Neural Networks for parameter estimation in microstructural MRI: a study with a high-dimensional diffusion-relaxation model of white matter microstructure

João P. de Almeida Martins[1,2]*, Markus Nilsson[1]*, Björn Lampinen[3], Marco Palombo[4], Peter T. While[2,5], Carl-Fredrik Westin[6,7], Filip Szczepankiewicz[1,6,7]

[1] *Department of Clinical Sciences, Radiology, Lund University, Lund, Sweden*
[2] *Department of Radiology and Nuclear Medicine, St. Olav's University Hospital, Trondheim, Norway*
[3] *Department of Clinical Sciences, Medical Radiation Physics, Lund University, Lund, Sweden*
[4] *Centre for Medical Image Computing and Department of Computer Science, University College London, London, UK*
[5] *Department of Circulation and Medical Imaging, NTNU - Norwegian Institute of Technology, Trondheim, Norway*
[6] *Radiology, Brigham and Women's Hospital, Boston, MA, United States*
[7] *Harvard Medical School, Boston, MA, United States*

* these authors have contributed equally to this study
Correspondence to: João P. de Almeida Martins (joao101martins@gmail.com)

ABSTRACT

Specific features of white-matter microstructure can be investigated by using biophysical models to interpret relaxation-diffusion MRI brain data. Although more intricate models have the potential to reveal more details of the tissue, they also incur time-consuming parameter estimation that may converge to inaccurate solutions due to a prevalence of local minima in a degenerate fitting landscape. Machine-learning fitting algorithms have been proposed to accelerate the parameter estimation and increase the robustness of the attained estimates. So far, learning-based fitting approaches have been restricted to lower-dimensional microstructural models where dense sets of training data are easy to generate. Moreover, the degree to which machine learning can alleviate the degeneracy problem is poorly understood. For conventional least-squares solvers, it has been shown that degeneracy can be avoided by acquisition with optimized relaxation-diffusion-correlation protocols that include tensor-valued diffusion encoding; whether machine-learning techniques can offset these acquisition requirements remains to be tested. In this work, we employ deep neural networks to vastly accelerate the fitting of a recently introduced high-dimensional relaxation-diffusion model of tissue microstructure. We also develop strategies for assessing the accuracy and sensitivity of function fitting networks and use those strategies to explore the impact of acquisition protocol design on the performance of the network. The developed learning-based fitting pipelines were tested on relaxation-diffusion data acquired with optimized and sub-sampled acquisition protocols. We found no evidence that machine-learning algorithms can by themselves replace a careful design of the acquisition protocol or correct for a degenerate fitting landscape.

1

MAIN TEXT

1 INTRODUCTION

Microstructure imaging aims at using diffusion MRI (dMRI) to map salient features of the tissue (Alexander et al., 2019; Nilsson et al., 2013; Novikov et al., 2019). A central goal in microstructure imaging has been to estimate volume fraction of different microstructure components (Lampinen et al., 2020; Lampinen et al., 2019; Veraart et al., 2018). Estimating volume rather than signal fractions is however challenging because it requires the simultaneous estimation of both diffusion and relaxation properties of the different compartments. This kind of inverse problem is sensitive to degeneracy issues (Jelescu et al., 2016; Lampinen et al., 2019), in which a multitude of different model parameters can describe the acquired data equally well. Parameter estimation can also be computationally slow, preventing real-time mapping. A potential solution is to employ machine learning to accelerate the parameter estimation process. However, the current literature lacks systematic descriptions of the gains and potential drawbacks of this approach, which is surprising considering the exponential increase in interest for such methods. In this work, we use neural networks to speed up the estimation process and investigate the veracity of the estimates as well as the potential for neural networks to alleviate problems that stem from degeneracy.

Neural networks and other machine learning approaches have been applied before to accelerate the estimation of microstructure parameters from dMRI data (Barbieri et al., 2020; Bertleff et al., 2017; Golkov et al., 2016; Grussu et al., 2020a; Gyori et al., 2019; Hill et al., 2021; Nedjati-Gilani et al., 2017; Palombo et al., 2020; Reisert et al., 2017). Examples include the use of a random forest regressor to compartment models with permeability for white matter microstructure imaging in presence of water exchange (Nedjati-Gilani et al., 2017) and the SANDI model to map gray matter properties (Palombo et al., 2020). Reisert *et al.* applied machine learning to a Bayesian estimation approach which dramatically accelerated the fitting of two- and three-compartment models (Reisert et al., 2017). Barbieri *et al* applied deep neural networks to the intra-voxel incoherent motion model (Barbieri et al., 2020). Nevertheless, an open question is what impact the training strategy has on the fitting performance, in particular when applied to models with many model parameters. Here, we will loosely refer to these as 'high-dimensional models.' For such models, generation of training data is challenging due to the poor scaling behaviour when a finite number of points are distributed across p parameter dimensions; to sample each combination of model parameters in m steps requires m^p samples. As p increases, it is unavoidable that a finite set of samples becomes sparse in the p-dimensional space. Here, we investigate the impact that the model parameter space sampling pattern has on the performance of the neural network.

Apart from accelerating model fitting, neural networks may in principle also reduce the requirements on the imaging protocol by exploiting parameter correlations. For example, priors learned from the training data have been observed to stabilise model fitting performance against substantial degrees of data down-sampling (Alexander et al., 2017; Golkov et al., 2016; Tian et al., 2020). However, we do not expect machine learning approaches to completely alleviate degeneracy issues. Indeed, for cases where the acquisition protocol does not provide sufficient information to resolve between different parameter values, the learning-based estimates will simply equal the mean of the model parameter distribution used for training (Reisert et al., 2017).

2

The aim of this study was to compare training strategies, propose tools to evaluate the performance of the neural network, and test to what degree neural networks could help solve the degeneracy problem. As a testbed, we use a high-dimensional relaxation-diffusion microstructure (Lampinen et al., 2020; Lampinen et al., 2019; Veraart et al., 2018). The parameter estimation was enabled by the use of state-of-the-art imaging protocols featuring so-called b-tensor encoding (Topgaard, 2017; Westin et al., 2016), combined with diffusion-relaxation correlations (de Almeida Martins et al., 2020; de Almeida Martins and Topgaard, 2018; Lampinen et al., 2019). We also investigated if neural network-based estimation of model parameters could offset the need for tensor-valued diffusion encoding, to enable this approach for data acquired with conventional diffusion encoding.

