NeuroImage 240 (2021) 118367

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

On the generalizability of diffusion MRI signal representations across acquisition parameters, sequences and tissue types: Chronicles of the MEMENTO challenge

Alberto De Luca[a,b,*], Andrada Ianus[c], Alexander Leemans[a], Marco Palombo[d], Noam Shemesh[c], Hui Zhang[d], Daniel C. Alexander[d], Markus Nilsson[e], Martijn Froeling[f], Geert-Jan Biessels[b], Mauro Zucchelli[g], Matteo Frigo[g], Enes Albay[g,h], Sara Sedlar[g], Abib Alimi[g], Samuel Deslauriers-Gauthier[g], Rachid Deriche[g], Rutger Fick[i], Maryam Afzali[j], Tomasz Pieciak[k,l], Fabian Bogusz[k], Santiago Aja-Fernández[l], Evren Özarslan[m,n], Derek K. Jones[j], Haoze Chen[o], Mingwu Jin[p], Zhijie Zhang[o], Fengxiang Wang[o], Vishwesh Nath[q], Prasanna Parvathaneni[r], Jan Morez[s], Jan Sijbers[s], Ben Jeurissen[s], Shreyas Fadnavis[t], Stefan Endres[u], Ariel Rokem[v], Eleftherios Garyfallidis[t], Irina Sanchez[w], Vesna Prchkovska[w], Paulo Rodrigues[w], Bennet A. Landman[x], Kurt G. Schilling[x,y]

[a] *PROVIDI Lab, Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands*
[b] *Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, the Netherlands*
[c] *Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal*
[d] *Centre for Medical Image Computing, Department of Computer Science, University College London, London, United Kingdom*
[e] *Clinical Sciences Lund, Radiology, Lund University, Lund, Sweden*
[f] *Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands*
[g] *Inria Sophia Antipolis – Méditerranée, Université Côte d'Azur, Sophia Antipolis, France*
[h] *Istanbul Technical University, Istanbul, Turkey*
[i] *TRIBVN Healthcare, Paris, France*
[j] *Cardiff University Brain Research, Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, United Kingdom*
[k] *AGH University of Science and Technology, Kraków, Poland*
[l] *LPI, ETSI Telecomunicación, Universidad de Valladolid, Valladolid, Spain*
[m] *Department of Biomedical Engineering, Linköping University, Linköping, Sweden*
[n] *Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden*
[o] *School of Instruments and Electronics, North University of China, Taiyuan, China*
[p] *Department of Physics, University of Texas at Arlington, Arlington, USA*
[q] *NVIDIA Corporation, Bethesda, USA*
[r] *National Institute of Health, Bethesda, USA*
[s] *Imec-Vision lab, Department of Physics, University of Antwerp, Antwerp, Belgium*
[t] *Intelligent Systems Engineering, Indiana University Bloomington, Indiana, USA*
[u] *Leibniz Institute for Materials Engineering — IWT, Faculty of Production Engineering, University of Bremen, Bremen, Germany*
[v] *Department of Psychology and the eScience Institute, University of Washington, Seattle, WA USA*
[w] *QMENTA Inc, Boston, USA*
[x] *Vanderbilt University Institute of Imaging Science, Vanderbilt University, Nashville, USA*
[y] *Department of Radiology and Radiological Science, Vanderbilt University Medical Center, Nashville, USA*

ABSTRACT

Diffusion MRI (dMRI) has become an invaluable tool to assess the microstructural organization of brain tissue. Depending on the specific acquisition settings, the dMRI signal encodes specific properties of the underlying diffusion process. In the last two decades, several signal representations have been proposed to fit the dMRI signal and decode such properties. Most methods, however, are tested and developed on a limited amount of data, and their applicability to other acquisition schemes remains unknown. With this work, we aimed to shed light on the generalizability of existing dMRI signal representations to different diffusion encoding parameters and brain tissue types. To this end, we organized a community challenge - named MEMENTO, making available the same datasets for fair comparisons

* Corresponding author.
E-mail address: a.deluca-2@umcutrecht.nl (A. De Luca).

https://doi.org/10.1016/j.neuroimage.2021.118367.
Received 3 March 2021; Received in revised form 9 June 2021; Accepted 4 July 2021
Available online 6 July 2021.
1053-8119/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

across algorithms and techniques. We considered two state-of-the-art diffusion datasets, including single-diffusion-encoding (SDE) spin-echo data from a human brain with over 3820 unique diffusion weightings (the MASSIVE dataset), and double (oscillating) diffusion encoding data (DDE/DODE) of a mouse brain including over 2520 unique data points. A subset of the data sampled in 5 different voxels was openly distributed, and the challenge participants were asked to predict the remaining part of the data. After one year, eight participant teams submitted a total of 80 signal fits. For each submission, we evaluated the mean squared error, the variance of the prediction error and the Bayesian information criteria. The received submissions predicted either multi-shell SDE data (37%) or DODE data (22%), followed by cartesian SDE data (19%) and DDE (18%). Most submissions predicted the signals measured with SDE remarkably well, with the exception of low and very strong diffusion weightings. The prediction of DDE and DODE data seemed more challenging, likely because none of the submissions explicitly accounted for diffusion time and frequency. Next to the choice of the model, decisions on fit procedure and hyperparameters play a major role in the prediction performance, highlighting the importance of optimizing and reporting such choices. This work is a community effort to highlight strength and limitations of the field at representing dMRI acquired with trending encoding schemes, gaining insights into how different models generalize to different tissue types and fiber configurations over a large range of diffusion encodings.

1. Introduction

Diffusion Magnetic Resonance Imaging (dMRI) is a powerful tool to investigate microstructural properties of biologic tissues in-vivo (Alexander et al., 2007; Tournier et al., 2011) with applications in neuroimaging studying brain development (Ouyang et al., 2019), plasticity (Blumenfeld-Katzir et al., 2011), aging (Baker et al., 2014), as well as changes upon disease for diagnostic and monitoring purposes in various conditions such as Alzheimer's disease (Doan et al., 2017, Weston et al., 2015), multiple sclerosis (De Santis et al., 2019, Inglese and Bester, 2010), Parkinson's disease (Atkinson-Clement et al., 2017), brain tumours (Costabile et al., 2019), etc. The signal measured in dMRI is sensitized to the microscopic motion of water molecules, which is hindered and restricted by the presence of biologic membranes, thus carrying information about the cellular organization. Over the last decade, an increasing number of techniques have been proposed in the literature to describe the dMRI signal and provide biomarkers of tissue microstructure and have been recently complemented with various machine learning approaches (Alexander et al., 2017, Ghosh et al., 2018; Novikov et al., 2019, Poulin et al., 2019, Ravi et al., 2019).

The standard acquisition strategy for dMRI data is single diffusion encoding (SDE), which employs a pair of diffusion weighting gradients with identical areas, usually embedded before and after the refocusing pulse in a spin echo preparation, a sequence widely known also as pulsed gradient spin-echo (Stejskal and Tanner, 1965). The SDE sequences are characterized by the gradient strength (G), duration (δ), time interval between the onset of the two gradients (Δ) and gradient orientation ($\hat{g}$). The scalar parameters (G, δ, Δ) are usually combined to describe the diffusion weighting of the sequence, also referred to as the b-value. For SDE sequences, $b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3)$, where γ is the gyromagnetic ratio. In the majority of SDE acquisitions δ and Δ are fixed and G is varied to change the b-value, although varying the gradient duration and diffusion time can provide additional orthogonal measurements (Novikov et al., 2019). Over the last decade, diffusion sequences which further vary the gradient waveform within one measurement, such as double diffusion encoding (DDE) (Henriques et al., 2020, Mitra, 1995) or b-tensor encoding approaches (Lasic et al., 2014; Westin et al., 2016), have been gaining interest as they can further improve the specificity of the measurements towards the underlying tissue microstructure. Other approaches replace the pulsed gradients with oscillating gradient waveforms to probe diffusion on a range of (shorter) time scales (Burcaw et al., 2015, Does et al., 2003), measurements which can also be performed with different gradient orientations in a double oscillating diffusion encoding (DODE) fashion (Ianus et al., 2018, Ianus et al., 2017). While the majority of recent dMRI studies employ SDE sequences, such advanced acquisitions are steadily gaining popularity.

The most widely used dMRI technique for brain imaging in the clinic is diffusion tensor imaging (DTI) (Basser et al., 1994), which assumes that water diffusion in the underlying tissue can be described by a Gaussian anisotropic process. As minimum requirements, the tensor parameters can be estimated from SDE sequences with a single b-value (usually about 1000 s/mm^2) and at least 6 non-collinear directions in addition to non-diffusion weighted data (b = 0 s/mm^2). Although simple and robust, DTI cannot describe the signal decay in correspondence of higher b-values (e.g., above about 1500 s/mm^2 in the living brain) and cannot distinguish between multiple fibre populations, for instance in areas of crossing fibres (Jeurissen et al., 2013). After DTI, a plethora of techniques have been introduced to capture the dMRI signal decay over a wider range of parameter values (Alexander et al., 2017; Novikov et al., 2019). dMRI models can generally be regarded as biophysical models, signal representations or somewhere in between (Ghosh et al., 2018; Jelescu and Budde, 2017). Biophysical models usually employ multiple water compartments to describe the dMRI signal in the tissue in order to capture microscopic metrics such as intracellular signal fraction, cell size, shape etc (Jespersen et al., 2007, Stanisz et al., 1997; Assaf et al., 2008; Alexander, 2008; Fan et al., 2020, Fieremans et al., 2013, Palombo et al., 2020, Panagiotaki et al., 2012, Zhang et al., 2012). Several biophysical models have been proposed in the literature and vary in terms of the number of compartments, diffusion model (hindered/restricted), number of fibre populations, fibre orientation distributions, etc. Signal representations on the other hand, usually provide a statistical description to capture the signal decay without explicitly modelling the underlying tissue composition (Yablonskiy et al., 2003; Jensen et al., 2005, Ozarslan et al., 2009; Fieremans et al., 2011, Ozarslan et al., 2013, Steven et al., 2014). Next to these two main families, there are also hybrid approaches that, for instance, aim to characterize the fibre orientation distribution without explicitly modelling the fibre composition (Tournier et al., 2008; Jeurissen et al., 2014), or use a statistical model for different compartments (De Luca et al., 2017, Pasternak et al., 2009, Scherrer et al., 2016), which can be defined a-priori or driven from the data (De Luca et al., 2018, Keil et al., 2017). Besides these "classical" approaches to model dMRI, the last couple of years have witnessed a vast increase in the number of machine learning techniques applied to dMRI to predict signal decay (Golkov et al., 2016, Grussu et al., 2020), fibre orientations (Nath et al., 2019, Poulin et al., 2019) or the underlying tissue parameters (Nedjati-Gilani et al., 2017).

The choice of dMRI technique depends on many factors, such as the purpose of the experiment, the amount and quality of the data, the number and strength of b-values, angular resolution, etc, and generally no consensus has been reached on what a state-of-the-art diffusion experiment should include. Nevertheless, for a given acquisition, comparing diffusion models can provide valuable information about which approaches best describe the signal and can be generalized to predict measurements outside the initial range. In the literature, there have been various studies which aimed to compare brain tissue models in terms of goodness of fit and signal prediction, with an emphasis on white matter (WM) (Panagiotaki et al., 2012; Ferizi et al., 2015; Jelescu et al., 2014, Rokem et al., 2015) and less on gray matter (GM)(Assaf, 2019). However, such studies usually focused on a certain group of models, for instance multi-compartment biophysical models were investigated in (Panagiotaki et al., 2012; Ferizi et al., 2015), while (Wang et al., 2017) looked in more detail at signal representations.

Open challenges play an important role to gain a better understanding of how various models capture the dMRI signal decay, as they

put forward rich datasets and well-defined tasks and usually receive submissions from across the modelling landscapes (Ferizi et al., 2017, Pizzolato et al., 2020, Schilling et al., 2019). The last diffusion microstructure challenge which included a comprehensive dMRI acquisition (Ferizi et al., 2017) was organized in 2015 and focused on modelling the dMRI signal acquired on the Connectome scanner for two ROIs in white matter: genu of the Corpus Callosum with mostly aligned fibres and fornix with a more complex fibre configuration. The challenge included a rich dataset acquired with many combinations of gradient strengths, durations and diffusion times and the goal was to predict unseen shells with parameter values within the range used for the provided data. Since the end of this challenge, many novel approaches have been proposed, including a booming application of machine learning techniques for data fitting and prediction (Golkov et al., 2016, Nath et al., 2019, Nedjati-Gilani et al., 2017, Poulin et al., 2019, Ravi et al., 2019). Moreover, previous challenges (Ferizi et al., 2017, Pizzolato et al., 2020, Schilling et al., 2019) included only diffusion data acquired with standard SDE sequences, and do not provide any insight into the different approaches available to analyse advanced sequences such as DDE.

In this challenge we set to evaluate the ability of different dMRI modeling approaches to capture the dMRI signal contrast from state-of-the-art acquisitions performed with SDE, including shells with an order of magnitude higher angular resolution compared to previous challenges (Ferizi et al., 2017, Pizzolato et al., 2020, Schilling et al., 2019), as well as DDE and DODE data. Further, we aim to investigate the relationship between the goodness of fit and tissue type, acquisition parameters, and diffusion sensitization. Finally, this challenge acts as a benchmark database for the evaluation of future models focusing on healthy tissue as the full datasets are made available (https://github.com/PROVIDI-Lab/MEMENTO.git).

