

# Robust and fast nonlinear optimization of diffusion MRI microstructure models



R.L. Harms<sup>a,b,\*</sup>, F.J. Fritz<sup>a</sup>, A. Tobisch<sup>c</sup>, R. Goebel<sup>a</sup>, A. Roebroeck<sup>a</sup>

<sup>a</sup> Dept. of Cognitive Neuroscience, Faculty of Psychology & Neuroscience, Maastricht University, The Netherlands

<sup>b</sup> Brain Innovation B.V., Maastricht, The Netherlands

<sup>c</sup> German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

## ARTICLE INFO

### Keywords:

Diffusion MRI  
Microstructure  
Biophysical compartment models  
Nonlinear optimization  
Graphics Processing Unit  
Parallel computing

## ABSTRACT

Advances in biophysical multi-compartment modeling for diffusion MRI (dMRI) have gained popularity because of greater specificity than DTI in relating the dMRI signal to underlying cellular microstructure. A large range of these diffusion microstructure models have been developed and each of the popular models comes with its own, often different, optimization algorithm, noise model and initialization strategy to estimate its parameter maps. Since data fit, accuracy and precision is hard to verify, this creates additional challenges to comparability and generalization of results from diffusion microstructure models. In addition, non-linear optimization is computationally expensive leading to very long run times, which can be prohibitive in large group or population studies. In this technical note we investigate the performance of several optimization algorithms and initialization strategies over a few of the most popular diffusion microstructure models, including NODDI and CHARMED. We evaluate whether a single well performing optimization approach exists that could be applied to many models and would equate both run time and fit aspects. All models, algorithms and strategies were implemented on the Graphics Processing Unit (GPU) to remove run time constraints, with which we achieve whole brain dataset fits in seconds to minutes. We then evaluated fit, accuracy, precision and run time for different models of differing complexity against three common optimization algorithms and three parameter initialization strategies. Variability of the achieved quality of fit in actual data was evaluated on ten subjects of each of two population studies with a different acquisition protocol. We find that optimization algorithms and multi-step optimization approaches have a considerable influence on performance and stability over subjects and over acquisition protocols. The gradient-free Powell conjugate-direction algorithm was found to outperform other common algorithms in terms of run time, fit, accuracy and precision. Parameter initialization approaches were found to be relevant especially for more complex models, such as those involving several fiber orientations per voxel. For these, a fitting cascade initializing or fixing parameter values in a later optimization step from simpler models in an earlier optimization step further improved run time, fit, accuracy and precision compared to a single step fit. This establishes and makes available standards by which robust fit and accuracy can be achieved in shorter run times. This is especially relevant for the use of diffusion microstructure modeling in large group or population studies and in combining microstructure parameter maps with tractography results.

## Introduction

Diffusion MRI (dMRI) is a tool for investigating the microstructure of biological tissue by probing the self-diffusion of water (Bihan et al. 1986). The conventional method for the analysis of white matter in dMRI imaging is the tensor model in Diffusion Tensor Imaging (DTI; Bassar et al. 1994). DTI has shown to be sensitive to microstructural changes due to, for example, development (Pfefferbaum et al. 2000; Neil et al. 2002; Assaf and Pasternak 2008; Lebel et al. 2010) and pathology (Werring et al. 1999; Horsfield and Jones 2002; Sotak 2002;

Sundgren et al. 2004). Although sensitive, DTI indices such as Fractional Anisotropy (FA) are also unspecific, since differences in FA can reflect different axonal properties such as axon density, diameter distribution and myelination (Beaulieu 2002; Assaf et al. 2004; Assaf and Pasternak 2008; Jones et al. 2013; Santis et al. 2014).

Recently, advances in biophysical multi-compartment modeling have gained popularity because they possess greater specificity than DTI in relating the dMRI signal to the underlying cellular microstructure (Assaf et al., 2004; Assaf and Bassar, 2005; Assaf et al., 2008; Alexander et al., 2010; Panagiotaki et al., 2012; Zhang et al., 2012;

\* Corresponding author at: Dept. of Cognitive Neuroscience, Faculty of Psychology & Neuroscience, Maastricht University, The Netherlands.  
E-mail address: [Robbert.harms@maastrichtuniversity.nl](mailto:Robbert.harms@maastrichtuniversity.nl) (R.L. Harms).

Assaf et al., 2013; Fieremans et al., 2013; De Santis et al., 2014; Santis et al., 2014; Jelescu et al., 2015). In these diffusion microstructure models the diffusion weighted signal is expressed as a combination of one or more biophysically inspired compartments. Although the compartment models are based on simple geometric shapes, and a strong assumption of no inter-compartment exchange is made (cf; Nilsson et al. 2013; Li et al. 2016), these models can provide specific measures such as fiber density, orientation dispersion, and axonal diameter distributions and have been shown to be sensitive to specific white matter alterations due to development (e.g. Kunz et al., 2014; Jelescu et al., 2015) and pathology (Bergers et al., 2002; Fieremans et al., 2013; Benitez et al., 2014; Timmers et al., 2015; Wen et al., 2015; Kamagata et al., 2016).

There is a large range of diffusion microstructure models, including popular models such as Neurite Orientation Dispersion and Density Imaging model (NODDI; Zhang et al., 2012), the Combined Hindered And Restricted Model of Diffusion (CHARMED; Assaf and Basser, 2005), the White Matter Tract Integrity model (WMTI; Fieremans et al., 2013), AxCaliber (Assaf et al., 2008), the Minimal Model of White Matter Diffusion (MMWMD in ActiveAx; Alexander et al., 2010; Zhang et al. 2011), including further developments of these models like NODDIDA (Jelescu et al., 2015), Bingham NODDI (Tariq et al., 2016), diffusion time dependent CHARMED (De Santis et al. 2016) and fiber-specific T<sub>1</sub> CHARMED (De Santis et al., 2016).

Each of these models needs to be fitted to the dMRI data to estimate parameter maps, which is commonly accomplished using non-linear optimization. This has two major challenges. First, it is computationally expensive, quickly leading to very long run times. Second, the quality, accuracy and precision of the data fit is often uncertain. To tackle these challenges, each of the popular models comes with its own, mostly different, optimization algorithm, noise model and initialization strategy to estimate its parameter maps. This creates challenges to comparability and generalization of results from diffusion microstructure models, additional to the model formulation itself. In this technical note we investigate the performance of several optimization algorithms and initialization strategies over a few of the more popular diffusion microstructure models. We investigate whether a single well performing approach exists that could be applied to many models and that equates both run time and fit aspects. To this end, we evaluate the fit, accuracy, precision and run time for different models of differing complexity against three common optimization algorithms and three parameter initialization strategies. All models, algorithms and strategies were implemented on the GPU to remove run time constraints.

Variability of the achieved quality of fit in actual data is evaluated on ten subjects of each of two large group studies with a different acquisition protocol, the MGH-USC part of the Human Connectome Project (Fan et al., 2016) and the Rhineland Study (www.rheinland-studie.de).

## Methods

All biophysical compartment models, noise models / likelihood functions, as well as optimization algorithms were implemented in a python based GPU accelerated toolbox (Maastricht Diffusion Toolbox or MDT, freely available under an open source L-GPL license at <https://github.com/cbclab/MDT>). Its object oriented modular design allows arbitrary combinations of individual single compartment models into composite compartment models, which can then be combined with a chosen likelihood function and optimization algorithm by Python scripting. The complete multi-compartment model, likelihood function and optimization algorithm is automatically compiled into OpenCL code executable on both CPUs and GPUs (or combinations thereof).

### Single compartment models

Table 1 defines the individual compartment models which are combined to construct the multi-compartment models (c.f. Panagiotaki et al., 2012; Ferizi et al., 2013) valid for (singly refocused) Pulsed Gradient Spin Echo (PGSE; Stepisnik, 1993) acquisitions.

The Tensor model was first described in (Basser et al., 1994), the Ball and Stick models are defined in (Behrens et al., 2003), NODDI<sub>in</sub> and NODDI<sub>ex</sub> are respectively the intra cellular and extra cellular models in (Zhang et al., 2012) and the CHARMED<sub>in</sub> compartment is the restricted compartment defined in (Assaf et al., 2004).

Each compartment has a signal function modelling the signal **S** depending on the unit norm gradient direction vector **g** with scalar  $b = (\Delta - \delta/3)(\gamma\delta G)^2$  and the vector  $\mathbf{q} = \gamma\delta\mathbf{g}G/2\pi$ . Here  $G$  is the gradient amplitude,  $\Delta$  the time between the start of the two gradient pulses,  $\delta$  the duration of the gradient pulse and  $\gamma$  the gyromagnetic ratio for <sup>1</sup>H in  $\text{rad}\cdot\text{s}^{-1}\cdot\text{T}^{-1}$ . Additionally, the CHARMED<sub>in</sub> compartments depend on the echo time  $TE$ .

Disregarding fixing of parameter values discussed below, the modelled fiber orientations **n** are optimized as spherical coordinates using the free parameters  $\theta$  and  $\phi$  with scalar  $d$  for the diffusivity. Some models have more than one diffusivity, these are all optimized

**Table 1**

The single compartment models, see Table 2 for an overview of the optimizable parameters. The primary direction of diffusivity **n**, is parameterized using polar coordinates with angles  $\theta$ ,  $\phi$  and radius  $d$ . The variables  $b$ ,  $\mathbf{q}$ ,  $\Delta$ ,  $\delta$ ,  $G$  and  $TE$  are sequence settings. In the Tensor compartment, the function  $\text{rotate}(\mathbf{n}, \psi)$  rotates the Tensor around **n** by the angle  $\psi$ . In the NODDI models, the function  $f(\mathbf{n}, \kappa)d\mathbf{n}$  gives the probability of finding fiber bundles along orientation **n** using a Watson distribution with parameter  $\kappa$  integrated over the unit sphere  $\mathbb{S}^2$ . In the NODDI<sub>ex</sub> model, the diffusion tensor  $D(\mathbf{n})$  is defined as a cylindrically symmetric Tensor (like the Tensor in this table except for the symmetry). In the CHARMED<sub>in</sub> compartment  $N$  is the number of gamma cylinders used,  $v_i$  is the weight per gamma distributed cylinders and  $R_i$  is the radius per cylinder. In previous work  $|\mathbf{q}|^2(\mathbf{n}\cdot\mathbf{g})^2$  is sometimes denoted as  $|\mathbf{q}_\parallel|^2$  and  $|\mathbf{q}|^2(1-(\mathbf{n}\cdot\mathbf{g})^2)$  as  $|\mathbf{q}_\perp|^2$ , we inlined these identities here in the CHARMED<sub>in</sub> equation.

Compartment	Signal function	Compartment model parameters
Tensor	$\mathbf{S} = e^{-b(d_\parallel(\mathbf{n}\cdot\mathbf{g})^2 + d_{\perp 1}(\mathbf{n}_{\perp 1}\cdot\mathbf{g})^2 + d_{\perp 2}(\mathbf{n}_{\perp 2}\cdot\mathbf{g})^2)}$ $\mathbf{n}_{\perp 1} = \text{rotate}(\mathbf{n}, \psi)$ $\mathbf{n}_{\perp 2} = \mathbf{n} \times \mathbf{n}_{\perp 1}$	$d_\parallel, d_{\perp 1}, d_{\perp 2}, \theta, \phi, \psi$
Ball	$\mathbf{S} = e^{-bd}$	$d$
Stick	$\mathbf{S} = e^{-bd(\mathbf{n}\cdot\mathbf{g})^2}$	$d, \theta, \phi$
NODDI <sub>in</sub>	$\mathbf{S} = \int_{\mathbb{S}^2} f(\mathbf{n}, \kappa) e^{-bd(\mathbf{n}\cdot\mathbf{g})^2} d\mathbf{n}$	$d, \theta, \phi, \kappa$
NODDI <sub>ex</sub>	$\mathbf{S} = e^{-b\mathbf{g}^T \left( \int_{\mathbb{S}^2} f(\mathbf{n}, \kappa) D(\mathbf{n}) d\mathbf{n} \right) \mathbf{g}}$	$d_\parallel, d_{\perp 1}, \theta, \phi, \kappa$
CHARMED <sub>in</sub>	$\mathbf{S} = \sum_{i=1}^N v_i [\mathbf{S}_\parallel(\mathbf{q}, \Delta) \cdot \mathbf{S}_{\perp i}(\mathbf{q}, TE)]$ $\mathbf{S}_\parallel(\mathbf{q}, \Delta) = e^{-4\pi^2  \mathbf{q} ^2 (\mathbf{n}\cdot\mathbf{g})^2 (\Delta - \delta/3)d}$ $\mathbf{S}_{\perp i}(\mathbf{q}, TE) = e^{-\left(4\pi^2  \mathbf{q} ^2 (1 - (\mathbf{n}\cdot\mathbf{g})^2) R_i^4 \left(\frac{d\cdot TE}{2}\right)\right) \left(\frac{7}{96}\right) \left(2 - \left(\frac{99}{112}\right) R_i^2 \left(\frac{d\cdot TE}{2}\right)\right)}$	$d, \theta, \phi$

**Table 2**

The parameter descriptions corresponding to Table 1.

