


RESEARCH ARTICLE

Diffusion MRI microstructure models with in vivo human brain Connectome data: results from a multi-group comparison

Uran Ferizi^{1,2,3}  | Benoit Scherrer⁴ | Torben Schneider^{3,5} | Mohammad Alipoor⁶ |
Odin Eufrazio⁷ | Rutger H.J. Fick⁸ | Rachid Deriche⁸ | Markus Nilsson⁹  |
Ana K. Loya-Olivas⁷ | Mariano Rivera⁷ | Dirk H.J. Poot¹⁰ |
Alonso Ramirez-Manzanares⁷ | Jose L. Marroquin⁷ | Ariel Rokem^{11,12} |
Christian Pötter¹² | Robert F. Dougherty¹² | Ken Sakaie¹³ |
Claudia Wheeler-Kingshott³ | Simon K. Warfield⁴ | Thomas Witzel¹⁴ |
Lawrence L. Wald¹⁴ | José G. Raya² | Daniel C. Alexander¹

¹Centre for Medical Image Computing, Department of Computer Science, University College London, UK

²Department of Radiology, New York University School of Medicine, USA

³Department of Neuroinflammation, Institute of Neurology, University College London, UK

⁴Computational Radiology Laboratory, Boston Children's Hosp., Harvard University, USA

⁵Philips Healthcare, Guildford, Surrey, UK

⁶Chalmers University of Technology, Gothenburg, Sweden

⁷Centro de Investigacion en Matematicas AC, Guanajuato, Mexico

⁸Athena Project-Team, INRIA Sophia Antipolis - Méditerranée, France

⁹Lund University Bioimaging Center, Lund University, Sweden

¹⁰Erasmus Medical Center and Delft University of Technology, the Netherlands

¹¹eScience Institute, University of Washington, USA

¹²Center for Cognitive and Neurobiological Imaging, Stanford University, USA

¹³Imaging Institute, The Cleveland Clinic, Cleveland, USA

¹⁴A.A. Martinos Center for Biomedical Imaging, MGH, Harvard University, USA

Correspondence

Uran Ferizi, PhD, Center for Biomedical Imaging, Department of Radiology, New York University Langone Medical Center, 660 First Avenue, 4th floor, New York, NY 10016, USA.

Email: uran.ferizi@med.nyu.edu

Funding information

The (UK) Engineering and Physical Sciences Research Council (EPSRC), Grant/Award Number: EP/G007748, EP/L022680/1, EP/I027084/01, EP/M020533/1 and EP/N018702/1; The (US) National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) of the National Institute of Health (NIH), Grant/Award Number: R21AR066897 and R01AR067789; NIH, Grant/Award Number: R01 NS079788, R01 EB019483, R01 EB018988 and BCH TRP Pilot and BCH CTREC K-to-R Merit Award; Swedish Strategic Research (SSF), Grant/Award Number: AM13-0090

A large number of mathematical models have been proposed to describe the measured signal in diffusion-weighted (DW) magnetic resonance imaging (MRI). However, model comparison to date focuses only on specific subclasses, e.g. compartment models or signal models, and little or no information is available in the literature on how performance varies among the different types of models. To address this deficiency, we organized the 'White Matter Modeling Challenge' during the International Symposium on Biomedical Imaging (ISBI) 2015 conference. This competition aimed to compare a range of different kinds of models in their ability to explain a large range of measurable in vivo DW human brain data. Specifically, we assessed the ability of models to predict the DW signal accurately for new diffusion gradients and b values. We did not evaluate the accuracy of estimated model parameters, as a ground truth is hard to obtain. We used the Connectome scanner at the Massachusetts General Hospital, using gradient strengths of up to 300 mT/m and a broad set of diffusion times. We focused on assessing the DW signal prediction in two regions: the genu in the corpus callosum, where the fibres are relatively straight and parallel, and the fornix, where the

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.

Copyright © 2017 The Authors. *NMR in Biomedicine* Published by John Wiley & Sons, Ltd.

configuration of fibres is more complex. The challenge participants had access to three-quarters of the dataset and their models were ranked on their ability to predict the remaining unseen quarter of the data. The challenge provided a unique opportunity for a quantitative comparison of diverse methods from multiple groups worldwide. The comparison of the challenge entries reveals interesting trends that could potentially influence the next generation of diffusion-based quantitative MRI techniques. The first is that signal models do not necessarily outperform tissue models; in fact, of those tested, tissue models rank highest on average. The second is that assuming a non-Gaussian (rather than purely Gaussian) noise model provides little improvement in prediction of unseen data, although it is possible that this may still have a beneficial effect on estimated parameter values. The third is that preprocessing the training data, here by omitting signal outliers, and using signal-predicting strategies, such as bootstrapping or cross-validation, could benefit the model fitting. The analysis in this study provides a benchmark for other models and the data remain available to build up a more complete comparison in the future.

KEYWORDS

brain microstructure, Connectome, diffusion MRI, fornix, genu, model selection

1 | INTRODUCTION

Diffusion-weighted (DW) magnetic resonance imaging (MRI) can provide unique insights into the microstructure of living tissue and is increasingly used to study the microanatomy and development of normal functioning tissue as well as its pathology. Mathematical models for analysis and interpretation have been crucial for the development and translation of DW-MRI. Even though diffusion tensor imaging (DTI),¹ which is based on a simple Gaussian model of the DW-MRI signal, has shown promise in clinical applications,² e.g. Alzheimer's disease,³ multiple sclerosis⁴ or brain tumors,⁵ a much wider variety of DW-MRI models has been proposed to extract more information from the DW signal.

Models generally fall between two extremes: 'models of the tissue' and 'models of the signal'. Models of the tissue⁶⁻¹⁷ describe the underlying tissue microstructure in each voxel explicitly with a multi-compartment approach.^{18,19,20} Models of the signal focus on describing the DW signal attenuation without describing the underlying tissue composition that gives rise to the signal explicitly.²¹⁻²⁹ Other approaches fall between these two classes and include some features of the tissue, such as the distribution of fibre orientations, but often describe the signal from individual fibres without modelling the fibre composition explicitly.³⁰⁻⁴⁰

Despite this explosion of DW-MRI models, a broad comparison on a common dataset and within a common evaluation framework is lacking, so little is understood about which models are more plausible representations or explanations of the signal. Panagiotaki et al.¹⁸ established a taxonomy of diffusion compartment models and compared 47 of them using data from the fixed corpus callosum of a rat acquired on a pre-clinical system. Later, Ferizi et al.³⁹ performed a similar experiment using data from a live human subject, while Ferizi et al.^{41,42} explored a different class of models, which aim to capture fibre dispersion. Rokem et al.⁴³ compared two classes of models using cross-validation and test-retest accuracy. All these studies^{18,43,44} aim to evaluate variations with specific classes of models with all other variables of the parameter estimation pipeline (i.e. noise model, fitting routine, etc.) fixed. While this provides fundamental insight into which compartments are important in compartment models, questions remain about the broader landscape of models; in particular, which classes of models explain the signal best and how strongly performance depends on the choice of parameter-estimation procedure.

Publicly organized challenges provide a unique opportunity to bring a research community together to gain a quantitative and unbiased comparison of a diverse set of methods applicable to a particular data-processing task. Such publicly organized challenges have helped to establish a common ground for the evaluation of competing methods in a variety of imaging-related tasks, e.g. in brain MR image registration⁴⁵ and segmentation.⁴⁶ In DW-MRI, public challenges have focused on recovering synthetic intra-voxel fibre configurations⁴⁷ or evaluating tractography techniques^{48,50} and have been very successful at driving research and translation forward. Another interesting comparison of reconstruction methods using DW-MRI data was based on the signal acquired from a physical phantom.⁴⁹ Here we report on such a community-wide challenge to model the variation of DW-MRI signals at the voxel level in the in vivo human brain.

Modelling the diffusion signal is a key step in realizing practical and reliable quantitative imaging techniques based on diffusion MRI. The challenge in the area is to extract the salient features from the diffusion signal and relate them to the principal features of the underlying tissue (e.g. in the case of brain white matter (WM) the fibre orientation, axonal packing and axonal size). Three distinct questions arise.

- i. Given the richest possible dataset that samples the space of achievable measurements as widely as possible, which mathematical model can capture best the intrinsic variation of the acquired signal?

*Uran Ferizi, Benoit Scherrer and Torben Schneider joint first co-authors.

Abbreviations used: CT, computerized tomography; CV, cross-validation; DBF, diffusion basis function; DF, diffusion function; DT, diffusion tensor; DTI, diffusion tensor imaging; DW, diffusion-weighted; EN, elastic net; ISBI, International Symposium on Biomedical Imaging; LASADD, Linear Acceleration of Sparse and Adaptive Diffusion Dictionary; LS, least-squares; MRI, magnetic resonance imaging; PDD, principal diffusion directions; RSI, restriction spectrum imaging; ROI, region of interest; SFM, sparse fascicle model; SNR, signal-to-noise ratio; SSE, sum of squared errors; TE, echo time; WM, white matter.

- ii. Which tissue features can be derived from the model?
- iii. What subset of those features can we estimate from limited acquisition time on a standard clinical scanner and what dataset best supports such estimates?

The intuition gained from (i) is generalizable over a wide range of applications, while (ii) and (iii) are highly dependent on the MRI study design and the available hardware. Therefore, our challenge focuses on question (i), as an understanding of (i) is necessary to inform (ii) and (iii). To that end, we acquire the richest possible dataset using the most powerful hardware available and the most motivated subject available (UF). Specifically, we use the Connectome scanner,⁵¹ which is unique among human scanners in having 300 mT/m gradients, rather than 40 mT/m as is typical of state-of-the-art human scanners. Preclinical work by Dyrby et al.¹³ highlights the benefits of such strong gradients and the first results from the Connectome scanner^{42,52-54} are now starting to verify those findings.

This kind of model comparison, based on prediction error, is a common and crucial part of the development of any statistical model-based estimation applications. Burnham and Anderson⁵⁵ explain how and why such comparisons should be performed to reject models that are theoretically plausible but not supported by the data. To that end, we used a uniquely rich dataset acquired on the Connectome system⁴² composed of around 5000 points in q space with, for each shell, a unique combination of gradient strength, diffusion time, pulse width and echo time. This offers the opportunity for the comparison of the many different types of models within a common framework, over a very wide range of measurement space. Using this rich dataset, we organized the White Matter Modeling challenge, held during the 2015 International Symposium on Biomedical Imaging (ISBI) in New York. The goal of the challenge was to evaluate and compare the models in two different tissue configurations that are common in the brain: (1) a WM region of interest where fibres are relatively straight and parallel, specifically the genu of the corpus callosum; and (2) a region in which the fibre configuration is more complex, specifically the fornix. Challenge participants had access to three-quarters of each whole dataset; the participating models were evaluated on how well they predicted the remaining 'unseen' part of the data. As announced before the challenge, the final ranking was based exclusively on the performance on the genu data. In this article, however, we include results from both the genu and the fornix.

The article is organized as follows. We first describe in section 2 the experimental protocol, data post-processing and preparation of the training and testing data for the challenge. We then present the methods for ranking the models and tabulate the various models involved in the competition succinctly. We report the challenge results in section 3 and discuss these results in section 4; a more detailed description of the models follows in the Appendix.

2 | MATERIAL AND METHODS

2.1 | The complete experiment protocol

One healthy volunteer was scanned over two non-stop 4 h sessions. The imaged volume comprised twenty 4 mm thick whole-brain sagittal slices covering the corpus callosum left-right. The image size was 110×110 and the in-plane resolution $2 \times 2 \text{ mm}^2$. 45 unique and evenly distributed diffusion directions (taken from <http://www.camino.org.uk>) were acquired for each shell, with both positive and negative polarities; these directions were the same in each shell. We also included 10 interleaved $b = 0$ measurements, leading to a total of 100 measurements per shell. Each shell had a unique combination of $\Delta = \{22, 40, 60, 80, 100, 120\}$ ms, $\delta = \{3, 8\}$ ms and $|G| = \{60, 100, 200, 300\}$ mT/m (see Table 1). The measurements were randomized within each shell, whereas the gradient strengths were interleaved. We inspected the images visually and did not observe any obvious shifts from gradient heating. The minimum possible echo time (TE) for each gradient duration and diffusion time combination was chosen to enhance signal-to-noise ratio (SNR) and potential estimation of compartment-specific relaxation constants. The SNR of $b = 0$ images was 35 at TE = 49 ms and 6 at TE = 152 ms. The SNR was computed by assessing the signal mean and noise variance across the selected WM voxels on multiple $b = 0$ images. In both cases these estimates matched reasonably well. More details about the acquisition protocol can be found in Ferizi et al.⁴²

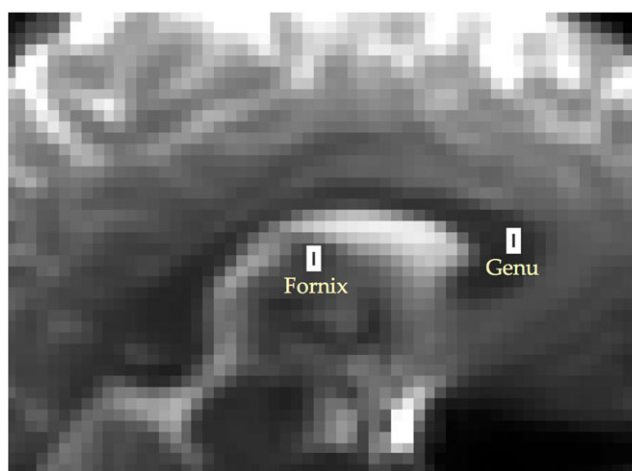
2.2 | Post-processing

All post-processing was performed using Software Library (FSL).⁵⁶ The DW images were corrected for eddy current distortions separately for each combination of δ and Δ using FSL's *Eddy* module (www.fmrib.ox.ac.uk/fsl/eddy) with its default settings. The images were then co-registered using FSL's *FNIRT* package. As the 48 shells were acquired across a wide range of TEs, over two days, we chose to proceed in two steps. First, within each quarter of the dataset (different day, different δ) we registered all the $b = 0$ images together. We then applied these transformations to their intermediary DW images, using a trilinear resampling interpolation. The second stage involved co-registering the four different quarters. To help the co-registration, especially between the two days images that required some through-plane adjustment as well, we omitted areas of considerable eddy-current distortions by reducing the number of slices from 20 to 5 (i.e. leaving two images either side of the mid-sagittal plane) and reducing the in-plane image size to 75×80 .