Table 1
A description of the SDE encoding acquired as part of the MASSIVE data, and their subdivision in training and evaluation data. No data points of SDE-MS at b = 4000 s/mm^2 and SDE-GRID at b > 7600 s/mm^2 were provided for training. The ratio between training and evaluation data was about 1:3 for SDE-MS and 3:5 for SDE-GRID.

Diffusion Encoding	Number of directions	Training	Evaluation
SDE-MS data of a healthy volunteer acquired at 3T in 18 sessions (13 shells)			
b = 0 s/mm^2	430	20	410
b = 5 s/mm^2	30	8	22
b = 10 s/mm^2	30	7	23
b = 25 s/mm^2	40	10	30
b = 40 s/mm^2	40	10	30
b = 60 s/mm^2	20	5	15
b = 80 s/mm^2	20	5	15
b = 140 s/mm^2	20	5	15
b = 250 s/mm^2	30	8	22
b = 500 s/mm^2	250	62	188
b = 1000 s/mm^2	500	125	375
b = 2000 s/mm^2	500	125	375
b = 3000 s/mm^2	500	125	375
b = 4000 s/mm^2	600	0	600
Total	3010	515	2495
SDE-GRID data of a healthy volunteer acquired at 3T in 18 sessions			
b = 0 s/mm^2	430	20	410
b = 141-563 s/mm^2	58	22	36
b = 750-938 s/mm^2	54	22	32
b = 1125-1875 s/mm^2	170	72	98
b = 2016-2766 s/mm^2	248	102	146
b = 3000-3938 s/mm^2	314	129	185
b = 4125-4875 s/mm^2	168	68	100
b = 5016-6891 s/mm^2	196	65	131
b = 7687-9000 s/mm^2	32	0	32
Total	1670	500	1170

2. Methods

Section 2.1 presents an overview of the data that was used in the MEMENTO challenge and of the reasoning behind the selection of specific brain locations. A description of the methods used in the received submissions is reported in section 2.2, whereas section 2.3 illustrates the analyses we performed on the collected signal predictions.

2.1. Challenge data

2.1.1. MRI acquisition

The main aim of the MEMENTO challenge was to investigate how well existing models can represent the dMRI signal collected with i) different gradient encoding schemes, and ii) from different tissue types.

To investigate how the existing models can predict data sampled with different diffusion sensitization, we selected two datasets containing extensively sampled dMRI brain data with 4 encoding schemes: multi-shell SDE (SDE-MS), SDE with gradients sampled in a cartesian grid (SDE-GRID), DDE and DODE. The SDE acquisitions were performed in a healthy volunteer with a 3T scanner as part of the MASSIVE datasets (Froeling et al., 2017a), a collection of 18 MRI sessions containing unique dMRI data performed with a 3T scanner (Philips Healthcare, The Netherlands) with voxel-size 2.5mm^3 isotropic, echo time TE=100ms and repetition time TR between 7 and 7.5s. The DDE and DODE data were sampled from an ex-vivo mouse brain imaged with a 16.4T scanner (Bruker) with imaging resolution $0.12 \times 0.12 \times 0.7$mm^3, TE=52ms, TR=3s (Ianus et al., 2018). The diffusion parameters of the acquired data are reported in Tables 1 and 2. To establish the signal to noise ratio (SNR) of the data, we considered the datapoints collected at b = 0 s/mm^2, removed eventual outliers and defined the average SNR as the ratio between the average non-weighted value and its standard deviation. A point was defined as an outlier when its value was outside the confidence interval defined by the median value ± 2 times the robust standard deviation of the data (Chang et al., 2005). The SNR of the 5 selected signals at b = 0 s/mm2 was 15 ± 3 for SDE-MS, 16 ± 3 for SDE-GRID, 76 ± 37 for DDE and 66 ± 28 for DODE.

2.1.2. Signals selection

Five signals were selected for each dataset from brain voxels exhibiting different microstructural organization. For the human MASSIVE dataset, the voxels aimed to include WM signals with an increasing number of crossing fiber configurations from 1 to 3, deep gray matter (DGM) and cortical gray matter (CGM) and the selection was based on visual inspection of the fiber orientation distribution (FOD). The FOD was derived with constrained spherical deconvolution (Tournier et al., 2007) of data at b = 0, 3000 s/mm^2 (500 directions) using a recursively calibrated response function (Tax et al., 2014) approach implemented in ExploreDTI. Given the relatively low imaging resolution of the MASSIVE dataset, 2.5mm^3 isotropic, care was taken in selecting voxels not located at tissue interfaces, also using the FODs as guidance. For the mouse brain, the five voxels were placed in white matter tracts with different microstructures and based on visual comparison with an anatomical atlas, as the number of collected gradient orientations per diffusion weighting was insufficient to reliably estimate the FOD. Specifically, the five voxels were placed in medial corpus callosum, lateral corpus callosum, internal capsule, a fanning region of the internal capsule and the fimbria, respectively. An illustration of the locations and tissue types of the selected voxels is shown in Fig. 1.

2.1.3. Training and evaluation data

The 20 selected measurement sets (5 signal locations x 4 diffusion encodings) were subdivided in training and evaluation data. The challenge participants were provided with the training data and the corresponding diffusion encoding information, and asked to predict the evaluation data. The proportion of training and evaluation data was not constant among data encodings. For SDE, the training data consisted of about 500 gradient directions uniformly subsampled from all available data except for the largest diffusion weightings, which corresponds to about

Table 2
The table describes the subdivision in training and evaluation data of the unique combinations of diffusion weighting b and diffusion time (Δ) / oscillation frequency (*f*) for the DDE and DODE data, respectively. In the DDE training set, the data acquired with Δ = 5ms was provided with the exception of the largest diffusion weighting (b = 4000 s/mm^2), and no data acquired with Δ = 10ms was provided for training. The mixing time for DDE sequences was 18.3ms. For DODE, the data points acquired with the three lower oscillation frequencies (66, 100, 133 Hz) were provided with the exception of b = 4000 s/mm^2, and no data acquired with f = 166 and 200 Hz was provided for training. The separation time between gradient waveforms was 5ms. For DDE and DODE data, the sequence description included information about gradient strengths, directions, timing parameters as well as the elements of the B-matrix.

Diffusion Encoding	Number of directions	Training	Evaluation
DDE data from a mouse brain ex-vivo acquired at 16.4T			
Δ=5ms, b = 0 s/mm^2	**40**	**32**	**8**
Δ=5ms, b = 1000 s/mm^2	**72**	**72**	**0**
Δ=5ms, b = 1750 s/mm^2	**72**	**72**	**0**
Δ=5ms, b = 2500 s/mm^2	**72**	**72**	**0**
Δ=5ms, b = 3250 s/mm^2	**72**	**72**	**0**
Δ=5ms, b = 4000 s/mm^2	**72**	**0**	**72**
Δ=10ms, b = 0 s/mm^2	**40**	**0**	**40**
Δ=10ms, b = 1000 s/mm^2	**72**	**0**	**72**
Δ=10ms, b = 1750 s/mm^2	**72**	**0**	**72**
Δ=10ms, b = 2500 s/mm^2	**72**	**0**	**72**
Δ=10ms, b = 3250 s/mm^2	**72**	**0**	**72**
Δ=10ms, b = 4000 s/mm^2	**72**	**0**	**72**
Total	**800**	**320**	**480**
DODE data from a mouse brain ex-vivo acquired at 16.4T			
f = 66, 100, 133Hz b = 0 s/mm^2	**3 × 40**	**3 × 32**	**3 × 8**
f = 66, 100, 133Hz b = 1000 s/mm^2	**3 × 72**	**3 × 72**	**0**
f = 66, 100, 133Hz b = 1750 s/mm^2	**3 × 72**	**3 × 72**	**0**
f = 66, 100, 133Hz b = 2500 s/mm^2	**3 × 72**	**3 × 72**	**0**
f = 66, 100, 133Hz b = 3250 s/mm^2	**3 × 72**	**3 × 72**	**0**
f = 66, 100, 133Hz b = 4000 s/mm^2	**3 × 72**	**0**	**3 × 72**
f = 166, 200Hz b = 0 s/mm^2	**2 × 40**	**0**	**2 × 40**
f = 166, 200Hz b = 1000 s/mm^2	**2 × 72**	**0**	**2 × 72**
f = 166, 200Hz b = 1750 s/mm^2	**2 × 72**	**0**	**2 × 72**
f = 166, 200Hz b = 2500 s/mm^2	**2 × 72**	**0**	**2 × 72**
f = 166, 200Hz b = 3250 s/mm^2	**2 × 72**	**0**	**2 × 72**
f = 166, 200Hz b = 4000 s/mm^2	**2 × 72**	**0**	**2 × 72**
Total	**2000**	**960**	**1040**

17% of SDE-MS, and 30% of SDE-GRID. . For DDE and DODE, we provided a larger amount of training data to take into account the need to model an additional encoding dimension (e.g., time and frequency), respectively 40% and 48%. To evaluate the ability of the tested models to predict unseen data points, all the data corresponding to specific diffusion weightings was removed from the training data, as reported in Tables 1 and 2.

2.2. Signal predictions

We received initial submissions from 9 teams, but 2 of the 9 teams did not provide valid submissions and were not included in this analysis. The remaining 7 teams submitted a total of 80 valid signal predictions that were considered in the subsequent analyses. Of these, 31 submissions predicted the SDE-MS signals (37%), 16 the SDE-GRID signals (19%), 15 the DDE signals (18%) and 18 the DODE signals (22%). When a model was applied more than once to predict a given set of signals, only the submissions corresponding to the best and worst prediction were analyzed, to simplify the presentation and interpretation of the results. The final selection of predictions included in this analysis is reported in Table 3. When multiple predictions with a given model were submitted, the best and worst predictions were identified by adding the labels "_best" and "_worst" to the model name.

To follow, we present an overview of the submissions we received grouped in four different categories.

2.2.1. Tensor and beyond

Diffusion tensor imaging (DTI, (Basser et al., 1994) is one of the most common quantification methods for dMRI data acquired with at least one diffusion-weighting and 6+ gradient directions. DTI is based on the three-dimensional generalization of the seminal works of Stejskal and Tanner (1965), and assumes the diffusion process to be Gaussian (i.e.,

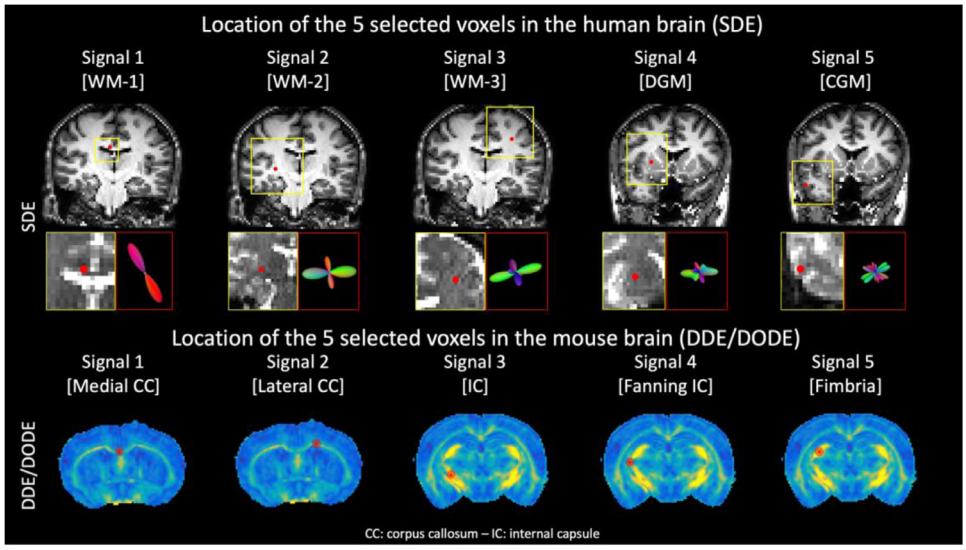

Fig. 1. The locations of the 5 voxels selected for the SDE (top) and DDE/DODE data (bottom). The SDE data are sampled from a human brain scanned at 3T with imaging resolution 2.5mm^3 isotropic, and is part of the MASSIVE dataset. The 5 locations were chosen in 5 different tissue types, such as white matter (WM) with increasing fiber complexity (signals 1-3, as exemplified by the shown fiber orientation distribution), deep gray matter (DGM) and cortical gray matter (CGM). DDE and DODE were sampled in a mouse brain at 16.4T with imaging resolution $0.12 \times 0.12 \times 0.7$mm^3. The selected signals are all taken from WM locations with well-established differences in fiber organization, as shown by the red dots overlaid on the fractional anisotropy map.

not restricted). In the living brain, such assumption is typically satisfied when collecting dMRI data with diffusion weightings in the range b = 800-1200 s/mm^2. While DTI typically does not accurately characterize complex diffusion environments where, for example, multiple diffusion mode (e.g., tissue diffusion vs blood pseudo-diffusion (Le Bihan et al., 1988) or "crossing-fibers" co-exist (Wedeen et al., 2005), it is one of the most common dMRI signal representations in clinical application, especially thanks to its sensitivity to microstructural changes in health and pathology. Keeping in mind all of the above, the DTI method was applied in this work to SDE-MS, DDE and DODE data to serve as baseline reference using a weighted-least-squares fit.

