NeuroImage 224 (2021) 117425

Contents lists available at ScienceDirect

NeuroImage

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

Machine learning based white matter models with permeability: An experimental study in cuprizone treated in-vivo mouse model of axonal demyelination

Ioana Hill[a,1], Marco Palombo[a,1,*], Mathieu Santin[b,c], Francesca Branzoli[b,c], Anne-Charlotte Philippe[b], Demian Wassermann[d,e], Marie-Stephane Aigrot[b], Bruno Stankoff[b,f], Anne Baron-Van Evercooren[b], Mehdi Felfli[b], Dominique Langui[b], Hui Zhang[a], Stephane Lehericy[b,c], Alexandra Petiet[b,c], Daniel C. Alexander[a], Olga Ciccarelli[g], Ivana Drobnjak[a]

[a] *Centre for Medical Image Computing and Dept of Computer Science, University College London, London, UK*
[b] *Institut du Cerveau et de la Moelle épinière, ICM, Sorbonne Université, Inserm 1127, CNRS UMR 7225, F-75013, Paris, France*
[c] *Institut du Cerveau et de la Moelle épinière, ICM, Centre de NeuroImagerie de Recherche, CENIR, Paris, France*
[d] *Université Côte d'Azur, Inria, Sophia-Antipolis, France*
[e] *Parietal, CEA, Inria, Saclay, Île-de-France*
[f] *AP-HP, Hôpital Saint-Antoine, Paris, France*
[g] *Dept. of Neuroinflammation, University College London, Queen Square Institute of Neurology, University College London, London, UK*

ABSTRACT

The intra-axonal water exchange time (τ_i), a parameter associated with axonal permeability, could be an important biomarker for understanding and treating demyelinating pathologies such as Multiple Sclerosis. Diffusion-Weighted MRI (DW-MRI) is sensitive to changes in permeability; however, the parameter has so far remained elusive due to the lack of general biophysical models that incorporate it. Machine learning based computational models can potentially be used to estimate such parameters. Recently, for the first time, a theoretical framework using a random forest (RF) regressor suggests that this is a promising new approach for permeability estimation. In this study, we adopt such an approach and for the first time experimentally investigate it for demyelinating pathologies through direct comparison with histology.

We construct a computational model using Monte Carlo simulations and an RF regressor in order to learn a mapping between features derived from DW-MRI signals and ground truth microstructure parameters. We test our model in simulations, and find strong correlations between the predicted and ground truth parameters (*intra-axonal volume fraction f:* R^2 =0.99, τ_i: R^2 =0.84, *intrinsic diffusivity d:* R^2 =0.99). We then apply the model in-vivo, on a controlled cuprizone (CPZ) mouse model of demyelination, comparing the results from two cohorts of mice, CPZ (N=8) and healthy age-matched wild-type (WT, N=8). We find that the RF model estimates sensible microstructure parameters for both groups, matching values found in literature. Furthermore, we perform histology for both groups using electron microscopy (EM), measuring the thickness of the myelin sheath as a surrogate for exchange time. Histology results show that our RF model estimates are very strongly correlated with the EM measurements (ρ = 0.98 for f, ρ = 0.82 for τ_i). Finally, we find a statistically significant decrease in τ_i in all three regions of the corpus callosum (splenium/genu/body) of the CPZ cohort ($<\tau_i>$=310ms/330ms/350ms) compared to the WT group ($<\tau_i>$=370ms/370ms/380ms). This is in line with our expectations that τ_i is lower in regions where the myelin sheath is damaged, as axonal membranes become more permeable. Overall, these results demonstrate, for the first time experimentally and in vivo, that a computational model learned from simulations can reliably estimate microstructure parameters, including the axonal permeability .

1. Introduction

The intra-axonal water exchange time (τ_i), a parameter associated with axonal permeability, is an important microstructural property of the tissue, which has been linked with myelination in the central nervous system (Nilsson et al., 2013a). Several neurological conditions such as Multiple Sclerosis (MS) cause a breakdown of the myelin sheath through a process known as demyelination, which may lead to a decrease in the exchange time as the intra-axonal water molecules encounter less barriers. Changes in permeability have also been linked with pathologies such as Parkinson's disease (Volles et al., 2001) and cancer Hu et al. (2006), leading to a widespread interest in developing permeability-based biomarkers. Due to its sensitivity to the motion of water molecules within tissue, modelling of Diffusion-Weighted MRI (DW-MRI) data enables the estimation of τ_i. However, measuring it has

[*] Corresponding author.
E-mail address: marco.palombo@ucl.ac.uk (M. Palombo).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neuroimage.2020.117425
Received 24 July 2019; Received in revised form 29 September 2020; Accepted 30 September 2020
Available online 6 October 2020

been problematic due to the intractability of the mathematical expressions which accurately incorporate τ_i into analytical models.

So far, the biophysical models that incorporate permeability rely on assumptions that are either too simplistic (Callaghan, 1997, Codd and Callaghan, 1999, Vangelderen et al., 1994) or do not hold in human tissue (Grebenkov et al., 2014, Kärger et al., 1988). The Kärger model (Kärger et al., 1988) is the most widely used analytical model that incorporates permeability (Nilsson et al., 2010, Stanisz et al., 2005, Lätt et al., 2009). However, its assumptions (i.e. the individual pools of water are well mixed and not restricted) do not hold in white matter and the model was shown to fail when applied to highly permeable tissue (Fieremans et al., 2010). A measurement technique for accessing exchange is the apparent exchange rate (AXR) imaging, however, it requires a specialised imaging protocol (Lasič et al., 2011, Nilsson et al., 2013b).

Computational models bypass the need for analytical expressions and incorporate permeability by creating a mapping between simulations of the DW-MRI signal and the ground truth microstructure parameters. Nilsson et al. (2010) use Monte Carlo simulations with known ground truth parameters including permeability to generate a synthetic library of DW-MRI signals. Given a previously unseen signal, they estimate permeability via a nearest-neighbour algorithm. However, their approach requires new libraries to be generated for each acquisition protocol - which in some cases may represent a problem - and the nearest-neighbour algorithm in general does not have a good generalisation capacity.

Recently, Nedjati et al. (2017) apply for the first time a machine learning approach using a random forest (RF) trained on a database of rotationally invariant features derived from the DW-MRI signals simulated using synthetic substrates of densely packed cylinders. Rotationally invariant metrics (e.g. MD and FA from DTI) are metrics calculated from DW-MRI data that do not depend on the particular orientation of the underlying tissue with respect to the scanner reference frame, thus providing valuable metrics for inter-subject and across-platform analyses. The model proposed by Nedjati et al. (2017) uses an RF instead of standard model-fitting approaches based on minimization of (non-linear) least-squares because it is more computationally efficient; it is less prone to local minimum problems; and it naturally encodes even complex constraints on parameter combinations through appropriate choice of training data, while guaranteeing good generalisation. The novel RF model is shown to outperform the Kärger's model on synthetic and in-vivo human data by providing more reproducible and robust estimates of τ_i (Nedjati et al., 2017). However, their in-vivo approach is tested only qualitatively on just two MS patients. Furthermore, Nedjati et al. (2017) hypothesise that τ_i is linked with demyelination in MS lesions, but they do not show whether other underlying processes such as axonal swelling or orientation dispersion affect the estimates. Here, we aim to address these limitations.

The aim of this study is to experimentally test a machine learning based computational model with permeability using a highly controlled cuprizone-treated, in-vivo mouse model of demyelination (CPZ), and a direct comparison to histology. We adopt the RF framework introduced in Nedjati et al (2017) to estimate tissue microstructure parameters. Prior to our in-vivo experiments, we use simulations representative for our mouse data to investigate the sensitivity of the PGSE protocol used to acquire the in-vivo data to τ_i, and select the most informative b shells (i.e. b values and directions) with respect to this parameter. We additionally establish a benchmark performance for our model by testing its performance on simulations. To test the in-vivo performance of the model, we use two cohorts of mice: CPZ and healthy age-matched wild-type (WT), with DW-MRI scans and histology data. Our demyelination model allows us to investigate the direct correlation between the estimated exchange time and histological measurements of myelin thickness. Furthermore, we investigate the potentially confounding effects of dispersion and axonal swelling to eliminate any potential bias in our estimates of the exchange time. Finally, we analyse the correlations between the estimations of our model and histology data.

2. Methods

This section first describes the imaging protocol, in-vivo data acquisition, histology analysis and the machine learning model and then outlines the principal steps of our experimental framework. Firstly, we investigate using synthetic data the sensitivity of our imaging protocol to changes in τ_i. Secondly, we optimise our computational model through a shell selection process and establish a benchmark performance for our model in simulations. We first ensure there is a good match between the synthetic and in-vivo data and we investigate any bias in our machine learning predictions of τ_i by looking at the effect of potential confounding factors. Finally, we test the in-vivo performance of our machine learning model on a cuprizone mouse model of demyelination and we analyse the correlations between the predictions and the ex-vivo histological measurements available.

2.1. *Mouse data*

2.1.1. *In-vivo data acquisition*

We image two cohorts of 8-week old C57BL/6J female mice, CPZ (N=8) and WT (N=8), using the same scanner and acquisition protocol as presented below. All animal experiments are performed in accordance with the European Council Directive (88/609/EEC). Eight mice were fed 0.2% cuprizone for 6 weeks, which corresponds to a demyelination without recovery phase, and eight healthy age-matched wild-type (WT) mice of the same background were fed a normal chow diet and used as controls. All mice are scanned on a Bruker BioSpec 11.7T scanner using the protocol described in Section 2.1.2 below. The WT data used in this study are available in the public domain and can be found at https://zenodo.org/record/996889#.WgH5E9vMx24 (Wassermann et al., 2017). The authors do not have permission to share the data used in this study for the CPZ treated mice. All the code used for the analysis is available upon request to the corresponding authors.

We post-process the images by correcting for eddy currents using FSL-eddy (Smith et al., 2004). No motion artefacts are observed. We restrict our analysis to white matter voxels within the corpus callosum (CC). To select the CC voxels, we compute maps of linearity (C_L), planarity (C_P) and sphericity (C_S) (Westin et al., 2002) from the diffusion tensor (DT) fit to the shell at b = 1241 s/mm^2. We create the CC maps by selecting the voxels with C_L>0.3, C_P<0.4, C_S<0.5 and fractional anisotropy (FA) > 0.40 (value chosen to distinguish WM form GM and CSF voxels also in the cuprizone treated mice, where FA values can be lower than the WT ones). Following this procedure, we obtain masks of the corpus callosum whose thickness varies slightly across all mice, randomly and with no statistically significant differences. Specifically, the mean ± s.d. of the number of voxels comprising the CC mask in the WT group is 161±13 and in the CPZ group is 175±18. Following a two-tail t-test we find this difference statistically insignificant (p>0.05). Previous studies, such as Wu et al. (2008), showed statistically significant increase in the volume of CC of CPZ intoxicated mice compared to WT. However, we do not measure a statistically significant increase and the further investigation of this observation is out of the scope of the present study.

2.1.2. *Diffusion imaging protocol*

We use the same DW-PGSE protocol for synthetic and in-vivo data, optimised to maximise signal reconstruction accuracy under realistic time constraints (Filipiak et al., 2019). Our imaging protocol has 25 shells, each with one b=0 measurement and a different combination of diffusion gradient strength G and diffusion gradient separation Δ as summarised in *Table 1*. The resulting protocol has 345 measurements in total, diffusion gradient duration δ=5 ms, $|G_{max}|$=500 m Tm^{-1} and shell

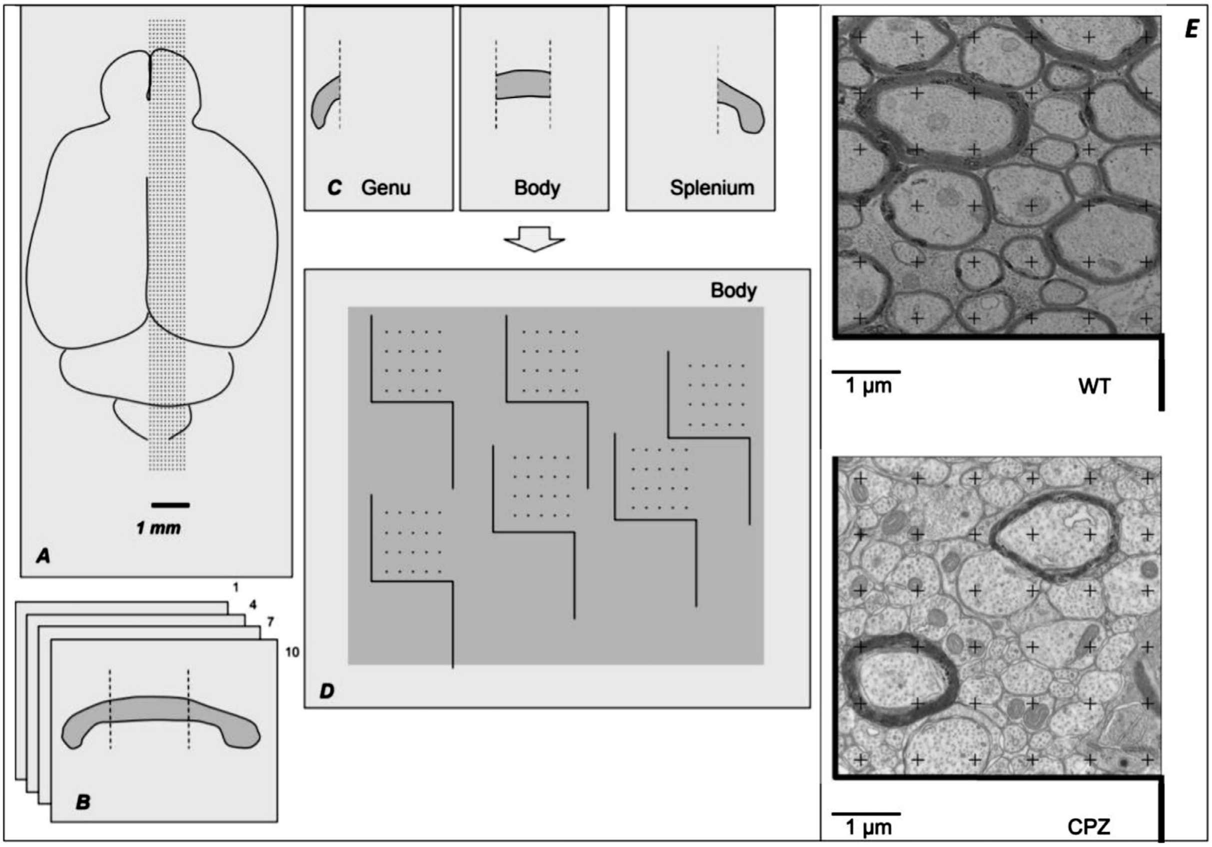

Fig. 1. Schematic pipeline of the stereological analysis to compute g_{ratios} and axonal diameters in the corpus callosum of the mice. First, Ten equally spaced slices are cut within the 1 millimeter from the middle of the corpus callosum in the sagittal section towards the edge of the brain (**A**). Then 4 slices are sampled starting from a random number. In this case, the randomly chosen starting number is 1, and the selected slices are #1, #4, #7 and #10 (**B**). Subsequently, these slices are used to localise the areas of interest (e.g., genu, body or splenium as shown in **C**), and each one of those is sliced ultra-thinly. On a randomly chosen ultra-thin slice for each of the ROIs, 30 spots are *also randomly chosen over the entire ROI at smaller magnification to assure that images are not intersected (**D** shows just 6 of those) before acquiring the final EM image at 62K magnification. Each of the 30 random spots are selected for stereological analysis using point grids of 36 regularly spaced crosses, each one representing an area of 0.5μm^2 (**E*** shows two of those 30 spots, one for WT and one for CPZ*). These point grids are used for quantification of the WT and CPZ mice.*

Table 1
DW-PGSE parameters with the corresponding nominal b-values in s/mm^2.