TABLE 1 The scanning protocol used, acquired in ~8 hours over two non-stop sessions. The protocol has 48 shells, each with 45 unique gradient directions ('blip-up-blip-down')

Acquisition Protocol									
Nr	$\delta = 3ms$				Nr	$\delta = 8ms$			
	Δ (ms)	TE (ms)	G (mT/m)	b (s/mm ²)		Δ (ms)	TE (ms)	G (mT/m)	b (s/mm ²)
1	22	49	61	50	25	22	58	58	300
2	22	49	86	100	26	22	58	95	800
3	22	49	192	500	27	22	58	190	3,200
4	22	49	285	1,100	28	22	58	275	6,700
5	40	67	63	100	29	40	72	59	600
6	40	67	100	250	30	40	72	100	1,700
7	40	67	200	1,000	31	40	72	200	6,850
8	40	67	289	2,100	32	40	72	292	14,550
9	60	87	63	150	33	60	92	34	300
10	60	87	103	400	34	60	92	100	2,650
11	60	87	199	1,500	35	60	92	200	10,500
12	60	87	290	3,200	36	60	92	292	22,350
13	80	107	63	200	37	80	112	61	1,300
14	80	107	99	500	38	80	112	100	3,550
15	80	107	201	2,050	39	80	112	200	14,150
16	80	107	291	4,300	40	80	112	292	30,200
17	100	127	63	250	41	100	132	60	1,600
18	100	127	101	650	42	100	132	100	4,450
19	100	127	200	2,550	43	100	132	200	17,850
20	100	127	291	5,400	44	100	132	292	38,050
21	120	147	63	300	45	120	152	60	1,950
22	120	147	99	750	46	120	152	100	5,350
23	120	147	199	3,050	47	120	152	200	21,500
24	120	147	291	6,500	48	120	152	292	45,900

Note. We provide signal for the parts of protocol marked in black. In red is the protocol for which the signal needs to be predicted.

**FIGURE 1** We only consider two ROIs, each containing six voxels from the genu in the corpus callosum, where the fibres are approximately straight and parallel, and from the fornix, where the configuration of fibres is more complex

2.3 | Training and testing data

The data for this work originated from two regions of interest (ROIs), each containing 6 voxels (see Figure 1). The first ROI was selected in the middle of the genu in the corpus callosum, where the fibres are mostly straight and coherent. The second ROI's fibre configuration is more complex: it lies in the body of fornix, where two bundles of fibres bend and bifurcate.

The dataset was split into two parts: the training dataset and the testing dataset. The training dataset was fully available for the challenge participants. The testing dataset was retained by one of the organizers (UF). The DW signal of the training dataset (36 shells, with acquisition parameters shown in black in Table 1) was provided together with the gradient scheme on the challenge website (<http://cmic.cs.ucl.ac.uk/wmmchallenge/>). This data was used by the participants to estimate their DW-MRI model parameters. The signal attenuation in the testing dataset (12 shells, with acquisition parameters shown in red in Table 1) was kept unseen. It contained one shell, chosen at random, from each TE-specific set of four shells (i.e. of the same combination of δ and Δ). The challenge participants were then asked to predict the signal for the corresponding gradient scheme. They were free to use as much or as little of the training data provided as they wished to predict the signal of the test dataset for the six voxels in each ROI.

Figure 2 shows the DW signal attenuation for each shell in the genu dataset, with stars in the legend indicating which shells were left out for testing. In this plot, a small number of data appear as 'outliers' (two such data are shown with arrows in the bottom-left subplot of Figure 2). Specifically, we counted about 10 of them among more than 4812 measurements, most of them being in the $b = 300 \text{ s/mm}^2$ shell. Since these outliers appear to be specific to the $b = 300 \text{ s/mm}^2$ shell and are not in other shells with similar b value, we attribute them to a momentary twitching of the subject rather than more systematic effects, such as perfusion.

Similarly, Figure 3 shows the signal for the fornix region, with the signal over the six voxels averaged out.

2.4 | Model ranking

Models were evaluated and ranked based on their ability to predict the unseen DW signal accurately. Specifically, the metric used was the sum of square differences between the hidden signal and the predicted signal, corrected for Rician noise:⁵⁷

$$SSE = \frac{1}{N} \sum_{i=1}^N \frac{(\tilde{S}_i - \sqrt{S_i^2 + \sigma^2})^2}{\sigma^2} \quad (1)$$

where N is the number of measurements, \tilde{S}_i is the i th measured signal, S_i its prediction from the model and σ the noise standard deviation.

2.5 | Competing models

Here we give a short summary of the competing models. Additionally, Table 2 provides a summary of their key characteristics. More details are included in the Appendix.

- *Ramirez-Manzanares*: a dictionary-based technique that accounts for multiple fibre bundles and models the distribution of tissue properties (axon radius, parallel diffusivity) and the orientation dispersion of fibres.
- *Nilsson*: a multi-compartment model that models isotropic, hindered and restricted diffusion and accounts for varying (T_1 , T_2) relaxation times for each compartment.⁵⁸
- *Scherrer*: a multi-compartment model in which each compartment is modelled by a statistical distribution of 3-D tensors.¹⁶
- *Ferizi₁* and *Ferizi₂*: two three-compartment models that account for varying T_2 relaxation times for each compartment. As regards the intracellular compartment, *Ferizi₁* models the orientation dispersion by using dispersed sticks as one compartment; *Ferizi₂* uses a single radius cylinder instead.⁴²
- *Poot*: a three-compartment model comprising an isotropic diffusion compartment, a tensor compartment and a model-free compartment in which an Apparent Diffusion Coefficient (ADC) is estimated for each direction independently. T_2 relaxation times are also estimated for each compartment.⁵⁹
- *Rokem*: a combination of the sparse fascicle model⁴³ with restriction spectrum imaging⁶⁰ that describes the signal arising from a multi-compartment model in a densely sampled spherical grid, using L1 regularization to enforce sparsity.
- *Eufrazio*: an extension of the Diffusion Basis Function (DBF) model that accounts for multiple b -value shells.
- *Loya-Olivas₁* and *Loya-Olivas₂*: two models based on the Linear Acceleration of Sparse and Adaptive Diffusion Dictionary (LASADD) technique. *Loya-Olivas₁* uses the DBF signal model, while *Loya-Olivas₂* uses a three-compartment tissue model. The optimization uses linearized signal models to speed up computation and sparseness constraints to regularize.
- *Alipoor*: a model of fourth-order tensors, corrected for T_2 -relaxation across different shells. A robust LS fitting was applied to mitigate influence of outliers.
- *Sakaie*: a two-compartment model of restricted and hindered diffusion with angular variation. A simple exclusion scheme based on the $b = 0$ signal intensity was applied to remove outliers.
- *Fick*: a spatio-temporal signal model to represent 3-D diffusion signal simultaneously over varying diffusion time. Laplacian regularization was applied during the fitting.⁶¹
- *Rivera*: a regularized linear regression model of diffusion encoding variables. This is intentionally built as a simplistic model to be used as a baseline for model comparison.

While the challenge organizers also had competing models (*Ferizi₁*, *Ferizi₂* and *Scherrer*), only *Ferizi* had access to the hidden data. The hidden data were never used to tune the results of his models.

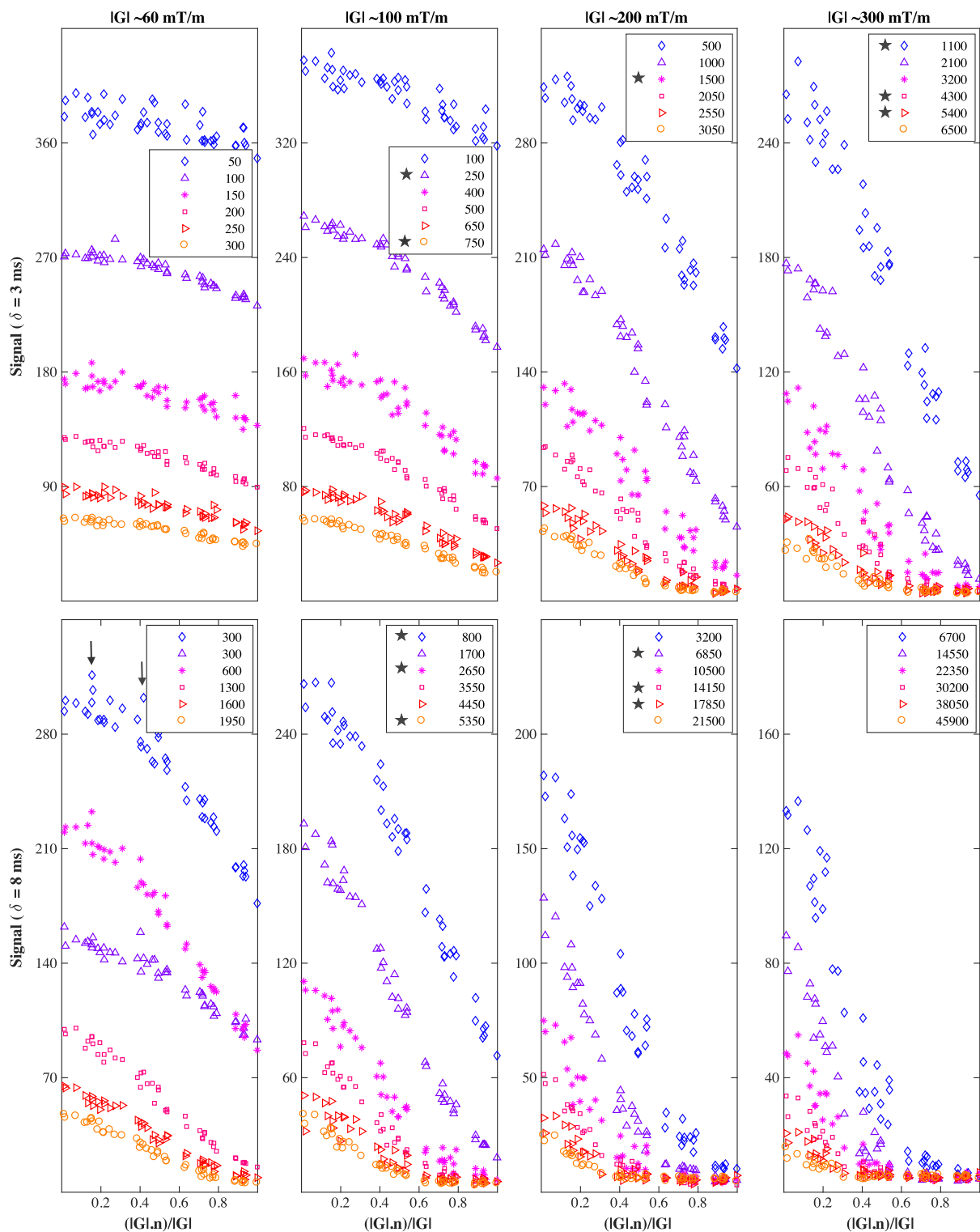


FIGURE 2 Diffusion-weighted signal from the genu ROI, averaged over the six voxels. Across each column and row, the signal pertains to one of the gradient strengths or pulse times δ used; in each subplot, the six shells shown in different colours are Δ -specific, increasing in value (22, 40, 60, 80, 100, 120 ms) from top to bottom. Inside the legend, the b value is in s/mm^2 units; here, the HARDI shells kept for testing are those marked with a star; the remaining shells comprise the training data. On the x-axis is the cosine of the angle between the applied diffusion gradient vector \mathbf{G} and the fibre direction \mathbf{n} . Some models in this study omit data outliers; two such data points are shown in the bottom-left subplot with vertical arrows — obviously each model has its own criteria for determining the outliers

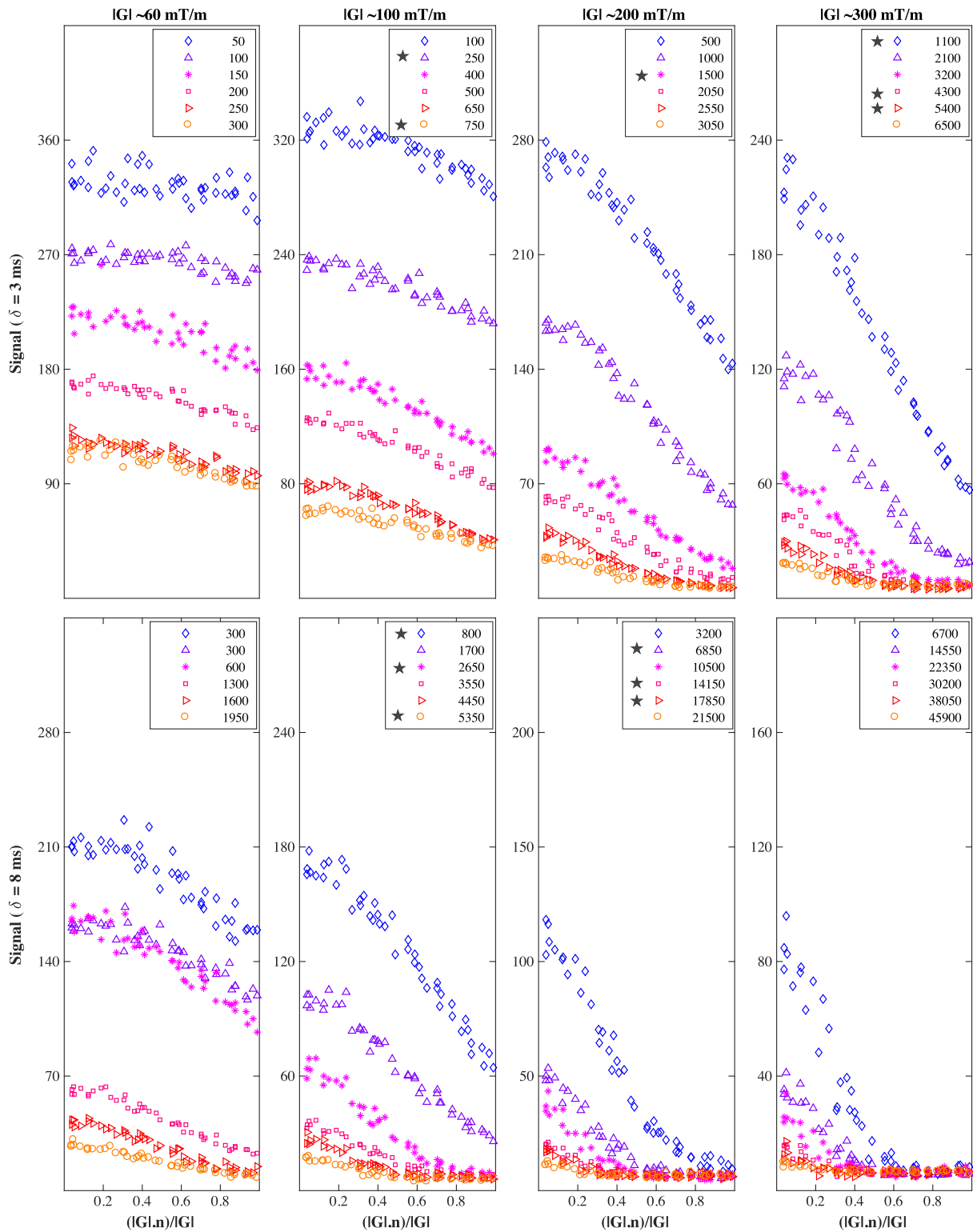


FIGURE 3 Diffusion-weighted signal from the fornix ROI, averaged over the six voxels. The legend's b value is in s/mm^2 units. Testing shells are marked with a star. On the x-axis is the cosine of the angle between the applied diffusion gradient vector \mathbf{G} and the fibre direction \mathbf{n}