2 THEORY

White matter (WM) microstructure can be modelled by multiple compartments with different microstructural properties but a common orientation distribution (Alexander et al., 2019; Novikov et al., 2019). In this description, the measured signal is the convolution between an orientation distribution function (ODF) $P(\hat{n})$ and a microstructural kernel $K(\hat{u} \cdot \hat{n})$

$$S(\hat{u}) = \int_{|\hat{n}|=1} P(\hat{n})K(\hat{u} \cdot \hat{n}) \, \mathrm{d}\hat{n} \,, \tag{1}$$

where $\hat{n}$ and $\hat{u}$ are unit vectors defining the symmetry axes of the ODF and of the diffusion encoding process, respectively. In this work, we assign an effective transverse relaxation time T_2 and an apparent microscopic diffusion tensor $\mathbf{D}$ to each microstructural component, and use exponentially decaying functions to model the effect of these microstructural properties on the relaxation-diffusion-weighted signal (Veraart et al., 2018). Under these assumptions, the microstructure kernel is written as a weighted sum of exponentials

$$K(\hat{u} \cdot \hat{n}) = S_0 \, \sum_{j=1}^{J} f_j \exp\left(-\mathbf{B}(\hat{u}) : \mathbf{D}_j(\hat{n})\right) \exp\left(-\frac{\tau_{\mathrm{E}}}{T_{2;j}}\right) \,, \tag{2}$$

corresponding to a mixture of J components each with signal fraction f_j, transverse relaxation time $T_{2;j}$, and diffusion tensor $\mathbf{D}_j$. Information about $T_{2;j}$ and $\mathbf{D}_j$ is encoded into the signal by the echo-time τ_{E} and diffusion encoding tensor $\mathbf{B}(\hat{u})$, respectively, both of which are experimental variables. To simplify the model, we only consider axisymmetric $\mathbf{B}(\hat{u})$ and additionally assume that the component-wise $\mathbf{D}_j$ are axially symmetric.

The convolution expressed in Eq. (1) can be simplified by factorizing both $P(\hat{n})$ and $K(\hat{u} \cdot \hat{n})$ in their spherical harmonic coefficients p_{lm} and k_{lm}, respectively:

$$P(\hat{u}) = \sum_l \sum_m p_{lm} Y_{lm}(\hat{u}) \,, \tag{3}$$

and

$$K(\hat{u} \cdot \hat{n}) = \sum_{l'} k_{l'0} Y_{l'0}(\hat{u} \cdot \hat{n}) \,, \tag{4}$$

where Y_{lm} are the spherical harmonics basis functions

3

$$Y_{lm}(\Theta, \Phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\, L_l^m(\cos\Theta)\exp(im\Phi)\,, \tag{5}$$

with the $L_l^m(x)$ term denoting the associated Legendre polynomials. The summations in Eqs. (3) are carried out for order $l = 0, 1, 2, \ldots$, and degree $m = -l, -l+1, \ldots, l$. In Eq. (4), we have taken the axial symmetry of the microstructural kernel $K(\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{n}})$ into account (Lampinen et al., 2020; Novikov et al., 2018). Symmetry around the polar axis implies $k_{l'm'} = 0$ for either $m' \neq 0$ or odd l'. Taken together, this means that the $k_{l'm'}$ coefficients are reduced to their 0th degree terms $k_{l'0}$ (typically written as $k_{l'}$) and only even-ordered spherical harmonic terms ($l' = 0, 2, \ldots$) provide non-trivial contributions. Using the spherical harmonics addition theorem, Eq. (4) can be rewritten as:

$$K(\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{n}}) = \sum_{l'} k_{l'0} \sum_{m'=-l'}^{l'} Y_{l'm'}(\hat{\boldsymbol{u}})\overline{Y}_{l'm'}(\hat{\boldsymbol{n}})\sqrt{\frac{4\pi}{2l'+1}}\,. \tag{6}$$

Inserting Eqs. (3) and (6) into Eq. (1) and making use of the orthogonality of the spherical harmonic basis finally yields (Driscoll and Healy, 1994; Healy et al., 1998):

$$S(\hat{\boldsymbol{u}}) = \sum_l \sum_m k_{l0} p_{lm} Y_{lm}(\hat{\boldsymbol{u}})\sqrt{\frac{4\pi}{2l+1}}\,, \tag{7}$$

where $\hat{\boldsymbol{u}}$ can be parameterized by the polar and azimuthal angles, θ and ϕ, describing the orientation of $\mathbf{B}$, $\hat{\boldsymbol{u}} \equiv (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$.

Exploiting the orthogonality of the spherical harmonic basis, the spherical harmonic coefficients of the ODF (p_{lm}) and the microstructure kernel (k_{l0}) can be determined by multiplying either $P(\hat{\boldsymbol{n}})$ or $K(\hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{n}})$, respectively, with the complex conjugate of Y_{lm} and then integrating over a sphere. For the microstructural kernel, such procedure results in:

$$k_{l0} \equiv k_l = S_0 \sum_{j=1}^J f_j \sqrt{4\pi(2l+1)}\,\mathrm{I}_{lj} \exp\left(-bD_{\mathrm{I};j}\left(1 - b_\Delta D_{\Delta;j}\right)\right)\exp\left(-\frac{\tau_\mathrm{E}}{T_{2;j}}\right)\,, \tag{8}$$

where b is the conventional b-value and b_Δ denotes the normalized anisotropy of the diffusion encoding tensor $\mathbf{B}$ (Eriksson et al., 2015). The isotropic diffusivity and the normalized diffusion anisotropy (D_I and D_Δ) are related to the axial and radial diffusivities ($D_\parallel$ and $D_\perp$) of the diffusion tensor according to $D_\mathrm{I} = (D_\parallel + 2D_\perp)/3$ and $D_\Delta = (D_\parallel - D_\perp)/3D_\mathrm{I}$ (Conturo et al., 1996). The I_{lj} factors are a function of the regular Legendre polynomials, L_l, and defined as:

$$I_{lj} = \int_0^1 \exp\left(-\alpha_j x^2\right)\cdot L_l(x)\,\mathrm{d}x\,, \tag{9}$$

with $\alpha_j = 3bD_{\mathrm{I};j}b_\Delta D_{\Delta;j}$.

Different diffusion MRI models vary in their number of components and the constraints imposed on their properties. Here we consider a two-compartment model ($J = 2$) comprising a "stick" component (S) with zero radial diffusivity and a "zeppelin" (Z) component with $D_{\Delta;\mathrm{Z}} \neq 1$:

4

$$S(\boldsymbol{e}, \boldsymbol{m}) = S_0 \Big[\, f_{\mathrm{S}} \exp\Big(-bD_{\mathrm{I;S}}(1 - b_\Delta)\Big)$$
$$\times \Big(\mathrm{I}_{0;\mathrm{S}} + 4\pi \mathrm{I}_{2;\mathrm{S}} \sum_m p_{2m} Y_{2m}(\theta, \phi) \Big) \exp\Big(-\frac{\tau_{\mathrm{E}}}{T_{2;\mathrm{S}}}\Big) \qquad (10)$$
$$+ (1 - f_{\mathrm{S}}) \exp\Big(-bD_{\mathrm{I;Z}}(1 - b_\Delta D_{\Delta;\mathrm{Z}})\Big)$$
$$\times \Big(\mathrm{I}_{0;\mathrm{Z}} + 4\pi \mathrm{I}_{2;\mathrm{Z}} \sum_m p_{2m} Y_{2m}(\theta, \phi) \Big) \exp\Big(-\frac{\tau_{\mathrm{E}}}{T_{2;\mathrm{Z}}}\Big) \Big]$$