Parameter	Compartments	Usage
$d_{  }$ (or $d$ )	Tensor, Ball, Stick, NODDI_in, NODDI_ex, CHARMED_in	Parallel diffusivity along the primary direction of diffusion $\mathbf{n}$
$d_{\perp 1}$	Tensor	Perpendicular diffusivity, perpendicular to both $d_{  }$ and $d_{\perp 2}$
$d_{\perp 2}$	Tensor	Perpendicular diffusivity, perpendicular to both $d_{  }$ and $d_{\perp 1}$
$\theta$	Tensor, Stick, NODDI_in, NODDI_ex, CHARMED_in	Polar angle used to parameterize $\mathbf{n}$ , the primary direction of diffusion
$\phi$	Tensor, Stick, NODDI_in, NODDI_ex, CHARMED_in	Azimuth angle used to parameterize $\mathbf{n}$ , the primary direction of diffusion
$\psi$	Tensor	Used to rotate the Tensor around its primary axis
$\kappa$	NODDI_in, NODDI_ex	The dispersion index of the Watson distribution

separately unless specified otherwise in the multi-compartment model. In the Tensor compartment the rotate function first rotates the fiber directions  $\mathbf{n}$  by 90 degrees in the (x, z) plane to make it perpendicular to the principal direction. The Tensor is then rotated around  $\mathbf{n}$  by the angle  $\psi$  resulting in  $\mathbf{n}_{11}$ , and, after a cross product  $\mathbf{n}_{12}$ . In the NODDI\_in and NODDI\_ex models, the function  $f(\mathbf{n}, \kappa)d\mathbf{n}$  gives the probability of finding fiber bundles along orientation  $\mathbf{n}$  using a Watson distribution integrated over the unit sphere  $\mathbb{S}^2$ . In the NODDI\_ex model, the diffusion tensor  $D(\mathbf{n})$  is defined as a cylindrically symmetric tensor. For more details on NODDI see (Zhang et al., 2012). The CHARMED\_in model is a Gamma Distributed Radii (GDR) Cylinder model (Panagiotaki et al., 2012) where the cylinder distribution is predefined with N cylinders (we take N=6 throughout). Our 6 cylinders have radii R of [1.5, 2.5, 3.5, 4.5, 5.5, 6.5]  $\times 10^{-6}$  meters and corresponding weights [0.0212, 0.1072, 0.1944, 0.2667, 0.2150, 0.1956] taken from a Gamma distribution derived from histological results (Aboitiz et al., 1992).

### Composite multi-compartment models

The general multi-compartment diffusion microstructure model has the form of a weighted sum of single compartments:

$$S = S_0 \sum_{i=0}^n w_i S_i$$

Where  $S_0$  is the signal for the non-diffusion weighted (or  $b_0$ ) acquisitions,  $w_i$  the volume fractions (signal weights, signal fractions or water fractions) and  $S_i$  is the signal function for the  $i$ 'th of  $n$  total compartments. For this work we selected the Tensor, Ball & Sticks, NODDI and CHARMED models. Table 3 shows these multi-compartment models (henceforth simply 'models'), their constituent compartments and total number of parameters including estimation of  $S_0$ .

Here we take the Stick compartment to be an idealized cylinder with zero radius and classify it as a restricted compartment (cf. Ferizi et al., 2014). We have added to some of the models the postfix '\_in[n]' which is used to identify the number of restricted compartments employed in a model for those models that allow multiple restricted compartments. For example, CHARMED\_in2 indicates a CHARMED model with 2 restricted compartments (and the regular single hindered compartment) for each of two unique fiber orientations in a voxel. The total number of parameters per model is the sum of the parameters per

compartment, the volume fractions (equal to the number of compartments minus one) and the scalar for  $S_0$ . The complete signal equation per model is then given as a volume fraction weighted sum of compartments, scaled by the non-diffusion weighted signal  $S_0$ .

### Evaluation function and likelihood

Models are optimized by finding the set of free parameter values  $\mathbf{x} \in \mathbb{R}^n$  that minimize the evaluation function or objective function of the modeling errors ( $\mathbf{O} - \mathbf{S}(\mathbf{x})$ ) with  $\mathbf{O}$  the observed data and  $\mathbf{S}(\mathbf{x})$  the model signal estimate. In general, the evaluation function is formulated as the negative log likelihood function which embeds a noise model for the data. Whereas the noise in the complex valued Fourier coefficient images can be described by a Gaussian distribution, the noise in the magnitude reconstructed MR data follows a non-zero mean (Rician or non-central  $\chi$ ) distribution (Gudbjartsson and Patz, 1995). For SNR > 2 an Offset Gaussian model can be used (Alexander 2009), with the advantage that it is potentially more stable than the Rician model (Panagiotaki et al., 2012). Therefore, we use the Offset Gaussian likelihood model for all optimizations. Minimizing the negative log likelihood leads to the following objective function (with constant terms dropped):

$$\bar{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \left( \frac{\sum (\mathbf{O} - \sqrt{(\mathbf{S}(\mathbf{x})^2 + \sigma^2)})^2}{2\sigma^2} \right)$$

Here we use  $\sigma$  as the standard deviation of the Gaussian distributed error of the complex valued Fourier images. We estimated this standard deviation  $\sigma$  from the reconstructed magnitude images using the  $\sigma_{\text{multi}}$  method in Dietrich et al. (2007; eq. A6).

### Parameter dependencies and global fixes

Some of the models have specific dependencies between their parameters, as well as parameter fixed to a specific value. Defining the dependencies at the level of the multi-compartment model has two advantages. First, this allows the compartments to be defined as general as possible and adapt them for the use in several multi-compartment models. Second, some dependencies relate the parameters of different compartments to each other and therefore must operate at the level of the multi-compartment model. The list of global fixes and dependencies is given by Table 4 where we use the 'dot'

**Table 3**

The used composite multi-compartment models, their compartments (divided into intra-, extra-axonal and isotropic) and total number of parameters.

Model	Restricted (intra-cellular) compartments	Hindered (extra-cellular) compartments	Isotropic compartments	Number of parameters	Acquisition requirements
Tensor	–	Tensor	–	7	$b < 1.5 \times 10^6 \text{ s/m}^2$
Ball & Sticks_in[n]	Stick (n times)	–	Ball	$1+3n$	–
NODDI	NODDI_in	NODDI_ex	Ball	6	$\geq 2$ b-values/shells
CHARMED_in[n]	CHARMED_in (n times)	Tensor	–	$7+4n$	$\geq 2$ b-values/shells $b_{\text{max}} \geq 4.0 \times 10^6 \text{ s/m}^2$

**Table 4**  
Parameter dependencies and global fixes.

Model	Parameter global fixes	Parameter dependencies
Tensor	–	–
Ball & Sticks_in[n]	Ball.d = $3 \times 10^{-9}$ m <sup>2</sup> /s Stick.d = $1.7 \times 10^{-9}$ m <sup>2</sup> /s	$\sum w_i = 1$
NODDI	NODDI_in.d = $1.7 \times 10^{-9}$ m <sup>2</sup> /s NODDI_ex.d = $3 \times 10^{-9}$ m <sup>2</sup> /s Ball.d = $3 \times 10^{-9}$ m <sup>2</sup> /s	1) $\sum w_i = 1$ 2) $\text{NODDI\_ex.d}_{\perp} = \text{NODDI\_ex.d}_{\parallel} \cdot \frac{w_{\text{ex}}}{(w_{\text{in}} + w_{\text{ex}})}$ 3) $\text{NODDI\_ex}\{\kappa, \theta, \phi\} = \text{NODDI\_in}\{\kappa, \theta, \phi\}$
CHARMED_in[n]	–	$\sum w_i = 1$

notation (e.g. Ball.d for the d parameter of the Ball compartment) to explicitly assign parameters to certain compartments.

In the NODDI model in Table 4, the second dependency defines the NODDI tortuosity assumption and the third dependency locks the orientations and dispersion of the NODDI\_in and NODDI\_ex model to each other (the dispersion because the NODDI model adjusts the ratio of parallel and perpendicular extra-axonal diffusivity on the basis of orientation dispersion of the intracellular compartment). Additionally, for all multi-compartment models, the weights (volume fractions) must sum to one. This is accomplished by normalizing the set of  $n-1$  volume fractions in each optimization iteration by the sum  $s = \sum_{i=1}^{n-1} w_i$  if that sum is larger than one. The last compartment weight, not explicitly optimized, is either set to zero, i.e.  $w_n=0$  or set as  $w_n=1-s$  if  $s$  is smaller than zero, for signal evaluation in the optimization process.

#### Indices calculated in post-processing

For every model various additional volumetric maps are calculated from the optimized model parameters. These can be model specific and model independent. Table 5 highlights the model specific maps.

FS (Fraction of Sticks) is the total fraction (sum of weights) of the  $n$  Stick compartments in the Ball & Sticks model. NDI (Neurite Density Index) and ODI (Orientation Dispersion Index) are defined in (Zhang et al., 2011) and FR (Fraction of Restricted) is the total fraction of the  $n$  restricted compartments in the CHARMED model (Assaf et al., 2004; De Santis et al., 2012).

In addition to these model specific maps, we add two model independent maps to the result set of every model. The first is the Log Likelihood (LL) map containing the complete (log) likelihood of the model given the data, as given by:

$$LL = -\frac{\sum (\mathbf{O} - \sqrt{(\mathbf{S}(\mathbf{x})^2 + \sigma^2)})^2}{2 \cdot \sigma^2} - m \cdot \log(\sigma \sqrt{2\pi})$$

With  $\mathbf{O}$  the observation,  $\mathbf{S}(\mathbf{x})$  the function value given the optimal set of parameters,  $\sigma$  the noise standard deviation and  $m$  the number of volumes in the dataset. This LL map gives an indication of the goodness of fit and a higher LL is preferred. The second model independent map we add is the Bayesian Information Criterion (BIC) map. The BIC (Schwarz, 1978) can be used to compare models with different numbers of parameters on the same dataset (Panagiotaki et al., 2012; Ferizi et al., 2014). It is defined as  $\text{BIC} = -2 \cdot LL + k \cdot \ln(m)$  where  $LL$  is the log likelihood value,  $k$  is the number of free parameters and  $m$

is the number of observations. The model with the lower BIC is preferred and features the best trade-off between model-fit and complexity (number of parameters).

#### Optimization algorithms

We explicitly separate the models from the optimization algorithms to enable evaluating model/optimization algorithm combinations. We choose three commonly used optimization algorithms, two generations of gradient free algorithms (Nelder-Mead Simplex and Powell's conjugate-direction) and one gradient-based algorithm (Levenberg-Marquardt). All optimization algorithms in this study were originally programmed in C and converted to OpenCL for the purpose of this work. OpenCL is a cross-vendor language standard and framework that enables highly parallelized calculations on the graphics card (graphical processing unit, GPU), as well as central processor unit (CPU), potentially allowing for large computational acceleration (c.f. Hernandez et al., 2013).

