (Schultz et al., 2010), a ball-and-stick model is estimated, which assumes that fascicles are all identical and that there is no diffusion in directions orthogonal to the fascicle orientation. In (Malcolm et al., 2010) and (Zhan and Yang, 2006), the signal fractions are arbitrarily fixed to an equal value. Arbitrarily fixing some parameters leads to suboptimal models whose properties may not represent the true underlying brain microstructure and cannot be compared across groups in population studies.

Since the ill-posedness problem affects both the signal fractions and the tensor eigenvalues, unreasonable assumptions about any of those parameters necessarily impact the estimation of the others. Assumptions about the signal fractions are arbitrary since the location of the boundary between white matter fascicles, or between white and gray matter, are not known in advance and depend on the position of the subject in the scanner and the imaging resolution. As for fascicle properties, there is a large body of evidence showing that they differ throughout the brain (Aboitiz et al., 1992; Lamantia and Rakic, 1990) and across subjects due to normal development (Gao et al., 2009) and disease (Heads et al., 1991; Zikopoulos and Barbas, 2010). These changes in axonal diameter, density and myelination directly impact diffusion properties of fascicles.

Some researchers have proposed to rely on spatial priors to regularize the estimation problem. Spatial priors express that DCI models in adjacent voxels are not independent and can be an excellent choice for modeling the underlying spatial homogeneity of the brain. However, most currently exploited spatial priors do not address the ill-posedness of the estimation problem caused by insufficient b-values (as we show in Section 2.2). Pasternak *et al.* proposed a piece-wise smoothness constraint to regularize the estimation of a single-tensor model with a free-water compartment (Pasternak et al., 2009, 2012a). In (Scherrer and Warfield, 2012), an anisotropic spatial prior is used to regularize a multi-tensor field. Importantly, we will show in Section 4.5 that relying solely on such spatial priors may mislead the conclusions drawn from population studies.

To regularize the estimation problem while keeping all the degrees of freedom of the model, Scherrer and Warfield proposed to acquire data at several b-values (Scherrer and Warfield, 2012). Pasternak *et al.* demonstrated that even for a simpler model including only one anisotropic tensor and an isotropic tensor, inclusion of multiple non-zero b-values improves the accuracy of the estimation (Pasternak et al., 2012a). In (Taquet et al., 2013), we considered the problem of estimating a multi-tensor model from single-shell HARDI data. However, it remains unclear whether any of these approaches are effective. In particular, there is no theoretical demonstration that the particular ill-posedness identified in (Kreher et al., 2005; Scherrer and Warfield, 2012) and further explored in (Taquet et al., 2013) is

alleviated when DWI are acquired at multiple b-values. Furthermore, it is unclear whether the approach proposed in (Taquet et al., 2013) leads to biased estimates of DCI models and whether they enable the detection of true brain microstructure differences in population studies. More work is thus needed to establish the appropriate theory and to demonstrate the efficacy of any solution.

Many large diffusion imaging datasets have been acquired with a single-shell HARDI acquisition sequence. For instance, the Pediatric Imaging, Neurocognition, and Genetics Study (PING) includes diffusion images from 800 children at $b = 0 s/mm^2$ and $b = 1000 s/mm^2$ (Fjell et al., 2012). Such datasets hold promise to better understanding the relation between brain microstructure and neuropsychological development. The inability to estimate multi-fascicle models from single b-value data, however, jeopardizes the potential of these data to reveal novel insight into the brain microstructure. There is therefore a strong need for a method to estimate DCI models from data acquired at a single non-zero b-value. A method to estimate a general N-tensor model can be applied to many DCI models since most of them include a multi-tensor of some kind (Panagiotaki et al., 2012), as, from first principles, individual spin packets exhibit monoexponential decay (Yablonskiy et al., 2003). In the DIAMOND model, the mode of each of the matrix-variate Gamma distribution is a tensor (Scherrer et al., 2013a). In NODDI, the model of the CSF compartment and the mode of the extracellular model is a multi-tensor (Zhang et al., 2012). In CHARMED, the multiple hindered compartments define a multi-tensor (Assaf et al., 2004).

This paper proposes a method to estimate a general N-tensor model from single-shell HARDI data and demonstrates its effectiveness in population studies of the brain microstructure. We start by considering and generalizing the ill-posedness problem proposed in (Kreher et al., 2005; Scherrer and Warfield, 2012) and provide, for the first time, a theoretical demonstration that this ill-posedness is alleviated insofar as at least two different non-zero b-values are used in the acquisition, regardless of the number N of tensors. To remove the ill-posedness in the estimation from data at a single non-zero b-value, we then propose to incorporate prior knowledge from data acquired in an external population of subjects which are imaged at multiple b-values. Section 2 describes the main contributions of this paper. Section 3 describes how the population-informed prior is used to estimate DCI models. Section 4 presents experimental results. Section 5 summarizes the main contributions of this paper.

2. Material

In this section, we first elucidate the ill-posedness of the estimation problem for N-fascicle models through a geometric argument (Section 2.1). We then demonstrate why most spatial priors cannot be used to resolve the ill-posedness from the estimation of the full DCI model (Section 2.2). Finally, we derive an expression for a population-informed prior to regularize the estimation problem (Section 2.3). We will denote random variables with sans-serif fonts (a for scalar, $\mathbf{a}$ for vector, and $\mathbf{A}$ for matrices), and constants or observations with serif fonts (a for scalars, $\boldsymbol{a}$ for vector and $\boldsymbol{A}$ for matrices).

2.1. Manifolds of Equivalent Models at a Given b-Value

Multi-tensor models represent the formation of the diffusion signal S for a b-value b and a gradient direction $\boldsymbol{g}$ (with $\|\boldsymbol{g}\| = 1$) by:

$$S(\boldsymbol{g}, b) = S_0 \sum_{i=1}^{N} f_i e^{-b\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}}, \quad (1)$$

where $\boldsymbol{D}_i$ and f_i are the tensor and the signal fraction of the i-th fascicle. The goal of the estimation problem is to determine the values of the parameters $\boldsymbol{D}_i$ and f_i from a set of measurements $S(\boldsymbol{g}, b)$ acquired for different gradient directions $\boldsymbol{g}$ and b-values b.

Since, for any $\gamma_i > 0$, $\gamma_i e^{-\log\gamma_i} = 1$, all multi-fascicle models with fractions $\gamma_i f_i$ and tensors $\boldsymbol{D}_i + \frac{\log\gamma_i}{b}\boldsymbol{I}$ generate the same values of the diffusion signal at a given b-value b:

$$S(\boldsymbol{g}, b) = S_0 \sum_{i=1}^{N} \gamma_i f_i e^{-b\boldsymbol{g}^T \left(\boldsymbol{D}_i + \frac{\log\gamma_i}{b}\boldsymbol{I}\right)\boldsymbol{g}}, \quad \text{with the constraint } \sum_{i=1}^{N} \gamma_i f_i = 1. \quad (2)$$

The tensors remain positive definite as long as $\gamma_i > e^{-b\lambda_i^{\min}}$, where $\lambda_i^{\min}$ is the lowest eigenvalue of $\boldsymbol{D}_i$. These models are all equally compatible with the observed measurements of the signal S at a single b-value b. Selection of a particular model based solely on the measurements at that particular single b-value is therefore purely arbitrary. Since the eigenvectors of $\boldsymbol{D}_i + \alpha\boldsymbol{I}$ are equal to those of $\boldsymbol{D}_i$, the principal directions of diffusion are not altered by the transformation (2). In terms of fractions and eigenvalues, the equivalence between models can be written:

$$\lambda_{i,j} \to \lambda_{i,j} + \frac{\log\gamma_i}{b} \quad \text{and} \quad f_i \to \gamma_i f_i \quad \Rightarrow \quad S(\boldsymbol{g}, b) \to S(\boldsymbol{g}, b), \quad \forall \boldsymbol{g}, \quad (3)$$

where $\lambda_{i,j}$ is the j-th eigenvalue ($j = 1, 2, 3$ with $\lambda_{i,1} \geq \lambda_{i,2} \geq \lambda_{i,3}$) of the i-th tensor. Varying all the γ_i under the constraint that $\sum_{i=1}^{N} \gamma_i f_i = 1$ defines an equivalence class of multi-tensor models that all present with the same values of the diffusion signal at a single b-value b. All the multi-tensor models in this equivalence class have a different set of lowest eigenvalues $(\lambda_{i,3})_{i=1,\ldots,N}$. Therefore each model satisfying (3) is uniquely identified by its vector $(\lambda_{1,3}, \ldots, \lambda_{N,3})$. In the N-dimensional space of lowest eigenvalues, the equivalence class defined by (3) is a manifold of dimension $(N - 1)$ defined by the implicit equations (we let $\lambda_i \triangleq \lambda_{i,3}$ for clarity):

$$\begin{cases} \lambda_i = \lambda_i^* + \frac{1}{b}\log(\gamma_i) & \text{for } i = 1, \ldots, N \\ \sum_{i=1}^{N} \gamma_i f_i^* = 1, \end{cases} \quad (4)$$

where $\{(f_i^*, \lambda_i^*)_{i=1,\ldots,N}\}$ is any fixed reference model on the manifold. This reference can be arbitrarily selected from the equivalence class. Without loss of generality, we therefore select it as the true (unknown) underlying model (Fig. 1(a)). An explicit equation of the

manifold can be obtained by eliminating the γ's between Equations (4) and expressing the lowest eigenvalue of the N-th tensor λ_N as a function of the $(N-1)$ others:

$$\lambda_N(\lambda_1,\ldots,\lambda_{N-1},b)=\lambda_N^*+\frac{1}{b}\log\left(\frac{1-\sum_{i=1}^{N-1}f_i^*e^{b(\lambda_i-\lambda_i^*)}}{f_N^*}\right). \quad (5)$$

This manifold describes the set of models that all generate the same diffusion signal at a single b-value b. Of course, any model on that manifold generates a diffusion signal whose value differs for different b-values. At b-values other than b, the signal generated by different models from that manifold may well be different so that these models could be distinguished if measurements at several b-values were available (Fig. 1(b)). In the formalism of manifolds, acquiring diffusion images at different b-values therefore amounts to defining different equivalence manifolds. Each such manifold is characterized by a single b-value b, which influences its shape (as defined by Equation (5)), and contains different models that are indistinguishable from measurements at that single b-value b. The true underlying model defined by $\lambda_i=\lambda_i^*$ must, by definition, be at the intersection of all these manifolds since it is compatible with the different measurements at the different b-values (Fig. 1(b)). This intersection is defined by a fixed vector of λ_i and therefore by a fixed set of tensors $\boldsymbol{D}_i$ and a fixed vector of fractions.

Remarkably, it can be demonstrated that, insofar as at least two non-zero b-values (b, $b'>0$, $b\neq b'$) are used in the acquisition, the true underlying model is the only one that satisfies the equations of $\lambda_N(\lambda_1,\ldots,\lambda_{N-1},b)$ and $\lambda_N(\lambda_1,\ldots,\lambda_{N-1},b')$, *i.e.*, the two manifolds described by two different b-values share only one intersection (Appendix A). This unicity property implies that the ill-posedness problem highlighted in Equation (3) is removed as soon as at least two non-zero b-values are used in the acquisition, regardless of the number N of fascicles and the actual b-values used.

Despite the unicity of the intersection of manifolds when two or more non-zero b-values are used, estimating a multi-fascicle model from data at several b-values remains sensitive to noise (Pasternak et al., 2012a; Scherrer and Warfield, 2012). This sensitivity can be explained by the geometry of the manifold of equivalent models. At their intersection, manifolds of equivalent models obtained for different b-values intersect tangentially (see proof in Appendix B). In other words, at the first-order approximation, the manifolds at all b-values coincide locally, explaining the high sensitivity to noise encountered when optimizing the parameters of a multi-fascicle model (Fig. 1(b)). This is akin to the sensitivity of the intersection point of two almost-parallel lines in the plane: small perturbations of their slope can have a dramatic effect.

From the analysis of the geometry of the manifolds in the vicinity of their intersection, it can also be shown that the difference in curvature between the manifolds at two different b-values (b and b') is proportional to the difference in b-values ($b-b'$) (Appendix B). It appears that a wider range of b-values therefore leads to a larger difference between their manifolds, which should in turn improve the accuracy of the estimation (ignoring the potential impact of b on noise). Continuing with a different analogy, this is akin to having a small ball stuck

inside a larger ball, such that they intersect at a single point. Moving the small ball gently displaces the intersection point somewhat. On the other hand, if the inner ball is nearly as large as the outer one (almost the same b-value), then small perturbations (due to noise) can lead to large displacements of the contact point.

When an isotropic compartment $f_{\text{iso}}e^{-bD_{\text{iso}}}$ is explicitly included in the model, one can show that the above development remains valid both with an unchanged N if D_{iso} is known (typically set to the diffusion of water at 37°C, that is $3 \times 10^{-3} mm^2/s$), and considering an (N +1)-fascicle model if D_{iso} is estimated from the DWI data. In summary, estimating multi-fascicle models from single b-value data is an ill-posed problem that has an infinite number of equivalent solutions. However, due to the particular geometry of the manifold of these solutions, acquiring data at two b-values is sufficient to resolve the ambiguity due to the equivalence relation (4). Acquiring additional b-values reduces the sensitivity of the system of equations to noise.

2.2. Ill-Posedness with Spatial Priors

This section demonstrates that the use of the most common spatial priors, as described in (Pasternak et al., 2009, 2012a; Scherrer and Warfield, 2012), does not alleviate the ill-posedness problem of the estimation. In their most general expression, these spatial priors impose a penalty on multi-fascicle models based on some discrepancy between the multi-tensor at one voxel $\mathcal{M}(\boldsymbol{x})$ and the multi-tensors in the neighborhood $\Omega_{\boldsymbol{x}}$:

$$p(\{\mathcal{M}(\boldsymbol{x}')|\boldsymbol{x}' \in \Omega_{\boldsymbol{x}}\}) = f(\{d(\mathcal{M}(\boldsymbol{x}'), \mathcal{M}(\boldsymbol{x}))|\boldsymbol{x}' \in \Omega_{\boldsymbol{x}}\}), \quad (6)$$

where $d(\mathcal{M}(\boldsymbol{x}'), \mathcal{M}(\boldsymbol{x}))$ is some discrepancy function between the multi-tensors $\mathcal{M}(\boldsymbol{x})$ and $\mathcal{M}(\boldsymbol{x}')$. The expressions for the spatial prior may vary (through the choice of f, d and/or $\Omega_{\boldsymbol{x}}$) but they share the property of reaching a maximum if the local neighborhood is constant (*i.e.*, if $\mathcal{M}(\boldsymbol{x}') = \mathcal{M}(\boldsymbol{x})$, $\forall \boldsymbol{x}' \in \Omega_{\boldsymbol{x}}$), since, typically, $d(\mathcal{M}, \mathcal{M}) = 0$, for any $\mathcal{M}$. For instance, the prior used in (Scherrer and Warfield, 2012) for a multi-fascicle model that has N tensors $\boldsymbol{D}_i(\boldsymbol{x})$ at location $\boldsymbol{x}$ is:

$$p_{\text{spatial}}(\mathbf{D}) \propto exp\left\{-\frac{1}{2\sigma_s^2}\sum_{\boldsymbol{x}}\sum_{i=1}^{N}\|\nabla \log \mathbf{D}_i(\boldsymbol{x})\|_{\text{F}}^2\right\}, \quad (7)$$

where ∇ is the spatial gradient with respect to $\boldsymbol{x}$ approximated by finite differences (Jordán, 1965), with $\mathbf{D} \triangleq (\mathbf{D}_1, \ldots, \mathbf{D}_N)$. The gradient ∇ is computed along the fascicle by selecting for each $\mathbf{D}_i$ at $\boldsymbol{x}$, the two tensors in $\Omega_{\boldsymbol{x}}$ most similar to $\mathbf{D}_i(\boldsymbol{x})$. In a constant field of multi-tensor models, all gradients are equal to zero and the prior is maximum. Therefore, two constant multi-tensor fields must have equal spatial priors. Given a constant multi-fascicle model, one can generate an infinite number of constant multi-fascicle models from (4) that the data alone cannot distinguish. Since the spatial prior is unable to distinguish them either, the ill-posedness problem persists.