Δ (ms) **G (mT/m)**	**10.8**	**13.1**	**15.4**	**17.7**	**20**	**#grad dirs**
150	358	445	533	620	707	**16**
200	620	775	930	1086	1241	**16**
300	1384	1733	2083	2432	2781	**8**
400	2489	3110	3731	4352	4973	**11**
500	3892	4862	5833	6803	7773	**13**

b-values as shown in *Table 1*. Additional protocol details are as follows: TE=33.6 ms, TR=2 s, FOV=16 × 16 mm, matrix size = 160 × 160, number of slices=5, slice thickness=0.5 mm. Total acquisition time 53 min.

2.1.3. Histology samples

The WT (n=8) and CPZ (n=8) animals are sacrificed by deep anaesthesia and perfused intracardially with 1% paraformaldehyde and 2.5% glutaraldehyde in phosphate buffer 0.12 M, pH 7.4 at the end of the 6-week CPZ treatment. The extracted brains are then post-fixed overnight at 4 °C in the same fixative and rinsed in phosphate buffer. Ten 100μm-thick sagittal sections are cut with a vibratome (Thermo Scientific Microm HM 650 V Vibration microtome) (*Fig. 1A*). The very first section closest to the brain midline is considered as #1 and sections #1, #4, #7, and #10 are selected (*Fig. 1B*). Sections are post-fixed with 1% osmium tetroxide in water for 1 h at room temperature (RT°), rinsed 3 × 5 min with water and contrasted "en bloc" for 1 h at RT° with 2% aqueous uranyl acetate. After rinsing, sections are progressively dehydrated with 50%, 70%, 90%, and 100% ethanol solutions for 2 × 5 min each. Final dehydration is achieved by immersing the sections twice in 100% acetone for 10 min. Embedding is performed in epoxy resin (Embed 812, EMS, Euromedex, France) overnight in 50% resin / 50% acetone at 4 °C followed by 2 × 2 h in pure resin at RT°, and polymerization is achieved at 56 °C for 48 h in a dry oven. Semi-thin sections (0.5 μm-thick) are collected with an ultramicrotome UC7 (Leica, Leica Microsystèmes SAS, France) and stained with 1% toluidine blue in 1% borax buffer (*Fig. 1D*). Ultra-thin sections (70 nm-thick) are contrasted with Reynold's lead citrate (Reynold ES, 1963), and observed with a transmission electron microscope (HITACHI 120 kV HT 7700), operating at 70 kV. Images (2048 × 2048 pixels) are acquired with an AMT41B camera (pixel size: 7.4 μm x 7.4 μm) (*Fig. 1E*).

2.1.4. Post-mortem analysis

From the electron microscopy (EM) samples obtained as outlined in *Section 2.1.3*, we estimate the mean and standard deviation of the g_{ratio}, myelin thickness, axonal diameter and the intra-axonal volume fraction of the WT and CPZ mice. The stereological analysis is performed in isolated regions of the CC (genu, body and splenium), where 4 random sections with uniform distance are quantified per animal (*Fig. 1B*), with 30 randomly located images per region and per animal acquired at 62,000 magnification. For volume fraction (*VF*) we proceed according to the Delesse principle Mouton (2002): volume fractions are calculated by dividing the total number of points hitting the structure (*P(Y)*) by the total number of points hitting the reference volume (*P(ref)*), following the equation: $VF(Y,ref) = \frac{\sum_{i=1}^{m} P(Y)_i}{\sum_{i=1}^{m} P(ref)_i}$.

A grid of 36 regularly spaced crosses (*Fig. 1*E) is generated with Fiji, an open-source platform for biological image analysis (Schindelin et al., 2012). To identify non-perpendicular axons and remove them from the analysis, we take into account the shape of the axons and the microtubules inside them. Perpendicular axons have a minimally elongated shape and their microtubules are small perfectly circular structures inside them. In contrast, non-perpendicular axons have more elongated shapes (e.g. more ellipsoid-like) and their microtubules appear like lines, depending on the angle of the section. Stereological analysis provides Myelin Volume Fractions (MVF), Axon Volume Fractions (AVF), and the total Axon Volume Fractions (tAVF), which includes both myelinated and unmyelinated axons. Total Axon Count (tAxCount) is manually quantified. The g_{ratio} of myelinated fibers is then calculated as $g_{\text{ratio}} = \sqrt{\frac{AVF}{(MVF+AVF)}}$ and the mean axon diameters (DAX) are calculated as $\text{DAX} = 2 \times \sqrt{\frac{(tAVF \times surface)}{(\pi \times tAxCount)}}$.

The outliers induced by the non-perpendicular axons in the images are not taken into consideration. From the g_{ratio} and the DAX, myelin thickness is computed as: myelin thickness = $\frac{DAX}{2g_{ratio}}(1 - g_{ratio})$.

We compare the estimates of the RF with the EM measurements by computing the group-wise mean in the CC ROIs of the myelin thickness and intra-axonal volume fraction (VF) and looking at the correlation between these and the RF estimations for τ_i and *f*.

2.2. Synthetic data

A machine learning regressor can be trained on different databases. In this work, we aim to compare the performance of training directly on simulated signals versus training on features obtained by modelling those signals. Therefore, we construct two training databases: one comprised of synthetic DW-MRI signals and the other of rotationally invariant features estimated from those signals.

Each entry in the database corresponds to a unique digital phantom which mimics the in-vivo data and for which the ground truth microstructure parameters are known. Each synthetic database is used to train a machine learning algorithm, here an RF, to build a mapping between the signal or features and the corresponding ground truth microstructure parameters. Please note that in this context we refer to “features” in a machine learning sense: measurable properties or characteristics of the DW-MRI signal, and some of the features used may depend on some of the others.

2.2.1. Synthetic signals database

We use Monte Carlo simulations of the DW-MRI signal to build our synthetic training database. The signals are generated using the open source Camino (Cook et al., 2006; http://camino.cs.ucl.ac.uk) simulation framework Hall and Alexander (2009) together with the imaging protocol in *Table 1*. Using the Camino toolbox, we generated synthetic signals by first simulating the diffusion of many spins as three-dimensional random walk using Monte Carlo methods for each synthetic substrate composed of randomly packed straight cylinders. Then, from the simulated spins trajectories, the diffusion-weighted signal was computed using the phase accumulation approach, according to the specific diffusion-sensitising gradient scheme chosen to match the experimental acquisition protocol. Thus, each simulated signal corresponds to a digital phantom which mimics the in-vivo mouse brain data introduced in *Section 2.3*. The digital phantoms are represented by synthetic substrates that model white matter as a collection of 100,000 non-abutting, parallel cylinders with gamma-distributed radii, a common choice in the brain literature (Aboitiz et al., 1992). The cylinders are randomly packed in the substrates as described in Hall and Alexander (2009), with example substrates shown in *Fig. 2*. We construct a database of 11,000 unique tissue substrates and their corresponding DW-MRI signals by randomly sampling from a range of histologically plausible substrate parameters for white matter tissue (Aboitiz et al., 1992, Barazany et al., 2009). A white matter synthetic substrate is defined through five parameters: the mean $\mu_R \in [0.2,1]$ μm and the standard deviation $\sigma_R \in [\min(0.1, \mu_R/5), \mu_R/2]$ μm of the axon radii distribution, the intra-axonal volume fraction $f \in [0.4, 0.7]$, the intra-axonal exchange time $\tau_i \in [2, 1000]$ ms and the intrinsic diffusivity $d \in [0.8, 2.2]$ $\mu\text{m}^2\text{ms}^{-1}$. To ensure the convergence and the high precision of the simulated signals, we generate our synthetic database using 100,000 spins and 2,000 time steps (Hall and Alexander, 2009). The Monte Carlo simulations use displacements in continuous space, with fixed step size in three dimensions $s = \sqrt{6d\delta t}$ Einstein (1905), with $\delta t = 10$ μs. The permeability of a substrate is specified within the Camino simulation framework via the probability parameter *p*. This parameter expresses the probability that a spin steps through a membrane encountered during the random walk (instead of always being reflected backwards as it is the case for impermeable substrates). The probability *p* is related to the permeability *k* through the expression:

$$p = \frac{2}{3}k\sqrt{6\frac{\delta t}{d}},$$

where *d* is the intrinsic diffusivity and δt is the temporal resolution. This expression is obtained by combining the Monte Carlo step length equation $s = \sqrt{6d\delta t}$ (Hall and Alexander, 2009) with the transition probability equation as derived in (Regan and Kuchel, 2000, Fieremans and Lee, 2018). Here, we measure permeability *k* via the intra-axonal water exchange time τ_i, which is inversely related to *k* through the expression $k = \frac{R}{2\tau i}$, where *R* is the axon radius (Fieremans et al., 2010).

To maximise the performance of our machine learning regressor, we aim to build a training database that resembles as closely as possible the in-vivo data. For this, we generate an additional set of synthetic signals to account for the noise present in the in-vivo data. We add Rician noise with a standard deviation σ corresponding to an SNR of 40, which reflects the noise level of the b=0 images with the longest $\boldsymbol{\Delta}$.

2.2.2. Synthetic features database

In order to make the method generalizable across different scans and scanners, we train the RF regressor using a convenient database of features extracted from the DW-MRI signals that are independent of the specific orientation of the brain within the scanner (i.e. rotationally invariant features) (Novikov et al., 2018, Reisert et al., 2017). Towards this goal, *we obtain an equivalent rotationally invariant database by computing for each of the synthetic signals generated in Section 2.2.1 a set of 15 rotationally invariant features (see Table A.1)*, as done in Nedjati et al. (2017). We compute the DT and the 4th order spherical harmonic (SH) fit for each b shell from the synthetic signals using the Camino toolkit (Cook et al., 2006). We then derive 15 rotationally invariant features for each b shell and build an equivalent rotationally invariant synthetic database. The first five signal-derived features are calculated from the DT fit and are the three eigenvalues $\lambda 1$, $\lambda 2$, $\lambda 3$, the mean diffusivity (MD) and the fractional anisotropy (FA). The remaining ten features are derived from the SH fit: the mean, peak, anisotropy, skewness and kurtosis of the apparent diffusion coefficient together with the peak dispersion (i.e. the standard deviation of the peaks of the SH functions over a set of evenly distributed points in space) and combinations of the first, second and fourth order SH (Nedjati et al., 2017). Section A.1 in the Appendix presents in more detail what each of the 15 features represents and how it is computed.

2.3. Machine learning

2.3.1. Random forest (RF)

Due to their interpretability, robustness to noise and easiness of tuning (Criminisi et al., 2011), RFs are widely used as regression or classification techniques in the medical field (Alexander et al., 2017,

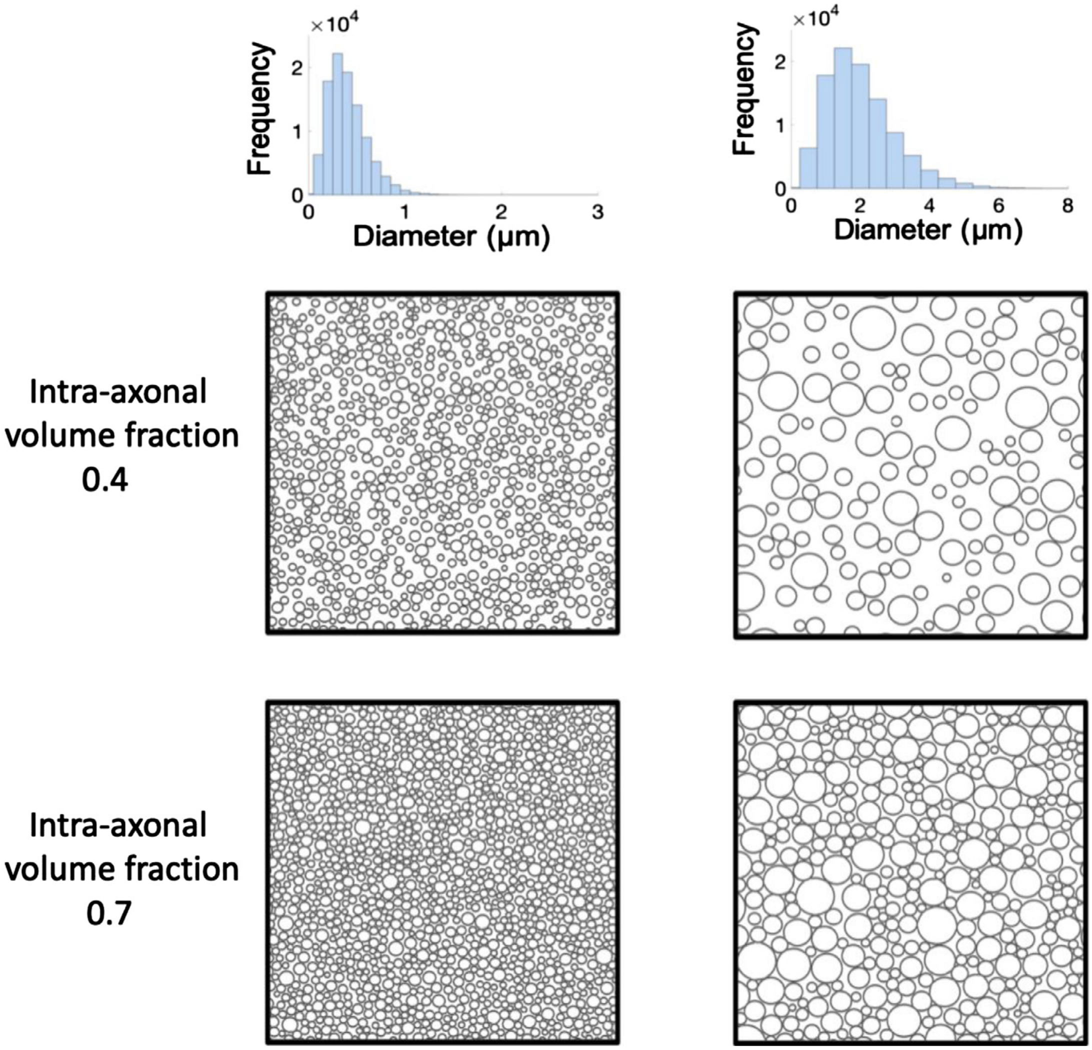

Fig. 2. Examples of the synthetic tissue used for our Monte Carlo simulations. From two given exemplar Gamma distributions of axon diameter (first row) four exemplar digital substrates are generated by packing straight non-overlapping cylinders up to two different intra-axonal volume fractions: 0.4 (second row) and 0.7 (third row).

Geremia et al., 2011, Nedjati-Gilani et al., 2017). An RF is an ensemble technique, built of a collection of decision trees, called *weak learners*. An RF regressor makes estimates by averaging the answers of all its decision trees, which are individually trained through a technique called *bagging*. This technique ensures the diversity of the trees by training each tree on a different random training subset. The randomness and diversity of the trees ensure their robustness to noise and good generalisation, resulting in the RF acting as a *strong learner* Breiman (2001). Here, we build an RF regressor that learns a mapping between the synthetic training database of DW-MRI signals/features and the ground truth microstructure parameters of the corresponding substrates. The mapping is learnt through a greedy splitting process of the input space (the synthetic signals/features) guided by the associated tissue parameters provided as labels during training.

