Machine learning for improved diffusion MRI parameter estimation with gradient nonlinearity correction

Paula Castro Ramírez
p.castroramirez@students.uu.nl
Supervisors: dr. ir. Chantal Tax, dr. ir. Alberto De Luca. Second reviewer: dr. dr.h.c.mult Alexander Leemans
MSc Medical Imaging, Graduate School of Life Sciences, Utrecht University
Image Sciences Institute, University Medical Center Utrecht (UMCU)

***Abstract*—White matter (WM) degradation is one of the most common lesions causing neurological disorders, where early diagnosis is crucial. Diffusion tensor imaging is a widely used method for the study of WM injuries in the brain, however its estimation using traditional methods such as non-linear least squares (NLLS) is time consuming and patient dependent. This work presents a machine learning (ML) approach for diffusion MRI parameter estimation. A feed forward network following two different strategies is presented to attempt the fitting of a symmetric diffusion tensor model. The possibility of real-time mapping is approached by four different experimental setups with increasing data complexity during the ML training, including synthetic and real signal simulations. Moreover, a novel approach on gradient nonlinearity (GNL) correction using ML is presented, opening the possibility of correcting for spatially varying b-values and b-vectors while training the model. The results showed an acceptable ML performance compared to NLLS when training and testing with one single subject, demonstrating that ML can be used for parameter estimation of diffusion images and GNL correction. However, the generalization of the network to accept more than one subject is still a challenge. Further hyperparameter tuning and architecture configuration experiments are needed to generate comparable results to NLLS. Nevertheless, these initial results highlighted crucial aspects in the fitting process that could be important for future research of the topic.**

***Index Terms*—Diffusion MRI, DTI, Machine Learning, Parameter estimation, Microstructural imaging, Zeppelin model.**

I. Introduction

White matter (WM) degradation is one of the most common lesions causing neurological diseases. These disorders mainly arise due anoxia, ischemia, trauma or autoimmune attacks, which result in swelling, damage or complete loss of cellular axons inside the brain [1]. Crucial early diagnosis of these kind of diseases has raised interest in the improvement of non-invasive techniques to effectively evaluate patient's conditions and develop more accurate treatment plans [2].

Magnetic resonance imaging (MRI) is a non-invasive imaging technique that uses powerful magnets to generate a strong magnetic field. When a individual is introduced in that magnetic field, the protons in the body are forced to align with it. Then, a radio frequency (RF) pulse is applied, making the protons turn into an excited state and temporarily spin. Lastly, the RF pulse is shut down, and the energy released by those protons is detected by RF receivers. That information will be used later to reconstruct the images [3]. For many years, conventional MRI has been the gold standard technique for the diagnosis of WM related diseases [4], however it may lack of the necessary sensitivity when detecting specific structures and pathologies [5].

Diffusion weighted imaging (DWI) can examine the microstructural configuration of tissues at a much lower scale than standard MRI. By focusing on the microscopic random motion of water molecules, it is possible to detect pathologies associated with WM more accurately [6]. The first method used to generate DWI was developed by Stejskal and Tanner [7], and was based on the original MRI spin echo sequence that consisted of a 90º pulse followed by a 180º pulse [8]. Pulsed gradient Spin Echo (PGSE) sequencing relies on two extra diffusion sensitizing gradients applied symmetrically to the 180º pulse, where the first one introduces a phase shift on the proton's position, and the second one reverses the change. The traditional PGSE method has been generalized to account for more complex direction configurations present inside the human brain. By repeating the simple experiment that describes the diffusion gradient in one direction in at least other six non-collinear directions, information about the flow and distance of water molecules could be provided [9][10].

Diffusion tensor imaging (DTI) is a strategy that models quantitatively structural and orientation information constructing the so-called diffusion tensor [11]. The correct estimation of diffusion parameters coming from the DTI formulation can effectively detect non-invasively and more accurately WM injuries. For instance, axial diffusivity (AD) has proved to accurately locate axonal damage in patients with Alzheimer's disease [12], and radial diffusivity (RD) has demonstrated to predict demyelination in multiple sclerosis [13]. However, this methodology relies on a simple single compartment assuming microscopically homogeneous environments in each voxel, and WM tissues might include multiple compartments with different behaviours [9]. Microstructural modeling relates the microscopic details of the tissue structures with the diffusion MR signals through the introduction of *compartment models* [14].

Parameter estimation of multiple microstructural components per voxel can be challenging and very time consuming, which makes the possibility of real-time mapping almost

impossible with traditional methods such as non-linear Least Squares (NLLS). Additionally, both DWI and DTI come with various artifacts that, if not accounted for, could lead to incorrect interpretation of parameter maps and signals. Some examples can be Gibbs ringing artifacts, integrated motion errors, eddy currents, susceptibility distortions, or gradient deviations [15]. Nowadays, advancements in technology have proven to be of use in this field, where machine learning (ML) networks could potentially address some of the previously mentioned limitations. ML networks consist of algorithms able to learn complex relationships between data by going through a series of hidden layers [16]. In the context of DTI, these networks could potentially fit biophysical models significantly faster and more precisely than conventional optimization techniques [17][18][19].

This project aims to use ML networks to predict DWI parameter maps by training and testing with different sets of data. We study the feasibility of fitting a symmetric tensor model using ML with multiple patients. Moreover, we also investigate the possibility of accounting for gradient nonlinearity (GNL) corrections during the network training.

II. Methods

This section will first explain the biophysical model chosen to fit the diffusion MRI data. Next, GNL deviations are introduced with an explanation on how to correct for them. Afterwards, we present the datasets used to execute the experiments, which consisted of simulations of synthetic and real data. Lastly, the three different strategies followed to fit the model and correct for the deviations are described, specifying the distinct experiments executed in each of them. All the codes developed can be found in GitHub github.com/PaulaCastroRamirez/ML_for_parameter_fitting.

A. Microstructural modeling

In order to model diffusion MRI data, mathematical representations are used to estimate different parameters and infer properties of the tissue microstructure. In this project, the biophysical model chosen is the so-called Zeppelin model, represented in Figure 1. The Zeppelin is a two-compartment cylindrically-symmetric diffusion tensor model with the shape of an oblate spheroid. It is commonly used in diffusion MRI modeling because it assumes restricted diffusion within elongated structures such as fiber bundles, where water does not follow spherical free motion anymore. The model has two principal eigenvalues (λ_1, λ_2), corresponding to axial diffusivity (AD) and radial diffusivity (RD), respectively [20][21][22].

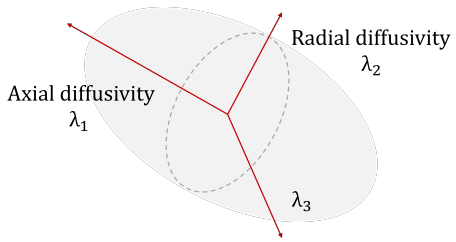

Fig. 1. Zeppelin Biophysical model with Axial (parallel) diffusivity (λ_1) and Radial (perpendicular) diffusivity ((λ_2+λ_3)/2). For Zeppelin model, λ_2=λ_3.

The parameters estimated are summarized in Table I, together with their corresponding physical bounds [18]. The baseline signal intensity (S_0) corresponds to the signal measured without any diffusion gradients. In diffusion MRI, it is used as a reference for the comparison between the different diffusion-weighted signal attenuations applied. AD (λ_1) describes the diffusion of water molecules parallel to the main direction of diffusion. Moreover, the water diffusivities perpendicular to the main direction of diffusion, λ_2 and λ_3, are averaged and denoted as RD ((λ_2+λ_3)/2). Lastly θ and ϕ represent the angle of the first eigenvalue of the diffusion tensor.

TABLE I
Physical parameter bounds
for the Zeppelin diffusion tensor model

Parameter	Bound
Baseline intensity (S_0)	[0.0 - inf]
Axial diffusivity (AD)	[0.0 - 3.2] μm^2/ms
Radial diffusivity (k*AD)	[0.0 - 3.2] μm^2/ms
Theta (θ)	[0.0 - π] rad
Phi (ϕ)	[0.0 - 2π] rad

According to the Zeppelin tensor model, the definition for the signal decay is given by Equation 1 [23]. Because the *in vivo* data used (detailed in Section II-D), did not contain echo time or T2 variations, the equation was modified as needed.

$$S\left(b,\Theta,\Phi,AD,k,S_0\right) = S_0 \exp -\left[\frac{1}{3}\, b\left(AD + 2kAD\right) + b\left(\theta_g,\phi_g\right)\cdot\left(\Theta,\Phi\right)\right)^2\left(AD - kAD\right)\right] \quad (1)$$

with b as the b-values, (Θ, Φ) defining the first eigenvector direction, AD as the axial diffusivity, kAD as the radial diffusivity, and S_0 as the baseline intensity.

B. Gradient Nonlinearity and Spatially varying L-matrix

Diffusion imaging exhibit relevant artifacts due to nonuniformities in the magnetic field gradients. New developments in diffusion MRI consider more advanced scanners which apply larger gradient amplitudes (300 mT/m) [24] compared to the conventional ones (30-45 mT/m) to increase resolution and get information at smaller scales. This can result in spatially dependent inaccuracies in the diffusion-encoding direction and orientation, leading to misinterpretation of the diffusion tensor. Given the information about the relative deviations of the magnetic field, GNL can be corrected pixel-wise. A gradient coil tensor *L(r)* can be defined for each location *r*, to relate the

effective gradient G_{act}, and the applied gradient G. Following the formulation from Bammer et al. (2003) [25]:

$$G_{\text{act}} = \begin{bmatrix} L_{xx} & L_{xy} & L_{xz} \\ L_{yx} & L_{yy} & L_{yz} \\ L_{zx} & L_{zy} & L_{zz} \end{bmatrix} G = L(r)G \tag{2}$$

Consequently, the effective b-matrix $B_{eff}(r)$ can also be related to the applied b-matrix B and computed for every voxel:

$$B_{\text{eff}}(r) = L(r)BL(r)^T \tag{3}$$

C. Simulations

Due to the complexity of diffusion MRI data, simulations were generated from different synthetic distributions of parameters. The simulations were obtained from parameters taken from uniformly distributed values scattered within their corresponding physical bounds (see Table I for bounds). Next, assuming those set of parameters as ground truth (S_0, AD, RD, θ and ϕ), we fed them into the Zeppelin model (Equation 1) to simulate ground truth signals. Finally, ground truth signals were corrupted with noise to generate the datasets that will be used for training, validation and testing of the methods. Three distinct noise distribution types were implemented, although only one was chosen to conduct the experiments. The noise types were Rician, Gaussian, and Noncentral Chi-Squared.

In total, we generated 8000 signals for the training set, 2000 for the validation set, and 1000 for the test following the same protocol (b-values) as for the *in vivo* experiments described in Section II-D. The noise type chosen was Gaussian with a signal-to-noise ratio (SNR) of 70.

D. In vivo experiments

The second part of the project consisted on testing the methodology using simulated data generated from real parameter distributions.

The datasets were taken from the Human Connectome Project (HCP) Q1 data release [26]. Diffusion MRI scans were acquired on a Siemens Skyra 3T scanner following a Spin-echo EPI sequence with a customized SC72 gradient insert, which allowed an important quality improvement of the diffusion images [27]. However, due to technical limitations associated with the gradient hardware and scanner design, GNL increased, exceeding the ones of the conventional 3T scanner. HCP datasets were corrected for geometric image distortions caused by eddy currents, susceptibility errors, and gradient field nonliearities [24], but not for voxel-wise changes in the direction and intensity of the diffusion encoding gradients. Effective b-values and b-vectors still contained slight variations from voxel-to-voxel that needed to be taken into consideration [28] (see representation of gradient deviations in Figure 2). A basic script provided in the HCP publication [29] was used to correct for GNL by calculating those individual b-values and gradient orientations in each voxel.