where the spherical harmonic summation is truncated at the second-ordered, $m \in \{-2, -1, 0, 1, 2\}$. The derivation of Eq. (10) uses the $p_{00} = Y_{00} = 1/\sqrt{4\pi}$ ODF normalization (Lampinen et al., 2020; Novikov et al., 2018). The vectors $\boldsymbol{e}$ and $\boldsymbol{m}$ capture the experiment-related parameters, $\boldsymbol{e} = (\tau_{\mathrm{E}}, b, b_\Delta, \theta, \phi)$, and scalar model parameters, $\boldsymbol{m} = (f_{\mathrm{S}}, D_{\mathrm{I;S}}, D_{\mathrm{I;Z}}, D_{\Delta;\mathrm{Z}}, T_{2;\mathrm{S}}, T_{2;\mathrm{Z}}, p_{20}, \mathrm{Re}(p_{21}), \mathrm{Im}(p_{21}), \mathrm{Re}(p_{22}), \mathrm{Im}(p_{22}))$. Besides setting $D_{\Delta;\mathrm{S}} = 1$, no other constraints were imposed on the compartment properties. We refer to the model expressed by Eq. (10) as the Relaxed Standard Model (RSM). This name is chosen to mark its descendance from the "standard model" of WM microstructure (Novikov et al., 2019) and to emphasize the fact that it accounts for compartment-specific T_2 times.

The RSM model parameters can be determined by fitting Eq. (10) directly to the acquired signals (Lampinen et al., 2020). An alternative strategy, followed in (Veraart et al., 2018), is to use a model fitting framework that effectively reduces the dimensionality of the parameter space by means of performing a hierarchical factorization of the voxel-wise ODFs (Novikov et al., 2018; Reisert et al., 2017). The initial step of such framework consists in projecting the measured signal onto a spherical harmonics' basis

$$S(\hat{\boldsymbol{u}}) = \sum_l \sum_m S_{lm} Y_{lm}(\hat{\boldsymbol{u}}) \,. \qquad (11)$$

The S_{lm} coefficients are subsequently converted to rotational invariants S_l, and fitted to the corresponding rotationally invariant terms of the $P(\hat{\boldsymbol{u}}) \otimes K(\hat{\boldsymbol{g}} \cdot \hat{\boldsymbol{u}})$ convolution:

$$S_l = p_l k_l \,, \qquad (12)$$

where k_l is the 0th degree term of the microstructural kernel as defined by Eq. (8). The rotationally invariant coefficients, S_l and p_l, are computed from (Novikov et al., 2018)

$$x_l = \sqrt{\frac{4\pi \sum_m |x_{lm}|^2}{(2l+1)}} \,, \qquad (13)$$

where x_{lm} are the spherical harmonic coefficients, and $x_l \equiv S_l$ or $x_l \equiv p_l$. Signal projections with $l > 2$ have small contributions to the measured signal (Jespersen et al., 2007), and the sum in Eq. (11) is typically truncated at the second order term ($l = 2$). The fitting framework summarized by Eqs. (11) and (12) is commonly referred to as the "RotInv" approach due to its use of rotational invariants. The $l = 2$ RotInv approach condenses the five $p_{2m, \, m \in \{-2,-1,0,1,2\}}$ parameters of the RSM model onto a single p_2 invariant capturing the orientation coherence of the sub-voxel diffusion domains, thus reducing the dimensionality of the fitting problem by four parameters.

3 METHODS

3.1 Neural network architecture and training

In this work, we constructed a feedforward deep neural network (DNN) using the *fitnet* function in MATLAB (The MathWorks, Inc.), and used it to fit vectors of scalar parameters, $\boldsymbol{m} = (f_S, D_{I;S}, D_{I;Z}, D_{\Delta;Z}, T_{2;S}, T_{2;Z}, p_{20}, \mathrm{Re}(p_{21}), \mathrm{Im}(p_{21}), \mathrm{Re}(p_{22}), \mathrm{Im}(p_{22}))$, to sets of $S(\tau_E, \boldsymbol{B})$ measurements. The network was configured to have 3 fully connected hidden layers with decreasing number of nodes (180, 80, and 55) and an output layer with 11 nodes, each of which representing a dimension within the $\boldsymbol{m}$ vectors. All hidden layers were activated by hyperbolic tangent (tanh) functions, while the output layer uses a linear activation function. The input consisted of a vector of E signal amplitudes sampled with a pre-defined relaxation-diffusion encoding protocol. We considered three different acquisition protocols comprising between $E = 164$ and $E = 270$ distinct $(\tau_E, \boldsymbol{B})$ points. To remove the influence of S_0 from the fitting problem, we normalized the input vector to the median of the signals measured at the point of maximal signal amplitude (minimum b and shortest τ_E).

The choice of a fully connected DNN follows the design of classic multilayer perceptrons (MLPs), which are thus well-suited for regression problems (Cybenko, 1989; Hornik et al., 1989). In this work, we employ the tanh activation function due its stronger gradients and faster convergence (LeCun et al., 2012).

Supervised network training was performed using a scaled conjugate gradient optimiser and a mean squared error loss

$$\mathrm{MSE} = \left\| \boldsymbol{m}_{\mathrm{targ}} - \boldsymbol{m}_{\mathrm{net}} \right\|_2^2 \tag{14}$$

where $\boldsymbol{m}_{\mathrm{targ}}$ is the ground-truth target vector, $\boldsymbol{m}_{\mathrm{net}}$ is the corresponding network output vector, and $\| \cdot \|_2$ denotes the Euclidean norm. The $\boldsymbol{m}_{\mathrm{targ}}$ parameters were rescaled between 0 and 1 using a min-max normalization strategy before being supplied to the network. The network was trained with a set of $5 \cdot 10^5$ voxels with randomly generated model parameters and noisy signal $S(\tau_E, \boldsymbol{B})$ (section 3.2 describes the training dataset generation). The training data was divided into different sub-sets before being supplied to the network such that 65% of the original data was used to update the weights and biases, 20% was used for cross-validation, and 15% was reserved for testing. In lieu of standard ℓ_1 or ℓ_2 regularizers, we prevented overfitting through an early stopping method and training was terminated following an increase of the MSE of the validation data for 5 consecutive epochs.

Network GPU training took approximately 3 hours on two parallel NVIDIA GeForce RTX 2080 SUPER graphic cards, each with 8 GB of memory. Both graphic cards were installed on a high-end consumer-grade desktop computer with an Intel i9-9900k 3.6 GHz CPU and 32 GB memory.

3.2 Generating training data

We generated training parameter vectors, $\boldsymbol{m}_{\mathrm{train}}$, from two distinct sets:
- parameter vectors obtained by uniform random sampling within the bounds described in **Table 1**, denoted $\boldsymbol{m}_{\mathrm{unif}}$
- parameter vectors estimated from a NLLS fit of Eq. (10) to *in vivo* brain data, denoted $\boldsymbol{m}_{\mathrm{brain}}$.

