

# Population-based Bayesian regularization for microstructural diffusion MRI with NODDIDA

Meghdoot Mozumder<sup>1,2</sup> | Jose M. Pozo<sup>3</sup> | Santiago Coelho<sup>3</sup> | Alejandro F. Frangi<sup>3,4</sup>

<sup>1</sup>Centre for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB), Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield, United Kingdom

<sup>2</sup>Department of Applied Physics, University of Eastern Finland, Kuopio, Finland

<sup>3</sup>Centre for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB), School of Computing and School of Medicine, University of Leeds, Leeds, United Kingdom

<sup>4</sup>LICAMM Leeds Institute of Cardiac and Metabolic Medicine, School of Medicine, University of Leeds, Leeds, United Kingdom

## Correspondence

Alejandro F. Frangi, Centre for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB), School of Computing and School of Medicine, University of Leeds, United Kingdom. and LICAMM Leeds Institute of Cardiac and Metabolic Medicine, School of Medicine, University of Leeds, Leeds, United Kingdom.

Email: a.frangi@leeds.ac.uk

Twitter: @affrangi

**Purpose:** Information on the brain microstructure can be probed by Diffusion Magnetic Resonance Imaging (dMRI). Neurite Orientation Dispersion and Density Imaging with Diffusivities Assessment (NODDIDA) is one of the simplest microstructural model proposed. However, the estimation of the NODDIDA parameters from clinically plausible dMRI acquisition is ill-posed, and different parameter sets can describe the same measurements equally well. A few approaches to resolve this problem focused on developing better optimization strategies for this non-convex optimization. However, this fundamentally does not resolve ill-posedness. This article introduces a Bayesian estimation framework, which is regularized through knowledge from an extensive dMRI measurement set on a population of healthy adults (henceforth *population-based prior*).

**Methods:** We reformulate the problem as a Bayesian maximum a posteriori estimation, which includes as a special case previous approach using non-informative uniform priors. A population-based prior is estimated from 35 subjects of the MGH Adult Diffusion data (Human Connectome Project), acquired with an extensive acquisition protocol including high *b*-values. The accuracy and robustness of different approaches with and without the population-based prior is tested on subsets of the MGH dataset, and an independent dataset from a clinically comparable scanner, with only clinically plausible dMRI measurements.

**Results:** The population-based prior produced substantially more accurate and robust parameter estimates, compared to the conventional uniform priors, for clinically feasible protocols, without introducing any evident bias.

**Conclusions:** The use of the proposed Bayesian population-based prior can lead to clinically feasible and robust estimation of NODDIDA parameters without changing the acquisition protocol.

## KEYWORDS

biophysical tissue models, diffusion MRI, microstructure imaging, modeling, parameter estimation

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Magnetic Resonance in Medicine* published by Wiley Periodicals, Inc. on behalf of International Society for Magnetic Resonance in Medicine

## 1 | INTRODUCTION

Diffusion magnetic resonance imaging (dMRI) allows in vivo and noninvasive mapping of water molecules' diffusive movement in biological tissues. This motion is constrained by the tissue microarchitecture.<sup>1</sup> Hence, combined with biophysical modeling, dMRI is potentially capable of capturing microstructural features related to tissue constituents. There exists several modeling techniques in the literature capable of capturing such information. *Signal models*, for instance, directly model the dMRI signal with a particular functional form. The most common of them is the diffusion tensor imaging,<sup>2</sup> which, despite its simplicity, can still provide meaningful biomarkers that are widely used as indications of microstructural tissue changes.<sup>3</sup> *Micro-structural models*,<sup>4</sup> instead, derive the dMRI signal from a physical model of the tissue microstructure (e.g.<sup>5-7</sup>). This allows capturing more specific information of individual tissue constituents.

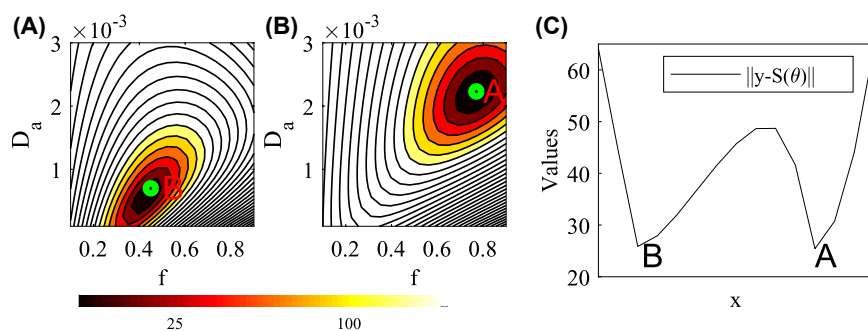
One of the most popular dMRI microstructural models is the neurite orientation dispersion and density imaging (NODDI).<sup>7</sup> NODDI describes the signal generated from a voxel as arising from three independent non-exchanging compartments: intra-neurite, extra-neurite, and cerebrospinal fluid (CSF). The intra-neurite compartment is modeled as a set of sticks, i.e. cylinders with zero radius, the extracellular compartment is modeled as set of cylindrically symmetric diffusion tensors, and the CSF as an isotropic compartment. Instead of estimating all model parameters directly from the data, NODDI makes some assumptions and constrains a few of its microstructural parameters for estimating the rest. These assumptions have been shown<sup>8-10</sup> to be not always valid and to result in biased estimates for the remaining model parameters.

Jelescu et al<sup>11</sup> suggested an alternative approach where these microstructural parameters were no longer considered fixed or constrained, called NODDIDA (NODDI with Diffusivity Assessment). This approach removes the incorrect assumptions of NODDI, but it makes the problem

ill-posed<sup>10</sup>: i.e. multiple parameter sets can describe the dMRI signal equally well (see Figure 1) and hence the solution is not unique. Also, it reduces NODDI from a three-compartment model to a two-compartment model, by eliminating the isotropic compartment. Hence, the model is only applicable in brain regions with minimal CSF contamination.

Despite these difficulties, these models describe several useful microstructural parameters for detecting pathological conditions, meriting further investigation. For example, the neurite orientation distribution has demonstrated potential in capturing white matter disarray in Alzheimer's disease (AD).<sup>12</sup> Extracellular radial diffusivity increases with reduced myelination,<sup>13,14</sup> and is observed in the development of multiple sclerosis (MS) and AD. Since demyelination, unlike axonal loss, is in many cases reversible, a method that differentiates between the two has potential clinical value. NODDIDA parameters are also prospective biomarkers for other brain disorders, such as axonal loss in MS and AD could be inferred from the intra-neurite fraction,<sup>13</sup> accumulation of fluids in intra- and extracellular spaces in cerebral edema and beading could be inferred from the intra- and extra-neurite diffusivities.<sup>15,16</sup>

The ill-posed parameter estimation in NODDIDA, sometimes referred to as *degeneracy*, requires some consideration. The NODDIDA degeneracy has been sometimes attributed to the non-convexity of the problem resulting in multiple local minima of the cost function. Non-convexity can be tackled using repeated local optimizations, starting from random parameter initializations, and selecting the solution with global minimum residual error.<sup>11</sup> But this strategy does not address the ill-posedness. The method produced reasonable estimates with extensive dMRI measurement protocols. However, for clinically feasible dMRI measurements, the method fails to resolve the NODDIDA degeneracy since there are multiple global minima with equal cost function value (as illustrated in Figure 1), making the optimum identification unstable for any noise level.



**FIGURE 1** Illustration of the degeneracy present in the likelihood function of the dMRI signal for the NODDIDA model. From the example NODDIDA parameters,  $\theta$ , in the Set B of Jelescu et al<sup>11</sup>, we generated the dMRI measurements,  $y$ , for a clinically feasible protocol and SNR. The residual norm,  $\|y-S(\theta)\|$ , presents 2 local minima with very similar values: the parameter point, **A**, generating the measurements, and another spurious minimum, **B**. Plots (A) and (B) display the cellular fraction,  $f$ , and intracellular diffusivity,  $D_a$ , plane (fixing the rest of the parameters) including the parameter points **B** and **A**, respectively. Plot (C) displays the line connecting both local minima

It was observed in Jelescu et al.<sup>11</sup> for a sample of cases, that the actual solution presented a wider basin of attraction in the parameter space than the spurious one. Assuming this observation is valid in general, they suggested to consider a corresponding alternative criterion for estimating the optimal solution: from several random initializations, selecting the local minimum with higher prevalence. This method was subsequently applied for a microstructural model similar to NODDIDA in Novikov et al.<sup>10</sup> However, this assumption is not correct in all situations. The case illustrated in Figure 1 is a counterexample, showing larger basin around the wrong minima. A systematic comparison of these methods, has not yet been performed.