During the learning phase, the training data is passed through the decision tree, starting at the root node towards the terminal nodes. At each node, the decision tree searches for a partition of the incoming data such that having separate partitions on either side of the node improves the estimation. If such a partition exists, the node is split and two child nodes are added on the level below. This procedure is repeated for every child node until splitting the data into smaller partitions does not improve the estimation anymore. If no better partition is found, the node becomes a terminal node. Mathematically, the training process is guided by the optimisation of a cost function, which is used to determine the best split at each node. The optimisation searches for the feature-threshold pairs (f_i,t_{fi}) that produce the best split. Here, we use the Classification and Regression Tree (CART) algorithm cost function J, defined as:

$$J\left(f_i, t_{fi}\right) = \frac{m_{left}}{m} MSE_{left} + \frac{m_{right}}{m} MSE_{right},$$

where $\mathrm{m}_{\mathrm{left/right}}$ is the number of training instances in the left/right subset and ‘MSE’ stands for the ’mean-squared-error’ between the ground-truth microstructure parameters (i.e. d, f and τ_i) known by design and the predicted ones.

There are two important parameters that need to be optimised to improve the learning performance of an RF: the number of trees and the maximum tree depth. The number of trees determines the smoothness of the decision boundary, and the tree depth parameter specifies the maximum levels that each decision tree can have. Too large a value can lead to overfitting while too low a value leads to underfitting, depending on the complexity of the data. Here, we run preliminary experiments and optimise these two parameters for our task in order to maximise the performance of our model.

2.3.2. *Training and testing*

We implement an RF regressor using the scikit-learn open source Python toolkit (Pedregosa et al., 2011). Following preliminary experiments, we build an RF with 200 trees of maximum depth 20 and bagging, as the setting that maximises the performance of the model. More general implementation details can be found at http://scikit-learn.org/. We train the RF for a multi-parameter regression task: we estimate the intra-axonal exchange time τ_i together with the intra-axonal volume fraction *f* and the intrinsic diffusivity *d*. Unlike the approach in Nedjati et al. (2017), we do not fit the axon radius index (Alexander et al., 2010) due to the lack of sensitivity of the signal to this parameter for our imaging protocol (Burcaw et al., 2015, Drobnjak et al., 2016).

The dimensionality of our synthetic databases is 11,000 by 345 for the signal database and 11,000 by 375 for the feature database. We set the size of training set to 11,000 as we did not find any improvements in performance above this number. The length of each synthetic training sample is reduced further during training according to the number of b shells selected in each training scenario. We train and test the RF on the synthetic databases using the associated ground truth parameters as labels for the supervised regression task. When predicting the parameter maps for the in-vivo data, we train the RF using the noisy databases as they are a more accurate representation of the in-vivo data. We split our synthetic database into a training set of 9,500 randomly selected signal/feature vectors and a test set formed of the remaining previously unseen 1,500 signal/feature vectors. As shown in Nedjati et al. (2017), the RF is not biased by the random selection of the training data as long as there is sufficient coverage of the parameter range, which we also ensure. To build the training set (to be done only once) it took approximately 3 days, using 50 nodes on our high-computing cluster of CPUs. The training of the machine learning model (to be done only once) took ~1 min and the prediction of the model parameters for $\sim 10^4$ exemplar voxels took ~1 min, on a 1.6 GHz dual-core Intel Core i5. Note that these times are just indicative, and they depend on the specific hardware used.

In this work we explore two possible ways of using machine learning for microstructure estimation: using a) signals or b) features of the signal to create the training database. The “signals training database” consists of standard DW-MRI signal intensities (normalized by the b=0) for a range of b values and gradient directions. The “Features training database” is created by replacing each signal at a given b value in the “signal training database” with 15 features, e.g. DTI and SH metrics, calculated using all the gradient directions for that voxel at that b value, as described in *Section 2.2.2*. While the first approach builds a direct mapping between the raw signals and the ground truth microstructure parameters, the second approach introduces an additional step of model fitting and constructs a mapping between DT and SH features of the raw signals and the microstructure parameters of interest. Because we chose to use rotationally invariant features, the second approach is generalizable across different scans and scanners.

2.4. Experiments

2.4.1. Sensitivity analysis

Firstly, we assess here that in the analysed data there is sufficient information about the targeted microstructural parameters, in particular τ_i. To ensure that there is enough information in the data, we investigate the sensitivity of our PGSE protocol to the intra-axonal exchange time by looking at the range of τ_i values for which the DW-MRI signal can be distinguished from that of an impermeable substrate. For this, we consider two synthetic substrates representative of mouse white matter tissue, with the following properties: the mean axonal diameter μ_D=0.4 μm and μ_D=2 μm, mimicking small and large axons in the CC, the intra-axonal volume fraction *f*=0.7 (Barazany et al., 2009), and the intrinsic diffusivity=1.2 μm^2 ms^{-1} (Wu et al., 2008). These substrates are a good representation of our in-vivo mice data, as shown by the histological measurements of μ_D in *Section 3.5*, all within the range of the gamma-distributions above. Note that the choice of fixing the diffusivity to 1.2 μm^2/ms is only made for the purpose of the sensitivity analysis to assess the suitability of the protocol. For all the other simulations in the machine learning analysis, the diffusivity is varied in the interval [0.8, 2.2] μm^2/ms, as done in Nedjati et al., 2017, and as shown in Wu et al., 2008 and Barazany et al., 2009 appropriate for rodents' brain. For application on human brain, higher diffusivity of ~2.2 μm^2/ms should be used for the sensitivity analysis, according to recent estimates of intra-axonal axial diffusivity in-vivo in the human brain (Dhital et al. 2019).

Using the Camino toolbox, we generate synthetic signals for each substrate and different values of δ, Δ and G, corresponding to the b shells in our PGSE protocol. The diffusion gradients are set perpendicular to the cylinders in the substrate to maximise sensitivity to τ_i. We investigate whether exchange time effects can be detected in the signal by looking at the difference in the normalised DW-MRI signal between impermeable (τ_i=∞) and permeable substrates. Moreover, we analyse the effect of noise by looking at a range of different SNRs: SNR=∞, SNR=40 and SNR=20, where SNR=40 corresponds to the level of noise present in our in-vivo data. By using synthetic substrates representative of our in-vivo data and the same imaging protocol, we expect the analysis in this section to provide an indicative range of exchange time values for which there is reasonable sensitivity in our in-vivo data.

2.4.2. Shell selection

As our imaging protocol uses an explorative range of imaging parameters, we select the b shells that maximise the performance of our RF model with respect to τ_i. For this, we evaluate the performance of our RF model for every possible combination of 4, 9 and 16 shells out of the 25 in our protocol. We first evaluate combinations of 4 shells using as a benchmark the 4-shell STEAM protocol (Nedjati et al., 2017) optimised Alexander (2008) for a two-compartment model with exchange and biophysically plausible tissue parameters. As there are 12650 possible combinations of 4 shells, we train the RF 12650 times, once on each different shell combination. Then, for each training scenario corresponding to a unique combination of shells, we compute the correlation coefficient R^2 for *f*, τ_i and *d* between the ground truth and the estimated values in the test set. Finally, we sort the different shell combinations according to their R^2 score for τ_i and choose the combination with the highest score as the one that maximises the performance of the model.

Furthermore, we investigate the effect of increasing the number of shells used for training. For this, we also look at combinations of 9 shells, as the minimum number of shells required to sample independently every unique G and **Δ** value in our PGSE protocol. Additionally, we look at combinations of 16 shells as a middle value between the 9-shell and the full protocol scenario. For this analysis, we use the synthetic feature-based dataset described in *Section 2.2.2*. Finally, we investigate the effect of noise on the performance of our model. For this, we look at a range of different SNRs: SNR=∞, SNR=40 and SNR=20.

2.4.3. Synthetic experiments

To assess the quality of the RF estimates after training is completed, we compute the Pearson correlation coefficient R^2 between the ground truth values and the RF estimates of the parameters in the previously unseen test set. To evaluate any potential bias in the estimates, we use Bland-Altman plots showing the mean of the estimated and ground truth values against their difference. We first analyse the performance of the model on the noise-free synthetic databases to establish a benchmark given our data and imaging protocol. Next, we apply our machine learning model to the SNR=40 database for a more accurate approximation of the performance we expect, given the noise present in our in-vivo data. For each experiment, we analyse both training scenarios outlined in *Section 2.4.2* (signal-based and feature-based) to test whether there are any differences in performance between the two approaches.

2.4.4. In-vivo imaging experiments

Before generating in-vivo parameter maps using our trained machine learning model, we first perform a data quality match to check that the dataset used to train our machine learning model represents well the characteristics of the in-vivo dataset.. In addition to this, we investigate

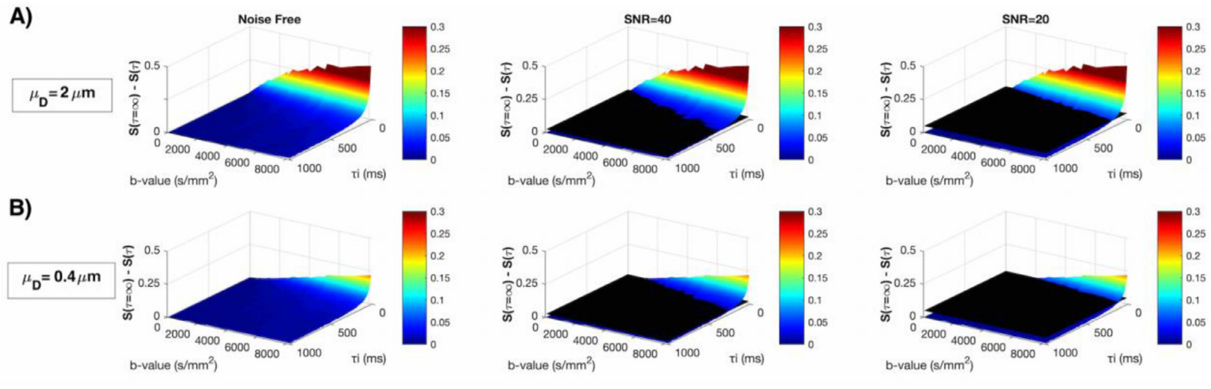

Fig. 3. Differences in the DW-MR normalized signal between impermeable (τ_i=∞) and the equivalent permeable (τ_i∈[20, 1000] ms) substrates at different b values, for different mean axonal diameter and SNRs and intra-axonal volume fraction f = 0.7. A) results for a substrate with mean axonal diameter μ_D=2 μm, representing large axons in the mouse brain. B) results for a substrate with μ_D=0.4 μm, mimicking small axons in the brain. The level of signal detectability is displayed for three SNR levels (∞, 40 and 20), represented by the black planes, below which any change in signal is undetectable.

any potential bias in our in-vivo estimates of τ_i due to changes in the orientation dispersion by computing maps of the NODDI orientation dispersion index (ODI) (Zhang et al., 2012a) using the NODDI Matlab (The MathWorks, Inc, Natick, MA) Toolbox[1]. Using the Camino toolbox, we additionally generate DTI maps at b=1241 s/mm^2 of axial diffusivity (AD), fractional anisotropy (FA) and radial diffusivity (RD) as measures of tissue properties that can be compared with already published works in cuprizone model (Boretius et al., 2012, Song et al., 2005, Zhang et al., 2012b).

Using the RF trained on the noisy database, we generate parameter maps for the CCs of the 16 mice for three parameters of interest: τ_i, *f* and *d*. To investigate the difference between the two groups (CPZ and WT), we compute box-and-whisker plots of region-specific comparisons between WT (8 mice) and CPZ (8 mice) for the DTI and NODDI metrics as well as for the RF estimates. Statistical significance is assessed by a two-tailed t-test, considering p-values<0.05. We run these experiments using the signals database. The Camino feature extraction of the in-vivo data did not produce histologically plausible results for the shells with very high gradient strengths (G>300 mT/m) in our protocol, and we therefore exclude this training approach from the analysis in this section. We discuss the potential explanations and the implications of this in *Section 4.1*.

3. Results

3.1. Sensitivity analysis

Fig. 3 shows the range of exchange time values for which the DW-MRI signal S(τ_i) can be distinguished from that of an impermeable substrate S(τ_i=∞) in the presence of noise. For this, we calculate the change in signal |S(τ_i=∞)-S(τ_i)| between an impermeable and an equivalent permeable substrate. To illustrate practically achievable sensitivities, we plot this difference against three noise levels, denoted by the black plane: SNR=∞ (1st column), SNR=40 (2nd column) and SNR=20 (3rd column). *Fig. 3A* illustrates the results for a substrate mimicking large axons in the white matter (μ_D=2 μm), while *Fig. 3B* corresponds to a substrate with smaller axons (μ_D=0.4 μm). The second column shows that, for substrates with large axons (*row A*) and an SNR of 40, matching that of our in-vivo data, it is possible to distinguish exchange time effects for values of $\tau_i \leq$ 400 ms. For substrates with small axons (*row B*), we can distinguish only permeable substrates with exchange times up to $\tau_i \leq$ 250 ms. As expected, when the SNR drops to 20, it becomes harder to distinguish between impermeable and permeable substrates. This trend can be observed in the 3rd column, where the range for distinguishable permeable substrates narrows from τ_i ϵ [0, 400] ms to τ_i ϵ [0, 200] ms for large axons and from τ_i ϵ [0, 250] ms to τ_i ϵ [0, 140] ms for small axons.

3.2. Shell selection

As our original 25-shell PGSE protocol uses an explorative range of imaging parameters, we choose the shells most sensitive to the exchange time (see *Section 2.5.2* for further details). In *Fig. 4,* each point on the *x-axis* represents one unique shell combination and the corresponding *y-axis* value indicates the R^2 score when the RF is trained on that particular shell combination. For example, the *x-axis* in *Fig. 4A* will have 12650 points, each one corresponding to one of the 12650 unique 4-shell combinations. As we are interested in the performance of the model with respect to τ_i (1*st* column), we rearrange the shell combinations in increasing order according to their R^2 for τ_i. This results in a monotonically increasing curve for τ_i, as seen in the first column. For *f* (2*nd* column) and *d* (3*rd* column)*,* we keep the x-axis ordering consistent with the results for τ_i in the 1*st* column.

The R^2 scores curves in the 1*st* column of *Fig. 4* show that only a limited number of shell combinations have a good correlation coefficient and are optimal for estimating τ_i, while the R^2 scores in the 2*nd* and 3*rd* column show that the majority of shell combinations provide good estimates of *f* and *d*. For example, in the noise free (blue curves) 4-shell case in the top row, we notice that the difference in R^2 score for τ_i between the best and the worst performing shell combinations is approximately 0.5. In contrast, this difference is much narrower for *f* and *d*: ≈0.02 for *f* and ≈0.01 for *d.* We observe the same trends for SNR=40 (orange) and SNR=20 (green).

By comparing the best R^2 scores on the blue curves in *Fig. 4A* and *Fig. 4B*, we can see that there is no difference in performance in the noise free scenario between using the best combination of 4 or 9 shells. However, this changes with the addition of noise. For example, for SNR=40 (orange curves) the R^2 score of the best 9-shell combination is 0.67, 0.07 higher than for the best 4 shells. This trend is similar for SNR=20 (green curves), with a difference of 0.1 between 9 and 4 shells. For the 16-shell scenario, we find no improvement in performance over using 9 shells.