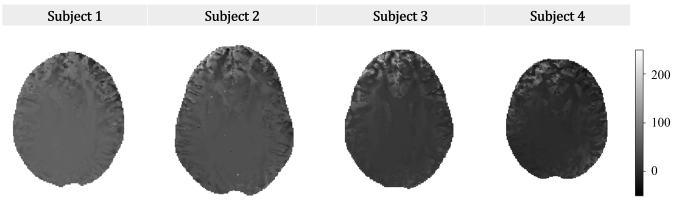

Fig. 2. Example of Gradient nonlinearities (GNL) from HCP subjects.

Four datasets were randomly chosen from the repository, all from a separate subject. Each diffusion MRI dataset was acquired with three different gradient tables, each table consisting of 90 diffusion weighting directions (shells with b-values 1000, 2000, and 3000 s/mm^2) and six b=0 images. Theoretically, more diffusion weighting (higher b-value) translate into different sensitivity towards microstructural features and diffusion processes inside the brain. The diffusion of water at higher b-values may be influenced by factors that were not taken into account in the simple model chosen [30]. Because of the simplicity of the Zeppelin tensor model, the resulting 288 measurements per signal (ordered from low to high b-value) were restricted to the first 108.

Ground truth parameters were simulated fitting each of the datasets from the HCP using non-linear least squares (NLLS) optimization. Taking those parameters, signals were back-simulated fitting the Zeppelin equation presented in Equation 1, and considered as ground truth signals. Lastly, noise was added to corrupt the ground truth signals and create the training, validation, and test sets in the same way we did with the synthetic signal simulations described in Section II-C. The noise distribution chosen to disrupt *in vivo* experiments was Gaussian with a SNR of 70.

E. Non-linear Least Squares optimization

NLLS optimization was chosen as the reference method in this work to compare and validate the novel ML approaches as it is one of the classical methods used for diffusion tensor fitting [17][31]. NLLS consists of solving a non-linear optimization problem with bounds (see Table I) on the fitted variables with the goal of finding the local minimum of a cost function [32]. The equation optimized was the Zeppelin as described in Equation 1. As the initial value for the variables in each pixel was randomly selected between the bounds, multiple initializations were needed during the process. At the end, the solution with lowest residuals was chosen as the right one.

Two different experiments were performed, with and without GNL correction. NLLS without GNL correction was executed as a normal optimization problem giving as an input the b-values and gradient directions without any customized preprocessing. Moreover, when taking into account GNL,

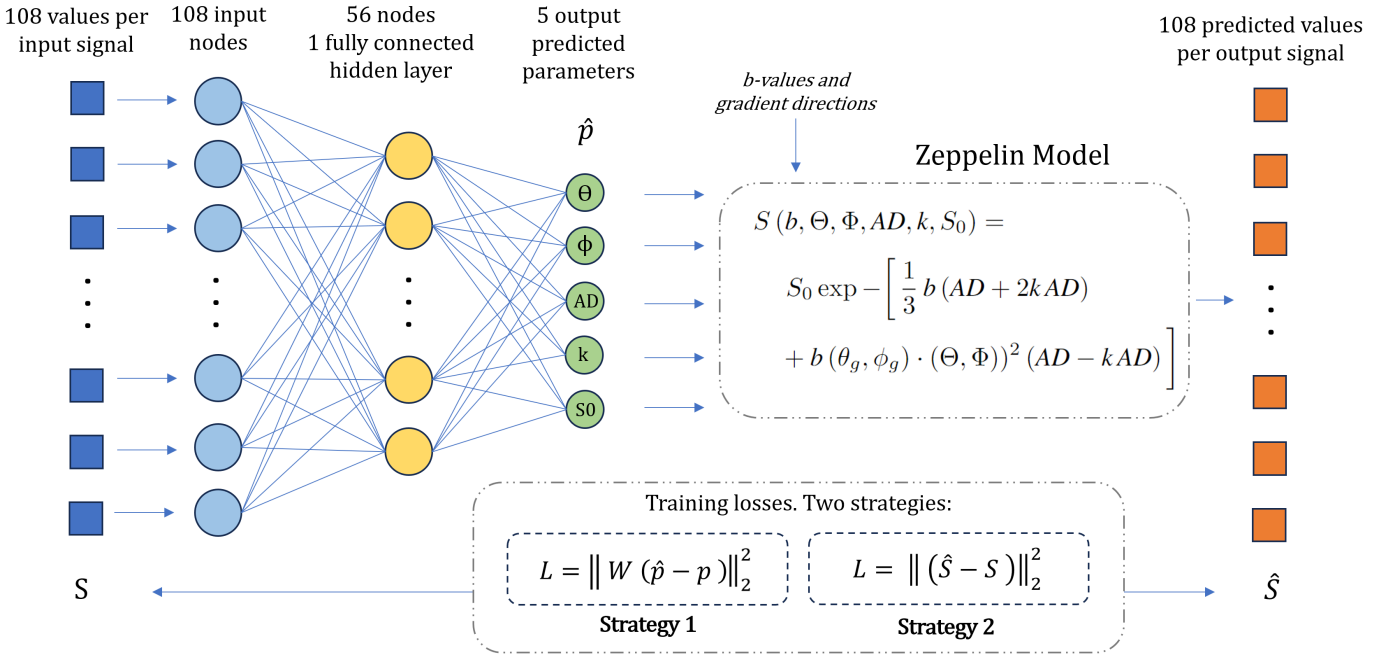

Fig. 3. Feed forward ML network proposed. Each artificial neuron in the hidden layer receives a number of input signals, x_1, x_2, ..., x_n, and gives as an output the parameters of interest (ϕ, θ, AD, k and S_0). Two strategies were used to calculate the training loss: Strategy 1 calculated the loss between tissue parameters and Strategy 2 calculated the loss between tissue signals [19].

b-values and gradient directions were corrected using the corresponding gradient deviation files provided by the HCP release and following the *L(r)* matrix correction from Bammer et al. (2003) [25] described in Section II-B.

F. Machine learning network

The network implemented in this paper was a feed forward network. Based on the two strategies presented on the paper from Grussu, Francesco, et al. (2021) [19], we propose a modified generalized network able to accept multiple subjects during the training, validation and testing. The default architecture consisted of 108 input nodes, 1 hidden layer with 56 nodes, and 5 output nodes, as depicted in Figure 3. Signals were fed into the network, each one consisting of 108 measurements per pixel (*S*), and passed through the hidden layer, which gave 5 parameters ($\hat{p}$) as the output. All the predicted parameters followed a min-max normalization after undergoing a series of steps which assured correct scaling between the desired bounds [19]. Lastly, if the second strategy was being followed, those $\hat{p}$ were given as an input to the Zeppelin model equation generating the final predicted signal $\hat{S}$. The training loss was calculated differently in two separate strategies. Strategy 1 calculated the mean squared error (MSE) loss between predicted parameters $\hat{p}$ and input parameters p, while Strategy 2 calculated the MSE loss between predicted signals $\hat{S}$, and input signals *S*. Moreover, to enable the network to include more than one dataset during the training, a customized data class was created giving signals as the input of each batch together with their corresponding b-values and gradient directions. The signals were fed into the network, while the b-values and gradient directions were used later for the calculation of the output signals from the predicted parameters. It must be highlighted that all the signals used for training were contained inside the brain, as the background was masked out using Otsu thresholding method [33].

1) ***Strategy 1****:* The first ML strategy calculated the training loss between predicted and input parameters as expressed in Equation 4.

$$L = \|W(\hat{p} - p)\|_2^2 \tag{4}$$

Following the scheme represented in Figure 3, Strategy 1 consisted of a supervised ML network. This method took input signals (*S*), predicted the corresponding parameters ($\hat{p}$), and compared those $\hat{p}$ to the real input parameters (p). To achieve this, real parameters were also given as an input during the training, but they were only used during the MSE loss calculation. In this approach, no GNL correction was conducted as it was not included directly in the parameter simulations. The correction was performed in the b-values and gradient directions pixel-wise, therefore, even if included, the MSE loss calculation would not have been influenced by them.

2) ***Strategy 2****:* The second ML strategy calculated the training loss between predicted and input signals as shown in Equation 5.

$$L = \|(\hat{S} - S)\|_2^2 \tag{5}$$

Following Figure 3, Strategy 2 consisted of a self-supervised ML network. This method took signals as an input (S), predicted the corresponding parameters ($\hat{p}$), forward-simulated signals with those $\hat{p}$ using the Zeppelin tensor model equation, and compared the predicted signal ($\hat{S}$) to the real input signal (S). For this strategy, GNL correction was included. All the experiments conducted following Strategy 2 compared results for correcting and not correcting for GNL.

The model was implemented in Python 3.10 using PyTorch 23.12 [34]. We used Adam as the optimizer, no dropout, constant learning rate of 0.0001, and batch size of 256. Hyperparameter tuning is one of the main steps when creating and regulating a ML model to get to its optimal capability. The default hyperparameter values that can be seen in Table II were defined based on several exhaustive experiments detailed in the following section.

TABLE II
DEFAULT NETWORK INFORMATION AND HYPERPARAMETERS

Hyperparameter	Default
Angle coordinates	Polar
Learning rate	0.0001
Nº of epoch	250
Batch size	256
Nº of workers	4
Network architecture	108-56-5
Signal normalization	Pixel-wise

G. Experimental setup

To study the generalizability of the methods proposed, four experiments were conducted. A summary can be found in Table III. For the first set of experiments, synthetic simulations of parameters and signals were used during the training and testing. For the second set, the training was performed with one dataset, and tested with slices from that same dataset. For the third set, the training was performed with four datasets, with the middle slice of each one taken out and used during the testing. Lastly, for the fourth, the training was conducted with three datasets, and tested with one completely new dataset. By separating the data in these four experiments, we were able to study how the methods proposed behaved in the presence of new data coming from three different sources: synthetic signals, signals from the same training subjects, and signals from new subjects.

a) ***Experiment 1. Synthetic simulations****:* Experiments were carried out using synthetic simulations (Section II-C), and default parameters were employed in the architecture setup for both ML strategies (see Table II). To study the influence of the data coordinate system and scattering, two key aspects were examined in while training Strategy 1: input distributions and angle transformations.

By changing how the parameter pairs were distributed, we were able to investigate how the data scattering affected the method proposed. Two different parameter distributions were configured in the simulations. The first approach was based on uniform distributions, which were generated individually for each parameter, and bounded according to the maximum and minimum values shown in Table I. Moreover, the second approach relied on von Mises distributions because of their good behaviour when dealing with directional data [35][36][32].

Furthermore, Strategy 1 calculated the training MSE loss between input and predicted parameters, which involved treating polar angles and diffusivities similarly. That matter brought up several aspects that MSE did not take into account, like periodicity or the spherical nature. Three angle transformations were tested. Apart from the default polar coordinate system, we also investigated using a sine transformation and a Cartesian coordinate system transformation before the MSE loss calculation.

b) ***Experiment 2. Single subject****:* Experiments were performed on real HCP data from a single subject, and tested with slices from the same subject. This experiment was taken as the proof of concept for testing the ML approaches with our dataset and model, as it proved to work with other setups and biophysical models [19][18]. From a total of 145 slices, 140 were divided into training and validation sets (70-30%, respectively), and 5 were used for testing (3% from the total). Default hyperparameters shown in Table II were employed for both strategies.

c) ***Experiment 3. Multiple subjects****:* Experiments were executed using data from four different subjects, and tested with slices from the same subjects. The middle slice was reserved from each dataset to create the test set, which consisted of a total of four slices. Hyperparameter tuning was conducted in this experiment, which involved the learning rate, the number of hidden layers and number of nodes per hidden layer, the dropout rate, and the input signal normalization.