3 | RESULTS

Figure 4 shows the averaged prediction error in each ROI (top subplot is for the genu, bottom subplot is for the fornix) and the corresponding overall ranking of the participating models in the challenge. The first six models in the genu ranking performed similarly, each higher ranked model marginally improving on the prediction error. The prediction error clearly increased at a higher rate for the subsequent models. In the fornix dataset, the prediction error was higher than in the genu. For both datasets, the first six models were the same, albeit permuted. Most of the models performed

TABLE 2 Summary of the various diffusion models evaluated. Tissue models are models that include an explicit description of the underlying tissue microstructure with a multi-compartment approach. In contrast, signal models focus on describing the DW signal attenuation without explicitly describing the underlying tissue and instead correspond to a 'signal processing' approach

	Type of model	Nb of free param. (genu/fornix)	Models effect of δ and Δ	Noise assumption	Optimization algorithm	Outliers strategy	Special signal prediction strategy
R-Manzanares	Tissue	N/A	Yes	Gaussian	weighted-LS bootstrapping	Yes	CV
Nilsson	Tissue	< 12/12	Yes	Gaussian	LM	Yes	CV
Scherrer	Tissue	10/16	No	Gaussian	Bobyqa	Yes	No
Ferizi_1	Tissue	< 12/12	Yes	approx.-Rician	LM	No	No
Ferizi_2	Tissue	< 10/10	Yes	approx.-Rician	LM	No	No
Alipoor	Signal	17/17	No	Gaussian	weighted-LS	Yes	No
Sakaie	Signal	N/A	No	Gaussian	nonlinear-LS	Yes	No
Rokem	Tissue	~20	No	Gaussian + Noise floor	Elastic net	No	CV
Eufracio	Tissue	7/7	No	Gaussian	bounded-LS Lasso, Ridge	No	No
Loya-Olivas_1	Tissue	11	No	Gaussian	bounded-LS & Lasso	No	No
Loya-Olivas_2	Tissue	11	No	Gaussian	bounded-LS	No	No
Poot	Signal	103	No	Rician	LM-like	No	No
Fick	Signal	475	Yes	Gaussian	Laplacian-reg-LS	No	partial-CV
Rivera	Signal	23	Yes	Gaussian	Weighted Lasso	Yes	CV

Abbreviations: LS=least-squares, LM=Levenberg–Marquardt, CV=cross-validation, reg=regularized

similarly in terms of ranking in both genu and fornix cases, i.e. Nilsson (second in genu/first in fornix), Scherrer (third/second) and Ferizi_2 (fourth/fourth). Others performed significantly better in one of the cases, with Ramirez-Manzanares (first/sixth) being the most notable.

Figure 4 also details the prediction error for different ranges of b values in the unseen dataset. Models inevitably vary in their prediction capabilities; some models perform better within a given b -value range but are penalized more in another. Across the models, as the figure shows, the ranking between models was dominated by the signal prediction accuracy for b values between 750 and 1400s/mm²; specifically, the shell that has the largest weight on this error is the $b = 1100$ s/mm² one. The top-ranking models, nevertheless, were better at predicting the signal for higher b -value images as well. The prediction performance of lower b -value images (<750s/mm²) in the genu was less consistent across ranks. For example, the models of Rokem and Sakaie outperformed most of the higher ranking models in this low b -value range. The fornix is a more complex region than the genu, hence the performance across the shells is less consistent. In the fornix, the prediction errors were generally larger than in the genu across all b values for all models, except Rivera's, which showed the opposite effect. The prediction errors of the $b = 0$ images were also larger than in the genu, especially for the highly ranked models of Poot and Ferizi. The prediction errors in other b -value shells followed the overall ranking of the models more closely.

Figure 5 shows the prediction error for each voxel independently. In the genu plot, the best performing models had high consistency of low prediction errors across all individual voxels. These were followed by the models with consistent larger prediction error in all voxels. Most of the lowest ranking models not only had largest prediction errors, they also showed large variations in prediction performance. For example, while the model of Loya-Olivas₂ was competitive in voxel 5, it ranked low due to large prediction errors in voxels 4 and 6. The results in the fornix show a lower consistency of prediction errors between the voxels than in the genu. Specifically, two voxels (3 and 4) showed substantially larger prediction errors and were likely responsible for much of the overall ranking.

Finally, we report in Figures 6–8 and 9 an illustration of the quality of fit of each model to four representative shells, including the $b=1100$ s/mm² shell mentioned above; Figures 6 and 7 concern the genu data and Figures 8 and 9 are for fornix data.

4 | DISCUSSION

The challenge set out to compare the ability of various kinds of models to predict the diffusion MR signal from WM over a very wide range of measurement parameters – exploring the boundaries of possible future quantitative diffusion MR techniques. The 14 challenge entries were a good representation of the many available models that are proposed in the literature. The acquired data aimed to cover the broadest spectrum of experimental parameters possible. The participating models use a variety of fitting routines and modelling assumptions, providing additional insight into the effects of algorithmic and modelling choices during parameter estimation. Although the set of methods tested is not sufficient to make a full

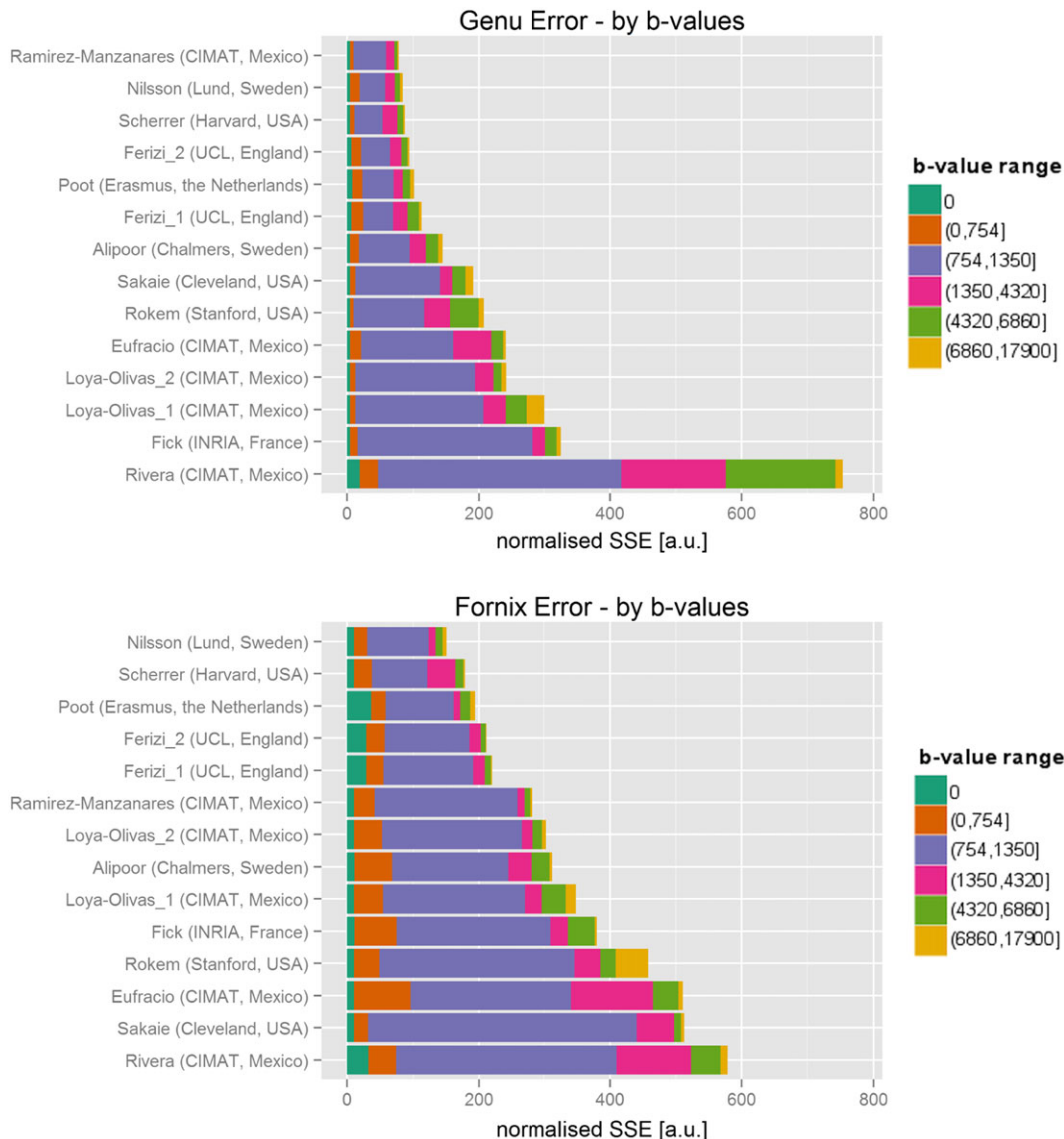


FIGURE 4 Overall ranking of models by sum-of-squared-errors (SSE) metric over all voxels in genu (top) and fornix (bottom) ROIs. The colors represent different ranges of b -value shells

comparison of each independent feature (diffusion model, noise model, fitting routine, etc.) and the number of combinations prohibits an exhaustive comparison, the results of the challenge do reveal some important trends.

In contrast with earlier model comparisons,^{18,43,44} the results provide new insight into which broad classes of model explain the signal best and what features of the estimation procedure are important. This information is very timely, as recent model-based diffusion MRI techniques, such as NODDI,¹⁵ SMT,^{17,40} DIAMOND,¹⁶ DKI⁶² and LEMONADE,⁶³ are starting to become widely adopted in clinical studies and trials. Despite their success, intense debate continues in the field about applicability of different models and fitting routines.^{64,65} The insights from this challenge provide key pointers to the important features of the next-generation of front-line imaging techniques of this type. Moreover, the data and evaluation routines remain available to form the basis of an expanding ranking of models and fitting routines and a benchmark for future model development.

4.1 | Main conclusions

The first insight is on the type of model used. Signal models do not necessarily outrank tissue models; indeed, using our dataset, models of the signal (Alipoor, Sakaie, Fick, Rivera) ranked on average lower than models of the tissues, despite their theoretical ability to offer more flexibility in describing the raw signal. This is quite surprising, as the current perception within the field is that, generally, we can capture the signal variation much better through a functional description of the signal (signal models) rather than via a biophysical model of the tissue (tissue models). The former generally consist of bases of arbitrary complexity, whereas the latter are generally very parsimonious models that rely on extremely crude descriptions of tissue (e.g. white matter as parallel impermeable cylinders). The results suggest that the flexibility of signal models can rapidly lead to overfitting.

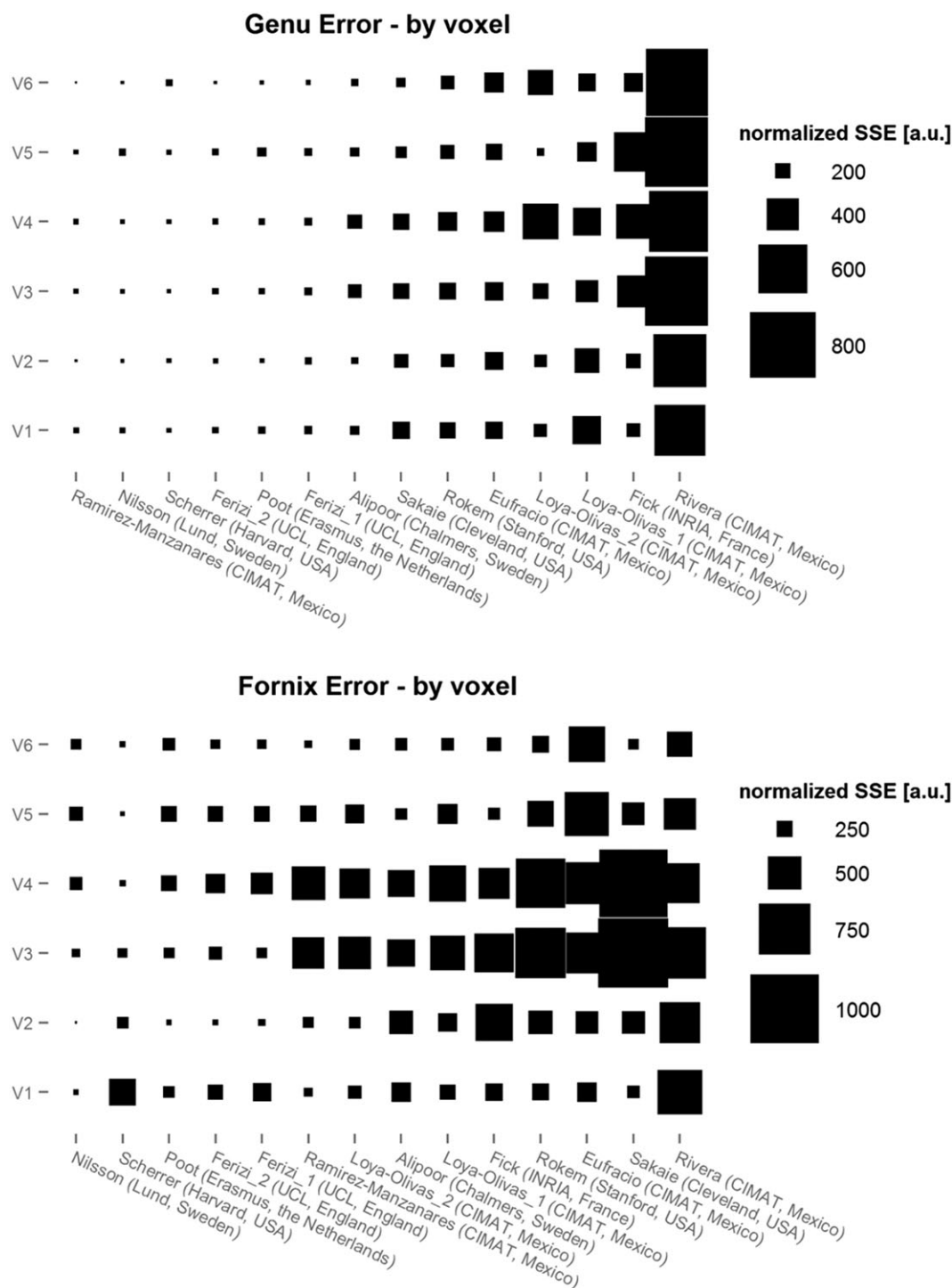


FIGURE 5 Sum-of-squared-errors (SSE) per voxel for each model in genu and fornix. The size of rectangles represent the SSE value per voxel

However, the tissue models can explain the signal relatively well even with just a few parameters (compare the quality-of-fit plots of the Rivera model in Figure 7 with the signal prediction of the top models in Figure 6: the higher the b value, the worse the prediction of the linear signal model). Certain underlying assumptions may cause the signal models to perform less well than expected. For example, they are often designed to work with data with a single diffusion time and do not generalize naturally to incorporate the additional dimension (although see Fick et al.⁶¹ for some steps towards generalization). Many of the tissue models, on the other hand, naturally account for finite δ , varying diffusion times and gradient strength (e.g. the Ramirez-Manzanares, Nilsson and Ferizi models in our collection). We cannot draw any conclusion about the benefits of an adjustable number of parameters in a model, because of the limited number of models in our study that do this and because the models differ in a range of other aspects.