A machine learning Bayesian approach, based on training using simulated data and assuming the traces of the intra- and extra-axonal diffusivities to be similar, has been proposed in Reisert et al.<sup>17</sup> The approach provided largely unreliable estimates for extracellular diffusivities, putting in doubt its assumption. In addition, this method estimates the parameters expected value, ignoring (not solving) the possible bimodality of the posterior. Moreover, the implementation of methods factoring out the neurite orientation distributions<sup>10,17</sup> is not straightforward, limiting its applicability.

In this paper, we demonstrate for the first time, that parameter estimation approaches proposed earlier<sup>10,11</sup> can be described within the Bayesian estimation formalism as maximum a posteriori (MAP) estimation with uniform probability density function (pdf) priors on the parameters. Subsequently, we used a prior on model parameters by estimating their probability density function from extensive dMRI measurements in a population of 35 healthy adult subjects. We coin these *population-based priors*, since they were estimated from a parameter sample in a reference population. We then compared MAP estimates with uniform and population-based priors using common optimization methods, using real dMRI data. We also compared these approaches to NODDI. We demonstrate that the estimation using the proposed population-based prior provides most reliable parameter estimates on clinical data that can be implemented on clinical dMRI protocols. However, no clear difference is observed between the performance of previously published optimization methods for each prior model. The proposed method is easily implemented by a straightforward modification of NODDI toolbox, increasing its potential application in clinical diagnosis.

## 2 | METHODS

### 2.1 | dMRI data

We have used the Massachusetts General Hospital (MGH) Adult Diffusion data,<sup>18</sup> available from the Human Connectome Project (HCP).<sup>19</sup> It is a high resolution, high  $b$ -value dMRI dataset, obtained from 35 healthy adult subjects between 20

and 59 years old, using a protocol that used a substantially large set of  $b$ -values and gradient directions (compared to clinical measurement protocols). Each subject of the MGH HCP has 40  $b_0$  images and measurements from 64 diffusion directions for  $b = 1000, 3000 \text{ s/mm}^2$ , 128 directions for  $b = 5000 \text{ s/mm}^2$ , and 256 directions for  $b = 10\,000 \text{ s/mm}^2$ .

We also used an independent dataset acquired by a 3T Siemens MAGNETOM Prisma<sup>Fit</sup> system with 80 mT/m maximum gradient strength, which is comparable to a clinical scanner. An EPI/spin echo (SE) diffusion-weighted pulse sequence was used with 75 ms echo time, 2700 ms pulse repetition time, and a  $128 \times 128$  acquisition matrix resulting in an isotropic voxel size of 2.5 mm. 11  $b$ -values were acquired ( $b = 0, 250, 350, 450, 550, 650, 750, 850, 1150, 1500, 2000 \text{ s/mm}^2$ ) with 12  $b_0$  images and 60 non-co-linear magnetic field gradient directions for the others. The total acquisition time was 33 minutes.

Only voxels with minimal CSF contamination were used in the analysis, selected by a threshold of 1% CSF fraction estimated by the “free water elimination” technique in Pasternak et al.<sup>20</sup>

### 2.2 | NODDIDA model

The NODDIDA microstructural model<sup>10,11,21</sup> considers a distribution of axonal fibers with two tissue components resulting in a biexponential kernel,

$$\mathcal{K}(b, \xi) = f e^{-b D_a \xi^2} + (1-f) e^{-b D_e^\perp - b(D_e^\parallel - D_e^\perp) \xi^2}, \quad (1)$$

dependent on the gradient strength,  $b$ , and the projection,  $\xi = \hat{\mathbf{g}} \cdot \hat{\mathbf{n}}$ , of the gradient direction,  $\hat{\mathbf{g}}$ , into the fiber direction,  $\hat{\mathbf{n}}$ . This kernel describes the diffusion due to a single axon and its extracellular matrix. Here,  $f$  describes the intracellular (axonal) fraction,  $D_a$  is the intracellular diffusivity, and  $D_e^\parallel$  and  $D_e^\perp$  are the parallel and perpendicular extracellular diffusivities. The dMRI signal attenuation is given by the convolution of the kernel with the fiber orientation distribution function  $p(\hat{\mathbf{n}})$ :

$$\mathbf{S} \equiv \mathbf{S}_{\hat{\mathbf{g}}}(b) = S_0 \cdot \int d\hat{\mathbf{n}} p(\hat{\mathbf{n}}) \mathcal{K}(b, \hat{\mathbf{g}} \cdot \hat{\mathbf{n}}). \quad (2)$$

NODDI<sup>7</sup> and NODDIDA models consider the Watson spherical distribution,  $p(\hat{\mathbf{n}}|\kappa, \hat{\boldsymbol{\mu}})$ , as fiber orientation distribution, parametrized by its main orientation  $\hat{\boldsymbol{\mu}}$  and the concentration  $\kappa$ , characterizing its anisotropy. Let  $\boldsymbol{\theta} = (f, D_a, D_e^\parallel, D_e^\perp, \kappa)$  denote the NODDIDA model parameters, and let  $\boldsymbol{\theta} \mapsto \mathbf{S}(\boldsymbol{\theta})$  be the deterministic forward model defined by (2). By deterministic, we mean, the model  $\mathbf{S}$  does not contain any signal noise or model inaccuracies. Typical clinical dMRI scanners show

SNRs  $>3$  for the human brain,<sup>22</sup> and as such the additive Rician noise,  $\epsilon$ , in measurements can be approximated by a Gaussian distribution.<sup>23</sup> Thus, the actual dMRI measurements can be modeled as

$$\mathbf{y} = \mathbf{S}(\boldsymbol{\theta}) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \boldsymbol{\Gamma}_\epsilon). \quad (3)$$

so that their conditional probability is

$$\pi(\mathbf{y}|\boldsymbol{\theta}) = \mathcal{N}(\mathbf{S}(\boldsymbol{\theta}), \boldsymbol{\Gamma}_\epsilon). \quad (4)$$

Assuming independent noise of the same standard deviation for all measurements, the covariance matrix will be proportional to the identity matrix,  $\boldsymbol{\Gamma}_\epsilon = \sigma^2 \mathbb{I}$ .

### 2.3 | Bayesian estimation

In the Bayesian approach to inverse problems, all unknowns and measured quantities are considered random variables and the uncertainty of their values is encoded into a probability density function (model). Using the Bayes theorem, we can express the posterior distribution

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (5)$$

in terms of the measurements model,  $\pi(\mathbf{y}|\boldsymbol{\theta})$ , and the prior pdf on the model parameters,  $\pi(\boldsymbol{\theta})$ . All these pdfs are probability densities on some high-dimensional space. One standard criterion for the estimation of the model parameters from the posterior probability, is the *maximum a posteriori* (MAP) estimate.

If a uniform prior distribution is considered for the model parameters,  $\boldsymbol{\theta} \sim \mathcal{U}(\mathbf{a}, \mathbf{b})$ , and the Gaussian measurement model (4) is explicitly expanded, the MAP estimate is given by

$$\begin{aligned} \boldsymbol{\theta}_{\text{MAP}} &= \arg \max_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta}|\mathbf{y}) \\ &= \arg \max_{\boldsymbol{\theta}} \exp\left(-\frac{1}{2}\|\mathbf{y} - \mathbf{S}(\boldsymbol{\theta})\|_{\boldsymbol{\Gamma}_\epsilon}^2\right) \mathcal{U}(\mathbf{a}, \mathbf{b}) \\ &= \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{S}(\boldsymbol{\theta})\|^2, \quad \text{subject to } \boldsymbol{\theta} \in [\mathbf{a}, \mathbf{b}]. \end{aligned} \quad (6)$$

Thus, this leads us to the least square cost function typically used<sup>10,11,21</sup> with box constraints

$$f \in [0, 1] \subset \mathbb{R}, \quad D_a, D_e^{\parallel}, D_e^{\perp} \in [0, 4] \subset \mathbb{R}, \quad \kappa \in [0, 64] \subset \mathbb{R}. \quad (7)$$

This estimator can be easily implemented as a modification to the NODDI toolbox<sup>7,11</sup> and is available for download from Mozumder.<sup>24</sup>

This cost function is usually interpreted as a maximum likelihood estimation (MLE) subject to constraints.<sup>11</sup> However, the Bayesian formulation allows us to consider diverse priors, incorporating available information on the problem at hand via the pdf of the model parameters.