Vectors derived from both sets were combined to create composite training datasets comprising a total of $\#m_{\mathrm{train}}$ vectors:

$$\#m_{\mathrm{train}} = \#m_{\mathrm{brain}} + \#m_{\mathrm{unif}}, \tag{15}$$

where $\#m_{\text{brain}}$ and $\#m_{\text{unif}}$ are the number of m_{brain} and m_{unif} vectors, respectively. To study the impact of different training data generation strategies on the accuracy of network-derived estimates, we compared the performance of networks trained with different ratios between $\#m_{\text{brain}}$ and $\#m_{\text{train}}$:

$$f_{\text{brain}} = \#m_{\text{brain}}/\#m_{\text{train}}, \tag{16}$$

i.e., networks trained with varying relative amounts of m_{brain} and m_{unif} vectors. The f_{brain} fractions were varied between 0 and 1 in increments of 0.05, thus resulting in a total of 21 training datasets. All sets contained a total of $\#m_{\text{train}} = 5 \cdot 10^5$ independent parameter vectors.

The m_{brain} vectors comprise the solutions of a nonlinear least-squares (NLLS) fit of Eq. (10) to *in vivo* signal data, m_{fit}, and an additional parameter set, m_{mut}, consisting of random mutations of the fitted solutions:

$$m_{\text{mut}} = X \circ m_{\text{fit}}, \tag{17}$$

where '∘' denotes the element-wise (Hadamard) product, and X is an 11-dimensional vector of normally distributed numbers. Each element of X is an independent and identically distributed random variable sampled from a normal distribution with mean 1 and standard deviation 0.2. The number of m_{fit} vectors was kept constant ($\#m_{\text{fit}} \sim 8 \cdot 10^4$), and the total of m_{mut} vectors was defined as:

$$\#m_{\text{mut}} = \#m_{\text{brain}} - \#m_{\text{fit}}. \tag{18}$$

The introduction of mutated parameters is a data augmentation technique, designed to simultaneously compensate for the relative low number of m_{fit} vectors and expand the (f_S, $D_{\text{I;S}}$, $D_{\text{I;Z}}$, $D_{\Delta;Z}$, $T_{2;S}$, $T_{2;Z}$, p_{20}, $\text{Re}(p_{21})$, $\text{Im}(p_{21})$, $\text{Re}(p_{22})$, $\text{Im}(p_{22})$) domain of the m_{brain} parameter targets.

Synthetic signal data were generated from m_{brain} and m_{unif} using Eq. (10) and one of three different (τ_E, **B**) acquisition protocols:
- Protocol A comprises tensor-valued encoding with full relaxation-diffusion-correlation optimized for minimal RSM parameter variance (Lampinen et al., 2020)
- Protocol B comprises tensor-valued encoding with relaxation-diffusion-correlations restricted to low *b*-values(Lampinen et al., 2019)
- Protocol C comprises diffusion-relaxation optimized for minimal RSM parameter variance but includes only linear **B** ($b_\Delta = 1$)(Lampinen et al., 2020).

Table 1 Relaxed standard model parameter bounds. The rationale behind the various model bounds is detailed in (Lampinen et al., 2020).

Bounds	f_S	$D_{\text{I;S}}$ [μm²/ms]	$D_{\text{I;Z}}$ [μm²/ms]	$D_{\Delta;Z}$	$T_{2;S}$ [ms]	$T_{2;Z}$ [ms]
Minimum	0	0.07 [a]	0.2 [a]	−0.46 [a]	30	30
Maximum	1	1.33 [a]	4.0 [a]	0.86 [a]	300	1000

[a] From $D_{\text{I;S}} = D_{\|;\text{S}}/3$ and $D_{\text{I;Z}} = (D_{\|;\text{Z}} + 2D_{\perp;\text{Z}})/3$, while enforcing $D_{\|;\text{S}}$, $D_{\|;\text{Z}}$ and $D_{\perp;\text{Z}} \in [0.2, 4.0]$ μm²/ms

Additional details on the various protocols can be found in their respective references and in **Table S1** of the Supporting Information.

Noise was sampled from the Rice distribution and added to the ground-truth synthetic signals. Because relaxation-diffusion MRI data displays spatially varying noise, the amplitude of the signal-to-noise ratio (SNR) was uniformly varied across voxels in the interval SNR $\in$ [20, 50]. This interval mimics characteristics observed for the brain *in vivo* (Lampinen et al., 2020). Finally, networks were trained using m_{train} vectors as targets and their corresponding *in silico* noisy signals as inputs.

3.3 Network evaluation

To find the optimal f_{brain} parameter, we trained networks with varying f_{brain}, deployed them on unseen *in silico* data, and compared the various networks in terms of accuracy of the resulting parameter estimates. Network accuracy was assessed via normalized root-mean-squared errors (NRMSE) and linear correlations with ground-truth values in terms of the Pearson correlation coefficient (ρ). Correlation plots were used to evaluate the network trained with the optimal f_{brain} value in further detail.

The effects of protocols A-C on network performance were evaluated in terms of NRMSE and sensitivity to parameter changes. The latter was gauged by modulating the parameters (f_S, $D_{\text{I;S}}$, $D_{\text{I;Z}}$, $D_{\Delta;\text{Z}}$, $T_{2;\text{S}}$, $T_{2;\text{Z}}$) of a RSM solution, one at a time by 10%, and measuring the response in all parameters. The original parameter set was based on *in vivo* data from the *corona radiata* where f_S = 0.45, $D_{\text{I;S}}$ = 0.58 μm^2/ms, $D_{\text{I;Z}}$ = 1.36 μm^2/ms, $D_{\Delta;\text{Z}}$ = 0.44, $T_{2;\text{S}}$ = 69 ms, $T_{2;\text{Z}}$ = 60 ms (Lampinen et al., 2020). *In silico* datasets were subsequently generated for each of the 6 modulated datasets, noise at SNR = 100 was added to the synthetic signals, and parameter estimates were finally retrieved with protocol-specific networks.

To investigate if the reduced parameter space of RotInv fitting impacts the performance of DNN fitting, we trained a network using rotationally invariant *in silico* datasets and the same optimal f_{brain} value found for the RSM network. RotInv training vectors, $m_{\text{train;RI}}$, were generated from the m_{train} vectors discussed in Section 3.2, using Eq. (13) to convert the full RSM parameter space to the (f_S, $D_{\text{I;S}}$, $D_{\text{I;Z}}$, $D_{\Delta;\text{Z}}$, $T_{2;\text{S}}$, $T_{2;\text{Z}}$, p_2) RotInv space. Subsequently, Eq. (12) was used to calculate $S_{l,\ l=\{0,2\}}$ signals from $m_{\text{train;RI}}$ and noise was added at SNR $\in$ [20, 50] to the *in silico* data. As with the full RSM model, training was performed using the $m_{\text{train;RI}}$ vectors as targets and their corresponding synthetic noisy signals as DNN inputs. Trained RSM and RotInv networks were applied to previously unseen $m_{\text{unif}}/m_{\text{unif;RI}}$ and $m_{\text{brain}}/m_{\text{brain;RI}}$ synthetic datasets, and finally compared in terms of their respective target-estimate correlations.