The second insight concerns the choice of noise modelling. Despite the fact that SNR at $b = 0$ and $TE = 152$ ms falls to about 6, use of the Rician noise model does not appear to be a significant benefit in predicting unseen signal; here, however, we do not investigate the effect on estimated

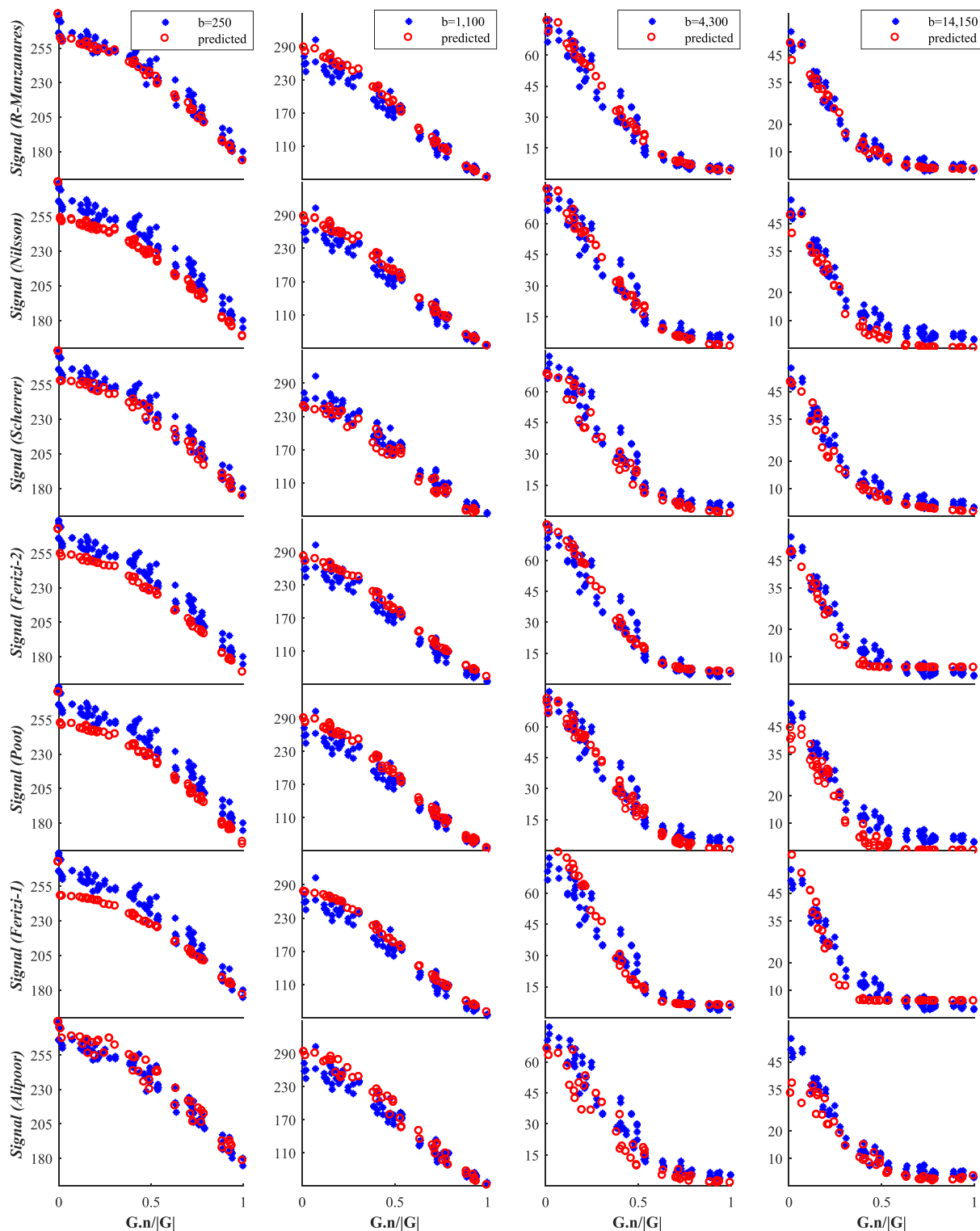


FIGURE 6 Genu signal for the group consisting of the best seven from 14 models. We show only four (of twelve) representative shells; these are shown by blue stars, while red circles denote the model-predicted data. The best models are listed first. The x-axis is the cosine of the angle between \mathbf{G} and \mathbf{n}

model parameters, which may still benefit from the more accurate noise model. In this challenge, most participants used non-linear least-squares or maximum-likelihood optimization. Additional regularization of the objective function (Eufrazio & Rivera/Lasso, Rokem/Elastic Net, Fick/Laplacian) appeared to have little benefit over non-regularized optimization.

The third observation is about removing signal outliers. Five of the eleven models preprocessed the training data by clearing out outliers, including the top two models. We tried this procedure with two good models that did not use such a procedure, Ferizi₁ and Ferizi₂, and observed that it did not affect the ranking, though it did improve the prediction error marginally. This is understandable, considering the relatively little weight

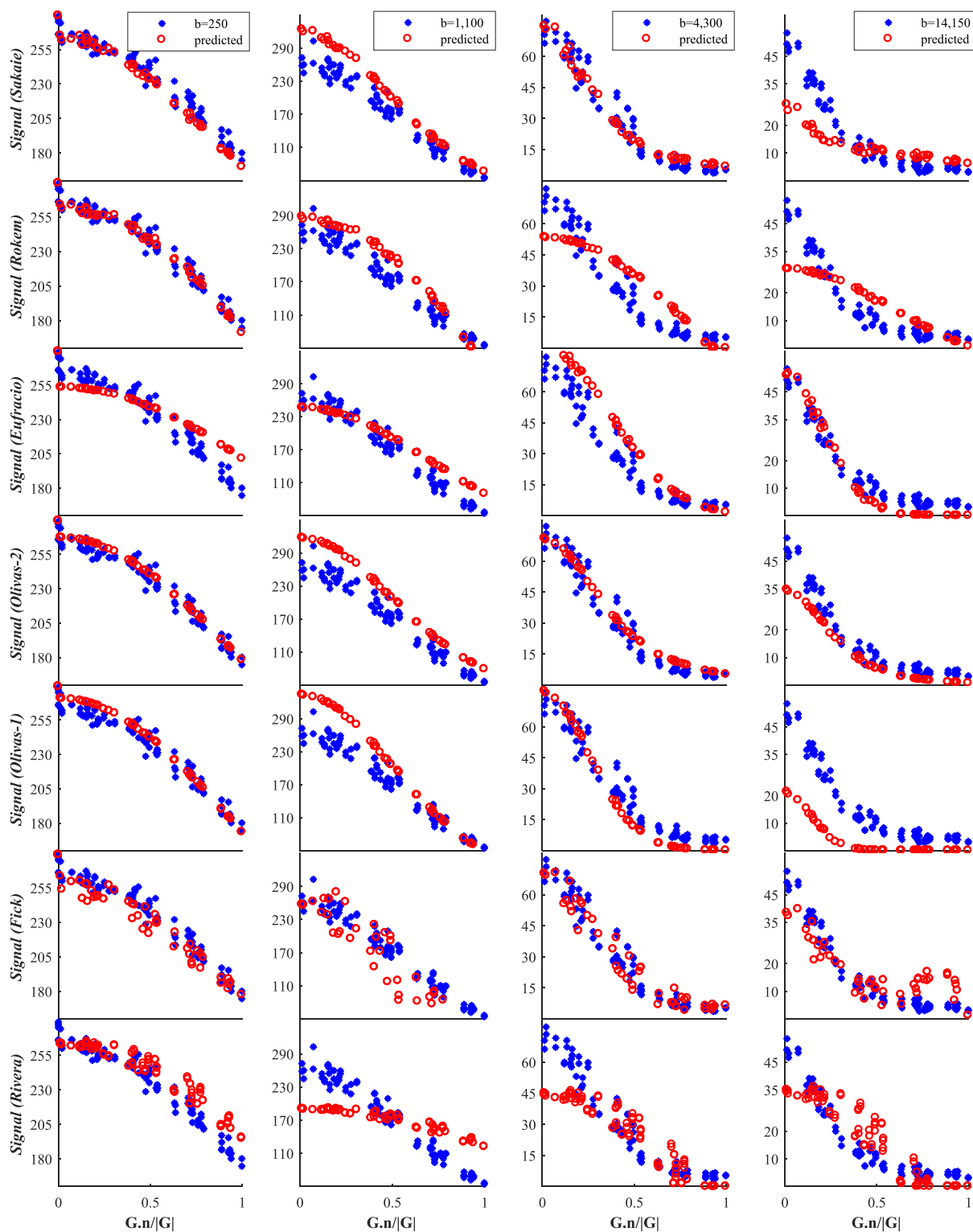


FIGURE 7 Genu signal for the second group of 14 models. Raw testing data are shown by blue stars, while red circles denote the model-predicted data. The x-axis is the cosine of the angle between \mathbf{G} and \mathbf{n}

these apparent outliers have on the total number of measurements (10 points from a 4812-strong dataset). Additionally, specific strategies for predicting the signal, e.g. bootstrapping or cross-validation, as used by the top two models of Ramirez-Manzanares and Nilsson, may also help the model ranking.

4.2 | Limitations and future directions

Although this challenge provides several new insights into the choice of model and fitting procedure for diffusion-based quantitative imaging tools, it has a number of limitations that future challenges might be designed to address. One limitation of the study is that we use a very rich acquisition

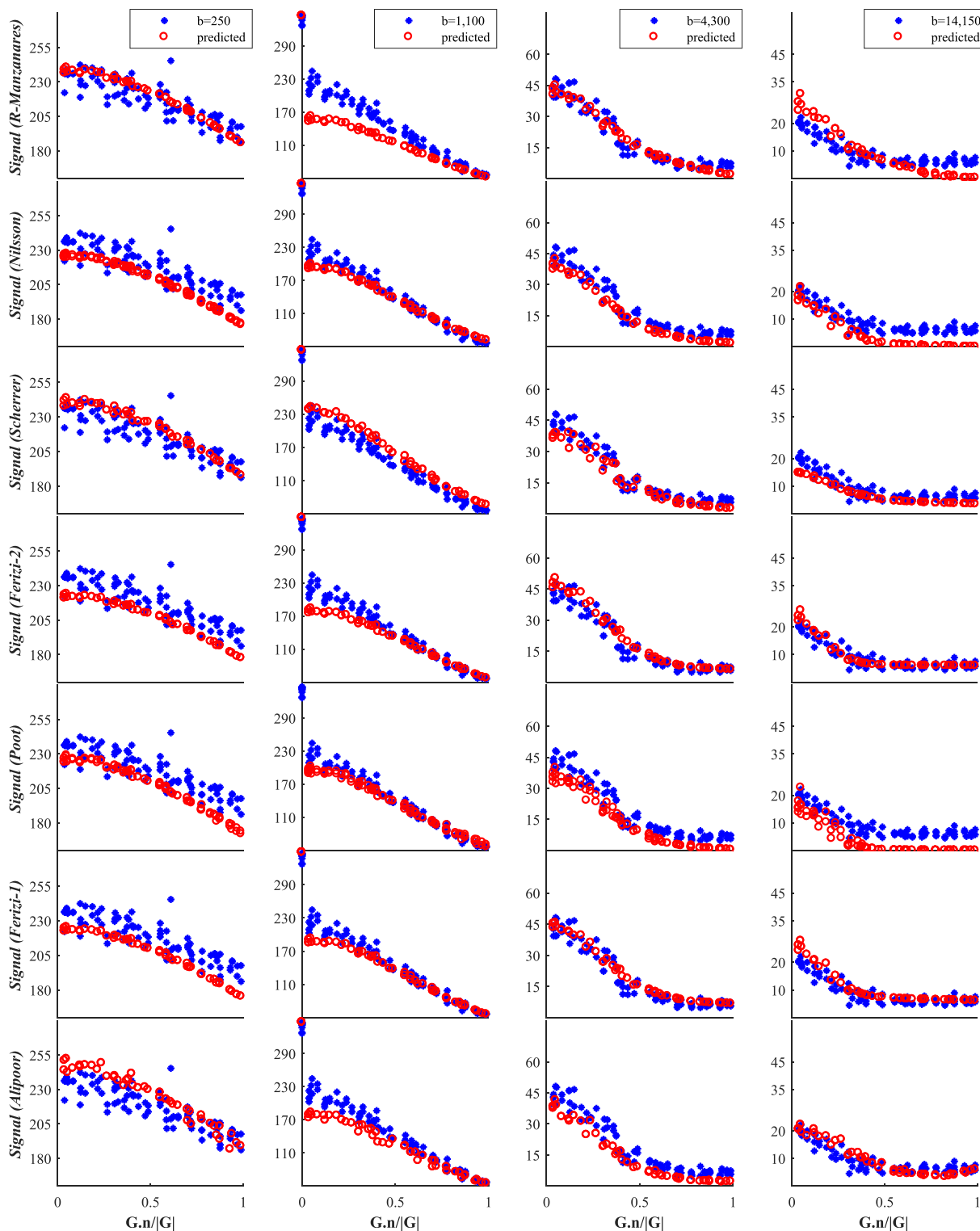


FIGURE 8 Fornix signal for the group consisting of the best 7 from 14 models. We show only four (of twelve) representative shells; these are shown by blue stars, while red circles denote model-predicted data. The best models are listed first. The x-axis is the cosine of the angle between G and n

protocol that is not representative of common or clinical acquisition protocols. In particular, we cover a very wide range of b values and the data acquisition (protocol) we use consists of many TEs, unlike many other multi-shell diffusion datasets that use a fixed TE. As stated in the Introduction, our intention is to sample the measurement space as widely as possible to support the most informative models possible. Varying the TE makes it possible to probe compartment-specific T_2 (the decay of which Ferizi et al.⁴² finds to be monoexponential at the voxel level), an investigation that would be impossible with a single TE. However, the good performance of DIAMOND also shows that a model with fixed δ and Δ can still capture the signal variation in multi-TE datasets and that, while the majority of the full data was ignored in each of the reconstructions, its prediction error compared favourably with other techniques.