Fig. 4 also shows the effect of noise on the estimation of each parameter. As expected, the addition of noise results in lower R^2 scores, a trend that holds for all parameters and across the 4 and 9-shell case. However, the estimation of τ_i is the most affected by the presence of noise: the maximum correlation coefficient drops from 0.82 in the noise free case to 0.67 for SNR=40 and even further to 0.52 for SNR=20. For *f* (2*nd* column), the effect of noise is considerably smaller: R^2 drops from 0.99 for

[1] http://mig.cs.ucl.ac.uk/index.php?n=Tutorial.NODDImatlab.

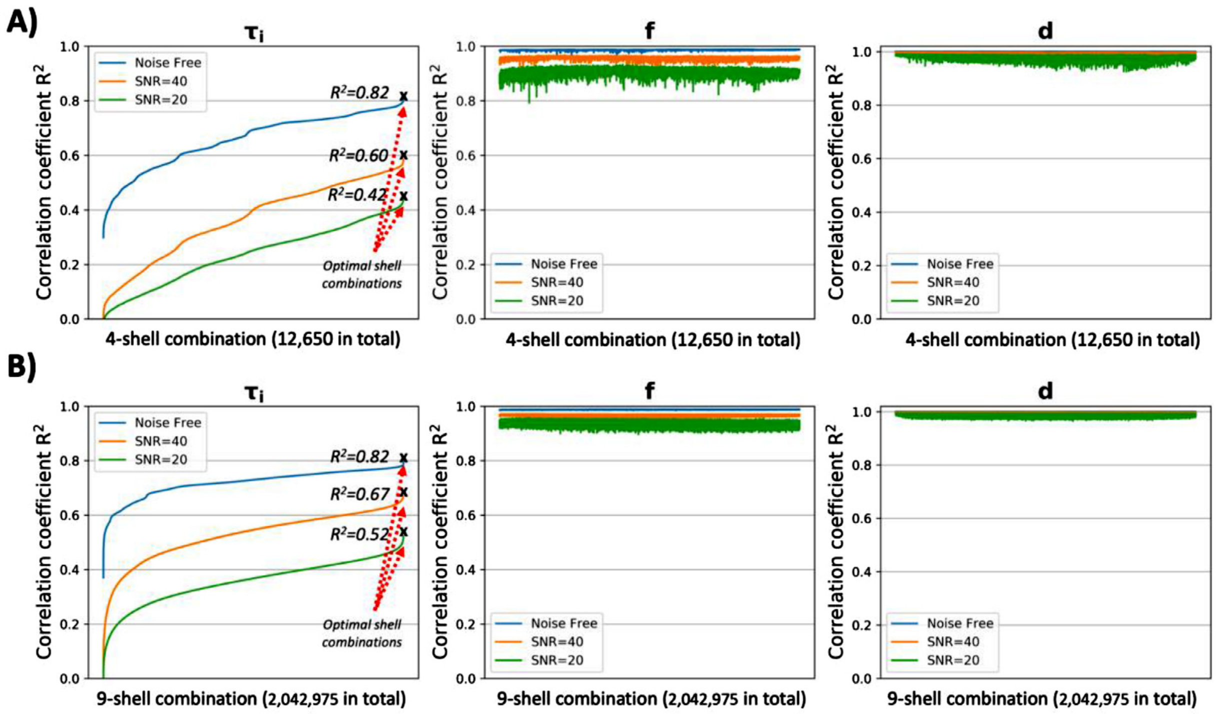

Fig. 4. Performance of the RF model prediction of $\tau_{i,}$ f and d, trained on different combinations of 4 **(A)** and 9 **(B)** shells. Each curve shows the R^2 score (y-axis) of the RF trained on a different combination of shells (x-axis). The shell combinations are sorted in increasing order according to their R^2 score. We show the results for three levels of noise: SNR=∞ (blue curve), SNR=40 (orange curve) and SNR=20 (green curve). The R^2 score for τ_i is calculated only for values ≤400 ms as this is the range over which we are sensitive to this parameter (see Section 3.1).

SNR=∞ to 0.94 for SNR=20. The estimation of the intrinsic diffusivity d is very robust to noise: the correlation coefficients remaining very high (0.99) even when training the model on the SNR=20 dataset. Furthermore, we find that all the top 100 combinations contain the two highest b-value shells (6,803 and 7,773 smm^{-2}) with the two longest Δs. Additionally, we find that high b-value shells only maximise the performance of the RF in combination with low b-value shells (775 and 930 smm^{-2}). For SNR=40 (orange curves), which we use when predicting on the in-vivo data, we find that the optimal combination of 9 shells sorted by b-value is [620, 775, 930, 1241, 1384, 2489, 4973, 6803, 7773] smm^{-2} with an R^2 score of 0.67, and the best combination of 4 shells is [775, 930, 6803, 7773] smm^{-2} with an R^2 score of 0.60. These results show that the optimal b-values for both 4 and 9 shells are a combination of low and high values, which sample both short and long **Δ**s. Similar results were also obtained for the "signals training database" (not shown). Since we are looking to optimise our framework for in-vivo estimation on the mouse data, we run the in-vivo experiments using the best 9-shell combination in the SNR=40 scenario, as the noise level which matches our in-vivo data.

3.3. Synthetic experiments

Fig. 5 shows the RF results obtained using the feature (top row) and the signal (bottom row) noise free databases. To assess the quality of our fit, we display the results using Bland-Altman plots and colour each data point according to how close the estimates are to the ground truth values. To aid visual interpretation, we cap the percentage error at 50%. The mean difference between the ground truth and the estimated values is shown by the black line and the 95% upper and lower limits of agreement by the dashed lines. For all three parameters of interest, we observe no overall estimation bias as the estimates are spread equally around the zero-difference black line. However, for τ_i, the parameter recovery is not perfect and the Bland-Altman plots show an overestimation bias for small values of τ_i and an underestimation bias for large values. The R^2 scores show a strong correlation between the estimates of our model and the ground truth parameter values: $R_{\tau i}{}^2$=0.82/0.84 (features/signals database), R$_\mathrm{f}{}^2$=0.99 (both databases), and $R_d{}^2$=0.99 (both databases). When assessing the model's performance with respect to the two training databases (features/signals), we observe no significant difference between the two approaches. The R^2 scores remain unchanged for f and d and show only a minor difference for τ_i: $R^2{}_\mathrm{features}$ = 0.82 / $R^2{}_\mathrm{signals}$ = 0.84. The advantages of each approach are discussed further in *Section 4*. The noise-free results in *Fig. 5* provide a benchmark performance of the model given our data and imaging protocol.

Fig. 6 shows the equivalent results for SNR=40. The presence of noise results in wider limits of agreement and affects differently the estimation of each parameter. The mean difference lines for all three parameters remain at zero, showing no general bias in the estimates. Intra-axonal volume fraction and diffusivity continue to be very well estimated and their correlation coefficients are only very mildly affected by the presence of noise: $R_\mathrm{f}{}^2$=0.97 and $R_d{}^2$=0.99, equal for both training databases. In contrast to this, the presence of noise has a stronger effect on the estimation of τ_i, resulting in a lower R^2 score and a more pronounced overestimation/underestimation bias for small and large values respectively. Despite this, we find that the RF works well within the sensitivity range computed in *Section 3.1*, with a very good correlation coefficient between the model's estimations and ground truth for τ_i≤400 ms (R^2=0.68). Outside this indicative sensitivity range, the correlation coefficient is very weak: R^2= 0.07 for τ_i≥400 ms. In line with the noise free case, we continue to see no significant difference between the signal and the feature approach: $R^2{}_\mathrm{features}$ = 0.67 / $R^2{}_\mathrm{signals}$ = 0.68.

3.4. In-vivo imaging experiments

To show that our in vivo data is well represented by our synthetic training database, we perform a data quality match (*Fig. 7*). We plot the signal intensity as a function of the angle between the diffusion gradients and the cylindrical fibres' axis θ (in degrees), for different diffusion

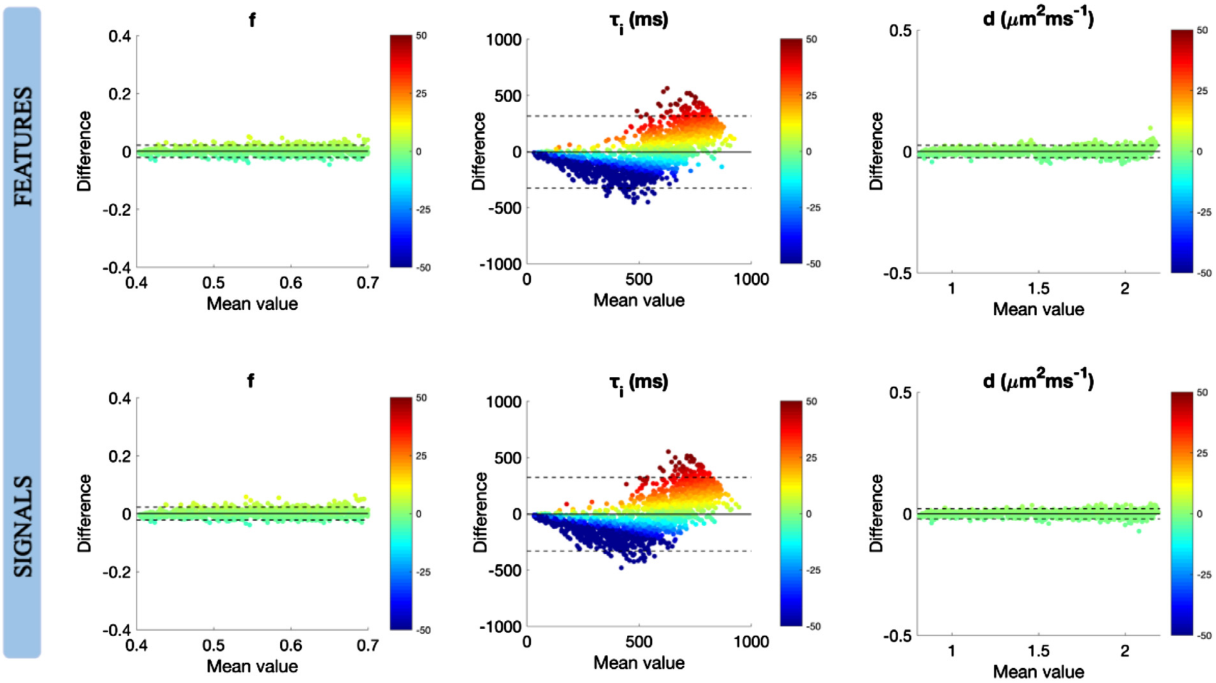

Fig. 5. Bland-Altman plots for the RF estimates of f, τ_i and d using the features (top row) and signals (bottom row) noise-free simulated database. To aid visual interpretation, the plots are color-coded with the percentage error capped at ±50%.

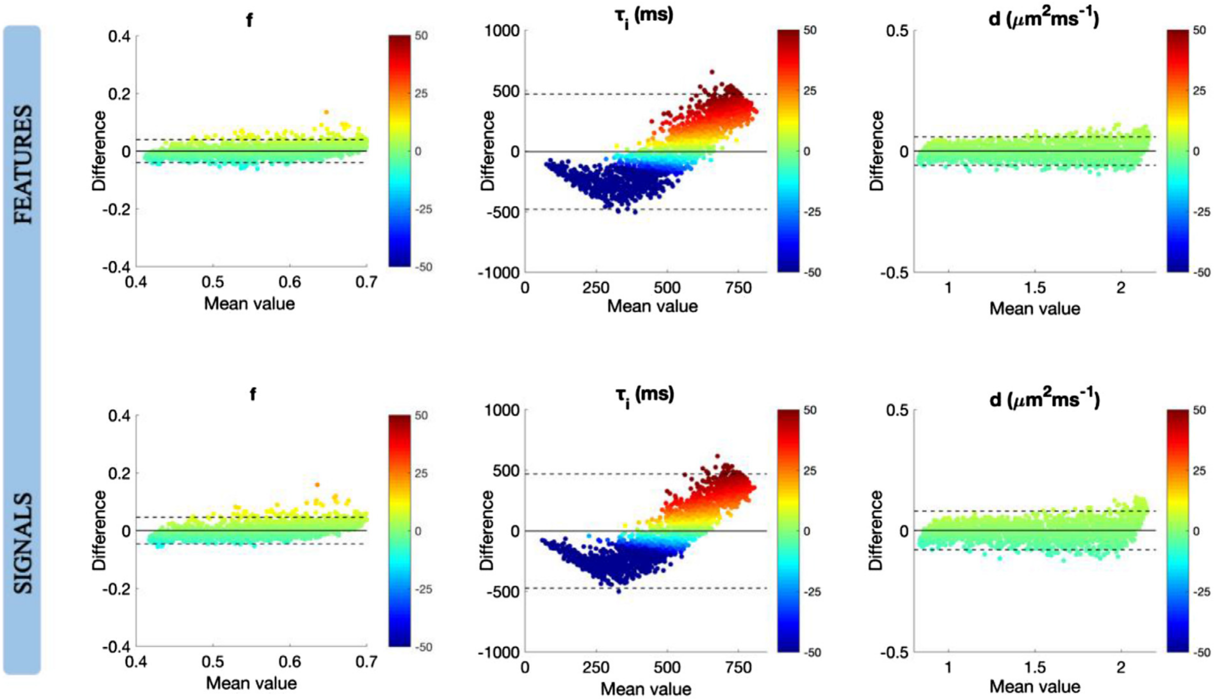

Fig. 6. Bland-Altman plots for the RF estimates of f, τ_i and d using the features (top row) and signals (bottom row) simulated database with SNR=40, matching the noise level in our in-vivo data. To aid visual interpretation, the plots are color-coded with the percentage error capped at ±50%

gradient strengths ($G_{1\text{-}5}$=150–500 mT/m) and for Δ={10.8, 20.0} ms. Fig. 7 provides a comparison between one of our simulated signals (at different gradient strengths and diffusion times) and the experimental signals measured from a voxel in the centre of the splenium of a WT mouse. We find a very good match between the simulated and in-vivo DW-MRI signals, demonstrating that our training data set is a good representation of the in-vivo mouse dataset. This is a necessary condition for our supervised learning approach to be valid and ensures that during the supervised learning we learn a training dataset which is similar to the test dataset. However, please note that similar DW-MRI signals do not necessarily imply similar underpinning microstructure. This very known ambiguity (Jelescu et al, 2016b, Novikov et al, 2019) is one of the main challenges in microstructure imaging, leading to higher uncertainty in the model parameter estimation.