The first routine we used is the Nelder-Mead simplex (NMSimplex) method (Nelder and Mead, 1965). This is a general derivative-free nonlinear local optimization algorithm, also used in the MATLAB fminsearch function. The NMSimplex routine starts with an  $n$ -dimensional simplex with at each vertex initial values for all  $n$  model parameters. During optimization, the values of the evaluation function at each vertex of the simplex are used to determine the worst corner point. The algorithm then tries to replace this worst point by transforming the simplex about its centroid using reflection, expansion, contraction and shrinking (Rios and Sahinidis, 2012). This algorithm has five coefficients that determine its behavior: the scale of the initial simplex ( $s$ ), and the coefficients for the reflection ( $\alpha$ ), contraction ( $\beta$ ), expansion ( $\gamma$ ), and shrink ( $\delta$ ) of the simplex during each iteration. Preliminary tests using the default coefficients by (Nelder and Mead, 1965), the table of coefficients in (Wang and Shoup, 2011) and the adaptive coefficients in (Gao and Han 2012) led us to use the adaptive coefficients in the latter, being:  $s = 1$ ,  $\alpha = 1$ ,  $\beta = 0.75 - 1.0/(2 \cdot k)$ ,  $\gamma = 1 + 2.0/k$  and  $\delta = 1 - 1.0/k$  with  $k$  being the number of free parameters in the model. The original implementation followed for translation into OpenCL was programmed in C by Michael F. Hutt (<http://www.mikehutt.com/neldermead.html>).

The second routine studied is the Powell conjugate-direction method (Powell, 1964). This is a general derivative-free nonlinear local optimization algorithm. It starts the search with  $n$  search vectors each containing all  $n$  model parameters. In every iteration, the algorithm tries to update each search vector in turn using a line search across each dimension. Doing so, it will continuously create new vector combinations (using line searches) and replace old combinations until a stop criteria is met. The final result is the search vector with the minimum function value. We used an OpenCL implementation translated from the C implementation described in Numerical Recipes 3<sup>rd</sup> edition (Press et al., 2007) while using the Brent algorithm (Brent, 1973) for the line search. We used the default settings (from the literature) with a bracket gold ratio of 1.618034 (default ratio by which

**Table 5**  
Additional volumetric maps per model.

Model	Additional maps
Tensor	–
Ball & Sticks_in[n]	FS = $1 - w_{\text{ball}}$
NODDI	1) $\text{NDI} = \frac{w_{\text{in}}}{w_{\text{in}} + w_{\text{ex}}}$ 2) $\text{ODI} = \arctan 2(1, \kappa \cdot 10) \cdot \frac{2}{\pi}$ FR = $1 - w_{\text{in}}$
CHARMED_in[n]	–



successive intervals are magnified in bracketing) and a g-limit of 100 (the maximum magnification allowed for a parabolic-fit step in function bracketing).

The third routine we studied is the Levenberg-Marquardt (LM) method (Levenberg, 1944). This method is a gradient based least squares optimization routine for which the gradients can either be given analytically or approximated numerically. We used numerical gradients for all models, as this is most commonly applied (Assaf et al., 2008; Alexander et al., 2010; De Santis et al., 2012; Zhang, et al. 2012). Our implementation is an OpenCL version of the lmfit (<http://apps.jens.fz-juelich.de/lmfit>) C implementation of this algorithm.

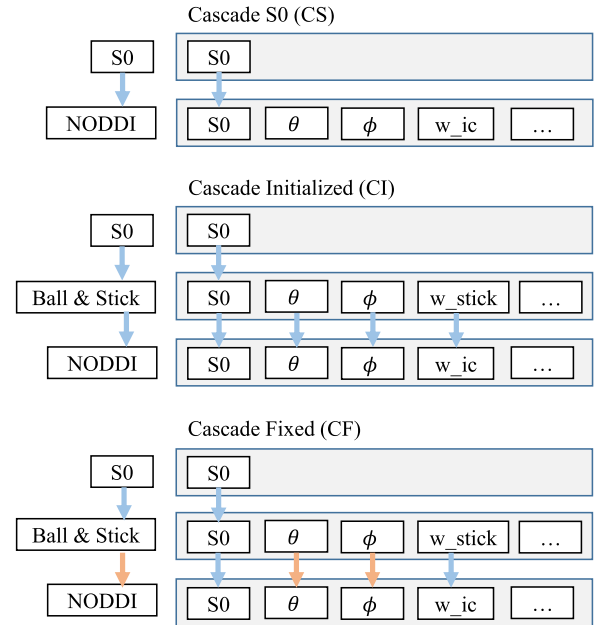
All three algorithms above have two main stop criteria. Either the difference between two consecutive iterations is below a certain threshold, or the predefined maximum number of iterations is reached. For the first we set the break threshold at thirty times the machine epsilon. For the second we set the so-called patience ( $i = p * (1 + k)$  where  $i$  is the number of iterations,  $p$  is the patience and  $k$  is the number of free parameters in the model) to a predefined setting per optimization algorithm. This construction allows fixing the patience while ensuring that models with more parameters are allowed more iterations. We did a preliminary study to determine per algorithm and model the optimal patience given goodness of fit and run time. We used, unless stated otherwise, for NMSimplex a patience of 200, for Powell a patience of 2 and for LM a patience of 100, which gave the optimal trade-off between run time and fit (objective function minimum) for each algorithm. All calculations were performed on a single 2015 AMD Fury X graphics card using single float (32-bit) precision, except for parts where value range (summation of log likelihoods, Euclidian norm in the LM algorithms) or precision (the Watson function in NODDI) are critical, which were evaluated with double precision.

#### Parameter transformations

During model fitting with the NMSimplex, Powell and LM method we use parameter transformations to limit the range of each parameter to biophysical meaningful values and to scale the parameters to a range better suited for optimization after Panagiotaki (2012). These transformations transform the parameters between two sets of parameters spaces: the model space and the optimization space, both in  $\mathbb{R}^n$ . The model space is the space in which the parameters can be used as input to the model. The optimization space is the set of parameters the optimization routine tries to optimize. We define the encoding transformation as transforming the parameters from model to optimization space and the decoding transformation from optimization to model space. Please see Appendix A for a complete list of parameter transformations used in this work.

#### Cascaded model optimization

A common strategy to optimize complex (i.e. high number of parameter) models is initializing subsets of parameters with results from earlier simpler model optimizations (Panagiotaki et al., 2012; Zhang et al., 2012; De Santis et al., 2016). We formalize this here as cascaded model optimization. We test three cascading strategies: Cascade  $S_0$  (CS), Cascade Initialized (CI) and Cascade Fixed (CF), see Fig. 1. In the CS cascade we first fit a one parameter  $S_0$  model to the non-diffusion weighted volumes to determine the scale of the non-diffusion weighted signal. We use this to initialize the  $S_0$  parameter of the model we are interested in. In the CI cascades we initialize a few specific parameters like the volume fractions and the fiber orientation parameters from the results of a simpler model. The CF cascades are similar to the CI cascades except that we now fix some of the parameters, reducing the dimensionality of the last optimization, instead of initializing them. For a complete overview of initialized and fixed parameters for all three cascade variants see Appendix B. In this work, we always use at least CS for all model fitting, as an alternative to dividing away the non-diffusion weighted signal or to



**Fig. 1.** Illustration of the three different cascading strategies (for the example of the NODDI model): CS, CI and CF. The blue arrows indicate initialization of a parameter, the orange arrows indicate fixing a parameter.

fixing it beforehand. Since the range of the  $S_0$  parameter is quite large and arbitrary, initializing with an  $S_0$  estimate improves the performance and run time of the optimization routines.

#### Datasets

For this study we used two groups of ten subjects coming from two studies. The first ten subjects are from the freely available fully preprocessed dMRI data from the USC-Harvard consortium of the Human Connectome project. Data used in the preparation of this work were obtained from the MGH-USC Human Connectome Project (HCP) database (<https://ida.loni.usc.edu/login.jsp>). The HCP project (Principal Investigators: Bruce Rosen, M.D., Ph.D., Martinos Center at Massachusetts General Hospital; Arthur W. Toga, Ph.D., University of California, Los Angeles, Van J. Weeden, MD, Martinos Center at Massachusetts General Hospital) is supported by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH) and the National Institute of Neurological Disorders and Stroke (NINDS). Collectively, the HCP is the result of efforts of co-investigators from the University of California, Los Angeles, Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH), Washington University, and the University of Minnesota (<http://www.humanconnectome.org/documentation/MGH-diffusion/>).

The data were acquired on a specialized Siemens Magnetom Connectom with 300mT/m gradient set (Siemens, Erlangen, Germany). These datasets were acquired at a resolution of 1.5mm isotropic with  $\Delta=21.8\text{ms}$ ,  $\delta=12.9\text{ms}$ ,  $TE=57\text{ms}$ ,  $TR=8800\text{ms}$ , Partial Fourier = 6/8, MB factor 1 (i.e. no simultaneous multi-slice), in-plane GRAPPA acceleration factor 3, with 4 shells of  $b=1000, 3000, 5000, 10,000, \text{s/mm}^2$ , with respectively 64, 64, 128, 393 directions to which are added 40  $b_0$  volumes leading to 552 volumes in total per subject, with an acquisition time of 89 minutes. We refer to these datasets as HCP MGH – 1.5mm – 552vol – b10k and to the multi-shell direction table as the HCP MGH table. These four-shell, high number of directions, and very high maximum b-value datasets allow a wide range of models to be fitted. The second set of ten subjects comes from the diffusion protocol pilot phase of the Rhineland Study ([www.rheinland-studie.de](http://www.rheinland-studie.de)) and was acquired on a Siemens Magnetom Prisma (Siemens, Erlangen, Germany) with the Center for Magnetic Resonance Research (CMRR) multi-band (MB) diffusion sequence

(Moeller et al., 2010; Xu et al., 2013). These datasets had a resolution of 2.0mm isotropic with  $\Delta=45.8\text{ms}$ ,  $\delta=16.3\text{ms}$  and  $\text{TE}=90\text{ms}$ ,  $\text{TR}=4500\text{ms}$  Partial Fourier = 6/8, MB factor 3, no in-plane acceleration with 3 shells of  $b=1000, 2000, 3000\text{ s/mm}^2$ , with respectively 30, 40 and 50 directions to which are added 14  $b_0$  volumes leading to 134 volumes in total per subject, with an acquisition time of 10 min 21 sec. Additional  $b_0$  volumes were acquired with a reversed phase encoding direction which were used to correct susceptibility related distortion (in addition to bulk subject motion) with the topup and eddy tools in FSL version 5.0.9. We refer to these datasets as RLS – 2mm – 134dir – b3k and to the multi-shell direction table as the RLS table. These three-shell datasets represent a relatively short time acquisition protocol that still allows many models to be fitted.

Since the Tensor model is only valid for  $b$ -values up to about  $1200\text{ s/mm}^2$ , it is estimated on only the  $b$ -value  $1000\text{ s/mm}^2$  shell and  $b_0$  volumes. All other models are estimated on all data volumes. For all datasets we created a whole brain mask, using BET from FSL (Smith, 2002), and a white matter (WM) mask. The whole brain mask is used during the model fitting, whereas averages over the WM mask are used in model or data comparisons. The WM mask was calculated by applying a threshold of 0.3 on the Tensor FA results, followed by a double pass 3D median filter of radius 2 in all directions. The Tensor estimate for this mask generation was calculated using a CI Ball Stick/Tensor cascade optimized with the Powell method.

To compare performance and robustness of different optimization algorithms and cascading strategies the per-subject LL averages over the WM mask were calculated. The mean differences between algorithms or strategies over ten subjects and their standard error of the mean (SEM) were then reported.

#### Ground truth simulations

We performed two sets of ground truth simulations (see Fig. 2) to quantify accuracy and precision of: 1) the optimization (Fig. 2A) and 2) the cascading results (Fig. 2B). First, to assess the accuracy and precision of the different optimization algorithms (Fig. 2A), we created ten thousand random fiber orientations for each of sixty linearly spaced volume fractions in  $[0.2, 0.8]$ , for the Ball & Sticks\_in1, CHARMED\_in1 and NODDI models using both a HCP MGH and a RLS multi-shell direction table. For the generated signals we created three copies with a Rician distributed noise

realization at SNRs 10, 20 and 40. We compute optimization error as optimized parameter value minus ground truth parameter value for the intra-axonal volume fraction, i.e. fraction of stick (FS) for Ball & Sticks\_in1, fraction of restricted (FR) for CHARMED\_in1 and fraction of restricted FR for NODDI. We compute a measure of accuracy as the inverse of the mean optimization error over ten thousand random fiber orientations per volume fraction (each with a unique noise realization). Similarly, as a measure of precision we compute inverse of the standard deviation of the optimization error over ten thousand random fiber orientations per volume fraction. Finally, we aggregated these results over the sixty volume fractions into a mean and standard error for both accuracy and precision.