In 2005, Jensen and colleagues introduced the diffusion kurtosis imaging (DKI) method (Jensen et al., 2005), an extension of DTI that allows to account for and quantify the amount of non-Gaussian diffusion that is observed at stronger diffusion weightings (e.g., b > 1400 s/mm^2 in the living brain). The DKI model requires the collection of at least 21 unique measurements, including 2 non-zero diffusion-weightings and 15+ unique gradient directions, and allows to quantify the amount of excess kurtosis of the diffusion process. While DKI is suitable for dMRI data acquired with a stronger diffusion weighting than DTI, nevertheless, there is a theoretical maximum to the diffusion weighting that can be fit (Jensen et al., 2005; Jensen and Helpern, 2010), and most brain applications use a maximum b-value equal to b = 3000 s/mm^2. Despite these theoretical limits, DKI was fit to all data included in this work using a constrained non-linear least-squares fit enforcing positivity in diffusion and kurtosis metrics and a monotonic decay of the signal. When fitting data acquired with strong diffusion weighting, it might be beneficial to take into account the presence of a minimum signal offset due to Rician noise in the measurements (Basu et al., 2006, Gudbjartsson and Patz, 1995). One of the submissions considered in this work extended the DKI method with an offset term to account for such effect (DKI+Offset) (Morez et al., 2020). This method was fit to SDE data by extending the classic DKI model with an additional degrees of freedom (22 + 1 = 23 free parameters), and to DDE and DODE data by extending a fourth order covariance tensor (28 + 1 = 29 free parameters) (Westin et al., 2016).

2.2.2. Multicompartment models

Multi-compartment models are a family of methods that allow to model the dMRI signal by means of biophysical features (Panagiotaki et al., 2012; Jelescu and Budde, 2017). The assumption behind these models is that the dMRI signal acquired in a voxel can be described as the linear combination of the signal profiles of each component that is present in a specific voxel.

All the submissions we received based on multicompartment models were computed with custom implementations of previously introduced methods with the "Diffusion Microstructure Imaging in Python" toolbox (Fick et al., 2019) (dmipy). The submissions considered only SDE data, and were based on three basic components: the intra-axonal compartment was modelled as a stick or a cylinder, whereas the extra-axonal anisotropic compartment was modelled as a zeppelin (axially symmetric tensor) and the cerebrospinal fluid contribution modelled as isotropic diffusion (sphere). Depending on the specific implemented model, the anisotropic compartments were optionally convolved with a Watson or a Bingham distribution to account for fiber orientation dispersion. A summary of the multicompartment models that were submitted and the components they are based on is shown in Table 4.

For all the above mentioned models, the parallel diffusivity of the anisotropic compartments was set to 1.7×10^{-3}mm^2/s, whereas the diffusivity of the isotropic compartment was set to 3×10^{-3}mm^2/s, as previously suggested (Behrens et al., 2003, Kaden et al., 2016, Sotiropoulos et al., 2012, Zhang et al., 2012). The perpendicular diffusivity of the anisotropic compartments was linked to the parallel diffusivity via the tortuosity constraint (Szafer et al., 1995, Zhang et al., 2012).

2.2.3. Parametric representations

This family of methods focuses on expressing the dMRI signal as a function of mathematical signal basis without biophysical hypotheses. A popular signal representation is the simple harmonic oscillator reconstruction (SHORE) (Ozarslan et al., 2009). The SHORE basis has some degrees of freedom such as the order and a scaling factor. The SHORE method was applied to predict all the 4 provided data types (SDE-MS, SDE-GRID, DDE, DODE) in combination with a BFGS fit (Nath et al., 2019) that can be utilized to achieve the best fit of the scaling parameter. The same method was used for all 4 types of data with harmonics of orders 6, 8 and 12.

MAP-MRI(Ozarslan et al., 2013) is a signal representation-based technique that expresses the diffusion signal in q-space and parametrizes its Fourier transform –the mean apparent propagator (MAP)– in terms of a series involving products of three Hermite functions, thus generalizing the 1D-SHORE technique to three dimensions. Unlike in 3D-SHORE, an anisotropic scaling parameter is employed in MAP-MRI, making it an extension of DTI for representing the signal at large q-values. In this challenge, we received submissions from two different teams based on a Laplacian regularized version of MAP-MRI. Both submissions (MAP-MRI, MAP-MRI+Reg) predicted the unseen data points using regularized least squares and SHORE basis of order 8.

Table 3
The valid signal predictions submitted to the MEMENTO challenge. For each method, we report the acronym and the main reference, the "category", special notes on the fit procedure, and the data it has been applied to. The following predictions were subdivided in the following categories: tensor-based (TENS), multi-compartment model (MCM), parametric representation (PAR), deep learning-based (DL).

Model name	Category	Notes	SDE-MS	SDE-GRID	DDE	DODE
DTI (Basser et al., 1994)	TENS	Linear-Least Squares	X	X	X	X
DKI (J. H. (Jensen et al., 2005)	TENS	Weighted-Least Squares	X	X	X	X
DKI+Offset (Morez et al., 2020)	TENS	Constrained Non-Linear fit	X	X	X	
DTD-cov (C. F. (Westin et al., 2016)	TENS	Constrained Non-Linear fit			X	X
DTD-cov (C. F. (Westin et al., 2016) + Offset	TENS	Constrained Non-Linear fit			X	X
Ball&Stick (Behrens et al., 2003)	MCM	Implemented in Dmipy (Fick et al., 2019)	X			
Ball&Racket (Sotiropoulos et al., 2012)	MCM	Implemented in Dmipy (Fick et al., 2019)	X			
NODDI-Watson (Zhang et al., 2012)	MCM	Implemented in Dmipy (Fick et al., 2019)	X			
NODDI -Bingham (Tariq et al., 2016)	MCM	Implemented in Dmipy (Fick et al., 2019)	X			
SMT (Kaden et al., 2016)	MCM	Implemented in Dmipy (Fick et al., 2019)	X			
NODDI-SMT	MCM	Implemented in Dmipy (Fick et al., 2019)	X			
MCMDI (Kaden et al., 2016)	MCM	Implemented in Dmipy (Fick et al., 2019)	X			
ActiveAx (D. C. (Alexander et al., 2010)	MCM	Implemented in Dmipy (Fick et al., 2019)	X			
SHORE (Ozarslan et al., 2009)	PAR	From DeepSHORE (Nath et al., 2019)	X	X	X	X
MAP-MRI Ozarslan et al., 2013	PAR	Custom implementation with regularization	X			
MAP-MRI+Reg (Fick et al., 2016)	PAR	Implemented in Dipy (Garyfallidis et al., 2014)	X	X		
NeuralNet	DL	Perceptron 1 Layer 50 nodes	X	X	X	X
NeuralNet+Reinf (Williams, 1992)	DL	Perceptron 7 Layers optimized with NAS (Zoph and Le, 2016)	X	X	X	X

Table 4
An overview of the diffusion models used to represent the individual components of the considered multicompartment models.

Model name	Intra-axonal component	Extra-Axonal component	Isotropic component	Orientation dispersion
ActiveAx	Cylinder	Zeppelin	Sphere	NA
Ball&Stick	Stick	NA	Sphere	NA
Ball&Racket	Stick	NA	Sphere	Bingham
MCMDI	Stick	Zeppelin	NA	NA
NODDI-Watson	Stick	Zeppelin	Sphere	Watson
NODDI-Bingham	Stick	Zeppelin	Sphere	Bingham
SMT	Zeppelin	NA	NA	NA
SMT-NODDI	Stick	Zeppelin	Sphere	Watson

2.2.4. Neural networks

Convolutional neural networks (NeuralNet) are increasingly being used for tasks such as quantification and signal representation. In this case, the output of the networks corresponded to the dMRI signal.

The first family of submissions that we received is based on feed-forward networks with a single hidden layer of 50 neurons and sigmoid activation functions. These networks were trained on 80% of the measurements and validated on 20% by minimizing the mean squared error of the predictions with the *AdamW* algorithm. For the learning phase, a

learning rate of 0.005 and 20000 epochs were used. To predict the SDE signals, the normalized components of the gradient (3 values) and the b-value were provided as inputs. For the DDE and DODE acquisitions, the gradient strength, the normalized components of the two gradients (6 values), the b-value, and the components of the b-matrix (6 values) were concatenated into one input vector of length 14. The training and prediction phase were repeated independently for each of the individual signal and data types.

The second family of submissions we received was based on neural networks with reinforcement learning (NN+Reinf) (Williams, 1992, Zoph and Le, 2016). A neural architecture search (NAS) was implemented to search the optimal 7-layer feed-forward model with ReLU activations for dMRI signal prediction given the acquisition parameters. The search space of NAS is the number of nodes in each of the seven layers in the set [8, 16, 32, 64, and 128]. that best fit the training data . For training, the initial learning rate was set to 0.01, and the adam optimizer was used.

2.3. Data analysis

The data analyses were performed separately for SDE-MS, SDE-GRID, DDE and DODE. For each encoding type and signal, we evaluated the min-max interval of all predictions, and their 25th to 75th percentile confidence interval. For each signal, we determined the best prediction as the one achieving the lowest mean squared residuals (MSE), and visually investigated the residuals. As the MSE only represents one of the possible metrics that can capture the goodness of the signal prediction, we also determined the variance of the residuals and the bayesian information criteria (BIC) associated with the prediction of each individual signal. These results can be found in the supplementary material Table S2 and S3.

Subsequently, the distribution of the residuals of each model were investigated by means of boxplots. At this stage, the residuals from all the 5 voxels were considered together. A ranking of all considered models was derived according to increasing MSE, then the best prediction model was established per encoding type. The residuals of the best prediction model were investigated as a function of the b-value, diffusion time (DDE only) and encoding frequency (DODE only), to understand whether all diffusion encodings were predicted with comparable accuracy and precision. In the subsequent analysis, the 5 voxels were studied separately. Specifically for the SDE-MS data, an additional analysis was performed to evaluate how the best model predicted the directional dependent information of the shell acquired at b = 4000s/mm^2. To this end, the measured data and the best signal prediction were projected on the unit sphere using spherical harmonics of order 12, then the prediction error was evaluated. The diffusion tensor imaging model was included in this step for reference.

3. Results

3.1. Signal representation of SDE-MS data

The geometric average of the SDE-MS signals and an overview of the predictions is shown in Fig. 2. In general, both the average and the best fitting method predicted the average signal decay without apparent biases, with the exception of the average fit of the low diffusion weightings of signals 3 and 5.

On average, the confidence interval of the submissions (blue error bar) is centered on the geometric average of the data for diffusion-weightings 400 < b < 4000 s/mm^2 for all 5 signals. The prediction of data at b = 4000 s/mm^2 (which was not provided in the training data) was overall accurate in WM (signals 1-3), but showed a small and consistent over-estimation in DGM and GM. The signal measured at low diffusion weightings (i.e. b < 200 s/mm^2) was on average accurately predicted in the WM voxel containing up to 2 crossing fibers and in deep gray matter, but not in the complex WM-configuration (Signal 3)

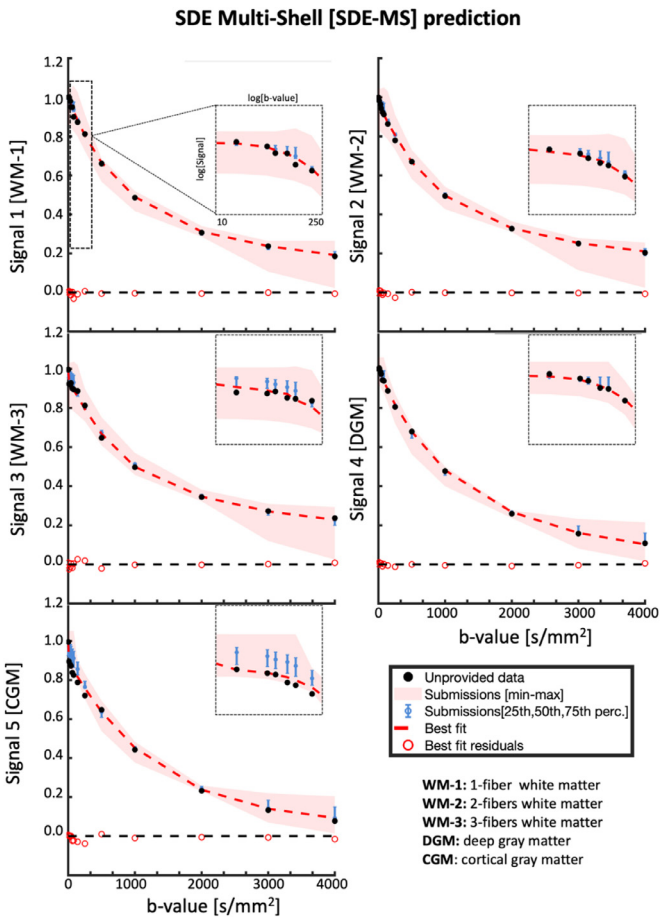

Fig. 2. Signal decay as a function of the b-value of the averaged SDE-MS data over different directions, for the unprovided measurements. The black dots represent the unprovided data, the red shaded area represents the min-max of the submissions, the blue error bars represent the 25-75 percentile, the solid red curve represents the best fit and the red circles represent the residuals of the best fitting model. The 5 different plots illustrate the predictions of the different signals.

and in cortical GM (Signal 5). The min-max range of the predictions highlights that data at b = 2000 s/mm^2 is predicted on average with the lowest uncertainty, whereas the largest spread is observed at b < 200 s/mm^2 and b = 4000 s/mm^2. The best predicting models for signals 1-5 are SHORE, MAP-MRI+Reg, MAP-MRI+Reg, Ball&Racket and NeuralNet, respectively.