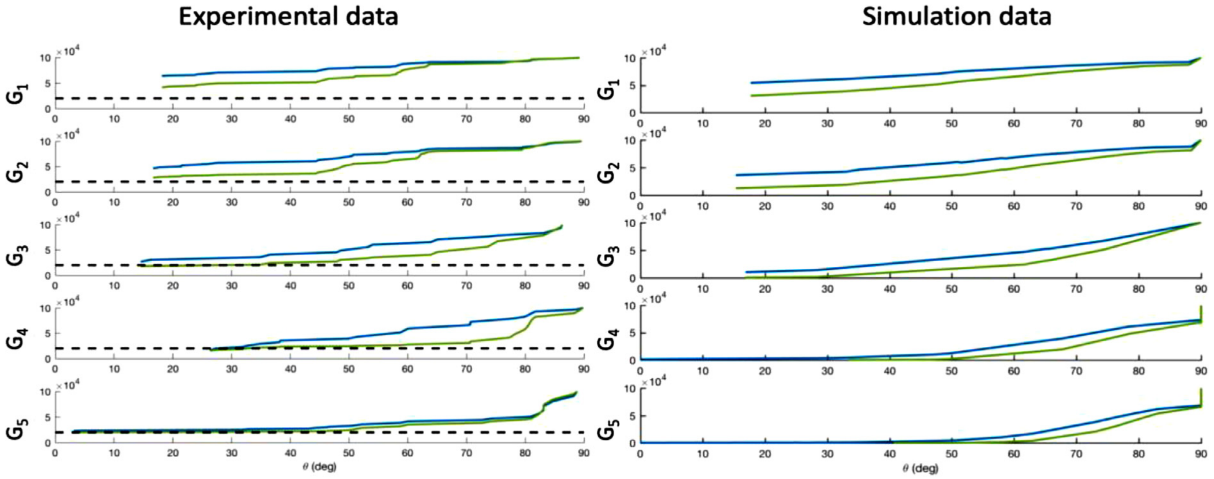

Fig. 7. Comparison between the in-vivo (left) and simulated (right) signal intensity as a function of the angle between diffusion gradients and the cylindrical fibres' axis θ (in degrees), for different diffusion gradient strengths ($G_{1\text{-}5}$=150-500mT/m) and two Δs: 10.8 ms (blue lines) and 20.0 ms (green lines). The dashed black line in the experimental data represents the noise floor level.

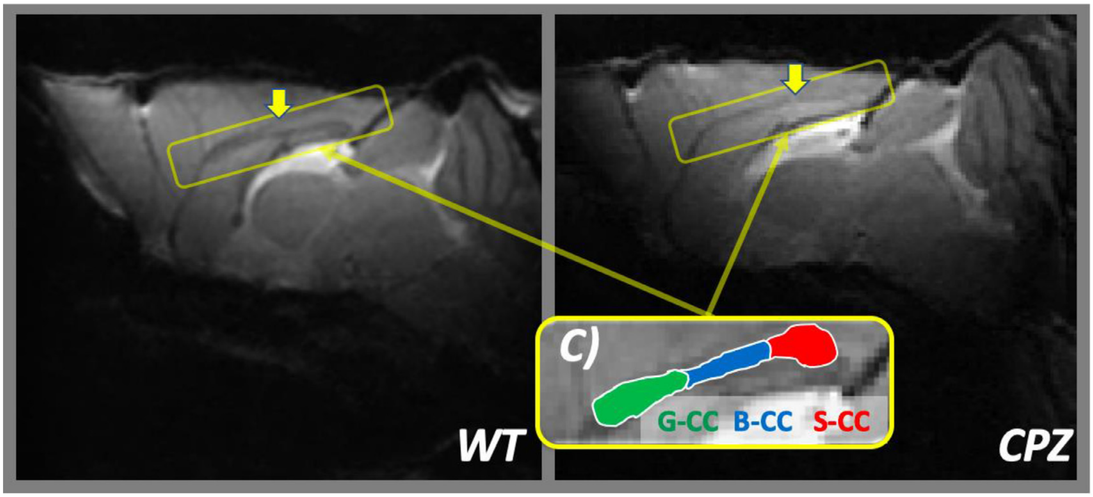

Fig. 8. Representative DW-MRI b=0 images of: **A)** a WT mouse scan in our cohort and **B)** a CPZ mouse scan in our cohort. **C)** ROIs of the CC overlaid on the zoomed in b=0 image of the WT mouse scan. The three ROIs are genu (G-CC), body (B-CC) and splenium (S-CC). The yellow square indicates the region in which the CC is found.

Fig. 8 shows examples of DW-MRI b=0 images for a WT (*Fig. 8A*) and for a CPZ (*Fig. 8B*) mouse. We can observe the appearance of the CC in the CPZ scan is different from the WT, showing the effect of demyelination. *Fig. 8C* shows the three ROIs of the CC overlaid on the b=0 image of the WT scan. We manually define three ROIs on the CC masks of each mouse scan: splenium (S-CC), body (B-CC) and genu (G-CC) by following the distribution of the RD values to help localize the central voxels of these three main regions: the genu and splenium of the corpus callosum show a lower RD than the body. We then calculate the mean parameter estimates for NODDI (ODI), DTI (AD, RD, FA) and RF (f, τ_i, d) in each ROI for every mouse, and study the differences between the WT and the CPZ groups. We present these results in the remainder of this section.

Fig. 9 shows CC maps for NODDI and DTI parameters for one exemplar healthy WT mouse (first column) and one exemplar CPZ mouse (second column). A visual inspection of the CC maps reveals no significant changes in ODI and AD between the two mice, together with a significant increase in RD and decrease in FA.

We observe the same trends in the DTI and NODDI parameters at group level, as shown in *Fig. 9B*. We illustrate the difference between the WT group and the CPZ group through box and whisker plots in the three ROIs of the CC: genu (G-CC), body (B-CC) and splenium (S-CC). We find the estimates of ODI in the two groups to be between 0.15 and 0.29, suggesting very low dispersion, in line with recently reported values in literature (Wang et al., 2019). Furthermore, we find no statistically significant difference in NODDI ODI between the two groups in the three regions of the CC, a finding that is also in line with Wang et al. (2019). The DTI estimates show negligible changes in AD, a significant increase in RD and a significant decrease in FA. These results are consistent with already published results (Boretius et al., 2012, Song et al., 2005, Zhang et al., 2012b).

The in-vivo RF estimates of *f*, τ_i and *d* obtained using the raw signal database are presented in *Fig. 10*.

The parametric CC maps shown in *Fig. 10A* correspond to the same WT mouse (first column) and CPZ mouse (second column) in *Fig. 9A*. The CC maps show a statistically significant decrease in *f* (first row) and τ_i (second row), and no significant change in *d* (third row). To provide a more quantitative analysis, we plot the box and whisker plots of region-specific parameter comparisons between the WT and the CPZ group over the three CC ROIs (*Fig. 10B)*. The trends observed visually in *Fig. 10A*

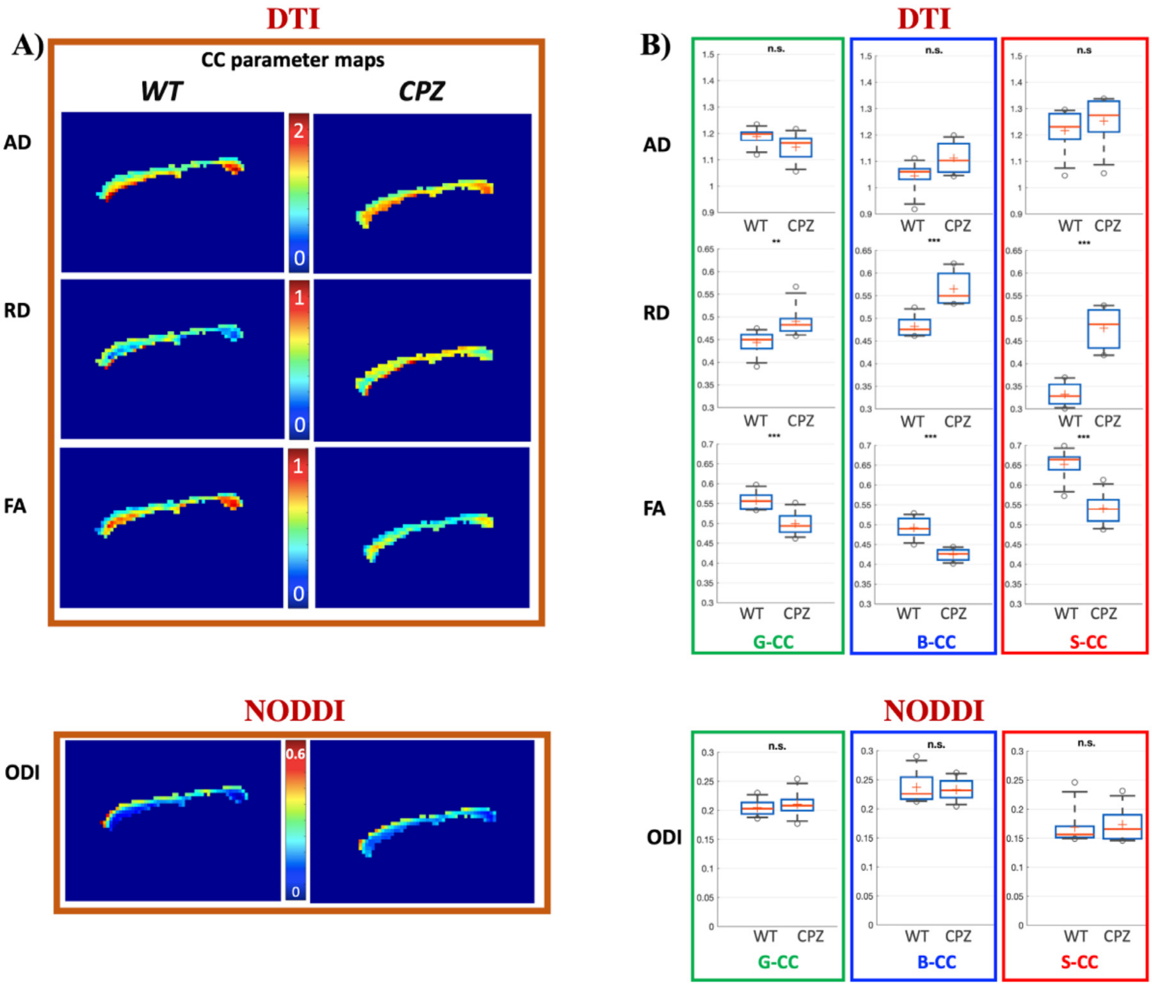

Fig. 9. **A)** Parametric maps of the CC in a healthy WT mouse (first column) and a CPZ mouse (second column) obtained from conventional DTI at b=1241 s/mm^2 and from NODDI ODI. **B)** Box and whisker plots of region-specific comparison between WT (N=8) and CPZ (N=8). DTI metrics (AD, RD, FA) are evaluated within the genu (G-CC), body (B-CC) and splenium (S-CC) of the CC. Statistical significance is assessed by using a 2-tailed t-test with equal variance and significance level: *=0.01, **=0.005, ***=0.001. 'n.s.' stands for non-significant.

Table 2
Mean and standard deviation of RF estimates for f, τ_i and d in the three CC ROIs for the WT and CPZ group. CPZ regions that are statistically different from WT regions are marked with * for p<0.01, ** for p<0.005 and *** for p<0.001.

	f		τ_i		*d*	
	WT	***CPZ***	***WT***	***CPZ***	***WT***	***CPZ***
S-CC	0.443 (0.005)	0.428(0.003)***	370 (7)	310 (15) ***	1.12 (0.07)	1.18 (0.07)
B-CC	0.430 (0.002)	0.424(0.001)***	370 (9)	330 (10) ***	1.10 (0.05)	1.15 (0.03)
G-CC	0.440 (0.006)	0.429(0.003)**	380 (14)	350 (12)**	1.15 (0.02)	1.11 (0.04)

hold for the group-wise quantitative comparison (WT versus CPZ): we observe statistically significant decreases in f and τ_i and a negligible and statistically insignificant increase in *d*. These trends are consistent across all three regions of the CC. The mean and standard deviations of the RF parameter estimates for each ROI are reported in Table 2.

3.5. Correlation with post-mortem analysis

The histological EM measurements in the splenium, body and genu of the CC over the cohort of WT (blue) and CPZ (black) mice are reported in the histograms of Fig. 11. Our histological data shows no axonal size changes (Fig. 11C) and no significant axonal loss (data not shown here) between the two cohorts. The axonal diameter measurements in Fig. 11C do not take into account the commonly accepted shrinkage factor of 30% (Barazany et al., 2009, Innocenti et al., 2015), after which the differences between the two groups continue to remain statistically non-significant. We also find a statistically significant decrease in myelin thickness (Fig. 11B) correlated with an increase in the g$_{\text{ratio}}$ (Fig. 11A) and a decrease in the intra-axonal volume fraction (Fig. 11D). Finally, we measure a weak but not statistically significant correlation between axonal diameter and intra-axonal volume fraction from the EM analysis (ϱ = 0.34 and p = 0.51 > 0.05).

Next, we study the correlation between these changes and the estimates of the RF model in Fig. 12. We assess the statistical significance of the linear correlation between τ_i and myelin thickness from EM with a two-tailed t-test by looking at the mean and the standard deviation

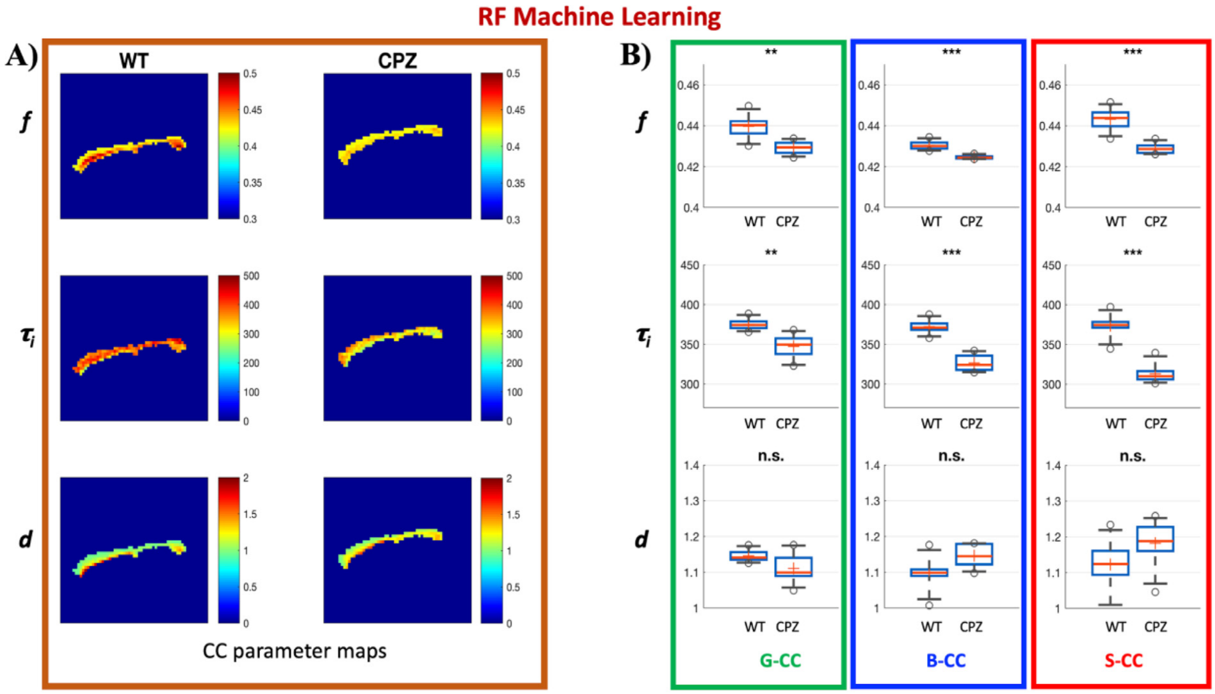

Fig. 10. **A)** Parametric maps with the RF estimates for f, τ_i and d in the CC of a healthy WT mouse (first column) and a CPZ mouse (second column). **B)** Box and whisker plots of region-specific comparison between WT (N=8) and CPZ (N=8). RF estimates for f, τ_i and d are computed independently for all voxels within the genu (G-CC), body (B-CC) and splenium (S-CC) of the CC. Statistical significance was assessed by using a 2-tailed t-test with equal variance and significance level: *=0.01, **=0.005, ***=0.001. 'n.s.' stands for non-significant. The difference in the morphology of the CC between the WT and the CPZ mice is mostly due to different masking, subject to different partial volume within the CSF of each mouse.

