FAST IDENTIFICATION OF OPTIMAL FASCICLE CONFIGURATIONS FROM STANDARD CLINICAL DIFFUSION MRI USING AKAIKE INFORMATION CRITERION

Aymeric Stamm[‡][⋆]   *Olivier Commowick*[⋆]   *Patrick Pérez*[†]   *Christian Barillot*[⋆]

[‡] Computational Radiology Laboratory, Boston Children's Hospital, Harvard Medical School, USA
[⋆] Visages INSERM/INRIA U746, IRISA - UMR CNRS 6074, Rennes, France
[†] Technicolor, Rennes, France

ABSTRACT

Analytic multi-compartment models have gained a tremendous popularity in the recent literature for studying the brain white matter microstructure from diffusion MRI. This class of models require the number of compartments to be known in advance. In the white matter however, several non-collinear bundles of axons, termed *fascicles*, often coexist in a same voxel. Determining the optimal fascicle configuration is a *model selection problem*. In this paper, we aim at proposing a novel approach to identify such a configuration from clinical diffusion MRI where only few diffusion images can be acquired and time is of the essence. Starting from a set of fitted models with increasing number of fascicles, we use Akaike information criterion to estimate the probability of each candidate model to be the best Kullback-Leibler model. These probabilities are then used to average the different candidate models and output an MCM with optimal fascicle configuration. This strategy is fast and can be adapted to any multi-compartment model. We illustrate its implementation with the ball-and-stick model and show that we obtain better results on single-shell low angular resolution diffusion MRI, compared to the state-of-the-art automatic relevance detection method, in a shorter processing time.

Index Terms— diffusion MRI, multi-compartment models, model selection, model averaging, semioval center.

1. INTRODUCTION

Diffusion magnetic resonance imaging (dMRI) is a tool of choice for the analysis of the brain white matter (WM). The diffusion of thermally agitated water molecules in the axons is indeed restricted by the cylindrical geometry of such cells with largest diameters around $20~\mu$m. The typical spatial resolution in clinical dMRI is $2 \times 2 \times 2$ mm^3 cubic voxels, which implies that hundreds of non-collinear axons can presumably populate a given voxel. Axons are usually assumed to be distributed in orientationally homogeneous bundles termed *fascicles*. Water in a voxel then populates the different fascicle environments and the extra-axonal space in unknown proportions. Multi-compartment models (MCM) rely on finite mixtures of distributions and are thus well suited to model the voxelwise diffusion-induced molecular displacements. Existing MCMs include the 47 models reviewed in [1] and Diffusion Directions Imaging [2]. The number of fascicles however needs to be known in advance, which is not the case in practice. So far, this model selection problem has been solved either by brute-force methods or using Bayesian frameworks.

In brute-force approaches, a set of nested candidate MCMs with increasing number of fascicles is fitted to the data. The *best* MCM is then identified as the candidate model that best fits the data, where the comparison usually relies on an F-test [3]. Since the more complex the model, the better the fit, the F-test often tends to favor MCMs that overfit the data when the same data is used for estimation and to assess goodness of fit. To limit overfitting, the Bayesian information criterion has been introduced to penalize model complexity that increases with the number of fascicles [4]. Recently, the generalization error has been proposed to choose the "optimal" MCM based on its ability to *predict new data* [5], thus avoiding the overfitting issue. These approaches limit their search of the "optimal" MCM to a predefined candidate set.

Differently, Bayesian frameworks try to estimate the *best* MCM as the one that maximizes a posterior distribution on the models. They rely on a careful choice of a prior distribution for MCMs. For instance, [6] uses Markov random fields (MRF) while [7] resorts to Automatic Relevance Detection (ARD) in which non-informative priors are assigned to all the MCM parameters except the mixture weights that are assumed to be Beta-distributed. Such priors automatically prune an entire compartment if it is not supported by the data. These methods simultaneously perform model estimation and selection. When translated to clinics, Bayesian methods however have limitations. They are prohibitively computationally expensive: ARD (resp., MRF) requires MCMC simulations (resp., optimization) to come up with a usable posterior distribution. Moreover, dMRI clinical protocols often include a single b-value and a maximum of 30 diffusion-sensitizing gradients (DSG) [8]. With such small sample sizes, the posterior distribution strongly depends on the prior, making the Bayesian information updating potentially ineffective.