- The learning rate defines the step size at which the objective function will converge during the training process [37]. The main tests were based on reducing the learning rate in each experiment by a factor of 10, from 0.01 gradually to 0.0001. Additionally, the possibility of including an exponential learning rate scheduler was examined. It was defined to reduce the learning rate at every epoch by a gamma value of 0.9 from an initial learning rate of 0.01.
- Concerning the network architecture, several tests were performed involving the number of hidden layers and nodes in the model. By default, the model was set to have only one hidden layer with a reduced number of nodes (108-56-5), due to the low complexity of the Zeppelin model (Equation 1). However, two other alternatives were assessed. The first approach consisted on changing the model's architecture to have one hidden layer with the same number of nodes as the input layer (108-108-5), and the second involved having three hidden layers with the same number of nodes as the input layer (108-108-108-5).

TABLE III
EXPERIMENT SPECIFICATIONS PER ML STRATEGY, DATA DISTRIBUTION, AND PURPOSE

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
ML Strategy 1 experiments	- Uniform vs von Mises - Polar vs Sine vs Cartesian	Default parameters	Default parameters	- Input normalization
ML Strategy 2 expetiments	Default parameters	Default parameters	- Learning rate - ML architecture - Dropout rate - Input normalization	- Input normalization
Train and Test data	- Train: synthetic simulations - Test: synthetic simulations	- Train: one subject - Test: same subject	- Train: multiple subjects - Test: same subjects	- Train: multiple subjects - Test: new subject
Purpose	- Evaluate ML methods - ML Strategy 1 analysis	Proof of concept ML methods	Hyperparameter tuning ML	Generalizability test ML

- The dropout rate was examined with the aim of regularizing the model and avoid overfitting [37]. By default, it was set to 0.0 (no dropout). However, several alternatives were studied by performing multiple tests and gradually reducing it from 0.3 to 0.0.
- Tests were performed on the normalization of input signal intensities. Two main strategies were explored. In the first one, all pixels were normalized independently dividing by their corresponding maximum b=0 signal intensity. In the second one, all the data was normalized based on the maximum b=0 global intensity of the whole image.

*d) **Experiment 4. Multiple subjects**:* Once the ML model was tuned with optimal hyperparameters, experiments were conducted using data from three different subjects (two for training and one for validation), and tested with one new subject. With this approach, we concluded our original attempt to generalize both strategies to accept more than one subject at a time. Default parameters seen in Table II were used for both strategies, except for the input signal normalization. For this set of experiments, an scaling factor was calculated and used to normalize the input signals. This scaling factor was individual for each subject and was calculated as follows. Taking one subject as a reference, we calculated its median intensity. Next, the median intensity of other subject was computed, and a linear scaling factor was determined between them. Finally, both subjects were normalized dividing by the maximum intensity of the reference subject.

H. Evaluation of results

To evaluate the performance of each experiment, different quantitative methods were carried out. For each of the four experiments, scatter plots were generated by plotting ground truth parameters (x-axis) vs predicted parameters (y-axis). This evaluation method allowed to visually show the quality of the relationship between the predictions and the real data, i.e. how well the model predicted the data. Additionally, the angles (θ and ϕ) were converted to Cartesian coordinates and evaluated using cosine similarity histograms. This method relies on the cosine concept, that is, cosine similarity values closer to 1 indicate similar vectors, and closer to 0 orthogonal vectors. For a more intuitive representation of the metric, we subtracted the absolute value of the cosine similarity to 1. Following Equation 6 with G as ground truth angles and P as predicted angles, values closer to 0 were interpreted as similar vectors, and values closer to 1 as orthogonal vectors.

$$1 - |\cos\alpha| = 1 - \left| \frac{G \times P}{|G| \times |P|} \right| \tag{6}$$

To evaluate the reconstruction of real data distributions (Experiments 2 to 4), maps were generated showing the percentage of the absolute differences between predicted and ground truth values.

Moreover, metrics were also calculated together with the scatter plots and difference maps. To quantify the different methods' performances, we used the R-squared (R^2) coefficient and the mean absolute error (MAE) [38].

III. RESULTS

In this section, the results from the different experimental setups are presented. Specifically, we present the results for the most optimal combination of parameters found (default values seen in Table II), based on the experiments detailed in Section II-G. For Experiments 1 and 2, results for NLLS, ML Strategy 1 and ML Strategy 2 are shown. Moreover, for Experiments 3 and 4, ML Strategy 1 and Strategy 2 results are presented.

A. Experiment 1: Synthetic simulations

The results from using simulations calculated from uniformly distributed parameters can be seen in Figure 4. Scatter plots demonstrate that the three methods worked, specially when predicting S_0, AD and RD. However, it can also be seen that as the ML model becomes more complex (Strategy 2), it performed worse using this data arrangement. Focusing on the cosine similarity histograms, while for NLLS the histogram was centered at 0, Strategy 1 histogram was uniformly scattered between 0 and 1, and Strategy 2 presented a considerable amount of values $>$0.5.

To asses numerically the behaviour of simulations, R^2 and MAE values are presented in Table IV. Focusing on the R^2 metric, we can see that apart from RD in Strategy 2, which had a value of 0.537, all parameters had a R^2 higher than 0.860. The same behaviour could be deduced from MAE values,

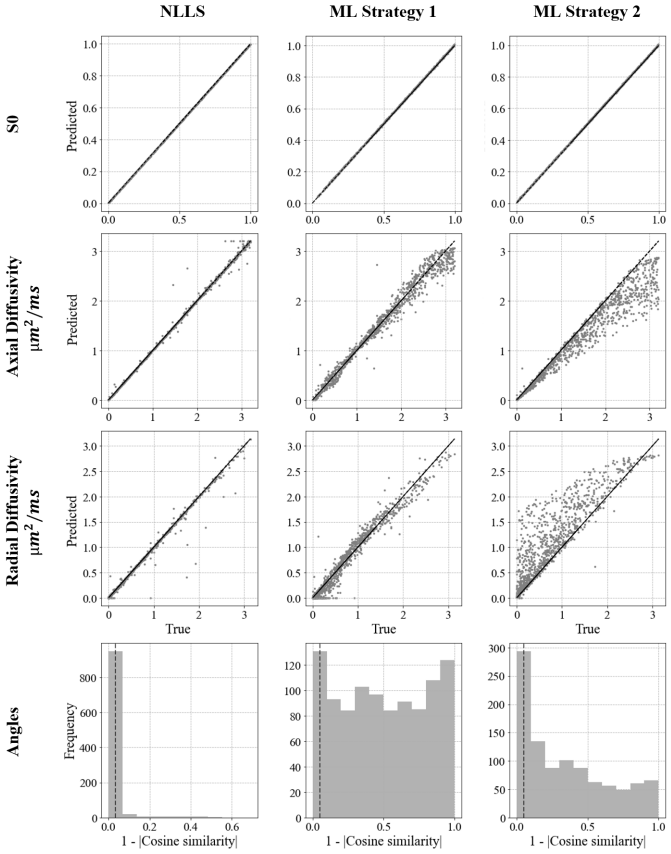

Fig. 4. Scatter plots and histograms 1-|cosine similarity| for **Experiment 1**. Columns correspond to methods while rows represent the specific parameters.

where mostly all methods obtained values <0.1. Following this results, the next step of the project was formulated based on simulations generated from real data.

TABLE IV
METRIC RESULTS FROM **EXPERIMENT 1**.

Parameter	Method	R^2	MAE
S0	NLLS	>0.999	0.001
	ML Strategy 1	0.997	0.006
	ML Strategy 2	>0.999	0.003
AD	NLLS	0.996	0.015
	ML Strategy 1	0.975	0.100
	ML Strategy 2	0.860	0.225
RD	NLLS	0.985	0.020
	ML Strategy 1	0.956	0.105
	ML Strategy 2	0.537	0.325

B. Experiment 2: Single subject

In Figure 5 we can see the scatter plots and cosine similarity histograms for the second experiment. The plots show that all the methods followed the expected straight trend line, and that the all cosine similarity histograms were centered at 0.

Moreover, Figure 6 displays the % of the absolute differences between the ground truth and the corresponding reconstructed maps for all the methods. NLLS optimization successfully corrected for GNL. Focusing the attention on the edges of the brain, without any correction we can appreciate a clear % of absolute differences (blue edges) of 10% that disappear when the correction was taken into consideration (<2%). Strategy 1 showed for S_0, AD and RD low a % of absolute differences, <20%. However, when looking at the error map for the angles, it showed a great number of areas with cosine similarity values >0.5. Lastly, Strategy 2 presented a similar behavior compared to Strategy 1 when predicting S_0, AD and RD, with % of absolute differences <20%. Despite that, an improvement on the angles' maps was demonstrated, with all values of cosine similarity metric <0.5.

Table V shows the R^2 and MAE values for the second set of experiments. Supporting the goal of achieving the method's proof of concept, both R^2 and MAE values confirmed the accuracy of the predictions, with values >0.975 and <0.06, respectively. However, it must be highlighted that, numerically, although GNL correction proved to work in NLLS, i.e. without GNL correction MAE value for AD was 0.014 and after GNL correction it reduced to 0.003, no significant difference could be appreciated in Strategy 2 due to the high range of errors coming from the ML model itself (0.033 without GNL correction compared to 0.043 with GNL correction).

TABLE V
METRIC RESULTS FROM **EXPERIMENT 2**.

Parameter	Method	R^2	MAE
S0	NLLS no GNL correction	>0.999	<0.001
	NLLS with GNL correction	>0.999	<0.001
	ML Strategy 1	0.998	0.002
	ML Strategy 2 no GNL correction	>0.999	0.001
	ML Strategy 2 with GNL correction	>0.999	0.001
AD	NLLS no GNL correction	0.997	0.014
	NLLS with GNL correction	>0.999	0.003
	ML Strategy 1	0.995	0.021
	ML Strategy 2 no GNL correction	0.987	0.033
	ML Strategy 2 with GNL correction	0.978	0.043
RD	NLLS no GNL correction	0.999	0.013
	NLLS with GNL correction	>0.999	0.004
	ML Strategy 1	0.991	0.034
	ML Strategy 2 no GNL correction	0.980	0.052
	ML Strategy 2 with GNL correction	0.975	0.060

C. Experiment 3: Multiple subjects

Figure 7 displays the scatter plots and cosine similarity histograms from the third experiment, where we trained with multiple subjects and tested with slices from that same set of subjects. For these experiments, a great difference in performance can be appreciated in between strategies. The scatter plots show that Strategy 1 could only accurately predict S_0, following the linear trend line, while it failed to predict AD and RD. For these last two parameters, it is clear that the observed relationship diverged from the ideal linear one. Moreover, the cosine similarity histogram was not centered at 0, showing also a great error range when predicting the angles. On the contrary, scatter plots and histograms from Strategy 2 confirmed an increased performance without and with GNL correction, following the expected behaviors.