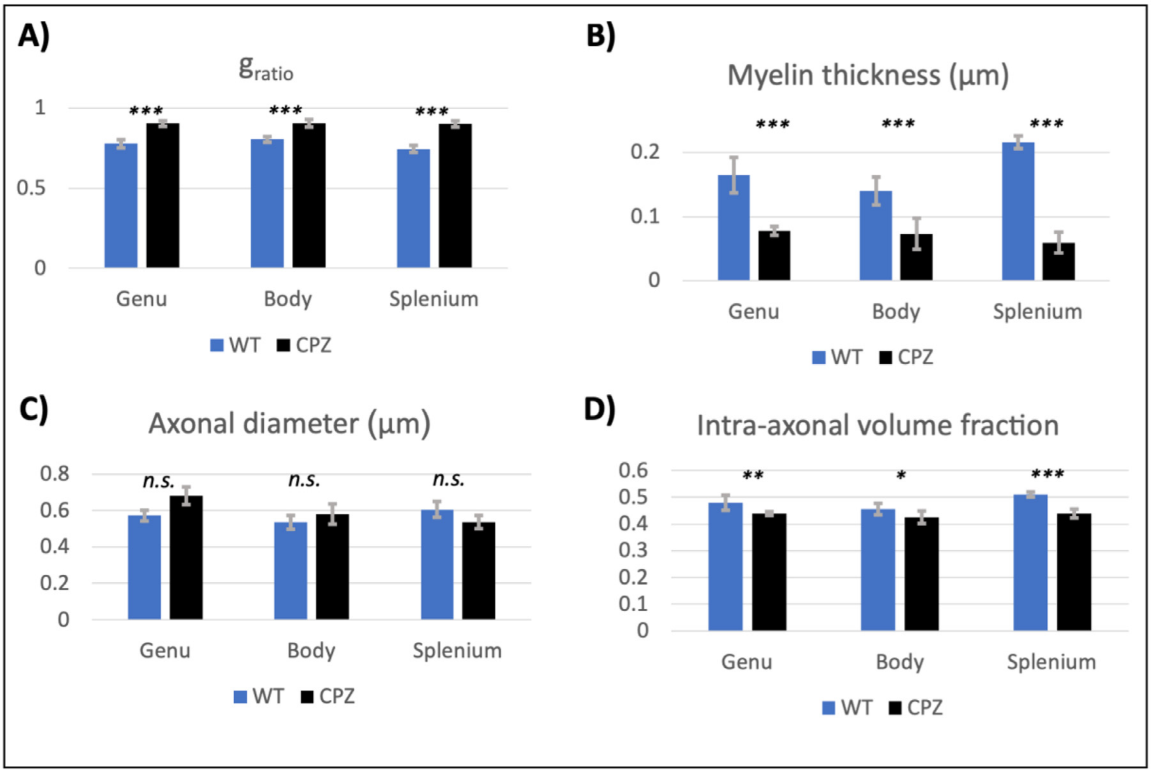

Fig. 11. ***Histology results.*** *The mean and the standard deviation of the EM measurements in the splenium, body and genu of the CC for the cohort of WT (blue) and CPZ (black) mice: the* g_{ratio} ***(A)****, myelin thickness* ***(B)****, mean axonal diameter* ***(C)*** *and intra-axonal volume fraction* ***(D)****.*

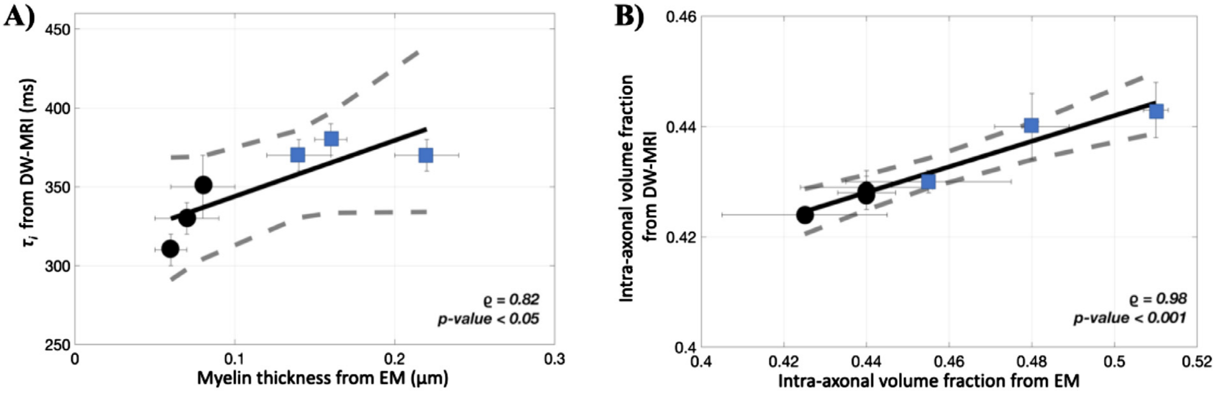

Fig. 12. Statistical significance and correlations between: **A)** the exchange time from DW-MRI (y-axis) and myelin thickness from EM (x-axis) and **B)** the intra-axonal volume fraction from DW-MRI (y-axis) and EM (x-axis). Each point represents the mean over one region of the CC for the WT (blue squares) and CPZ (black circles) group. Error bars indicate the standard deviation over the region.

of each CC ROI of the WT (blue squares) and CPZ (black circles) group (*Fig. 12A)*. We find a Pearson linear correlation coefficient ϱ of 0.82 and a p-value < 0.05 for τ_i, showing a good correlation between the RF estimates of the exchange time from DW-MRI *(y-axis)* and histological measurements of myelin thickness *(x-axis)*.

Similarly, we investigate the statistical significance of the linear correlation between intra-axonal volume fraction f as estimated from DW-MRI *(y-axis)* and from EM *(x-axis)* *(Fig. 12B)*. We find a Pearson correlation coefficient ϱ of 0.98 and a p-value<0.001, showing a strong correlation between the RF estimates and the histological measurements of the intra-axonal volume fraction. Note that the lower ϱ value for the analysis in *Fig. 12A* is likely due to the sensitivity limit of the current experimental protocol to changes in τ_i. Moreover, the fact that EM measurements of intra-axonal volume fraction are consistently higher than the RF estimation from in vivo DW-MRI may be due to unaccounted shrinkage effects, which affect mostly the extracellular space and thus can lead to an increase in the intra-axonal volume fraction.

Discussion

In this work, we focus on the experimental study of a RF based computational model for axonal permeability estimation using an in-vivo cuprizone mouse model of demyelination. Because no analytical model is available for permeability characterisation in the general case of neither very fast or very slow exchange, here we use the computational approach proposed in Nedjati et al. (2017) . Specifically, we use Monte Carlo simulations of the DW-MRI signal and train our model to estimate microstructure parameters with a focus on the intra-axonal water exchange time τ_i, a parameter inversely related to axonal permeability. Using synthetic substrates mimicking our in-vivo data, we show that our imaging protocol has good sensitivity to exchange times $\tau_i \leq$ 400 ms for large axons (mean diameter of 2 μm) and to $\tau_i \leq$ 250 ms for small axons (mean diameter of 0.4 μm) under the noise conditions of our in-vivo data (SNR=40). Following from this, we find that the RF model we developed works very well in this range: we find a good correlation between RF estimates and the ground truth for $\tau_i \leq$ 400 ms (R^2=0.87 for SNR=inf and R^2=0.68 for SNR=40), and a weak correlation for $\tau_i >$ 400 ms (R^2=0.3 for SNR=inf and R^2=0.07 for SNR=40) due to the low sensitivity in our protocol for values above 400 ms. In our in-vivo imaging experiments, we find that the RF estimates of τ_i are within the sensitivity range and in line with literature values of the exchange time reported in healthy rat brain tissue (Prantner, 2008, Quirk et al., 2003). Furthermore, we find that the RF estimates of τ_i in the CPZ group are significantly lower than in the WT group, a finding that one would intuitively expect to see in a model of demyelination. Furthermore, we find that our intra-axonal volume fraction estimates in CPZ mice are also significantly lower than in controls. These results are in strong agreement ($\varrho_{\tau i}$ = 0.82, ϱ_f = 0.98) with our EM histology results of myelin thickness and intra-axonal volume fraction, respectively. Finally, we show that potentially confounding factors such as axonal swelling and dispersion have a negligible effect on the estimated differences between the WT and CPZ group. These results suggest for the first time, quantitatively and in-vivo, that machine learning based computational models could act as a suitable biomarker to detect and track changes in demyelinating pathologies. Furthermore, they support the application of τ_i as more sensitive and specific marker of demyelination.

4.1. Simulations

Sensitivity analysis. Our sensitivity analysis shows that our imaging protocol has good sensitivity for exchange times in the range $\tau_i \in$ [0, 400] ms for substrates with large axons (mean diameter μ_D=2 μm) and in the range $\tau_i \in$ [0, 250] ms for substrates with small axons (mean diameter μ_D=0.4 μm), under noise conditions matching that of our in-vivo data (SNR=40). Generally speaking, the noise in the data affects the sensitivity differently, depending on the mean axon diameter in the substrate. For substrates with large axons (μ_D=2 μm), the sensitivity halves from $\tau_i \in$ [0, 400] ms for SNR=40 to $\tau_i \in$ [0, 200] ms for SNR=20. For substrates with smaller axons (μ_D=0.4 μm), decreasing the SNR from 40 to 20 has a smaller effect on the sensitivity range, reducing it by 44% from $\tau_i \in$ [0, 250] ms (SNR=40) to $\tau_i \in$ [0, 140] ms (SNR=20). Furthermore, we find that the larger the axons in the substrate, the better the sensitivity range. Substrates with μ_D=2 μm have a sensitivity range wider by 60% (for SNR=40) and by 43% (for SNR=20) than substrates with μ_D=0.4 μm.

Shell selection. To optimise the performance of the machine learning model, we explore the wide range of parameters in our PGSE protocol and select the best combination of 4 and 9 shells. We show that for our in-vivo data with SNR=40 the number of shells that maximises the performance of the model is 9, with the b-values [620, 775, 930, 1241, 1384, 2489, 4973, 6803, 7773] smm^{-2} and an R^2 score of 0.67. When analysing the best combinations of 4 and 9 shells, we observe that they sample every value of $\mathbf{\Delta}$ in our sequence, resulting in a combination of low and high b-value shells. This finding is in accordance with the optimised STEAM protocol in Nedjati et al. (2017), which contains two long $\mathbf{\Delta}$ and two short $\mathbf{\Delta}$ shells. This suggests that to maximise sensitivity to the intra-axonal exchange time, it is necessary to include a combination of short and long $\mathbf{\Delta}$**s**.

We show that noise is an important factor for the performance of our model. We find that in the noise free case, it is sufficient to use only 4 shells as introducing more shells does not improve performance. However, in the presence of noise, we find that increasing the number of shells from 4 to 9 improves the R^2 score between the estimated and the ground truth τ_i. A potential explanation for this is that the addition of noise corrupts the information in each shell, and having more shells to corroborate information from helps the RF model learn better. Our analysis also reveals that increasing the number of shells above 9 does not offer any additional benefits even in the presence of noise. Moreover, we show that noise has a stronger effect on the estimation of τ_i, for which the R^2 score drops from 0.84 in the noise free case to ≈0.5 for SNR=20. The estimation of *f* and *d* is considerably more robust: $R^2_{noise\text{-}free}$=0.99 versus $R^2_{SNR=20}$ =0.94 for *f* and no drop for *d*. This suggests that SNR plays an important role in a protocol's suitability for permeability estimation using our approach.

Feature extraction. When extracting the rotationally invariant features from our synthetic signals, we obtain meaningful values for all b shells in the synthetic data. When we apply the same method to in-vivo data, the feature extraction becomes difficult and does not give meaningful results for b shells with high gradient strength (above 300 mT/m) and high b-values. We believe that this difference is most likely due to the effect of fibre dispersion, present in the in-vivo data but not included in our simulations. As the gradient strength increases, the dispersed fibres would cause larger drops in the signal, as can also be seen in (*Fig. 7*), where we notice that the drop in the signal intensity relative to the gradient direction is less prominent in the synthetic signals than in the in-vivo data. Moreover, we note that theoretically the diffusion tensor features at b values higher than 2000 s/mm^2 lose their physical meaning. However, here we do not interpret the diffusion tensor features at high b values in terms of tissue microstructure, but we rather use them just as convenient metrics to represent the signal. Note that we include all of the diffusion tensor features even if some of them are not mutually independent. We prefer to work with a comprehensive set of features to ensure that our machine learning algorithm finds the most informative split criteria.

Synthetic data experiments. The RF model estimates in the noise free case have very strong correlations with the ground truth values, providing an excellent benchmark performance for our model and imaging protocol (*f*: R^2=0.99, τ_i: R^2=0.84 *d:* R^2=0.99). We show that the addition of noise with SNR=40, matching our in-vivo data, does not affect much the estimation of *f* and *d* (*f*: R^2=0.97, *d:* R^2=0.99), however, it has a stronger effect on the estimation of τ_i. In line with our sensitivity results, for τ_i<400 ms the effect is present, however, the performance is still sufficiently good (R^2=0.68), while for τ_i>400 ms the performance of the model is severely affected (R^2=0.07). The estimation of *f* and *d* is considerably more robust than that of τ_i due to the use of a range of gradient strengths from 50 to 300 mT/m, which has been shown to improve the sensitivity to *f* and *d* (Huang et al, 2015, Sepehrband et al 2016). Moreover, the robust estimation of *f* and *d* is in agreement with what has been shown by Fieremans et al. (2011) about the estimation of *f* and *d* in the case of parallel fibres. Indeed, the case of parallel fibers is solvable analytically using only four estimated parameters: diffusivity and kurtosis in the directions parallel and perpendicular to the fibres. Since all the information for computing these parameters is present in the data, this explains the high fidelity of the prediction. However, we note that Fieremans et al.'s model is confounded by the fiber orientation dispersion, which is known to be present in white matter (Ronen et al., 2014) and therefore, in the case of non-negligible fibre dispersion, our parameter estimates may be biased.

In addition to this, we compare for the first time the signal and feature training approaches and show that there is no significant difference in the RF performance according to which database is used for training. This is a significant result as it shows that when extracting the rotationally invariant features from the raw signals we do not lose information that is essential for training our model. Consequently, we can use the features database without affecting the performance of our model. The advantage of a rotationally invariant feature approach is that it does not require the generation of a new library for every new acquisition protocol as long as the b-values and the TE of the protocols match. Nevertheless, as discussed above, caution should be applied with this approach when the acquisition protocol uses high gradient strengths (G≥300 mT/m) and the SNR is low, such as conditions often found in the pre-clinical setting, and then using signals database might be the preferable choice. On the other hand, in the clinical setting, imaging protocols have much lower gradient strengths and sufficient SNR to fit the DTI and SH model parameters in the feature extraction approach, and consequently, we expect the rotationally invariant feature approach to be a better choice (as used in Nedjati et al. (2017)). Irrespective of the training approach, we expect our model's performance to be similar in both the clinical and preclinical setting.