Spatial priors that do not follow Equation (6) may be proposed. For instance, one may propagate properties of tensors from single-tensor to multi-tensor areas (Schultz, 2012b). Since the estimation in single-tensor areas is not ill-posed, such a prior would, in theory,

avoid the ill-posedness problem at the expense of strong assumptions about the homogeneity of diffusion properties along the fascicles. More work is needed to assess whether such priors would be beneficial for population studies of the brain microstructure.

2.3. Posterior Predictive Distribution of the Parameters

While all models of (4) are equally compatible with the observed DWI at a given b-value, they are not all equally likely from a biological point of view. This knowledge can be learned from available observations of multi-fascicle models at multiple non-zero b-values in other subjects. Since these subjects are imaged at multiple non-zero b-values, their multi-fascicle models are not affected by this ill-posedness problem and they are therefore reliable observations of the anatomy.

For a fascicle i at a specific voxel, we denote by $\mathscr{F}_i=\{f_i^0,\dots,f_i^m\}$ the observations of its i-th fraction and by $\mathscr{D}_i=\{\boldsymbol{D}_i^0,\dots,\boldsymbol{D}_i^{m_i}\}$ the observations of its i-th tensor in m other subjects. Notice that the number of observations for the fraction is always equal to the number m of subjects while the number of observed tensors is $m_i \leq m$ since a compartment with a null fraction has undetermined tensor. We let $\boldsymbol{\mathscr{F}}=(\mathscr{F}_1,\dots,\mathscr{F}_N)$ and $\mathscr{D}=(\mathscr{D}_1,\dots,\mathscr{D}_N)$ for the observations of all fractions and all tensors in one voxel. These observations can be incorporated into the estimation of a multi-fascicle model in a new subject, as a prior over the parameters $(\mathrm{f}_i, \mathbf{D}_i)$ (Fig. 1(c)). This section derives an expression for the prior probability $p(\mathbf{f}, \mathbf{D}|\boldsymbol{\mathscr{F}}, \mathscr{D})$ for the fractions and tensors in one voxel given the observations at the same voxel in other subjects. We assume that the fascicle properties are independent of partial voluming and that the properties of one fascicle are independent of those of another (these assumptions are experimentally tested in Section 4.1). The prior can therefore be expressed as:

$$p(\mathbf{f},\mathbf{D}|\boldsymbol{\mathscr{F}},\mathscr{D})=p(\mathbf{f}|\boldsymbol{\mathscr{F}})\prod_{i=1}^{N}p(\mathbf{D}_i|\mathscr{D}), \quad (8)$$

We employ the following strategy to derive an expression for this prior. First, we express a likelihood over the model parameters $\mathbf{f}$, $\mathbf{D}_i$, whose expression depends on unknown parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}_i$. Second, we express non-informative hyperpriors over these unknown parameters. Finally, we derive, from the resulting hierarchical model, an expression for the posterior predictive distribution over new observations given past observations. This posterior predictive distribution is the population-informed prior.

The fractions are not independent since they sum to 1. However, we assume that any fraction f_i is independent of the relative proportions of others $\mathrm{f}_j/(1-\mathrm{f}_i)$. This *neutral vector* assumption naturally leads to the Dirichlet distribution for the likelihood:

$$p(\mathbf{f}|\boldsymbol{\alpha})=\frac{\mathbb{1}_{\mathbf{f}\in\mathscr{S}}}{B(\boldsymbol{\alpha})}\prod_{i=1}^{N}\mathrm{f}_i^{\alpha_i-1},\ \text{where}\ \mathscr{S}=\left\{x\in\mathbb{R}^N : x_i>0, \sum_{i=1}^{N}x_i=1\right\}, \quad (9)$$

where $\mathbb{1}_{\mathbf{f}\in\mathscr{S}}$ equals one if and only if $\mathbf{f}\in\mathscr{S}$ and zero otherwise. Calculus of diffusion tensor data is typically performed in the log-Euclidean domain since it resembles the Euclidean

space and avoids nonsensical negative eigenvalues (Arsigny et al., 2006). The prior knowledge about $\mathbf{D}_i$ can be described as a multivariate Gaussian distribution over their logarithm (Arsigny et al., 2006; Schwartzman et al., 2008):

$$\mathbf{L}_i = \log \mathbf{D}_i \sim \mathcal{N}(\mathbf{M}_i, \sum\nolimits_i). \quad (10)$$

In this notation, $\mathbf{L}_i$ is vectorized as $(\mathrm{diag}(\mathbf{L}_i), \sqrt{2}\,\mathrm{offdiag}(\mathbf{L}_i))$. In general, $\mathbf{\Sigma}_i$ has 21 free parameters, which may overfit the usually small training dataset. For DTI, Schwartzman *et al.* suggested in (Schwartzman et al., 2008) to constrain $\mathbf{\Sigma}_i$ to be orthogonally invariant, imposing the following structure that depends only on σ_i and τ_i:

$$\sum\nolimits_i = \sigma_i^2 \begin{pmatrix} \boldsymbol{I}_3 + \frac{\tau_i}{1-3\tau_i}\boldsymbol{I}_3 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_3 \end{pmatrix} \triangleq \boldsymbol{B}(\sigma_i, \tau_i).$$

Orthogonal invariance implies that a deviation $\mathbf{\Delta}_i$ from the mean $\mathbf{M}_i$ is as likely to occur as its rotated versions $\boldsymbol{Q}\boldsymbol{\Delta}_i\boldsymbol{Q}^T$ for any orthogonal matrix $\boldsymbol{Q}$. This yields a closed-form solution for the maximum likelihood (ML) estimator (Schwartzman et al., 2008):

$$\hat{\boldsymbol{M}}_i = \overline{\boldsymbol{L}}_i = \frac{1}{m_i}\sum_{k=1}^{m_i} \boldsymbol{L}_i^k \quad \text{and} \quad \hat{\sum\nolimits_i} = \boldsymbol{B}(\hat{\sigma}_i, \hat{\tau}_i), \quad (11)$$

$$\text{with} \quad \hat{\tau}_i = \frac{-\sum_{i=1}^{m_i} \|\boldsymbol{L}_i^k - \overline{\boldsymbol{L}}_i\|_2^2}{5\sum_{i=1}^{m_i}\left[\mathrm{Tr}(\boldsymbol{L}_i^k - \overline{\boldsymbol{L}}_i)\right]^2} \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{6m_i}\sum_{i=1}^{m_i} \|\boldsymbol{L}_i^k - \overline{\boldsymbol{L}}_i\|_{\hat{\tau}_i}^2, \quad (12)$$

where $\|A\|_t^2 = \langle A, A\rangle_t$ and $\langle \boldsymbol{A}, \boldsymbol{B}\rangle_t = \mathrm{Tr}(\boldsymbol{AB}) - t\,\mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{B})$. The ML estimator may be unreliable for compartments with only a few observations. This uncertainty is accounted for by replacing point estimates of $\boldsymbol{\theta}$ by posterior distributions and integrating over all possible $\boldsymbol{\theta}$. This yields the *posterior predictive distribution* which encodes all the knowledge about new observations that we learn from previous observations. Its derivation requires the definition of hyperpriors over $\boldsymbol{\theta}$ and has closed-form if we select conjugate hyperpriors. $\mathbf{M}_i \sim \mathcal{N}(\boldsymbol{M}_0, \mathbf{\Lambda}_0)$ is a conjugate hyperprior for the tensor part of (8) assuming a deterministic $\mathbf{\Sigma}_i = \hat{\mathbf{\Sigma}}_i$. We set $\mathbf{\Lambda}_0 = \boldsymbol{B}(1, 0)$ and $\boldsymbol{M}_0 = \log \boldsymbol{D}_{\mathrm{iso}}$ to keep it weakly informative (this hyperprior merely encodes the order of magnitude of diffusivity at 37°*C*). The posterior predictive distribution over the tensors becomes:

$$\mathbf{L}_i | \mathcal{D}_i \sim \mathcal{N}\left(\tilde{\boldsymbol{M}}_i, \tilde{\sum\nolimits_i}\right), \quad (13)$$

$$\text{with} \quad \tilde{\sum\nolimits_i} = \hat{\sum\nolimits_{\boldsymbol{i}}} + \left(\mathbf{\Lambda}_0^{-1} + m_i \hat{\sum\nolimits_{\boldsymbol{i}}}^{-1}\right)^{-1} \triangleq \boldsymbol{B}(\tilde{\sigma}_i, \tilde{\tau}_i), \quad (14)$$

$$\text{and} \quad \tilde{\boldsymbol{M}}_i = \left(\boldsymbol{\Lambda}_0^{-1} + m_i \hat{\sum}_{\boldsymbol{i}}^{-1}\right)^{-1} \left(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{M}_0 + m_i \hat{\sum}_{\boldsymbol{i}}^{-1} \overline{L}_i\right). \quad (15)$$

For the parameters α_i, a conjugate hyperprior is the Dirichlet distribution. We set all its parameters to 1, making it uniform over the simplex $\mathscr{S}$. The resulting posterior predictive distribution is a Dirichlet with parameters $1+\sum_{k=1}^{m} f_i^k$. The complete posterior predictive distribution is represented as a graphical model in Fig. 2 and its expression is (with $C_{\mathscr{F},\mathscr{D}}$ constant):

$$p(\mathbf{f}, \mathbf{D}|\mathscr{F}, \boldsymbol{D}) = C_{\mathscr{F},\mathscr{D}} \mathbb{1}_{\mathbf{f}\in\mathscr{S}} \prod_{i=1}^{N} \mathrm{f}_i^{\sum_{k=1}^{m} f_i^k} \prod_{i=1}^{N} \exp\left\{-\frac{\|\mathbf{L}_i - \tilde{\boldsymbol{M}}_i\|^2_{\tilde{\tau}_i}}{2\tilde{\sigma}_i^2}\right\}. \quad (16)$$

This prior pertains to a single voxel. The population-informed prior for the entire multi-fascicle model is obtained by multiplying this prior for all location $\boldsymbol{x}$. Dependence between adjacent voxels can be accounted for by a separate spatial prior. This population-informed prior encodes our *a priori* knowledge of the brain microstructure at every location. In the next section (Section 2.4), we will see how this prior is incorporated in a maximum a posteriori estimate of the model parameters.

2.4. Maximum A Posteriori Estimation of the Model Parameters

The population-informed prior is incorporated in the estimation as a prior over the model parameters. The maximum a posteriori (MAP) estimate of the parameters is:

$$\boldsymbol{f}^*, \boldsymbol{D}^* = \arg\max_{\mathbf{f},\mathbf{D}} \left(p(\mathbf{f}, \mathbf{D}|\mathscr{F}, \mathscr{D}) \bullet p_{\text{spatial}}(\mathbf{f}, \mathbf{D}) \bullet p_{\text{likelihood}}(\mathbf{y}|\mathbf{f}, \mathbf{D})\right).$$

We use the spatial prior $p_{\text{spatial}}(\mathbf{f}, \mathbf{D})$ defined in (Scherrer and Warfield, 2012) and recalled in (7). The likelihood density is over measurements $\mathbf{y} = (\mathrm{y}_1, \ldots, \mathrm{y}_K)$, that is the set of K DWI. The expression of the likelihood density depends on the noise model. This noise model pertains to data acquired at a single b-value. This b-value is typically low so that we assume Gaussian noise with variance $\boldsymbol{\sigma}^2_{\text{noise}}$ (Gudbjartsson and Patz, 1995) (other models can be incorporated in the same way in the estimation):

$$p_{\text{likelihood}}(\mathbf{y}|\mathbf{f}, \mathbf{D}) \propto \prod_{\boldsymbol{x}} \prod_{k=1}^{K} \exp\left(-\frac{1}{2\sigma^2_{\text{noise}}} \left(S_k(\mathbf{f}(\boldsymbol{x}), \mathbf{D}(\boldsymbol{x})) - \mathrm{y}_k(\boldsymbol{x})\right)^2\right),$$

where S_k is the modeled signal obtained with (1) for volume fractions $\mathbf{f}(\boldsymbol{x})$ and tensors $\mathbf{D}(\boldsymbol{x})$, given the b-value and the gradient orientation of the measured y_k.

The MAP estimate at each voxel amounts to solving the following optimization:

$$\boldsymbol{f}^*, \boldsymbol{D}^* = \arg\max_{\mathbf{f},\mathbf{D}} \sum_{\boldsymbol{x}} \Bigg[\log(\mathbb{1}_{\mathbf{f}\in\mathscr{S}}) + \sum_{i=1}^{N}\sum_{j=1}^{m} f_i^j(\boldsymbol{x}) \log \mathrm{f}_i(\boldsymbol{x}) - \sum_{i=1}^{N} \frac{\|\mathbf{L}_i(\boldsymbol{x}) - \tilde{\boldsymbol{M}}_i(\boldsymbol{x})\|^2_{\tilde{\tau}_i}}{2\tilde{\sigma}_i^2} - \frac{1}{2\sigma_s^2} \sum_{i=1}^{N} \|\nabla \log \mathbf{D}_i(\boldsymbol{x})\|^2_{\mathrm{F}} - \frac{1}{2\sigma^2_{\mathrm{noise}}} \sum_{k=1}^{K} \left(S_k(\mathbf{f}(\boldsymbol{x}), \mathbf{D}(\boldsymbol{x})) - \mathrm{y}_k(\boldsymbol{x})\right)^2 \Bigg]. \quad (17)$$

The first two lines are the population-informed prior, the third line is the spatial prior and the fourth line is the likelihood. The population-informed prior is learned in closed form and therefore does not introduce additional parameters. The estimation of the optimal values for σ_s^2 and $\sigma^2_{\mathrm{noise}}$ is investigated in the next section. As in (Scherrer and Warfield, 2012), Equation (17) is maximized using the BOBYQA algorithm (Powell, 2009) that allows the introduction of constraints. The values of all BOBYQA parameters are set to those of (Scherrer and Warfield, 2012). To select the number *N* of fascicles, we use an *F*-test with a threshold t=25 as in (Scherrer and Warfield, 2012) due to its widespread use (Alexander et al., 2002; Kreher et al., 2005; Scherrer and Warfield, 2012) to isolate the impact of the population-informed prior when comparing the results with those of earlier methods. Recent developments have, however, enabled more reliable selection of the number of fascicles (Scherrer et al., 2013b; Schultz, 2012a) and can be used with the proposed estimation.

2.5. Estimation Bias and Detectability of Group Differences

The introduction of an informed prior in the estimation implies that estimated models will be closer to the mean of the population than they would be if the estimation problem was well-posed and if no prior was used. This is the *shrinkage towards the mean* effect that is common to any parameter estimation expressed as a maximum a posteriori. In this section, we investigate, theoretically, the impact of this effect on the bias of the estimates and on their capability to detect group differences. In Section 4.2, we present empirical evidence demonstrating the main conclusions of this section.