### 2.4 | Population-based priors

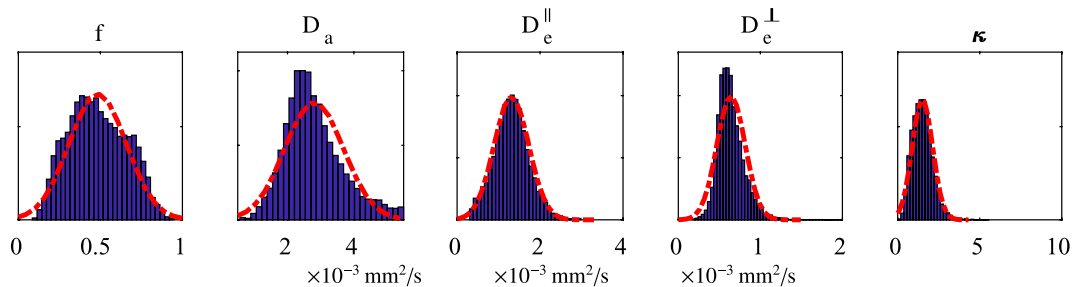
We use informative priors estimated from a sample of dMRI datasets with extended diffusion protocols. We used horizontal midbrain dMRI slices from the 35 subjects from the MGH HCP database. In contrast to more common clinical dMRI protocols, this protocol used to acquire the MGH HCP, had a substantially larger set of  $b$ -values and gradient directions. This makes the parameter estimation problem better posed, and the larger number of directional measures increases accuracy.<sup>10,11</sup>

We considered the method outlined in Jelescu et al<sup>11</sup> for solving Equation (6) using several parameter initializations and choosing the solution with minimum residual errors. The obtained parameters have been considered as ground truth (GT). A total of  $N = 44\,931$  horizontal midbrain voxels have been analyzed, obtaining the corresponding  $N$  sets of NODDIDA parameters:

$$\{\boldsymbol{\theta}^{(l)}, l = 1 \dots N\}. \quad (8)$$

The parameter distribution in this sample (shown in Figure 2) provides an estimate of their population pdf, which is used as prior in MAP estimation. All parameters display near-symmetric unimodal pdfs. Thus, we model them jointly as a multidimensional Gaussian pdf,

$$\pi(\boldsymbol{\theta}) = \mathcal{N}(\bar{\boldsymbol{\theta}}, \boldsymbol{\Gamma}_\theta), \quad (9)$$



**FIGURE 2** Distribution of NODDIDA parameters in the population estimated from midbrain sections of 35 patients from the MGH HCP dataset. They have been approximated by a multivariate Gaussian distributions (red dashed lines), defining the population-based prior. Their cross-correlations are not shown here



which allows an easy treatment of their correlation, with the mean parameter vector,  $\bar{\theta}$ , and covariance matrix,  $\Gamma_{\theta}$ , estimated from the sample (8). Mean and covariance calculated from the 35 subjects are provided in the Supporting Information Equation S1. Proceeding as in Equation (6), but using this Gaussian prior, the MAP estimate is given by

$$\begin{aligned} \theta_{\text{MAP}} &= \arg \max_{\theta} \exp\left(-\frac{1}{2} \|\mathbf{y} - S(\theta)\|_{\Gamma_{\epsilon}^{-1}}^2 - \frac{1}{2} \|\theta - \bar{\theta}\|_{\Gamma_{\theta}^{-1}}^2\right) \\ &= \arg \min_{\theta} (\|\mathbf{y} - S(\theta)\|_{\Gamma_{\epsilon}^{-1}}^2 + \|\theta - \bar{\theta}\|_{\Gamma_{\theta}^{-1}}^2), \end{aligned} \quad (10)$$

which is a generalized least mean square cost function.

The covariance  $\Gamma_{\epsilon}$  depends on the noise of the dMRI experimental data. We estimated it from the value of  $S_0$  and assumed a SNR = 50. This SNR was arbitrarily chosen, since typical clinically feasible protocols does not acquire many  $S_0$  images,<sup>11</sup> and hence, the SNR cannot be directly estimated from the standard deviation of  $S_0$  signals.

## 2.5 | Optimization strategies

The calculation of the maximum posterior estimate,  $\theta_{\text{MAP}}$ , with both uniform (6) and population-based Gaussian (10) priors, requires solving a nonlinear optimization problem and concomitant iterative optimization. We have used the Levenberg-Marquardt method, following the same approach as in Jelescu et al.<sup>11</sup> The MAP estimations with each type of prior, was carried out using three different optimization strategies:

### 2.5.1 | Single random initialization

The optimization algorithm is run only once, considering randomly sampled initialization parameters in the interval (7). This strategy only guarantees the localization of a random local optimum. Thus, the instability of the result evidences the non-convexity of the optimization problem.

### 2.5.2 | Maximum a posteriori selection

The optimization algorithm is run multiple times with different initializations.<sup>11</sup> The estimate yielding the maximum a posteriori probability is selected as estimate of the global optimum. We have considered 20 initializations, the same number as in Novikov et al.<sup>10</sup> This allows its direct comparison with their results, and we have observed no significant difference using a larger number of initializations.

### 2.5.3 | Highest prevalence selection

Optimization from multiple initializations. The most common solution is considered as the estimate representing the local optimum with wider basin of attraction.<sup>10</sup> As for the previous strategy, 20 initializations were used.

These three optimization strategies, combined with the two pdfs, give a total of six estimation strategies. For completeness, we have also compared the estimates obtained via NODDI (as in the NODDI toolbox<sup>7,11</sup>). NODDI fixes two diffusivities ( $D_a = D_e^{\parallel} = 1.7 \times 10^{-3} \text{ mm}^2/\text{s}$ ) and assumes that the third diffusivity is constrained ( $D_e^{\perp} = (1-f)D_e^{\parallel}$ ). Hence, it estimates only  $f$ ,  $\kappa$  out of the five NODDIDA parameters. Table 1 lists the seven estimation strategies with the corresponding acronyms used in this work.

## 2.6 | Experiments

From the complete dMRI extended acquisition protocol of the MGH HCP dataset, we considered a subset emulating a clinically feasible protocol.<sup>11</sup> This subset included 3  $b$ -values ( $b = 0, 1000, 3000 \text{ s/mm}^2$ ), consisting of 1  $b_0$  image and 30 non-co-linear gradient directions for other shells. The  $b = 3000 \text{ s/mm}^2$  is the closest value to  $b = 2000 \text{ s/mm}^2$ , used in Jelescu et al.,<sup>11</sup> available in the MGH HCP database. Despite this increased  $b$ -value, it is still clinically feasible.<sup>25</sup>

The parameters estimated from the complete extended protocol were used for two purposes. First, the population-based prior,  $\pi(\theta)$ , was estimated from them. Second, they were considered as the GT for the evaluation of the seven approaches considered, when applied on the clinically feasible subset. For a fair comparison, for the analysis of each subject, the prior in methods G1, Gm, and Gp did not include information from the same subject. That is, the prior estimation and accuracy evaluation were performed in a leave-one-subject-out fashion.

### 2.6.1 | Evidencing the degeneracy

In order to illustrate the presence of a degenerated estimation in NODDIDA parameter and to explore in detail how this affects different estimation strategies, we considered two random voxels selected one from the corpus callosum (CC) and another from the posterior limb of the internal capsule (PLIC), from subject *MGH\_1001*.

**TABLE 1** Acronyms for the models and estimation methods considered in this paper

Acronym	Prior	Optimization strategies
N	Uniform	NODDI
U1	Uniform	Single initialization.
Um	Uniform	Maximum a posteriori (20 init.)
Up	Uniform	Highest prevalence (20 init.)
G1	Population-based	Single initialization.
Gm	Population-based	Maximum a posteriori (20 init.)
Gp	Population-based	Highest prevalence (20 init.)

## 2.6.2 | Brain maps illustration

Two different brain regions of subjects *MGH\_1001* and *MGH\_1002* were considered to explore the spatial continuity of the model parameters and the importance of the generated noise with each of the seven estimation strategies. The NODDIDA parameter maps using the clinically feasible subset were also compared to the GT.