4.2. In-vivo mouse data and correlation with post-mortem analysis

Our data quality match shows that our synthetic training data is a good representation of the in-vivo data. Our DTI results show an increase in RD and a decrease in FA between the two groups. This could be explained by the breakdown of the myelin layer which allows water to diffuse more in the radial direction, leaving AD unchanged and having the overall effect of reducing FA. These changes in DTI metrics are in agreement with those reported in several studies of the CPZ mouse model of demyelination (Boretius et al., 2012, Song et al., 2005, Zhang et al., 2012b). Nevertheless, the DTI metrics are not specific because they provide only indirect measures of the underlying microstructural changes in the CPZ model. For instance, the observed increase in RD may be due to the increase spaces between the axons and not to the higher permeability of less myelinated axons.

On the other hand, our RF estimates of τ_i provide a more direct and specific measure of permeability. In fact, in our computational model, diffusivity (via d) and permeability (via τ_i) are decoupled and individually estimated from the data. We find that our estimations of τ_i in the healthy mice compare well with literature values. Studies on sphingomyelin membranes found in axonal membranes suggest values between 300 ms and 600 ms for axons with radii between 0.5 and 1 μm (Finkelstein, 1976). Contrast agent and relaxometry studies in the rat brain estimate the intracellular water exchange lifetime in the rat brain to be between 200 ms (Prantner, 2008) and 550 ms (Quirk et al., 2003). It is worthwhile to note that our experimental protocol does not provide enough sensitivity to detect exchange times > 0.4 s. Hence, our method would estimate τ_i ~ 0.4 s for any actual exchange time ≥ 0.4 s. Nevertheless, we have high sensitivity to reliably measure any changes in τ_i occurring below 0.4 s due to demyelination. As accurate histology measurements of τ_i are not available due to tissue fixation altering the membrane permeability, we compare our estimates of τ_i with EM measurements of myelin thickness. We compute myelin thickness from myelinated axons only, and it includes both the effect of demyelination induced by CPZ and some remyelination that happens spontaneously in the CPZ model (Matsushima and Morell, 2001). We find a strong correlation between the RF estimates of τ_i and myelin thickness ($\varrho_{\tau i}$ = 0.82). This is in very good agreement with a recently published simulation work investigating the link between exchange time and myelin thickness (Brusini et al., 2019), further supporting the findings that myelin wrapping can meaningfully contribute to the signal in DW-MRI and impact τ_i. Furthermore, our RF estimates of *d* lie in the range 1–1.3 μm^{2}sm^{-1}, an expected range for the mouse CC (Wu et al., 2008), and our estimates of *f* correlate very strongly with the EM intra-axonal volume fraction measurements (ϱ_f = 0.98).

When comparing the two groups, we observe the following general trends: a statistically significant decrease in the intra-axonal volume fraction *f* and in the intra-axonal exchange time τ_i, together with a negligible and statistically insignificant increase in the intrinsic diffusivity *d.* We expect *f* to be lower in the CPZ group as there is an increase in the extracellular space due to the breakdown of myelin. Demyelination is also thought to cause a decrease in the intra-axonal exchange time as the water molecules encounter less barriers when moving from the intracellular to the extracellular space. In line with this, the RF estimations of τ_i in the CPZ group are significantly lower than in the WT group. However, it is worth noting that the actual values of the estimated parameters may be biased by the presence of fibre dispersion, undulation or beading (Budde and Frank, 2010, Nilsson et al., 2012, Palombo et al., 2018), which we do not account for in our simulations. Nevertheless, we expect that these factors have negligible effects on our differential analysis of control and cuprizone groups, as we explain in more details in the next paragraph.

The strong correlation between myelin thickness and the estimated τ_i suggests that demyelination could be one of the main factors behind our measured decrease in τ_i. To strengthen this hypothesis, we analyse the potential confounding effect of other underlying processes. Our AD measurements from the DTI fit suggest that, if undulation or beading effects are present, they have a negligible effect (Budde and Frank, 2010, Nilsson et al., 2012, Palombo et al., 2018). Furthermore, we analyse the effect fibre dispersion can have on our results. Previous work investigating the impact of dispersion on axonal permeability estimation shows that the presence of orientation dispersion could result in underestimated values of τ_i (Nilsson et al., 2013a), and, therefore, could affect our estimates. In order to investigate this, we estimate NODDI ODI, and find, as shown in our results, that dispersion is very low (around 0.2), and, hence, the effect on the estimates of τ_i will also be low. In addition to this, we find no significant difference in the ODI between the WT and CPZ groups (*Fig. 9*). These findings are in line with recently published work by Wang et al. (2019) and, together with the non-significant change in DTI AD, suggest that we can rule out the effect of dispersion on our estimation of the difference in τ_i between the two groups. We also rule out the potential confounding effect of axonal swelling by looking at the statistically non-significant changes in axonal diameter as measured by EM. Finally, we note that, according to the findings of Jelescu et al. (2016a), we expect the effect of changes in extra-axonal transverse diffusivity for 6-week cuprizone intoxicated mice to be negligible at the diffusion times experimentally probed in this study, ruling out changes in extra-axonal transverse diffusivity as another potential confounding factor to our estimation of exchange time. This, together with the measured changes in RD, FA and the RF estimations of τ_i suggest that demyelination is the main process underpinning our DW-MRI contrast. In particular, our histological data strongly supports τ_i as a biomarker directly related to the thickness of the myelin sheath, which suffers degeneration in demyelinating diseases such as Multiple Sclerosis.

4.3. Limitations

Tissue model. One limitation of the present work is the simplicity of the white matter substrates used for the Monte Carlo simulations. Here we use the same computational model proposed by Nedjati et al. (2017) because the aim of this work is to validate that model and method. Other models, perhaps more complex (e.g. including dispersion, along-axon axonal size variance etc.), may improve the estimation of the exchange time, and will be explored in future work. Due to current limitations in our simulation system, we make several assumptions about the geometry of the tissue such as representing axons as non-abutting parallel cylinders. Such simplified sketch of the more complex white matter microstructure may lead to slight differences between the simulated DW-MRI signals and the measured ones, especially at very strong diffusion gradients (see for example case G4 and G5 in *Fig. 7*). Future work should aim to train the machine learning model on more realistic simulations, which account for different effects such as myelin water (Harkins and Does, 2016, Brusini et al., 2019), axonal undulation (Nilsson et al., 2012), dispersion (Ginsburger et al., 2019, Callaghan et al., 2019, Callaghan et al., 2020), neurons and glial cells (Palombo et al., 2018). Such effects, once included in the simulations, can easily be incorporated in the machine learning framework used in this paper. Using more complex and realistic simulations will also narrow down the gap between the synthetic and in-vivo data and increase the robustness of the parameter estimates. The gap between simulations and in-vivo data could also be addressed by using domain adaptation techniques, which can be integrated within other machine learning approaches such as neural networks.

Sensitivity of the imaging protocol. The sensitivity of our imaging protocol to the exchange time in the presence of noise, as we show in our simulation experiments, is not ideal. Although we perform some level of optimisation by choosing the most optimal shells in our large explorative protocol, the sensitivity might have been better if we optimised the protocol using our computational framework with respect to τ_i prior to imaging, as done in Nedjati et al.. Nevertheless, even with the protocol we use we can estimate values of $\tau_\mathrm{i} \leq 400$ ms, which is sufficient for the in-vivo mouse application used in this paper. The machine learning model here can also easily be adapted to incorporate more specialised diffusion encoding sequences such as OGSE for more sensitivity to axon diameter (Drobnjak et al., 2016; Kakkar et al., 2018) or STEAM for longer diffusion times (Fieremans et al., 2016). This limitation can be addressed by using the machine learning framework in this paper with refined Monte Carlo simulations and an imaging protocol optimised with respect to the intra-axonal exchange time.

RF model validation. Another limitation that stems from our simulation system is the testing of the model on the same type of data as it is trained on. Nevertheless, it is worth mentioning that despite using the same tissue model, we test our model using previously unseen parameter values. Future work should include the training and testing of the machine learning model using different types of substrates once these become available.

4.4. Implications for clinical applications

The extension of our approach to clinical systems can potentially be important for numerous white matter pathologies of the human nervous system. Here, we demonstrate the potential of our model in a cuprizone mouse model of demyelination, which is extensively used in the MS literature due to its close similarity to the demyelination and remyelination processes occurring in MS lesions Ransohoff (2012). The extension to clinical systems is challenging and most likely requires a reduced and optimized acquisition protocol and the use of STEAM based sequences to explore longer diffusion times. Nevertheless, even once reduced and optimised, we expect the acquisition to still take longer time than conventional DW imaging. However, as shown in Nedjati et al. (2017), that is feasible in approximately 30 min and has been done for both control subjects and MS patients. This suggests that our approach may be suitable for clinical and biomedical research applications.

The applicability of our approach extends to other myelin damaging pathologies such as spinal cord injury or leukodystrophies due to the hypothesised correlation between τ_i and the condition of the myelin sheath (Nilsson et al., 2013a, Ford and Hackney, 1997, Hwang et al., 2003). The current key limitation is the reduced sensitivity to the intra-axonal exchange time of clinically available imaging protocols. This can be addressed by using more specialised sequences such as the AXR sequence (Nilsson et al., 2013a) or optimised STEAM pulse sequences as in Nedjati et al. (2017), to which our framework can easily be adapted. With the continually increasing SNR in the clinical scanners, we expect the clinical applicability of this approach to also improve.

Our machine learning approach can be easily extended to a range of other intractable parameters such as undulation or extracellular space. Furthermore, our approach can be trained also on databases of experimental data, for which a precise correspondence between measured DW-

MRI signal features and ground-truth microstructure features is known. Another important further development for clinical practice is the introduction of uncertainty measures on the estimates of τ_i. Uncertainty could be included via a Bayesian approach as in Tanno et al. (2016) and it would help highlight areas of the brain where the estimates are less reliable due to unfamiliar signals. From a clinical perspective, these developments can have a great impact on the understanding and diagnosis of neurological conditions of the white matter in the near future.

CRediT authorship contribution statement

Ioana Hill: Formal analysis, Software, Investigation, Visualisation, Writing – Original Draft, Writing - Review & Editing. **Marco Palombo:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - Review & Editing, Visualization, Supervision, Project administration. **Mathieu Santin:** Methodology, Investigation, Data curation, Writing - Review & Editing. **Francesca Branzoli:** Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - Review & Editing. **Anne-Charlotte Philippe:** Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - Review & Editing. **Demian Wassermann:** Conceptualization, Methodology, Validation, Investigation, Data curation, Resources, Funding acquisition, Project administration, Writing - Review & Editing. **Marie-Stephane Aigrot:** Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - Review & Editing. **Bruno Stankoff:** Conceptualisation, Methodology, Resources, Writing – Review & Editing. **Anne Baron-Van Evercooren:** Conceptualisation, Methodology, Resources, Writing – Review & Editing. **Mehdi Felfli:** Validation, Formal analysis, Investigation, Data curation, Writing - Review & Editing. **Dominique Langui:** Validation, Formal analysis, Investigation, Data curation, Writing - Review & Editing. **Hui Zhang:** Writing – Review & Editing. **Stephane Lehericy:** Conceptualisation, Methodology, Resources, Writing – Review & Editing. **Alexandra Petiet:** Conceptualization, Methodology, Validation, Investigation, Data curation, Resources, Funding acquisition, Project administration, Writing - Review & Editing. **Daniel C. Alexander:** Writing – Review & Editing. **Olga Ciccarelli:** Writing – Review & Editing. **Ivana Drobnjak:** Methodology, Supervision, Project administration, Writing - original draft, Writing - Review & Editing.

Acknowledgements

This work was supported by EPSRC (EP/G007748, EP/I027084/01, EP/L022680/1, EP/M020533/1, EP/N018702-1), EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1), Spinal Research, UK MS Society, NIHR and NIHR UCLH Biomedical Research Centre (BRC). MP is supported by UKRI Future Leaders Fellowship MR/T020296/1.

The research leading to these results received funding from the programs ’Institut des neurosciences translationnelles’ ANR-10-IAIHU-06 and ’Infrastructure d’avenir en Biologie Santé’ ANR-11-INBS-0006 and ‘Translational research infrastructure for biotherapies in neuroscience’ ANR-11-INBS-0011. We thank the ICM Quant Cellular Imaging platform for their assistance in EM.

Data and Code availability statement

The data concerning healthy wild-type mice can be found at https://zenodo.org/record/996889#.WgH5E9vMx24. The authors do not have the permission to share the data used in this study for the cuprizone treated mice. All the code used for the analysis is available upon request to the corresponding authors.

Appendix

A.1. The table below presents information about the 15 DT and SH features extracted from each DW-MRI signal to construct the rotationally invariant database

Ftr No.	Feature/Model	Feature Information
1	λ_1 (DTI)	First eigen value of the diffusion tensor, representing the first main direction of diffusion, Obtained as the first output of the *dteig* command in Camino.
2	λ_2 (DTI)	Second eigenvalue of the diffusion tensor, representing the second main direction of diffusion. Obtained as the fifth output of the *dteig* command in Camino.
3	λ_3 (DTI)	Third eigenvalue of the diffusion tensor, representing the third main direction of diffusion. Obtained as the ninth output of the *dteig* command in Camino.
4	MD (DTI)	Mean diffusivity, an estimate of the overall diffusion in a voxel, computed as $\lambda_1+\lambda_2+\lambda_3$.
5	FA (DTI)	Fractional anisotropy, an estimate of the anisotropic diffusion in a voxel. It takes values between 0 and 1 and is computed as: $\frac{3}{2}\sqrt{\frac{\sum(\lambda_j-MD)^2}{\sum\lambda_j}}$
6	I_0 (SH)	A combination of the SH coefficients $a_{k,i}$ of order k=0 and i ndex i, calculated as $I_0=\sum_{i=-k}^{k}\lvert a_{k,i}\rvert^2$, where k=0.
7	I_2 (SH)	A combination of the SH coefficients $a_{k,i}$ of order k=2 and index i, calculated as $I_2=\sum_{i=-k}^{k}\lvert a_{k,i}\rvert^2$, where k=2.
8	I_4 (SH)	A combination of the SH coefficients $a_{k,i}$ of order k=4 and index i, calculated as $I_4=\sum_{i=-k}^{k}\lvert a_{k,i}\rvert^2$, where k=4.
9	mean ADC (SH)	This feature is computed by calculating the values of the spherical functions f of the voxel at a set of evenly distributed sample points on a unit sphere S and taking the mean of these. Obtained directly from the *sfpeaks* command in Camino.
10	peak ADC (SH)	The maximum value of the spherical functions f over the points of the unit sphere S (see feature 9 for more details). Obtained as the 10th output of the *sfpeaks* command in Camino.
11	λ_1 (SH)	The first eigenvalue of the Hessian matrix at the peak.