3.4 *In vivo* data acquisition

We analysed data from three adult volunteers previously reported in (Lampinen et al., 2020). The study was approved by the regional ethical review board in Lund and written informed consent was obtained from all volunteers prior to scanning. Measurements were performed on a MAGNETOM Prisma 3T system (Siemens Healthcare, Erlangen, Germany) using a prototype spin-echo EPI sequence (Szczepankiewicz et al., 2019a) that facilitates user-defined gradient waveforms for diffusion encoding (Szczepankiewicz et al., 2021). Data was collected using a 2.5 mm^3 isotropic spatial-resolution, 40 slices, a matrix-size of 88×88, parallel imaging factor 2 (GRAPPA), partial Fourier of 3/4,

8

a bandwidth = 1775 Hz/pixel, and "strong" fat saturation. Diffusion encoding was performed with gradient waveforms numerically optimized to maximize the encoding strength per unit time (Sjölund et al., 2015) and to suppress concomitant field effects (Szczepankiewicz et al., 2019b). Data was acquired for a total of 270 different combinations of τ_E and **B**, sampled according to protocol A in **Table S1** of the Supporting Information. Simultaneous multi-slice with a multiband factor of 2 was used to accelerate the acquisition (Setsompop et al., 2012), resulting in a repetition time of 3.4 s and a total acquisition time of 15 minutes.

3.5 *In vivo* data processing & parameter estimation

Prior to any DNN training or model fitting, all acquired data were corrected for eddy-currents and subject motion using ElastiX (Klein et al., 2009) with extrapolated target volumes (Nilsson et al., 2015). Moreover, susceptibility-induced geometric distortions were corrected using the TOPUP tool in FMRIB software library (FSL) (Smith et al., 2004), and Gibbs ringing artefact correction was performed according to the method described in (Kellner et al., 2016). To further improve the smoothness of the sought RSM parameter maps, we followed the procedure used in (Lampinen et al., 2020) and filtered the corrected data with a Gaussian kernel with a standard deviation of 0.45 times the voxel size.

The RSM model parameters were estimated from a voxel-by-voxel NLLS fit of Eq. (10) to the post-processed data. The fitting process was performed with the multidimensional dMRI toolbox (https://github.com/markus-nilsson/md-dmri) (Nilsson et al., 2018), with MATLAB's built-in *lsqcurvefit* function being used to solve the NLLS minimization problem. To supress the frequency of outliers, model fitting was performed twice and the result with lowest residual was retained (Lampinen et al., 2020). The initial guesses were sampled uniformly from the broad parameter bounds in **Table 1**. The voxel-wise RSM solution yielding the lowest residuals was stored and used to compute *in silico* signal data following the procedure detailed in Section 3.2. In a previous study, the probability of finding the global fitting solution using two initial uniformly random guesses was estimated to approximately 99.96% for *in vivo* brain data, meaning that the final RSM solution is expected to be robust in respect to the initial random guesses (Lampinen et al., 2020). NLLS fitting of a single *in vivo* brain dataset took approximately 8 hours (~ 5.5 s per voxel) on the CPU described in Section 3.1.

Finally, Eq.(10) was also fitted to the *in vivo* data using a DNN, which took approximately 3 seconds. Training was performed on *in silico* $\boldsymbol{m}_{train}$ data with an optimal f_{brain} fraction. The network was trained using synthetic data generated from a single subject and deployed on the two other previously unseen subjects. Neural network fitting provided voxel-wise parameter maps that were compared to the ones obtained from a traditional NLLS fitting approach.

4 RESULTS

4.1 Impact of training set composition on network performance

We first investigated the influence of the f_{brain} parameter, and the result is shown in **Figure 1**. Large errors—quantified by the normalized root mean squared error (RMSE)—were found between true and estimated parameters in the WM range when the training was performed exclusively with uniformly distributed random samples ($f_{brain} = 0$). This finding can likely be attributed to the vastness of

9

the 11-dimensional space of model parameters and that the model parameters associated with WM occupy a very small part of this space. Indeed, when drawing uniformly distributed random samples from the intervals in **Table 1** only $\sim 2\%$ of the samples will resemble the expected properties of healthy WM tissue, i.e., $D_I = f_S D_{I;S} + (1-f_S) D_{I;Z} \in [0.8, 1.1]$ µm²/ms and $T_2 = f_S T_{2;S} + (1-f_S) T_{2;Z} \in [60, 120]$ ms. Mixing in small amounts of m_{brain} led to a drastic reduction of the error, with near-constant performance obtained already for $f_{\text{brain}} \geq 0.1$.

We now turn our attention to the correlations between estimated and ground-truth parameters. For networks deployed on data derived from m_{brain} vectors, higher Pearson correlation coefficients (ρ) are observed for higher f_{brain} fractions. Conversely, networks deployed on m_{unif} data resulted in a monotonically decreasing relationship between ρ and f_{brain}. The opposing trends are not surprising, and simply indicate that better results are obtained whenever the training and test datasets are generated with comparable strategies.

Networks trained with high f_{brain} fractions provide better estimates whenever the underlying data falls within the expected range of WM parameters, but high performance is constrained to a relatively small domain of model parameters. The f_{brain} hyper-parameter then controls a trade-off between accuracy to WM-relevant parameters and network generalizability. To achieve a balance between accuracy and generalizability, we selected the $f_{\text{brain}} = 0.35$, because it maximizes the sum between the parameter-wide mean ρ of m_{brain} data (solid blue line in right plot of **Figure 1**) and the parameter-wide mean ρ of m_{unif} data (solid red line in right plot of **Figure 1**). From this point onward, we concentrate on networks trained with $f_{\text{brain}} = 0.35$ datasets unless stated otherwise.

4.2 Neural network parameter estimates

Network-based parameter estimation was $\sim 10^4$ times faster than NLLS fitting on the same computer, and yielded parameters that are in good agreement with the ground-truth targets and preserve contrast between regions characterized by distinct (T_2, **D**) properties (see **Figure 2**). For example, the estimated f_S and p_2 maps are similar to their targets, being high in WM regions and highest in orientationally coherent WM regions such as the *corpus callosum*. An example of a slight degradation of contrast can however be observed in the $T_{2;Z}$ parameter maps, where the distinction between WM (darker) and cortical GM (brighter) regions is more clear in the original map than in its NN estimate. The $T_{2;Z}$ estimates are also characterized by considerable differences between ground-truth and estimated parameters in the long T_2 regions such as ventricles. The largest discrepancy between estimated and target parameters was found for $D_{\Delta;Z}$, likely because the signal is insensitive to this parameter below values of 0.5 (Eriksson et al., 2015). Using a DNN trained on synthetic data to fit *in vivo* experimental data resulted in noisier maps that nevertheless preserve anatomically plausible contrast. The noisier appearance of the *in vivo* parameter maps is attributed to the relatively high residuals of the RSM model (Lampinen et al., 2020).