## 2.6.3 | Global accuracy evaluation

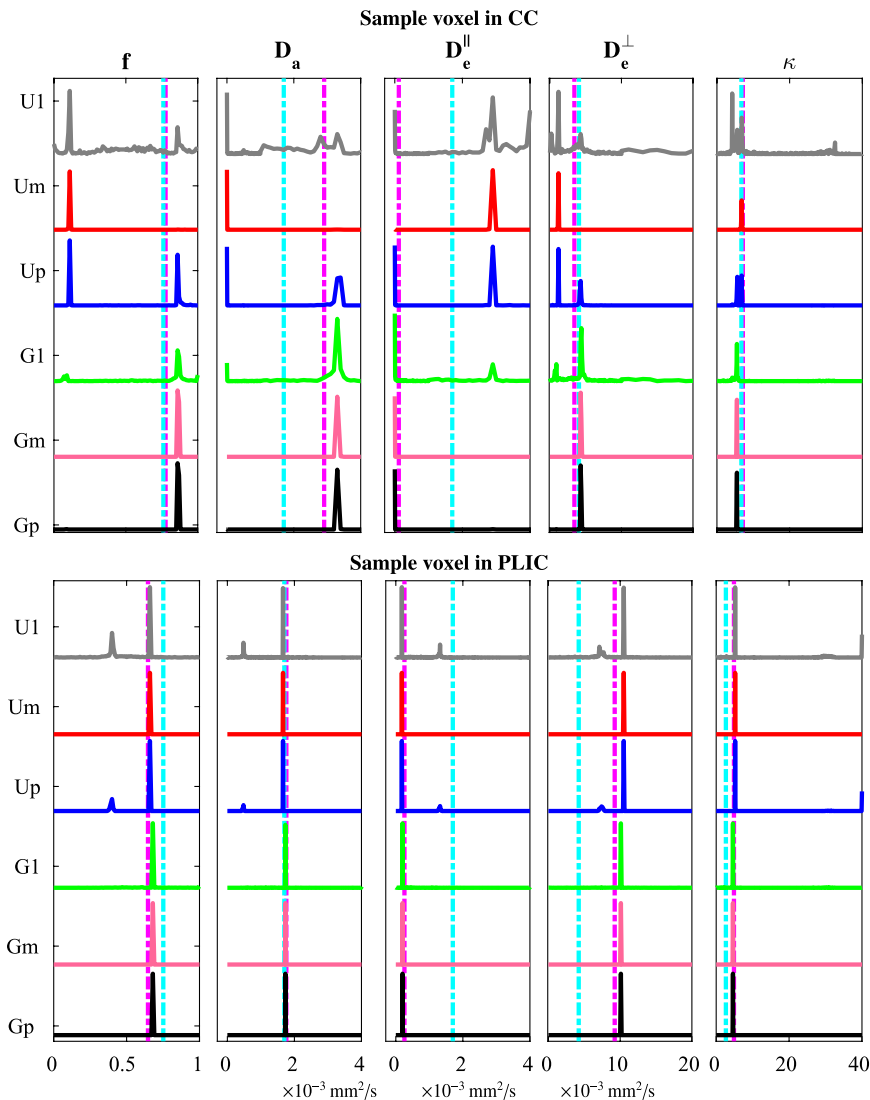
The midbrain slices of the full set of 35 subjects of the MGH HCP database was considered to evaluate and compare the accuracy of the seven estimation strategies when applied to the clinically feasible subset. To further explore the impact of the noise level on the parameter estimation accuracy, we subdivided the set of voxels into three groups, according to the SNR calculated using the standard deviation of the  $S_0$  data:  $\text{SNR} < 25$ ,  $25 \leq \text{SNR} \leq 35$ , and

$35 < \text{SNR}$ . The estimation accuracy for each of the parameters,  $\theta = (f, D_a, D_e^{\parallel}, D_e^{\perp}, \kappa)$ , was measured by the relative error

$$E_{\alpha} = |\theta_{\alpha}^{(Estimated)} - \theta_{\alpha}^{(GT)}| / \theta_{\alpha}^{(GT)} \times 100\%, \quad \text{for } \alpha = 1, \dots, 5. \quad (11)$$

## 2.6.4 | Using an independent dataset

We used the independent dMRI dataset from the MAGNETOM PrismaFit system to test the performance of the seven estimation strategies. First, the whole dataset (all  $b$ -values and gradient directions) were used to estimate the GT parameters. Then, a clinically feasible subset with three  $b$ -values ( $b = 0, 1150, 2000 \text{ s/mm}^2$ ), with 1  $b_0$  image and 30 directions from other shells were used for the evaluation of the estimation strategies. The population-based prior used in this case was based on the MGH HCP data.



**FIGURE 3** Illustration of the degeneracy in NODDIDA parameter estimation in two random voxels from CC and PLIC. The estimation of each of the NODDIDA parameters from a clinically feasible subset of dMRI measurements using the different estimation methods produce histogram of solutions, which can be multimodal. The cyan dotted lines represent the estimates using NODDI. The parameter estimates using the full extended protocol (GT) is shown in magenta dotted lines

### 3 | RESULTS

#### 3.1 | Evidencing degeneracy

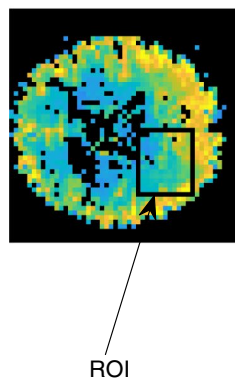
Figure 3 displays the histogram of estimates, for each model parameters obtained with each estimation method. It presents separately the results for a random voxel of CC and for a random voxel of PLIC. The multimodal distribution shown by U1 evidences the non-convexity of the dMRI likelihood for NODIDDA, presenting many local optima. The results for G1, show that this non-convexity is already partially mitigated for the posterior with the population-based prior. The global optimum selected by Um is close to the GT for the PLIC voxel, but identifies the wrong value in the CC. Only a slight improvement is obtained by the higher prevalence criteria in Up. In contrast, the population-based prior allows both Gm and Gp to localize the correct optimum. We can also observe that the values assumed by NODDI for the diffusivities are suboptimal, thus biasing the estimates for  $f$  and  $\kappa$ , in agreement with,<sup>8</sup> especially in the PLIC.

#### 3.2 | Brain dMRI example

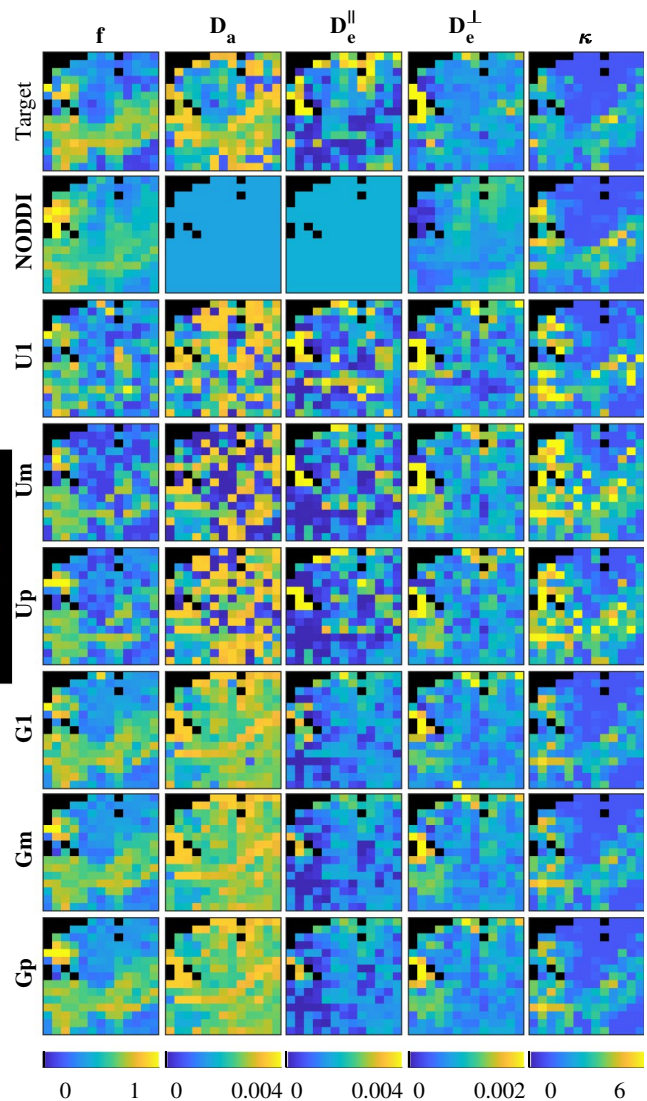
Figures 4 and 5 show model parameter maps in two different regions of the brain, each from a different subject. Parameter map of an entire brain slice is displayed in Supporting Information Figure S1. The estimates with U1, Um, and Up are very noisy. In contrast, the estimates obtained with population-based priors (methods G1, Gp, and Gm) are smoother and present more realistic patterns alike GT maps. In addition, the obtained values are in the ranges expected for human brains, qualitatively matching parameter maps in Novikov et al,<sup>10</sup> obtained with more extensive dMRI acquisition protocols, and showing higher ( $f, D_a, \kappa$ ) and lower ( $D_e^{\parallel}, D_e^{\perp}$ ) in white matter compared to gray matter. The maps obtained by NODDI are also smooth and qualitatively similar for  $f$  and  $\kappa$ , but cannot estimate the diffusivities.