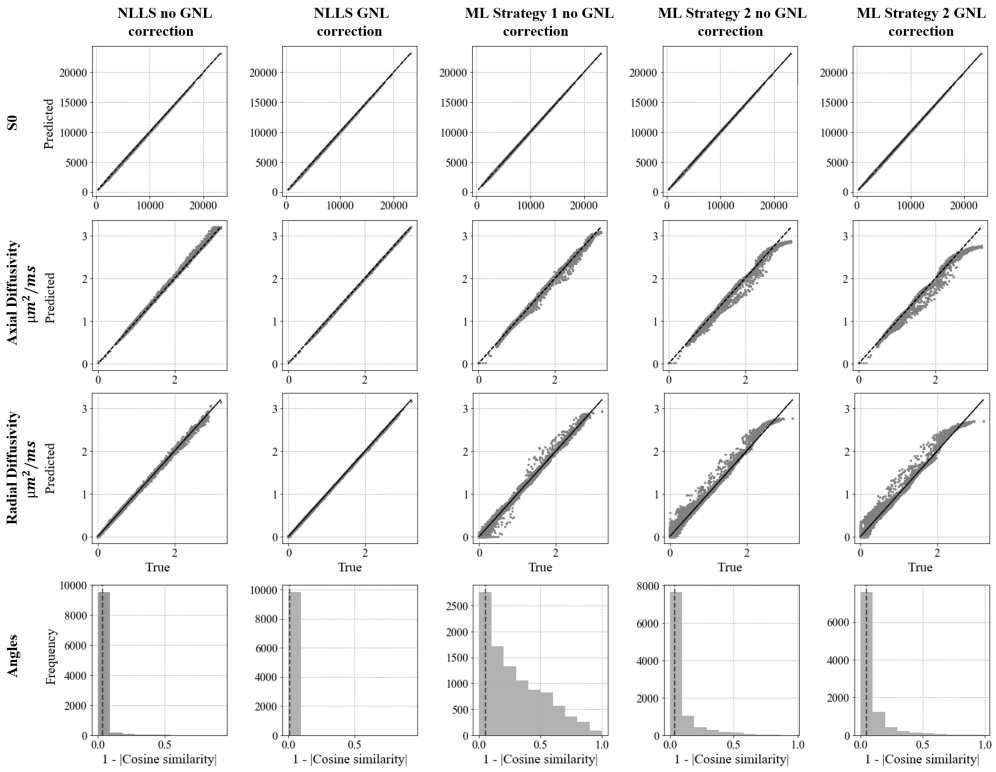

Fig. 5. Scatter plots and histograms 1-|cosine similarity| for **Experiment 2**. Columns correspond to methods while rows represent the specific parameters.

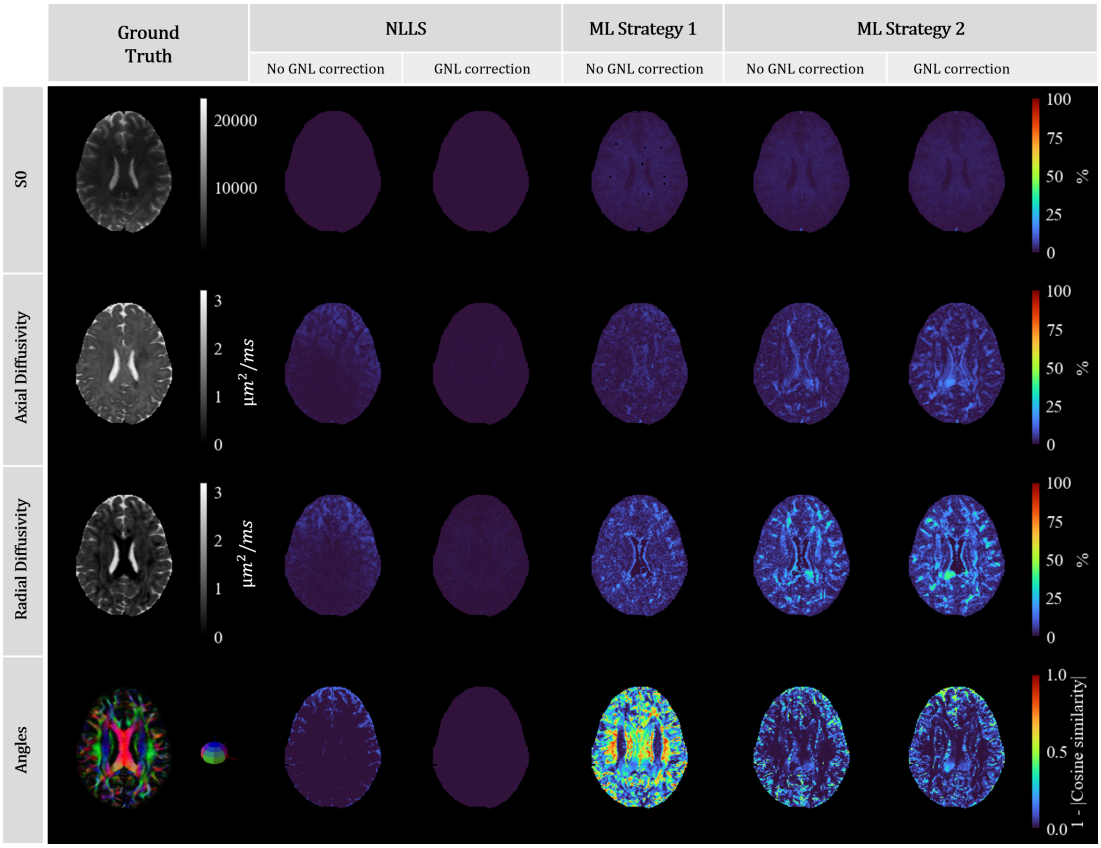

Fig. 6. Reconstruction results for **Experiment 2**. First column shows the ground truth parameter maps. The following columns show the % of absolute differences between predicted and ground truth simulations for each of the strategies followed. Rows correspond to the specific parameters.

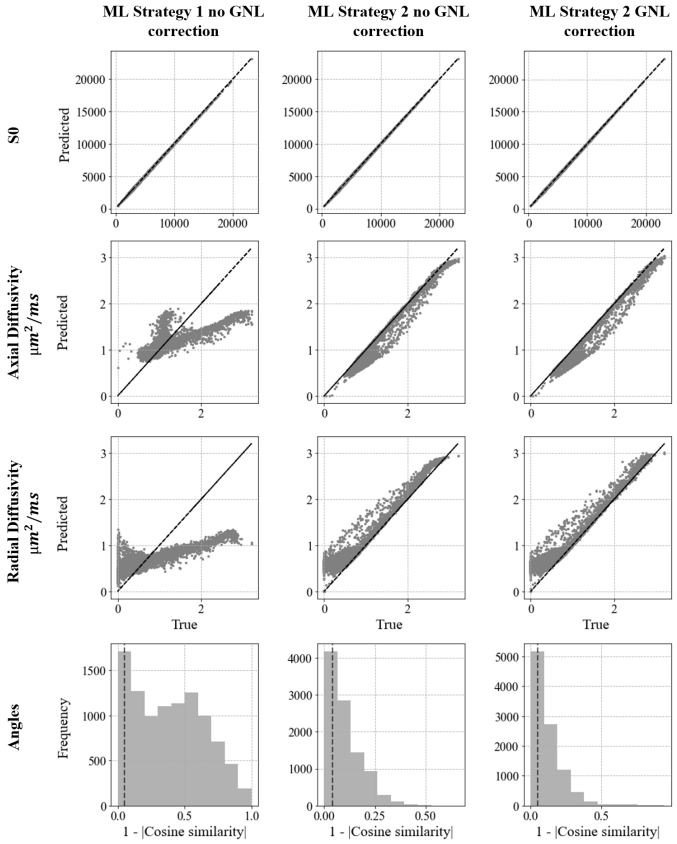

Fig. 7. Scatter plots and histograms 1-|cosine similarity| for **Experiment 3**. Columns correspond to methods while rows represent the specific parameters.

The % of absolute differences between ground truth and predicted maps are presented in Figure 9. From a first visual inspection, it is clear that adding more datasets to the training, made the model predict less accurately the parameter maps. Strategy 1 showed a maximum % of absolute differences in AD and RD of 60% and >80%, respectively. Moreover, the angles presented high cosine similarity as well, obtaining values >0.5. Regarding Strategy 2, AD and angle maps had better predictions compared to Strategy 1 (AD <35%, and angles <0.4). However, RD maps showed very high errors in between the ventricles with values >90%.

Metrics from Experiment 3 are presented in Table VI. Strategy 1 showed a considerably low R^2 for AD and RD maps (0.551 and 0.434, respectively) compared to Strategy 2 (0.991 and 0.78, respectively). Consequently, MAE values presented the same behavior. Strategy 1 had significant high errors for both diffusivities (AD and RD MAE values of 0.199, 0.268, respectively), while Strategy 2, although still with high, had sightly lower values (AD and RD MAE values of 0.106, 0.209, respectively). As exposed in Experiment 2, the high range of errors originated from the ML network itself made it impossible to appreciate the GNL correction performed in each of the parameter maps using either visual inspection (Figure 9) or quantitative metrics (Table VI).

TABLE VI
METRICS RESULTS FROM **EXPERIMENT 3**.

Parameter	Method	R^2	MAE
S0	ML Strategy 1	>0.999	0.003
	ML Strategy 2 no GNL correction	>0.999	0.002
	ML Strategy 2 with GNL correction	>0.999	0.002
AD	ML Strategy 1	0.551	0.199
	ML Strategy 2 no GNL correction	0.991	0.106
	ML Strategy 2 with GNL correction	0.893	0.119
RD	ML Strategy 1	0.434	0.268
	ML Strategy 2 no GNL correction	0.780	0.209
	ML Strategy 2 with GNL correction	0.800	0.195

D. Experiment 4: Multiple subjects

The scatter plots and histograms of cosine similarity from Experiment 4 are shown in Figure 8. Following the hyperparameter settings chosen for this experiment, both Strategy 1 and Strategy 2 showed the expected linear relationship for S_0, AD and RD. Also, for Strategy 2, both histograms of cosine similarity were centered at 0, demonstrating a good relationship between ground truth and predicted angles, while Strategy 1 failed to accurately predict the angles (histogram scattered between 0 and 1).

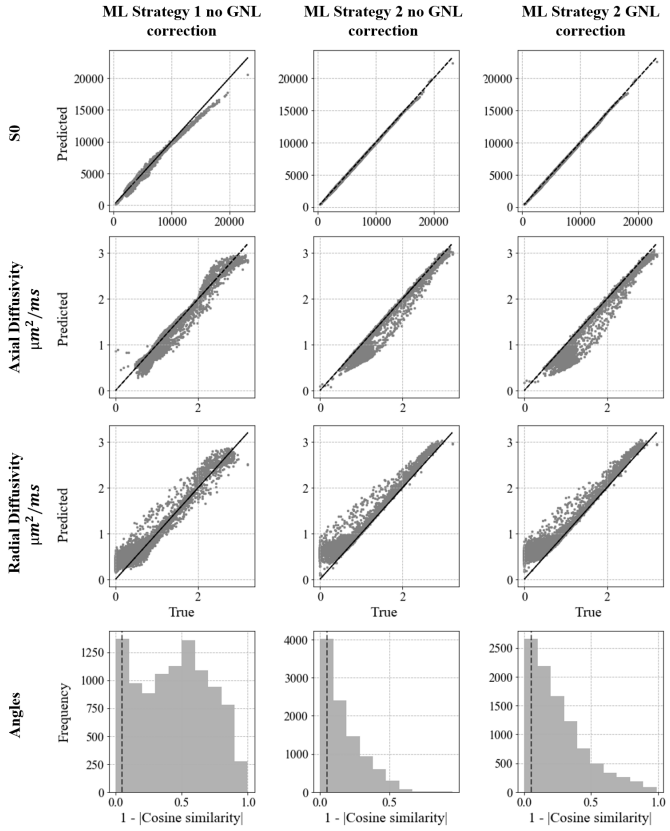

Fig. 8. Scatter plots and histograms 1-|cosine similarity| for **Experiment 4**. Columns correspond to methods while rows represent the specific parameters.