The use of a population-informed prior assumes that the population is well represented in the samples used to build the prior. If a sufficient coverage of the population is used to generate the prior, then the use of the prior does not bias further estimates in the same population. In other words, estimates of multi-fascicle models from new control subjects are closer to the mean of the prior population but the average estimated model in new control subjects corresponds to the average model in the prior population which is asymptotically the true average of the control population. In contrast, maximum likelihood estimates (MLE) are ill-posed and their average corresponds to the average model of a manifold of equivalent models. This average model may substantially differ from the true average control subject, resulting in higher biases in the estimates (Fig. 3).

When the population-informed prior is used to estimate multi-tensor models of patients in voxels with a group difference, the estimation is biased because the average patient model is

brought closer to the average control model. Importantly, for population studies, the prior ought to be built from control subjects only. Otherwise, blindness in the experimental design is lost and the reported p-values may not be valid. With a population-informed prior based upon control subjects only, the resulting inference of group differences is more conservative and no additional false positives are expected. To appraise the extent of this bias, let us consider virtual DWI from an average patient in a single voxel. To identify the contribution of the prior in the estimation error, let us consider that measurements are noise-free. In this scenario, all models on the equivalence manifold have equal likelihood and all models outside of the manifold have zero likelihood. The MAP is therefore the model from the equivalence manifold that maximizes the population-informed prior. From the MAP expression (17) and from the implicit equation of the equivalence manifold (4), this optimal model is characterized by an optimal vector $\boldsymbol{\gamma}^{\text{opt}}$ given by:

$$\boldsymbol{\gamma}^{\text{opt}} = \arg\max_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \left[\left(\sum_{k=1}^{m} f_i^k \right) \log(\gamma_i f_i^*) - \frac{\| \log\left(\boldsymbol{D}_i^* + \frac{1}{b}\log(\gamma_i)\boldsymbol{I}\right) - \tilde{\boldsymbol{M}}_i \|_{\tilde{\tau}_i}^2}{2\tilde{\sigma}_i^2} \right], \quad (18)$$

under the constraint

$$\sum_{i=1}^{N} \gamma_i f_i^* = 1.$$

Optimizing the terms in f_i^* tends to bring the fractions of the estimated model $\gamma_i f_i^*$ close to the average fractions in the prior population $\frac{1}{m}\sum_k f_i^k$ whereas optimizing the term in $\boldsymbol{D}_i^*$ tends to bring the *i*-th log-tensor close to the average log-tensor in the prior population $\tilde{\boldsymbol{M}}_i$. Group differences typically affect specific properties of the microstructure (*e.g.*, the radial diffusivity of one fascicle). In terms of unaffected properties, both the population-informed prior and the likelihood are simultaneously maximized at the true underlying model since the average patient has, by definition, the same values of unaffected properties as the average control subject. Importantly, because the number of degrees of freedom ($N - 1$) is smaller than the number of parameters of the model ($7N - 1$), reducing the bias in unaffected properties also decreases the bias in the estimation of affected properties. The terms in (18) that relate to unaffected properties therefore drive the model towards the true model. The bias in the estimation thus comes exclusively from the terms in (18) that pertain to affected properties and is mitigated by the other terms. Furthermore, even within the terms that contain affected properties (*e.g.*, the term in $\boldsymbol{D}_i^*$ whose principal eigenvalue presents with a group difference), the bias is mitigated by the other properties of the same term that are not affected (*e.g.*, the second and third eigenvalues of $\boldsymbol{D}_i^*$). To illustrate the mitigation of the estimation bias by unaffected model parameters, the different factors of the population-informed prior are depicted in Fig. 4 for a two-tensor model whose principal eigenvalue of the first tensor is significantly decreased in patients. As expected, the terms related to $\boldsymbol{D}_2$ and to the fractions are maximized at the true underlying model, whereas the term in $\boldsymbol{D}_1$ is offset due to the presence of a group difference.

3. Methods

In this section, we specify how the population-informed prior is constructed and we provide details about the experimental setup used to validate our method.

3.1. Construction and Evaluation of the Population-Informed Prior

The population-informed prior is built from data acquired in completely different subjects at several b-values. These data are used to compute multi-fascicle models for each subject. These multi-fascicle models are then registered to a multi-fascicle atlas, using the method described in (Taquet et al., 2012b). In applying the transformation to the multi-fascicle models, the reorientation of tensors is performed using the finite-strain rationale first proposed in (Alexander et al., 1999, 2001; Ruiz-Alzola et al., 2000) adapted to multi-tensor models (Taquet et al., 2014a). Following alignment, tensors from all subjects at each voxel are clustered in N compartments by minimizing the cumulative relative differential entropy as described in (Taquet et al., 2012a). The number N of compartments is set to the maximum number of tensors observed among subjects at that location. Each cluster represents the sets $\mathscr{F}_i$ and $\mathscr{D}_i$ of available observations for the i-th fascicle and the prior is built following the procedure described in Section 2.3.

To evaluate the prior at a given voxel $\boldsymbol{x}$ for a given candidate model ($\boldsymbol{f}(\boldsymbol{x})$, $\boldsymbol{D}(\boldsymbol{x})$), we need the prior to be aligned to the subject whose multi-fascicle model is being estimated. We also need to identify which fascicle of the prior corresponds to what fascicle of the model being estimated. Alignment of the prior is performed by first estimating a constrained multi-fascicle model without the population-informed prior and by performing multi-fascicle registration (Taquet et al., 2012b) between this first estimate and the multi-fascicle atlas. The constrained model is a fixed response model whose description is provided in Section 4.4. Associations between the fascicles of the model and the fascicles of the prior is achieved by computing the prior for all possible associations of fascicles and recording the maximum value obtained.

3.2. Comparison Metrics

A multi-fascicle model ($\boldsymbol{f}$, $\boldsymbol{D}$) is compared to a ground truth ($\tilde{\boldsymbol{f}}$, $\tilde{\boldsymbol{D}}$), with the following six metrics at every voxel ($\boldsymbol{e}_{1,i}$ is the principal eigenvector of $\boldsymbol{D}_i$ with unit norm):

$$\Delta^2_{\mathrm{FA}} = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \left(\mathrm{FA}(\boldsymbol{D}_i) - \mathrm{FA}(\tilde{\boldsymbol{D}}_i) \right)^2, \quad (19)$$

$$\Delta^2_{\mathrm{MD}} = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \left(\mathrm{MD}(\boldsymbol{D}_i) - \mathrm{MD}(\tilde{\boldsymbol{D}}_i) \right)^2, \quad (20)$$

$$\mathrm{Fro}^2 = \sum_{i=1}^{N} \frac{f_i + \tilde{f}_i}{2} \| \boldsymbol{D}_i - \tilde{\boldsymbol{D}}_i \|_F^2, \quad (21)$$

$$\Delta_{\mathrm{Dir}}=\sum_{i=1}^{N}\frac{f_i+\tilde{f}_i}{2}(1-|\boldsymbol{e}_{1,i}\cdot\tilde{\boldsymbol{e}}_{1,i}|), \quad (22)$$

$$\Delta_{\mathrm{F}}^2=\sum_{i=1}^{N}(f_i-\tilde{f}_i)^2, \quad (23)$$

$$\Delta_{\mathrm{iso}}^2=(f_{\mathrm{iso}}-\tilde{f}_{\mathrm{iso}})^2, \quad (24)$$

3.3. Validation with a Synthetic Phantom and In Vivo Data

To validate the methods, experiments are performed with a synthetic phantom and *in vivo* data in both healthy subjects and children with autism. The synthetic phantom consists of a 16 × 16 × 16 multi-fascicle model containing an isotropic compartment and 1, 2 or 3 tensors with various properties (Fig. 7, see caption for parameter details). The non attenuated signal S_0 is set to 400. This phantom enables the generation of synthetic DWI at different noise levels with different gradient directions and b-values.

In vivo DWI were acquired on a Siemens 3T Trio scanner with a 32 channel head coil using the CUSP-45 gradient sequence (Scherrer and Warfield, 2012). This sequence includes 30 diffusion-encoding gradients on a shell at b = 1000 s/mm^2 and 15 extra gradients in the enclosing cube of constant echo time with b-values up to 3000 s/mm^2. Eddy current distortion was minimized using a twice-refocused spin echo sequence (Reese et al., 2003). Other acquisition parameters were set to FOV= 220mm, matrix= 128×128, number of slices=68, resolution = 1.7×1.7×2mm^3. Data acquisition was conducted using a protocol approved by the Institutional Review Board (IRB). Images were acquired for 31 healthy controls and 10 children diagnosed with a syndromic form of autism spectrum disorder (ASD) related to Tuberous Sclerosis Complex. No significant difference in age and gender were observed between the groups. A total of 13 healthy controls were used to construct the population-informed prior and the remaining 18 healthy controls as well as the children with ASD were used to test the performance of the estimation method. The computational time for the estimation of the DCI model with the prior on an 8 Core 3Ghz Intel Xeon machine running on Linux was approximately 100min, which is mostly driven by the registration of the prior to the subject (approximately 30min) and the estimation with the aligned prior (approximately 70min).

4. Results

In this section, we validate the proposed method to estimate multi-fascicle models from single non-zero b-value data. We first investigate the validity of the major assumption of our model. We then conduct a set of experiments on synthetic single-voxel data to assess the bias of the population-informed prior. We then carry out a set of experiments on the synthetic phantom where the true model is known by construction. We subsequently conduct experiments on *in vivo* data of both healthy controls and children with autism. We show that

the proposed approach performs significantly better than estimations without prior or with a spatial prior only. Finally, we show that the use of a population-informed prior enables reliable detections of group differences in the white matter microstructure.

4.1. Testing the independence between fractions and tensors

Our model assumes that the fractions **f** and the tensors **D** are independent variables at every location. The intuition behind this assumption is that the volumetric fractions and the properties of the tensors represent two independent properties of the microstructure (Taquet et al., 2014b). To test the validity of this hypothesis, we sampled the aligned images of all healthy controls at 520 locations (on a regular grid) and, at each location, we recorded the values of fractions and eigenvalues of all tensors for all subjects. For every location, we then tested the independence between a fraction and an eigenvalue by computing the distance correlation between the two (Székely et al., 2007). The process was repeated for every possible pair (fraction, eigenvalue). Ideally, the distance correlation would be null for independent variables but non-null values occur due to chance. To test whether chance could account for the distance correlations observed in our data, we estimated, at every location, the null distribution of the distance correlation by randomly permuting 1000 times the identification of the subjects and computing the distance correlation for each permuted dataset. This process enables estimating the p-value for the null hypothesis that the two variables are independent by permutation testing.

For every pair of variables, a minimum of 80% and an average of 87% of the locations had a distance correlation whose value could be explained by chance ($p > 0.05$) and hence whose independence could not be rejected. This result implies that, for the vast majority of voxels, independence between fractions and fascicle properties cannot be rejected. The remaining 13% of locations can be explained by two phenomena. First, as with any statistical test, the threshold at $p < 0.05$ means that the hypothesis will be rejected at 5% of locations due to chance. Second, we hypothesize that the remaining 8% of locations at which independence is statistically rejected present with registration or estimation errors since such errors affect all properties of the multi-tensor model simultaneously so that dependence between variables may appear. This small rate of locations with rejected independence does not warrant incorporating dependence in the generative model.

4.2. Bias Assessment and Detection of Synthetic Group Differences

In this section, we investigate the impact of the population-informed prior on the bias of estimations and on the detection of group differences. Thirteen different configurations of group differences were simulated by altering the microstructure properties of a two-fascicle control template according to the differences specified in Table 1 (illustrated in Fig. 6(a)). For each configuration, 1000 control templates and 1000 patient templates were generated. Control templates were generated by randomly drawing the principal eigenvectors uniformly on a sphere, the relative direction of the second eigenvector uniformly on a circle, the axial diffusivities from a normal distribution centered at 3.5×10^{-3} mm^2/s with a standard deviation of 0.5×10^{-3} mm^2/s and the radial diffusivities from a uniform distribution between 0.3×10^{-3} mm^2/s and $\lambda_1 - 0.9 \times 10^{-3}$ mm^2/s where λ_1 is the randomly drawn axial diffusivity. For each of the 13000 pairs of template, 20 control subjects were simulated to

build the prior, 20 other control subjects were generated to be compared to patients and 20 patients were simulated. Intra-group variability was simulated by adding independent Gaussian noise to the eigenvalues of the template (standard deviation: 0.1×10^{-3} mm^2/s), adding independent Gaussian noise to the components of the eigenvectors (standard deviation: 0.05) and re-normalizing them, and adding Gaussian noise to the fraction (standard deviation: 0.05).

Those simulated subjects correspond to the ground truth. Estimations of these models at a single b-value (we let b = 1000 s/mm^2) were simulated first by randomly selecting a model from the corresponding equivalence manifold, while enforcing the constraint of positive eigenvalues (estimation without prior) and then by optimizing the value of the prior (estimation with prior). The estimation errors are depicted in Figure 5. Since the equivalence manifold induces the same errors to all three eigenvalues of a same tensor (regardless of the use of a prior), these are represented as λ_1 (for tensor 1) and λ_2 (for tensor 2). As expected from theoretical consideration in Section 2.5, estimations in control subjects (configuration ①) are unbiased with the prior and biased when the prior is not used. When group differences are present, the bias of the estimations with the prior is non-zero but remains smaller than the bias resulting from estimations without prior (for all properties and all configurations of group differences). When the prior is not used, the mean bias observed across all 13 configurations of group differences for the estimation of fractions is 0.22 and for the estimation of eigenvalues is 6.3×10^{-4} mm^2/s. In contrast, the mean bias with our approach is 0.016 for fractions and 6.1×10^{-5} mm^2/s for eigenvalues. The inclusion of the prior thus reduces the estimation bias by an order of magnitude for all properties.

Using the estimated models in the 13000 simulated populations, group comparisons were performed with two-sample t-tests. Significance thresholds at 10^{-4} were then applied to the p-values to detect group differences (the choice of the threshold is motivated by the actual thresholds typically applied to each voxel of an image when FDR correction is applied). As expected, no additional false positives are observed when the prior is used in the estimation under the null hypothesis (configuration ①). The vast majority of true group differences are detected when the prior is used whereas most group differences vanish by Type II error when no prior is included in the estimation (Fig. 6(b)). For estimation with the prior, increased number of properties affected by group differences increase estimation bias, thereby making inference of group differences more conservative (configurations ⑪ – ⑬) as illustrated in Figure 6(c). This results in slightly decreased detection rates (*e.g.*, differences in $\lambda_{2,3}$ under configuration ⑬ are detected 75% of the times). In these configurations, however, the number of properties affected by group differences increases the likelihood of one property being detected as abnormal and therefore our ability to consider that the microstructure in the voxel is abnormal. This likelihood is estimated by counting the number of times any property from the voxel is detected as different. With the prior, this rate is equal to 100% for configurations ③ to ⑬, to 93% for configuration ② and to 0% for configuration ①.

4.3. Synthetic Phantom Experiment

In this experiment, we evaluated the impact of noise, group differences, and registration error on the models estimated with and without the population-informed prior. We constructed the prior from 20 sets of 95 DWI (5 at b =0 s/mm^2, 30 at b =1000 s/mm^2, 30 at b =2000 s/mm^2, and 30 at b =3000 s/mm^2) simulated by the synthetic phantom under a Rician noise with squared scale σ^2 = 80. This prior was then used to estimate multi-fascicle models from 30 DWI simulated at a single non-zero b-value (b =1000 s/mm^2) and 5 DWI at b =0 s/mm^2. The estimation was performed in three different scenarios: (i) under the influence of different levels of Rician noise, (ii) under the simulation of group differences and (iii) under the simulation of registration errors. Each estimation was repeated 20 times, each time with 35 DWI simulated to test the performance of the estimation. In these experiments, no spatial prior was used (*i.e.*, $\sigma_s \to \infty$ in (17)). The parameter σ^2_{noise} in Experiments (ii) and (iii) was set to 80. Summary statistics for the comparison metrics are reported in Table 2.