#### 3.3 | Evaluating global accuracy

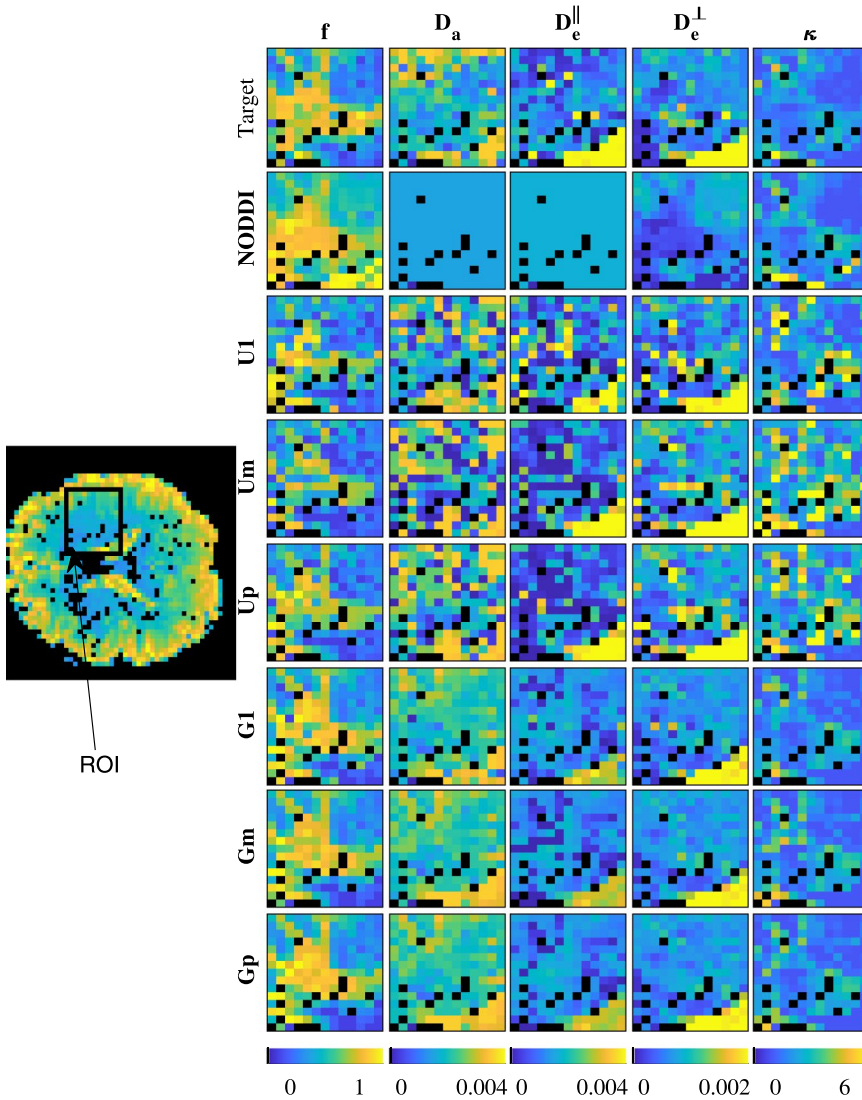
Figure 6 displays the distribution of relative errors obtained using each of the seven estimation strategies, for different



**FIGURE 4** Map of NODDIDA parameters estimated from a brain region of interest (ROI) from subject *MGH\_1001* (indicated within a rectangle in its  $S_0$  image on the left). Top row show the GT parameters estimated using the whole set of extended dMRI measurements. The rows below show the results from the seven different methods applied on a clinically feasible subset of measurements







**FIGURE 5** Map of NODDIDA parameters estimated from a brain region of interest (ROI) from subject *MGH\_1002* (indicated within a rectangle in its  $S_0$  image on the left). Top row show the GT parameters estimated using the whole set of extended dMRI measurements. The rows below show the results from the seven different methods applied on a clinically feasible subset of measurements

SNR ranges. The proposed population-based priors gave lower errors for all parameters and all SNRs, for the three optimization strategies. Both highest prevalence (Gp) and minimum cost (Gm) displayed similar errors. The only slightly higher errors obtained with a single initialization (G1) suggest, in agreement with Figure 3, that the non-convexity of the problem is largely mitigated. This could indicate that the global optima could be robustly found with fewer initializations than the 20 used here for Gm and Gp, making the estimation faster.

For the uniform prior, the maximum prevalence criterion (Up), provides in general worse results than minimum cost (Um). This contrasts with the observations in Jelescu et al,<sup>11</sup> subsequently applied in Novikov et al,<sup>10</sup> suggesting that minimum cost is a better criterion for the estimation of NODDIDA parameters.

The errors provided by NODDI were smaller than the ones from the NODDIDA estimation with uniform prior for all parameters. This happens even for the diffusivities, which have given fixed values in NODDI. In contrast, the inclusion

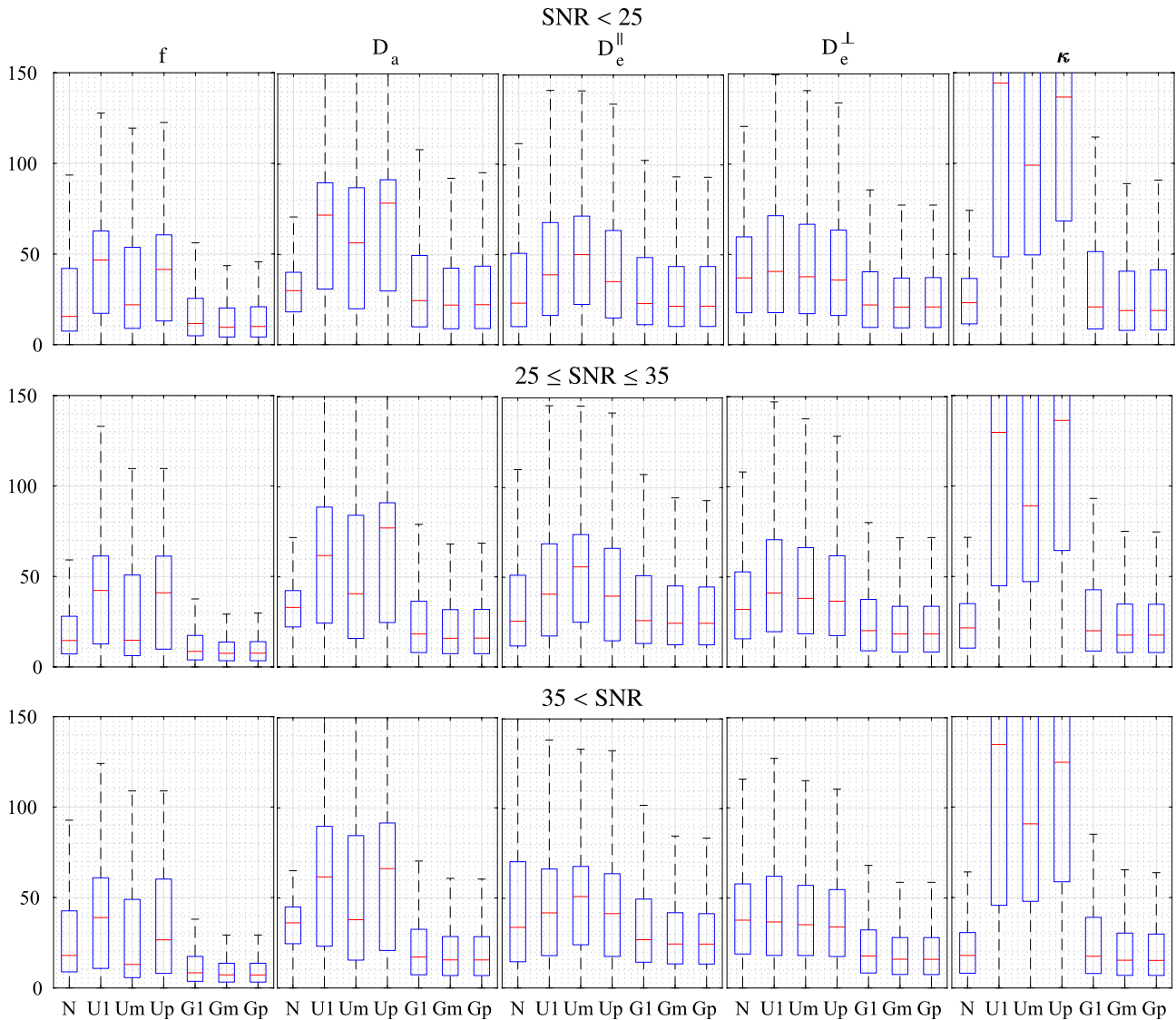
of the population-based prior leads to smaller errors also in comparison to NODDI.