Fig. 3 shows the average residuals of all tested signal predictions when considering the 5 provided signals together. The predictions of the first 7 methods were remarkably accurate, as shown by the tight confidence interval of the residuals of about 0.03 of the measured data, which occasionally reached values up to 0.1. MAP-MRI provided the best overall signal prediction with average MSE 0.00236 ± 0.00035 (Supplementary Material Table S1). When comparing the residuals of this prediction to those from the remaining models, we found that only NeuralNet provided equally distributed residuals (sign test, p≤0.05). When looking at the average residuals of each signal individually (colored dots in the boxplots of Fig. 3), a tendency towards a better prediction of WM signals as compared to CGM and DGM was observed.

To understand how the fit of the best prediction method (MAP-MRI) varied as a function of the diffusion-weighting, we evaluated the value of its average residuals for all 5 signals together for each shell independently. In general, the average residuals of most shells were close to zero, but a significant overestimation for b-shells 60 ≤ b ≤ 250 s/mm^2 was observed (one sample t-test, $p \leq 0.05$). The average prediction error

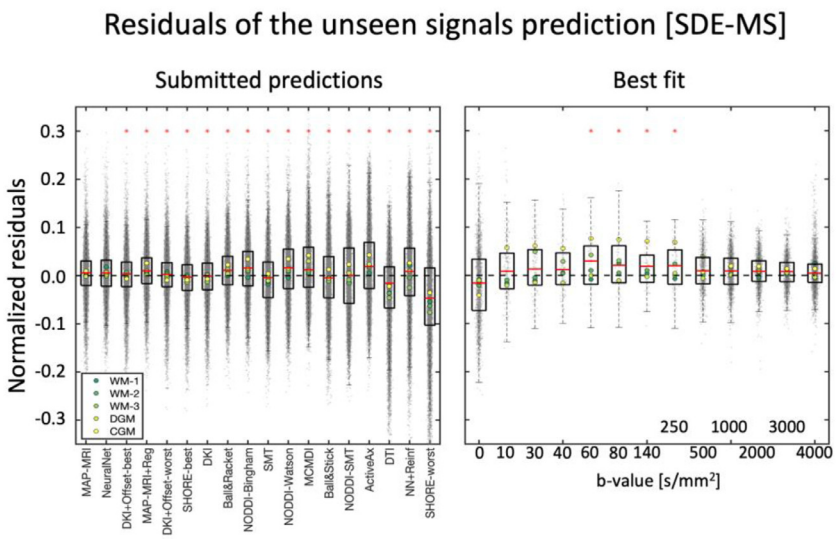

Fig. 3. Left) Boxplots of the normalized residuals (gray dots) of each prediction of SDE-MS data, when pooling together all 5 signals. Right) The normalized residuals of the best prediction (MAP-MRI) over individual diffusion weightings. The red asterisks on the left panel indicate predictions significantly different from the best prediction, whereas those on the right indicate that residuals at a specific diffusion weighting show a significantly non-zero mean.

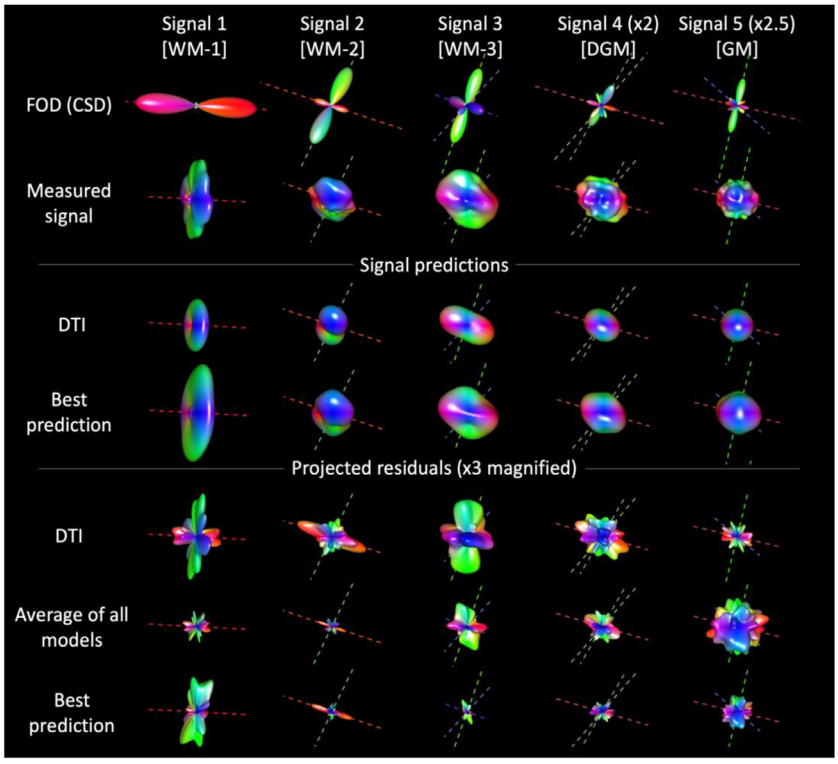

Fig. 4. 3D visualisation of the fiber orientation distribution (FOD), a projection on the unit sphere of the measured signal, of the signal predictions with DTI and with the best prediction model as well as of the residuals determined with DTI, average of all models and with the overall best predicting model for the unprovided SDE-MS data at b = 4000 s/mm^2

was on average less than 0.05 for all shells, but errors up to about 0.2 can be observed for the non-weighted data and the unprovided shell at b = 4000 s/mm^2.

Having established that model MAP-MRI provided the best average fit, we set to investigate how this method could predict the angular information of the unprovided shell at b = 4000 s/mm^2, and included a prediction with DTI and the average prediction of all methods for reference. This angular resolution analysis is shown in Fig. 4.

All methods could well-predict on average the donut-shaped 3D representation of WM-1, and the average residuals of this signal are smaller than those of the best fitting method. The DTI prediction of WM-1 well captured the overall shape of the signal, but also showed large errors in specific directions, likely due to unaccounted signal restrictions at this diffusion weighting. The best prediction was remarkably better than the average prediction as well as the DTI one in the more complex configurations WM-2 and WM-3. In the WM-signals, the largest angular errors were observed in directions parallel and perpendicular to the main diffusion directions derived with CSD, as expected. In DGM and CGM, all methods performed overall equally well and the residuals had an almost isotropic distribution.

3.2. Signal representation of SDE-GRID data

Fig. 5 reports the average signal predictions for SDE-GRID after binning closely spaced diffusion-weightings to enhance clarity. In general, the SDE-GRID was well predicted by most submissions, as highlighted by the tight min-max and confidence intervals. Larger prediction variance can be observed at low (b < 1000 s/mm^2) and large diffusion weightings (b > 6000 s/mm^2) than in the intermediate range. The best predictions of signals 1 and 2 were achieved with DKI+Offset, whereas MAP-

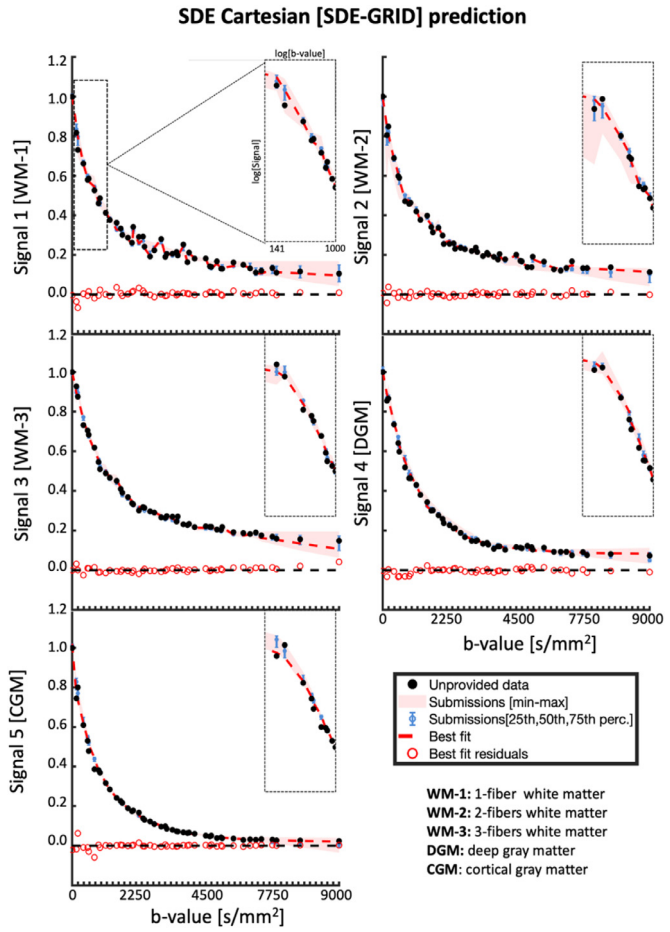

Fig. 5. Signal decay as a function of b-value of the averaged SDE-GRID data over different directions, for the unprovided measurements. The diffusion weightings were rounded to the closest multiple of 100 before averaging to enhance clarity. The black dots represent the unprovided data, the red shaded area represents the min-max of the submissions, the blue error bars represent the 25-75 percentile, the solid red curve represents the best fit and the red circles represent the residuals of the best fitting model. The 5 different plots illustrate the fits of the different signals.

MRI+Reg provided the best prediction of signals 3-5 and the best overall prediction with MSE 0.00260 ± 0.00043 (Supplementary Table S1).

The boxplots of residuals of the SDE-GRID predictions ranked by MSE are reported in Fig. 6. The first 7 submissions predicted the signals accurately, without visible biases both at average level as well as in specific tissue-types, and most prediction errors were in the range of ±0.03 with values occasionally up to 0.1, similar to what was previously observed for SDE-MS. When considering all predictions together, MAP-MRI+Reg provided the best overall prediction, but predictions with DKI+Offset and NeuralNet can be considered comparable according to a signed rank test. When analyzing the average prediction residuals of MAP-MRI+Reg for the binned diffusion weightings, it is appreciable that most data was well predicted without biases and errors below 0.05, with the exception of b ≤ 800 s/mm^2, b around 3000 s/mm^2 and b > 6000 s/mm^2.

3.3. Signal representation of DDE and DODE data

Fig. 7 shows the best and average signal predictions of the unseen DDE and DODE data for the 5 different voxels. Fig. 8 presents the normalized residuals for the different submissions, averaged over voxels, b-values and diffusion times/frequencies while Fig. 9 shows the normalized residuals for the best fitting model as a function of the b-value.

For DDE we see that the prediction of the directionally averaged signal is well aligned with the measured data with the DTD-cov+Offset providing the best prediction with MSE 0.00072 ± 0.00023 (Supplementary Table S1). Nevertheless, other methods such as DKI, SHORE and neural networks also performed reasonably well in providing unbiased predictions, but with visibly larger errors. In general, we see that the prediction of the higher b-values (> 2500 s/mm^2) is better than the prediction of the lower b-values (< 2500 s/mm^2). For DODE data, the best prediction comes from NeuralNet-best with MSE 0.00070 ± 0.00036, whereas the majority of the submissions overestimate the signal, especially for b-values larger than 1750 s/mm^2. For both DDE and DODE, the predictions show similar trends in the 5 different white matter voxels. For the DODE data, both frequencies also show similar trends of the predicted signal.

4. Discussion

We have evaluated the generalizability of existing dMRI methods at predicting diffusion-weighted data measured with SDE-MS, SDE-GRID, DDE and DODE by analyzing 80 submissions to the MEMENTO challenge from 7 different teams. In general, our analysis suggests that models predicting SDE-MS and SDE-GRID data generalized the easiest to unseen diffusion encodings, whereas the prediction of DDE and DODE seems more challenging. Within the domain of SDE, the worst prediction was observed in correspondence of low and very strong diffusion weightings.

4.1. Trends in SDE data predictions

The large majority of the analyzed submissions predicted SDE data, with a considerable preference for SDE-MS over SDE-GRID, which also reflects the overall larger number of studies which employ (multi-)shell data. When looking at SDE-MS, we can observe that the majority of submissions could well predict the global signal decay, and 14 out of 18 predictions had a median error smaller than 0.04. Of these, however, only 7 had an interquartile range (25th-75th percentile) of the residuals smaller than 0.05 in absolute value, which suggests how the prediction of the isotropic component of the signal decay (which captures the average decay of a given diffusion weighting) is an easier task than the prediction of the anisotropic component. Interestingly, the 7 predictions with the lowest MSE can account for complex fiber configurations such as 2+ crossing fibers, whereas predictions with single-fiber based methods result in higher MSE. Looking at the angular analysis reported in Fig. 4, it becomes clear that MAP-MRI provides the best prediction of both SDE-MS and SDE-GRID by well representing the signal in voxels with complex fiber configurations (WM-2, WM-3) as well as in DGM, whereas the prediction error in voxels with a single fiber population (WM-1) or almost isotropic diffusion (CGM) is worse than the average submission.