Influence of noise—The influence of noise level was assessed by simulating Rician noise of squared scale varying from 40 to 120 by steps of 20. As depicted in Fig. 8, the population-informed prior significantly improves the accuracy of the estimation for the six comparison metrics (one-tail paired t-test: $p < 10^{-10}$). The difference is less pronounced for Δ_{Dir}, since directions are not affected by the ill-posedness problem, as predicted by Equation (2).

Impact of group differences—Estimating multi-fascicle models from patients' data implies that the prior (built in healthy controls) may be offset in terms of fascicle properties. To simulate the impact of these group differences, we offset the FA of the tensors in the phantom by −10% to +10%, compared to the population-informer prior, with steps of 2.5%. The datasets of DWI used for testing were simulated from these offset phantoms while the population-informed prior was left unchanged. The results, depicted in Fig. 8, show that the population-informed prior significantly improves the accuracy of the estimation for the six metrics (one-tail paired t-test: $p < 10^{-5}$ for Δ_{Dir} and $p < 10^{-10}$ for the other metrics). As for the previous experiment, the improvement in Δ_{Dir} is less pronounced than improvements in other metrics.

Impact of registration errors—The prior needs to be registered to the subject whose multi-fascicle model is to be estimated. To simulate the impact of registration errors on the estimated multi-fascicle model, we applied random deformations to the prior. The three components of the deformation at each voxel were sampled from an uniform distribution between 0 and a = 0, 0.5, 1, 1.5, and 2 voxels which spans beyond the typical accuracy obtained with the multi-tensor registration used for the in-vivo experiment (Taquet et al., 2012b, 2014a). The results, depicted in Fig. 8, show that the population-informed prior significantly improves the accuracy of the estimation for the five metrics other than Δ_{Dir} (one-tail paired t-test: $p < 10^{-5}$). For Δ_{Dir}, a small decrease in accuracy is observed because estimates without prior are not affected by registration errors. The angular magnitude of this decrease, however, is 0.048°, which is negligible in practice.

These experiments demonstrate that the population-informed prior significantly improves the estimation of multi-fascicle models, even when registration errors and group differences occur. In Section 4.4, we will see that the same is true for *in vivo* data.

4.4. In Vivo Data Experiment

For *in vivo* data, there is no absolute reference standard. However, the CUSP acquisition sequence allows us to estimate the multi-fascicle model from the *full set* of DWI (5 DWI at b = 0 s/mm^2, 30 DWI on a shell at b = 1000 s/mm^2 and 15 DWI on the enclosing cube with b-values up to 3000 s/mm^2) and from the *subset* of those DWI at a single non-zero b-value (5 DWI at b = 0 s/mm^2, 30 DWI on a shell at b = 1000 s/mm^2). The multi-fascicle models estimated with the *full sets* of DWI are not affected by this ill-posedness problem. We consider them as reference standards for the experiments and we compare estimations from the *restricted sets* to it. In this section, we optimize the parameters for the estimation and we compare our approach with the results obtained with (i) no prior, (ii) a spatial prior only, (iii) a fixed response function as proposed in the literature (Tuch et al., 2002) and (iv) by simply using the population mode without leveraging data (this dummy estimator is used to demonstrate how the proposed approach harnesses both prior information and likelihood from data). A summary of the performance of all methods is depicted at the end of this section, in Fig. 10.

Optimization of the weight ratio—Two parameters need to be fixed for the estimation of the maximum a posteriori in (17): the weight associated with the likelihood, $1/2\sigma^2_{\text{noise}}$, and the weight associated with the spatial prior, $1/2\sigma^2_{\text{s}}$. When no population-informed prior is used, only the ratio of these weights, $\sigma^2_{\text{noise}}/\sigma^2_{\text{s}}$, needs to be fixed. The ground truths (based on the full sets of DWI) were estimated using a ratio $\sigma^2_{\text{noise}}/\sigma^2_{\text{s}}=0.4$ as we observed that it yields smooth multi-fascicle models while preserving important structures of the white matter.

We first estimated multi-fascicle models from single b-value data using a spatial prior only, removing the first term of (17). Because of the ill-posedness problem, the optimal weight ratio ($\sigma^2_{\text{noise}}/\sigma^2_{s}$) may differ from that used to build the ground truth. We therefore estimated multi-fascicle models with weight ratios of $\sigma^2_{\text{noise}}/\sigma^2_{\text{s}}=0.2$, 0.4, 0.6, 0.8, 1.0, 1.2 and 1.4. The evolutions of the six similarity metrics with the weight ratio are depicted in Fig. 9. These results show that the best overall results are obtained for a weight ratio of 0.6, which is optimal for Δ_{MD} and Δ_{iso} and close to optimal for Δ_{Dir}, and Fro. For a weight ratios above 0.6, the spatial prior is overweighted resulting in too large a spatial regularization.

Comparison between the population-informed prior and a spatial prior—As explained in Section 2.2, spatial priors alone do not resolve the ill-posedness problem in the estimation of multi-fascicle models from single b-value data. In this experiment, we explored whether the population-informed prior increases the accuracy of the estimation. We kept $\sigma^2_{\text{noise}}/\sigma^2_{\text{s}}$ to its optimal value of 0.6 and we estimated multi-fascicle models for $\sigma^2_{\text{noise}}=0$, 3.7, 4.5, 5.75, 7.83, 12, 24.5, 32.83, 49.5 and 99.5. These values stem from our implementation which consists in fixing the trade-off between the population-informed prior

(assigning it a weight w) and the likelihood (assigning it a weight $(1 - w)$) with linearly increasing w (0, 0.88, 0.9, 0.92, 0.94, 0.96, 0.98, 0.985, 0.99, 0.995). When $\sigma^2_{\text{noise}}=0$, the population-informed prior is mute and the estimation corresponds to the estimation with the spatial prior only (dashed lines in Fig. 9).

Results in Fig. 9 demonstrate that the population-informed prior improves the accuracy of the estimation for all metrics except for Δ_{Dir}. Optimal performances occur for $\sigma^2_{\text{noise}}=49.5$ (corresponding to a trade-off $w = 0.99$). Importantly, the performances improve monotonically and smoothly with σ_{noise} so that slightly suboptimal estimates of the noise level will not dramatically affect the results. For $\sigma^2_{\text{noise}}=49.5$, the improvement in the group of healthy controls ranges from 7.8% for Fro to 28.8% for Δ_{iso}. One-tail paired t-tests indicate that these improvements are significant ($p < 0.005$ for all five metrics). Importantly, similar results were obtained for the estimation of multi-fascicle models in patients with autism (Fig. 10). Improvements were observed for all five non-directional metrics with improvements ranging from 3.2% for Δ_{FA} to 18.2% for Δ_{iso}. One-tail paired t-tests indicate that these improvements are significant ($p < 0.005$ for all metrics except $p = 0.05$ for Δ_{F}).

To better appraise the improvement brought by the population-informed prior, Fig. 11 depicts the spatial distribution of the improvement in estimation error (as measured by Δ_{F}). While the improvement is widespread, the patterns show that it is more important in areas with crossing fascicles than in areas with a single fascicle (in those areas, estimating the model with single b-value data is not ill-posed). Importantly, even in areas with low diffusion contrast, the estimation remains better with a population-informed prior, owing to the reliable registration accuracies in those areas where the registration is driven by reproducible patterns of isotropic diffusion (Taquet et al., 2014a).

The slight increase in Δ_{Dir} observed with the population-informed prior corresponds to an angular difference of 0.4°, which is negligible in most practical applications. The equivalent accuracies in Δ_{Dir} is explained by Equation (2) which implies that directions are not affected by the ill-posedness problem. Figure 13(d–e) illustrate the difference in directions between estimations with and without the population-informed prior.

Consistently for both synthetic data and *in vivo* data in healthy controls and patients with autism, the isotropic fraction is the metric most beneficial of the introduction of the population-informed prior. The reason for this dramatic improvement can be understood from Equation (4). Equation (4) implies that models with large γ_i have the i-th fascicle close to an isotropic tensor (due to the offset of all eigenvalues by $\log(\gamma_i)/b$) and a signal fraction ($\gamma_i f_i$). Due to noise, those fascicles may be associated with an isotropic compartment, thereby increasing the isotropic fraction. Conversely, a multi-fascicle model that has a large isotropic fraction may be misinterpreted as one with a small isotropic fraction and a tensor with a low fractional anisotropy. The choice between the two models is arbitrary. The spatial prior only imposes that a consistent choice be made between adjacent voxels, and will lead to globally biased isotropic fractions if the wrong model is picked. This phenomenon is observed in Fig. 12(a), where, in the first row, a negative bias is observed throughout the map of isotropic fraction, and, in the second row, a positive bias is observed. The same

phenomenon is observed, to a lesser extent, for maps of the maximum FA from all tensors (Fig. 12(b)) and of the maximum radial diffusivity from all tensors (Fig. 12(c)). These global biases may mislead population studies as will be demonstrated in the next section.

Comparison with a fixed response function—One way to remove the ill-posedness problem is to fix *a priori* the tensor eigenvalues and to focus on optimizing the directions (Anderson, 2005; Tuch et al., 2002). While such a strategy does not aim to accurately estimate eigenvalues, it may lead to lower errors in orientation by further constraining the optimization problem. We assessed this strategy by estimating, for each subject individually, a typical response function, *i.e.*, a tensor that represents the typical fascicle within the individual's brain. This response was estimated in a similar fashion as (Tournier et al., 2007). First, we selected the 300 voxels with the highest FA. We assumed that a single fascicle was present in these voxels. A tensor was then estimated in each of these voxels and we computed its principal eigenvector. This eigenvector was used to re-orient the fascicle with the *z*-axis. The rotation was then applied to all DWI voxels and the corresponding gradients were reoriented accordingly. Finally, the typical response tensor was estimated from all 300 voxels of all the reoriented DWI. This response function was plugged into the maximum a posteriori estimate as a hard constraint on the tensor eigenvalues. Results in Fig. 10 demonstrate that such a fixed response strategy does not reduce the error made on orientation while the estimated FA is on average significantly worse than that estimated with our population-informed prior. These results demonstrate that the assumption that all fascicles within the brain can be represented by the same response function is not respected in *in vivo* experiments.

Assessment of the shrinkage towards the population mean—Bayesian parameter estimation leads to models that are closer to the mean of the population than they would be without a prior. One could be concerned that this *shrinkage*, if too severe, would jeopardize the usefulness of the models for population studies. To assess the extent of the shrinkage induced by the population-informed prior, we compared the models estimated with our method with the mode of the population-informed prior. If the shrinkage is large, the models would be very close to this mode. Results in Fig. 10 show that the estimation accuracies of the population mode are significantly worse than those obtained with the proposed population-informed prior. The departure of our results from those obtained with the population mode shows that the Bayesian inference operates in a regime that trades off information from both the prior and the likelihood. Remarkably, our estimation performs as well as the spatial prior estimate in terms of Δ_{Dir} despite the poor results obtained by the population mode. This suggests that our maximum a posteriori expression correctly integrates information from the likelihood, the spatial prior and the population-informed prior. In particular, if the population-informed prior indicates a large dispersion in fascicle orientations in one voxel, the corresponding variance $\tilde{\sigma_i}$ will be large and more weight will be put on the spatial prior and the likelihood.

4.5. Application to Population Studies

Multi-fascicle models enable population studies of various properties of the brain microstructure. In this section, we propose two population studies of the brain

microstructure comparing patients with autism spectrum disorders to typically developing controls. The increased accuracy brought by the population-informed prior translates —as we will see— into a more reliable inference of group differences. All results presented in this section were based on $\sigma^2_{\text{noise}}/\sigma^2_s = 0.6$ and $\sigma^2_{\text{noise}} = 49.5$.

Isotropic diffusion analysis in the white matter—Isotropic diffusion analysis (Taquet et al., 2013) allows whole-brain inspection of differences in isotropic fraction f_{iso}. Excessive f_{iso} relates to the presence of neuroinflammation and edema among others (Pasternak et al., 2012b; Wang et al., 2011). We performed isotropic diffusion analysis to compare the group of patients with ASD to the group of typically developing controls. The statistics of interest was the t-statistics transformed with threshold-free cluster enhancement (Smith and Nichols, 2009). This statistics is more sensitive to group differences occurring in clusters of neighboring voxels, without relying on the (typically arbitrary) choice of a threshold. The *p*-values were corrected for family-wise error rate using exact tests with 5000 permutations.

Widespread regions of significantly higher isotropic fraction were observed in patients with autism, as compared to controls (see the maps of corrected *p*-values thresholded at $p < 0.05$ depicted in Fig. 14). These regions correspond to the arcuate fasciculi, the corpus callosum and cortico-spinal tracts. These widespread differences are consistent with finding of widespread white matter abnormalities (Ridler et al., 2001) and widespread disconnection (Peters et al., 2013b) in patients with tuberous sclerosis complex as well as recent findings of impaired language pathways (Lewis et al., 2013) and loss of corpus callosum integrity (Peters et al., 2012) in children with syndromic autism. However, those recent results were based on single-tensor DTI analysis which conflates differences in fascicle properties (such as the FA) and differences in isotropic fraction. DTI studies therefore cannot separate axon/myelin injury from increased cellularity associated with neuroinflammation (Peters et al., 2013a; Wang et al., 2011). Our findings of increased isotropic fraction in children with autism suggest that previous findings of impaired white matter in children with syndromic autism may be in part due to increased cellularity or edema that may point to a neuroinflammatory process in those regions, as suggested by post-mortem studies of autism (Vargas et al., 2005).

Remarkably, the differences found with the population-informed prior match the differences found with the ground truth models. By contrast, no significant differences were detected with the spatial prior, probably owing to the arbitrariness of the choice of a model from the manifold of equivalent models as explained in the previous section and depicted in Fig. 12. Changes in signal fractions across the manifold of equivalent models are compensated by changes in fascicle properties. Group difference in signal fraction may therefore be incorrectly interpreted as differences in fascicle properties. In the next experiment, we will see that this misleading effect occurs in our population when spatial priors are being used alone.

Fascicle-based spatial statistics of the dorsal language circuit—Fascicle-based spatial statistics (FBSS) was proposed in (Taquet et al., 2012b) to analyze white matter properties along individual fascicles in the presence of crossing fascicles. FBSS proceeds in

four steps. First, all multi-fascicle models are registered to a multi-fascicle atlas. Second, tractography is performed on the atlas and specific fascicles are selected. Third, properties of the fascicles are extracted from each subject by selecting at every point the tensor most aligned with the fascicle. Fourth, statistical analysis is performed on the fascicle properties by computing t-tests at every point of the fascicle and by performing permutation tests on the threshold-free cluster enhanced t-statistics, similar to (Smith et al., 2006), to control for multiple comparisons.