In this paper, we propose to combine the two previous approaches. Starting with a candidate set, we search for the best Kullback-Leibler (KL) MCM (i.e., at minimal KL divergence to the true unknown diffusion model) as a combination of the candidate models. In section 2.1, we show that we can estimate the probability of any MCM to be the best KL MCM using only the Akaike Information Criterion (AIC) and we explain how these probabilities can be used to generate the best KL MCM from the candidate set. In section 2.2, we implement our method for the ball-and-stick model [7], which allows us to set up a pilot study (Section 3) to compare our results to the ARD method on standard clinical data. Results are presented in section 4 and discussed in section 5.

2. PROPOSED APPROACH

2.1. Model averaging from Akaike information criterion

The AIC is widely used in brute force approaches where the *best* model is often chosen as the one with minimal AIC. However, by definition, it has a fundamental role in model averaging [9] that this section aims at summarizing.

The AIC is an asymptotically unbiased estimator of the relative expected KL divergence between an estimated model and the true unknown model. It is defined as follows:

$$\text{AIC} = -2 \log L + 2K, \quad (1)$$

where L is the maximized value of the likelihood for the estimated model and K is the number of parameters of the estimated model. When the sample size N is small, the following corrected AIC is recommended to avoid overfitting:

$$\text{AIC}_c = -2 \log L + 2K + \frac{2K(K+1)}{N-K-1}. \quad (2)$$

Since the true model is unknown, its AIC is approximated by the minimal AIC among those of candidate models. Let $\{\mathcal{M}_\ell\}_{\ell=0,\dots,L}$ be the set of candidate models. The AIC difference $\Delta_\ell = \text{AIC}_c(\ell) - \min_{k \in [\![0,L]\!]} \text{AIC}_c(k)$ of model $\mathcal{M}_\ell$ can then be used to define its so-called Akaike weight w_ℓ:

$$w_\ell = \frac{\exp\{-\Delta_\ell/2\}}{\sum_{k=0}^{L} \exp\{-\Delta_k/2\}}. \quad (3)$$

The Akaike weight w_ℓ approximates the probability for $\mathcal{M}_\ell$ to be the best KL model. Such probabilities happen to be very useful. For instance, if we sort the candidate models from largest to smallest Akaike weight, we can establish a 95% *confidence set* of models by keeping the first p models such that $\sum_{\ell=0}^{p} w_\ell = 0.95$. An *evidence ratio* between model $\mathcal{M}_{\ell_1}$ and $\mathcal{M}_{\ell_2}$ can also be estimated as w_{ℓ_1}/w_{ℓ_2} and used to assess how strongly $\mathcal{M}_{\ell_1}$ is more likely to be the best KL model compared to $\mathcal{M}_{\ell_2}$. Finally, a *model-averaged estimate* of any parameter θ across all candidate models can be

obtained as follows:

$$\widehat{\theta} = \frac{\sum_{\ell=0}^{L} w_\ell \mathbb{1}_{\Theta_\ell}(\theta)\widehat{\theta}_\ell}{\sum_{\ell=0}^{L} w_\ell \mathbb{1}_{\Theta_\ell}(\theta)}, \ \forall \theta \in \Theta_0 \cup \Theta_1 \cup \cdots \cup \Theta_L, \quad (4)$$

where $\mathbb{1}_{\Theta_\ell}$ is the characteristic function of the parameter set Θ_ℓ of model $\mathcal{M}_\ell$ and $\widehat{\theta}_\ell$ is the estimate of θ under model $\mathcal{M}_\ell$.

2.2. Ball-and-stick model averaging

The ℓ-fascicle ball-and-stick model. It assumes that the expected MR signal $S_i^{(\ell)}$ induced by a DSG with b-value b and direction $\boldsymbol{u}_i$ has the following parametric form for $\ell \geq 1$ [7]:

$$\frac{S_i^{(\ell)}}{S_0} = \left(1 - \sum_{j=1}^{\ell} f_j^{(\ell)}\right) e^{-bd^{(\ell)}} + \sum_{j=1}^{\ell} f_j^{(\ell)} e^{-bd^{(\ell)}\left(\boldsymbol{u}_i^{\mathsf{T}}\boldsymbol{\mu}_j^{(\ell)}\right)^2},$$

$$(5)$$

in which S_0 is the expected MR signal in absence of DSG and $\Theta_\ell = \left\{ \left(\pm\boldsymbol{\mu}_1^{(\ell)}, f_1^{(\ell)}\right), \dots, \left(\pm\boldsymbol{\mu}_\ell^{(\ell)}, f_\ell^{(\ell)}\right), d^{(\ell)}\right\}$ is the parameter set of the ℓ-fascicle ball-and-stick MCM where:

- the fascicle orientations $\pm\boldsymbol{\mu}_1^{(\ell)}, \dots, \pm\boldsymbol{\mu}_\ell^{(\ell)} \in \mathbb{S}^2$;
- the fascicle occupancies $f_1^{(\ell)}, \dots, f_\ell^{(\ell)} \in [0,1]$;
- the free diffusivity $d^{(\ell)} \geq 0$.

