

Estimation of a Multi-Fascicle Model from Single B-Value Data with a Population-Informed Prior

Maxime Taquet^{1,2}, Benoît Scherrer¹, Nicolas Boumal²,
Benoît Macq², and Simon K. Warfield¹

¹ Computational Radiology Laboratory, Harvard Medical School, Boston, USA

² ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Abstract. Diffusion tensor imaging cannot represent heterogeneous fascicle orientations in one voxel. Various models propose to overcome this limitation. Among them, multi-fascicle models are of great interest to characterize and compare white matter properties. However, existing methods fail to estimate their parameters from conventional diffusion sequences with the desired accuracy. In this paper, we provide a geometric explanation to this problem. We demonstrate that there is a manifold of indistinguishable multi-fascicle models for single-shell data, and that the manifolds for different b-values intersect tangentially at the true underlying model making the estimation very sensitive to noise. To regularize it, we propose to learn a prior over the model parameters from data acquired at several b-values in an external population of subjects. We show that this population-informed prior enables for the first time accurate estimation of multi-fascicle models from single-shell data as commonly acquired in clinical context. The approach is validated on synthetic and in vivo data of healthy subjects and patients with autism. We apply it in population studies of the white matter microstructure in autism spectrum disorder. This approach enables novel investigations from large existing DWI datasets in normal development and in disease.

Keywords: Diffusion, Single-Shell, Generative Models, Estimation

1 Introduction

Diffusion tensor imaging is unable to represent the signal arising from crossing fascicles. Various approaches have been proposed to overcome this limitation. Among them, generative models such as multi-tensor models [2,3] seek to represent the signal contribution from different populations of water molecules. Based on biological modelling, they are of great interest to characterize and compare white-matter properties. However, estimating their parameters from conventional diffusion data has proven inefficient.

Recent works have suggested that part of this inaccuracy is explained by the ill-posedness of the problem and not only by the imaging nuisance [3,4]. To

MT and NB are research fellows of the F.R.S.-FNRS. MT is also research fellow of the B.A.E.F. This work was supported in part by NIH grants 1U01NS082320, R01 EB008015, R03 EB008680, R01 LM010033, R01 EB013248, P30 HD018655, BCH TRP, R42 MH086984 and UL1 TR000170.

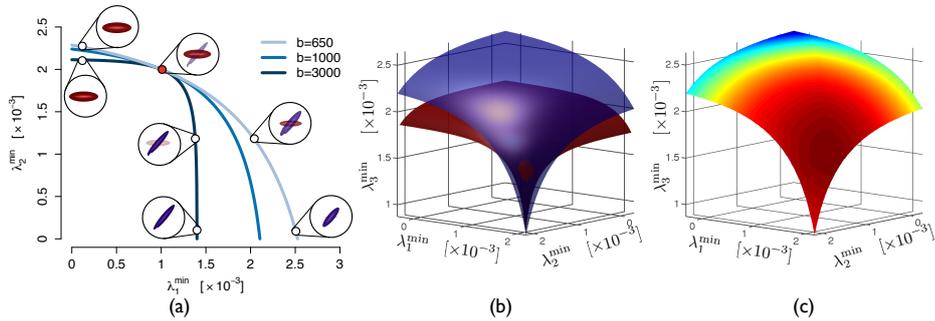


Fig. 1. (a) Infinitely many models produce the same diffusion signal at a given b -value and form a manifold. The manifolds for different b -values intersect at the true underlying model. (b) For N -fascicle models (here $N=3$), manifolds are $(N-1)$ -dimensional hypersurfaces that intersect tangentially, making the estimation sensitive to noise. (c) The population-informed prior assigns different probabilities to models on the manifold.