We performed FBSS for the FA of the dorsal language circuit (Fig. 15). This set of pathways involved in language is organized around three main fascicles (Lewis et al., 2013): the posterior, anterior and long fascicles. These fascicles are thought to connect Broca's area in the frontal lobe, Wernicke's area in the temporal lobe, and Geschwind's territory in the parietal lobe. FBSS was performed on these three fascicles individually. Along the posterior fascicle, a significantly lower FA was observed in two clusters ($p < 0.05$ after correction, Fig. 16). Similar findings were obtained by using the models estimated with the population-informed prior and the spatial prior alone. Along the anterior fascicle, no significant difference was observed (Fig. 16). Estimations with the population-informed prior only found a small cluster of false positive difference covering 1.4% of the fascicle. With the spatial prior, three large clusters of false positive differences were observed, covering 39.4% of the fascicle. This high prevalence of false discoveries with the spatial prior can be explained by its inability to detect the significant differences in isotropic fractions. These differences are incorrectly reflected on the FA of the fascicle. A similar effect is observed for the long fascicle. A cluster of significantly lower FA is observed using the ground truth models. This cluster was not detected by the population-informed prior and only partially detected (30% overlap) by the spatial prior. This partial overlap, however, comes with a high rate of false discoveries (38.9%).

5. Discussion and Conclusion

Multi-fascicle models cannot be estimated from conventional single-shell HARDI data alone because a manifold of different models are equally compatible with the measurable diffusion signals, making them indistinguishable. Estimation with the most common spatial priors arbitrarily selects a model from the manifold of equivalent models and thereby conflates differences in fascicle properties and differences in signal fractions. This conflation misleads conclusions of population studies in a similar way as single tensor DTI does. With single b-value data, these spatial priors alone therefore fail to harness the novel insight provided by multi-fascicle models.

We showed that, when data at two non-zero b-values are acquired, the ill-posedness encoded by Equation (3) disappears because the intersection of the manifolds of equivalent solutions at two b-values is unique. Importantly, there might be other sources of ill-posedness not accounted for by Equation (3) which two non-zero b-values would not solve. In the typical case of a single isotropic tensors and multiple anisotropic ten-sors (such as those exploited in this paper), no such other source could be found in the literature. In the particular case of multiple isotropic tensors, however, the model of Equation (1) is no more dependent of the gradient orientation $\boldsymbol{g}$ (so that $S(\boldsymbol{g}, b) \equiv S(b)$) and estimating its parameters boils down to

estimating a weighted sum of exponentials. In that degenerate case, it is known since (Prony, 1795) that two non-zero b-values is not sufficient to estimate the parameters of the exponentials. Andersson *et al*. demonstrated that a sufficient condition for the fitting of a sum of exponentials is that the rank of the Hankel matrix generated by the observations $S(b)$ is equal to the number N of components in the model (Andersson et al., 2014). The reason why two non-zero b-values are not sufficient in this case can be understood from the fact that the expression for two models $(\boldsymbol{D},\boldsymbol{f})$ and $(\boldsymbol{D}^*,\boldsymbol{f}^*)$ to be equivalent:

$$\sum_{i=1}^{N} f_i e^{-b\boldsymbol{g}^T \boldsymbol{D}_i \boldsymbol{g}} = \sum_{i=1}^{N} f_i^* e^{-b\boldsymbol{g}^T \boldsymbol{D}_i^* \boldsymbol{g}}, \forall \boldsymbol{g} \in \mathscr{S}_2$$

becomes, for isotropic tensors,

$$\sum_{i=1}^{N} f_i e^{-bd_i} = \sum_{i=1}^{N} f_i^* e^{-bd_i^*}$$

and only imposes a single non-linear constraint while, in the general case, the constraint must hold true for any gradient orientation $\boldsymbol{g}$.

We showed that the introduction of a population-informed prior generates more accurate multi-fascicle models by removing the ill-posedness problem from the estimation. This is achieved by learning a prior distribution over model parameters from a population of subjects scanned at multiple b-values. This prior is estimated in the space of an atlas and is then registered to the subject. Accurate registration is therefore required to achieve accurate estimation. In this paper, we leveraged the full multi-tensor information to drive the registration which provides the best accuracy.

Throughout the paper, we have targeted the estimation of a true multi-fascicle model. Such a true model must be understood as the optimal representation of the brain microstructure at the probed diffusion length scale for the data at hand. It may be that other multi-fascicle models (with other values for the parameters) may better represent the microstructure for other length scales, as it is the case for single tensor models (Yoshiura et al., 2001). Caution must therefore be taken in achieving the additional b-values required to remove the ill-posedness in the estimation when learning the prior: the additional b-value must be achieved by manipulating the gradient strengths without altering the diffusion times or the consequences of changing the diffusion times, in terms of length scales probed and the microstructure of interest, must be negligible. In this paper, we have opted for the former through a CUSP acquisition protocol (Scherrer and Warfield, 2012).

Achieving reliable estimations of Diffusion Compartment Imaging models from single-shell HARDI data leads, in turn, to more reliable inferences in population studies, distinguishing differences in isotropic fractions from differences in fascicle properties. The population-informed prior therefore enables novel investigations of properties extracted from DCI

models with single-shell diffusion data. This method thus opens new opportunities for population studies with the large number of available clinical diffusion images.

Acknowledgments

Maxime Taquet thanks the WBI.WORLD for financial support. This work was supported in part by NIH grants 1U01NS082320, R01 EB008015, R01 NS079788, R03 EB008680, R01 LM010033, R01 EB013248, P30 HD018655, a research grant from Boston Children's Hospital Translational Research Program, R42 MH086984 and UL1 TR000170.

References

Aboitiz F, Scheibel AB, Fisher RS, Zaidel E. Fiber composition of the human corpus callosum. Brain research. 1992; 598:143–153. [PubMed: 1486477]

Alexander DC, Barker GJ, Arridge SR. Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data. Magnetic Resonance in Medicine. 2002; 48:331–340. [PubMed: 12210942]

Alexander, DC.; Gee, JC.; Bajcsy, R. Medical Image Computing and Computer-Assisted Intervention–MICCAI99. Springer; 1999. Strategies for data reorientation during non-rigid warps of diffusion tensor images; p. 463-472.

Alexander DC, Pierpaoli C, Basser PJ, Gee JC. Spatial transformations of diffusion tensor magnetic resonance images. Medical Imaging, IEEE Transactions on. 2001; 20:1131–1139.

Anderson AW. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. Magnetic Resonance in Medicine. 2005; 54:1194–1206. [PubMed: 16161109]

Andersson F, Carlsson M, Tourneret JY, Wendt H. A new frequency estimation method for equally and unequally spaced data. Signal Processing, IEEE Transactions on. 2014; 62:5761–5774.

Arsigny V, Fillard P, Pennec X, Ayache N. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic resonance in medicine. 2006; 56:411–421. [PubMed: 16788917]

Assaf Y, Basser PJ. Composite hindered and restricted model of diffusion (charmed) mr imaging of the human brain. Neuroimage. 2005; 27:48–58. [PubMed: 15979342]

Assaf Y, Freidlin RZ, Rohde GK, Basser PJ. New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. Magnetic Resonance in Medicine. 2004; 52:965–978. [PubMed: 15508168]

Daducci A, Canales-Rodríguez EJ, Zhang H, Dyrby TB, Alexander DC, Thiran JP. Accelerated microstructure imaging via convex optimization (amico) from diffusion mri data. NeuroImage. 2015a; 105:32–44. [PubMed: 25462697]

Daducci A, Dal Palu A, Lemkaddem A, Thiran JP. Commit: convex optimization modeling for microstructure informed tractography. Medical Imaging, IEEE Transactions on. 2015b; 34:246–257.

Fjell AM, Walhovd KB, Brown TT, Kuperman JM, Chung Y, Hagler DJ, Venkatraman V, Roddey JC, Erhart M, McCabe C, et al. Multimodal imaging of the self-regulating developing brain. Proceedings of the National Academy of Sciences. 2012; 109:19620–19625.

Gao W, Lin W, Chen Y, Gerig G, Smith J, Jewells V, Gilmore J. Temporal and spatial development of axonal maturation and myelination of white matter in the developing brain. American journal of neuroradiology. 2009; 30:290–296. [PubMed: 19001533]

Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. Magnetic Resonance in Medicine. 1995; 34:910–914. [PubMed: 8598820]

Heads T, Pollock M, Robertson A, Sutherland W, Allpress S. Sensory nerve pathology in amyotrophic lateral sclerosis. Acta neuropathologica. 1991; 82:316–320. [PubMed: 1662002]

Jeurissen B, Leemans A, Tournier JD, Jones DK, Sijbers J. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. Human Brain Mapping. 2012

Jones DK. The effect of gradient sampling schemes on measures derived from diffusion tensor mri: a monte carlo study. Magnetic Resonance in Medicine. 2004; 51:807–815. [PubMed: 15065255]

Jordán, K. Calculus of finite differences. AMS Bookstore; 1965.

Kreher B, Schneider J, Mader I, Martin E, Hennig J, Il'yasov K. Multitensor approach for analysis and tracking of complex fiber configurations. Magnetic resonance in medicine. 2005; 54:1216–1225. [PubMed: 16200554]

Lamantia AS, Rakic P. Cytological and quantitative characteristics of four cerebral commissures in the rhesus monkey. Journal of Comparative Neurology. 1990; 291:520–537. [PubMed: 2329189]

Lewis WW, Sahin M, Scherrer B, Peters JM, Suarez RO, Vogel-Farley VK, Jeste SS, Gregas MC, Prabhu SP, Nelson CA, Warfield SK. Impaired language pathways in tuberous sclerosis complex patients with autism spectrum disorders. Cerebral Cortex. 2013; 23:1526–1532. [PubMed: 22661408]

Malcolm JG, Michailovich O, Bouix S, Westin CF, Shenton ME, Rathi Y. A filtered approach to neural tractography using the watson directional function. Medical Image Analysis. 2010; 14:58. [PubMed: 19914856]

Mishra V, Guo X, Delgado MR, Huang H. Toward tract-specific fractional anisotropy (tsfa) at crossing-fiber regions with clinical diffusion mri. Magnetic Resonance in Medicine. 2014

Panagiotaki E, Schneider T, Siow B, Hall MG, Lythgoe MF, Alexander DC. Compartment models of the diffusion mr signal in brain white matter: a taxonomy and comparison. Neuroimage. 2012; 59:2241–2254. [PubMed: 22001791]

Pasternak O, Shenton M, Westin CF. Estimation of extracellular volume from regularized multi-shell diffusion MRI. Medical Image Computing and Computer-Assisted Intervention–MICCAI. 2012a; 2012:305–312.

Pasternak O, Sochen N, Gur Y, Intrator N, Assaf Y. Free water elimination and mapping from diffusion MRI. Magnetic Resonance in Medicine. 2009; 62:717–730. [PubMed: 19623619]

Pasternak O, Westin CF, Bouix S, Seidman L, Goldstein J, Woo T, Petryshen T, Mesholam-Gately R, McCarley R, Kikinis R, et al. Excessive extracellular volume reveals a neurode-generative pattern in schizophrenia onset. The Journal of Neuroscience. 2012b; 32:17365–17372. [PubMed: 23197727]

Peters JM, Sahin M, Vogel-Farley VK, Jeste SS, Nelson CA III, Gregas MC, Prabhu SP, Scherrer B, Warfield SK. Loss of white matter microstructural integrity is associated with adverse neurological outcome in tuberous sclerosis complex. Academic radiology. 2012; 19:17–25. [PubMed: 22142677]

Peters JM, Taquet M, Prohl AK, Scherrer B, van Eeghen AM, Prabhu SP, Sahin M, Warfield SK. Diffusion tensor imaging and related techniques in tuberous sclerosis complex: review and future directions. Future Neurology. 2013a; 8:583–597. [PubMed: 24489482]

Peters JM, Taquet M, Vega C, Jeste SS, Sanchez Fernandez I, Tan J, Nelson CA, Sahin M, Warfield SK. Brain functional networks in syndromic and non-syndromic autism: a graph theoretical study of eeg connectivity. BMC medicine. 2013b; 11:54. [PubMed: 23445896]

Powell, MJ. The BOBYQA algorithm for bound constrained optimization without derivatives. Cambridge NA: University of Cambridge, UK; 2009. Report NA2009/06

Prony R. Essai experimental. J de lEcole Polytechnique. 1795

Ramirez-Manzanares A, Rivera M, Vemuri BC, Carney P, Mareci T. Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry. Medical Imaging, IEEE Transactions on. 2007; 26:1091–1102.

Reese TG, Heid O, Weisskoff RM, Wedeen VJ. Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. Magn Reson Med. 2003; 49:177–182. [PubMed: 12509835]

Ridler K, Bullmore E, De Vries P, Suckling J, Barker G, Meara S, Williams S, Bolton P. Widespread anatomical abnormalities of grey and white matter structure in tuberous sclerosis. Psychological medicine. 2001; 31:1437–1446. [PubMed: 11722158]

Ruiz-Alzola, J.; Westin, CF.; Warfield, SK.; Nabavi, A.; Kikinis, R. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000. Springer; 2000. Nonrigid registration of 3d scalar, vector and tensor medical data; p. 541-550.

Scherrer, B.; Schwartzman, A.; Taquet, M.; Prabhu, SP.; Sahin, M.; Akhondi-Asl, A.; Warfield, SK. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer; 2013a.

Characterizing the distribution of anisotropic micro-structural environments with diffusion-weighted imaging (diamond); p. 518-526.

Scherrer, B.; Taquet, M.; Warfield, SK. Reliable selection of the number of fascicles in diffusion images by estimation of the generalization error. Information Processing in Medical Imaging, 23rd biennial International Conference on; Springer; 2013b. p. 742-753.

Scherrer, B.; Warfield, SK. Why multiple b-values are required for multi-tensor models. evaluation with a constrained log-Euclidean model. Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on; IEEE; 2010. p. 1389-1392.

Scherrer B, Warfield SK. Parametric representation of multiple white matter fascicles from cube and sphere diffusion MRI. PloS one. 2012; 7:e48232. [PubMed: 23189128]

Schultz, T. Learning a reliable estimate of the number of fiber directions in diffusion MRI. Medical Image Computing and Computer-Assisted Intervention–MICCAI; 2012; Springer; 2012a. p. 493-500.

Schultz, T. New Developments in the Visualization and Processing of Tensor Fields. Springer; 2012b. Towards resolving fiber crossings with higher order tensor inpainting; p. 253-265.

Schultz, T.; Westin, CF.; Kindlmann, G. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010. Springer; 2010. Multi-diffusion-tensor fitting via spherical deconvolution: a unifying framework; p. 674-681.

Schwartzman A, Mascarenhas WF, Taylor JE. Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. The Annals of Statistics. 2008:2886–2919.

Smith S, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols T, Mackay C, Watkins K, Ciccarelli O, Cader M, Matthews P, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage. 2006; 31:1487–1505. [PubMed: 16624579]

Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage. 2009; 44:83–98. [PubMed: 18501637]

Székely GJ, Rizzo ML, Bakirov NK, et al. Measuring and testing dependence by correlation of distances. The Annals of Statistics. 2007; 35:2769–2794.

Taquet, M.; Scherrer, B.; Benjamin, C.; Prabhu, S.; Macq, B.; Warfield, SK. Interpolating multi-fiber models by Gaussian mixture simplification. Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on; IEEE; 2012a. p. 928-931.

Taquet, M.; Scherrer, B.; Boumal, N.; Macq, B.; Warfield, SK. Medical Image Computing and Computer-Assisted Intervention (MICCAI) Part I, LNCS 8149. Springer; 2013. Estimation of a multi-fascicle model from single b-value data with a population-informed prior; p. 695-702.

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. 2014a; 33:504–517.