In Figure 10 the % of absolute differences from each parameter map is illustrated. Strategy 1 presented errors <30% in AD, and <65% in RD. However, the angle's cosine

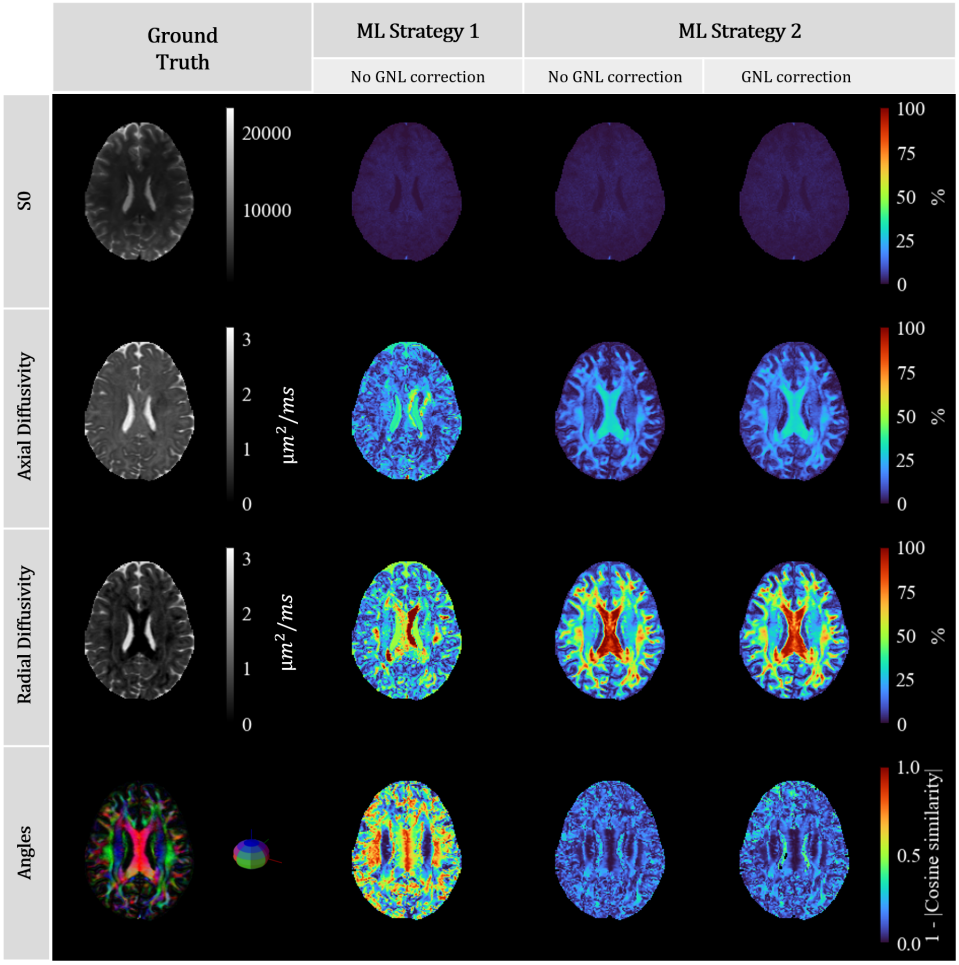

Fig. 9. Reconstruction results for **Experiment 3**. First column shows the ground truth parameter maps. The following columns show the % of absolute differences between predicted and ground truth simulations for each of the strategies followed. Rows correspond to the specific parameters.

similarity measure showed values up to 0.8, as expected from the histograms illustrated in Figure 8. On the other hand, Strategy 2 had a comparable performance compared to Strategy 1 predicting AD maps, but got a higher % of absolute differences (>0.8) in RD maps. In spite of this, the angles' cosine similarity maps showed significantly better estimations, with values <0.5.

Lastly, moving on to the metrics presented in Table VII, it can be seen that all R^2 values for the three methods got a value >0.771. However, MAE metric did not reflect such good predictions, with values up to 0.214 in the worst map. Once more, the difference between not correcting and correcting for GNL was not visually perceptible due to the high errors coming from the ML model.

TABLE VII
METRIC RESULTS FROM **EXPERIMENT 4**.

Parameter	Method	R^2	MAE
S0	ML Strategy 1	0.923	0.025
	ML Strategy 2 no GNL correction	0.993	0.007
	ML Strategy 2 with GNL correction	0.996	0.005
AD	ML Strategy 1	0.935	0.092
	ML Strategy 2 no GNL correction	0.897	0.111
	ML Strategy 2 with GNL correction	0.855	0.131
RD	ML Strategy 1	0.912	0.118
	ML Strategy 2 no GNL correction	0.771	0.214
	ML Strategy 2 with GNL correction	0.781	0.209

IV. DISCUSSION

In this paper, the possibility of fitting a signal representation using ML networks was investigated. Moreover, we also studied the correction for GNL while predicting the parameter maps, and the feasibility of generalizing the proposed ML model strategies to train and test with multiple subjects. By accepting more than one subject during the training, the possibility of real-time mapping could become a reality, allowing us to have networks pretrained with several datasets ready to predict new patient maps in seconds.

A. Experiment 1

The first set of experiments of the project were based on a simulation study performed with different distributions

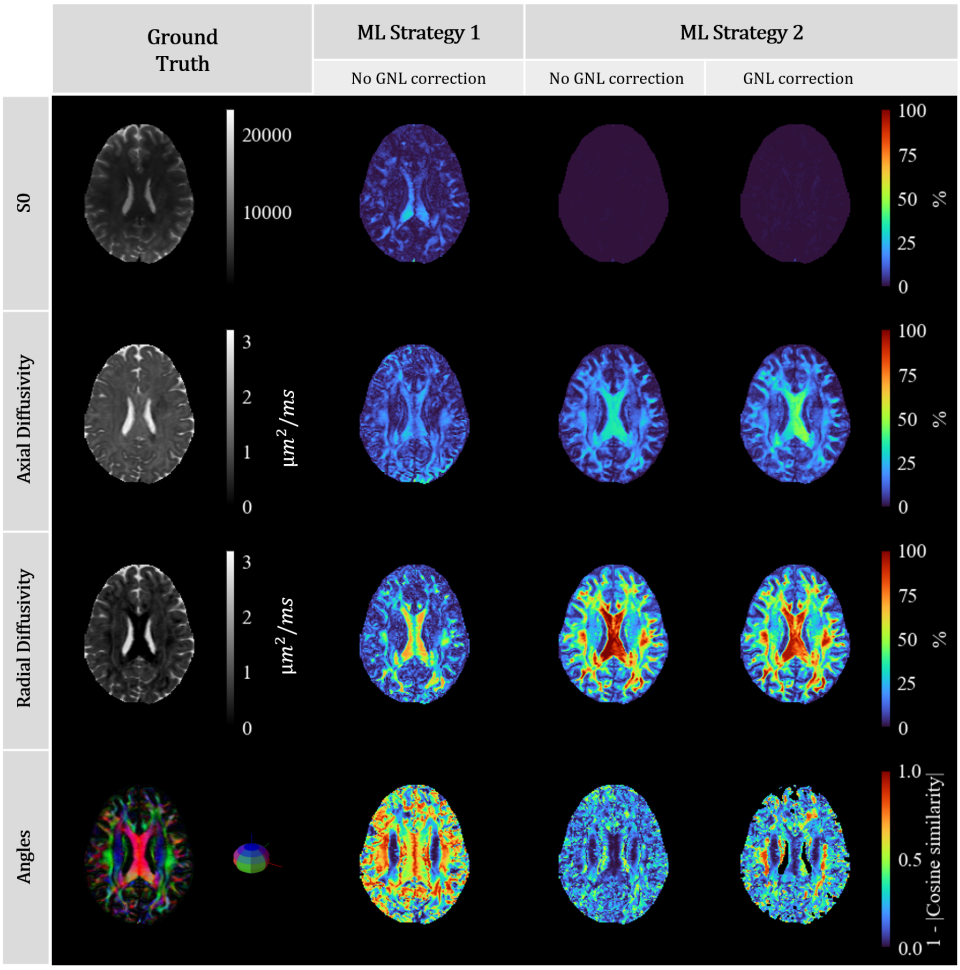

Fig. 10. Reconstruction results for **Experiment 4**. First column shows the ground truth parameter maps. The following columns show the % of absolute differences between predicted and ground truth simulations for each of the strategies followed. Rows correspond to the specific parameters.

of parameters to test the approaches with simpler data. As mentioned in Section II-C, diffusion MRI data is inherently complex, containing information about the movement of water molecules inside the brain. By simplifying parameter distributions, we could study in a more controlled way the behaviour of our proposed setups.

For ML Strategy 1, two different input distributions were tested, uniform and von Mises distributions. The experiments tested on uniform distributions exposed the main limitation of this strategy, which was the spherical behavior of polar angles (Figure 4). Von Mises distributions were implemented, however, results revealed that changing input distributions did not effectively solve the problem. Since the MSE loss was calculated between input and predicted parameters, the model handled in the same way the linear relationship of the S_0, AD and RD, and the spherical relationship of the angles θ and ϕ. The MSE loss did not take into account periodicity, where angles that differ by 2π radians should be interpreted as the same, or positional errors in the sphere. This problem was further addressed changing the angles themselves. Considering that the input distribution of angles were in polar coordinates, two main transformations were examined. A sine transformation was implemented to handle the problem of high positional errors due to the spherical distribution. As the sine function is periodic with a period of 2π it could potentially solve the problem described earlier. However, this did not completely unravel the issue, as these parameters also had orientation. For example, a vector that is at position 0 and other that is at position π have the same direction but different orientations. Moreover, the possibility of transforming from polar to Cartesian coordinates was also explored, but results were non conclusive either. Following all the previous reasoning, synthetic simulation studies concluded that Strategy 1 can give accurate predictions for parameters following linear relationships, but fail to account for spherical ones.

Using synthetic simulations, NLLS performed as expected from literature [17][31], but Strategy 2 did not give such good results. A reason for this could be the non realistic scattering of the synthetic uniformly distributed parameters, which is why the next step was to perform simulations based on *in vivo* data simulations calculated from HCP subjects.

B. Experiment 2

Training and testing with one unique subject proved to work in all the methods proposed, as expected from previous studies [18][19]. NLLS confirmed its superior performance for fitting the Zeppelin model and for correcting for GNL. However, the main disadvantage of the computational time still persisted. On the same line, ML demonstrated an acceptable performance for ML Strategy 1 and Strategy 2 (leaving out the angular parameters for Strategy 1 as previously discussed en Experiment 1). GNL correction was also correctly applied, although the metrics did not reflect the improvement because of the high error values coming from the ML predictions compared to the small values related to the correction itself. The correction was visually confirmed by reconstructing and computing the absolute difference between the corresponding signals from predicted parameters without and with GNL correction.