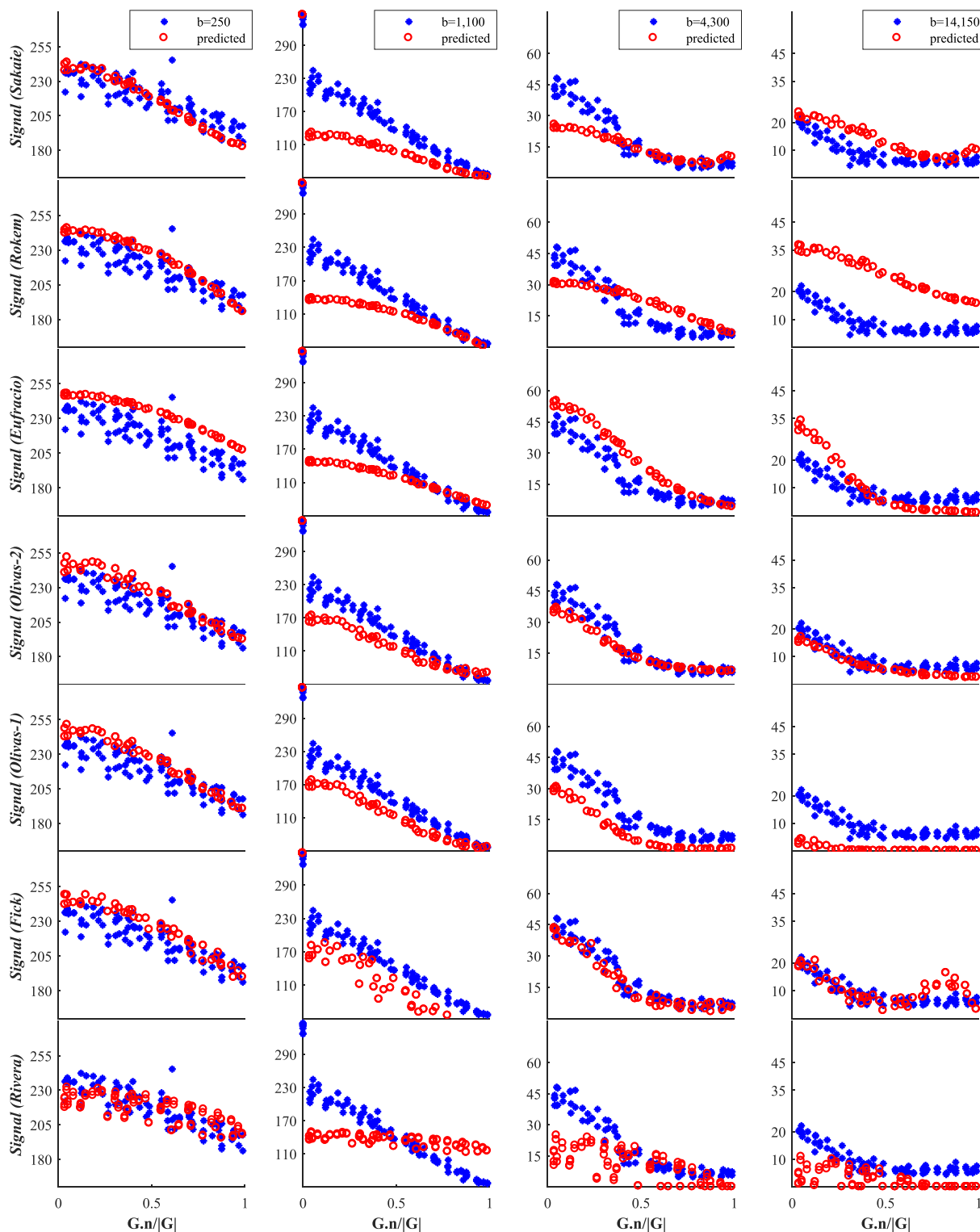


FIGURE 9 Fornix signal for the second group of 14 models. Raw testing data are shown by blue stars, while red circles denote the model-predicted data. The x-axis is the cosine of the angle between \mathbf{G} and \mathbf{n}

We use the unique human Connectome scanner⁵¹ to acquire a dataset with gradients of up to 300mT/m, which is not readily available in most current MR machines. However, previous preclinical work by Dyrby et al.¹³ suggests that high diffusion gradients enrich the signal, which helps model fitting and comparison. Future challenges might be designed that focus on explaining the signal and estimating parameters from data more typical of clinical acquisitions.

Assessing the prediction performance on unseen data as in this challenge is different from assessing the fitting error: it implicitly penalizes models that overfit the data. However, since most of the missing shells lie in between other shells (in terms of b values, TEs, etc.), the quality of signal extrapolation was not assessed. We get a glimpse of this from Figure 4, where the SSE is unevenly distributed between the b values. Here, the shell

that bore the largest error is the $b = 1100$ s/mm² one; see also Figures 6 and 7. Of all 'unseen' shells, this shell combines the lowest Δ and highest $|G|$, placing it on the edge of the range of the measurement space sampled. Such a b -value shell combines high signal magnitude with high sensitivity, i.e. the gradient of signal against b -value is highest in this range, which makes it hard to predict. (We stress that this observation is in the context of the wider multi-shell acquisition, and is not to be seen in isolation for its potential impact on single-shell acquisition methods.) On the other hand, the variability of prediction errors in the $b < 750$ s/mm² range could arise from the varying sensitivity of different models to the free water component, which is challenging to estimate as it can easily be confounded with hindered water, or physiological effects, which are mostly observable in this low b -value range. Future work can take this further, by selecting unseen shells outside the min–max range of experimental parameters. This is likely to penalize more complex models that overfit the data even more strongly.

We did not take into account the computational demand of each model, and this might limit the generalization of the results. Models that use bootstrapping generally have a higher computational burden and may not be feasible for large datasets, e.g. whole brain coverage.

The dataset used in this challenge is specific to one subject who underwent a long-duration acquisition, which adds to the question of generalizability. The subsequent preprocessing of the data is also a factor to bear in mind: the registration of two 4h datasets, across such a broad range of echo times, poses its own challenges for certain non-homogenous regions in the brain, such as the fornix (compared with, for example, the relatively large genu). Thus the results may be somewhat subject-specific and may be affected by residual alignment errors.

Another limitation is that we only look at isolated voxels inside the corpus callosum and the fornix. Questions still remain about which models are viable even in the most coherent areas of the brain with the simplest geometry, so we believe our focused challenge on well-defined areas is an informative first step necessary before extending the idea to the whole of the white matter, which would make for an extremely complex challenge. We note, however, recent work by Ghosh et al.⁶⁶ that illustrates such an approach with Human Connectome Project (HCP) data.

We focused here on comparing models based on their ability to predict unseen data. Although models that reflect true underlying tissue structure should explain the data well, we cannot infer in general that models that predict unseen data better are mechanistically closer to the tissue than those that do not. As we discuss in the Introduction, the main power of evaluating models in terms of prediction error is to reject models that cannot explain the data. Thus, while the identification of parsimonious models that explain the data certainly has great benefit, further validation is necessary through comparison of the parameters that they estimate with independent measurements, e.g. obtained through microscopy (our challenge makes no attempt to assess the integrity of parameter estimates themselves, but future challenges might use such performance criteria). Models can be evaluated to some extent by sanity checking the realism of their fitted parameter values, as in for example Jelescu et al.⁶⁴ or Burcaw et al.⁶⁷ However, obtaining accurate ground-truth values for quantitative evaluation remains a hard and yet unsolved problem for diffusion MRI in general. In particular, histology can only roughly approximate the in vivo ground truth and introduces its own set of challenges in sample preparation, acquisition and biophysical interpretation.^{12,13,65,68–71} This challenge highlights the need for improved model comparison and validation methods.

5 | CONCLUSION

Challenges such as this have great value in bringing the community together and provide an unbiased comparison of wide-ranging solutions to key data-processing problems. They raise new insights and ideas, motivating more directed future studies. The data are publicly available for others to use, with more details of the dataset given on the Challenge website at <http://cmic.cs.ucl.ac.uk/wmmchallenge/>. On this website, an up-to-date ranking of the models will be available, where additional models can be added after the publication of the article and where the community will be able to evaluate further the impact of noise correction, compartment-specific T_2 estimation, inter-class model assumptions, e.g. tissue versus signal models, or indeed intra-class model assumptions, e.g. whether cylinders or sticks are optimal models for the given dataset.⁴² This will provide an important benchmark for future models and parameter estimation routines.

ACKNOWLEDGEMENTS

Research reported in this manuscript was supported by the following organizations. EPSRC supported this work through grants EP/G007748, EP/L022680/1, EP/I027084/01, EP/M020533/1 and EP/N018702/1. U. Ferizi is also supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) of the National Institute of Health (NIH) under award numbers R21AR066897 and R01AR067789. B. Scherrer was supported in part by NIH R01 NS079788, R01 EB019483, R01 EB018988 and BCH TRP Pilot and BCH CTREC K-to-R Merit Award. M. Nilsson is supported by the Swedish Strategic Research (SSF) Grant AM13-0090. R. Deriche and R. Fick were supported in part by the European Research Council (ERC) under the Horizon 2020 research and innovation program (ERC Advanced Grant agreement No 694665 : CoBCoM).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding bodies (EPSRC or NIH).

Lastly, we thank the ISBI 2015 challenge organizers, in particular Stephen R. Aylward (Kitware Inc., USA) and Badri Roysam (University of Houston, USA).

REFERENCES

1. Basser PJ, Mattiello J, LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson, Ser B*. 1994;103(3):247–254.
2. Assaf Y, Pasternak O. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: A review. *J Mol Neurosci: MN*. 2008;34(1):51–61.

3. Rose SE, Chen F, Chalk JB, et al. Loss of connectivity in Alzheimer's disease: An evaluation of white matter tract integrity with colour coded MR diffusion tensor imaging. *J Neurol, Neurosurg Psychiatry*. 2000;69(4):528–530.
4. Werring D, Brassat D, Droogan A, et al. The pathogenesis of lesions and normal-appearing white matter changes in multiple sclerosis. *Brain*. 2000;123(8):1667–1676.
5. Price S, Jena R, Burnet N, et al. Improved delineation of glioma margins and regions of infiltration with the use of diffusion tensor imaging: An image-guided biopsy study. *Amer J Neuroradiol*. 2006;27(9):1969–1974.
6. Mitra PP, Sen PN, Schwartz LM, Le Doussal P. Diffusion propagator as a probe of the structure of porous media. *Phys Rev Lett*. 1992;68(24):3555.
7. Behrens T, Woolrich M, Jenkinson M, et al. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn Reson Med*. 2003;50(5):1077–1088.
8. Behrens T, Berg HJ, Jbabdi S, Rushworth M, Woolrich M. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*. 2007;34(1):144–155.
9. Assaf Y, Basser PJ. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage*. 2005;27(1):48–58.
10. Jespersen SN, Kroenke CD, Østergaard L, Ackerman JJH, Yablonskiy DA, Modeling dendrite density from magnetic resonance diffusion measurements. *NeuroImage*. 2007;34(4):1473–1486.
11. Jespersen SN, Bjarkam CR, Nyengaard JR, et al. Neurite density from magnetic resonance diffusion measurements at ultrahigh field: Comparison with light microscopy and electron microscopy. *NeuroImage*. 2010;49(1):205–216.
12. Alexander DC, Hubbard PL, Hall MG, et al. Orientationally invariant indices of axon diameter and density from diffusion MRI. *NeuroImage*. 2010;52(4):1374–1389.
13. Dyrby TB, Hall MG, Ptito M, et al. Contrast and stability of the axon diameter index from microstructure imaging with diffusion MRI. *Magn Reson Med*. 2013;70(3):711–721.
14. Sotiropoulos SN, Behrens T, Jbabdi S. Ball and rackets: Inferring fiber fanning from diffusion-weighted MRI. *NeuroImage*. 2012;60(2):1412–1425.
15. 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(4):1000–1016.
16. Scherrer B, Schwartzman A, Taquet M, Sahin M, Prabhu SP, Warfield SK. Characterizing brain tissue by assessment of the distribution of anisotropic microstructural environments in diffusion-compartment imaging (DIAMOND). *Magn Reson Med*. 2016;76(3):963–977. <https://doi.org/10.1002/mrm.25912>
17. Kaden E, Kelm ND, Carson RP, Does MD, Alexander DC. Multi-compartment microscopic diffusion imaging. *NeuroImage*. 2016;139:346–359.
18. Panagiotaki E, Schneider T, Siow B, Hall MG, Lythgoe MF, Alexander DC. Compartment models of the diffusion MR signal in brain white matter: A taxonomy and comparison. *NeuroImage*. 2012;59(3):2241–2254.
19. Stanisz GJ, Wright GA, Henkelman RM, Szafer A. An analytical model of restricted diffusion in bovine optic nerve. *Magn Reson Med*. 1997;37(1):103–111.
20. Nilsson M, van Westen D, Ståhlberg F, Sundgren PC, Lätt J. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. *Magn Reson Mater Phys, Biol Med*. 2013;26(4):345–370.
21. Liu C, Bammer R, Moseley ME. Generalized diffusion tensor imaging (GDTI): A method for characterizing and imaging diffusion anisotropy caused by non-Gaussian diffusion. *Isr J Chem*. 2003;43(1-2):145–154.
22. Tuch DS. Q-ball imaging. *Magn Reson Med*. 2004;52(6):1358–1372.
23. Jensen JH, Helpert JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med*. 2005;53(6):1432–1440.
24. Descoteaux M, Angelino E, Fitzgibbons S, Deriche R. Regularized, fast, and robust analytical Q-ball imaging. *Magn Reson Med*. 2007;58(3):497–510.
25. Assemblal HE, Tschumperlé D, Brun L. Efficient and robust computation of pdf features from diffusion MR signal. *Med Image Anal*. 2009;13(5):715–729.
26. Yablonskiy DA, Sukstanskii AL. Theoretical models of the diffusion weighted MR signal. *NMR Biomed*. 2010;23(7):661–681.
27. Kiselev VG. The cumulant expansion: An overarching mathematical framework for understanding diffusion NMR. In: Derek KJ, ed. *Diffusion MRI: Theory, Methods, and Applications*. OUP USA; 2011:152–168.
28. Özarslan E, Koay C, Basser P. Simple harmonic oscillator based estimation and reconstruction for one-dimensional q-space MR. In: Proc. Intl. Soc. Mag. Reson. Med, Vol. 16; 2008:35.
29. Özarslan E, Koay CG, Shepherd TM, et al. Mean apparent propagator (MAP) MRI: A novel diffusion imaging method for mapping tissue microstructure. *NeuroImage*. 2013;78:16–32.
30. Tournier JD, Calamante F, Gadian DG, Connelly A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*. 2004;23(3):1176–1185.
31. Alexander DC. Maximum entropy spherical deconvolution for diffusion MRI. *Biennial International Conference on Information Processing in Medical Imaging: 19th International Conference, IPMI 2005*. Glenwood Springs, CO, USA. Berlin: Springer; 2005:76–87.
32. Anderson AW. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magn Reson Med*. 2005;54(5):1194–1206.
33. Jian B, Vemuri BC. Multi-fiber reconstruction from diffusion MRI using mixture of Wisharts and sparse deconvolution. *Information Processing in Medical Imaging*. Berlin: Springer; 2007:384–395.
34. Jian B, Vemuri BC, Özarslan E, Carney PR, Mareci TH. A novel tensor distribution model for the diffusion-weighted MR signal. *NeuroImage*. 2007;37(1):164–176.
35. Sakaie KE, Lowe MJ. An objective method for regularization of fiber orientation distributions derived from diffusion-weighted MRI. *NeuroImage*. 2007;34(1):169–176.
36. Dell'Acqua F, Rizzo G, Scifo P, Clarke RA, Scotti G, Fazio F. A model-based deconvolution approach to solve fiber crossing in diffusion-weighted MR imaging. *IEEE Trans Biomed Eng*. 2007;54(3):462–472.
37. Jbabdi S, Sotiropoulos S, Savio A, Graña M, Behrens T. Model-based analysis of multishell diffusion MR data for tractography: How to get over fitting problems. *Magn Reson Med*. 2012;68(6):1846–1855.