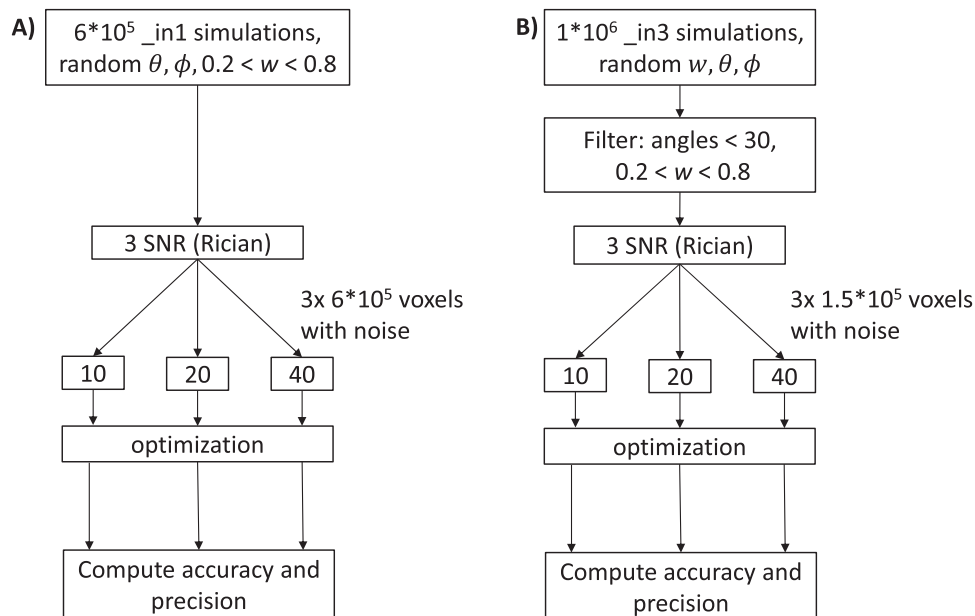
Second, to assess the accuracy and precision of the different cascading strategies (see Fig. 2B), we created for the CHARMED\_in3 model  $1\text{e}6$  instances with randomly distributed fiber orientations and volume fractions, using both a HCP MGH and a RLS multi-shell direction table. Instead of generating consistent instances which meet inter-fiber angle and volume fraction constraints, we chose to generate many random instances and filter for the constraints. Therefore, after generating  $1\text{e}6$  instances we filtered out: i) instances with any angle between the three orientations smaller than 30 degrees, and ii) individual volume fractions outside the range  $[0.2, 0.8]$ , leaving  $1.5\text{e}5$  instances. For the generated signals we created three copies with a Rician distributed noise realization at SNRs 10, 20 and 40. After optimization, we calculated for largest, middle and smallest volume fractions the accuracy and precision as above. We aggregated these results over the three volume fractions within the model into a mean and standard error for both accuracy and precision.

## Results

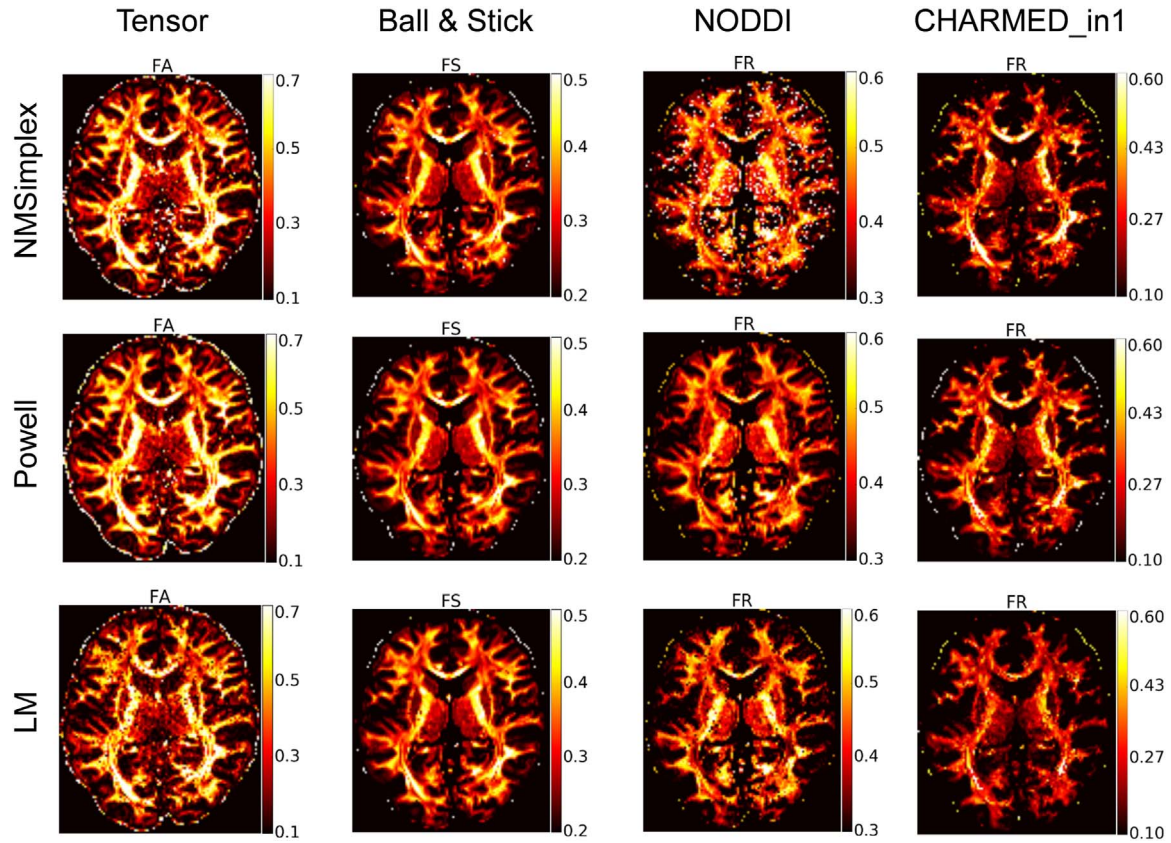
We focus first on the effect of optimization algorithms when applied to the entire diffusion model estimation problem. Subsequently, we focus on initialization strategies by cascaded optimization, dividing the estimation problem into sequential lower dimensional problems.

#### Effect of optimization algorithm

Fig. 3 shows estimated parameter maps for a single HCP MGH subject (HCP MGH – 1.5mm –552vol – b10k, subject 1003) for four different models (Tensor, Ball & Sticks\_in1, NODDI, CHARMED\_in1) each estimated with three optimization algorithms (NMSimplex,



**Fig. 2.** Illustration of the ground truth simulation workflow to quantify the accuracy and precision of A) the optimization routine comparisons, B) the cascading results.



**Fig. 3.** Models (in columns) against optimization algorithms (in rows) for a single HCP MGH dataset. FA: Fractional Anisotropy; FS: Fraction of stick compartment; FR: fraction of restricted (fraction of intra-cellular/restricted compartment). Each model parameter map (column) is individually scaled.

Powell, LM). Note that the Tensor model is mainly shown for comparison and is only estimated on the b-value 1000 s/mm<sup>2</sup> shell. Table 6 shows the corresponding Log Likelihoods (LL), Bayesian Information Criteria (BIC) and run times (RT) for these results, along with the results of a single RLS – 2mm – 134dir – b3k subject. It can be seen that a GPU based implementation gives full model optimization results and estimated whole brain parameter maps in matters of seconds to minutes even for complex models and a large dataset. Powell almost always outperforms other methods in terms of LL and BIC (except for CHARMED\_in1 for the RLS dataset). The Powell algorithm shows the most consistently high fitting performance over models and datasets, especially on the complex 552 volume HCP MGH dataset. On the simpler 134 volume RLS dataset the fit is sometimes matched (with NMSimplex in Tensor-RLS and with LM in Ball & Sticks\_in1- RLS) or slightly exceeded (by NMSimplex in CHARMED\_in1 - RLS) by one of the other algorithms. In addition, Powell is often the fastest algorithm, and when not the fastest always

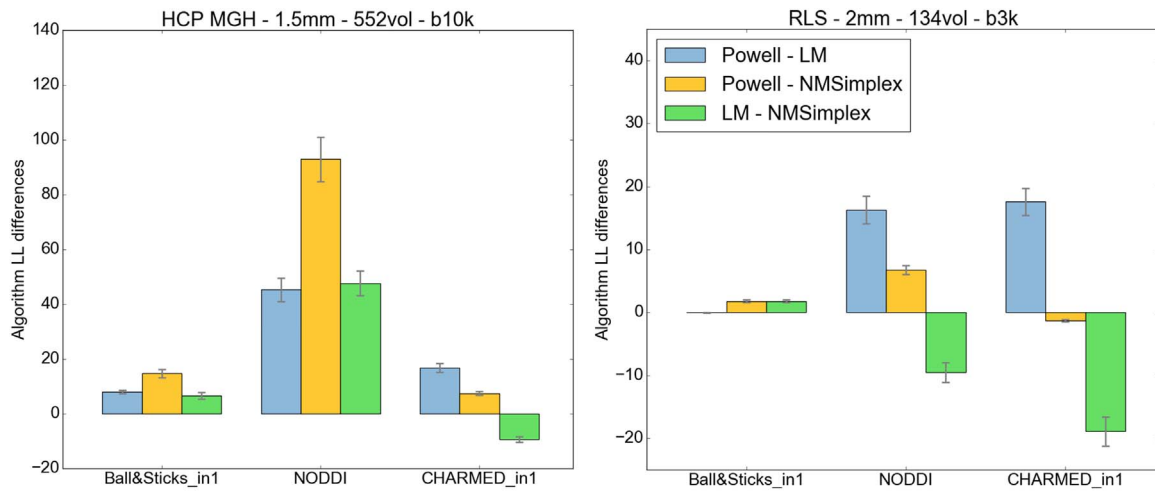
better fitting than the fastest algorithm. To assess the variability of the achieved fit, Fig. 4 reports mean differences in fit between algorithms and its standard error over the mean (SEM) over ten HCP MGH and ten RLS subjects. Powell outperforms the other algorithms except for the case of the CHARMED\_in1 model with the RLS data by a very small margin. Error bars show that the variability of the fit differences over subjects is relatively low with the SEM generally at about ten percent of the fit difference.

To assess accuracy and precision for different optimization algorithms, Fig. 5 reports ground truth simulation results over three models (Ball & Sticks\_in1, NODDI and CHARMED\_in1), the two multi-shell direction tables and three simulated SNRs. A general trend of increasing accuracy and precision with SNR can be observed. In general, Powell has a tendency for higher accuracy and precision than the other algorithms over all models, which is often more pronounced at higher SNRs and more complex models. In terms of accuracy, LM can keep up with Powell in Ball & Sticks\_in1 and NMSimplex in the

**Table 6**

Models (in columns) against optimization algorithms (in rows) for a single HCP MGH and RLS dataset. Reported LL and BIC are the mean over a whole brain white matter mask. The run times are over the entire brain mask and are in units of (h:mm:ss). For the BIC, the number of observations is for Ball & Sticks\_in1, NODDI and CHARMED\_in1 given as (HCP MGH: 552, RLS: 134) and for Tensor given as (HCP MGH: 104, RLS: 44), while the number of free parameters is given as (Tensor: 7, Ball & Sticks\_in1: 4, NODDI: 6, CHARMED\_in1: 11).

		Tensor			Ball & Sticks_in1			NODDI			CHARMED_in1		
		LL	BIC	RT	LL	BIC	RT	LL	BIC	RT	LL	BIC	RT
HCP MGH	NMSimplex	-528	1099	0:00:41	-3000	6025	0:01:33	-2795	5627	1:07:50	-2698	5466	2:30:53
	Powell	-514	1072	0:01:08	-2985	5995	0:00:50	-2714	5467	0:12:08	-2691	5452	0:21:17
	LM	-519	1082	0:01:30	-2992	6009	0:00:49	-2759	5557	0:14:48	-2708	5486	0:28:10
	NMSimplex	-324	682	0:00:07	-1083	2185	0:00:09	-984	1998	0:05:32	-962	1977	0:06:36
RLS	Powell	-324	682	0:00:03	-1081	2182	0:00:03	-976	1980	0:01:23	-963	1979	0:01:19
	LM	-328	689	0:00:28	-1081	2182	0:00:04	-997	2023	0:01:31	-972	1998	0:02:59



**Fig. 4.** Algorithm Log Likelihood differences per model over 10 HCP MGH (left) and 10 RLS subjects (right). Bars depict the mean LL difference and error bars give the standard error of the mean for the difference over subjects.