A second aspect regarding the analysis of SDE data is the dependence of the prediction accuracy and precision on the diffusion weighting and on the specific tissue type. Our results suggest that current dMRI methods can well represent and predict dMRI data with commonly used diffusion weightings. Indeed, we observed that most submissions could predict SDE data remarkably well within the range of commonly employed diffusion weightings (e.g., 1000 ≤ b ≤ 4000 s/mm^2), whereas the prediction of low (b < 800 s/mm^2) and strong (b > 6000 s/mm^2) diffusion weightings was generally less accurate. While the latter might originate from Rician-related biases, it might also highlight the limited specificity of existing models to genuine components such as perfusion contributions at low diffusion weightings (Le Bihan et al., 1988, Pasternak et al., 2009) and WM-restriction at strong diffusion weightings (Cohen and Assaf, 2002). In this context, a trend towards a worse prediction of the DGM and CGM signals as compared to WM signals emerges with SDE-MS and, to a lesser extent, with SDE-GRID. This seems to be mostly driven by the inaccurate prediction of the signal measured at low diffusion

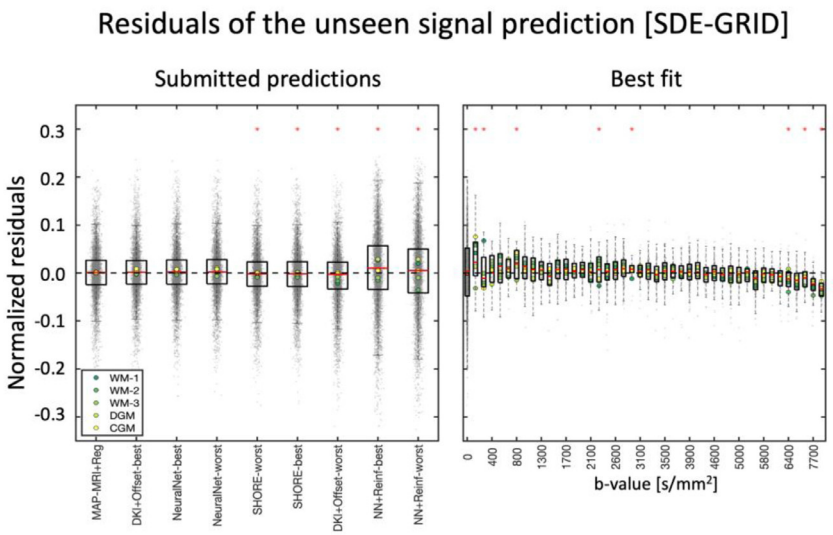

Fig. 6. Left) boxplots of the normalized residuals (gray dots) of each prediction of SDE-MS data, when pooling together all 5 signals. Right) The normalized residuals of the best prediction (MAP-MRI+Reg) over individual diffusion weightings. The red asterisks on the left panel indicate predictions significantly different from the best prediction, whereas those on the right indicate that residuals at a specific diffusion weighting show a significantly non-zero mean.

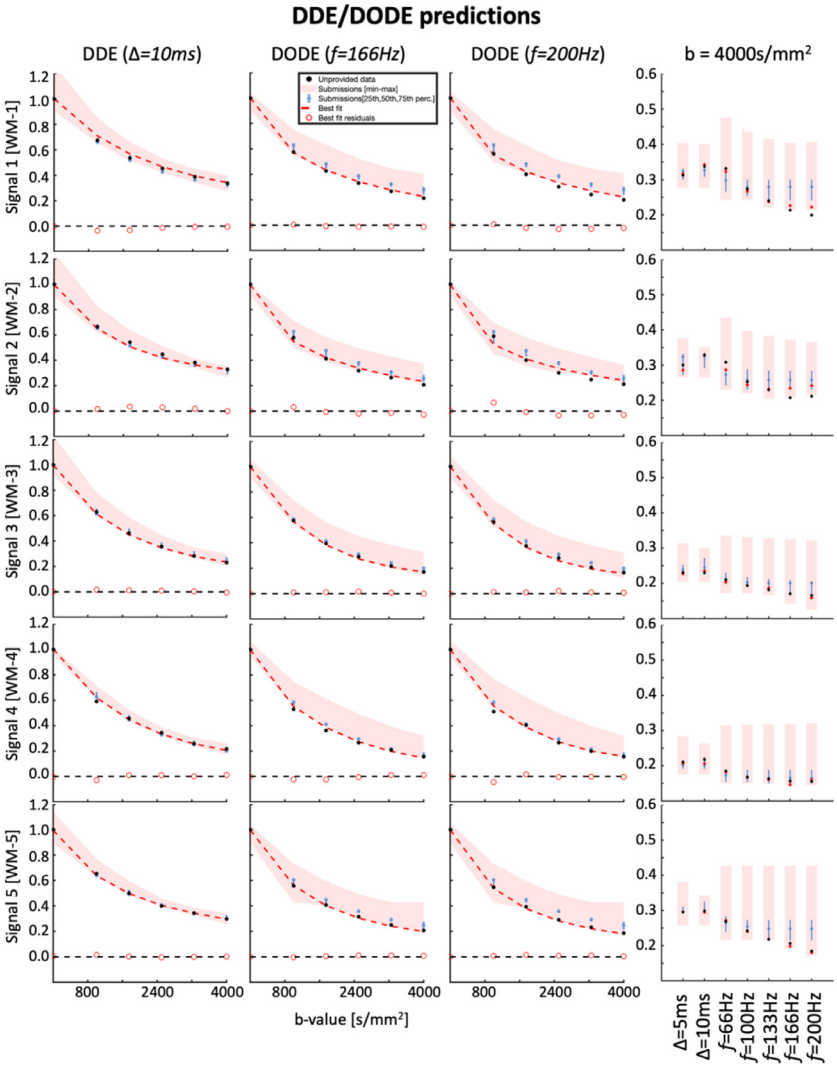

Fig. 7. The first three columns show the signal decay as a function of b-value of the geometrically averaged DDE and DODE data over different directions, for the unprovided measurements. The fourth column shows the geometric average of the signal measured at b = 4000 s/mm^2 for different diffusion times Δ (DDE) and oscillation frequencies f (DODE). The black dots represent the unprovided data, the red shaded area represents the min-max of the submissions, the blue error bars represent the 25-75 percentile, the solid red curve represents the best fit and the red circles represent the residuals of the best fitting model. The 5 different plots illustrate the fits of the different signals.

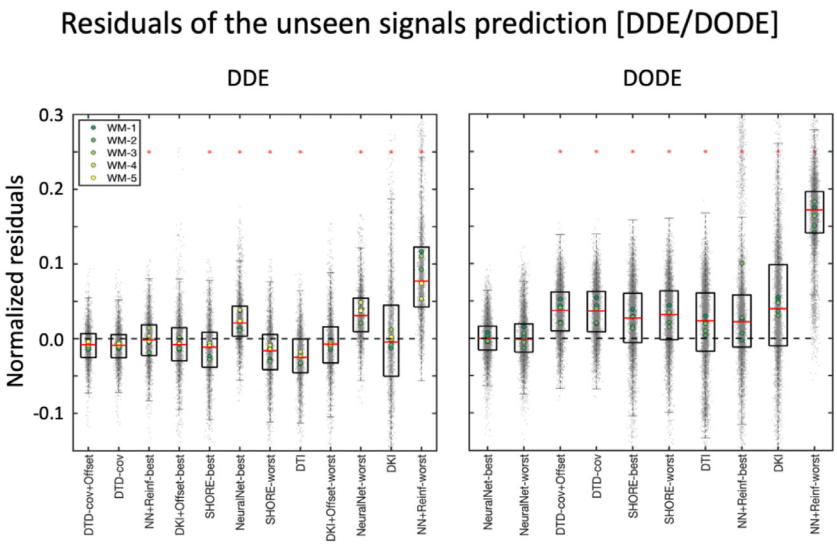

Fig. 8. The boxplots of the normalized residuals (gray dots) of the DDE (left) and DODE (right) predictions. The red asterisks on the panels indicate predictions significantly different from the best prediction. The first 5 DDE predictions perform reasonably well as shown by the value of most residuals being well-below 0.1, although a trend towards the overestimation of the signal could generally be observed.

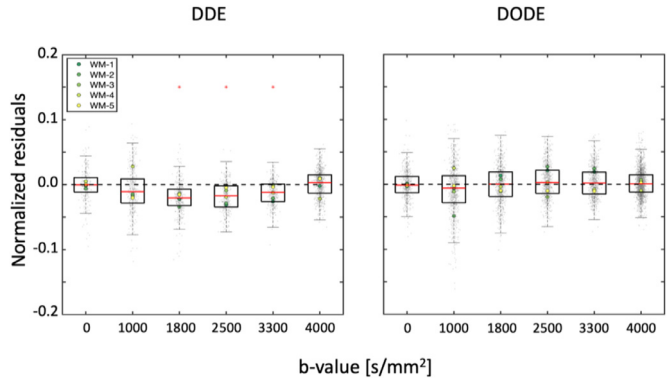

Fig. 9. Right) The normalized residuals of the best prediction of DDE (DTD-cov+Offset) and DODE (NeuralNet-best) over individual diffusion weightings. The red asterisks indicate that residuals at a specific diffusion weighting show a significantly non-zero mean. With DDE data, this was observed only for b-values between 1800 and 3300 s/mm^2, whereas no biases were observed for DODE.

weightings where the sensitivity to blood pseudo-diffusion is maximal, which once again suggests a lack specificity at taking into account specific properties of the GM like its higher perfusion as compared to WM (Ahlgren et al., 2016). This holds also for the best prediction (MAP-MRI) as shown by the significant overestimation of the signal at b < 250s/mm^2, where the contribution of pseudo-diffusion effects becomes non-negligible. A bias in the prediction of SDE-GRID with MAP-MRI is also revealed for b > 6000 s/mm^2. While this effect might be partially explained by Rician noise, the observation that its effect is larger in WM than GM, and that it grows in magnitude with fiber complexity being the largest for WM-3, suggests the presence of a genuine unaccounted trend in the signal. Interestingly, smaller errors are observed for the prediction of SDE-GRID than of SDE-MS on average, and even methods providing visibly biased SDE-MS predictions such as SHORE-worst, performed well at predicting SDE-GRID. This might be explained by the larger range of unique diffusion weightings included in SDE-GRID providing less redundant information than many measurements in few shells, or to the larger minimum diffusion weighting included in SDE-GRID (b = 141 s/mm^2) as compared to SDE-MS (b = 10 s/mm^2).

4.2. Trends in DDE and DODE data predictions

The prediction of DDE and DODE data seems more challenging than that of SDE. Indeed, the signal measured with DDE and DODE is encoded with additional dimensions as compared to SDE, namely parallel and orthogonal gradient orientations within one measurement, leading to linear and planar b-tensors, respectively, as well as different diffusion times and oscillation frequencies. In the challenge design, this aspect was stressed by requiring the prediction of unseen diffusion-weightings and gradient directions for 1 completely unseen diffusion time (DDE) and 2 unseen oscillation frequencies (DODE) provided a training set encoded with a different diffusion time and 3 different oscillation frequencies, which is arguably a harder task than the prediction of SDE data. The first take home message from the analysis of the DDE and DODE predictions is the need to take into account the additional encoding dimension, which is in line with our expectations given that the previous analysis of the data showed a clear diffusion time/frequency dependence over this parameter range (e.g. Fig. 9 in Ianus et al., 2018). For models which do not account for diffusion time / frequency (e.g. DTI / DKI / DTD, etc), we expect the predictions for the unseen DDE with $\Delta = 10$ ms to be the same as the model prediction for the provided DDE data with $\Delta = 5$ ms. Since for DDE only one diffusion time was provided, the neural networks could not learn the effect of this parameter, so in practice none of the submissions could account for diffusion time. For the unseen DDE data with $\Delta = 10$ ms, the best model from the current pool of submissions is DTD-cov+Offset, which nevertheless resulted in a small bias between the measurements and the predictions. The prediction errors of the directionally averaged signal provided by the best model are larger for intermediate b-values (e.g., b = 1800 – 3300 s/mm^2) than for the highest b-value of 4000 s/mm^2. This might be the case because the effect of diffusion time on the signal over this parameter range becomes smaller at higher b-value, for instance due to a smaller contribution of extracellular space.

For DODE data, multiple frequencies were provided for training. The best DODE predictions were achieved with methods able to account for the frequency implicitly, such as in the case of neural networks, whereas most submissions overestimated the measured signal. Interestingly, the error increases on average with the oscillation frequency, and even the best prediction, NeuralNet-best, could not predict well the data at b = 4000 s/mm^2 for f=200Hz. All submissions but one achieved a 25th to 75th percentile error below 0.1, but 8 out of 10 submissions exhibited a consistent median error of about 0.04. Altogether, this suggests in our opinion that we currently do not fully understand how to properly model the effect of frequency, and highlights the need for further research in the optimal modelling of DODE data.

4.3. Deep learning-based methods

The application of machine learning and deep learning is currently booming across all science fields dealing with large data. MRI and dMRI

are no exception, and in this very own challenge we have received 32 submissions based on deep learning methods next to established signal representations and biophysical models, which represents 40% of the total submissions. Interestingly, neural network-based methods provided accurate predictions for different diffusion encodings, achieving the 2nd best prediction for SDE-MS, the 3rd best prediction for SDE-GRID and DDE, and the best prediction of DODE data. The latter performance is remarkable, because NeuralNet-best and NeuralNet-worst provided the only two unbiased DODE predictions, showing ability to learn the relation between the diffusion signal and its encoding parameters, including the oscillation frequency, without the need for explicit modelling. These results certainly showcase the potential of these methods, and support their applicability as excellent interpolators, able to learn data features from a rich dataset and to provide good predictions of unseen data - within the boundaries of their training set. Most deep-learning based methods do not quantify metrics that can be used to extract properties of in-vivo tissues, and are thus unlikely to spread into clinical use at the current moment. Nevertheless, their good prediction performance make them favorable for tasks where a direct manipulation of the signal is required, such as denoising, artefact removal or even data augmentation, and their application in combination with classic methods might prove advantageous to enhance the quality of the results or to shorten acquisition time.