regularize the estimation of models with a single anisotropic tensor, elaborate spatial priors have been proposed [2], and it was shown that acquiring additional b -values improves the analysis of isotropic fraction [1]. For N -tensors, it was proposed to fix the tensor eigenvalues [4], solving the ill-posedness problem but reducing the amount of microstructural information contained in the model. No method has proposed to regularize the estimation of an N -fascicle model while keeping all its degrees of freedom. Furthermore, there is a strong need for a strategy to estimate multi-fascicle models from conventional single-shell data due to their wide availability in clinical setting. Section 2 analyzes the estimation problem from a geometric point of view. Section 3 develops an estimator based on a prior informed by an external population of subjects. Section 4 presents results and Section 5 concludes. Conclusions about estimating an N -tensor model can be applied to all generative models that include a multi-tensor as part thereof.

2 Manifolds of Equivalent Models at a Given B-Value

A multi-fascicle model is represented as a mixture of single fascicle models. In the multi-tensor formalism, the generative model for the formation of the diffusion signal S for a b -value b and a gradient direction \mathbf{g} is:

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

where \mathbf{D}_i and f_i are the tensor and the volumetric fraction of fascicle i . Since $\gamma_i e^{-\log \gamma_i} = 1$, all multi-fascicle models with fractions $\gamma_i f_i$ and tensors $\mathbf{D}_i + \frac{\log \gamma_i}{b} \mathbf{I}$ produce the same signal:

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

The tensors remain positive definite as long as $\gamma_i > e^{-b\lambda_i^{\min}}$, where λ_i^{\min} is the lowest eigenvalue of \mathbf{D}_i . Each of these models is uniquely identified by its vector $(\lambda_1^{\min}, \dots, \lambda_N^{\min})$. The set of all models respecting Equation (2) is a manifold of dimension $(N - 1)$ defined by the implicit equations (we let $\lambda_i := \lambda_i^{\min}$):

$$\begin{cases} \lambda_i = \lambda_i^{\text{true}} + \frac{1}{b} \log(\gamma_i) & \text{for } i = 1, \dots, N \\ \sum_{i=1}^N \gamma_i f_i = 1, \end{cases} \quad (3)$$

where $(\lambda_1^{\text{true}}, \dots, \lambda_N^{\text{true}})$ is the true unknown model (Fig. 1(a)). Since these equations depend on b , so will the manifold. Acquiring diffusion images at different b-values amounts to defining different such manifolds. Let us investigate how those manifolds intersect at the point of interest $\lambda_i = \lambda_i^{\text{true}}$. The explicit equation of the hypersurface $\lambda_N(\lambda_1, \dots, \lambda_{N-1})$ obtained by eliminating the γ 's between equations (3) is:

$$\lambda_N(\lambda_1, \dots, \lambda_{N-1}) = \lambda_N^{\text{true}} + \frac{1}{b} \log \left(\frac{1 - \sum_{i=1}^{N-1} f_i e^{b(\lambda_i - \lambda_i^{\text{true}})}}{f_N} \right). \quad (4)$$

The normal vector to the hypersurface is $\boldsymbol{\eta} = \left(\frac{\partial \lambda_N}{\partial \lambda_1}, \dots, \frac{\partial \lambda_N}{\partial \lambda_{N-1}}, -1 \right)$. Its k -th component evaluated at the true model is:

$$\eta_k \Big|_{\lambda_i = \lambda_i^{\text{true}}, \forall i} = \frac{\partial \lambda_N}{\partial \lambda_k} \Big|_{\lambda_i = \lambda_i^{\text{true}}, \forall i} = \frac{-f_k}{f_N}. \quad (5)$$

Remarkably, this normal vector (and thereby the tangent hyperplane) does not depend on b at the point of interest. 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)).