CHARMED\_in1 model, but precision is almost always considerably higher for Powell. When comparing acquisition tables for the Powell method, little difference can be seen for Ball & Sticks\_in1, the simplest model. For NODDI there seems to be a tendency for increased precision when using the extensive HCP MGH direction table instead of the RLS table at the same SNR. For CHARMED\_in1, a clear increase in accuracy and precision can be observed when going to the more complex HCP protocol, which may reflect the presence of high b-values ( $> 3000 \text{ s/mm}^2$ ) as previously reported (Santis et al., 2014).

#### Effects of cascaded optimization

We next focus on the effect of different cascading strategies for a very complex model, CHARMED\_in3 with three different restricted compartments per voxel and a total number of 19 parameters. This is a very difficult model to fit, even on a very large number of volumes (such as the HCP MGH multi-shell direction table), but nonetheless a very useful model for microstructure quantification as it separates multiple fiber orientations per voxel (De Santis et al., 2012). Fig. 6 shows FR maps and multiple local restricted compartment orientations for CS (Cascade  $S_0$ ), CI (Cascade Initialized) and CF (Cascade Fixed). CS will optimize the full 19 parameters in the final cascade step; CI will also optimize the full 19 parameters in the final cascade step but the orientations are initialized from an earlier Ball & Sticks\_in3 cascade step; CF will optimize 13 parameters in the final cascade step where the 6 restricted orientation parameters are fixed from an earlier cascade step. Transverse maps and orientation plots are shown at the level of the centrum semiovale where three distinct fiber orientations are expected in white matter. Both FR and orientation estimates appear more structured and smooth in CI and CF, compared to CS. Orientation estimates, in particular, show missing identification of the green A-P orientation (corresponding to the superior longitudinal fasciculus) in some places, for CS and to a lesser degree CI (see white arrow). Table 7 reports run times and fit parameters for different models of increasing complexity for a single HCP MGH and RLS dataset. The run times and fit parameters show that quantitatively CI and CF have a better fit than CS. In terms of run time, CF outperforms both CS and CI, especially for the CHARMED\_in3 model by a factor of 3. These results show that a straightforward optimization of a large 19 parameter model as in CS, achieves a suboptimal fit, even when using a large multi-shell direction table. However, with a good initialization of a subset of the parameters,

the fit can be improved considerably. Finally, although results are more subtle, fixing (in CF) rather than initializing (in CI) the orientation parameters, provides similarly good results even though the degrees of freedom in the final optimization step have been reduced. For simpler models (NODDI and CHARMED\_in1 and \_in2) with less parameters, the fit advantages of CI and CF strategies are almost absent and the speed advantages less pronounced, although speed-ups of 1.5x to 3x are still achieved.

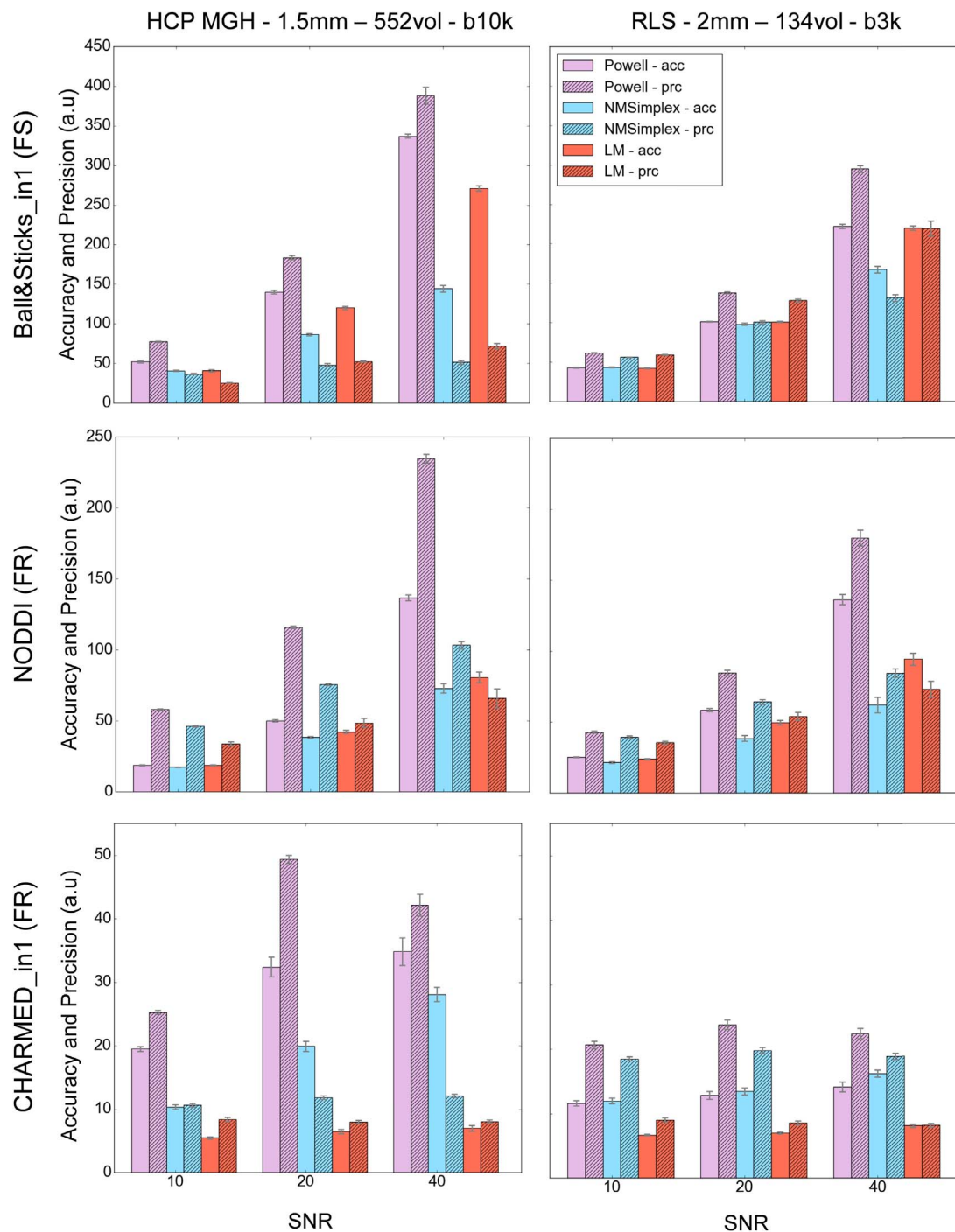
To assess the robustness of the fit improvements, Fig. 7 reports mean difference in fit between cascading strategies and its SEM over ten HCP MGH and ten RLS subjects. CI and CF always have a higher LL than CS (with the exception of NODDI for the HCP MGH data). More generally, the fit improvement of CI and CF over CS increases with model complexity (i.e. with CHARMED\_in2, CHARMED\_in3). In addition, the fit of CI is always slightly higher than that of CF, with this fit difference generally being smaller than that separating CS from both of these. Error bars show that the variability of the fit differences over subjects is relatively low.

To assess accuracy and precision for different cascading strategies, Fig. 8 reports ground truth simulation results for the two multi-shell direction tables and three SNRs for the most complex CHARMED\_in3 model. The cascading strategies CI and CF outperform CS in terms of accuracy and precision, especially for the higher SNRs of 20 and 40. For the more extensive HCP protocol, accuracy and precision for CI and CF is very similar, irrespective of SNR. For the less extensive RLS protocol, CF has a tendency of outperforming CI, especially for lower SNRs. Also, the accuracy and precision is clearly higher for the more complex HCP MGH multi-shell direction table.

#### Discussion

Using an efficient GPU based implementation, we show that run time can be removed as a fundamental constraint for optimization of multi-compartment models, achieving whole brain dataset fits in seconds to minutes. We find that optimization algorithms and multi-step optimization approaches have a considerable influence on performance and stability over subjects and over acquisition protocols. The gradient-free Powell conjugate-direction algorithm was found to outperform other common algorithms in terms of run time, fit, accuracy and precision. Parameter initialization approaches were found to be relevant especially for more complex models, such as those involving several fiber orientations per voxel.





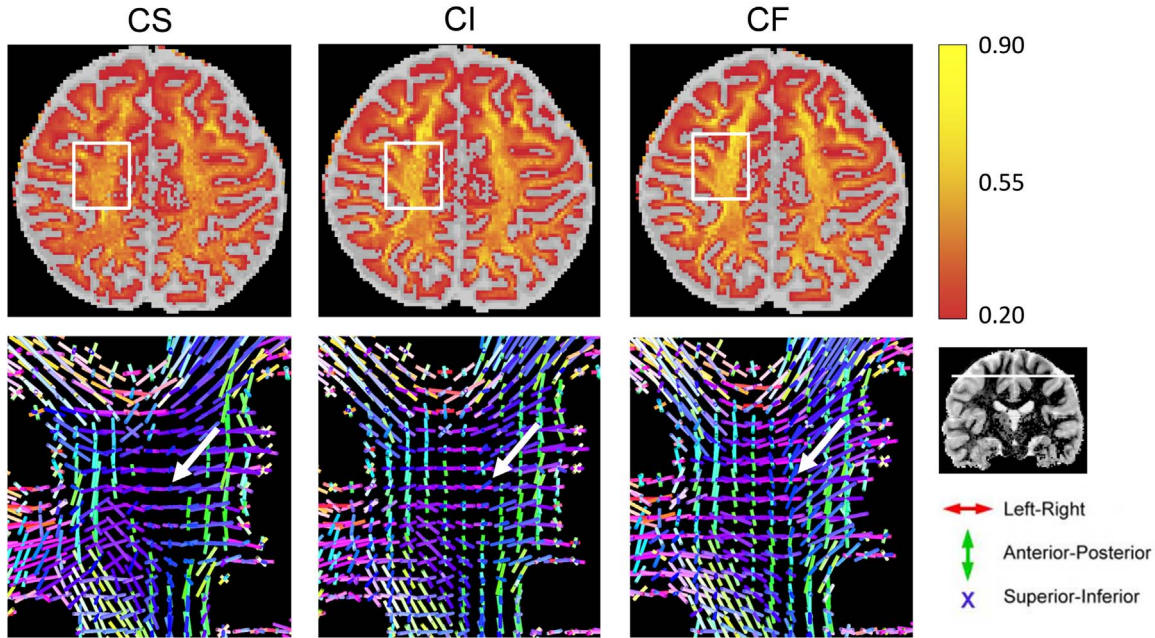
**Fig. 5.** Ground truth simulation results for three models (rows) and two multi-shell direction tables (columns) for three SNRs. The unstriped bars represent the accuracy (acc) per optimization algorithm, the striped bars represent the precision (prc), computed over  $10^4$  realizations per volume fraction. The bars give the mean and the error bars give the standard error computed over seven different simulated volume fractions between 0.2 and 0.8.

### Optimization algorithms

The Powell conjugate-direction algorithm almost always outperforms the other methods in terms of LL and BIC, over multiple models and two different multi-shell acquisition protocols. In addition, Powell is often the fastest algorithm, and when not the fastest it always provides better fitting than a faster algorithm. The variability over ten subjects in the fit differences between algorithms is relatively low with the standard error generally at a

third to half of the fitting difference. This shows that the fitting improvement is a significant and robust one likely exceeding the variability of the model parameters in experimental subject groups and therefore possibly of influence on statistical power. The BIC differences between optimization algorithms are of a similar size as the BIC differences between models. This means that the choice of optimization algorithm could have an influence on later model comparisons and model rankings. Going beyond data fit and considering accuracy and precision using ground truth simulations we find

## Effects of cascaded optimization



**Fig. 6.** Parameter maps (CHARMED\_in3, FR map; top row) and local restricted compartment 3D color-coded orientations (bottom row) for different cascading strategies (in columns) for a single HCP MGH subject (1003). The top row shows a thresholded FR map superimposed on a hindered fraction map in gray-scale.

that Powell outperforms the other algorithms, which is even more pronounced at higher SNRs and with more complex models. When comparing accuracy and precision between acquisition protocols for the Powell method, little effect is observed for simpler models such as Ball & Stick\_in1 and NODDI, but an increase in accuracy can be observed for CHARMED\_in1 when going to the more complex HCP protocol.