Taquet, M.; Scherrer, B.; Commowick, O.; Peters, J.; Sahin, M.; Macq, B.; Warfield, SK. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012. Springer; Berlin/Heidelberg: 2012b. Registration and analysis of white matter group differences with a multi-fiber model; p. 313-320.

Taquet, M.; Scherrer, B.; Peters, JM.; Prabhu, SP.; Warfield, SK. A fully bayesian inference framework for population studies of the brain microstructure. Medical Image Computing and Computer-Assisted Intervention–MICCAI; 2014; Springer; 2014b. p. 25-32.

Tournier J, Calamante F, Connelly A, et al. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. NeuroImage. 2007; 35:1459–1472. [PubMed: 17379540]

Tuch D, Reese T, Wiegell M, Makris N, Belliveau J, Wedeen V. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine. 2002; 48:577–582. [PubMed: 12353272]

Vargas DL, Nascimbene C, Krishnan C, Zimmerman AW, Pardo CA. Neuroglial activation and neuroinflammation in the brain of patients with autism. Annals of neurology. 2005; 57:67–81. [PubMed: 15546155]

Vos SB, Jones DK, Jeurissen B, Viergever MA, Leemans A. The influence of complex white matter architecture on the mean diffusivity in diffusion tensor MRI of the human brain. NeuroImage. 2012; 59:2208–2216. [PubMed: 22005591]

Wang Y, Wang Q, Haldar JP, Yeh FC, Xie M, Sun P, Tu TW, Trinkaus K, Klein RS, Cross AH, et al. Quantification of increased cellularity during inflammatory demyelination. Brain. 2011; 134:3590–3601. [PubMed: 22171354]

Yablonskiy DA, Bretthorst GL, Ackerman JJ. Statistical model for diffusion attenuated mr signal. Magnetic resonance in medicine. 2003; 50:664–669. [PubMed: 14523949]

Yoshiura T, Wu O, Zaheer A, Reese TG, Sorensen AG. Highly diffusion-sensitized mri of brain: Dissociation of gray and white matter. Magnetic resonance in medicine. 2001; 45:734–740. [PubMed: 11323798]

Zhan W, Yang Y. How accurately can the diffusion profiles indicate multiple fiber orientations? a study on general fiber crossings in diffusion MRI. Journal of magnetic resonance (San Diego, Calif: 1997). 2006; 183:193.

Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. Noddi: Practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage. 2012

Zikopoulos B, Barbas H. Changes in prefrontal axons may disrupt the network in autism. The Journal of Neuroscience. 2010; 30:14595–14609. [PubMed: 21048117]

Appendix A. Unicity of the intersection of two manifolds

In this section, we demonstrate that the equations of two manifolds for two different b-values (b and b' with $b' > b > 0$) have one and only one intersection regardless of the values b and b' and regardless of the number N of tensors. The equation of the manifold is (5) :

$$\lambda_N(\lambda_1, \ldots, \lambda_{N-1}, b) = \lambda_N^* + \frac{1}{b}\log\left(\frac{1-\sum_{i=1}^{N-1} f_i^* e^{b(\lambda_i-\lambda_i^*)}}{f_N^*}\right).$$

For the sake of notations, we let $\Delta_i = \lambda_i - \lambda_i^*$ and $\tilde{\boldsymbol{\Delta}} = [\Delta_1, \ldots, \Delta_{N-1}]$, so that $\tilde{\boldsymbol{\Delta}} = 0$ corresponds to the true underlying model. For all values of b, $\lambda_N(\lambda_1^*, \ldots, \lambda_{N-1}^*, b) = \lambda_N^*$ and hence does not depend on b. This encodes the fact that all manifolds intersect at the true model. We now show that this is the only point that any two manifolds have in common. The principal idea to demonstrate the unicity of the intersection of the manifolds lies in the following proposition.

Proposition 1

The manifolds at two different b-values do not intersect at $\tilde{\boldsymbol{\Delta}}$ *if the partial derivative of* $\lambda_N(\tilde{\boldsymbol{\Delta}}, b)$ *with respect to b is strictly negative for all* $b > 0$.

Proof

This follows from the definition of strictly decreasing functions. If $f(x)$ is strictly decreasing, then $f(x_1) > f(x_2)$ if $x_1 < x_2$ and $f(x_1) < f(x_2)$ if $x_1 > x_2$. If $\frac{\partial \lambda_N(\tilde{\boldsymbol{\Delta}}, b)}{\partial b} < 0$ for a given $\tilde{\boldsymbol{\Delta}}$ and for all $b > 0$, then $\lambda_N(\tilde{\boldsymbol{\Delta}}, b)$ as a function of b is a strictly decreasing function and therefore cannot take on the same value for two different b-values.

From this proposition, it suffices to demonstrate that the derivative of $\lambda_N(\tilde{\boldsymbol{\Delta}}, b)$ with respect to b is strictly negative for any $\tilde{\boldsymbol{\Delta}} \neq 0$. This derivative is given by:

$$\frac{\partial \lambda_N(\tilde{\boldsymbol{\Delta}}, b)}{\partial b} = \frac{-1}{b^2}\log(w) - \frac{1}{wb^2 f_N^*}\sum_{i=1}^{N-1} f_i^* b\Delta_i e^{b\Delta_i}, \quad \text{(A.1)}$$

where

$$w = \frac{1 - \sum_{i=1}^{N-1} f_i^* e^{b\Delta_i}}{f_N^*} > 0. \quad \text{(A.2)}$$

For all positive w, it holds that

$$\frac{w-1}{w} \leq \log(w) \leq w-1.$$

Multiplying (A.1) by $-wb^2$ on both sides and combining with the first inequality yields:

$$-b^2 w \frac{\partial \lambda_N(\tilde{\boldsymbol{\Delta}}, b)}{\partial b} \geq w - 1 + \frac{1}{f_N^*}\sum_{i=1}^{N-1} f_i^* b\Delta_i e^{b\Delta_i}. \quad \text{(A.3)}$$

From (A.2) and since $\sum_{i=1}^{N} f_i^* = 1$, we have:

$$w - 1 = \frac{1}{f_N^*}\left(\sum_{i=1}^{N-1} f_i^* - \sum_{i=1}^{N-1} f_i^* e^{b\Delta_i}\right).$$

Substituting this expression into (A.3), it follows that

$$-b^2 w f_N^* \frac{\partial \lambda_N(\tilde{\boldsymbol{\Delta}}, b)}{\partial b} \geq \sum_{i=1}^{N-1} f_i^* (1 + e^{b\Delta_i}(b\Delta_i - 1)) \quad \text{(A.4)}$$

$$\triangleq \sum_{i=1}^{N-1} f_i^* h(b\Delta_i) \quad \text{(A.5)}$$

where $h(x) \triangleq 1 + e^x(x-1)$. Finally, we can prove that the right-hand side of this inequality is strictly greater than zero for $\tilde{\boldsymbol{\Delta}} \neq 0$ based on the following proposition.

Proposition 2

The function $h(x) = 1 + e^x(x-1)$ *is strictly positive for* $x \neq 0$ *and equal to zero for* $x = 0$.

Proof

We have $h(0) = 1 + e^0(-1) = 0$ which proves the second part of the proposition. The derivative of $h(x)$ is $h'(x) = xe^x$ which is strictly negative for $x < 0$ and strictly positive for x

> 0. The function $h(x)$ is thus strictly decreasing for $x < 0$ and strictly increasing for $x > 0$ and therefore $h(x) > h(0) = 0$ for all x, which proves the first part of the proposition.

Since the weighted sum of strictly positive functions with strictly positive weights is strictly positive, it follows from (A.4) that $-b^2 w f_N^* \frac{\partial \lambda_N(\tilde{\mathbf{\Delta}}, b)}{\partial b} > 0$ for $\tilde{\mathbf{\Delta}} \neq 0$ and therefore, for $\tilde{\mathbf{\Delta}} \neq 0$:

$$\frac{\partial \lambda_N(\tilde{\mathbf{\Delta}}, b)}{\partial b} < 0,$$

which, from the first proposition, implies that two manifolds at two different b-values do not intersect for $\tilde{\mathbf{\Delta}} \neq 0$.

Appendix B. Geometry of the manifolds around the true underlying model

In this section, we analyze the geometry of the manifolds at the first and second order in the vicinity of the true underlying model. At the first order, the manifold is characterized by its tangent hyperplane or, equivalently, by its normal vector. This normal vector is: $\boldsymbol{\eta} = \left(\frac{\partial \lambda_N}{\partial \lambda_1}, \ldots, \frac{\partial \lambda_N}{\partial \lambda_{N-1}}, -1 \right)$. Its k-th component is:

$$\eta_k = \frac{\partial \lambda_N}{\partial \lambda_k} = \frac{-f_k^* e^{b(\lambda_k - \lambda_k^*)}}{1 - \sum_{i=1}^{N-1} f_i^* e^{b(\lambda_i - \lambda_i^*)}}, k = 1, \ldots, N-1.$$

At the true underlying model, the normal vector is:

$$\eta_k \big|_{\lambda_i = \lambda_i^*, \forall i} = \frac{\partial \lambda_N}{\partial \lambda_k} \Big|_{\lambda_i = \lambda_i^*, \forall i} = \frac{-f_k^*}{f_N^*}. \quad \text{(B.1)}$$

This normal vector (hence the tangent hyperplane) does not depend on b at the point of interest. This finding explains why manifolds coincide locally and why the estimation of multi-fascicle models is sensitive to noise.

At the second-order approximation, the manifold is characterized by the Hessian matrix of $\lambda_N(\lambda_1, \ldots, \lambda_{N-1})$:

$$\boldsymbol{H}(b) \triangleq \boldsymbol{H}\big|_{\lambda_i = \lambda_i^*} = \frac{-b}{(f_N^*)^2} \left(\tilde{\boldsymbol{f}}^* \tilde{\boldsymbol{f}}^{*T} + f_N^* \operatorname{diag}(\tilde{\boldsymbol{f}}^*) \right),$$

where $\tilde{\boldsymbol{f}}^* = [f_1^*, \ldots, f_{N-1}^*]^T$. The difference between the Hessian matrices at the intersection point for two different b-values, b and $b' > b$, is positive definite since, for all $\boldsymbol{s} \neq 0$, we have

$$\boldsymbol{s}^T \left(\boldsymbol{H}(b) - \boldsymbol{H}(b') \right) \boldsymbol{s} = \frac{b' - b}{(f_N^*)^2} \left((\tilde{\boldsymbol{f}}^{*T} \boldsymbol{s})^2 + f_N^* \boldsymbol{s}^T \operatorname{diag}(\tilde{\mathbf{f}}^*) \boldsymbol{s} \right) > 0. \quad \text{(B.2)}$$

This equation indicates that manifolds with large b-values are strictly more curved than manifolds with small b-values, in all directions. The difference in curvature is larger if the difference in b-value is larger. This in turn translates into lower sensitivity of the intersection point (the targeted model) to small perturbations of the manifolds.

Highlights

- We analyze the geometry of diffusion compartment imaging (DCI) estimation problem.
- Estimating DCI models from single b-value data is an ill-posed problem.
- Spatial priors fail to regularize the estimation from single b-value data.
- We derive a population-informed prior that regularizes the estimation problem.
- Our approach enables reliable estimations of DCI models from single b-value data.

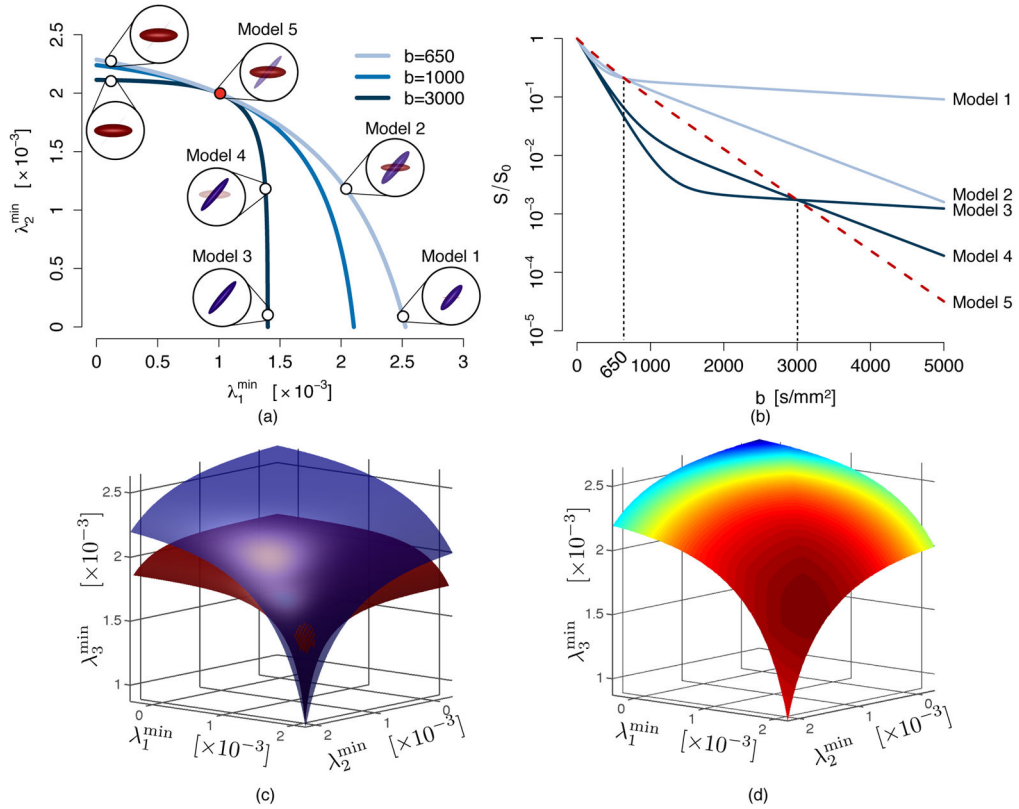


Figure 1.

Illustration of intersection of diffusion signal arising from different model parameters. (a) An infinite number of models with very different parameters present with the same values of the diffusion signal decay at a single specified b-value, forming a manifold of equivalent models at that b-value. There is only one unique model, (illustrated by the red dot), at which each of the manifolds formed at *different* b-values intersect tangentially. (b) Illustration of the diffusion MRI signal for five models selected from (a) in a gradient direction orthogonal to the red tensor. In each case, the signal depends on the b-value as given by Equation (1). Models 1, 2 and 5 generate the same diffusion signal at b = 650 s/mm^2 and thus cannot be distinguished from a single-shell acquisition of only b = 650 DWI (irrespective of the number of gradient directions acquired). Models 3, 4 and 5 generate the same diffusion signal at b = 3000 s/mm^2, thus cannot be distinguished from a single-shell acquisition of only b = 3000 DWI. Only Model 5 is compatible with observations at the two shells of b = 650 and b = 3000. In principle, an acquisition of two shells would allow determination that the parameters of Model 5 uniquely explains the observed imaging data. (c) For an N-fascicle model (here N=3), the manifolds of equivalent models are (N−1)-dimensional hypersurfaces which share the same tangent hyperplane at their intersection making the estimation problem very sensitive to noise. (d) The population-informed prior (here restricted to the manifold) assigns different probabilities to otherwise equivalent models.