C. Experiment 3

As the goal was to generalize for more subjects, Experiment 3 was the main focus for optimizing the model's performance. Training with multiple subjects and testing with data from the same subjects was a valid intermediate step which permitted a gradual increase in the complexity of the model, while at the same time maintaining the familiarity of the test distribution. Moreover, it also allowed to experiment in a more controlled way its performance, minimizing the possible negative influence of introducing new data without first optimizing the network. The focus of the hyperparameter tuning was set in the learning rate, dropout rate, network architecture, and input signal normalization.

The most optimal learning rate was found to be 0.0001. By using a small constant learning rate that constrained the model to take small steps during the training, we effectively encouraged the network to avoid abrupt oscillations and prevent overfitting. Results showed that, in our specific network, applying learning rate schedulers with gradual learning rate reduction leaded to sub-optimal training with slow convergence and sharp loss changes, not reaching appropriately to the optimal solution. Additionally, no dropout proved to give the best results. Considering our simple model with only one hidden layer, experiments demonstrated that adding a dropout rate only introduced unwanted uncertainty to the training process, and did not improve the network's performance. Next, based on literature [18], different network architectures were tested. The analysis of the results after comparing networks 108-56-5, 108-108-5, 108-108-108-5 yielded to the following interpretations. The addition of hidden layers and nodes allowed the model to capture more complex patterns in the data, increasing its capability to extract deeper and more abstract features. In our experimental setup, because most of the relationships between parameters were linear, and only 5 parameters were predicted, the number of hidden layers and nodes that gave better predictions was 108-56-5. To finish with the parameter optimization, two alternative methods were studied to normalize the input signals: pixel-wise and global dataset normalization. Results confirmed that pixel-wise normalization was the most effective since scaling the intensity values individually preserved the relative contrast intensity of the pixels within the image. By dividing each dataset by the maximum b=0 intensity value, the standardization was too strict and the individual intensity characteristics of the images were lost.

The results obtained using the optimized parameters in this experiment were not optimal, specially for ML Strategy 1 which failed to accurately predict AD, RD, and the angles. Strategy 2 performed slightly better when predicting AD and the angles, but failed to predict RD maps too. One reason that could explain this behavior is linked to the first limitation of the proposed network, which is the definition of the input training signals. To generalize the model and enable it the accept multiple subjects, we defined a data class including signals together with their corresponding b-values and gradient directions. While the signals were given as an input to the model's first layer, the b-values and gradient directions were not shown to the network. They were reserved and input directly in the forward-simulation of the signal from the predicted parameters. This made the network unaware of the situation of having different sets of subjects in the training set, and unable to differentiate between them without the posterior reconstruction of the signal. The proposed method worked better in Strategy 2, because for that strategy the b-values and gradient directions were not fully excluded from the training loop. Their influence was taken into account in the MSE loss calculated between the signals.

D. Experiment 4

Lastly, Experiment 4 was conducted to test the network's ability to generalize to new subjects. Although Experiment 3 did not generate optimal results, this last experiment was needed to analyse and understand the predictions more in depth. Moreover, the application of real-time mapping needed to be examined, involving testing with completely new different subjects. In this experiment, a new approach for input signal normalization was considered as pixel-wise normalization did not generate adequate predictions. An individual scaling factor was calculated for each dataset based on their median values, followed by the division by the maximum signal intensity of the reference dataset. With this new approach, we were able to normalize the training signals while accounting for differences in intensity distributions between subjects, and preserving the relative intensity within each dataset.

The ML predictions for this last experiment were similar to the ones in Experiment 3. However, we must highlight a considerable improvement in ML Strategy 1 maps. As discussed in the previous experiment, Strategy 1 was unaware of the different b-values and gradient directions during the whole training process. In Experiment 3, pixel-wise normalization did not consider normalization between subjects, leading to wrong map outputs. In this final experiment, the scaling factor

calculation helped the model to generalize better by addressing the variability before the training, taking as an input signals that were already normalized across subjects.

E. Limitations and future work

Overall, the project had some limitations which could be solved in future experiments. Firstly, RD was theoretically defined based on AD (*RD = k*AD*), explaining why RD maps were systematically worse in every experiment. This definition directly affected RD maps, as every source of error coming from predicted AD maps was propagated to RD, and added to its own set of errors. Independent variables for AD and RD inside the training loop could be considered in new experiments to avoid error propagation between predictions, always constraining RD<= AD. Moreover, hyperparameter tuning experiments were not very extensive, covering only a few values and three different network architectures. In the future, a more exhaustive study could be executed using approaches such as grid search or random search [37] to consider more tests, and find more optimal parameters. In relation with the data class and the input signals definition drawback, some changes could also be implemented. Instead of keeping the b-values and gradient directions aside until the forward-simulation of the predicted signals, they could also be input to the network. This way, the ML network could generalize better to multiple subjects using all the available information that was not directly explored in our approach. Also, it must be highlighted that the model chosen to fit the HCP data was very simple, fitting only 5 parameters. NLLS optimization has proved to fit remarkably good simple models with a few number of parameters. However, as the model's complexity increases, the optimization problem becomes more challenging. Studies have shown higher performance of ML networks compared to traditional optimization techniques in more complicated models such as the White Matter model [39]. As a future step, other biophysical models could be considered and added to the methodology proposed. Lastly, the presented methods were only tested on healthy control subjects taken from a public dataset. Once further investigation is conducted to solve the main issues in the proposed methodology, real data from patients suffering from pathologies such as Alzheimer's disease or multiple sclerosis could be used. In particular, one alternative could be the usage of data acquired with an ultra-strong gradient head insert at the Utrecht Medical Center Utrecht (UMCU).

Regarding the individual experiments performed in this project, although Experiment 3 served as a valid first generalization step, the focus for future research should be placed in Experiment 4. A more exhaustive hyperparameter tuning should be executed, more signal normalization techniques could be investigated, such as White Stripe, or Z-score normalization [40][41], and new implementations for including the b-values and b-vectors should be examined. Lastly, with respect to the ML network implemented, we must highlight the superior performance of ML Strategy 2 compared to Strategy 1. Self-supervised networks have proven to mitigate the influence of data distribution, specially when dealing with directional parameters (periodicity issue) [18], and have been successful in correcting for GNL during training.

V. Conclusion

This study proposed a ML approach for diffusion MRI parameter estimation. Images from the HCP were fitted using the Zeppelin diffusion tensor model by two ML strategies which were compared to traditional NLLS optimization. We investigated the possibility of real-time mapping by generalizing the model to accept multiple subjects during training, validation and testing. Additionally, the incorporation of GNL correction in the ML network was also explored. The results showed that ML approaches worked when training and testing with one single subject. However, predictions were not optimal when introducing multiple datasets. Nevertheless, S_0, AD and angular parameters revealed promising preliminary results, giving a valid starting point for future experiments. Moreover, GNL correction was properly implemented and included in the methodology, but the high errors require future thorough investigation before such networks could be used in e.g. population studies.

References

[1] Peter Stys. "White Matter Injury Mechanisms". In: *Current Molecular Medicine* 4 (2 Mar. 2004), pp. 113–130. ISSN: 15665240. DOI: 10.2174/1566524043479220.

[2] S Love. "Demyelinating diseases". In: *Journal of Clinical Pathology* 59 (11 May 2006), pp. 1151–1159. ISSN: 0021-9746. DOI: 10.1136/jcp.2005.031195.

[3] Robert-Jan M. van Geuns et al. "Basic principles of magnetic resonance imaging". In: *Progress in Cardiovascular Diseases* 42 (2 Sept. 1999), pp. 149–156. ISSN: 00330620. DOI: 10.1016/S0033-0620(99)70014-9.

[4] DW Paty et al. "MRI in the diagnosis of MS: a prospective study with comparison of clinical evaluation, evoked potentials, oligoclonal banding, and CT". In: *Neurology* 38.2 (1988), pp. 180–180.

[5] H. Jiang et al. "Early Diagnosis of Spastic Cerebral Palsy in Infants with Periventricular White Matter Injury Using Diffusion Tensor Imaging". In: *American Journal of Neuroradiology* 40 (1 Jan. 2019), pp. 162–168. ISSN: 0195-6108. DOI: 10.3174/ajnr.A5914.

[6] Benedetta Bodini and Olga Ciccarelli. *Diffusion MRI in Neurological Disorders*. Elsevier, 2014, pp. 241–255. DOI: 10.1016/B978-0-12-396460-1.00011-1.

[7] E. O. Stejskal and J. E. Tanner. "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient". In: *The Journal of Chemical Physics* 42 (1 Jan. 1965), pp. 288–292. ISSN: 0021-9606. DOI: 10.1063/1.1695690.

[8] H. Y. Carr and E. M. Purcell. "Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments". In: *Physical Review* 94 (3 May 1954), pp. 630–638. ISSN: 0031-899X. DOI: 10.1103/PhysRev.94.630.

[9] Jan Hrabe, Gurjinder Kaur, and DavidN Guilfoyle. "Principles and limitations of NMR diffusion measurements". In: *Journal of Medical Physics* 32 (1 2007), p. 34. ISSN: 0971-6203. DOI: 10.4103/0971-6203.31148.

[10] Michael A. Jacobs et al. "Diffusion-Weighted Imaging With Apparent Diffusion Coefficient Mapping and Spectroscopy in Prostate Cancer". In: *Topics in Magnetic Resonance Imaging* 19 (6 Dec. 2008), pp. 261–272. ISSN: 0899-3459. DOI: 10.1097/RMR.0b013e3181aa6b50.

[11] Peter J Basser, James Mattiello, and Denis LeBihan. "Estimation of the effective self-diffusion tensor from the NMR spin echo". In: *Journal of Magnetic Resonance, Series B* 103.3 (1994), pp. 247–254.

[12] J. Huang, R.P. Friedland, and A.P. Auchus. "Diffusion Tensor Imaging of Normal-Appearing White Matter in Mild Cognitive Impairment and Early Alzheimer Disease: Preliminary Evidence of Axonal Degeneration in the Temporal Lobe". In: *American Journal of Neuroradiology* 28 (10 Nov. 2007), pp. 1943–1948. ISSN: 0195-6108. DOI: 10.3174/ajnr.A0700.

[13] Eric C. Klawiter et al. "Radial diffusivity predicts demyelination in ex vivo multiple sclerosis spinal cords". In: *NeuroImage* 55 (4 Apr. 2011), pp. 1454–1460. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.01.007.

[14] Daniel C. Alexander et al. "Imaging brain microstructure with diffusion MRI: practicality and applications". In: *NMR in Biomedicine* 32 (4 Apr. 2019). ISSN: 0952-3480. DOI: 10.1002/nbm.3841.

[15] Chantal M.W. Tax et al. "What's new and what's next in diffusion MRI preprocessing". In: *NeuroImage* 249 (Apr. 2022), p. 118830. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2021.118830.

[16] Vladimir Golkov et al. "Q-space deep learning: twelve-fold shorter and model-free diffusion MRI scans". In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1344–1351.

[17] Cheng Guan Koay et al. "A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging". In: *Journal of Magnetic Resonance* 182 (1 Sept. 2006), pp. 115–125. ISSN: 10907807. DOI: 10.1016/j.jmr.2006.06.020.

[18] Jason P. Lim et al. "Fitting a Directional Microstructure Model to Diffusion-Relaxation MRI Data with Self-Supervised Machine Learning". In: (Oct. 2022).

[19] Francesco Grussu et al. "Deep learning model fitting for diffusion-relaxometry: a comparative study". In: *Computational Diffusion MRI: International MICCAI Workshop, Lima, Peru, October 2020*. Springer. 2021, pp. 159–172.