When $\ell = 0$, it reads $S_i^{(0)} = S_0 e^{-bd^{(0)}}$ and $\Theta_0 = \left\{d^{(0)}\right\}$.

Model averaging. We can observe that, for $\ell \in [\![0,L]\!]$, only the parameter d appears in more than one MCM with the same interpretation as the free diffusivity (it actually appears in all of them). Indeed, MCMs with an increasing number ℓ of fascicles are nested and there is thus no pairwise matching of the fascicle compartments between two different MCMs, making model-averaged estimates of the fascicle orientations/occupancies hard to define. For example, when $L = 3$, the unique fascicle of the 1-fascicle MCM $\mathcal{M}_1$ can be averaged either with the first or with the second fascicle of the 2-fascicle MCM $\mathcal{M}_2$. Each of these two combinations can further be averaged with one of the 3 fascicles of the 3-fascicle MCM $\mathcal{M}_3$, leading to a total of 6 averaged fascicles. In general, a candidate set made of MCMs from 0 to L fascicle compartments can generate up to $L!$ fascicles. To address this issue, we express the candidate MCMs with a same fixed number $L!$ of fascicle compartments, while ensuring a pairwise matching of the compartments *between* the MCMs. This is achieved by the indexing $k = (m-1)\frac{L!}{(\ell-1)!} + (j-1)\frac{L!}{\ell!} + p$ where $p \in [\![1, \frac{L!}{\ell!}]\!]$ and $m \in [\![1, (\ell-1)!]\!]$, which yields the following reformulation of eq. (5):

$$\frac{S_i^{(\ell)}}{S_0} = \left(1 - \sum_{k=1}^{L!} f_{L,k}^{(\ell)}\right) e^{-bd^{(\ell)}} + \sum_{k=1}^{L!} f_{L,k}^{(\ell)} e^{-bd^{(\ell)}\left(\boldsymbol{u}_i^{\mathsf{T}}\boldsymbol{\mu}_{L,k}^{(\ell)}\right)^2},$$

$$(6)$$

where $f_{L,k}^{(\ell)} = \frac{\ell}{L!} f_j^{(\ell)}$ and $\boldsymbol{\mu}_{L,k}^{(\ell)} = \boldsymbol{\mu}_j^{(\ell)}$. Note that this formulation remains valid for $\ell = 0$. Finally, we can apply eq. (4)

to eq. (6) to obtain model-averaged estimates of the ball-and-stick parameters as follows:

$$\widehat{d} = \sum_{\ell=0}^{L} w_\ell \widehat{d}^{(\ell)} \,, \ \widehat{f}_{L,k} = \sum_{\ell=0}^{L} w_\ell \widehat{f}_{L,k}^{(\ell)} \text{ and}$$

$$\widehat{\boldsymbol{\mu}}_{L,k} = \mathrm{E}_1 \left[D_{L,k} \right] \text{ with } D_{L,k} = \frac{\sum_{\ell=1}^{L} w_\ell \widehat{\boldsymbol{\mu}}_{L,k}^{(\ell)} \widehat{\boldsymbol{\mu}}_{L,k}^{(\ell)^{\mathrm{T}}}}{\sum_{\ell=1}^{L} w_\ell} \,, \quad (7)$$

where $\mathrm{E}_1 [\cdot]$ is the principal eigenvector operator.

3. MATERIAL & METHODS

3.1. Description of the data

Raw data. We conducted a pilot study in which two healthy volunteers (S1 and S2) underwent a series of 10 diffusion MRI scans on the same MR scanner (Siemens 3T Verio) with the same protocol. This protocol lasted 7 min and comprised a single non-weighted diffusion image $B0$ and 30 diffusion-weighted (DW) images acquired at $b = 1000$ s/mm^2 along 30 non-collinear DSG directions $\boldsymbol{u}_i$ uniformly spread over the north hemisphere. The following parameters were used: $128 \times 128 \times 60$ image resolution with $2 \times 2 \times 2$ mm^3 voxels, TR$= 11$ s and TE$= 99$ ms.