Figure 3 shows that network-based estimates correlate well with the ground-truth parameter targets, with most parameters yielding linear correlation coefficients above 0.9. The referenced figure focuses on the performance of a network trained with *in silico* $S(\tau_E, \mathbf{B})$ data generated with protocol A and $f_{\text{vivo}} = 0.35$, and distinguishes between performance on parameters obtained by uniform random sampling (light blue points) and parameters derived from *in vivo* non-cortical brain data (dark blue

10

points). Red points correspond to parameter vectors derived from low compartment-specific signal fractions, as described in the figure caption. A considerably poor performance is observed for low $D_{\Delta;Z}$ values, where the network yields $D_{\Delta;Z} \sim 0.3$ regardless of the underlying ground-truth. This can be attributed to an intrinsic difficulty in distinguishing between the diffusion-weighted signals of $|D_{\Delta;Z}| < 0.5$ components (Eriksson et al., 2015). Moreover, weak correlations are observed for $T_{2;Z}$-times much longer than the maximal τ_E of protocol A.

The errors and prediction-target correlations of the network-based estimates are compiled in **Table S2** of the Supporting Information, where they are additionally compared to the errors and correlations obtained with a conventional NLLS solver. The NLLS approach has a higher accuracy for *in silico* datasets designed to capture non-cortical (T_2, $\mathbf{D}$) properties. By contrast, the function-fitting network is observed to be more accurate than the NLLS approach for synthetic $\boldsymbol{m}_{\text{unif}}$ datasets.

4.3 Effect of acquisition protocol on network accuracy and sensitivity

In this section, we focus on the relationship between acquisition protocol design and network performance. **Figure 4** shows that network-based fitting could not improve the known fit degeneracy in protocols B and C. Indeed, networks based on the full relaxation-diffusion-correlation optimized protocol (protocol A) consistently provide lower estimation errors. Comparing protocols B and C, we note that protocol B results in a better performance when the test data is generated from $\boldsymbol{m}_{\text{unif}}$ and observe a mixed performance for *in silico* test data based on $\boldsymbol{m}_{\text{brain}}$ targets (protocol B yields more accurate estimates of $D_{\text{I};S}$, $D_{\text{I};Z}$, $T_{2;Z}$, and p_2, while protocol C yields more accurate estimates of f_S, $D_{\Delta;Z}$, and $T_{2;S}$).

Figure 5 shows the sensitivity of the various protocols to small parameter changes. Networks trained on data generated with protocol A are sensitive to 10% parameter modulations, but underestimate the magnitude of the change slightly. The parameter-specific modulations did not have a major effect on the estimation of the remaining unmodulated parameters. An exception occurs when the underlying $T_{2;Z}$ is increased by 10%, which results in a 4% underestimation of the unperturbed f_S. Compared to protocol A, protocols B and C exhibit a lower sensitivity to the small parameter modulations and appear to be unresponsive to changes in $D_{\Delta;Z}$ and $D_{\text{I};S}$, respectively. Besides its lower sensitivity, protocol C also results in less accurate estimations of the unmodulated parameters, with a 10% modulation of $T_{2;Z}$ leading to 6% increase of the estimated f_S.

4.4 Neural network fitting of Rotationally Invariant microstructural features

Figure 6A displays parameter maps of f_S, $D_{\text{I};S}$, $D_{\text{I};Z}$ and $D_{\Delta;Z}$, obtained by applying a network trained with rotational invariants to an unseen *in vivo* $S_{l=\{0,2\}}$ dataset. The resulting maps have a smooth appearance and exhibit anatomically plausible contrast. For example, regions with high f_S correspond to WM regions, the lateral ventricles are characterized by low f_S and high $D_{\text{I};Z}$ values, and darker/brighter $D_{\Delta;Z}$ regions demarcate cortical/non-cortical parenchyma. While it is tempting to favour the seductively 'robust' maps of **Figure 6A** over the noisier maps of **Figure 2** (fourth column), we note that the RotInv-based network fitting approach results in worse correlations between target and estimated parameters (compare the scatter plots of **Figure 6B** with those of **Figure 3**). For example, network-based estimates of $D_{\Delta;Z}$ might yield a smooth map that appeals to our intuition, but

11

a closer look reveals that the $D_{\Delta;Z}$ estimates in WM and deep GM regions are equal to the mean of the target $D_{\Delta;Z}$ distribution and constitute a very poor estimate of the underlying ground-truth. A similarly poor correlation performance has been reported in a previous machine learning study of RotInv model fitting (Reisert et al., 2017).

5 DISCUSSION & CONCLUSIONS

Replacing traditional NLLS solvers with function-fitting neural networks provides a means to vastly reduce the time required for parameter estimation with high-dimensional microstructural models. Parameter estimation with networks trained on *in silico* data was achieved within seconds on a consumer-grade desktop computer. The resulting parameter estimates were generally observed to be in good agreement with both synthetic test datasets designed to simulate the relaxation-diffusion properties of healthy WM microstructure and test datasets spanning the entire space of allowed model parameters. When deployed on unseen *in vivo* brain data, neural networks provide maps that are consistent with known brain anatomy and preserve contrast between regions with different relaxation-diffusion properties. Our findings are encouraging and in line with recent advanced dMRI modelling studies that use machine learning techniques for parameter estimation (Barbieri et al., 2020; Bertleff et al., 2017; Golkov et al., 2016; Grussu et al., 2020a; Gyori et al., 2019; Hill et al., 2021; Nedjati-Gilani et al., 2017; Palombo et al., 2020; Reisert et al., 2017). A combination of simple error metrics, correlation analysis, and sensitivity matrices was found to provide a useful set of tools for quantitatively assessing parameter-specific accuracy/sensitivity and for identifying the limitations of learning-based approaches. These tools allowed the identification of a heterogeneous performance across the various dimensions of the RSM model, with $D_{\Delta;Z}$ estimates being consistently more inaccurate than other parameter estimates. However, this was unsurprising given the known difficulties of accurately estimating the anisotropy of microscopic **D** in general (Eriksson et al., 2015), and the anisotropy of the "zeppelin" compartment in particular (Lampinen et al., 2020; Lampinen et al., 2019). By contrast, simple visual inspection of machine-learned parameter maps was found to provide limited insight on the general performance of the networks. Indeed, smooth and anatomically plausible maps can be achieved even if the correlation between estimation and target is weak; a deceptive pitfall that has also been reported by (Reisert et al., 2017).

We found no evidence that learning-based fitting pipelines can by themselves navigate a degenerate fitting landscape (Jelescu et al., 2016) or replace an exhaustive and careful probing of all relevant experimental dimensions (Coelho et al., 2019; Lampinen et al., 2020). Comparison between networks based on optimal protocols and networks based on sub-optimal protocols revealed clear differences in both accuracy and sensitivity of the resulting parameter estimates. The network trained on an optimized protocol (Lampinen et al., 2020) consistently outperformed networks trained on less adequate sampling schemes. Our results suggest that the learning approach cannot substitute for a rich set of data, and learning-based fitting of diffusion-based microstructural models should be complemented with b-tensor encoding strategies (Szczepankiewicz et al., 2021) and optimized (τ_E, **B**) sampling schemes. In this work, we tested the learning-based approach on protocols designed via Cramer-Rao lower-bound optimization (Alexander, 2008; Coelho et al., 2019; Lampinen et al., 2020). While not yet tested in conjunction with learning-based fitting pipelines, alternative optimization methods based on either efficient signal decomposition schemes (Bates et al., 2020; Song and Xiao, 2020) or deep learning algorithms for feature selection (Grussu et al., 2020b; Pizzolato et al., 2020) are also expected to have a positive impact on the performance of the DNN fitting approach.