4.4. Importance of hyperparameters and user choices

Our analysis highlights how user choices and hyperparameters can remarkably affect the prediction accuracy. The SHORE method, for example, achieved both one of the best predictions of SDE-MS data as well as the worst, with an average error difference between the two of about 8%. Similarly, the addition of a degree of freedom to DKI or DTD-cov (i.e., +Offset) appreciably improved its accuracy. DKI+Offset, for example, predicted the SDE-MS and DDE data with an average prediction error 10% and 92% smaller than DKI, respectively. Large variability in the prediction performance was also observed for neural networks methods, which flexibility in design allows the implementation of very diverse architectures with performance strongly influenced by the optimization of hyperparameters. Altogether, we believe that this highlights the importance of not only reporting the specific method used for data analysis, but also to explain the choice of its hyperparameters and, where possible, to share its implementation to maximize the comparability of results obtained in different studies. Next to user settings and hyperparameters, model assumptions are an important factor that might limit the generalizability of a prediction method. This is especially the case for multi-compartment models based on physiological assumptions, including fixing the intra-cellular diffusivity to literature values or using tortuosity principles to constraint the cellular fraction, as recently suggested (Lampinen et al. 2019; Henriques et al. 2019).

4.5. Link to previous challenges

Various dMRI challenges have been organized in the past (Ferizi et al., 2017, Pizzolato et al., 2020, Schilling et al., 2019), with a focus on different data aspects. For instance the ones described by Schilling et al. (2019) focused on tractography, while others have focused on data prediction, including diffusion measurements at multiple echo times (Ferizi et al., 2017) as well as inversion times (Pizzolato et al., 2020). In terms of challenge requirements, the one organized by Ferizi et al. (2017) is the most similar to the current one, as it also evaluated the submissions based on the prediction of unseen SDE data shells.

Nevertheless, the majority of models submitted to the SDE part of this challenge do not overlap with the models included in the previous challenge, making a direct comparison difficult. One notable exception is MAP-MRI, which performed very well for the current SDE datasets, but not for the one considered by Ferizi et al. (2017) which included multiple diffusion and echo times. In that case, models which accounted for multiple compartments with different T2 properties provided better estimates, an effect not considered here. In addition to the development of more models and signal representations, the current work investigates generalizability over a wider range of diffusion weightings, angular directions, and diffusion times.

4.6. Limitations

Some limitations of this study should be acknowledged for a more comprehensive interpretation of the presented results.

Data limitations: Firstly, the SDE-MS/SDE-GRID and DDE/DODE data have been acquired in very different settings. The SDE data were acquired as part of the MASSIVE data (Froeling et al., 2017b) and represent an unique collection of thousands of unique diffusion measurements in an in-vivo human brain at 3T, but were spreaded in 18 acquisition sessions - which might introduce additional variability in the data - and are characterized by an overall modest SNR at $b = 0 s/mm^2$ ($\sim$15). Conversely, the DDE/DODE data were acquired in an ex-vivo mouse brain with a state-of-the-art 16.4T scanner, and characterized by very high SNR. Consequently, the generalizability of our results to DDE/DODE data acquired in in-vivo humans at 3T requires further research. While we expect our findings to generalize to individuals with similar characteristics, e.g., healthy adult humans (SDE) and mice (DDE/DODE), some results might be driven by the unique characteristics of these brains, and will likely not extrapolate well in presence of pathology, as well as in infants and elderlies.

Data selection: A further element of variability in the comparison of the two datasets is introduced by the different criteria used for the selection of the signals: on the SDE data, we sampled signals from WM voxels with different configurations (1 to 3 fibers), but also included GM voxels, which allowed us to investigate tissue-specific performance. As a consequence of this choice, any submissions regarding SDE-MS and SDE-GRID needed to perform well in both WM and GM to achieve a good score. Differently, all signals sampled from the DDE/DODE datasets were located in WM and offer a more thorough overview of the prediction performances across different fiber configurations - including regions with known fiber fanning - but no insights into their applicability to GM. For all of the above, the prediction performance obtained by the submissions on SDE and DDE/DODE should not be directly compared. In the selection of the voxels, we attempted to avoid tissue interfaces in order to minimize partial volume effects between different tissue types, but these cannot be ruled out and are expected to be more detrimental for methods not explicitly dealing with partial volume effects. Nevertheless, we argue that some extent of partial volume is ubiquitous in brain dMRI applications, and that taking it into account is likely part of the challenge to accurately model the dMRI signal.

Challenge evaluation: As a community challenge, we chose to calculate a single metric (the mean squared error) in order to determine a "winning" algorithm. Other choices of the score criteria were possible, and would likely result in a different ranking. For example, according to modelling theory it would seem more appropriate to investigate a goodness of fit criteria as the Bayesian information criteria rather than considering the signal residuals alone, to penalize signal overfitting (Supplementary Material Table S2 and S3). However, it is arguable that these kinds of metrics are not suitable to characterize methods based on machine learning / deep learning where thousands to millions of parameters are fitted, and that the mean squared error captures, in its simplicity and limitation, the basic ability to predict an unseen signal. Nevertheless, doing well in the current challenge does not automatically guarantee that these algorithms are the most appropriate models in all cases. Here, we have focused on the ability to explain (i.e. predict) the signal over a wide range of diffusion weightings, diffusion times, and frequencies. Furthermore, some modelling approaches in this study may be suitable only for a subset of the wide range of acquisitions in this database, and may be more/less sensitive at different areas of the diffusion sensitization space. Tensor-based models such as DTI and DKI, for

example, are known to well fit data in the range b = 800-1200 s/mm^2 and b = 1000-3000 s/mm^2, respectively. Unsurprisingly, in this challenge we indeed observe very large residuals for the DTI model at b < 500 s/mm^2 and b > 2000 s/mm^2, and for the DKI model at b < 800 s/mm^2 and b > 5800 s/mm^2, respectively, which penalize the final scoring of these methods. Another aspect that might influence the evaluation is the pre-processing of the data, which is well-established to have a major impact on the subsequent data analysis. To rule out its potential confounding effect on our results, we have provided the participants with standardized - already pre-processed data, but the inclusion of additional pre-processing steps (i.e., denoising, Gibbs ringing correction, outliers replacement, etc.) might have resulted in different prediction performance and "winners".

Lack of validation: The ability of a method to accurately predict unseen data offers a measure of fidelity to the underlying tissue microstructure, but it is by no means a substitute for validation efforts that compare signal models to the actual biological tissue structure obtained through orthogonal measurements such as, for example, high-resolution microscopy. Thus, we argue that the appropriateness and specificity of the tested methods cannot be adequately captured by signal fitting alone, and requires external validation, which is particularly critical in the case of biophysical multicompartment models. In addition to empirically assessing data, future work should continually strive to validate these measures against orthogonal information through simulations, physical phantoms, and animal models of tissue microstructure in order to paint the complete picture of the models successes and abilities.

4.7. What is a good model?

As described in past challenges (Panagiotaki et al., 2012; Ferizi et al., 2015; Jelescu et al., 2020) and in reviews (Jelescu et al., 2020; Novikov et al., 2018), a good model or signal representation must well-capture trends in the signal (explain seen signal and predict unseen signal), and also have stability and robustness of fit (Jelescu et al., 2016), *for the appropriate signal regime*.

On the other hand, a good model fit to the data, and ability to predict unseen data, does not guarantee that the estimated model parameters have a sensible physiological meaning. Similarly, a visually appealing map of quantitative indices also does not equate to a "good model". While the "best" model is the one that well-explains the underlying physiology that the signal is sensitive to (within the experimental design), the process of converging on the most-appropriate model is complex, and examining the generalizability of the model to various diffusion sensitizations is only one step in that process. This specific step lends insight into the information uncovered and captured in the signal, and successes and limitations of various attempts to describe the signal.

5. Conclusions

We have reported the results of a community effort to investigate the generalizability of existing methods at predicting unseen diffusion MRI signals collected over a large range of diffusion encodings. Our results highlight that existing models perform well at predicting SDE data in white matter and, to a lesser extent, in grey matter. Conversely, future work is needed to better understand and model the information content of DDE and DODE data. Next to the method choice, hyperparameters play a key role in the generalizability of fit methods, highlighting the importance of their optimization, and of reporting their values to support reproducibility. These challenge results serve not only as a snapshot of the current status quo in the field, but also as an openly available benchmark to support the development of novel methods.

Availability Statement

The provided and unprovided data of the Challenge are freely available together with the evaluation code at the following link: (https://github.com/PROVIDI-Lab/MEMENTO.git).

Credit authorship contribution statement

Alberto De Luca: Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Project administration, Writing – original draft, Writing – review & editing. **Andrada Ianus:** Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Project administration, Writing – original draft, Writing – review & editing. **Alexander Leemans:** Conceptualization, Funding acquisition, Writing – review & editing. **Noam Shemesh:** Conceptualization, Funding acquisition, Writing – review & editing. **Hui Zhang:** Conceptualization, Funding acquisition, Writing – review & editing. **Daniel C. Alexander:** Conceptualization, Funding acquisition, Writing – review & editing. **Markus Nilsson:** Conceptualization, Writing – review & editing. **Martijn Froeling:** Conceptualization, Writing – review & editing. **Geert-Jan Biessels:** Conceptualization, Funding acquisition, Writing – review & editing. **Mauro Zucchelli:** Formal analysis, Methodology, Writing – review & editing. **Matteo Frigo:** Formal analysis, Methodology, Writing – review & editing. **Enes Albay:** Formal analysis, Methodology, Writing – review & editing. **Sara Sedlar:** Formal analysis, Methodology, Writing – review & editing. **Abib Alimi:** Formal analysis, Methodology, Writing – review & editing. **Samuel Deslauriers-Gauthier:** Formal analysis, Methodology, Writing – review & editing. **Rachid Deriche:** Formal analysis, Funding acquisition, Writing – review & editing. **Rutger Fick:** Formal analysis, Methodology, Writing – review & editing. **Maryam Afzali:** Formal analysis, Methodology, Writing – review & editing. **Tomasz Pieciak:** Formal analysis, Methodology, Writing – review & editing. **Fabian Bogusz:** Formal analysis, Methodology, Writing – review & editing. **Santiago Aja-Fernández:** Formal analysis, Funding acquisition, Methodology, Writing – review & editing. **Evren Özarslan:** Formal analysis, Funding acquisition, Methodology, Writing – review & editing. **Derek K. Jones:** Formal analysis, Funding acquisition, Methodology, Writing – review & editing. **Haoze Chen:** Formal analysis, Methodology, Writing – review & editing. **Mingwu Jin:** Formal analysis, Methodology, Writing – review & editing. **Zhijie Zhang:** Formal analysis, Methodology, Writing – review & editing. **Fengxiang Wang:** Formal analysis, Methodology, Writing – review & editing. **Vishwesh Nath:** Formal analysis, Methodology, Writing – review & editing. **Prasanna Parvathaneni:** Formal analysis, Methodology, Writing – review & editing. **Jan Morez:** Formal analysis, Methodology, Writing – review & editing. **Jan Sijbers:** Formal analysis, Funding acquisition, Methodology, Writing – review & editing. **Ben Jeurissen:** Formal analysis, Funding acquisition, Methodology, Writing – review & editing. **Shreyas Fadnavis:** Formal analysis, Methodology, Writing – review & editing. **Stefan Endres:** Formal analysis, Methodology, Writing – review & editing. **Ariel Rokem:** Formal analysis, Methodology, Writing – review & editing. **Eleftherios Garyfallidis:** Formal analysis, Methodology, Writing – review & editing. **Irina Sanchez:** Formal analysis, Methodology, Software, Writing – review & editing. **Vesna Prchkovska:** Formal analysis, Methodology, Software, Writing – review & editing. **Paulo Rodrigues:** Formal analysis, Methodology, Software, Writing – review & editing. **Bennet A. Landman:** Conceptualization, Project administration, Funding acquisition, Writing – review & editing. **Kurt G. Schilling:** Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Project administration, Writing – original draft, Writing – review & editing.

Acknowledgements

The work by Inria co-authors was partially funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC Advanced Grant agreement No

694665: CoBCoM - Computational Brain Connectivity Mapping, P.I. Rachi Deriche) and by the French government, through the 3IA Côte D'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002.

Marco Palombo, Daniel C Alexander and Hui Zhang were supported by the EPSRC grant EP/N018702/1 and Marco Palombo by the UKRI Future Leaders Fellowship MR/T020296/1.

Jan Morez is supported by a grant (ISLRA-2009) from the European Space Agency, by Belgian Science Policy Office-Prodex. Ben Jeurissen and Jan Sijbers received funding from the Research Foundation Flanders (FWO Vlaanderen: 12M3119N; G0D7216N).

Maryam Afzali and Derek K Jones are supported by a Wellcome Trust Investigator Award (096646/Z/11/Z) and DKJ by a Wellcome Trust Strategic Award (104943/Z/14/Z).

Tomasz Pieciak acknowledges the Polish National Agency for Academic Exchange for grant PN/BEK/2019/1/00421 under the Bekker programme and the Ministry of Science and Higher Education (Poland) under the scholarship for outstanding young scientists (692/STYP/13/2018).

Fabian Bogusz acknowledges AGH Science and Technology, Poland (16.16.120.773).

Evren Özarslan is financially supported by Linköping University (LiU) Center for Industrial Information Technology (CENIIT), LiU Cancer, VINNOVA/ITEA3 17021 IMPACT, Analytic Imaging Diagnostic Arena (AIDA), and the Swedish Foundation for Strategic Research (RMX18-0056).

Andrada Ianus' work received the support of a fellowship from "la Caixa" Foundation (ID 100010434) and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 847648, fellowship code LCF/BQ/PI20/11760029.

Santiago Aja-Fernandez acknowledges the "Ministerio de Ciencia e Innovación" of Spain for research grant RTI2018-094569-B-I00. The work of Ariel Rokem and Eleftherios Garyfallidis is supported by grant 5R01EB027585-02 from the National Institute for Biomedical Imaging and Bioengineerring (PI: Eleftherios Garyfallidis).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118367.