The errors in the group with higher SNR are, in general, smaller. But this effect is more prominent for estimates with population-based priors. Thus, Gm estimations could further benefit from increased dMRI signal quality. The estimation with uniform priors is apparently not improved by the increase in SNR, probably due to the presence of bimodality in the estimations.

### 3.4 | Independent dataset

Figure 7 displays parameter maps from an independent dMRI dataset. We observe that the proposed population-based prior still provide smooth realistic parameter maps, closer to the target parameter maps. Figure 8 displays the distribution of relative errors obtained using the seven estimation strategies on the whole brain slice. The proposed population-based priors gave lower errors for all parameters.





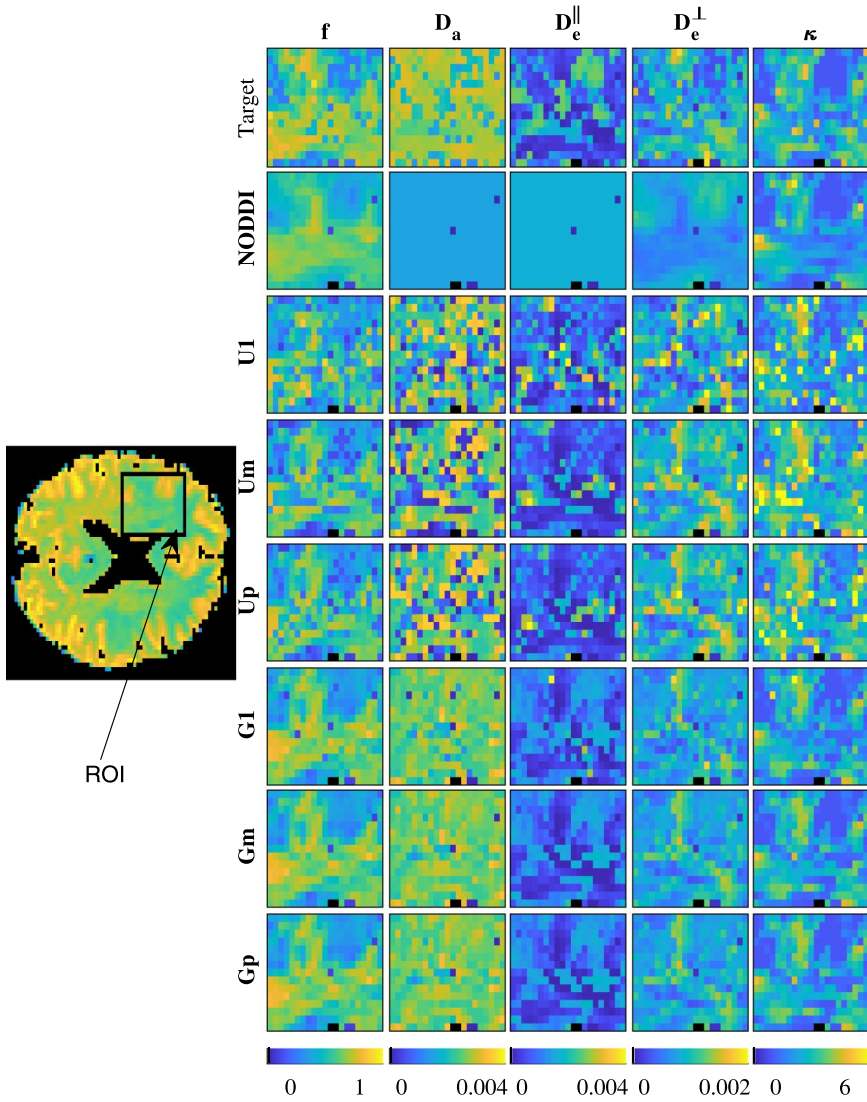
**FIGURE 6** Accuracy of the estimation of the NODDIDA parameters by each of the estimation methods applied to a clinically feasible subset of measurements. The boxplots represent the distribution of the relative estimation error for each NODDIDA parameter and for each method. The results have been stratified in three groups of SNR ranges

## 4 | DISCUSSION

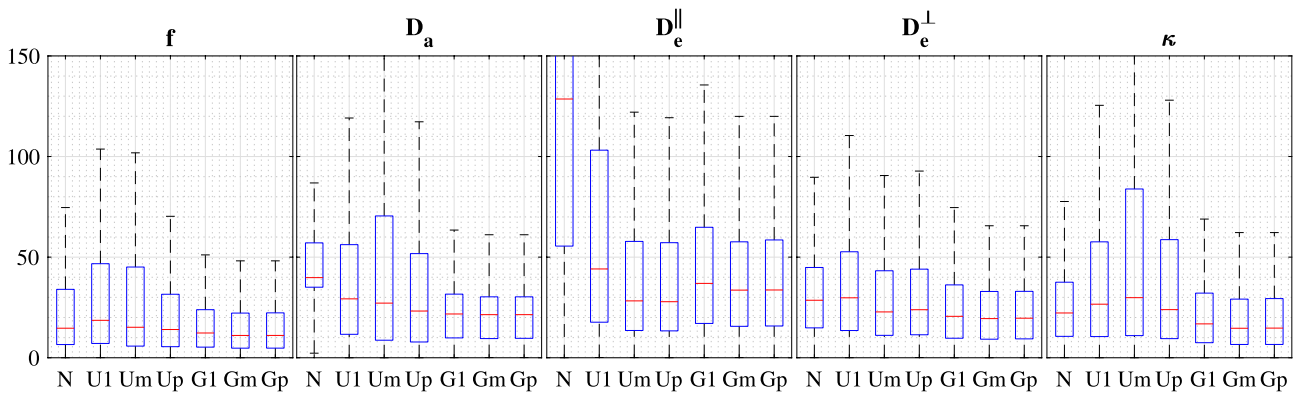
This work addresses parameter estimation in microstructural and in general, multi-compartmental dMRI models. Estimation of parameters in multi-compartmental dMRI models, such as NODDIDA, is inherently ill-posed. This problem fundamentally arises when microstructural parameter values for different compartments are exchanged resulting in indistinguishable dMRI signals. Approaches to resolve this degeneracy have focused on developing better optimization strategies addressing non-convex optimization. This, however, does not solve the ill-posed nature of the problem for clinically relevant dMRI acquisition protocols and SNR, where acquiring extensive diffusion orientations is unfeasible. The use of priors in the estimation of NODDIDA parameters has been suggested earlier in Jelescu et al.<sup>11</sup> However,

its use has been ill-advised in Novikov et al.,<sup>10</sup> following the prevalent belief among the diffusion community that priors are equivalent to constraints introducing bias in parameter estimation.