12

The 11-dimensional parameter space of the RSM model is hard to sample densely and thus presents a challenge when designing training datasets that are representative of the vast fitting landscape. In this work, we addressed this challenge by defining composite training datasets that combine varying relative amounts of parameter vectors derived from *in vivo* healthy brain data (m_{brain}) and parameter vectors obtained from uniform random sampling of the entire model parameter space (m_{unif}). Networks trained exclusively with m_{brain} vectors display the best accuracy in terms of expected WM properties, but their domain of validity is restricted to the relatively small space spanned by m_{brain} solutions. The constrained domain of m_{brain} networks raises questions about their usefulness in practical applications focusing on WM microstructural alterations, where atypical microscopic tissue structures may lead to significant deviations from the RSM parameters found in the healthy human brain (Alexander et al., 2019). To define a good trade-off between accuracy and generalizability we optimized the fraction f_{brain} between the number of m_{brain} and m_{unif}. However, we expect that more work is needed to define a truly optimal strategy network training.

While the high-dimensional model parameter space introduces difficulties in the network training process, we found that directly reducing the model dimensionality through the computation of rotational invariants resulted in a reduced performance. Indeed, learning based on a RotInv framework yielded convincing parameter maps (reproducible, smooth, anatomically plausible), but closer inspection revealed both poor accuracy and sensitivity. Fitting an $l = 0$ RotInv network to "powder-averaged" diffusion-weighted data (Jespersen et al., 2013; Kaden et al., 2016; Lasič et al., 2014) resulted in a similarly poor performance. These observations suggest that the additional orientational information present in non-rotationally invariant data contributes valuable information to the learning-based fitting procedure.

A potential limitation of the present study is the relatively simple architecture of the trained networks, especially when compared to the complex high-dimensional RSM model. Given the impact of network design on the accuracy of its predictions (Isensee et al., 2020), future work should focus on a more elaborate design and thorough investigation of the effects of learning rate, network depth, optimization algorithms, and regularization approaches on network performance. Another improvement can be gained from incorporating mini-batch processing methods in the training process. This would allow training with larger datasets providing a denser sampling of the model parameter space and hopefully improve the accuracy and generalizability of the network-based estimates. Moreover, the supervised learning strategy used in this work should be compared against unsupervised deep learning strategies, which, when applied to bi-exponential modelling of low-*b* diffusion-weighted data, have been observed to provide more accurate estimates than NLLS or traditional Bayesian estimator (Barbieri et al., 2020; Kaandorp et al., 2020). Alternatives or complements to the fully-connected DNN architecture should also be explored. Promising avenues include the use of dropout (Gal and Ghahramani, 2016; Tanno et al., 2021) or deep ensemble strategies (Lakshminarayanan et al., 2016; Qin et al., 2021) as a means to derive uncertainty metrics, the use of network structures inspired by non-learning-based iterative fitting frameworks (Ye, 2017), or use denoising networks (Fadnavis et al., 2020; Wang et al., 2019) to minimize the amount of noise present in the data that is supplied to the function-fitting DNN. However, the sensitivity analysis presented here should be applied to any new fitting strategy to test specificity to change in single model parameters. Despite the clear room for improvement, we note that the estimated target-estimate correlations were stronger than those reported by Reisert *et al.* (Reisert et al., 2017), where a supervised learning in terms of a Bayesian estimator was used to fit a three-compartment diffusion model, and are equivalent to the correlations

13

reported in more recent works focusing on machine-learning fitting of diffusion (Gyori et al., 2019; Palombo et al., 2020) and diffusion-relaxation MRI models (Grussu et al., 2020a).

In conclusion, function fitting neural networks can be used to vastly accelerate parameter estimation with high-dimensional microstructural MRI models. However, accurate estimation is achieved only if the measurement protocol samples adequate information; we found no evidence that learning-based approaches can replace the need for a rich set of data. Therefore, deep learning methodology in MRI microstructure modelling should be matched with comprehensive data acquisition. The adequacy of a given measurement protocol, in combination with a network for parameter estimation, can be evaluated by a suite of error metrics, estimate-target correlation plots, and sensitivity matrices. The learning-based fitting framework and the test tools developed herein may be used to evaluate network performance.

CODE AVAILABILITY

Once the manuscript is accepted for publication, MATLAB code for training and deploying the networks discussed in this work will be shared in open source via a GitHub repository.

ACKNOWLEDGMENTS

14
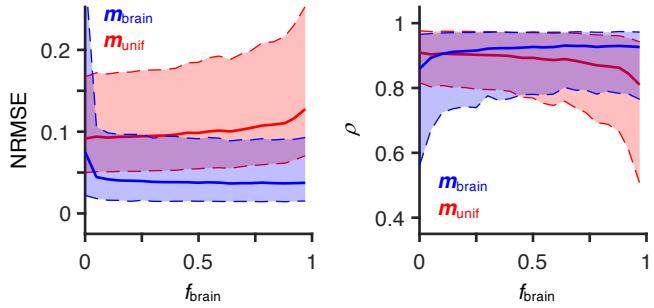
FIGURES

Figure 1. Accuracy of parameters estimated from networks trained on datasets containing varying fractions of targets derived from *in vivo* brain measurements (f_{brain}). There is a marked trade-off between accuracy and generalizability dependent on f_{brain}. For example, networks trained with high f_{brain} provide better estimates of WM-like parameters, but are constrained to a limited domain of the model space. The accuracy of network-based estimates is assessed by deploying the various networks on two previously unseen datasets where one is derived from least-squared model fitting to brain data (m_{brain}, blue), and another obtained by uniform random sampling (m_{unif}, red). The differences and correlations between estimated and ground-truth parameters is quantified by the normalised root mean square error (NRMSE) and Pearson's correlation coefficient (ρ). Solid lines represent the parameter-wide mean NRMSE (left plot) and the parameter-wide mean ρ (right plot). Computation of parameter-wide mean metrics was performed with two steps: 1) estimate NRMSE and ρ for each individual model parameter, and 2) average the parameter-specific metrics (NRMSE or ρ). The dashed lines identify the maximum and minimum estimates of parameter-specific metrics, and the shaded regions illustrate the range of metrics estimated for different model parameters.