38. Rathi Y, Michailovich O, Laun F, Setsompop K, Grant PE, Westin CF. Multi-shell diffusion signal recovery from sparse measurements. *Med Image Anal.* 2014;18(7):1143–1156.
39. Ferizi U, Schneider T, Panagiotaki E, et al. A ranking of diffusion MRI compartment models with in vivo human brain data. *Magn Reson Med.* 2014;72(6):1785–1792.
40. Kaden E, Kruggel F, Alexander DC. Quantitative mapping of the per-axon diffusion coefficients in brain white matter. *Magn Reson Med.* 2016;75(4):1752–1763. <https://doi.org/10.1002/mrm.25734>
41. Ferizi U, Schneider T, Tariq M, Wheeler-Kingshott C, Zhang H, Alexander D. The importance of being dispersed: A ranking of diffusion MRI models for fibre dispersion using in vivo human brain data. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Lecture Notes in Computer Science, vol. 8149. Berlin: Springer; 2013:74–81.
42. Ferizi U, Schneider T, Zhang H, Wheeler-Kingshott CA, Alexander DC. White matter compartment models for in vivo diffusion MRI at 300 mT/m. *Neuroimage.* 2015;118:468–483. <https://doi.org/10.1016/j.neuroimage.2015.06.027>
43. Rokem A, Yeatman JD, Pestilli F, et al. Evaluating the accuracy of diffusion MRI models in white matter. *PLoS One.* 2015;10(4):e0123272.
44. Ferizi U. Compartment Models and Model Selection for in-vivo Diffusion-MRI of Human Brain White Matter [PhD thesis]. London: University College London; 2014.
45. Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage.* 2009;46(3):786–802.
46. Mendrik A, Vincken K, Kuijf H, et al. MRBrainS challenge: Online evaluation framework for brain image segmentation in 3T MRI scans. *Comput Intell Neurosci.* 2015;2015:813696. <https://doi.org/10.1155/2015/813696>
47. Daducci A, Canales-Rodriguez EJ, Descoteaux M, et al. Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI. *IEEE Trans Med Imaging.* 2014;33(2):384–399.
48. Fillard P, Descoteaux M, Goh A, et al. Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. *Neuroimage.* 2011;56(1):220–234.
49. Ning L, Laun F, Gur Y, et al. Sparse reconstruction challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use? *Med Image Anal.* 2015;26(1):316–331.
50. Pujol S, Wells W, Pierpaoli C, et al. The DTI challenge: Toward standardized evaluation of diffusion tensor imaging tractography for neurosurgery. *J Neuroimaging.* 2015;25(6):875–882.
51. Setsompop K, Kimmlingen R, Eberlein E, et al. Pushing the limits of in vivo diffusion MRI for the Human Connectome project. *NeuroImage.* 2013;80:220–233.
52. McNab JA, Edlow BL, Witzel T, et al. The human connectome project and beyond: Initial applications of 300 mT/m gradients. *NeuroImage.* 2013;80:234–245.
53. Duval T, McNab J, Setsompop K, et al. In vivo estimation of axon diameter in the human spinal cord using 300 mT/m gradients. *International Society of Magnetic Resonance in Medicine (ISMRM) 24th Annual Meeting & Exhibition.* Milan, Italy; 2014.
54. Huang SY, Nummenmaa A, Witzel T, et al. The impact of gradient strength on in vivo diffusion MRI estimates of axon diameter. *NeuroImage.* 2015;106:464–472.
55. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Berlin: Springer; 2002.
56. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage.* 2002;17(2):825–841.
57. Jones DK, Basser PJ. Squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted MR data. *Magn Reson Med.* 2004;52(5):979–993.
58. Nilsson M, Lätt J, Ståhlberg F, Westin D, Hagglätt H. The importance of axonal undulation in diffusion MR measurements: A Monte Carlo simulation study. *NMR Biomed.* 2012;25(5):795–805.
59. Poot D, Klein S. Detecting statistically significant differences in quantitative MRI experiments, applied to diffusion tensor imaging. *IEEE Trans Med Imaging.* 2015;34(5):1164–1176.
60. White NS, Leergaard TB, D'Arceuil H, Bjaalie JG, Dale AM. Probing tissue microstructure with restriction spectrum imaging: Histological and theoretical validation. *Hum Brain Mapp.* 2013;34(2):327–346.
61. Fick R, Wassermann D, Pizzolato M, Deriche R. A unifying framework for spatial and temporal diffusion in diffusion MRI. *International Conference on Information Processing in Medical Imaging.* Berlin: Springer; 2015:167–178.
62. Fieremans E, Jensen JH, Helpert JA. White matter characterization with diffusional kurtosis imaging. *Neuroimage.* 2011;58(1):177–188.
63. Novikov DS, Veraart J, Jelescu IO, Fieremans E. Mapping orientational and microstructural metrics of neuronal integrity with in vivo diffusion MRI. arXiv preprint arXiv:1609.09144. 2016 Sep 28.
64. 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.
65. Jelescu IO, Veraart J, Fieremans E, Novikov DS. Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue. *NMR Biomed.* 2016;29(1):33–47.
66. Ghosh A, Zhang H, Alexander DC. To be dispersed or not to be dispersed: A study using HCP data. *International Society of Magnetic Resonance in Medicine (ISMRM) 24th Annual Meeting & Exhibition.* Singapore; 2016.
67. Burcaw LM, Fieremans E, Novikov DS. Mesoscopic structure of neuronal tracts from time-dependent diffusion. *NeuroImage.* 2015;114:18–37.
68. Barazany D, Basser PJ, Assaf Y. In vivo measurement of axon diameter distribution in the corpus callosum of rat brain. *Brain.* 2009;132(5):1210–1220.
69. Fieremans E, Burcaw LM, Lee HH, Lemberskiy G, Veraart J, Novikov DS. In vivo observation and biophysical interpretation of time-dependent diffusion in human white matter. *NeuroImage.* 2016;129:414–427.
70. Guglielmetti C, Veraart J, Roelant E, et al. Diffusion kurtosis imaging probes cortical alterations and white matter pathology following cuprizone induced demyelination and spontaneous remyelination. *Neuroimage.* 2016;125:363–377.

71. Kelm ND, West KL, Carson RP, Gochberg DF, Ess KC, Does MD. Evaluation of diffusion kurtosis imaging in ex vivo hypomyelinated mouse brains. *NeuroImage*. 2016;124:612–626.
72. Yablonskiy DA, Bretthorst LG, Ackerman JJ. Statistical model for diffusion attenuated MR signal. *Magn Reson Med*. 2003;50(4):664–669.
73. Axer H, Axer M, Krings T, Graf Diedrich v, Keyserlingk D. Quantitative estimation of 3-d fiber course in gross histological sections of the human brain using polarized light. *J Neurosci Methods*. 2001;105(1):121–131.
74. Ramirez-Manzanares A, Rivera M, Vemuri BC, Carney P. Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry. *IEEE Trans Med Imaging*. 2007;26(8):1091–1102.
75. Van Gelderen P, DesPres D, van Zijl P, Moonen CT. Evaluation of restricted diffusion in cylinders. phosphocreatine in rabbit leg muscle. *J Magn Reson, Ser B*. 1994;103(3):255–260.
76. Stone M. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Ser B (Methodological)*. 1974;1:111–147.
77. Efron B. Bootstrap methods: Another look at the jackknife. *Ann Stat*. 1979:1–26.
78. van Gelderen P, DesPres D, Vanzijl P, Moonen C. Evaluation of restricted diffusion in cylinders. phosphocreatine in rabbit leg muscle. *J Magn Reson, Ser B*. 1994;103(3):255–260.
79. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*. 2005;67(2):301–320.
80. Garyfallidis E, Brett M, Amirbekian B, et al. DIPY, a library for the analysis of diffusion MRI data. *Front Neuroinform*. 2014;8:8.
81. Pedregosa F, Varoquaux G, Gramfort A, et al. SCIKIT-LEARN: Machine learning in Python. *J Mach Learning Res*. 2011;12:2825–2830.
82. Zou H, Hastie T, Tibshirani R, et al. On the 'degrees of freedom' of the lasso. *Ann Stat*. 2007;35(5):2173–2192.
83. Loya-Olivas A, Aranda R, Rivera M. LASADD: linear acceleration method for adapting diffusion dictionaries. In: International Society of Magnetic Resonance in Medicine (ISMRM) 23rd Annual Meeting and Exhibition; 2015; Toronto, Ontario, Canada.
84. Loya AK. Algoritmos para solución de modelos no lineales en DW-MRI [Master's thesis]. Guanajuato, Mexico: Centro de Investigación en Matemáticas; 2015.
85. Aranda R, Ramirez-Manzanares A, Rivera M. Recovering detailed intra-voxel white matter structure by using an adaptive diffusion dictionary. In: International Society of Magnetic Resonance in Medicine (ISMRM) 23rd Annual Meeting and Exhibition; Toronto, Ontario, Canada:2015.
86. Aranda R, Ramirez-Manzanares A, Rivera M. Sparse and Adaptive Diffusion Dictionary (SADD) for recovering intra-voxel white matter structure. *Med Image Anal*. 2015;26(1):243–255.
87. Sherbondy AJ, Rowe MC, Alexander DC. Microtrack: An algorithm for concurrent projectome and microstructure estimation. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2010, Lecture Notes in Computer Science*, vol. 6361. Berlin: Springer; 2010:183–190.
88. Stejskal EO, Tanner JE. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J Chem Phys*. 1965;42(1):288–292. <https://doi.org/10.1063/1.1695690>
89. Özarslan E, Mareci TH. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magn Reson Med*. 2003;50(5):955–965.
90. Alipoor M, Gu IY, Mehnert AJ, Lilja Y, Nilsson D. On high order tensor-based diffusivity profile estimation. *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. Osaka, Japan; 2013:93–96.
91. Alexander D, Barker G, Arridge S. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magn Reson Med*. 2002;48(2):331–340.
92. Callaghan PT. Pulsed-gradient spin-echo NMR for planar, cylindrical, and spherical pores under conditions of wall relaxation. *J Magn Reson, Ser A*. 1995;113(1):53–59.
93. Craven P, Wahba G. Smoothing noisy data with spline functions. *Numer Math*. 1978;31(4):377–403.

How to cite this article: Ferizi U, Scherrer B, Schneider T, et al. Diffusion MRI microstructure models with in vivo human brain Connectome data: results from a multi-group comparison. *NMR in Biomedicine*. 2017;30:e3734. <https://doi.org/10.1002/nbm.3734>

APPENDIX: COMPETING MODELS

A.1 | Tissue models

A.1.1 | Ramirez-Manzanares (CIMAT, Mexico): Empirical Diffusion-and-Direction Distributions (ED³)

This work builds on the statistical modelling of the apparent diffusion coefficient⁷² and tackles the modelling of axon fibre dispersion in single^{15,73} and multiple fibre bundle cases. The method estimates empirically (rather than imposes) the distribution of tissue properties (axon radius, parallel diffusion, etc.), as well as the orientational distribution of the bundles. The general framework is as follows:

- estimation of mean principal diffusion directions (PDD) per axon bundle;
- selection of a dense set of orientationally focused basis directions that capture the discrete non-parametric fibre dispersion;
- design of a dictionary of intra/extracellular synthetic DW signals, which are precomputed along the basis directions (see the DBF method in Ramirez-Manzanares et al.⁷⁴);
- computation of the size compartments per diffusion atom of the dictionary (model fitting).