Preprocessing. We preprocessed the data with FSL [10]: (i) For each scan, we performed (a) a rigid registration of the DW images on the $B0$ to correct for subject motion and rotated the gradient tables accordingly and (b) an affine registration of the DW images on the $B0$, guided by the previously estimated rigid transformation, to correct for distortions. (ii) We performed a rigid registration of the $B0$ images of the different scans on the $B0$ image of the first scan and applied the corresponding transformation to the subsequent DW images. We rotated the gradient tables accordingly. (iii) For each scan, we reduced the noise in the images using the Rician-adapted non-local means filter [11]. (iv) We extracted the brain using the BET algorithm. (v) We computed a WM mask using the FAST algorithm.

3.2. Experimental setup

Maximal number of candidate MCMs. According to known anatomy [12], the semioval center is an area where association, commissural and projection fascicles cross. We thus need to fit MCMs up to at least 3 fascicles ($L \geq 3$). We aim at showing that setting $L = 3$ yields a sufficient candidate set for model averaging. To this end, we observed that the estimate of free water occupancy is highly biased for low L and gets better and stabilizes as L increases. Hence, we performed the model averaging procedure for $L = 3, 4$ and compared the resulting free water occupancy maps by means of Dice coefficient for different occupancy thresholds ranging in $]0, 1[$: if these two maps are highly similar, the contribution of the 4-fascicle MCM to the best MCM is negligible.

Comparison to ARD. Next, we propose to compare our approach to the ARD method as implemented in FSL [10] using $L = 3$. ARD alters the estimation of the fascicle occupancies in such a fashion that f_L^{ARD} estimates the probability that the L-fascicle MCM contributes to the *best* unknown MCM. With our approach, this probability is given by the Akaike weight w_L of the L-fascicle MCM. In order to compare f_3^{ARD} and w_3, we computed these indices on 100 bootstrap DW images generated out of the original 10. For both indices, we then performed an FDR-corrected z-test of nullity in all WM voxels. The aim is to compare the areas where the methods detect a statistically significant contribution of the 3-fascicle MCM to the best MCM, with respect to known anatomy. To this end, we focused on the semioval center and computed the mean of both indices wherever significantly non null. We also provide a qualitative visualization of the estimated fascicles.
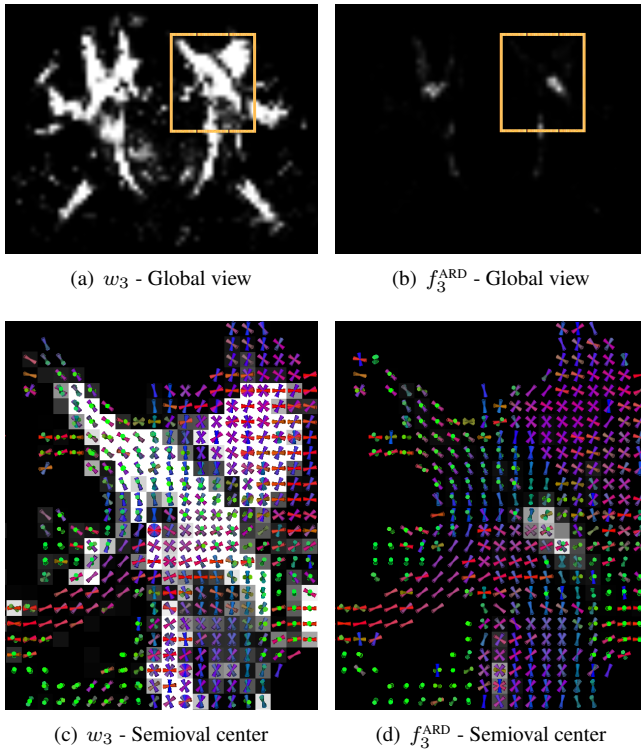
4. RESULTS

Maximal number of candidate MCMs. We computed the Dice coefficient between binarized free water occupancy maps obtained with $L = 3, 4$. Results are summarized in table 1. For all thresholds employed, the Dice coefficient remains greater than 0.95, which means that including a 4-fascicle MCM in the averaging does not bring much additional information. We can thus reasonably conclude that the contribution of the 4-fascicle MCM to the best MCM is negligible in the WM.

f_0 Map Threshold		0.2	0.4	0.6	0.8
Dice Score	S1	0.9987	0.9899	0.9818	0.9568
	S2	0.9993	0.9940	0.9953	0.9957

Table 1. Dice similarity index between free water occupancy (f_0) maps obtained assuming $L = 3, 4$ fascicle compartments. Different thresholds were used to binarize the maps.