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

$$\mathbf{H} \Big|_{\lambda_i = \lambda_i^{\text{true}}} = \frac{-b}{f_N^2} \left(\tilde{f} \tilde{f}^T + f_N \text{diag}(\tilde{f}) \right),$$

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

$$x^T (\mathbf{H}(b) - \mathbf{H}(b')) x = \frac{b' - b}{f_N^2} \left((\tilde{f}^T x)^2 + f_N x^T \text{diag}(\tilde{f}) x \right) > 0. \quad (6)$$

Therefore, there exists no direction x along which the two manifolds have the same curvature. Consequently, the true model is locally the only intersection of all manifolds. Given the difference (6), it appears that a wider range of b-values 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).

When an isotropic compartment $f_{\text{iso}} e^{-bD_{\text{iso}}}$ is added to the model, one can show that the above development remains valid with an unchanged N if D_{iso} is known and considering an $(N + 1)$ -fascicle model if D_{iso} needs also be optimized.

3 Posterior Predictive Distribution of the Parameters

While all models of (3) are equally compatible with the observed DWI at a given b-value, they are not all as likely from a biological point of view. This knowledge can be learnt from available observations at several b-values of a fascicle i in m_i subjects $\mathcal{F}_i = \{f_i^0, \dots, f_i^{m_i}\}$, $\mathcal{D}_i = \{\mathbf{D}_i^0, \dots, \mathbf{D}_i^{m_i}\}$, and incorporated in the estimation as a prior over the parameters (f_i, \mathbf{D}_i) (Fig. 1(c)). If the effect of the fascicle properties on partial voluming is negligible, and if the properties of one fascicle are independent of those of another, then the prior can be expressed as:

$$P_{\mathbf{f}, \mathbf{D}}(\mathbf{f}, \mathbf{D}; \boldsymbol{\theta}) = P_{\mathbf{f}}(\mathbf{f}; \boldsymbol{\theta}_f) \prod_{i=1}^N P_{\mathbf{D}_i}(\mathbf{D}_i; \boldsymbol{\theta}_i). \quad (7)$$

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 $f_j/(1 - f_i)$. This *neutral vector* assumption naturally leads to the Dirichlet distribution:

$$P_{\mathbf{f}}(\mathbf{f}; \boldsymbol{\alpha}) = \frac{\mathbb{1}_{\mathbf{f} \in \mathcal{S}}}{B(\boldsymbol{\alpha})} \prod_{i=1}^N f_i^{\alpha_i - 1}, \text{ where } \mathcal{S} = \left\{ \mathbf{x} \in \mathbb{R}^N : x_i > 0, \sum_{i=1}^N x_i = 1 \right\}. \quad (8)$$

To prevent negative eigenvalues of the tensors, the prior knowledge about \mathbf{D}_i can be described as a multivariate Gaussian distribution over their logarithm [5]:

$$\mathbf{L}_i = \log \mathbf{D}_i \sim \mathcal{N}(\mathbf{M}_i, \boldsymbol{\Sigma}_i). \quad (9)$$

In general, $\boldsymbol{\Sigma}_i$ has 21 free parameters, which may overfit the usually small training dataset. For DTI, it is suggested in [5] to constrain $\boldsymbol{\Sigma}_i$ to be orthogonally invariant, imposing the following structure that depends only on σ_i and τ_i :

$$\boldsymbol{\Sigma}_i = \sigma_i^2 \begin{pmatrix} I_3 + \frac{\tau_i}{1-3\tau_i} \mathbb{1}_3 & \mathbf{0} \\ \mathbf{0} & I_3 \end{pmatrix} \triangleq \mathbf{B}(\sigma_i, \tau_i).$$