The PDDs are estimated from the Diffusion Tensor (DT) (single bundle case) and DBF⁷⁴ (complex structure cases). The 120 orientations closest to the PDDs are selected from a set of 1000 evenly distributed orientations. The intra-axonal signals S^i are precomputed from the model in Van Gelderen et al.⁷⁵ for restricted diffusion within a cylinder with radius $R = 1, 2, \dots, 10 \mu\text{m}$ and parallel diffusion $d_{\parallel} = 1, 1.1, \dots, 2.1 \mu\text{m}^2/\text{ms}$. The extra-axonal signals are generated as follows: S^e from *zeppelins* with combinations of parallel and radial diffusion, $d_{\parallel} = 1, 1.1, \dots, 2.5 \mu\text{m}^2/\text{ms}$ and $d_{\perp} = 2, 3, \dots, 8 \mu\text{m}^2/\text{ms}$; isotropic diffusion compartment signals $S_j^{\text{iso}} = \exp^{-\alpha r d_{\text{iso}}}$ for $d_{\text{iso}} = 2, 2.1, 2.2, \dots, 4 \mu\text{m}^2/\text{ms}$; and the *dot* signal, which takes into account static proton density. The values of the dictionary atoms above were tuned by cross-validation.⁷⁶ The size compartments $\beta \geq 0$ computed in the weighted non-negative LS formulation,

$$\left\| \left\| W \left[S - S_0^{\text{TE}} \left(\sum_{i=1}^{N_i} \beta_i^i S_i^i - \sum_{j=1}^{N_e} \beta_j^e S_j^e - \sum_{k=1}^{N_{\text{iso}}} \beta_k^{\text{iso}} S_k^{\text{iso}} + \beta_{\text{dot}} \right) \right] \right\| \right\|_2^2 \quad (\text{A1})$$

indicate the atoms that explain the signal; the W weights are proportional to SNR. Overfitting is reduced by a *bootstrap*⁷⁷ procedure.

The cross-validation experiments indicate that the reconstructions given by the robust fitting of this rich multi-compartment diffusion dictionary allow us to predict accurately non-acquired MR signals for different machine protocols. This is of most interest in the development of methods able to detect the complex microstructure heterogeneity associated with the different compartments within the voxels. The atoms with coefficients $\beta > 0$ depict the empirical distributions and their orientations indicate non-parametrical bundle-dispersion configurations (such as fanning or radially symmetrical). The recovered distributions reveal, for instance, an axon radius of between 1 and $4 \mu\text{m}$. One should take into account, however, that, since the heterogeneous intra/extra-axonal T_2 relaxation feature is not modelled explicitly, the method may compensate for T_2 variations by using, for instance, large isotropic d_{iso} coefficients to fit the signal accurately. For this reason, a direct interpretation of the fitted parameters may be misleading. The use of more specific models is a part of ongoing work.

A.1.2 | Nilsson (Lund, Sweden): multi-compartment model outlier rejection and separate fitting of b_0 data

This multiple compartment model was developed specifically for the ISBI WM challenge and built up by relaxation-weighted and time-dependent diffusion tensors according to

$$S = S_0 \sum_i w_i e^{-\mathbf{B} \cdot \mathbf{D}_i} e^{-\text{TE}/T_2} \left(1 - e^{-(\text{TR}-\text{TE})/T_1} \right) \quad (\text{A2})$$

where $\mathbf{B} = b \vec{n} \otimes \vec{n}$ and $b = (\gamma \delta g)^2 t_d$. The diffusion time t_d was corrected for rise times (ξ) according to $t_d = \Delta - \delta/3 + \xi^3/30\delta^2 - \xi^2/6\delta$. Each component was also described by a weight (w_i) and relaxation times (T_1 and T_2). The model featured three types of component, with either isotropic, hindered or restricted diffusion. Diffusion in the isotropic component was modelled by a single diffusion coefficient. The hindered and restricted components were modelled by cylinder-symmetric tensors described by axial and radial diffusivities together with the polar and azimuth angles. In the restricted component, the apparent diffusion coefficient of the radial component depended on δ and Δ , as well as on the cylinder radius, according to van Gelderen et al.⁷⁸

Three modifications were performed to this very general model. First, to accommodate for potential bias in the b_0 images (which was the case for fornix data, where deviations of up to 20σ was observed), the prediction for b_0 data was obtained from the median of all signals acquired with identical TE instead of from Equation A2. Second, opposite direction acquisitions were rescaled by a free model parameter, in order to allow for potential gradient instabilities inducing differences between the directions and their opposite directions. Third, models were generated dynamically during fitting by randomly selecting up to four hindered components and up to three restricted components. One isotropic component was always included.

The model was first fitted to half of the diffusion-weighted data (randomly selected), after which outliers were rejected ($> 2.5\sigma$). Thereafter a second fit was performed. Both fit steps assumed Gaussian noise and utilized the 'lsqcurvefit' function in Matlab. The procedure was repeated 100 times for different randomly generated models.

To prepare for submission of the results, only the models that best predicted the hidden half of the data were selected, after which the median of the selected predictions was used for the final prediction.

A.1.3 | Scherrer (Harvard, USA): distribution of anisotropic microstructural environments in diffusion compartment imaging (DIAMOND)

DIAMOND models the set of tissue compartments in each voxel by a finite sum of unimodal continuous distributions of diffusion tensors. This corresponds to a hybrid tissue model that combines biophysical and statistical modelling. As described by Scherrer et al.,¹⁶ the DW signal S_k for a gradient vector \mathbf{g}_k and b value b_k is modelled by

$$S_k = S_0 \left[\sum_{j=0}^N f_j \left(1 + \frac{b_k \mathbf{g}_k^T \mathbf{D}_j^0 \mathbf{g}_k}{\kappa_j} \right)^{-\kappa_j} \right]$$

where S_0 is the non-attenuated signal, N is the number of compartments, f_j the relative fraction of occupancy of the j th compartment and κ_j and \mathbf{D}_j^0 are respectively the concentration and expectation of the j th continuous tensor distribution. DIAMOND enables assessment of compartment-specific

diffusion characteristics such as the compartment FA (cFA), compartment RD (cRD) and compartment MD (cMD). It also provides a novel measure of microstructural heterogeneity for each compartment.

The estimation of a continuous distribution of diffusion tensors requires DW data acquired with the same timing parameters δ and Δ .¹⁶ To compare DIAMOND with other models in this dataset, we fitted one DIAMOND model separately for each $\{\delta, \Delta\}$ group (i.e. for each TE group), leading to 12 DIAMOND models. One shell was missing in each TE group; we predicted its signal using the corresponding DIAMOND model. The model estimation was achieved as follows. We first computed the mean and standard deviation of S_0 (μ_{S_0} and σ_{S_0}) within each TE group and discarded DW signals with intensity larger than $\mu_{S_0} + 3\sigma_{S_0}$ (simple artefact correction). We then estimated DIAMOND parameters as described in Scherrer et al.,¹⁶ considering Gaussian noise and cylindrical anisotropic compartments. For the genu, we considered a model with one freely diffusing and one anisotropic compartment; for the fornix, we considered a model with one freely diffusing compartment and two anisotropic compartments.

A.1.4 | Ferizi_1 and Ferizi_2 (UCL, England)

This submission uses two three-compartment models, as described in previous studies.^{39,41} These models consist of (1) either a Bingham distribution of sticks or a cylinder for the intracellular compartment; (2) a diffusion tensor for the extracellular compartment; (3) an isotropic cerebrospinal fluid (CSF) compartment. The T_2 relaxation element is fitted beforehand to the (variable echo time) $b = 0$ measurements. The signal model is as follows:

$$S = \tilde{S}_0 \left[f_i \exp\left(-\frac{TE}{T_2^i}\right) S_i + f_e \exp\left(-\frac{TE}{T_2^e}\right) S_e + f_c \exp\left(-\frac{TE}{T_2^c}\right) S_c \right] \quad (\text{A3})$$

where f_i , f_e and f_c are the weights of the intracellular, extracellular and third normalized compartment signals S_i , S_e and S_c , respectively; the values of compartmental T_2 are indexed similarly; \tilde{S}_0 is the proton density signal (which is TE-independent and obtained from fitting to the $b = 0$ signal). These models emerged from previous studies.^{41,44} Here, however, a single white matter T_2 and separate compartmental diffusivities are additionally fitted.

There is a two-stage model fitting procedure. The first step estimates the T_2 decay rate of tissue separately in each voxel, by fitting a bi-exponential model to the $b = 0$ intensity as a function of TE, in which one component is from tissue and the other from CSF. A preliminary analysis of voxels fully inside WM regions shows no significant departure from mono-exponential decay; equal T_2 values are then assumed within the intra and extracellular compartments. When fitting the bi-exponential model, the value of T_2 in CSF is fixed to 1000 ms (a more precise value of CSF is unlikely to be estimated with this protocol). Thus, for each voxel, the volume fraction of CSF, \tilde{S}_0 and T_2 of the tissue are estimated. These three estimates are then fixed for all the subsequent model fits. Then, each model is fitted using the Levenberg–Marquardt algorithm with an offset Gaussian noise model. The model parameters obtained were similar to earlier estimates obtained using the full dataset,⁴² differing by between 5–10% from the original.

A.1.5 | Poot (Erasmus, the Netherlands)

This submission uses a three-compartment model, with for each compartment a different complexity of the diffusion model and an individual T_2 value. This model was developed specifically for the ISBI WM challenge and is the result of iteratively visualizing different projections of the residuals and trying to infer the maximum complexity that the rich data supports.

The first compartment models isotropic diffusion and, through the initialization procedure, it captures the fast diffusion components. The second compartment is modelled by a second-order (diffusion) tensor and models intermediate diffusion strengths. The third compartment is model-free, as the ADC is estimated for each direction independently. Each compartment additionally has an individual T_2 value and signal intensity at $b = 0$, $TE = 0$ (which could easily be translated into volume fractions). Hence, the complete model of a voxel in image j is given by

$$S_j(\theta) = \sum_{i=1}^3 A_i e^{-TE_j R_{2,i}} e^{-b_j \text{ADC}_{j,i}} = \sum_{i=1}^3 e^{M_{i,j} \theta} \quad (\text{A4})$$

where S_j is the predicted signal intensity of image j , A_i is the non-diffusion weighted signal intensity of compartment i at zero TE, TE is the echo time, R_2 is the reciprocal of the T_2 relaxation time of compartment i , $b = (\Delta - \delta/3)\delta^2 |G|^2 \gamma^2$, with $\gamma = 42.5781$ MHz/T, $\text{ADC}_{j,1} = c$, $\text{ADC}_{j,2} = \mathbf{g}_j^T \mathbf{D} \mathbf{g}_j$, $\text{ADC}_{j,3} = \mathbf{d} \mathbf{h}_j^T$, where \mathbf{d} is a vector with the ADC value of each orientation group and \mathbf{h}_j is a vector that selects the orientation group to which image j belongs (90 groups in total). Note that \mathbf{h}_j has at most one non-zero element and that element has a value of one. As displayed in the rightmost part of Equation A4, the model can be written as a multiplication of matrices \mathbf{M}_i , containing all rows $\mathbf{M}_{i,j}$, with $\theta = [\ln A_1, R_{2,1}, c, \ln A_2, R_{2,2}, D_{11}, D_{12}, D_{13}, D_{22}, D_{23}, D_{33}, \ln A_3, R_{2,3}, \mathbf{d}]^T$, which combines all 103 parameters into a single parameter vector. All parameters are estimated simultaneously from the 3311 measurements provided per voxel by a maximum-likelihood estimator that assumes a Rician distribution of the measurements and simultaneously optimizes the noise level.⁵⁹ Finally, the signal intensities of the ‘unseen’ data are predicted by substituting the estimate into Equation A4.

A.1.6 | Rokem (Stanford, USA): a restriction-spectrum sparse fascicle model (RS-SFM)

The sparse fascicle model (SFM)⁴³ is a member of the large family of models that account for the diffusion MRI signal in the white matter as a combination of signals due to compartments corresponding to different axonal fibre populations (fascicles) and other parts of the tissue. Model fitting proceeds in two steps. First, an isotropic component is fitted. We model the effects of both the measurement echo time (TE) and the measurement b

value on the signal. These are fitted as a $\log(TE)$ -dependent decay with a low-order polynomial function and a b -value-dependent multi-exponential decay (also including an offset to account for the Rician noise floor). The residuals from the isotropic component are then deconvolved with the perturbations in the signal due to a set of fascicle kernels, each modelled as a radially symmetric ($\lambda_2 = \lambda_3$) diffusion tensor. The putative kernels are distributed in a dense sampling grid on the sphere. Furthermore, restriction spectrum imaging (RSI)⁶⁰ is used to extend the model, by adding a range of fascicle kernels at each sampling point, with different axial and radial diffusivities, capturing diffusion at different scales. To restrict the number of anisotropic components (fascicles) in each voxel and to prevent overfitting, the RS-SFM model employs the Elastic Net algorithm (EN),⁷⁹ which applies a tunable combination of L1 and L2 regularization on the weights of the fascicle kernels. We used elements of the SFM implemented in the *dipy* software library⁸⁰ and the EN implemented in *scikit-learn*.⁸¹ In addition, to account for differences in SNR, we implemented a weighted least-squares strategy, whereby each signal's contribution to the fit was weighted by its TE, as well as the gradient strength used. EN has two tuning parameters, determining (1) the ratio of L1-to-L2 regularization and (2) the weight of the regularization relative to the least-squares fit to the signal. To find the proper values of these parameters, we employed k -fold cross-validation,⁴³ leaving out one shell of measurement in each iteration for cross-validation. We determined that the tuning parameters with the lowest Least Squares Error (LSE)¹⁸ provide an almost-even balance of L1 and L2 penalty with weak overall regularization. Because of the combination of a dense sampling grid (362 points distributed on the sphere) and multiple restriction kernels (45 per sampling point), the maximal number of parameters for the model is approximately 16 300, more than the number of data points. However, because regularization is employed, the effective number of parameters is much smaller, resulting in an active set of approximately 20 regressors.⁸² We have made the code to reproduce our results fully available at <https://arokem.github.io/ISBI2015>.