### Cascaded optimization

Different forms of cascaded optimization have already been applied to diffusion microstructure modeling. For instance, Zhang et al. (2012) and the accompanying NODDI MATLAB toolbox first do an initial grid search over a feasible range of parameter values after which the parameters are optimized twice, once with the fiber orientations and dispersion locked and the second time with all parameters free. De Santis et al. (2016) uses a two-step optimization approach in which the number of intra-axonal compartments and their orientations are fixed to the orientations found by the CSD method (Tournier et al., 2007; Jeurissen et al., 2014). We formalize and generalize multi-step optimization here as cascaded model optimization. Results show that

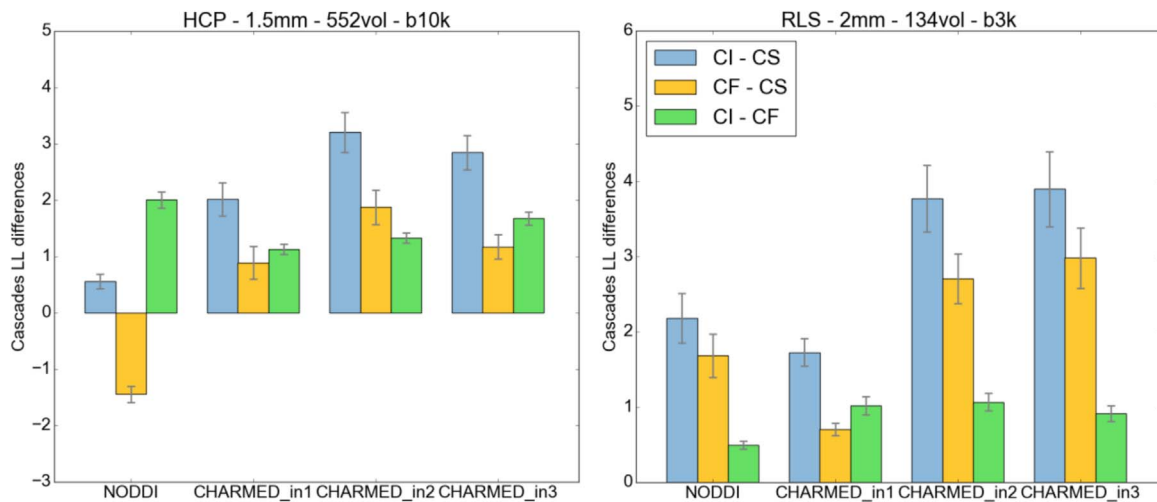
CI and CF always have a better fit (higher Log Likelihood) than CS and that the fit improvement of CI and CF over CS increases with model complexity. This shows that with a good initialization of a subset of the parameters, the fit can be improved considerably over one-step optimization (not considering  $S_0$ , as it is common practice to either work on  $S_0$ -normalized signal or to estimate and fix it beforehand). In terms of accuracy and precision CI and CF also outperform CS, especially for the higher SNRs of 20 and 40.

Comparing between CI and CF for the complex CHARMED\_in3 model (19 parameters) and for the more extensive HCP protocol, the accuracy and precision for CI and CF is very similar, irrespective of SNR. For the less extensive RLS protocol, CF has a tendency of outperforming CI especially for lower SNRs. These results are interesting given that CF has lower degrees of freedom in its final cascade step than CI. More degrees of freedom potentially allow for a better fit since the model can better adapt to the given data, which is shown in the superior fit performance in CI in the context of high SNR and the rich HCP MGH acquisition protocol. However, in the context of few data points or lower SNR, more degrees of freedom are a disadvantage, as is shown in the superior accuracy of CF in the context of lower SNRs

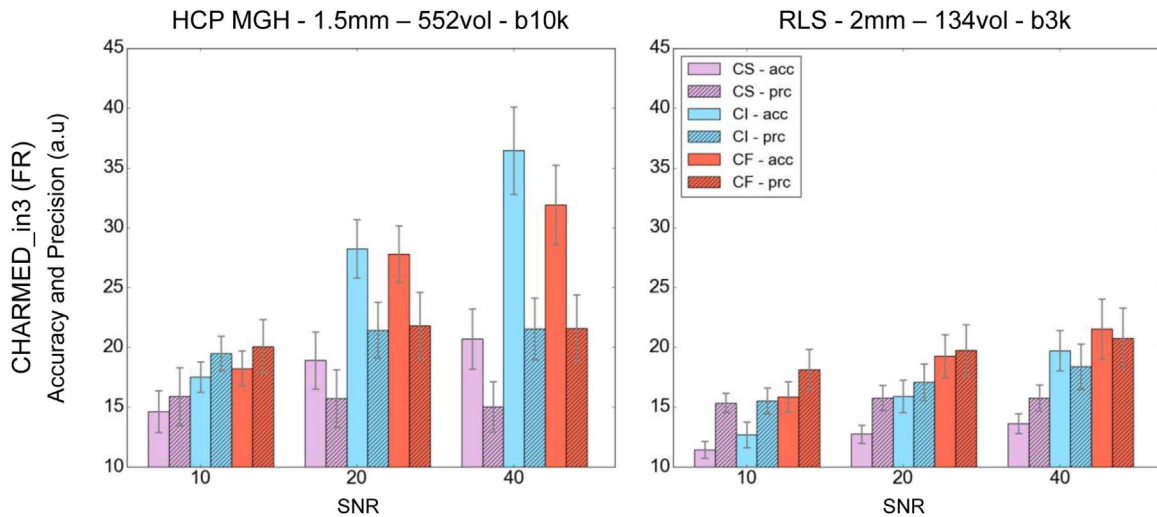
**Table 7**

Models (columns) against optimization algorithms (rows) for a single HCP MGH and RLS dataset. Reported LL and BIC are the mean over a whole brain white matter mask. The run times are over the entire brain mask and are in units of (h:mm:ss). For the BIC, the number of observations is given by the number of volumes in the datasets (552 for HCP MGH and 134 for RLS), while the number of free parameters is given as (NODDI: 6, CHARMED\_in1: 11, CHARMED\_in2: 15, CHARMED\_in3: 19).

		NODDI			CHARMED_in1			CHARMED_in2			CHARMED_in3		
		LL	BIC	RT	LL	BIC	RT	LL	BIC	RT	LL	BIC	RT
HCP MGH	CS	-2714	5467	0:12:08	-2691	5452	0:21:17	-2680	5455	1:03:38	-2677	5474	1:29:34
	CI	-2716	5470	0:11:31	-2689	5447	0:21:02	-2677	5449	1:00:56	-2675	5469	2:07:14
	CF	-2715	5468	0:07:47	-2690	5449	0:13:17	-2678	5451	0:33:44	-2676	5472	1:03:09
	CS	-976	1980	0:01:23	-963	1979	0:01:19	-962	1997	0:03:51	-961	2015	0:11:12
RLS	CI	-974	1977	0:01:09	-961	1977	0:01:42	-959	1991	0:04:57	-958	2009	0:12:31
	CF	-974	1978	0:00:48	-962	1978	0:00:51	-960	1993	0:02:13	-959	2011	0:03:38



**Fig. 7.** Inter-subject differences of the LL for each model and cascading technique over 10 MGH and 10 RLS subjects. The error bars give the standard error of the mean for the inter-subject differences.



**Fig. 8.** Ground truth simulation results for the CHARMED\_in3 model for the HCP MGH and RLS multi-shell direction tables for three SNRs. The unstriped bars represent the accuracy (acc) per cascading strategy, the striped bars represent the precision (prc), computed over a total number of  $10^6$  realizations. The bars give the mean and the error bars give the standard error computed over the volume fractions of the three simulated orientations.

and the RLS acquisition protocol. Hence, overall, CF has very appealing properties for general use as a cascaded optimization strategy for complex models. Besides providing a good fit, accuracy and precision in many contexts, it is also, by a large margin, invariably, the fastest cascading strategy. Since the Powell algorithm provided the best performance when compared to NMSimplex and LM we focused here on comparing cascading results using Powell only. Comparisons of cascading strategies using the other optimization algorithms (see [Supplementary Table 1](#)) show that CI and CS strategies can provide considerable benefits for NMSimplex and LM as well, but also that the Powell algorithm continues to outperform NMSimplex and LM when combined with CF and CI. Other initialization strategies like random multi-start and initializing with a preliminary grid search have been applied to diffusion microstructure model optimization. Comparisons of these initialization strategies to the cascading strategies investigated here using all optimization algorithms (see [Supplementary Table 2](#)) show that considerable fit improvements can indeed be achieved with multi-start and grid search compared to CS for NMSimplex and LM. However, these results are almost invariably inferior to, or as good as,

the CF and CI results, especially for the Powell algorithm. In terms of runtime/performance tradeoff the cascading strategies perform vastly better than grid search and multi-start.

An additional argument for the use of CF is that it allows combining microstructure parameter maps with tractography results in a structured way (e.g.; [Bells et al., 2011](#)). One can use the fiber orientations from a Ball & Sticks\_in[n] fit for tractography and subsequently use CF with the same orientations for a CHARMED\_in[n] model to ensure that microstructure parameters mapped onto these tracks agree with the local track orientation. Alternatively, when using CI or when fitting the tractography orientations and the microstructure models separately, one would not have this assurance. Future research could focus on more ways of initializing or fixing parameters over cascading steps, such as limiting the extra-axonal Tensor compartment in CHARMED to the orientations of the intra axonal orientations. Additionally, one could also use other methods for orientation initialization, such as CSD ([Tournier et al., 2007](#); [Jeurissen et al., 2014](#)). Additionally, one can think of cascades where the final model is initialized with the results of more than one model, for example, initializing the orientations with a Ball & Sticks model and initializing a (possible) T1 or T2

compartment from a different model. These extensions fit within the framework proposed here.

### Run times

All run times reported here were achieved with a single 2015 AMD Fury X graphics card with a theoretical peak performance of 8.6 teraflops single precision. Higher performing hardware, as well as use of multiple devices in parallel should deliver proportional run time gains. In practice, OpenCL driver performance, number of parallel streaming processors and streaming processor architecture, device memory and memory bandwidth (here 4GB and 512GB/sec, respectively) also determine computational performance.

### Limitations

In this work we compared three optimization algorithms: Nelder-Mead Simplex, Powell's conjugate-direction and Levenberg-Marquardt which covers a set of generally used algorithms. Other optimization algorithms could be considered. Since all the tested optimization routines are local optimizers, it would be interesting to also use global optimization algorithms such as the Self-Organizing Migrating Algorithm (SOMA; Zelinka, 2004) to evaluate their capacity of avoiding local minima. Also Markov Chain Monte Carlo sampling of the full Bayesian posterior would be an interesting extension since it can avoid local minima and would provide precision or uncertainty estimates for the obtained parameters. However MCMC sampling is a different type of algorithm than the optimization algorithms covered in this work, in two ways. First, it has a different goal: it aims to sample the full posterior distribution of parameters, rather than provide a (maximum likelihood) point estimate of the optimal parameter value. Generating the full posterior inevitably also takes more runtime, which is often only justified if more than just the maximum of the posterior is used (e.g. the standard deviation of the posterior). Second, it requires more information. All algorithms in this work are local optimization routines that only require an objective (likelihood) function to perform the minimization. MCMC on the other hand additionally requires the user to define a prior probability function and a proposal function for each of the model parameters. Other optimization approaches avoid non-linear optimization altogether by linearizing the optimization problem (Daducci et al., 2015; Novikov, 2015). For instance, AMICO (Daducci et al., 2015) constructs a convex formulation with a single global minimum by choosing a finite dictionary of atoms and has been successfully applied to the NODDI, ActiveAx (Anna Auer, 2015, Daducci et al., 2015) and VERDICT (Bonet-Carne et al., 2016) models. This different approach comes with its own trade-offs. For instance, atom dictionary size can increase exponentially in more complex models when sampling interdependent parameters sufficiently dense (Pew-Thian Yap, 2015) and it is limited to a Gaussian noise model, which embeds assumptions almost never met in standard multichannel acquisitions (Gudbjartsson and Patz, 1995; Dietrich et al., 2008; Daducci et al., 2015).