This structure yields a closed-form solution for the maximum likelihood [5]:

$$\hat{\mathbf{M}}_i = \bar{\mathbf{L}}_i = \frac{1}{m_i} \sum_{k=1}^{m_i} \mathbf{L}_i^k \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_i = \mathbf{B}(\hat{\sigma}_i, \hat{\tau}_i), \quad (10)$$

$$\text{with } \hat{\tau}_i = \frac{-\sum_{i=1}^{m_i} \|\mathbf{L}_i^k - \bar{\mathbf{L}}_i\|_2^2}{5 \sum_{i=1}^{m_i} [\text{Tr}(\mathbf{L}_i^k - \bar{\mathbf{L}}_i)]^2} \quad \text{and} \quad \hat{\sigma}_i^2 = \frac{1}{6m_i} \sum_{i=1}^{m_i} \|\mathbf{L}_i^k - \bar{\mathbf{L}}_i\|_{\hat{\tau}_i}^2, \quad (11)$$

where $\|\cdot\|_t^2$ is defined by $\langle \mathbf{A}, \mathbf{B} \rangle_t = \text{Tr}(\mathbf{A}\mathbf{B}) - t \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{B})$. 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* (PPD) which contains all the knowledge about new observations that we learn from previous observations. Its derivation requires the definition of hyperpriors over $\boldsymbol{\theta}$ and is closed-form if we select conjugate hyperpriors.

$\mathbf{M}_i \sim \mathcal{N}(\mathbf{M}_0, \mathbf{A}_0)$ is a conjugate hyperprior for the tensor part of (7) assuming a deterministic $\boldsymbol{\Sigma}_i = \hat{\boldsymbol{\Sigma}}_i$. We set $\mathbf{A}_0 = \mathbf{B}(1, 0)$ and $\mathbf{M}_0 = \log \mathbf{D}_{\text{iso}}$ to keep it weakly informative (this hyperprior merely encodes the order of magnitude of diffusivity at 37°C). The PPD over the tensors is $\mathbf{D}_i | \mathcal{D}_i \sim \mathcal{N}(\mathbf{M}_i^{m_i}, \mathbf{A}_i^{m_i})$ with

$$\mathbf{A}_i^{m_i} = \hat{\boldsymbol{\Sigma}}_i + \left(\mathbf{A}_0^{-1} + m_i \hat{\boldsymbol{\Sigma}}_i^{-1} \right)^{-1} \triangleq \mathbf{B}(\tilde{\sigma}_i, \tilde{\tau}_i), \quad (12)$$

$$\text{and } \mathbf{M}_i^{m_i} = \left(\mathbf{A}_0^{-1} + m_i \hat{\boldsymbol{\Sigma}}_i^{-1} \right)^{-1} \left(\mathbf{A}_0^{-1} \mathbf{M}_0 + m_i \hat{\boldsymbol{\Sigma}}_i^{-1} \bar{\mathbf{L}}_i \right). \quad (13)$$

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 \mathcal{S} . The resulting PPD is a Dirichlet with parameters $1 + \sum_{k=1}^{m_i} f_i^k$. In this expression, we consider f_i^k as frequency counts since they are samples of \mathbf{f}_i rather than samples from a multinomial parameterized by \mathbf{f}_i . The complete PPD is (with $C_{\mathcal{F}, \mathcal{D}}$ constant):

$$P_{\mathbf{f}, \mathbf{D}}(\mathbf{f}, \mathbf{D} | \mathcal{F}, \mathcal{D}) = C_{\mathcal{F}, \mathcal{D}} \mathbb{1}_{\mathbf{f} \in \mathcal{S}} \prod_{i=1}^N f_i^{\sum_{k=1}^{m_i} f_i^k} \prod_{i=1}^N \exp \left\{ -\frac{\|\mathbf{L}_i - \mathbf{M}_i^{m_i}\|_{\tilde{\tau}_i}^2}{2\tilde{\sigma}_i^2} \right\}. \quad (14)$$

We incorporate this PPD as a prior in the estimation. We assume Gaussian noise on the DWI measurements y_k since they are acquired on a single shell typically at $b=1000$ for which noise is approximately Gaussian. The maximum a posteriori estimator at each voxel amounts to maximizing the following for \mathbf{f} and \mathbf{D} :

$$L = \log(\mathbb{1}_{\mathbf{f} \in \mathcal{S}}) + \sum_{i=1}^N \sum_{k=1}^{m_i} f_i^k f_i - \sum_{i=1}^N \frac{\|\mathbf{L}_i - \mathbf{M}_i^{m_i}\|_{\tilde{\tau}_i}^2}{2\tilde{\sigma}_i^2} - \sum_{k=1}^K \frac{(S_k(\mathbf{f}, \mathbf{D}) - y_k)^2}{2\sigma_k^2}. \quad (15)$$