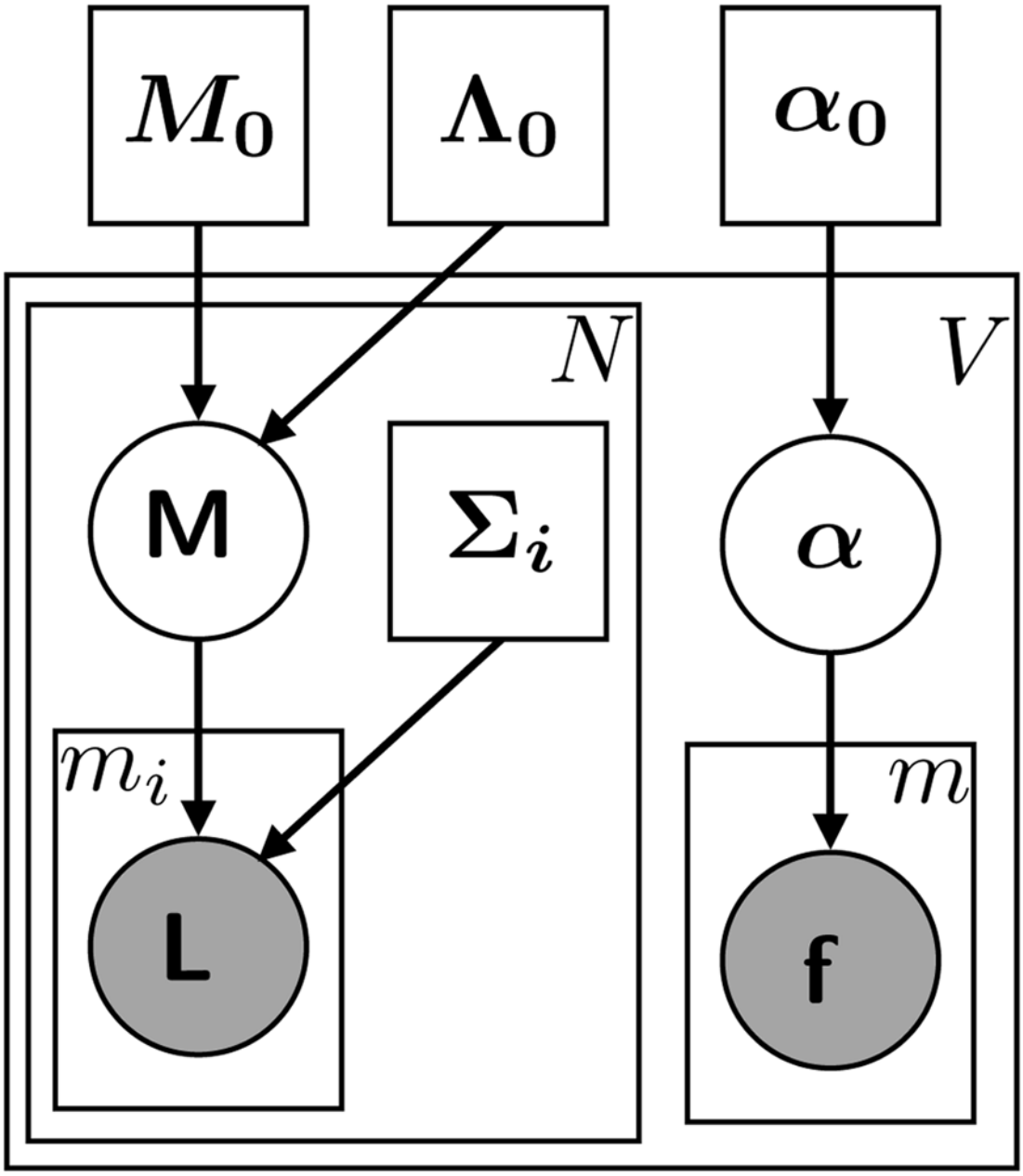


Figure 2.
Graphical model of the posterior predictive distribution model that serves as a prior for new observations. Circles indicate random variables, shaded circles indicate observed variables, and boxes indicate parameters. The graphical model is repeated for all *V* voxels.

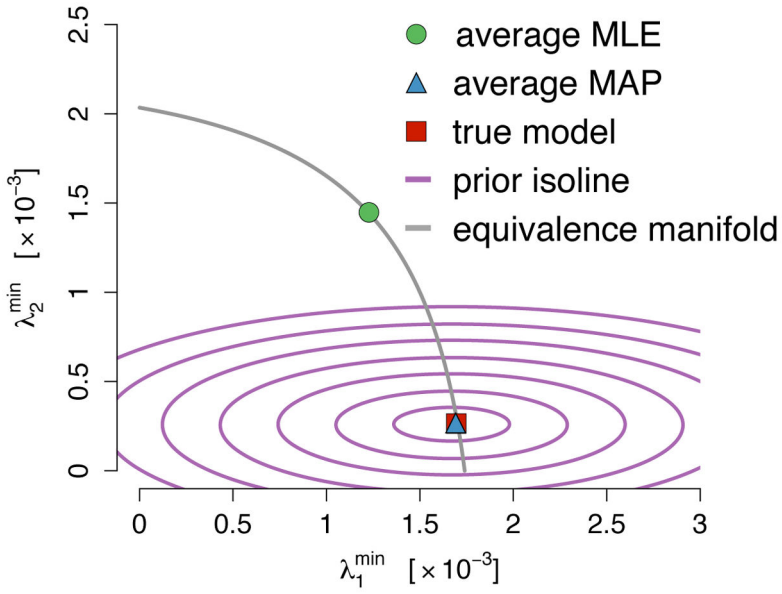


Figure 3.

Maximum a posteriori (MAP) estimates with the population-informed prior are unbiased insofar as the targeted population is well represented in the prior population. On the other hand, maximum likelihood estimates (MLE) arbitrarily pick a model from the manifold so that the average estimate lies at the center of the manifold leading to a high bias if the true population average lies away from the center of the manifold.

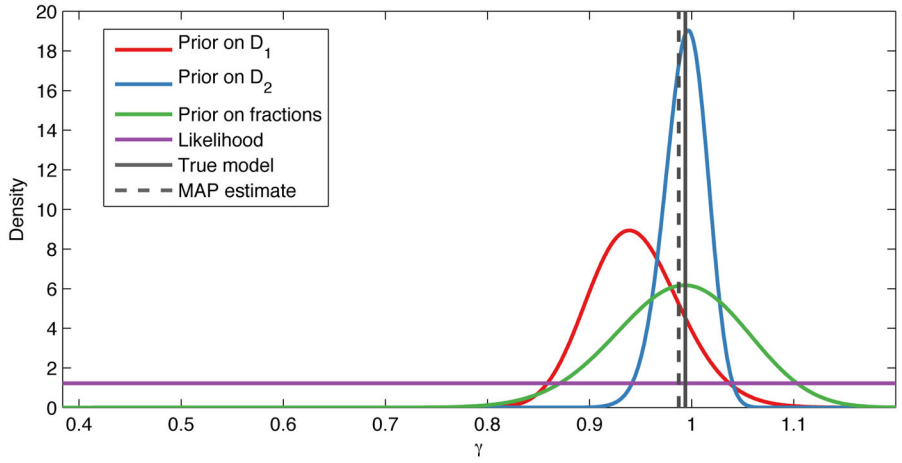


Figure 4.

Relative contribution of each factor of the prior in Equation (18) for the estimation of a two-tensor model in patients presenting with decreased axial diffusivity in the first fascicle. The prior on fractions and on $\mathbf{D}_2$ are centered on the true model since no group difference is present for those properties. The prior on $\mathbf{D}_1$ is offset because of the group difference in $\lambda_{1,1}$. This offset, however, is mitigated by the absence of differences in the other eigenvalues of $\mathbf{D}_1$. Multiplying all contributions to the prior thus leads to a MAP estimate that is close to the true model.

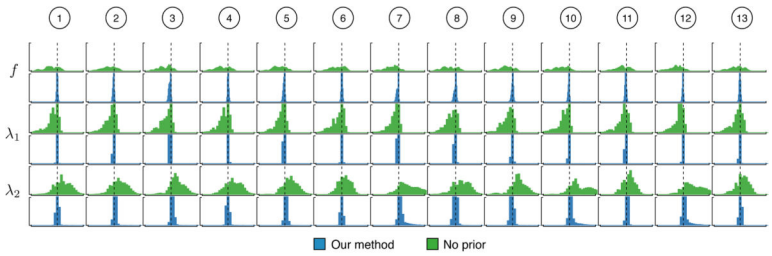


Figure 5.

Error distributions for the estimation with the population-informed prior and without prior for 20000 synthetic two-fascicle templates under 13 configurations of group differences summarized in Table 1 and illustrated in Fig. 6(a). Both estimation bias and variance are significantly smaller when the population-informed prior is used. Vertical dashed lines indicate zero errors. Axes span from −1 to +1 for f and from -3×10^{-3} mm^2/s to $+3 \times 10^{-3}$ mm^2/s for λ_1 and λ_2.

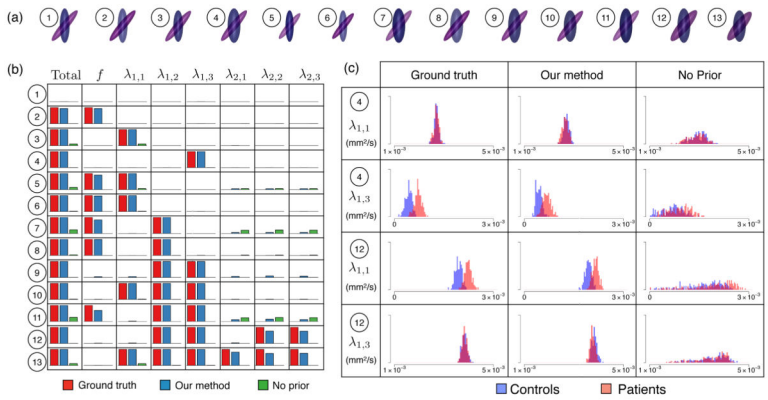


Figure 6.

Results of the experiments on 13000 simulated population studies of 20 control vs. 20 patients in 13 configurations of group differences. (a) Illustration of the synthetic group differences introduced (fractions are encoded as transparency of the tensor), as described in Table 1. (b) The detection rate when the population-informed prior is used is close to that of the ground truth. In particular, no additional false positive is observed under the null hypothesis (Configuration 1). The false negative rate slightly increases as the number of properties affected by group differences increases (Configurations ⑪–⑬). However, in those cases, the voxel as a whole can still reliably be detected as affected (leftmost column). (c) Illustration of the shrinkage towards the mean effect observed with the population-informed prior and the large variance observed when no prior are used, for two different configurations.

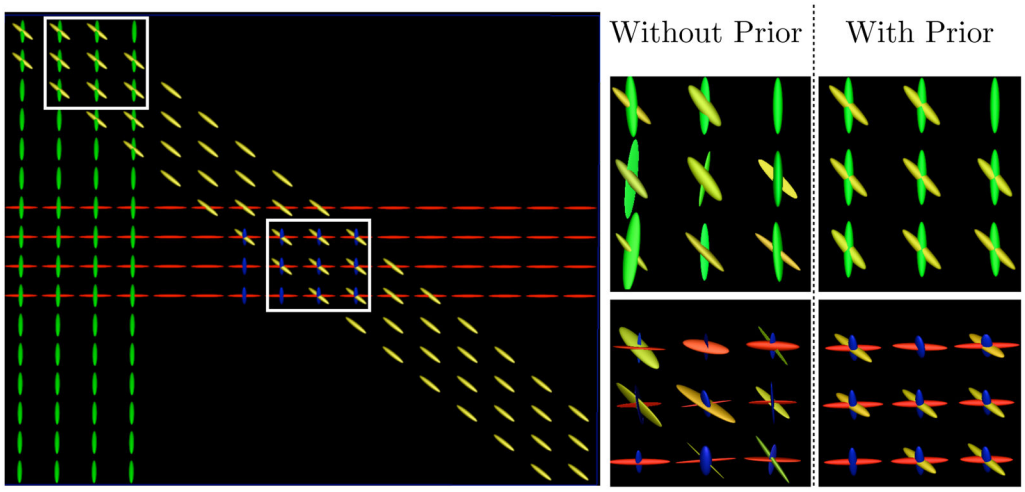


Figure 7.

(Left) The synthetic phantom used to conduct the experiment includes areas with zero (isotropic diffusion only), one, two and three fascicles. The blue, green and yellow tensors have axial diffusivity of 1.55×10^{-3} mm^2/s and radial diffusivity of 2.73×10^{-4} mm^2/s. The red tensors have axial diffusivity of 1.77×10^{-3} mm^2/s and radial diffusivity of 1.64×10^{-4} mm^2/s. The fraction of isotropic diffusion f_{iso} is equal to 0.15 throughout the image except in areas with no tensor where it is equal to 1. Fractions of anisotropic diffusion are equally split into the different tensors present.(Right) The better accuracy mostly affects the diffusion properties of tensors and not their directions, as predicted by Equation (2).

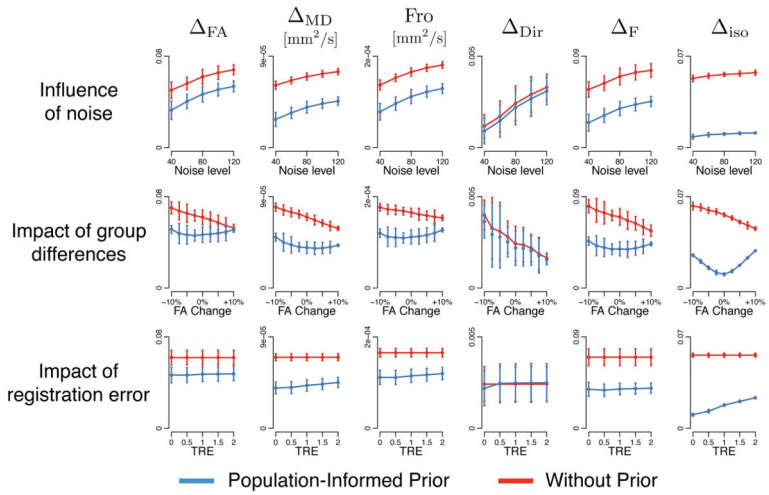


Figure 8.

Evolution of the six comparison metrics under the influence of noise, group differences and registration error for the estimation of the synthetic phantom. Significantly better results were obtained by utilizing the population-informed prior. The directions are not affected by the absence of a population-informed prior, as predicted by Equation (2).

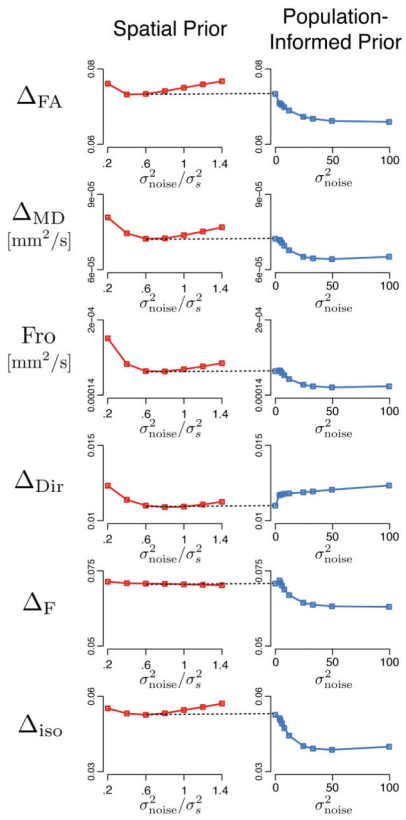


Figure 9.

(Left) Evolution of the six comparison metrics when only a spatial prior is used, for different weight ratios $\sigma^2_{\text{noise}}/\sigma^2_s$. (Right) Evolution of the six comparison metrics when both a spatial prior and the population-informed prior are used, for different values of the noise level σ^2_{noise} with $\sigma^2_{\text{noise}}/\sigma^2_s=0.6$. The use of a population-informed prior significantly improves the estimation accuracy for all five non-directional metrics. The evolution is smooth and monotonic with respect to σ^2_{noise} so that moderate errors in the estimation of σ^2_{noise} do not have a large impact on the result.