## Appendix A

To enforce boundary conditions on the parameters and to scale the parameters to a range better suited for optimization, we use encoding and decoding transformations on the model parameters. Table 8 shows the model encoding and decoding transformations used in this work. For models that share the same parameter and that have a different transformation we have listed the specific transformation per parameter. The transformations containing a cosine or sine are used to restrain the parameter within a given set of lower and upper bounds. The general encoding transformation is (for example for the sine variant):  $\bar{x} = \text{asin}(\sqrt{\frac{x - \text{lb}}{\text{ub} - \text{lb}}})$ , where  $x$  is the parameter in model space,  $\bar{x}$  is the parameter in optimization space,  $\text{lb}$  is the lower bound of that parameter (in model space) and  $\text{ub}$  the upper bound of that parameter (in model space). The decoding operation reverses this transformation using  $x = \sin^2(\bar{x}) \cdot (\text{ub} - \text{lb}) + \text{lb}$ .

## Conclusions and recommendations

When the aim is optimization of a point estimate, rather than sampling a full parameter posterior, (e.g. with MCMC), we would advise the use of the Powell algorithm for microstructure modelling in dMRI. For low complexity models up to about 8 parameters, including NODDI and Ball & Sticks\_in1 a simple CS strategy can be used, although a CF strategy can yield run time gains. For intermediate complexity models, such as Ball & Sticks\_in3, we recommend using a successive CI cascade on an increasing number of fiber orientations. Finally, CF is recommended for the higher complexity models such as CHARMED\_in3 yielding run time, fit, accuracy and precision advantages as well as advantages for combined tractography and microstructure.

### Software implementation note

All biophysical compartment models, noise models / likelihood functions, as well as optimization algorithms were modularly implemented in a python based GPU accelerated toolbox (Maastricht Diffusion Toolbox or MDT) freely available under an open source L-GPL license at <https://github.com/cbclab/MDT>. It comes with a graphical user interface for easy usage as an analysis tool. MDT can be used to perform all model optimizations reported in this work and the modular design of the toolbox allows easy extension with new single compartment models and composite multi-compartment models.

### Acknowledgements

AR was supported by an ERC Starting Grant (MULTICONNECT, #639938) and Dutch science foundation (NWO) VIDI Grant (#14637). RG was supported by an ERC Advanced Grant (# 269853). RLH and RG have received funding from the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement n°602450. This paper reflects only the author's views and the European Union is not liable for any use that may be made of the information contained therein. Data collection and sharing for this project was provided, in part, by the MGH-USC Human Connectome Project (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D., Van J. Weeden, MD). HCP funding was provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). HCP data are disseminated by the Laboratory of Neuro Imaging at the University of California, Los Angeles. Collectively, the HCP is the result of efforts of co-investigators from the University of California, Los Angeles, Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH), Washington University, and the University of Minnesota. Data collection and sharing for this project was provided, in part, by the Rhineland Study (Principal Investigator: Monique M.B. Breteler, M.D., Ph.D.; German Center for Neurodegenerative Diseases (DZNE), Bonn).



**Table 8**Parameter transformations, the units of the diffusivities is in  $\text{m}^2/\text{s}$ , other units are dimensionless.

Parameter	Model (& compartment)	Encoding transformation	Decoding transformation
$d_{  }$	Tensor	$\bar{d}_{  } = \text{asin} \left( \sqrt{\frac{d_{  }}{1 * 10^{-8}}} \right)$	$d_{  } = \sin^2(\bar{d}_{  }) * 1 * 10^{-8}$
$d_{\perp 1}$	Tensor	$\bar{d}_{\perp 1} = \text{asin} \left( \sqrt{\frac{d_{\perp 1}}{1 * 10^{-8}}} \right)$	$d_{\perp 1} = \sin^2(\bar{d}_{\perp 1}) * 1 * 10^{-8}$
$d_{\perp 2}$	Tensor	$\bar{d}_{\perp 2} = \text{asin} \left( \sqrt{\frac{d_{\perp 2}}{1 * 10^{-8}}} \right)$	$d_{\perp 2} = \sin^2(\bar{d}_{\perp 2}) * 1 * 10^{-8}$
$\psi$	–	$\bar{\psi} = \lfloor \psi \rfloor \% \pi$	$\psi = \lfloor \bar{\psi} \rfloor \% \pi$
$\theta$	–	$\bar{\theta} = \lfloor \theta \rfloor \% \pi$	$\theta = \lfloor \bar{\theta} \rfloor \% \pi$
$\phi$	–	$\bar{\phi} = \lfloor \phi \rfloor \% \pi$	$\phi = \lfloor \bar{\phi} \rfloor \% \pi$
$\kappa$	–	$\bar{\kappa} = \text{acos} \left( \sqrt{\frac{\kappa - 1 * 10^{-5}}{2\pi - 1 * 10^{-5}}} \right)$	$\kappa = \cos^2(\bar{\kappa}) * (2\pi - 1 * 10^{-5}) + 1 * 10^{-5}$
$d_{  }$	CHARMED (Tensor)	$\bar{d}_{  } = \text{asin} \left( \sqrt{\frac{d_{  } - 1 * 10^{-9}}{5 * 10^{-9} - 1 * 10^{-9}}} \right)$	$d_{  } = \sin^2(\bar{d}_{  }) * (5 * 10^{-9} - 1 * 10^{-9}) + 1 * 10^{-9}$
$d_{\perp 1}$	CHARMED (Tensor)	$\bar{d}_{\perp 1} = \text{asin} \left( \sqrt{\frac{d_{\perp 1} - 1 * 10^{-10}}{5 * 10^{-9} - 3 * 10^{-10}}} \right)$	$d_{\perp 1} = \sin^2(\bar{d}_{\perp 1}) * (5 * 10^{-9} - 3 * 10^{-10}) + 3 * 10^{-10}$
$d_{\perp 2}$	CHARMED (Tensor)	$\bar{d}_{\perp 2} = \text{asin} \left( \sqrt{\frac{d_{\perp 2} - 3 * 10^{-10}}{3 * 10^{-9} - 3 * 10^{-10}}} \right)$	$d_{\perp 2} = \sin^2(\bar{d}_{\perp 2}) * (3 * 10^{-9} - 3 * 10^{-10}) + 3 * 10^{-10}$
$d$	CHARMED(CHARMED_in)	$\bar{d} = \text{asin} \left( \sqrt{\frac{d - 3 * 10^{-10}}{3 * 10^{-9} - 3 * 10^{-10}}} \right)$	$d = \sin^2(\bar{d}) * (3 * 10^{-9} - 3 * 10^{-10}) + 3 * 10^{-10}$
$\eta$	CHARMED(Johnson noise)	$\bar{\eta} = \text{acos} \left( \sqrt{\frac{\eta}{1 * 10^6}} \right)$	$\eta = \cos^2(\bar{\eta}) * 1 * 10^6$

## Appendix B

Table 9, Table 10 and Table 11 show the model initializations in use for respectively CS, CI and CF. The target model (first column) is the final model we wish to optimize, the source model (second column) is the model we use as a source of initialization. For CS and CI the last column shows which parameters are initialized from source model to target model. The last column in the CF table shows which parameters are fixed instead initialized. That is, CF has the same initialization as CI except for some parameters that are fixed instead of initialized.

Please note that the cascades are defined recursively. For example, if we optimize CHARMED\_in[n] using CI, we use Ball & Sticks\_in[n] as a source model. Ball & Sticks\_in[n] in turn uses Ball & Sticks\_in[n-1] as a source of initialization, and so forth until the source model is S0.

The initialization and fixes are indicated with an assignment operation. For example, Tensor.  $\theta = \text{Stick0}.\theta$  indicates that the  $\theta$  parameter of the Tensor compartment in the target model is initialized with the optimized  $\theta$  parameter maps of the Stick0 model. In all the tables we use the convention that initializations and fixes with similar named models are abbreviated to a single statement. For example, initializing/fixing Stick0.  $\theta$  of the previous model to the Stick0.  $\theta$  of the next model is abbreviated as just Stick0.  $\theta$ .

**Table 9**The initialization in a Cascade S<sub>0</sub> (CS).

Target model	Source model	Cascade parameters inits
Ball & Sticks_in[n]	S0	S0.s0
CHARMED_in[n]	S0	S0.s0
Tensor	S0	S0.s0
NODDI	S0	S0.s0

**Table 10**  
The initializations in a Cascade Initialize (CI) cascade.

Target model	Source model	Cascade parameters inits
Ball & Sticks_in1	S0	S0.s0
Ball & Sticks_in[n]	Ball & Sticks_in [n-1]	S0.s0, Stick[n].{ $\theta, \phi$ }, $w_{stick\{0,1,2,\dots,n\}}$
CHARMED_in[n]	Ball & Sticks_in [n]	S0.s0 CHARMED_in[n].{ $\theta, \phi$ } = Stick[n].{ $\theta, \phi$ } Tensor.{ $\theta, \phi$ } = Stick0.{ $\theta, \phi$ } $w_{res\{0,1,2,\dots,n\}} = w_{stick\{0,1,2,\dots,n\}}$
Tensor	Ball & Sticks_in1	S0.s0, Tensor.{ $\theta, \phi$ } = Stick0.{ $\theta, \phi$ }
NODDI	Ball & Sticks_in1	S0.s0 NODDI_in.{ $\theta, \phi$ } = Stick0.{ $\theta, \phi$ } $w_{csf} = w_{ball}$ $w_{in} = w_{stick0}/2$ $w_{ex} = w_{stick0}/2$

**Table 11**  
The fixes overriding the initializations in a Cascade Fixed (CF) cascade.

Target model	Source model	Overriding cascade fixes
CHARMED_in[n]	Ball & Sticks_in [n]	CHARMED_in[n].{ $\theta, \phi$ } = Stick[n].{ $\theta, \phi$ }
NODDI	Ball & Sticks_in1	NODDI_in.{ $\theta, \phi$ } = Stick0.{ $\theta, \phi$ }

Appendix C. Supplementary material

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.neuroimage.2017.04.064](https://doi.org/10.1016/j.neuroimage.2017.04.064).

References

Aboitiz, F., Scheibel, A.B., Fisher, R.S., Zaidel, E., 1992. Fiber composition of the human corpus callosum. *Brain Res.* 598 (1-2), 143–153.

Alexander, D.C., 2009. In: Laidlaw, D., Weickert, J. (Eds.), *Modelling, Fitting and Sampling in Diffusion MRI. Visualization and Processing of Tensor Fields: Advances and Perspectives*. Springer Berlin Heidelberg, Berlin, Heidelberg, 3–20.

Alexander, D.C., Hubbard, P.L., Hall, M.G., Moore, E.A., Ptito, M., Parker, G.J., Dyrby, T.B., 2010. Orientationally invariant indices of axon diameter and density from diffusion MRI. *Neuroimage* 52 (4), 1374–1389.

Anna Auria, E.C.-R., Wiaux, Yves, Dirby, Tim, Alexander, Daniel, Thiran, Jean-Philippe, Daducci, Alessandro, 2015. Accelerated Microstructure Imaging via Convex Optimization (AMICO) in Crossing Fibers. *ISMRM*.

Assaf, Y., Alexander, D.C., Jones, D.K., Bizzi, A., Behrens, T.E., Clark, C.A., Cohen, Y., Dyrby, T.B., Huppi, P.S., Knoesche, T.R., Lebihan, D., Parker, G.J., Poupon, C., consortium, C., Anaby, D., Anwender, A., Bar, L., Barazany, D., Blumenfeld-Katzir, T., De-Santis, S., Duclap, D., Figini, M., Fisch, E., Guevara, P., Hubbard, P., Hofstetter, S., Jbabdi, S., Kunz, N., Lazeyras, F., Lebois, A., Liptrot, M.G., Lundell, H., Mangin, J.F., Dominguez, D.M., Morozov, D., Schreiber, J., Seunarine, K., Nava, S., Poupon, C., Riffert, T., Sasson, E., Schmitt, B., Shemesh, N., Sotiropoulos, S.N., Tavor, I., Zhang, H.G., Zhou, F.L., 2013. The CONNECT project: combining macro- and micro-structure. *Neuroimage* 80, 273–282.

Assaf, Y., Basser, P.J., 2005. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *Neuroimage* 27 (1), 48–58.

Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y., Basser, P.J., 2008. AxCaliber: a method for measuring axon diameter distribution from diffusion MRI. *Magn. Reson. Med.* 59 (6), 1347–1354.