Appendix A

A.1. Tensor-based models

- DTI: The diffusion tensor imaging method was fitted with a linear least squares procedure to determine the diffusion tensor (6 parameters) and the average non-weighted signal (1 parameter).
- DKI: The diffusion kurtosis imaging extends the DTI method to account for restricted diffusion. It was fitted with a weighted least squares procedure using ExploreDTI to determine 22 parameters: 6 for the diffusion tensor, 15 for the kurtosis tensor and the non-weighted signal. No additional constraints were considered in this fit.
- DKI+Offset: The DKI model was extended to accommodate an additional degree of freedom modelling a positive constant bias in the signal due to, for example, Rician noise. The 23 free parameters of this model were fitted with a non-linear least squares procedure implemented in MATLAB, constraining a monotonic signal decay and enforcing both the diffusion and kurtosis tensor to be positive definite.
- DTD-cov: The diffusion tensor distribution (DTD) method describes the diffusion signal as the sum of a distribution of microscopic tensors. The 28 parameters of the fourth order covariance tensor method were fitted to the data with a non-linear least squares procedure implemented in MATLAB, constraining a monotonic signal decay and enforcing both the diffusion and kurtosis tensor to be positive definite.
- DTD-cov+Offset: The DTD-cov method was extended with one additional degree of freedom modelling a positive constant bias in the signal due to, for example, Rician noise. The 29 free parameters of this model were fit with a non-linear least squares procedure implemented in MATLAB, constraining a monotonic signal decay and enforcing both the diffusion and kurtosis tensor to be positived definite.

A.2. Multi-compartment models

- Ball&Stick: originally proposed from Behrens and colleagues, this model consists of two compartments: a stick (impermeable cylinder with zero radius) to model anisotropic restricted intra-cellular diffusion, and a ball to model isotropic hindered extra-cellular diffusion. The model was implemented in Python using the Dmipy package, and its 4 parameters fitted to the data using a two stages procedure consisting of an initial grid search, followed by a constrained non-linear fit procedure based on a limited-memory quasi-Newton method.
- Ball&Racket: this model is an extension of the Ball&Stick that explicitly takes into account fanning configurations. The 7 parameters of the model were fitted to the data using the same procedure described for the Ball&Stick model.
- NODDI-Watson: originally introduce from Zhang et al., this model accounts for intra-cellular diffusion modelled as a tensor convolved with a Watson distribution to account for axonal dispersion, an extracellular compartment modelled with a Zeppelin, and an isotropic free water component to account for partial volume with the cerebrospinal fluid. The volumes of the intracellular and extracellular compartments are linked with a tortuosity principle, and the parallel diffusivity of the tensor is set to 1.7×10^{-3}mm^2/s. The 5 parameters of the model were fitted to the data using the same procedure described for the Ball&Stick model.
- NODDI-Bingham: this model extends the NODDI-Watson model to account for asymmetric fiber dispersion using a Bingham distribution. The 7 parameters of the model were fitted to the data using the same procedure described for the Ball&Stick model.
- SMT: The spherical mean technique (SMT) model provides estimates of neurite density and of the intrinsic tissue diffusivity unconfounded by fibre crossings and orientation dispersion. The 51 parameters of the model were fitted to the data using the same procedure described for the Ball&Stick model.
- NODDI-SMT: This is a reformulation of the NODDI-Watson model using the SMT technique. The 50 parameters of the model were fitted to the data using the same procedure described for the Ball&Stick model.
- MCMDI: this model describes intra-cellular diffusion with a stick, and extra-cellular diffusion with a Zeppelin. The SMT technique is used to achieve invariance to fibre crossing and orientation dispersion. The 50 parameters of the model were fitted to the data using the same procedure described for the Ball&Stick model.
- ActiveAx: introduced from Dyrby and colleagues, this model describes intra-cellular diffusion as a cylinder with finite radius, extracellular diffusion as a zeppelin, and accounts for isotropic contamination due to cerebrospinal fluid. The 7 parameters of the model were fitted to the data using the same procedure described for the Ball&Stick model.

A.3. Parametric representations

- SHORE: The method is based on the original simple harmonic oscillator reconstruction (SHORE). SHORE with optimized reconstruction

was tested at different orders of 6, 8 and up to 12. However, the best results or lower errors were determined to be at either order 6 or 8. The 50 parameters of the model were fitted to the data using a linear least-squares approach.

- MAP-MRI: Mean Apparent Propagator Magnetic Resonance Imaging (MAP-MRI) is a linear representation of the diffusion signal that uses a 3D generalization of the SHORE basis. The 95 parameters of the method were fitted using a penalized least-squares procedure with generalized cross validation implemented.
- MAP-MRI+Reg: This submission used the Laplacian-regularized MAP-MRI method of order 8 implemented in the Dipy software library with no positivity constraint in the propagator and a regularization weight of 0.47 to fit the 95 free parameters of the method.

A.4. Deep-learning methods

- NeuralNet: A fully connected neural network with a single hidden layer of 50 neurons and using sigmoid activation functions was trained to predict the unprovided signal amplitudes for each measurement independently. The 50 parameters of the network were optimized based on the mean squared error of the predictions using the ADAM algorithm with a learning rate of 0.005 over 20000 epochs. For SDE-MS and SDE-GRID the normalized components of the gradient (3 values) and the b-value were provided as inputs to the network. For the DDE and DODE acquisitions, the gradient strength, the normalized components of the two gradients (6 values), the b-value, and the components of the b-matrix (6 values) were concatenated into one input vector of length 14.

Table A1
The valid signal predictions submitted to the MEMENTO challenge. For each method, we report the acronym and the main reference, the "category", special notes on the fit procedure, and the data it has been applied to. The following predictions were subdivided in the following categories: tensor-based (TENS), multi-compartment model (MCM), parametric representation (PAR), deep learning-based (DL).

Model name	Category	Implementation details	Computation time [voxel]	Number of free parameters	Noise assumptions	Optimization algorithm
DTI (Basser et al., 1994)	TENS	-	< 1s	7	Gaussian	Linear Least Squares
DKI (J. H. (Jensen et al., 2005)	TENS	Implemented in ExploreDTI	< 1s	22	Gaussian	Weighted Least Squares
DKI+Offset (Morez et al., 2020)	TENS	Monotonic signal decay, positive definite tensor	< 1s	23	Rician	Non-linear least-squares
DTD-cov (C. F. (Westin et al., 2016)	TENS	Monotonic signal decay, positive definite tensor	< 1s	28	Gaussian	Non-linear least-squares
DTD-cov (C. F. (Westin et al., 2016) + Offset	TENS	Monotonic signal decay, positive definite tensor	< 1s	29	Rician	Non-linear least-squares
Ball&Stick (Behrens et al., 2003)	MCM	Implemented in Dmipy (Fick et al., 2019)	< 1s	4	Gaussian	Constrained non-linear least-squares
Ball&Racket (Sotiropoulos et al., 2012)	MCM	Implemented in Dmipy (Fick et al., 2019)	< 1s	7	Gaussian	Constrained non-linear least-squares
NODDI-Watson (Zhang et al., 2012)	MCM	Implemented in Dmipy (Fick et al., 2019)	< 1s	5	Gaussian	Constrained non-linear least-squares
NODDI-Bingham (Tariq et al., 2016)	MCM	Implemented in Dmipy (Fick et al., 2019)	< 1s	7	Gaussian	Constrained non-linear least-squares
SMT (Kaden et al., 2016)	MCM	Implemented in Dmipy (Fick et al., 2019)	< 1s	51	Gaussian	Constrained non-linear least-squares
NODDI-SMT	MCM	Implemented in Dmipy (Fick et al., 2019)	< 1s	50	Gaussian	Constrained non-linear least-squares
MCMDI (Kaden et al., 2016)	MCM	Implemented in Dmipy (Fick et al., 2019)	< 1s	50	Gaussian	Constrained non-linear least-squares
ActiveAx (D. C. (Alexander et al., 2010)	MCM	Implemented in Dmipy (Fick et al., 2019)	< 1s	7	Gaussian	Constrained non-linear least-squares
SHORE (Ozarslan et al., 2009)	PAR	From DeepSHORE (Nath et al., 2019)	< 1s	50	Gaussian	Regularized least squares
MAP-MRI Ozarslan et al., 2013	PAR		< 1s	95	Gaussian	Constrained quadratic programming
MAP-MRI+Reg (Fick et al., 2016)	PAR	Implemented in Dipy (Garyfallidis et al., 2014)	33s	95	Gaussian	Regularized least squares
NeuralNet	DL	Perceptron 1 Layer 50 nodes	Training: ~ 70s, Prediction: <1s	Signal dependent	Gaussian	Adam
NeuralNet+Reinf (Williams, 1992)	DL	Perceptron 7 Layers optimized with NAS (Zoph and Le, 2016)	NA	Up to 896	Gaussian	Adam

- NeuralNet+Reinf: A fully connected neural network with reinforcement learning. The authors adopted a neural architecture search (NAS) to identify the optimal 7-layer perceptron model for dMRI signal prediction with either 8. 16, 32, 64 or 128 nodes per layer. The free parameters of the network ranged between 56 and 896, and their values was optimized with the ADAM method using an initial learning rate equal to 0.01 and 200 training epochs.

Table A1.

References

Ahlgren, A., Knutsson, L., Wirestam, R., Nilsson, M., Ståhlberg, F., Topgaard, D., Lasič, S., 2016. Quantification of microcirculatory parameters by joint analysis of flow-compensated and non-flow-compensated intravoxel incoherent motion (IVIM) data. NMR Biomed. 29 (5), 640–649.

Alexander, A.L., Lee, J.E., Lazar, M., Field, A.S., 2007. Diffusion tensor imaging of the brain. Neurotherapeutics 4, 316–329.

Alexander, D.C., 2008. A general framework for experiment design in diffusion mri and its application in measuring direct tissue-microstructure features. Magn. Reson. Med. 60, 439–448.

Alexander, D.C., Dyrby, T., Nilsson, M., Zhang, H., 2017. Imaging brain microstructure with diffusion MRI: practicality and applications. NMR Biomed. 32, e3841.

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, 1374–1389.

Assaf, Y., 2019. Imaging laminar structures in the gray matter with diffusion MRI. Neuroimage 197 (August), 677–688.

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, 1347–1354.

Atkinson-Clement, C., Pinto, S., Eusebio, A., Coulon, O., 2017. Diffusion tensor imaging in Parkinson's disease: review and meta-analysis. NeuroImage. Clin. 16 (July), 98–110.

Baker, LM., Laidlaw, DH., Conturo, TE., Hogan, J., Zhao, Yi, Luo, Xi, Correia, S., et al., 2014. White matter changes with age utilizing quantitative diffusion MRI. Neurology 83 (3), 247–252.

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

Basu, S., Fletcher, T., Whitaker, R., 2006. Rician noise removal in diffusion tensor MRI. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006. Springer Berlin Heidelberg, pp. 117–125.

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, 1077–1088.

Blumenfeld-Katzir, T., Pasternak, O., Dagan, M., Assaf, Y., 2011. Diffusion MRI of structural brain plasticity induced by a learning and memory task. PLoS One 6 (6), e20678.

Burcaw, L., Fieremans, E., Novikov, D.S., 2015. Mesoscopic structure of neuronal tracts from time-dependent diffusion. Neuroimage 114, 18–37.

Chang, L.-.C., Jones, DK., Pierpaoli, C., 2005. RESTORE: robust estimation of tensors by outlier rejection. Magn. Reson. Med. 53 (5), 1088–1095.

Cohen, Y., Assaf, Y., 2002. High B-value Q-space analyzed diffusion-weighted mrs and mri in neuronal tissues - a technical review. NMR Biomed. 15 (7–8), 516–542.

Costabile, JD., Alaswad, E., D'Souza, S., Thompson, JA., Ryan Ormond, D., 2019. Current applications of diffusion tensor imaging and tractography in intracranial tumor resection. Front. Oncol. 9 (May), 426.

De Luca, A., Bertoldo, A., Froeling, M., 2017. Effects of perfusion on DTI and DKI estimates in the skeletal muscle. Magn. Reson. Med. 78 (1), 233–246.

De Luca, A., Leemans, A., Bertoldo, A., Arrigoni, F., Froeling, M., 2018. A robust deconvolution method to disentangle multiple water pools in diffusion MRI. NMR Biomed. 31 (11), e3965.

De Santis, S., Granberg, T., Ouellette, R., Treaba, CA., Herranz, E., Fan, Q., Mainero, C., Toschi, N., 2019. Evidence of early microstructural white matter abnormalities in multiple sclerosis from multi-shell diffusion MRI. NeuroImage. Clinical 22 (January), 101699.

Doan, N.T., Engvig, A., Persson, K., Alnæs, D., Kaufmann, T., Rokicki, J., Córdova-Palomera, A., et al., 2017. Dissociable diffusion MRI patterns of white matter microstructure and connectivity in Alzheimer's disease spectrum. Sci. Rep. 7 (March), 45131.

Does, M.D., Parsons, E.C., Gore, J.C., 2003. Oscillating gradient measurements of water diffusion in normal and globally ischemic rat brain. Magn. Reson. Med. 49, 206–215.

Fan, Q., Nummenmaa, A., Witzel, T., Ohringer, N., Tian, Q., Setsompop, K., Klawiter, E.C., Rosen, B.R., Wald, L.L., Huang, S.Y., 2020. Axon diameter index estimation independent of fiber orientation distribution using high-gradient diffusion MRI. Neuroimage 222, 117197.