[20] Stefanie Eriksson et al. "NMR diffusion-encoding with axial symmetry and variable anisotropy: Distinguishing between prolate and oblate microscopic diffusion tensors with unknown orientation distribution". In: *The Journal of Chemical Physics* 142 (10 Mar. 2015). ISSN: 0021-9606. DOI: 10.1063/1.4913502.

[21] Daniel Topgaard. "Multidimensional diffusion MRI". In: *Journal of Magnetic Resonance* 275 (Feb. 2017), pp. 98–113. ISSN: 10907807. DOI: 10.1016/j.jmr.2016.12.007.

[22] Uran Ferizi et al. "A robust diffusion tensor model for clinical applications of MRI to cartilage". In: *Magnetic Resonance in Medicine* 79 (2 Feb. 2018), pp. 1157–1164. ISSN: 0740-3194. DOI: 10.1002/mrm.26702.

[23] Chantal M.W. Tax et al. "Measuring compartmental T2-orientational dependence in human brain white matter using a tiltable RF coil and diffusion-T2 correlation MRI". In: *NeuroImage* 236 (Aug. 2021), p. 117967. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2021.117967.

[24] Matthew F. Glasser et al. "The minimal preprocessing pipelines for the Human Connectome Project". In: *NeuroImage* 80 (Oct. 2013), pp. 105–124. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.04.127.

[25] R. Bammer et al. "Analysis and generalized correction of the effect of spatial gradient field distortions in diffusion-weighted imaging". In: *Magnetic Resonance in Medicine* 50 (3 Sept. 2003), pp. 560–569. ISSN: 0740-3194. DOI: 10.1002/mrm.10545.

[26] D.C. Van Essen et al. "The Human Connectome Project: A data acquisition perspective". In: *NeuroImage* 62 (4 Oct. 2012), pp. 2222–2231. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2012.02.018.

[27] K. Setsompop et al. "Pushing the limits of in vivo diffusion MRI for the Human Connectome Project". In: *NeuroImage* 80 (Oct. 2013), pp. 220–233. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.05.078.

[28] Hamed Y. Mesri et al. "The adverse effect of gradient nonlinearities on diffusion MRI: From voxels to group studies". In: *NeuroImage* 205 (Jan. 2020), p. 116127. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2019.116127.

[29] Stamatios N. Sotiropoulos et al. "Advances in diffusion MRI acquisition and processing in the Human Connectome Project". In: *NeuroImage* 80 (Oct. 2013), pp. 125–143. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.05.057.

[30] Lei Tang and Xiaohong Joe Zhou. "Diffusion MRI of cancer: From low to high b-values". In: *Journal of Magnetic Resonance Imaging* 49 (1 Jan. 2019), pp. 23–40. ISSN: 1053-1807. DOI: 10.1002/jmri.26293.

[31] Ileana O. Jelescu et al. "Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue". In: *NMR in Biomedicine* 29 (1 Jan. 2016), pp. 33–47. ISSN: 0952-3480. DOI: 10.1002/nbm.3450.

[32] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[33] Eleftherios Garyfallidis et al. "Dipy, a library for the analysis of diffusion MRI data". In: *Frontiers in Neuroinformatics* 8 (Feb. 2014). ISSN: 1662-5196. DOI: 10.3389/fninf.2014.00008.

[34] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[35] DJ Best and Nicholas I Fisher. "Efficient simulation of the von Mises distribution". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28.2 (1979), pp. 152–157.

[36] Kantilal Varichand Mardia. "Statistics of directional data". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 37.3 (1975), pp. 349–371.

[37] Li Yang and Abdallah Shami. "On hyperparameter optimization of machine learning algorithms: Theory and practice". In: *Neurocomputing* 415 (Nov. 2020), pp. 295–316. ISSN: 09252312. DOI: 10.1016/j.neucom.2020.07.061.

[38] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[39] João P. de Almeida Martins et al. "Neural networks for parameter estimation in microstructural MRI: Application to a diffusion-relaxation model of white matter". In: *NeuroImage* 244 (Dec. 2021), p. 118601. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2021.118601.

[40] Alexandre Carré et al. "Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics". In: *Scientific reports* 10.1 (2020), p. 12340.

[41] Alberto De Luca et al. "Multimodal tract-based MRI metrics outperform whole brain markers in determining cognitive impact of small vessel disease-related brain injury". In: *Brain Structure and Function* 227 (7 Sept. 2022), pp. 2553–2567. ISSN: 1863-2653. DOI: 10.1007/s00429-022-02546-2.

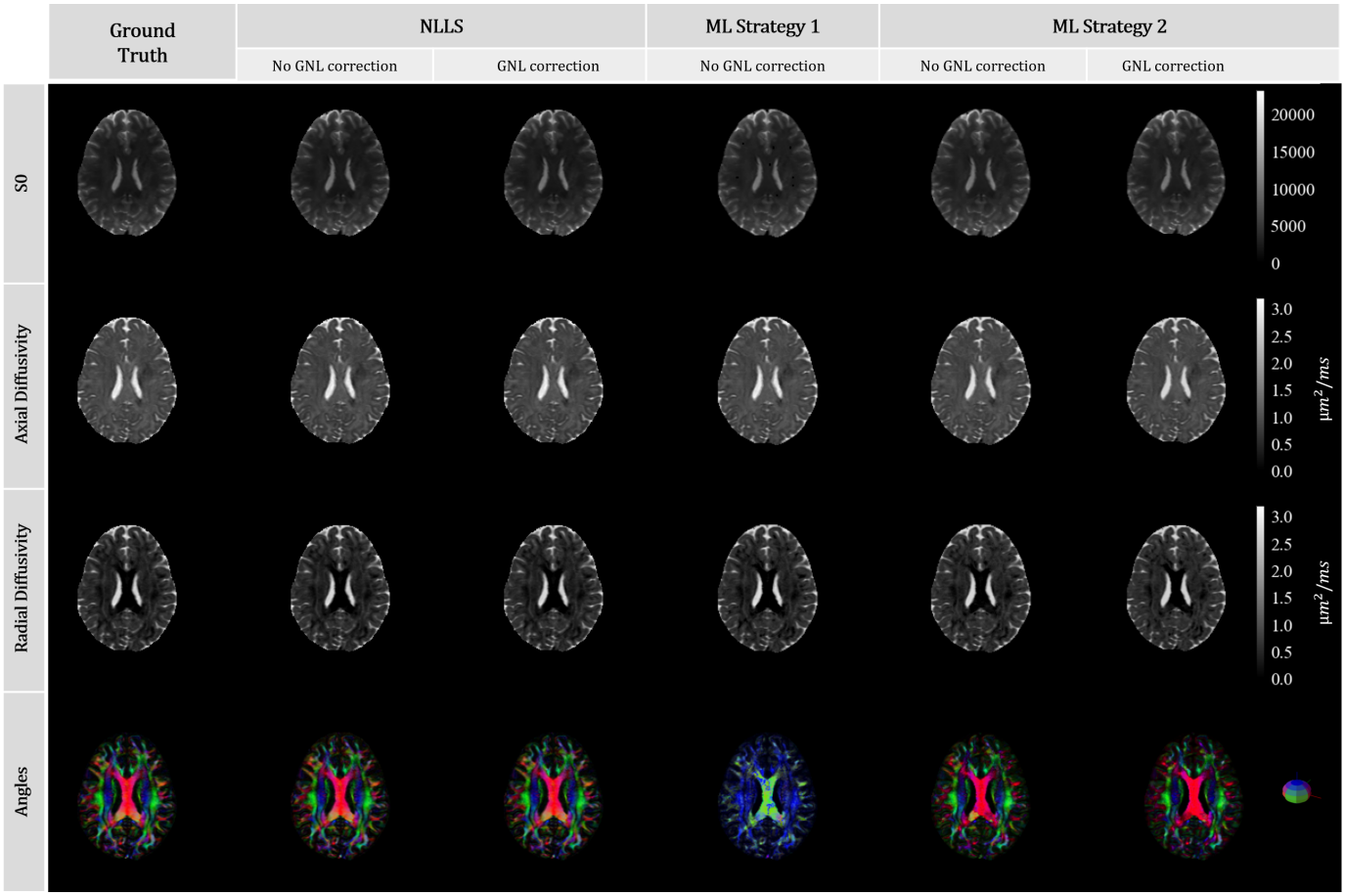

Fig. A.1. Reconstructed parameter maps from methods executed with **Experiment 2**. Methods presented in columns, and parameters in rows

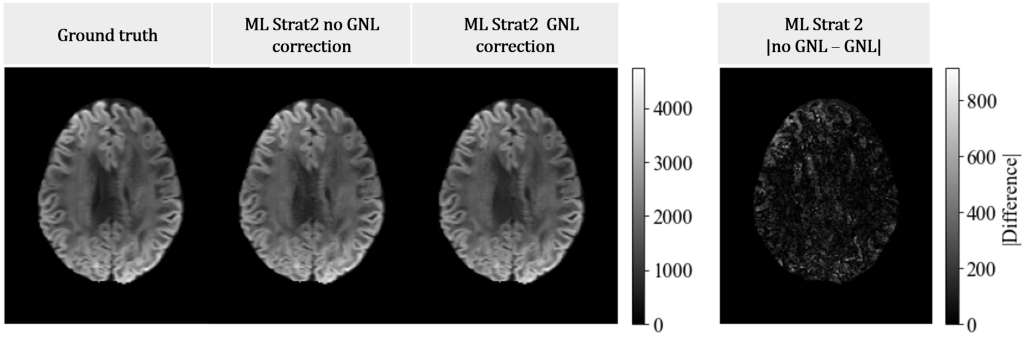

Fig. A.2. Absolute difference map showing GNL correction for Strategy 2 **Experiment 2** predictions. Signals reconstructed with parameters with and without GNL correction, and subtracted. Signals shown for b-value of 1.005 $\mu m^2/ms$