Assaf, Y., Freidlin, R.Z., Rohde, G.K., Basser, P.J., 2004. New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magn. Reson. Med.* 52 (5), 965–978.

Assaf, Y., Pasternak, O., 2008. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *J. Mol. Neurosci.* 34 (1), 51–61.

Basser, P.J., Mattiello, J., LeBihan, D., 1994. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66 (1), 259–267.

Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system—a technical review. *NMR Biomed.* 15 (7-8), 435–455.

Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50 (5), 1077–1088.

Bells, S., Cercignani, M., Deoni, S., Assaf, Y., Pasternak, O., Evans, C.J., Leemans, A., Jones, D.K., 2011. Tractometry: comprehensive multi-modal quantitative assessment of white matter along specific tracts. *Proceedings International Soc. Magn. Res. Med.*, Montreal.

Benitez, A., Fieremans, E., Jensen, J.H., Falangola, M.F., Tabesh, A., Ferris, S.H., Helpert, J.A., 2014. White matter tract integrity metrics reflect the vulnerability of late-myelinating tracts in Alzheimer's disease. *Neuroimage Clin.* 4, 64–71.

Bergers, E., Bot, J.C., De Groot, C.J., Polman, C.H., Lycklama a Nijeholt, G.J., Castelijns, J.A., van der Valk, P., Barkhof, F., 2002. Axonal damage in the spinal cord of MS patients occurs largely independent of T2 MRI lesions. *Neurology* 59 (11), 1766–1771.

Bihan, D.L., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M., 1986. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology* 161 (2), 401–407.

Bonet-Carne, E.D., Alessandro, Panagiotaki, Eleftheria, Johnston, Edward, Stevens, Nicola, Atkinson, David, Punwani, Shonit, Alexander, D.C., 2016. Non-invasive Quantification of Prostate Cancer using AMICO Framework for VERDICT MRI. *ISMRM*.

Brent, R.P., 1973. *Algorithms for Minimization Without Derivatives*. Dover Publications.

Daducci, A., Canales-Rodriguez, E.J., Zhang, H., Dyrby, T.B., Alexander, D.C., Thiran, J.P., 2015. Accelerated Microstructure Imaging via Convex Optimization (AMICO) from diffusion MRI data. *Neuroimage* 105, 32–44.

De Santis, S., Assaf, Y., Jeurissen, B., Jones, D.K., Roebroeck, A., 2016. T1 relaxometry of crossing fibres in the human brain. *Neuroimage* 141, 133–142.

De Santis, S., Assaf, Y., Jones, D.K., 2012. Using the biophysical CHARMED model to elucidate the underpinnings of contrast in diffusional kurtosis analysis of diffusion-weighted MRI. *Magn. Reson. Mater. Phys. Biol. Med.* 25 (4), 267–276.

De Santis, S., Drakesmith, M., Bells, S., Assaf, Y., Jones, D.K., 2014. Why diffusion tensor MRI does well only some of the time: variance and covariance of white matter tissue microstructure attributes in the living human brain. *Neuroimage* 89, 35–44.

De Santis, S., Jones, D.K., Roebroeck, A., 2016. Including diffusion time dependence in the extra-axonal space improves in vivo estimates of axonal diameter and density in human white matter. *Neuroimage* 130, 91–103.

Dietrich, O., Raya, J.G., Reeder, S.B., Ingrisch, M., Reiser, M.F., Schoenberg, S.O., 2008. Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. *Magn. Reson. Imaging* 26 (6), 754–762.

Dietrich, O., Raya, J.G., Reeder, S.B., Reiser, M.F., Schoenberg, S.O., 2007. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *J. Magn. Reson. Imaging* 26 (2), 375–385.

- Fan, Q., Witzel, T., Nummenmaa, A., Van Dijk, K.R., Van Horn, J.D., Drews, M.K., Somerville, L.H., Sheridan, M.A., Santillana, R.M., Snyder, J., Hedden, T., Shaw, E.E., Hollinshead, M.O., Renvall, V., Zanzonico, R., Keil, B., Cauley, S., Polimeni, J.R., Tisdall, D., Buckner, R.L., Wedeen, V.J., Wald, L.L., Toga, A.W., Rosen, B.R., 2016. MGH-USC Human Connectome Project datasets with ultra-high b-value diffusion MRI. *Neuroimage* 124 (Pt B), 1108–1114.
- Ferizi, U., Schneider, T., Panagiotaki, E., Nedjati-Gilani, G., Zhang, H., Wheeler-Kingshott, C.A., Alexander, D.C., 2014. A ranking of diffusion MRI compartment models with in vivo human brain data. *Magn. Reson. Med.* 72 (6), 1785–1792.
- Ferizi, U., Schneider, T., Tariq, M., Wheeler-Kingshott, C.A., Zhang, H., Alexander, D.C., 2013. The importance of being dispersed: a ranking of diffusion MRI models for fibre dispersion using in vivo human brain data. *Med. Image Comput. Comput. Assist. Interv.* 16 (Pt 1), 74–81.
- Fieremans, E., Benitez, A., Jensen, J.H., Falangola, M.F., Tabesh, A., Deardorff, R.L., Spampinato, M.V., Babb, J.S., Novikov, D.S., Ferris, S.H., Helpert, J.A., 2013. Novel white matter tract integrity metrics sensitive to Alzheimer disease progression. *AJNR Am. J. Neuroradiol.* 34 (11), 2105–2112.
- Gao, F., Han, L., 2012. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Comput. Optim. Appl.* 51 (1), 259–277.
- Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI Data. *Magn. Reson. Med.* 34 (6), 910–914.
- Hernandez, M., Guerrero, G.D., Cecilia, J.M., Garcia, J.M., Inuggi, A., Jbabdi, S., Behrens, T.E., Sotiropoulos, S.N., 2013. Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using GPUs. *PLoS One* 8 (4), e61892.
- Horsfield, M.A., Jones, D.K., 2002. Applications of diffusion-weighted and diffusion tensor MRI to white matter diseases—a review. *NMR Biomed.* 15 (7–8), 570–577.
- Jelescu, I.O., Veraart, J., Adisetiyo, V., Milla, S.S., Novikov, D.S., Fieremans, E., 2015. One diffusion acquisition and different white matter models: how does microstructure change in human early development based on WMTI and NODDI? *Neuroimage* 107, 242–256.
- Jeurissen, B., Tournier, J.D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *Neuroimage* 103, 411–426.
- Jones, D.K., Knosche, T.R., Turner, R., 2013. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *Neuroimage* 73, 239–254.
- Kamagata, K., Hatano, T., Okuzumi, A., Motoi, Y., Abe, O., Shimoji, K., Kamiya, K., Suzuki, M., Hori, M., Kumamaru, K.K., Hattori, N., Aoki, S., 2016. Neurite orientation dispersion and density imaging in the substantia nigra in idiopathic Parkinson disease. *Eur. Radiol.* 26 (8), 2567–2577.
- Kunz, N., Zhang, H., Vasung, L., O'Brien, K.R., Assaf, Y., Lazeyras, F., Alexander, D.C., Huppi, P.S., 2014. Assessing white matter microstructure of the newborn with multi-shell diffusion MRI and biophysical compartment models. *Neuroimage* 96, 288–299.
- Lebel, C., Caverhill-Godkewitch, S., Beaulieu, C., 2010. Age-related regional variations of the corpus callosum identified by diffusion tensor tractography. *Neuroimage* 52 (1), 20–31.
- Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. *Q. J. Appl. Mathematics* II (2), 164–168.
- Li, H., Jiang, X., Xie, J., Gore, J.C., Xu, J., 2016. Impact of transcytolemmal water exchange on estimates of tissue microstructural properties derived from diffusion MRI. *Magn. Reson. Med.*
- Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Ugurbil, K., 2010. Multiband multislice GE-EPI at 7 T, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn. Reson. Med.* 63 (5), 1144–1153.
- Neil, J., Miller, J., Mukherjee, P., Huppi, P.S., 2002. Diffusion tensor imaging of normal and injured developing human brain—a technical review. *NMR Biomed.* 15 (7–8), 543–552.
- Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. *Comput. J.* 7 (4), 308–313.
- Nilsson, M., van Westen, D., Ståhlberg, F., Sundgren, P.C., Lätt, J., 2013. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. *Magn. Reson. Mater. Phys. Biol. Med.* 26 (4), 345–370.
- Novikov, D.S.J., Ileana, O., Fieremans, E., 2015. From Diffusion Signal Moments to Neurite Diffusivities, Volume Fraction and Orientation Distribution: An Exact Solution. *ISMRM*.
- Panagiotaki, E., Schneider, T., Siow, B., Hall, M.G., Lythgoe, M.F., Alexander, D.C., 2012. Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. *Neuroimage* 59 (3), 2241–2254.
- Pew-Thian Yap, Y.Z., Shen, Dinggang, 2015. Iterative Subspace Screening for Rapid Sparse Estimation of Brain Tissue Microstructural Properties. *MICCAI*.
- Pfefferbaum, A., Sullivan, E.V., Hedehus, M., Lim, K.O., Adalsteinsson, E., Moseley, M., 2000. Age-related decline in brain white matter anisotropy measured with spatially corrected echo-planar diffusion tensor imaging. *Magn. Reson. Med.* 44 (2), 259–268.
- Powell, M.J.D., 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* 7 (2), 155–162.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. *Numerical Recipes: The Art of Scientific Computing* 3rd Edition. Cambridge University Press.
- Rios, L.M., Sahinidis, N.V., 2012. Derivative-free optimization: a review of algorithms and comparison of software implementations. *J. Glob. Optim.* 56 (3), 1247–1293.
- Santis, S., Assaf, Y., Evans, C.J., Jones, D.K., 2014. Improved precision in CHARMED assessment of white matter through sampling scheme optimization and model parsimony testing. *Magn. Reson. Med.* 71 (2), 661–671.
- Schwarz, G., 1978. *Estim. Dimens. A Model*, 461–464.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17 (3), 143–155.
- Sotak, C.H., 2002. The role of diffusion tensor imaging in the evaluation of ischemic brain injury—a review. *NMR Biomed.* 15 (7–8), 561–569.
- Stepisnik, J., 1993. Time-dependent self-diffusion by NMR spin-echo. *Physica B* 4 (193), 343–350.
- Sundgren, P.C., Dong, Q., Gomez-Hassan, D., Mukherji, S.K., Maly, P., Welsh, R., 2004. Diffusion tensor imaging of the brain: review of clinical applications. *Neuroradiology* 46 (5), 339–350.
- Tariq, M., Schneider, T., Alexander, D.C., Gandini Wheeler-Kingshott, C.A., Zhang, H., 2016. Bingham-NODDI: mapping anisotropic orientation dispersion of neurites using diffusion MRI. *Neuroimage* 133, 207–223.
- Timmers, I., Zhang, H., Bastiani, M., Jansma, B.M., Roebroek, A., Rubio-Gozalbo, M.E., 2015. White matter microstructure pathology in classic galactosemia revealed by neurite orientation dispersion and density imaging. *J. Inherit. Metab. Dis.* 38 (2), 295–304.
- Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* 35 (4), 1459–1472.
- Wang, P.C., Shoup, T.E., 2011. Parameter sensitivity study of the Nelder-Mead Simplex Method. *Adv. Eng. Softw.* 42 (7), 529–533.
- Wen, Q., Kelley, D.A., Banerjee, S., Lupo, J.M., Chang, S.M., Xu, D., Hess, C.P., Nelson, S.J., 2015. Clinically feasible NODDI characterization of glioma using multiband EPI at 7 T. *Neuroimage Clin.* 9, 291–299.
- Werring, D.J., Clark, C.A., Barker, G.J., Thompson, A.J., Miller, D.H., 1999. Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology* 52 (8), 1626–1632.
- Xu, J., Moeller, S., Auerbach, E.J., Strupp, J., Smith, S.M., Feinberg, D.A., Yacoub, E., Ugurbil, K., 2013. Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *Neuroimage* 83, 991–1001.
- Zelinka, I., 2004. *SOMA—Self-Organizing Migrating Algorithm*. New Optimization Techniques in Engineering. Springer Berlin Heidelberg, Berlin, Heidelberg, 167–217.
- Zhang, H., Hubbard, P.L., Parker, G.J., Alexander, D.C., 2011. Axon diameter mapping in the presence of orientation dispersion with diffusion MRI. *Neuroimage* 56 (3), 1301–1315.
- Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C., 2012. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* 61 (4), 1000–1016.