Nevertheless, we demonstrate in this work that the use of population-based priors can substantially increase the robustness and accuracy of the NODDIDA parameters estimates without introducing noticeable bias. The mean estimation errors reduced to below 16% using the proposed prior, compared to below 42% using previously proposed priors. There are two aspects to highlight here. First, in contrast with NODDI constraints, the prior distribution can model parameter correlations without fixing them to exact parameter values. Second, prior information is encoded based on the observation of the model parameter distribution across distinct patient populations, not in hypothetical parameter values.



**FIGURE 7** Map of NODDIDA parameters estimated from a brain region of interest (ROI) from the independent dMRI dataset (indicated within a rectangle in its  $S_0$  image on the left). Top row show the GT parameters estimated using the whole set of extended dMRI measurements. The rows below show the results from the seven different methods applied on a clinically feasible subset of measurements



**FIGURE 8** Accuracy of the estimation of the NODDIDA parameters by each of the estimation methods applied to a clinically feasible subset of measurements from the independent dMRI dataset. The boxplots represent the distribution of the relative estimation error for each NODDIDA parameter and for each method

In our methods we chose SNR 50, since clinically feasible datasets doesn't have multiple  $b_0$  values to allow the estimation of SNRs. A good guess of SNR can further improve the estimation accuracy, see Supporting Information Figure S2 and the discussion therein.

We evidence that previously proposed approaches in Novikov et al<sup>10</sup> and Jelescu et al<sup>11</sup> are akin to using non-informative uniform priors within our Bayesian estimation framework. To develop more informative priors integrating information from specific populations, we considered

the distribution of the model parameters estimated from 35 subjects of the MGH dataset from the HCP.<sup>19</sup> These data comprise dMRI measurements with an extensive protocol including high  $b$ -values, for which the degeneracy issue is minimal. The resulting parameters distribution was sufficiently well approximated by a multivariate Gaussian. The use of a Gaussian prior leads also to Gaussian posterior, convenient for computational optimization.

We tested previously proposed estimation approaches using uniform and population-based priors in experimental datasets with clinically achievable dMRI acquisition protocols. For this, we selected appropriate subsets of measurements from the MGH dataset. This allowed evaluating the accuracy of the estimations from clinically feasible acquisitions, considering the model parameters estimated from the extensive protocol as GT. Observe that, unlike the previous NODDI/NODDIDA studies,<sup>7,10,11,17</sup> the evaluation was carried out against GT values calculated from real experimental data.

First, we explored in two sample voxels the behavior of the estimations for each method. Second, we qualitatively investigated the smoothness and plausibility of the spatial map of estimated parameters in two ROIs from two subjects. Next, we quantified the global accuracy of the estimations by the distribution of the relative errors in the dataset of 35 subjects. Finally, we used an independent dMRI measurement from a clinically comparable scanner to evaluate the estimation accuracies. Consistently across the four experiments, the introduction of the population-based priors, effectively deals with the ill-posedness of the problem and produce substantially more accurate results than previous techniques.

The multimodality of the parameter estimates using a single parameter initialization with the uniform prior (U1), observed in Figure 3, evidences the non-convexity of the problem, as earlier reported in Jelescu et al.<sup>11</sup> The known, more severe, ill-posedness of the problem is evidenced in the instability of the global optimum with minimum residual (Um). In contrast with the observation and suggestion in Novikov et al<sup>10</sup> and Jelescu et al,<sup>11</sup> the selection of the wider optimum by the higher prevalence criterion (Up) does not solve the problem. This is further evidenced in the noisy parameter maps (Figures 4 and 5) obtained from any of the three criteria with uniform prior. Further, the largest errors observed (Figure 6) for the higher prevalence criteria (Up) discourages the use of this criteria. A few more examples are presented in Supporting Information Figures S3-S6.

The non-convexity of the estimation problem is largely reduced by the introduction of the population-based prior, as evidenced in the removing of most of the multimodality in Figure 3 for a single initialization (G1). This is further reflected in the significant error reduction (Figures 6 and 8) for all methods including the population-based prior (G1, Gm, and Gp) but with only slightly larger errors for G1, indicating that the search for the global optimum is much simplified.

Figure 6 points out the impact on the estimation accuracy of the level of noise in the three SNR ranges considered is negligible for NODDI and for the methods with uniform prior. In contrast, the gain in parameter accuracy with the quality of the signal is clear for the methods with the population-based prior. This can be interpreted considering that the existing bimodality in the likelihood function is a persistent obstruction to the accuracy even with infinitesimal noise, but the removal of the bimodality by the population-based prior releases the achievable accuracy from this obstruction.

NODDI provides accuracies (Figure 6) similar to the best method (Gm) for the dispersion parameter,  $\kappa$ , and, to some extent, for the diffusivities,  $D_a$  and  $D_e^{\parallel}$ , especially for measurements with low SNR. However, NODDI presents substantially larger errors for the cellular fraction,  $f$ , and cannot estimate the variations in the diffusivities, since, in contrast with NODDIDA, they are fixed to predefined values (see Figures 4,5 and 7).

The population-based prior in this work was based on measurements from the Connectome scanner which has very high 300 mT/m maximum gradient strength, compared to clinical scanners which have 40-60 mT/m maximum gradient strength. Since the observed diffusion values in dMRI depends partially on the diffusion gradient strengths and other scanning parameters,<sup>26</sup> the presented prior might cause some bias in the estimation of the extra and intracellular diffusion coefficients, when applied to data obtained with other scanners or protocols. The presented experiment on an independent dMRI data acquired with a clinical scanner, demonstrates that the use of the prior still results in the same reduction of the ill-posedness. However, since no ground truth is available in this case, we cannot discard a possible bias.

In the context of disease detection and treatment, changes in NODDIDA parameters ( $\kappa$ ,  $D_e^{\perp}$ ) have been reported<sup>12-14</sup> in multiple sclerosis and Alzheimer's disease, and is known to be potentially useful for several other disorders such as beading, edema, and inflammation.<sup>10</sup> The parameter variations of  $D_a$ ,  $D_e^{\parallel}$ , not accessible by NODDI, are also potentially useful in detection of edema and inflammation. Figure 6, shows how utilizing the conventional uniform prior model or NODDI would lead to large estimation errors and could make estimation of NODDIDA parameter changes unfeasible in clinical settings. Using the population-based prior, the errors are decreased in all parameters, for all SNR ranges. This decrease is particularly marked in  $D_a$ ,  $\kappa$  and  $f$ , with mean errors below 16% in the more clinically relevant parameters ( $D_a$ ,  $D_e^{\perp}$ ,  $\kappa$ ,  $f$ ) for  $\text{SNR} \geq 35$ , indicating the potential of the use of population-based prior in robust NODDIDA parameter estimation and subsequent clinical diagnosis development.

In this work, we have estimated the prior from a population of healthy subjects. For clinical applications investigating any pathology, a more general prior should be estimated from a population including also cases with such pathologies.



We expect this will be feasible in the future, when extensive acquisition protocol-based dMRI data from pathology cases will be available. We plan to construct such general priors in the future, based on pathology cases.

## 5 | CONCLUSIONS

In this work, we introduced a novel framework of NODDIDA estimation, which allowed the use of a population-based prior information in a Bayesian formulation. The prior was estimated from the distribution of NODDIDA parameters obtained from the publicly accessible MGH HCP dataset<sup>19</sup> with dMRI from 35 subjects with an exceptionally extensive acquisition protocol. The parameter distribution was approximated by a multivariate Gaussian distribution defining the prior. This leads also to a Gaussian posterior, whose optimization is easily implemented as a modification of the NODDI toolbox. The code used for estimation can be downloaded from Mozumder<sup>24</sup> and used in conjunction to the NODDI toolbox.

The results indicate that, contrary to previous claims,<sup>10</sup> the use of priors within the Bayesian estimation framework, can lead to accurate and robust estimation of the NODDIDA parameters. The integration of the population-based prior, with the minimum cost criteria to find the global optimum, provides parameter estimation accuracies surpassing the ones from previous methods. The results also suggest that this prior largely removes the ill-posedness, and even partially the non-convexity of the estimation problem. This method provides mean relative errors below 16% for the more clinically relevant parameters ( $D_a, D_e^+, \kappa, f$ ) using clinically feasible SNRs and datasets. The proposed method can be potentially developed for clinical diagnosis of brain disorders using the estimated NODDIDA parameters.