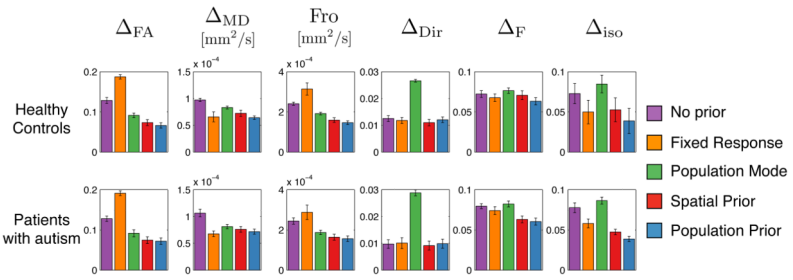


Figure 10.

Accuracies of the estimation of a multi-fascicle model from single-shell HARDI data with the five methods under comparison: (1) estimation without prior, (2) estimation with a fixed response (fixed eigenvalues), (3) estimation by setting all parameters to the mode of the population, (4) estimation with a spatial prior only and (5) estimation with a population-informed prior. Results with the population-informed prior have overall better performances than all other methods.

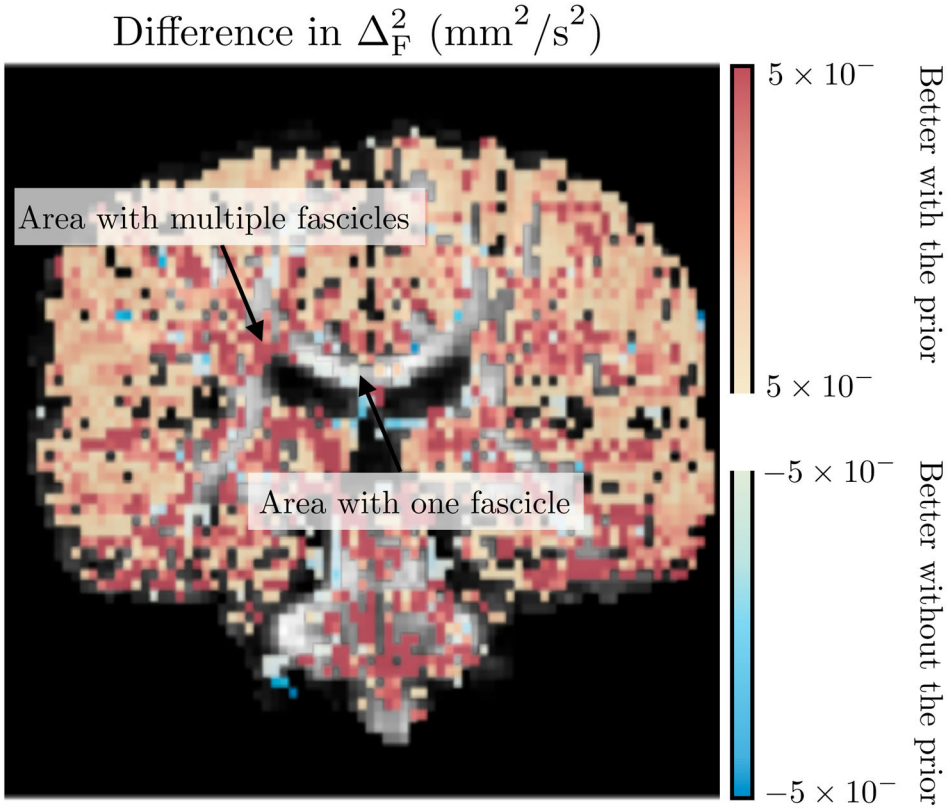


Figure 11.

Difference in Δ_F^2 between the estimation without the prior and with the population-informed prior. Values larger than zero indicate voxels where the population-informed prior improves the estimation. As indicated by the arrows, the improvement is more important in areas with crossing fascicles. In areas without crossing fascicles, the estimation without prior is not ill-posed and results in similar accuracies as the estimation with a population-informed prior.

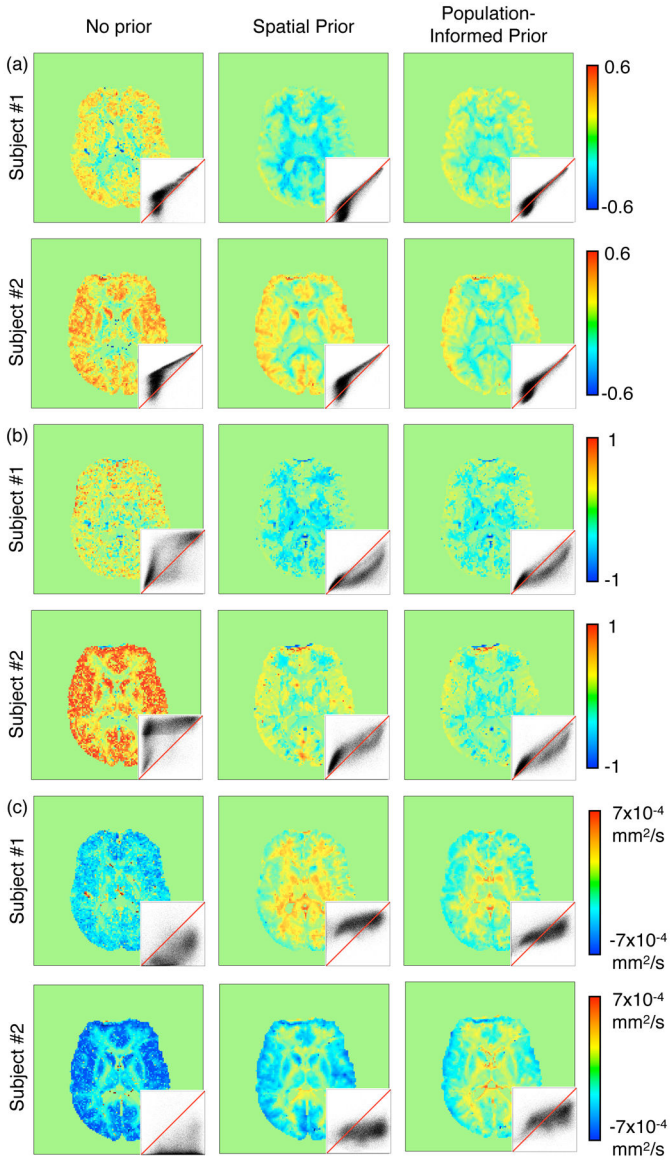


Figure 12.

Maps of estimation errors in (a) isotropic fraction, (b) maximum FA, (c) maximum radial diffusivity, for different estimation methods. The spatial prior leads to maps that are visually plausible but affected by a global bias (*e.g.*, bias in isotropic fraction is negative for Subject 1 and positive for Subject 2). This bias is due to the arbitrary choice of a model and the propagation of this choice through neighboring voxels. By contrast, the population-informed prior reduces the estimation bias as already observed in synthetic examples. Scatter plots next to every image depicts the joint distribution of the corresponding property in the ground truth (*x*-axis) and in the estimation (*y*-axis) highlighting the bias observed in estimations with a spatial prior.

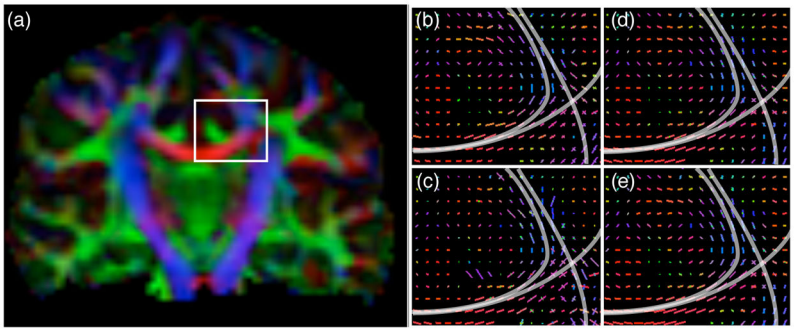


Figure 13.

In terms of direction, the spatial prior and the population-informed prior perform equally well. (a) Coronal slice of a subject's color FA highlighting the centrum semiovale. (b) Directions estimated with the DWI at multiple b-values. Directions estimated with DWI at a single b-value (c) without prior, (d) with a spatial prior only, (e) with both spatial and population-informed priors.

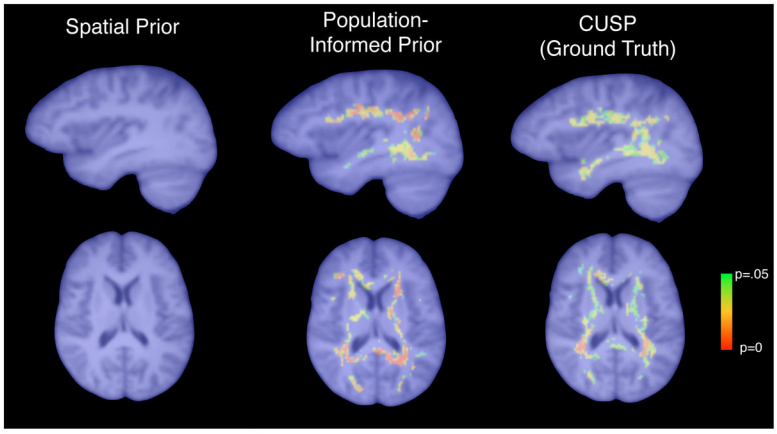


Figure 14.

Maps of the *p*-values of isotropic diffusion analysis, thresholded at $p < 0.05$ after correction for family-wise error rate. The population-informed prior leads to inference of group differences that are remarkably close to those obtained with the data at several b-values (CUSP). A spatial prior alone fails to detect any difference due to biases in estimations of the isotropic fraction.

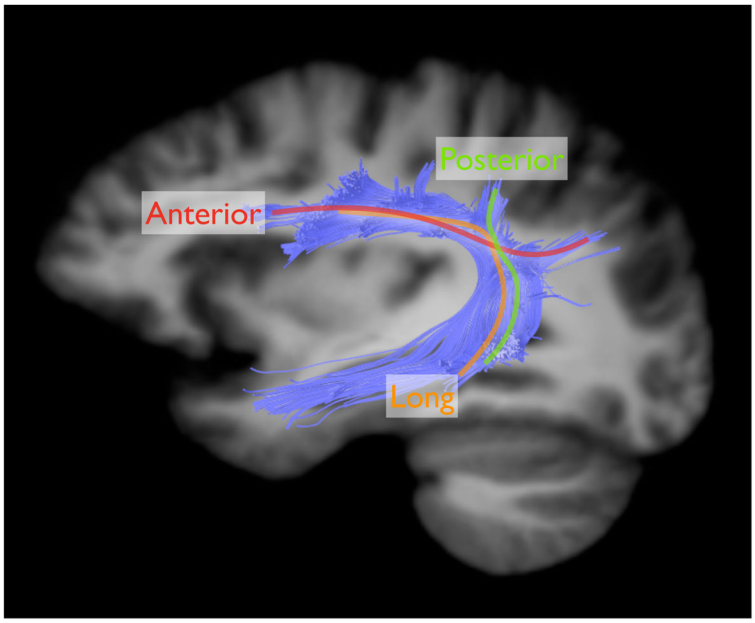


Figure 15.
The dorsal language circuit is composed of white matter fascicles thought to connect Broca's area in the frontal lobe, Wernicke's area in the temporal lobe, and Geschwind's territory in the parietal lobe.

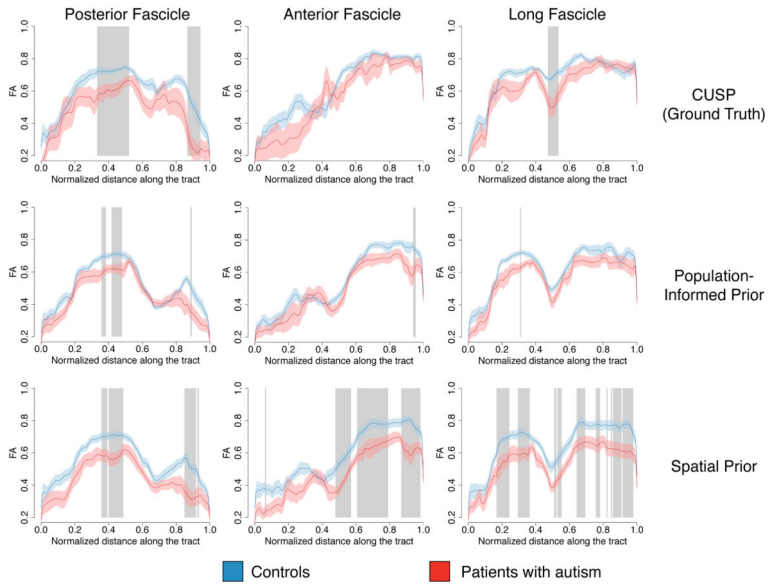


Figure 16.

Results of fascicle-based spatial statistics showing the mean FA (dark line) and its confidence intervals (colored shadowed curves) along the three main fascicles of the dorsal language circuit. Segments of significant differences are depicted as grey rectangles. Estimations with a population-informed prior detect most differences observed with the ground truth and does not lead to excessive false discovery rates. By contrast, estimations with a spatial prior generate false positives at a high rate, covering the anterior fascicle and the long fascicle by over 35%.

Table 1

Simulated alterations in terms of fraction f and eigenvalues of the first ($\lambda_{1,j}$) and second ($\lambda_{2,j}$ b) tensor.

	Δf	$\Delta\lambda_{1,1}$ $\times 10^{-3}$ mm^2/s	$\Delta\lambda_{1,2}$ $\times 10^{-3}$ mm^2/s	$\Delta\lambda_{1,3}$ $\times 10^{-3}$ mm^2/s	$\Delta\lambda_{2,1}$ $\times 10^{-3}$ mm^2/s	$\Delta\lambda_{2,2}$ $\times 10^{-3}$ mm^2/s	$\Delta\lambda_{2,3}$ $\times 10^{-3}$ mm^2/s
①	–	–	–	+0.3	–	–	–
②	−0.1	–	–	–	–	–	–
③	–	−0.3	–	–	–	–	–
④	–	–	–	+0.3	–	–	–
⑤	+0.1	+0.3	–	–	–	–	–
⑥	−0.1	+0.3	–	–	–	–	–
⑦	+0.1	–	+0.3	–	–	–	–
⑧	−0.1	–	+0.3	–	–	–	–
⑨	–	–	+0.3	+0.3	–	–	–
⑩	–	−0.3	+0.3	+0.3	–	–	–
⑪	+0.1	–	+0.3	+0.3	–	–	–
⑫	–	–	+0.3	+0.3	–	+0.3	+0.3
⑬	–	−0.3	+0.3	+0.3	−0.3	+0.3	+0.3

Table 2

Average improvement achieved by utilizing the population-informed prior in the estimation of the synthetic phantom (in percents of the value obtained without the prior).

Metric / Experiment	Δ_{FA}	Δ_{MD}	Fro	Δ_{Dir}	Δ_{F}	Δ_{iso}
Influence of noise	26.2%	45.0%	34.1%	12.2%	46.9%	81.7%
Group differences	20.4%	38.7%	29.3%	8.3%	39.6%	64.0%
Registration error	23.6%	38.7%	29.9%	−2.4%	44.6%	66.5%