The influence of the noise σ_k^2 is analyzed in the next section. In practice, the prior is built from data acquired in completely different subjects at several b-values. All these subjects are registered to a multi-fascicle atlas as in [7]. Following alignment, tensors from all subjects at each voxel are clustered in N compartments as in [6]. Each cluster represents the sets \mathcal{F}_i and \mathcal{D}_i of available observations. The prior is then aligned with an initial estimate (without prior) of the multi-fascicle model. To evaluate (14), all assignments of compartments to tensors are considered and the highest prior value is recorded. BOBYQA algorithm is used to maximize (15) and the number of fascicles is estimated by an F-test as in [3].

4 Results

We compare the models estimated by our method to a ground truth $\{g_i, \mathbf{G}_i\}$ with five root mean square metrics, Δ_{FA} , Δ_{MD} , Fro , Δ_{F} and Δ_{iso} defined by:

$$\Delta_{\text{FA}}^2 = \sum_{i=1}^N \frac{f_i + g_i}{2} (\text{FA}(\mathbf{F}_i) - \text{FA}(\mathbf{G}_i))^2, \quad \Delta_{\text{MD}}^2 = \sum_{i=1}^N \frac{f_i + g_i}{2} (\text{MD}(\mathbf{F}_i) - \text{MD}(\mathbf{G}_i))^2$$

$$\text{Fro}^2 = \sum_{i=1}^N \frac{f_i + g_i}{2} \|\mathbf{D}_i - \mathbf{G}_i\|_F, \quad \Delta_{\text{F}}^2 = \sum_{i=1}^N (f_i - g_i)^2 \quad \text{and} \quad \Delta_{\text{iso}}^2 = (f_{\text{iso}} - g_{\text{iso}})^2.$$

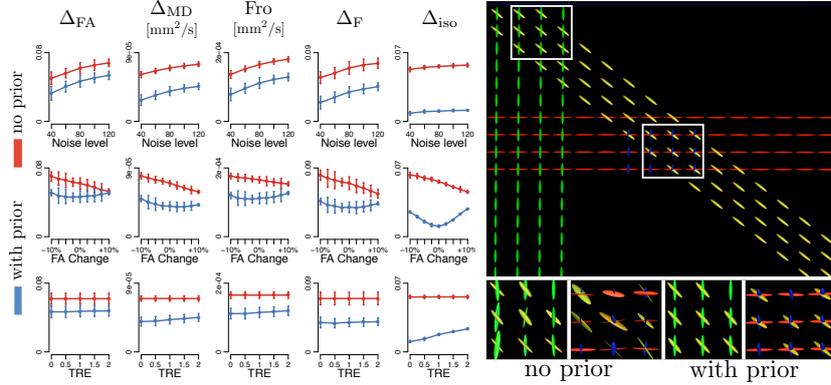


Fig. 2. (Left) Incorporating the prior in the estimation significantly improves the accuracy of the estimated model under the three simulated scenarios and for all five comparison metrics (distributions are shown for 20 datasets simulated for each set of parameters). (Right) The better accuracy mostly affects the diffusion properties of tensors (other than their directions), as predicted by Equation (3).