A.2. Camino and scikit-learn commands used to simulate the synthetic signals database to train and apply the RF machine learning algorithm

A. The datasynth Camino command was used to generate synthetic diffusion MRI data. For each synthetic DW-MRI signal, the command was run using a different combination of parameters, sampled uniformly at random over the ranges specified in Section 2.2.1. The intra-axonal exchange time was specified through the probability -p, described in detail in the same section. More details on the other parameters of the datasynth command can be found at: http://camino.cs.ucl.ac.uk/index.php?n=Man.Datasynth
datasynth -walkers 100000 -tmax 2000 -voxels 1 -increments 1 -substrate inflammation -numcylinders 100000 -separateruns -schemefile

```
PGSE.scheme -initial uniform -p param1 -latticesize param2 -gamma
param3 param4 -diffusivity param5 >simulated_signal.Bfloat
```

B. Python scikit-learn code

```
# Import toolboxes
from sklearn.model_selection import train_test_split
from sklearn.ensemble
from sklearn.ensemble
# Load parameters to fit
params_to_fit = scipy.io.loadmat('parameters_to_fit.mat')
# Load the database of synthetic signals
synth_signals = scipy.io.loadmat('synthetic_signals_database.mat')
# Initialise RF regressors. Details on parameters can be accessed at
https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.
RandomForestRegressor.html.
rf_reg = RandomForestRegressor(n_estimators=no_trees, max_depth=
tree_depth,max_features="sqrt",
                              random_state=rndm_seed)
# Divide database into train and test set
test_set_size = 0.13
sim_train, sim_test, params_sim_train, params_sim_test = train_test_
split(synth_signals, params_to_fit,
                                         test_size=test_set_size,
                                         random_state=rndm_seed)
# Train the RF
rf_reg.fit(sim_train,y_sim_train)
# Predict parameter values for the test set
estimated_params = rf_reg.predict(sim_test)
```

References

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

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., 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.

Alexander, D.C., Zikic, D., Ghosh, A., Tanno, R., Wottschel, V., Zhang, J., Kaden, E., Dyrby, T.B., Sotiropoulos, S.N., Zhang, H., 2017. Image quality transfer and applications in diffusion MRI. Neuroimage 152, 283–298.

Barazany, D., Basser, P.J., Assaf, Y., 2009. In vivo measurement of axon diameter distribution in the corpus callosum of rat brain. Brain 132, 1210–1220.

Boretius, S., Escher, A., Dallenga, T., Wrzos, C., Tammer, R., Brück, W., Nessler, S., Frahm, J., Stadelmann, C., 2012. Assessment of lesion pathology in a new animal model of MS by multiparametric MRI and DTI. Neuroimage 59, 2678–2688.

Breiman, L., 2001. Random forests. Machine Learn. 45, 5–32.

Brusini, L., Menegaz, G., Nilsson, M., 2019. Monte Carlo simulations of water exchange through myelin wraps: implications for diffusion MRI. IEEE Trans. Med. Imaging.

Budde, M.D., Frank, J.A., 2010. Neurite beading is sufficient to decrease the apparent diffusion coefficient after ischemic stroke. Proc. Natl. Acad. Sci. 107, 14472–14477.

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

Callaghan, P.T., 1997. A simple matrix formalism for spin echo analysis of restricted diffusion under generalized gradient waveforms. J. Magn. Reson. 129, 74–84.

Callaghan, R., Alexander, D.C., Zhang, H., Palombo, M., 2019. Contextual fibre growth to generate realistic axonal packing for diffusion mri simulation. Information Processing in Medical Imaging: IPMI 2019. Lecture Notes in Computer Science 11492, 429–440.

Callaghan, R., Alexander, D.C., Palombo, M., Zhang, H., 2020. ConFiG: Contextual Fibre Growth to generate realistic axonal packing for diffusion MRI simulation. Neuroimage 220, 117107. doi:10.1016/j.neuroimage.2020.117107.

Codd, S.L., Callaghan, P.T., 1999. Spin echo analysis of restricted diffusion under generalized gradient waveforms: planar, cylindrical, and spherical pores with wall relaxivity. J. Magn. Reson. 137, 358–372.

Cook, P., Bai, Y., Nedjati-Gilani, S., Seunarine, K., Hall, M., Parker, G., Alexander, D.C., 2006. Camino: open-source diffusion-MRI reconstruction and processing. In: 14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine, Seattle WA, USA, p. 2759.

Criminisi, A., Shotton, J., Konukoglu, E., 2011. Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning [internet]. Microsoft Res..

Dhital, B., Reisert, M., Kellner, E., Kiselev, V., 2019. Intra-axonal diffusivity in brain white matter. Neuroimage 189, 543–550.

Drobnjak, I., Zhang, H., Ianuş, A., Kaden, E., Alexander, D.C., 2016. PGSE, OGSE, and sensitivity to axon diameter in diffusion MRI: insight from a simulation study. Magn. Reson. Med. 75, 688–700.

Einstein, A., 1905. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. Ann. Phys. 17, 549–560.

Fieremans, E., Burcaw, L.M., Lee, H.-H., Lemberskiy, G., Veraart, J., Novikov, D.S., 2016. In vivo observation and biophysical interpretation of time-dependent diffusion in human white matter. Neuroimage 129, 414–427.

Fieremans, E., Novikov, D.S., Jensen, J.H., Helpern, J.A., 2010. Monte Carlo study of a two-compartment exchange model of diffusion. NMR Biomed. 23, 711–724.

Fieremans, E., Jensen, J.H., Helpern, J.A., 2011. White matter characterization with diffusional kurtosis imaging. Neuroimage 58 (1), 177–188.

Fieremans, E., Lee, HH., 2018. Physical and numerical phantoms for the validation of brain microstructural MRI: a cookbook. Neuroimage 182, 39–61.

Filipiak, P., Fick, R., Petiet, A., Santin, M., Philippe, A.C., Lehericy, S., Ciuciu, P., Deriche, R., Wassermann, D., 2019. Reducing the number of samples in spatiotemporal dMRI acquisition design. Magn. Reson. Med. 81, 3218–3233.

Finkelstein, A., 1976. Water and nonelectrolyte permeability of lipid bilayer membranes. J. Gen. Physiol. 68, 127–135.

Ford, J.C., Hackney, D.B., 1997. Numerical model for calculation of apparent diffusion coefficients (ADC) in permeable cylinders—comparison with measured ADC in spinal cord white matter. Magn. Reson. Med. 37, 387–394.

Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N., 2011. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. Neuroimage 57, 378–390.

Ginsburger, K., Matuschke, F., Poupon, F., Mangin, J.-F., Axer, M., Poupon, C., 2019. MEDUSA: a GPU-based tool to create realistic phantoms of the brain microstructure using tiny spheres. Neuroimage.

Grebenkov, D.S., Van Nguyen, D., Li, J.-R., 2014. Exploring diffusion across permeable barriers at high gradients. I. Narrow pulse approximation. J. Magn. Reson. 248, 153–163.

Hall, M.G., Alexander, D.C., 2009. Convergence and parameter choice for Monte-Carlo simulations of diffusion MRI. IEEE Trans. Med. Imaging 28, 1354–1364.

Harkins, K.D., Does, M.D., 2016. Simulations on the influence of myelin water in diffusion-weighted imaging. Phys. Med. Biol. 61, 4729.

Hu, J., Verkman, A., Hu, J., Verkman, A., 2006. Increased migration and metastatic potential of tumor cells expressing aquaporin water channels. FASEB J. 20, 1892–1894.

Huang, S.Y., Nummenmaa, A., Witzel, T., Duval, T., Cohen-Adad, J., Wald, L., Mcnab, J.A., 2015. The impact of gradient strength on in vivo diffusion MRI estimates of axon diameter. Neuroimage 106, 464–472.

Hwang, S.N., Chin, C.L., Wehrli, F.W., Hackney, D.B., 2003. An image-based finite difference model for simulating restricted diffusion. Magnet. Reson. Med. 50, 373–382.

Innocenti, G.M., Caminiti, R., Aboitiz, F., 2015. Comments on the paper by Horowitz et al. (2014). Brain Struct. Func. 220, 1789–1790.

Jelescu, I.O., Zurel, M., Winters, K.V., Veraart, J., Rajaratnam, A., Kim, N.S., Babb, J.S., Sheperd, T.M., Novikov, D.S., Kim, S.G., Fieremans, E., 2016a. In vivo quantification of demyelination and recovery using compartment-specific diffusion MRI metrics validated by electron microscopy. Neuroimage 132, 104–114.

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

Kakkar, L.S., Bennett, O.F., Siow, B., Richardson, S, Ianus, A., Quick, T., Atkinson, D., Phillips, J.B., Drobnjak, I., 2018. Low frequency oscillating gradient spin-echo sequences improve sensitivity to axon diameter: an experimental study in viable nerve tissue. Neuroimage 182, 314–328.

Kärger, J., Pfeifer, H., Heink, W., 1988. Principles and application of self-diffusion measurements by nuclear magnetic resonance. Advances in Magnetic and Optical Resonance. Elsevier.

Lasič, S., Nilsson, M., Lätt, J., Ståhlberg, F., Topgaard, D., 2011. Apparent exchange rate mapping with diffusion MRI. Magn. Reson. Med. 66, 356–365.

Lätt, J., Nilsson, M., Van Westen, D., Wirestam, R., Ståhlberg, F., Brockstedt, S., 2009. Diffusion-weighted MRI measurements on stroke patients reveal water-exchange mechanisms in sub-acute ischaemic lesions. NMR Biomed. 22, 619–628.

Matsushima, G.K., Morell, P., 2001. The neurotoxicant, cuprizone, as a model to study demyelination and remyelination in the central nervous system. Brain Pathol. 11, 107–116.

Mouton, P., 2002. Principles and Practices of Unbiased Stereology: an Introduction for Bioscientists. Johns Hopkins University Press, Baltimore.

Nedjati-Gilani, G.L., Schneider, T., Hall, M.G., Cawley, N., Hill, I., Ciccarelli, O., Drobnjak, I., Wheeler-Kingshott, C.A.G., Alexander, D.C., 2017. Machine learning based compartment models with permeability for white matter microstructure imaging. Neuroimage 150, 119–135.

Nilsson, M., Alerstam, E., Wirestam, R., Sta, F., Brockstedt, S., Lätt, J., 2010. Evaluating the accuracy and precision of a two-compartment Kärger model using Monte Carlo simulations. J. Magn. Reson. 206, 59–67.

Nilsson, M., Lätt, J., Ståhlberg, F., Van Westen, D., Hagslätt, H., 2012. The importance of axonal undulation in diffusion MR measurements: a Monte Carlo simulation study. NMR Biomed. 25, 795–805.

Nilsson, M., Van Westen, D., StåHLBERG, F., Sundgren, P.C., Lätt, J., 2013a. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. Magn. Reson. Mater. Phys., Biol. Med. 26, 345–370.

Nilsson, M., Lätt, J., Van Westen, D., Brockstedt, S., Lasič, S., Ståhlberg, F., Topgaard, D., 2013b. Noninvasive mapping of water diffusional exchange in the human brain using filter-exchange imaging. Magn. Reson. Med. 69, 1572–1580.

Novikov, D.S., Veraart, J., Jelescu, I.O, Fieremans, E., 2018. Rotationally-invariant mapping of scalar and orientational metrics of neuronal microstructure with diffusion MRI. Neuroimage 2018, 518–538.

Palombo, M., Ligneul, C., Hernandez-Garzon, E., Valette, J., 2018. Can we detect the effect of spines and leaflets on the diffusion of brain intracellular metabolites? Neuroimage 182, 283–293.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. J. Machine Learn. Res. 12, 2825–2830.

Prantner, A.M., 2008. Re-evaluation of Transmembrane Water Exchange in the Rat Brain. Washington University in St. Louis.

Quirk, J.D., Bretthorst, G.L., Duong, T.Q., Snyder, A.Z., Springer Jr, C.S., Ackerman, J.J., Neil, J.J., 2003. Equilibrium water exchange between the intra-and extracellular spaces of mammalian brain. Magnet. Resonance Med. 50, 493–499.

Ransohoff, R.M., 2012. Animal models of multiple sclerosis: the good, the bad and the bottom line. Nat. Neurosci. 15, 1074.

Regan, D.G., Kuchel, P.W., 2000. Mean residence time of molecules diffusing in a cell bounded by a semi-permeable membrane: Monte Carlo simulations and an expression relating membrane transition probability to permeability. Eur. Biophys. J. 29, 221–227.

Reisert, M., Kellner, E., Dhital, B., Hennig, J., Kiselev, V.G., 2017. Disentangling micro from mesostructure by diffusion MRI: a Bayesian approach. Neuroimage 147, 964–975.

Ronen, I., et al., 2014. Microstructural organization of axons in the human corpus callosum quantified by diffusion-weighted magnetic resonance spectroscopy of N-acetylaspartate and post-mortem histology. Brain Struct. Funct. 219 (5), 1773–1785.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676.

Sepherband, F., Alexander, D.C., Kurniawan, N.D., Reutens, D.C., Yang, Z., 2016. Towards higher sensitivity and stability of axon diameter estimation with diffusion-weighted MRI. NMR Biomed. 29 (3), 293–308.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219.

Song, S.-K., Yoshino, J., Le, T.Q., Lin, S.-J., Sun, S.-W., Cross, A.H., Armstrong, R.C., 2005. Demyelination increases radial diffusivity in corpus callosum of mouse brain. Neuroimage 26, 132–140.

Stanisz, G.J., Odrobina, E.E., Pun, J., Escaravage, M., Graham, S.J., Bronskill, M.J., Henkelman, R.M., 2005. T1, T2 relaxation and magnetization transfer in tissue at 3T. Magnet. Resonance Med. 54, 507–512.

Tanno, R., Ghosh, A., Grussu, F., Kaden, E., Criminisi, A., Alexander, D.C., 2016. Bayesian image quality transfer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 265–273.

Vangelderen, P., Despres, D., Vanzijl, P., Moonen, C., 1994. Evaluation of restricted diffusion in cylinders. Phosphocreatine in rabbit leg muscle. J. Magnet. Resonance, Series B 103, 255–260.

Volles, M.J., Lee, S.-J., Rochet, J.-C., Shtilerman, M.D., Ding, T.T., Kessler, J.C., Lansbury, P.T., 2001. Vesicle permeabilization by protofibrillar α-synuclein: implications for the pathogenesis and treatment of Parkinson's disease. Biochemistry 40, 7812–7819.

Wang, N., Zhang, J., Cofer, G., Qi, Y., Anderson, R.J., White, L.E., Johnson, G.A., 2019. Neurite orientation dispersion and density imaging of mouse brain microstructure. Brain Struct. Funct. 1–17.

Wassermann, D., Santin, M., Philippe, A.C., Fick, R., Deriche, R., Lehericy, S., & Petiet, A. 2017. Test-Retest qt-dMRI datasets for "Non-Parametric GraphNet-Regularized Representation of dMRI in Space and Time"; [Data set]. Zenodo. doi:10.5281/zenodo.996889.

Westin, C.-F., Maier, S.E., Mamata, H., Nabavi, A., Jolesz, F.A., Kikinis, R., 2002. Processing and visualization for diffusion tensor MRI. Med. Image Anal. 6, 93–108.

Wu, Q.Z., Yang, Q., Cate, H.S., Kemper, D., Binder, M., Wang, H.X., Fang, K., Quick, M.J., Marriott, M., Kilpatrick, T.J., 2008. MRI identification of the rostral-caudal pattern of pathology within the corpus callosum in the cuprizone mouse model. J. Magnet. Resonance Img. 27, 446–453.

Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C., 2012a. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 61, 1000–1016.

Zhang, J., Jones, M.V., Mcmahon, M.T., Mori, S., Calabresi, P.A., 2012b. In vivo and ex vivo diffusion tensor imaging of cuprizone-induced demyelination in the mouse corpus callosum. Magn. Reson. Med. 67, 750–759.