## ACKNOWLEDGMENTS

The work has been supported by the project OCEAN (EP/M006328/1) funded by the Engineering and Physical Sciences Research Council and project VPH-DARE@IT (FP7-ICT-2011-9-601055) funded by the European Union's Seventh Framework Programme. AFF acknowledges support from the Royal Academy of Engineering Chair in Emerging Technologies Scheme (CiET1819\19). Patient dMRI data were comes from the Human Connectome Project, MGH-USC Consortium (Principal Investigators: Bruce R. Rosen, Arthur W. Toga and Van Wedeen; U01MH093765) funded by the NIH Blueprint Initiative for Neuroscience Research grant; the National Institutes of Health grant P41EB015896; and the Instrumentation Grants S10RR023043, 1S10RR023401, 1S10RR019307. The authors thank Paul M. Parizel and Pim Pullens from the University Hospital Antwerp and Quinten Collier, Jelle Veraart, Ben Jeurissen, and Jan Sijbers from iMindsVision Lab of the University of

Antwerp for their help in the setup, acquisition and processing of the multishell dMRI dataset.

## REFERENCES

1. Jones DK, *Diffusion MRI Theory, Methods, and Applications*. Oxford, NY: Oxford University Press; 2011.
2. Basser PJ, Jones DK. Diffusion-tensor MRI: theory, experimental design and data analysis—a technical review. *NMR Biomed*. 2002;15:456–467.
3. Johansen-Berg H, Behrens TE. *Diffusion MRI: From Quantitative Measurement to in Vivo Neuroanatomy*. London, United Kingdom: Academic Press; 2013.
4. Novikov DS, Kiselev VG, Jespersen SN, On modeling. *Magn Reson Med*. 2018;79:3172–3193.
5. Assaf Y, Basser PJ. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage* 2005;27:48–58.
6. Zhang H, Hubbard PL, Parker GJ, Alexander DC. Axon diameter mapping in the presence of orientation dispersion with diffusion MRI. *NeuroImage* 2011;56:1301–1315.
7. 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;61:1000–1016.
8. Hutchinson EB, Avram AV, Irfanoglu MO, et al. Analysis of the effects of noise, DWI sampling, and value of assumed parameters in diffusion MRI models. *Magn Reson Med*. 2017;78:1767–1780.
9. Lampinen B, Szczepankiewicz F, Mårtensson J, van Westen D, Sundgren PC, Nilsson M. Neurite density imaging versus imaging of microscopic anisotropy in diffusion MRI: a model comparison using spherical tensor encoding. *NeuroImage* 2017;147:517–531.
10. Novikov DS, Veraart J, Jelescu IO, Fieremans E. Rotationally-invariant mapping of scalar and orientational metrics of neuronal microstructure with diffusion MRI. *NeuroImage* 2018;174:518–538.
11. Jelescu IO, Veraart J, Fieremans E, Novikov DS. Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue. *NMR Biomed*. 2016;29:33–47.
12. Colgan N, Siow B, O'Callaghan JM, et al. Application of neurite orientation dispersion and density imaging (NODDI) to a tau pathology model of alzheimer's disease. *NeuroImage* 2016;125:739–744.
13. Fieremans E, Jensen JH, Helpert JA, et al. Diffusion distinguishes between axonal loss and demyelination in brain white matter. In 20th Annual Meeting of the International Society for Magnetic Resonance in Medicine. Melbourne, Australia; 2012:465.
14. Jelescu IO, Zurek M, Winters KV, et al. In vivo quantification of demyelination and recovery using compartment-specific diffusion MRI metrics validated by electron microscopy. *NeuroImage* 2016;132:104–114.
15. Unterberg A, Stover J, Kress B, Kiening K. Edema and brain trauma. *Neuroscience* 2004;129:1019–1027.
16. Budde MD, Frank JA. Neurite beading is sufficient to decrease the apparent diffusion coefficient after ischemic stroke. *Proc Nat Acad Sci*. 2010;107:14472–14477.
17. Reisert M, Kellner E, Dhital B, Hennig J, Kiselev VG. Disentangling micro from mesostructure by diffusion MRI: a bayesian approach. *NeuroImage* 2017;147:964–975.
18. Fan Q, Witzel T, Nummenmaa A, et al. MGH-USC human connectome project datasets with ultra-high b-value diffusion MRI. *NeuroImage* 2016;124:1108–1114.



19. Van Essen DC, Smith SM, Barch DM, et al. The WU-Minn human connectome project: an overview. *NeuroImage* 2013;80:62–79.
20. Pasternak O, Sochen N, Gur Y, Intrator N, Assaf Y. Free water elimination and mapping from diffusion MRI. *Magn Reson Med*. 2009;62:717–730.
21. Jelescu IO, Veraart J, Adisetiyo V, Milla SS, Novikov DS, Fieremans E. One diffusion acquisition and different white matter models: how does microstructure change in human early development based on WMTI and NODDI. *NeuroImage* 2015;107:242–256.
22. Polders DL, Leemans A, Hendrikse J, Donahue MJ, Luijten PR, Hoogduin JM. Signal to noise ratio and uncertainty in diffusion tensor imaging at 1.5, 3.0, and 7.0 Tesla. *J Magn Reson Imaging* 2011;33:1456–1463.
23. Gudbjartsson H, Patz S. The rician distribution of noisy MRI data. *Magn Reson Med*. 1995;34:910–914.
24. Mozumder M. NODDIDA Various Estimation Codes, GitHub repository. <https://doi.org/10.6084/m9.figshare.6626792.v2>. Accessed October 21, 2018.
25. Seo H, Chang K-H, Na D, Kwon B, Lee D. High b-value diffusion ( $b = 3000 \text{ s/mm}^2$ ) MR imaging in cerebral gliomas at 3T: visual and quantitative comparisons with  $b = 1000 \text{ s/mm}^2$ . *Am J Neuroradiol*. 2008;29:458–463.
26. Sasaki M, Yamada K, Watanabe Y, et al. Variability in absolute apparent diffusion coefficient values across different platforms may be substantial: a multivendor multi-institutional comparison study. *Radiology* 2008;249:624–630.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**FIGURE S1** Map of NODDIDA parameters estimated from a midbrain slice from subject *MGH\_1001*. Top row show the GT parameters estimated using the whole set of extended dMRI measurements. The rows below show the results from the seven different methods applied on a clinically feasible subset of measurements

**EQUATION S1** The mean  $\theta$  of the NODDIDA model parameters  $\theta = (f, D_a, D_e^{\parallel}, D_e^{\perp}, \kappa)$ , calculated from midbrain slices of the 35 subjects of MGH HCP database

**FIGURE S2** Accuracy of the estimation of the NODDIDA parameters by each of the previously proposed estimation methods (N, U1, Um, Up) and G1, Gm, Gp assuming SNR 50, assuming SNR 20, and using their true (calculated) SNRs. The boxplots represent the distribution of the relative estimation error for each NODDIDA parameter and for each method, with a clinically feasible subset of measurements. We observe a slight improvement in the estimation of  $\kappa$ ,  $D_e^{\parallel}$  and  $D_e^{\perp}$ , using the proposed methods, when SNR 20 or the estimated SNRs were used. This is because, the mean SNR from all the MGH HCP data was around 35, and hence choosing SNR 20 was a closer approximation to the true SNR, than choosing SNR 50

**FIGURE S3** Estimates from a voxel of corona radiata with clinical subset of measurements. The cyan dotted lines represent the estimates using NODDI. The parameter estimates using the full extended protocol (GT) is shown in magenta dotted lines

**FIGURE S4** Estimates from a voxel of corona radiata with a different subset of measurements. The cyan dotted lines represent the estimates using NODDI. The parameter estimates using the full extended protocol (GT) is shown in magenta dotted lines

**FIGURE S5** Estimates from a gray matter voxel with clinical subset of measurements. The cyan dotted lines represent the estimates using NODDI. The parameter estimates using the full extended protocol (GT) is shown in magenta dotted lines

**FIGURE S6** Estimates from a different gray matter voxel with clinical subset of measurements. The cyan dotted lines represent the estimates using NODDI. The parameter estimates using the full extended protocol (GT) is shown in magenta dotted lines

**How to cite this article:** Mozumder M, Pozo JM, Coelho S, Frangi AF. Population-based Bayesian regularization for microstructural diffusion MRI with NODDIDA. *Magn Reson Med*. 2019;82:1553–1565. <https://doi.org/10.1002/mrm.27831>