Synthetic Phantom Experiment DWI were simulated under Rician noise from a phantom containing an isotropic compartment and 0 to 3 tensors of various properties with $S_0=400$. The prior was built from 20 datasets of 90 DWI at $b=1000, 2000$ and 3000 . The accuracy was evaluated with 20 datasets of 30 DWI at $b=1000$ in three scenarios. First, the noise variance increased from 40 to 120. Second, the FA of the phantom was offset by -10% to $+10\%$ without changing the prior to simulate patient’s data with a prior built from healthy subjects. Third, random deformations of 0 to 2 voxels were applied to the prior to simulate registration errors. In the last two scenarios, the noise variance was 80. In all scenarios and for all metrics, incorporating the prior significantly improved the accuracy of the estimation (one-tail paired t-test: $p < 10^{-6}$) (Fig. 2)

In Vivo Data Experiment Eighteen healthy subjects and 10 subjects with autism were imaged to test the method and an extra 13 healthy subjects were imaged to build the prior. For all subjects, DWI at resolution $1.7 \times 1.7 \times 2\text{mm}^3$ were acquired with a Siemens 3T Trio with a 32 channel head coil using the CUSP-45 sequence [3]. This includes 30 gradients on a single-shell at $b = 1000$ and 15 gradients with b-values up to 3000. For each test subject, all 45 DWI were first used to estimate a multi-fascicle model considered as a ground truth. Estimations using the single-shell subset only were then compared to it. Four strategies were compared: estimation without prior, estimation by fixing all tensors to a globally optimized value, and estimation with the prior assuming a noise level σ_k^2 of 20 and 500. Results in Fig. 3 show that estimations which incorporate a prior outperform other strategies. Estimations with $\sigma_k^2 = 500$ are significantly better than estimations without prior for all metrics and, remarkably, for both healthy controls and ASD patients (one-tail paired t-test: $p < 10^{-6}$). The true

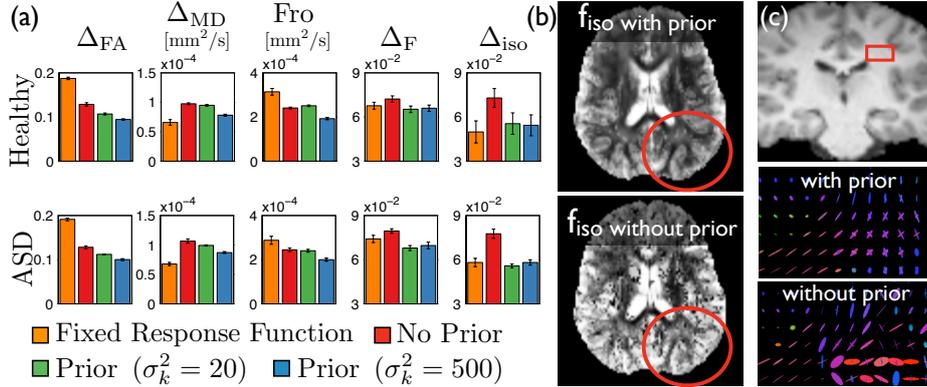


Fig. 3. (a) Incorporating prior knowledge significantly improves the quality of the model estimation for all five metrics and for both healthy controls and ASD patients. This improvement implies that (b) the extracellular water fraction can be visualized with more contrast and less noise in smaller details of the white matter up to its boundary with the grey matter, and (c) properties of the fascicles in crossing areas (shown is the corona radiata) are better represented and do not suffer the arbitrary choice of a model from Equation (3).

noise level of DWI is arguably closer to 500 than 20. However, estimations with $\sigma_k^2 = 20$ remain more accurate than estimations without prior, indicating that the population-informed prior improves the model accuracy even for crude estimates of the noise level. Empirical estimates of this noise level is kept for future work. Finally, fixing the fascicle response results in accuracies that strongly vary among quality metrics and, furthermore, only provides average information about the brain microstructure, which is not suitable in most studies.