A.1.7 | Eufrazio (CIMAT, Mexico): diffusion basis functions for multi-shell scheme

This model is based on the Diffusion Basis Functions (DBF) model,⁷⁴ a discrete version of the Gaussian Mixture Model for the sphere: $\hat{s}_i = \sum_{j=1}^m \alpha_j \phi_{ij} + \epsilon$, with $\hat{s}_i = s_i/s_0$, $\phi_{ij} = \exp(-bq_i^T T_j q_i)$ and $T_j = (\chi_1 v_j v_j^T + \chi_2 I)$. The DBF model is reformulated by substituting ϕ_{ij} and T_j : $\hat{s}_i = \sum_{j=1}^m \alpha_j \exp(-b_i \chi_2 g_i^T g_i) \exp(-b_i \chi_1 (v_j^T g_i)^2) + \epsilon$. The first exponential can be defined as a scale factor that depends on the b values, $\beta_i = \exp(-b_i \chi_2 q_i^T q_i)$. In this way, the β_i factors are associated with different b values, so the new model includes information for multi-shell schemes. The coefficients α and the shell scale factor β are computed by solving the optimization problem:

$$\min_{\alpha, \beta_c} f(\alpha, \beta_c; \lambda_\alpha, \lambda_\beta) = \|B\Phi\alpha - S\|_2^2 + \lambda_\alpha \|\alpha\|_1 + \lambda_\beta \|\beta_c^0 - \beta_c\|_2^2 \quad \text{s.t. } \mathbf{1}^T \alpha = 1, \quad \alpha \geq 0 \quad (\text{A5})$$

where $B = \text{diag}(\beta_c)$,

$$\beta_c = \frac{1}{\#C} \sum_{i \in C} \exp(-b_i \chi_2 (q_i^T q_i))$$

and C is the set of indices grouped by different b values ($\#C$ is the number of elements in it). The regularization term weighted by λ_α demands sparseness and the term weighted by λ_β prevents overfitting. The problem in Equation A5 is solved in three steps. First, the active atoms are predicted ($\alpha_i > 0$) with $\tilde{\alpha} = \text{argmin}_\alpha f(\alpha, \beta_c; \lambda_\alpha, \lambda_\beta)$. Second, the active atoms are corrected with $\alpha = \text{argmin}_{\{\alpha_i: \tilde{\alpha}_i > 0\}} f(\alpha, \beta_c; 0, \lambda_\beta)$. Finally, the factors β_c are updated with $\beta_c = \text{argmin}_{\beta_c} f(\alpha, \beta_c; \lambda_\alpha, \lambda_\beta)$. To solve each step, the active sets algorithm for quadratic programming is used.

To train the model for the WMM'15 data, Equation A5 is solved for each voxel with the training data to find the optimal weights α_j and scale factors β_c that best reproduce the training data. For this challenge, the β_c factors are grouped by the 36 training shells and the method parameters are set by hand: $\lambda_\alpha = 0.5$, $\lambda_\beta = 0.02$, $\chi_1 = 9.5 \times 10^{-4}$ and $\chi_2 = 5 \times 10^{-5}$. To predict the unseen signal at each voxel, the reformulated model is used with the optimal weights α_j and the 12 scale factors for the unseen β_c are calculated by interpolation with the 36 optimal β_c of the training data.

A.1.8 | Loya-Olivas_1 and Loya-Olivas_2 (CIMAT, Mexico): Linear Acceleration of Sparse and Adaptive Diffusion Dictionary (LASADD)

LASADD is a multi-tensor based technique to adapt dynamically the diffusion functions (DFs) dictionary to a DW-MRI signal.^{83,84} The method changes the size and orientation of relevant diffusion tensors (DTs). The optimization algorithm uses a special DT expression and assumptions to reduce the computational cost.

The one-compartment version (LASADD-1C) is based on a DBF multi-tensor model:⁷⁴ $s_i^* = \sum_{j=1}^n \alpha_j \phi_{ij}$, where $s_i^* = s_i/s_0$, $\phi_{ij} = \exp(-b_i g_i^T T_j g_i)$, $\alpha_j > 0$ and $\sum_{j=1}^n \alpha_j = 1$. LASADD expresses the DT as

$$T_j = \chi_{1j} v_j v_j^T + \chi_{2j} I, \quad (\text{A6})$$

where $\chi_{\{1,2\}j}$ are scalars associated with the eigenvalues, v_j is the principal diffusion direction (PDD) and I is the identity matrix. The algorithm iterates three steps, like Aranda et al.:^{85,86} Predict, Correct, and Generate, until convergence. Prediction selects the relevant DFs using LASSO to regulate the number to choose. Correction adjusts the volume fraction, size and orientation of the DTs. Taking advantage of the DT expression and Taylor first-order series approximation of the exponential, the optimizations are reduced to bounded least-squares problems, which are solved by a Projected Gauss-Seidel scheme. Generation controls the overestimation of fibres by adding to the basis the DTs resulting from combining two and three DFs for the new iteration.

An extra refinement to the computed results, named LASADD-3C, splits each detected DF into three compartments:⁸⁷ intracellular (IC), extra-cellular (EC) and CSF. The multi-tensor model is $s_i^* = \sum_{j=1}^n \alpha_j^{IC} \psi_{ij} + \sum_{j=1}^n \alpha_j^{EC} \theta_{ij} + \alpha^{CSF} \omega_i$ with $\sum_{j=1}^n (\alpha_j^{IC} + \alpha_j^{EC}) + \alpha^{CSF} = 1$. The ψ_{ij} models the directional IC compartment diffusion for each fibre bundle using $\mathbf{T}_j^{IC} = \chi_{0j} \mathbf{v}_j \mathbf{v}_j^T$. The EC compartment with hindered diffusion uses the representation (A6) for θ_{ij} . The isotropic diffusion ω_i uses $\mathbf{T}^{CSF} = \chi_3 \mathbf{I}$. This stage keeps the PDDs fixed and only adjusts the α 's and χ 's of the three compartments.

The parameters of the models were estimated using the training dataset: the b values using the equation by Stejskal and Tanner⁸⁸ and the S_0 values as the median of the gradient-free signals with equal echo time per voxel. The initial basis comprises 33 PDDs distributed in the unitary sphere. The bounds $\chi_{\{0,1\}} \in [1, 39] \times 10^{-4}$ and $\chi_{\{2,3\}} \in [1, 9] \times 10^{-4} \text{ mm}^2/\text{s}$ and the LASSO regularization parameter (equals 1.7) was tuned by hand such as to provide the minimum error. The best multi-tensorial model for both algorithms was used for each voxel to predict the corresponding unseen data.

A.2 | Signal models

A.2.1 | Alipoor (Chalmers, Sweden)

The diffusion MRI signal is modelled as a fourth-order symmetric tensor as proposed by Özarlan and Mareci.⁸⁹ Let $\mathbf{g}_i = [x_i y_i z_i]$ and $\mathbf{a}_i = [z_i^4 \ 4y_i z_i^3 \ 6y_i^2 z_i^2 \ 4y_i^3 z_i \ y_i^4 \ 4x_i z_i^3 \ 12x_i y_i z_i^2 \ 12x_i^2 y_i z_i \ 4x_i y_i^3 \ 6x_i^2 y_i^2 \ 12x_i^2 y_i z_i \ 6x_i^2 y_i^2 4x_i^3 z_i \ 4x_i^3 y_i \ x_i^4]^T$ be a gradient encoding direction and the corresponding design vector, respectively. The diffusion signal is then described by

$$S(\mathbf{g}_i) = S_0 \exp\left(\frac{-TE}{T_2}\right) \exp(-b\mathbf{t}^T \mathbf{a}_i) \quad (\text{A7})$$

where $S(\mathbf{g}_i)$ is the measured signal when the diffusion sensitizing gradient is applied in the direction \mathbf{g}_i , S_0 is the observed signal in the absence of such a gradient, b is the diffusion-weighting factor and $\mathbf{t} \in \mathbb{R}^{15}$ contains the distinct entries of a fourth-order symmetric tensor. Note that $d(\mathbf{g}_i, \mathbf{t}) = d(\mathbf{g}_i)$ is used for simplification. Given measurements in $N > 15$ different directions, the least-squares (LS) estimate of the diffusion tensor is $\hat{\mathbf{t}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{y}$, where \mathbf{G} is an $N \times 15$ matrix defined by $\mathbf{G} = [\mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_N]^T$ and $y_i = -b^{-1} \ln(S(\mathbf{g}_i)/S_0)$. We use the weighted LS tensor estimation method in Alipoor et al.⁹⁰ to mitigate the influence of outliers.

To estimate the diffusion signal for a given acquisition protocol with $TE = TE_x$, $b = b_x$ and $\delta = \delta_x$, the two non-diffusion-weighted measurements with the closest TE to TE_x (among measurements with $\delta = \delta_x$) are used to estimate T_2 and S_0 for each voxel. Then, data from the closest shell to b_x (among shells with $\delta = \delta_x$) are used to estimate the tensor describing the underlying structure.

A.2.2 | Sakaie-Tatsuoka-Ghosh (Cleveland, USA): an empirical approach

As the extent of q space in the dataset is unusually comprehensive, we chose a simple, generic approach to gain intuition. Visual inspection suggested use of a restricted and hindered component, each with angular variation:

$$S_i = A_{TE_i} (f R_i + (1 - f) \exp(-b_i D_i)) \quad (\text{A8})$$

where S_i is the predicted signal for a signal acquired with TE_i , b_i , A_{TE_i} is the median signal at a given TE with no diffusion weighting. Fitted parameters are f , the volume fraction of R_i , the restricted component, and D_i , the diffusivity. R_i and D_i are modelled as spherical harmonics with real, antipodal symmetry⁹¹ with maximum degree 4. The model has 31 fit parameters for each voxel. Data were fitted using using a non-linear least-squares algorithm (lsqcurvefit, MATLAB). Prior to the fit, data points with non-zero b value that had a signal higher than the median of the $b = 0$ signal plus 1.4826 times the median absolute deviation were excluded. Shells with a normalized median signal smaller than that of shells with lower b values were also excluded. Normalization was performed by dividing by the median of the $b = 0$ signal with the same TE.

A.2.3 | Fick (INRIA, France): a spatio-temporal functional basis to represent the diffusion MRI signal

We use our recently proposed spatio-temporal (3D+t) functional basis⁶¹ to represent the diffusion MRI signal simultaneously over three-dimensional wave vector \mathbf{q} and diffusion time τ . Based on Callaghan's theoretical model of spatio-temporal diffusion in pores,⁹² our basis represents the 3D+t diffusion signal attenuation $E(\mathbf{q}, \tau)$ as a product of a spatial and temporal functional basis as

$$E(\mathbf{q}, \tau) = \sum_{N=0}^{N_{\max}} \sum_{\{jlm\}}^{O_{\max}} c_{\{jlm\}} S_{jlm}(\mathbf{q}, u_s) T_o(\tau, u_t) \quad (\text{A9})$$

where T_o is our temporal basis with basis order o and S_{jlm} is the spatial isotropic MAP-MRI basis²⁹ with radial and angular basis orders j , l and m . Here, N_{\max} and O_{\max} are the maximum spatial and temporal order of the bases, which can be chosen independently. We formulate the bases themselves as

$$S_{jlm}(\mathbf{q}, u_s) = \sqrt{4\pi} i^{-l} (2\pi^2 u_s^2 q^2)^{l/2} e^{-2\pi^2 u_s^2 q^2} \times L_{j-1}^{l+1/2} (4\pi^2 u_s^2 q^2) Y_l^m(\mathbf{u}) \quad (\text{A10})$$

$$T_o(\tau, u_t) = \exp(-u_t \tau / 2) L_o(u_t \tau)$$

with u_s and u_t the spatial and temporal scaling functions, Y_l^m the spherical harmonics and L_o a Laguerre polynomial. We calculate the spatial scaling u_s by fitting an isotropic tensor to the TE-normalized signal attenuation $E(\mathbf{q}, \cdot)$ for all \mathbf{q} . Similarly, we compute u_t by fitting an exponential $e^{-u_t \tau / 2}$ to $E(\cdot, \tau)$

for all τ . We fit our basis using Laplacian-regularized least squares in the following steps. We first denote $\Xi_i(\mathbf{q}, \tau, u_s, u_t) = S_{jlm(i)}(\mathbf{q}, u_s) T_{o(i)}(\tau, u_t)$ with $i \in \{1 \dots N_{\text{coef}}\}$, with N_{coef} the number of fitted coefficients. We then construct a design matrix $\mathbf{Q} \in \mathbb{R}^{N_{\text{data}} \times N_{\text{coef}}}$ with $\mathbf{Q}_{ik} = S_{N_i}(\mathbf{A}, \mathbf{q}_k) T_{o_i}(\tau_k, u_t)$. The signal is then fitted as $\mathbf{c} = \text{argmin}_{\mathbf{c}} \|\mathbf{y} - \mathbf{Q}\mathbf{c}\|^2 + \lambda U(\mathbf{c})$ with \mathbf{y} the measured signal, \mathbf{c} the fitted coefficients and λ the weight for our analytic Laplacian regularization $U(\mathbf{c})$. We used generalized cross-validation⁹³ to find the optimal regularization weighting λ in every voxel. In our submitted results, we used a spatial order of 8 and a temporal order of 4, resulting in 475 fitted coefficients.

A.2.4 | Rivera (CIMAT, Mexico): baseline method: robust regression

We regard this very simplistic model as a baseline for other model-based methods. It assumes as little information as possible from the diffusion signal. The vector of independent variables is $\mathbf{x}_i = [g_i, |G_i|, \Delta_i, \delta_i, TE_i, b_i]$, containing the gradient strength g , the echo time TE and the b value b . Given signal s_i , we then estimate the parameters of the linear regression model:

$$s = X\theta + \epsilon \quad (\text{A11})$$

where $\theta \in \mathbb{R}^{23}$ is the unknown vector of coefficients, ϵ is the residual error and

$$X = [x, x^2, \Delta, \delta, \Delta TE, \Delta b, \delta TE, \delta b, TE b, 1]$$

is the *matrix design* (x^2 is obtained from squaring each element of the matrix x). To account for outliers, we estimate θ with a weighted (robust) least-squares approach using the Lasso regularization:

$$\theta^{t+1} = \text{argmin}_{\theta} \|W^t(X\theta - y)\|_2^2 + \lambda \|\theta\|_1 \quad (\text{A12})$$

where W^0 is the identity matrix and each subsequent W is computed via

$$W^{t+1} = \text{diag}(v_i^{t+1} w_i^{t+1}) \quad (\text{A13})$$

with outlier weighting in $\omega_i^{t+1} = \kappa^2 / (\kappa^2 + (y_i - X\theta_i^{t+1})^2)$ through κ , an arbitrary parameter that controls the outlier sensitivity. The protocol weight

$$v_i^{t+1} = \text{mean}_{j \in \Omega_i} \{w_j^{t+1}\} \quad (\text{A14})$$

and

$$\Omega_i = \{j : TE_j = TE_i, |G_j| = |G_i|\}$$

computes a confidence factor for the complete protocol.

Equations A12 and A13 are iterated three times. The final estimated signal is computed using (A11), using the protocol of the unseen signal.