Ferizi, U., Scherrer, B., Schneider, T., Alipoor, M., Eufracio, O., Fick, RH.J., Deriche, R., et al., 2017. Diffusion MRI microstructure models with in vivo human brain connectome data: results from a multi-group comparison. NMR Biomed. 30 (9). doi:10.1002/nbm.3734.

Ferizi, U., Schneider, T., Witzel, T., Wald, L.L., Zhang, H., Wheeler-Kingshott, C.A., Alexander, D.C., 2015. White Matter Compartment Models for in Vivo Diffusion MRI at 300mT/m. Neuroimage 118, 468–483.

Fick, RH.J., Wassermann, D., Caruyer, E., Deriche, R., 2016. MAPL: tissue microstructure estimation using Laplacian-regularized MAP-MRI and its application to HCP data. Neuroimage 134, 365–385.

Fick, RH.J., Wassermann, D., Deriche, R., 2019. The dmipy toolbox: diffusion MRI multi–compartment modeling and microstructure recovery made easy. Front. Neuroinform. 13 (October), 64.

Fieremans, E., Benitez, A., Jensen, J.H., Falangola, M.F., Tabesh, A., Deardorff, R.L., Spampinato, M.V., et al., 2013. Novel white matter tract integrity metrics sensitive to alzheimer disease progression. AJNR 34, 2105–2112.

Fieremans, E., Jensen, JH., Helpern, JA., 2011. White matter characterization with diffusional kurtosis imaging. Neuroimage 58, 177–188.

Froeling, M., Tax, CM.W., Vos, SB., Luijten, PR., Leemans, A., 2017a. 'MASSIVE' brain dataset: multiple acquisitions for standardization of structural imaging validation and evaluation. Magn. Reson. Med. 77 (5), 1797–1809.

Froeling, M., Tax, CM.W., Vos, SB., Luijten, PR., Leemans, A., 2017b. 'MASSIVE' brain dataset: multiple acquisitions for standardization of structural imaging validation and evaluation. Magn. Reson. Med. 77 (5), 1797–1809.

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Walt, S.V.D., Descoteaux, M., Nimmo-Smith, I., 2014. Dipy, a library for the analysis of diffusion MRI data. Frontiers in Neuroinformatics 8. doi:10.3389/fninf.2014.00008.

Ghosh, A., Ianus, A., Alexander, D.C., 2018.

Golkov, V., Dosovitskiy, A., Sperl, JI., Menzel, MI., Czisch, M., Samann, P., Brox, T., Cremers, D., 2016. Q-Space Deep learning: twelve-fold shorter and model-free diffusion MRI scans. IEEE Trans. Med. Imaging 35 (5), 1344–1351.

Grussu, F., Blumberg, S.B., Battiston, M., Kakkar, L.S., Lin, H., 2020. ' Select and Retrieve via Direct Upsampling' Network (SARDU-Net): a data-driven, model-free, deep learning approach for quantitative MRI protocol design. bioRxiv. https://www.biorxiv.org/content/10.1101/2020.05.26.116491v1.abstract.

Gudbjartsson, H., Patz, S., 1995. The rician distribution of noisy MRI data. Magn. Reson. Med. 34 (6), 910–914.

Henriques, RN., Palombo, M., Jespersen, SN., Shemesh, N., Lundell, H., Ianuş, A., 2020. Double diffusion encoding and applications for biomedical imaging. J. Neurosci. Methods November, 108989.

Ianus, A., Jespersen, S.N., Duarte, T.S., Alexander, D.C., Drobnjak, I., Shemesh, N., 2018. Accurate estimation of microscopic diffusion anisotropy and its time dependence in the mouse brain. Neuroimage 183, 934–949.

Ianus, A., Shemesh, N., Alexander, D.C., Drobnjak, I., 2017. Double oscillating diffusion encoding and sensitivity to microscopic anisotropy. Magn. Reson. Med. 78, 550–564.

Inglese, M., Bester, M., 2010. Diffusion imaging in multiple sclerosis: research and clinical implications. NMR Biomed. 23 (7), 865–872.

Jelescu, IO., Budde, MD., 2017. Design and validation of diffusion MRI models of white matter. Front. Phys. 28 (November). doi:10.3389/fphy.2017.00061.

Jelescu, IO., Palombo, M., Bagnato, F., Schilling, KG., 2020. Challenges for biophysical modeling of microstructure. J. Neurosci. Methods 344 (October), 108861.

Jelescu, IO., Veraart, J., Fieremans, E., Novikov, DS., 2016. Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue. NMR Biomed. 29 (1), 33–47.

Jelescu, I.O., Veraart, J., Adisetiyo, V., Milla, S.S., Novikov, D.S., Fieremans, E., 2014. 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.

Jensen, JH., Helpern, JA., 2010. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR Biomed. 23, 698–710.

Jensen, J.H., Helpern, J.A., Ramani, A., Lu, H., Kaczynski, K., 2005. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. Magn. Reson. Med. 53, 1432–1440.

Jespersen, S.N., Kroenke, C.D., Østergaard, L., Ackerman, J.J.H., Yablonskiy, D.A., 2007. Modeling dendrite density from magnetic resonance diffusion measurements. Neuroimage 34, 1473–1486.

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.

Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K., Sijbers, J., 2013. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. Hum. Brain Mapp. 34, 2747–2766.

Kaden, E., Kelm, N.D., Carson, R.P., Does, M.D., Alexander, D.C., 2016. Multi-compartment microscopic diffusion imaging. Neuroimage 139, 346–359.

Keil, VC., Mädler, B., Gielen, GH., Pintea, B., Hiththetiya, K., Gaspranova, AR., Gieseke, J., Simon, M., Schild, HH., Hadizadeh, DR., 2017. Intravoxel incoherent motion MRI in the brain: impact of the fitting model on perfusion fraction and lesion differentiability. J. Magn. Reson. Imaging 46 (4), 1187–1199.

Lasic, S., Szczepankiewicz, F., Eriksson, S., Nilsson, M., Topgaard, D., 2014. Microanisotropy imaging: quantification of microscopic diffusion anisotropy and orientational order parameter by diffusion MRI with magic-angle spinning of the Q-Vector. Front. Phys. 2.

Le Bihan, D., Breton, E., Lallemand, D., Aubin, M.L., Vignaud, J., Laval-Jeantet, M., 1988. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology 168.

Mitra, P.P., 1995. Multiple wave-vector extensions of the NMR pulsed-field-gradient spin-echo diffusion measurement. Phys. Rev. B 51, 15074–15078.

Morez, J., Sijbers, J., Vanhevel, F., Jeurissen, B., 2020. Constrained spherical deconvolution of nonspherically sampled diffusion MRI data. Hum. Brain Mapp. November. doi:10.1002/hbm.25241.

Nath, V., Lyu, I., Schilling, KG., Parvathaneni, P., Hansen, CB., Huo, Y., Janve, VA., et al., 2019. Enabling multi-shell B-value generalizability of data-driven diffusion models with deep SHORE. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer International Publishing, pp. 573–581.

Nath, V., Schilling, KG., Parvathaneni, P., Hansen, CB., Hainline, AE., Huo, Y., Blaber, JA., et al., 2019. Deep learning reveals untapped information for local white-matter

fiber reconstruction in diffusion-weighted MRI. Magn. Reson. Imaging 62 (October), 220–227.
Nedjati-Gilani, GL., Schneider, T., Hall, MG., Cawley, N., Hill, I., Ciccarelli, O., Drobnjak, I., Wheeler-Kingshott, CA.M.G, Alexander, DC., 2017. Machine learning based compartment models with permeability for white matter microstructure imaging. Neuroimage 150 (April), 119–135.
Novikov, DS., Kiselev, VG., Jespersen, SN., 2018. On modeling. Magn. Reson. Med. 79 (6), 3172–3193.
Novikov, D.S., Fieremans, E., Jespersen, S.N., Kiselev, V.G., 2019. Quantifying brain microstructure with diffusion MRI: theory and parameter estimation. NMR Biomed. 32, e3998.
Ouyang, M., Dubois, J., Yu, Q., Mukherjee, P., Huang, H., 2019. Delineation of early brain development from fetuses to infants with diffusion MRI and beyond. Neuroimage 185 (January), 836–850.
Ozarslan, E., Koay, C.G., Shepherd, T.M., Komlosh, M.E., Irfanoglu, M.O., Pierpaoli, C., Basser, P.J., 2013. Mean apparent propagator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure. Neuroimage 78, 16–32.
Ozarslan, E., Koay, C., Shepherd, T., Blackband, S., Basser, P., 2009. Simple harmonic oscillator based reconstruction and estimation for three-dimensional Q-Space MRI. ISMRM 17, 1396.
Palombo, M., Ianus, A., Guerreri, M., Nunes, D., Alexander, D.C., Shemesh, N., Zhang, G., 2020. SANDI: a compartment-based model for non-invasive apparent soma and neurite imaging by diffusion MRI. Neuroimage 215, 116835.
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, 2241–2254.
Pasternak, O., Sochen, N., Gur, Y., Intrator, N., Assaf, Y., 2009. Free water elimination and mapping from diffusion MRI. Magn. Reson. Med. 62, 717–730.
Pizzolato, M., Palombo, M., Bonet-Carne, E., Tax, CM.W., Grussu, F., Ianus, A., Bogusz, F., et al., 2020. Acquiring and predicting multidimensional diffusion (MUDI) data: an open challenge. In: Computational Diffusion MRI. Springer International Publishing, pp. 195–208.
Poulin, P., Jörgens, D., Jodoin, P.-M., Descoteaux, M., 2019. Tractography and machine learning: current state and open challenges. Magn. Reson. Imaging 64 (December), 37–48.
Ravi, D., Ghavami, N., Alexander, DC., Ianus, A., 2019. Current applications and future promises of machine learning in diffusion MRI. In: Computational Diffusion MRI. Springer International Publishing, Cham, pp. 105–121 Mathematics and Visualization..
Rokem, A., Yeatman, JD., Pestilli, F., Kay, KN., Mezer, A., Walt, S., Wandell, BA., 2015. Evaluating the accuracy of diffusion MRI models in white matter. PLoS One 10 (4), e0123272.
Scherrer, B., Schwartzman, A., Taquet, M., Sahin, M., Prabhu, S.P., Warfield, S.K., 2016. Characterizing brain tissue by assessment of the distribution of anisotropic microstructural environments in diffusion-compartment imaging (DIAMOND). Magn. Reson. Med. 76, 963–977.
Schilling, KG., Daducci, A., Maier-Hein, K., Poupon, C., Houde, J.-C., Nath, V., Anderson, AW., Landman, BA., Descoteaux, M., 2019. Challenges in diffusion MRI tractography – lessons learned from international benchmark competitions. Magn. Reson. Imaging doi:10.1016/j.mri.2018.11.014.
Sotiropoulos, SN., Behrens, TE.J., Jbabdi, S., 2012. Ball and rackets: inferring fiber fanning from diffusion-weighted MRI. Neuroimage 60, 1412–1425.
Stanisz, G.J., Szafer, A., Wright, G.A., Henkelman, R.M., 1997. An analytical model of restricted diffusion in bovine optic nerve. Magn. Reson. Med. 37, 103–111.
Stejskal, E.O., Tanner, J.E., 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J. Chem. Phys. 42, 288 – +.
Steven, AJ., Zhuo, J., Melhem, ER., 2014. Diffusion kurtosis imaging: an emerging technique for evaluating the microstructural environment of the brain. Am. J. Roentgenol. 202 (January), W26–W33.
Szafer, A., Zhong, J., Gore, J.C., 1995. Theoretical model for water diffusion in tissues. Magn. Reson. Med. 33, 697–712.
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.
Tax, CM.W., Jeurissen, B., Vos, SB., Viergever, MA., Leemans, A., 2014. Recursive calibration of the fiber response function for spherical deconvolution of diffusion MRI data. Neuroimage 86 (February), 67–80.
Tournier, J.D., Mori, S., Leemans, A., 2011. Diffusion tensor imaging and beyond. Magn. Reson. Med. 65, 1532–1556.
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.
Tournier, J.D., Yeh, C.H., Calamante, F., Cho, K.H., Connelly, A., Lin, C.P., 2008. Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. Neuroimage 42, 617–625.
Wang, S., Peterson, DJ., Wang, Y., Wang, Q., Grabowski, TJ., Li, W., Madhyastha, TM., 2017. Empirical comparison of diffusion kurtosis imaging and diffusion basis spectrum imaging using the same acquisition in healthy young adults. Front. Neurol. 8 (March), 118.
Wedeen, VJ., Hagmann, P., Tseng, W.-Y.I., Reese, TG., Weisskoff, RM., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. 54 (6), 1377–1386.
Westin, C.-F., Knutsson, H., Pasternak, O., Szczepankiewicz, F., Özarslan, E., van Westen, D., Mattisson, C., et al., 2016. Q-Space trajectory imaging for multidimensional diffusion MRI of the human brain. Neuroimage 135 (July), 345–362.
Weston, PS.J., Simpson, IJ.A., Ryan, NS., Ourselin, S., Fox, NC., 2015. Diffusion imaging changes in grey matter in Alzheimer's disease: a potential marker of early neurodegeneration. Alzheimer's Res. Ther. 7 (1), 47.
Williams, RJ., 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach. Learn. 8 (3), 229–256.
Yablonskiy, DA., Bretthorst, G.L, Ackerman, JJ.H., 2003. Statistical model for diffusion attenuated MR signal. Magn. Reson. Med. 50 (4), 664–669.
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, 1000–1016.
Zoph, B., Le, QV., 2016. Neural architecture search with reinforcement learning. arXiv [cs.LG]. arXiv http://arxiv.org/abs/1611.01578 .