Comparison to ARD. Figure 1 shows the results of the model selection obtained with our approach (left) and with ARD (right) on S1. A global coronal view of the statistically significant mean probabilities w_3 (Figure 1(a)) and f_3^{ARD} (Figure 1(b)) with same window level reveals that our approach offers a more robust detection of statistically significant contributions of the 3-fascicle MCM to the best MCM. We also zoomed on the semioval center (Figures 1(c) and 1(d)) where one expects to see predominantly 3 distinct fascicles. Results show that ARD hardly identifies voxels with 3 fascicles. It also requires to set a threshold on the fascicle occupancies (here, 0.05) to discard fascicles with too low occupancy, on which the resulting fascicle configurations highly depend. In contrast, our approach does not need to set any parameters and, most of the time, correctly identifies 1 fascicle in the corpus callosum and 3 fascicles in the semioval center. Comparative performances were identical for S2.

(a) w_3 - Global view

(b) f_3^{ARD} - Global view

(c) w_3 - Semioval center

(d) f_3^{ARD} - Semioval center

Fig. 1. **Model selection results in the semioval center.** Coronal view of the mean probabilities w_3 (left) and f_3^{ARD} (right) masked with the result of FDR-corrected z-test of nullity. Second row focuses on the semioval center and estimated fascicles are overlaid as equally shaped cones. Fascicle orientations are color-coded and match the cone orientations.

5. DISCUSSION

We presented a new approach to the model selection problem that enables the use of MCMs in a clinical setting. We combined ideas from brute-force approaches starting with a set of candidate models and ideas from Bayesian estimation quantifying the probability of each candidate to be the best one. We focused on the best KL model, which only requires AIC computation which is instantaneous. The use of a bias-corrected AIC makes the procedure robust to small sample sizes thus suited to clinical data. On such data, our method proved to be more efficient than ARD in correctly identifying 3-fascicle crossing areas. It is worth noting that no user-defined parameters are required contrary to ARD. From a computational point of view, the averaging procedure is very fast (a few seconds) but requires the estimation of MCMs up to 3 fascicles, which is a longer step. Nonetheless, estimation and averaging took about 5 hours for ARD and only 1 hour for our approach on an 8-core computer, which makes it closer to a clinically acceptable running time. Future works will focus on the accuracy of the estimated fascicle orientations. This is a tricky problem since (i) no ground truth is available and (ii) the op-

timal fascicle configurations might be model-dependent.

6. REFERENCES

[1] E. Panagiotaki *et al.*, "Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison.," *NeuroImage*, vol. 59, no. 3, pp. 2241–54, 2012.

[2] A. Stamm, P. Pérez, and C. Barillot, "A new multi-fiber model for low angular resolution diffusion MRI.," in *ISBI*, 2012, pp. 936–39.

[3] D. Alexander, G. Barker, and S. Arridge, "Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data.," *Magn. Reson. Med.*, vol. 48, no. 2, pp. 331–40, 2002.

[4] T. Schultz, C. Westin, and G. Kindlmann, "Multi-diffusion-tensor fitting via spherical deconvolution: a unifying framework.," in *MICCAI*, 2010, vol. 13, pp. 674–81.

[5] B. Scherrer, M. Taquet, and S. Warfield, "Reliable selection of the number of fascicles in diffusion images by estimation of the generalization error," in *IPMI*, 2013, vol. 7917, pp. 742–753.

[6] C. Demiralp and D. Laidlaw, "Generalizing diffusion tensor model using probabilistic inference in Markov random fields.," in *MICCAI CDMRI Workshop*, 2011.

[7] T. Behrens *et al.*, "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?," *NeuroImage*, vol. 34, no. 1, pp. 144–55, 2007.

[8] D. Jones, "The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study.," *Magn. Reson. Med.*, vol. 51, no. 4, pp. 807–15, 2004.

[9] D. Posada and T. Buckley, "Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests.," *Syst. Biol.*, vol. 53, no. 5, pp. 793–808, 2004.

[10] M. Jenkinson *et al.*, "FSL.," *NeuroImage*, vol. 62, no. 2, pp. 782–90, 2012.

[11] N. Wiest-Daesslé *et al.*, "Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: applications to DT-MRI.," in *MICCAI*, 2008, vol. 11, pp. 171–9.

[12] H. Kretschmann and W. Weinrich, *Cranial neuroimaging and clinical neuroanatomy: magnetic resonance imaging and computed tomography.*, Thieme, 2004.