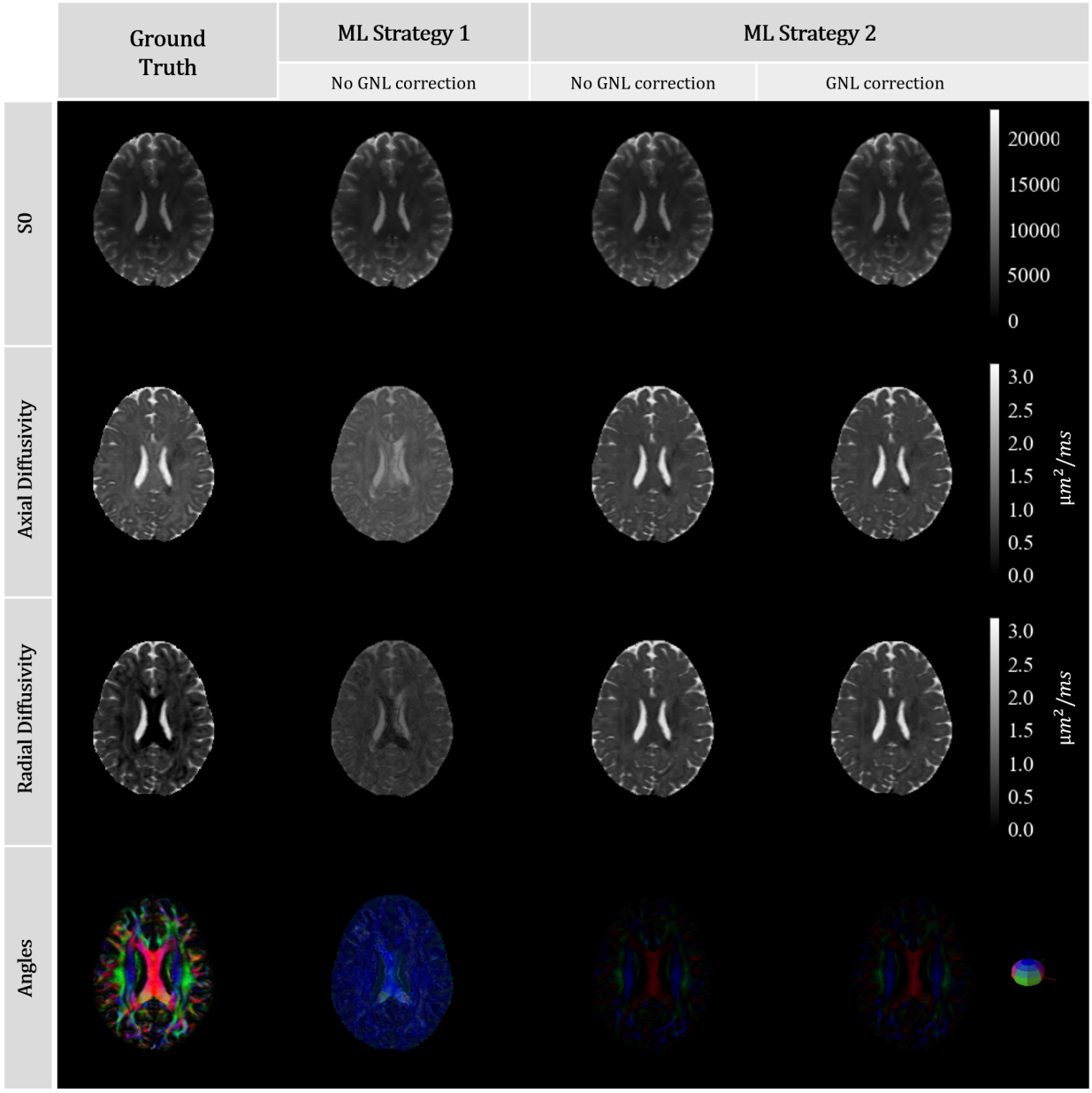

Fig. A.3. Reconstructed parameter maps from methods executed with **Experiment 3**. Methods presented in columns, and parameters in rows.

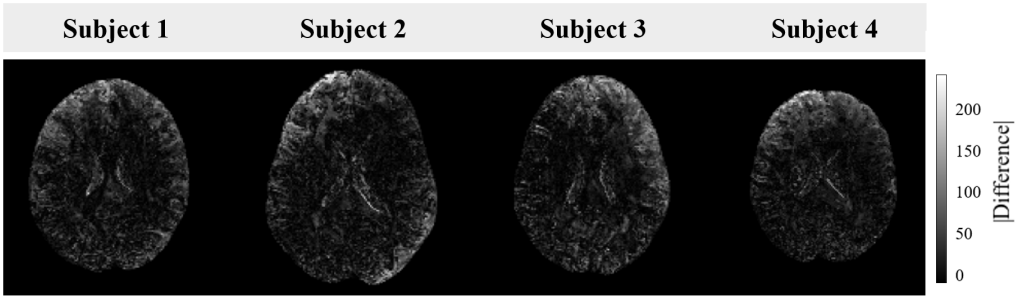

Fig. A.4. Absolute difference maps showing GNL correction for Strategy 2 **Experiment 3** predictions. Signals reconstructed with parameters with and without GNL correction, and subtracted. Middle slice of each dataset (the 4 slices comprising the test set) is shown for b-value of 1.005 μm^2/ms

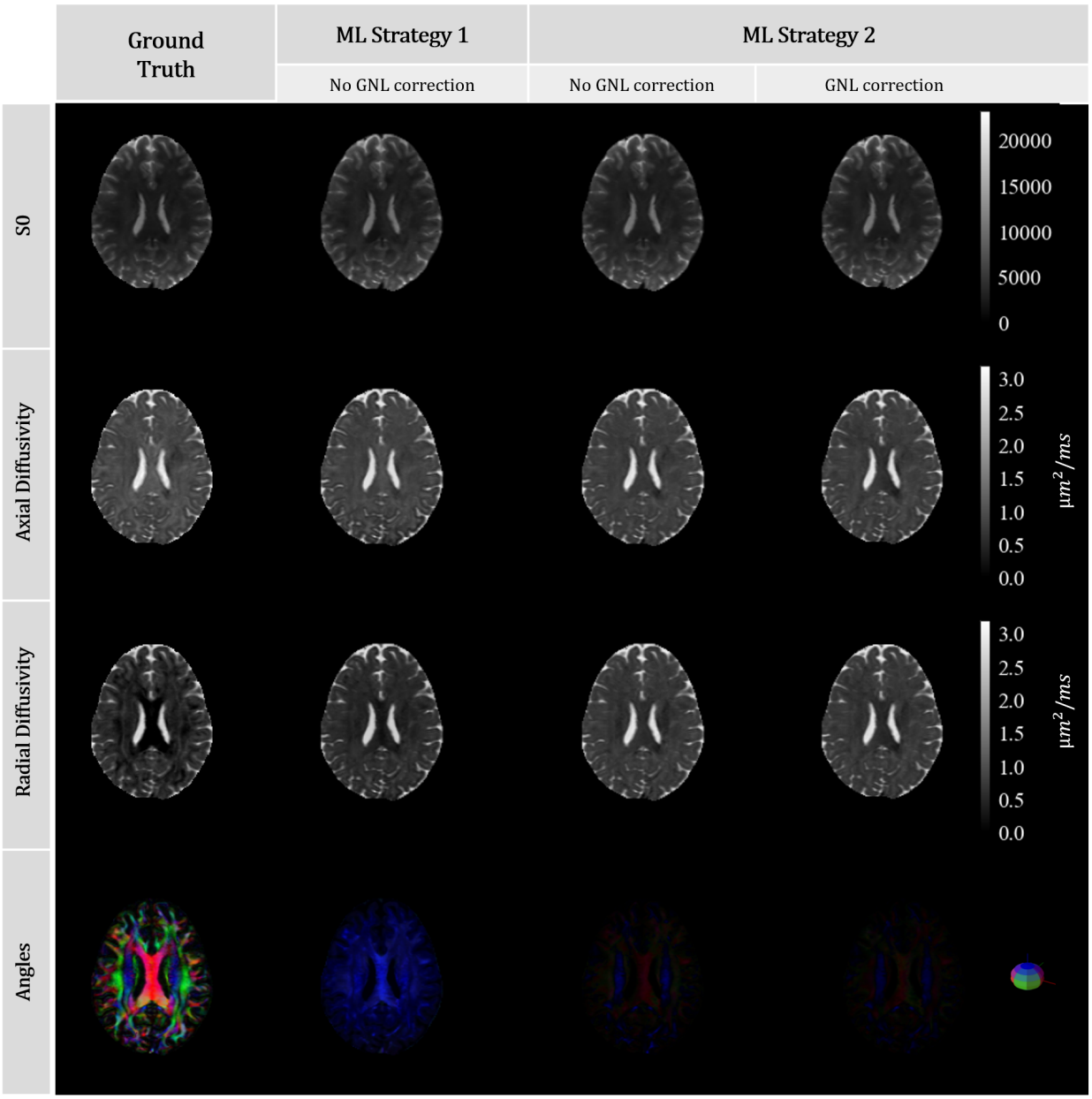

Fig. A.5. Reconstructed parameter maps from methods executed with **Experiment 4**. Methods presented in columns, and parameters in rows.

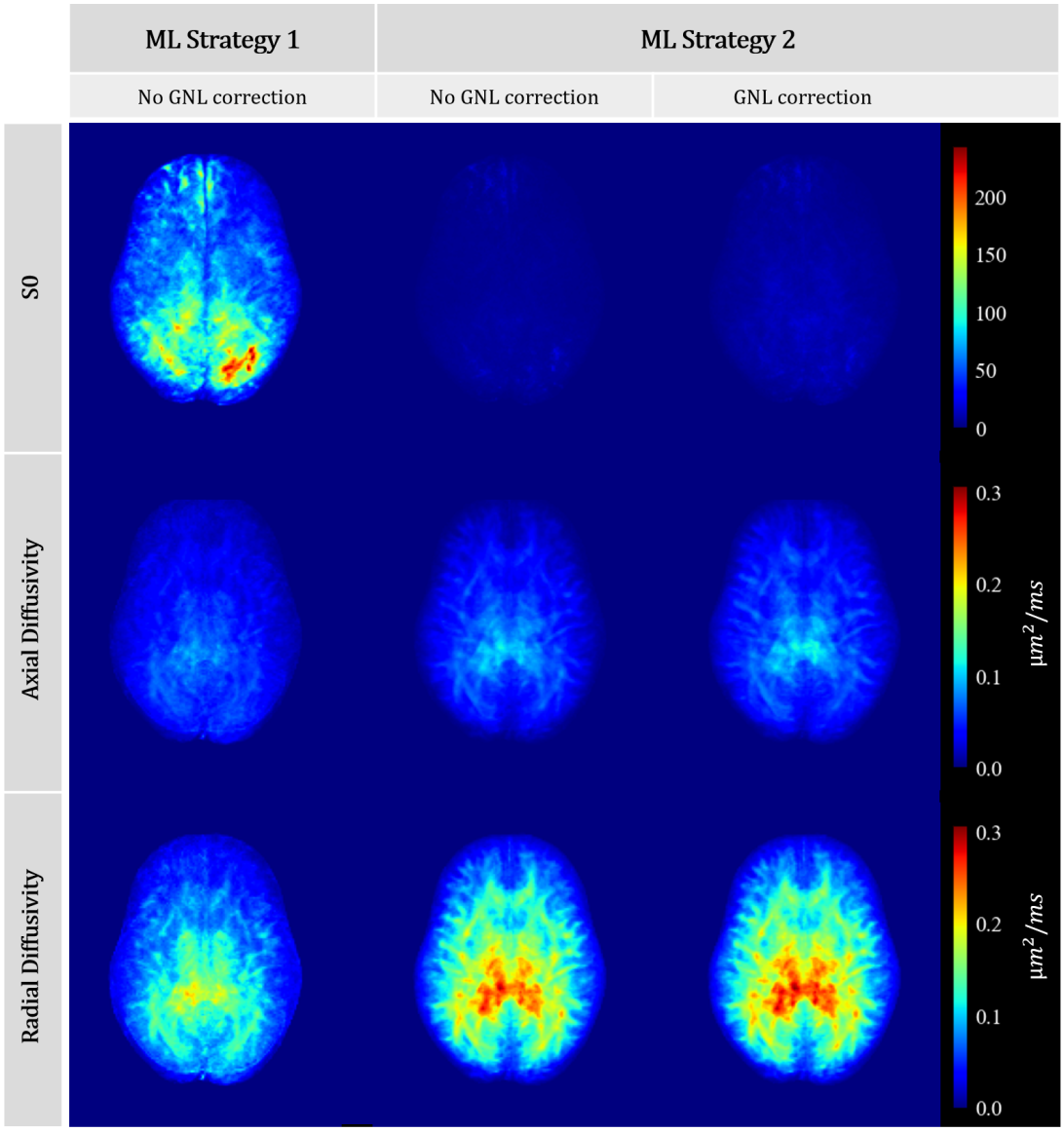

Fig. A.6. Average intensity projection of the absolute errors calculated between ground truth and predicted parameters from the test dataset following ML strategies in **Experiment 4**.