Application to Population Studies One could be concerned that the improved accuracy brought by the prior would come with a severe shrinkage of the estimated parameters towards the mean of the population. This would prevent its use in population studies. To address this concern, we conducted two population studies of autism spectrum disorder (ASD) using the proposed estimator. The first one focused on fascicle properties in the left arcuate fasciculus by analyzing the FA along the median tract. The second study investigated whether an increased extracellular volume fraction f_{iso} is observed in ASD. Less restricted diffusion may be related to the presence of edema, thinner axons, and neuroinflammation [1]. The latter has been proposed as a possible cause of autism. Corrections for multiple comparisons were based on cluster-size statistics in 1000 permutations with a threshold on t -scores of 3. As presented in Fig. 4, the first study revealed decreased FA integrity in the arcuate fasciculus of patients with ASD, in line with most recent studies of autism. The second study revealed one clusters of significantly increased unrestricted diffusion (permutation test: $p < 0.003$). Without the prior, none of these findings were observed ($p > 0.1$).

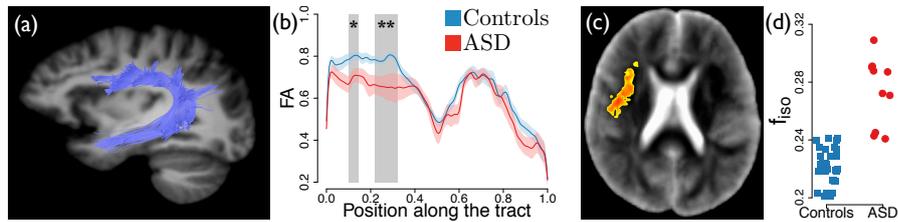


Fig. 4. The population-informed prior enables population studies of multi-fascicle models from single-shell HARDI data. (a-b) The first study reveals significantly decreased FA related to autism in the left arcuate fasciculus (* $p < .05$, ** $p < .01$). (c-d) The second study reveals a cluster of significantly higher f_{iso} . (d) Average f_{iso} in the cluster.

These studies show that the use of a prior in the estimation preserves contrasts of diffusion properties between groups, so that single-shell HARDI data can be used in large population studies based on multi-fascicle models.

5 Conclusion

Multi-fascicle models cannot be estimated from conventional single-shell HARDI data because a manifold of models produce the same diffusion signals. However, we showed that a posterior predictive distribution over the model parameters can be learnt from data acquired at several b-values in an external population. By incorporating this population-informed prior in the maximum a posteriori estimator of the parameters, we are able to estimate accurate multi-fascicle models from data at a single b-value. This method thus opens new opportunities for population studies with the large number of available clinical diffusion images.

References

1. Pasternak, O., Shenton, M., Westin, C.F.: Estimation of extracellular volume from regularized multi-shell diffusion MRI. In: MICCAI. pp. 305–312 (2012)
2. Pasternak, O., Sochen, N., Gur, Y., Intrator, N., Assaf, Y.: Free water elimination and mapping from diffusion MRI. *Magnet Reson Med* 62(3), 717–730 (2009)
3. Scherrer, B., Warfield, S.K.: Parametric representation of multiple white matter fascicles from cube and sphere diffusion MRI. *PLoS one* 7(11), e48232 (2012)
4. Schultz, T., Westin, C.F., Kindlmann, G.: Multi-diffusion-tensor fitting via spherical deconvolution: a unifying framework. In: MICCAI. pp. 674–681 (2010)
5. Schwartzman, A., Mascarenhas, W.F., Taylor, J.E.: Inference for eigenvalues and eigenvectors of gaussian symmetric matrices. *Ann Stat* pp. 2886–2919 (2008)
6. Taquet, M., Scherrer, B., Benjamin, C., Prabhu, S., Macq, B., Warfield, S.: Interpolating multi-fiber models by gaussian mixture simplification. In: IEEE ISBI. pp. 928–931 (2012)
7. Taquet, M., Scherrer, B., Commowick, O., Peters, J., Sahin, M., Macq, B., Warfield, S.: Registration and analysis of white matter group differences with a multi-fiber model. In: MICCAI. pp. 313–320 (2012)