15
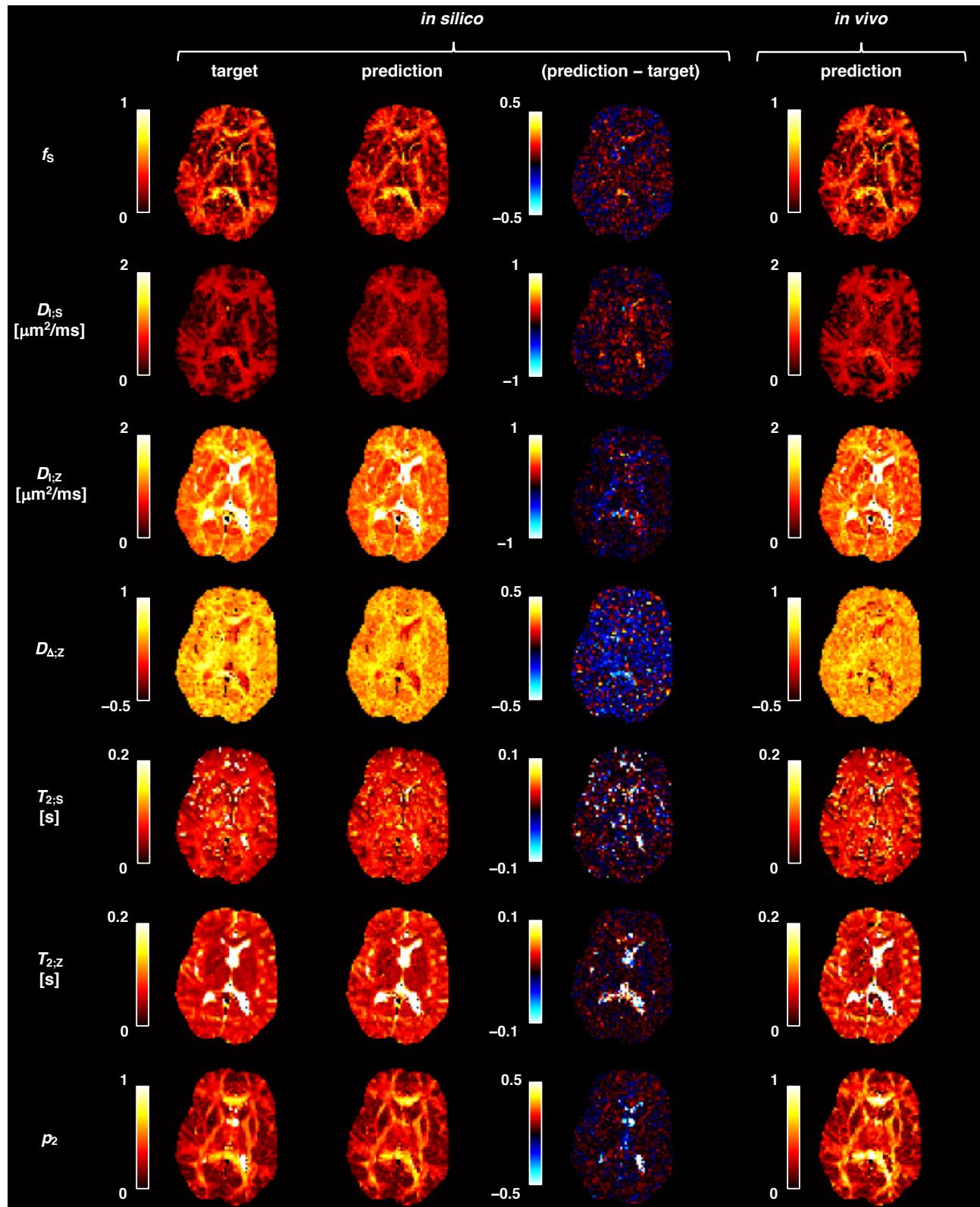
Figure 2. Deploying trained networks on previously unseen *in silico* and *in vivo* data provides anatomically plausible parameter maps in under 10 s (including data management times). The first and second columns compare the ground-truth targets and network predictions, respectively, of the *in*

silico dataset. Difference maps are shown in the third column. Parameter maps obtained from applying a trained network on *in vivo* brain data are displayed in the fourth column.
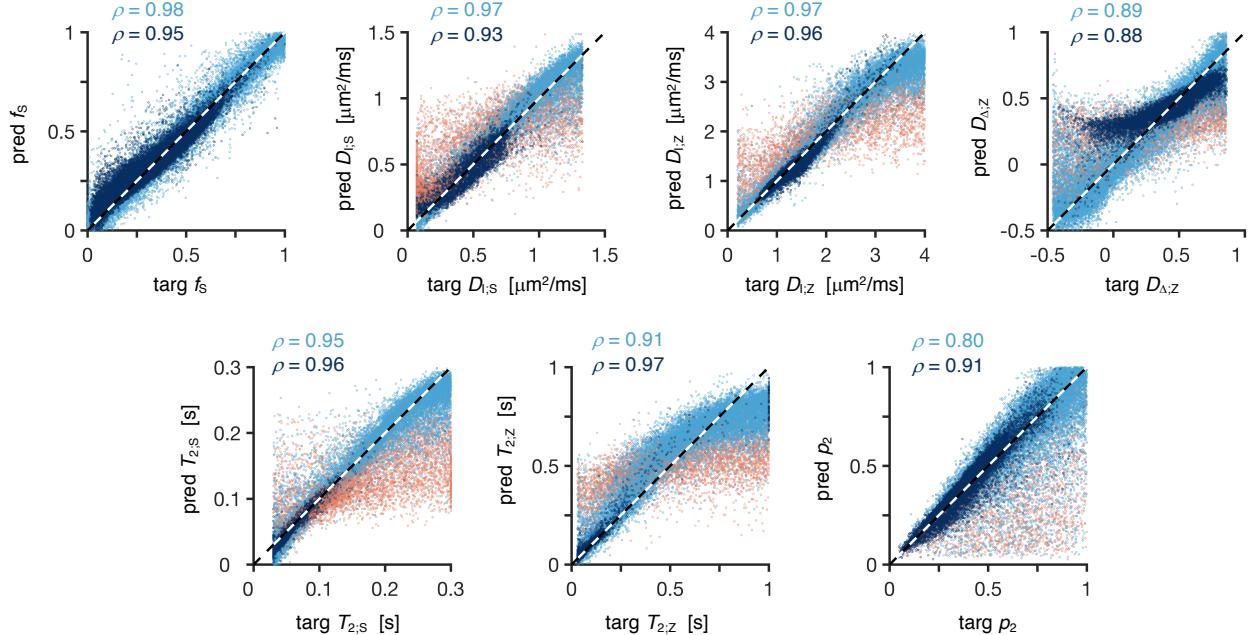

Figure 3. Scatter plots of ground-truth parameters vs. neural network predictions. Light blue points show results when the network is deployed on uniformly distributed random model parameters. The dark blue points correspond to an *in silico* dataset derived from a nonlinear least-squared fit to measured brain data where voxels within CSF and cortical GM were excluded by masking out regions where microscopic anisotropy (Lasič et al., 2014), μFA, is lower than 0.6. The red points correspond to regions where poor accuracy is expected, i.e., where the signal fraction of the relevant component ("stick" or "zeppelin" depending on the parameter) accounts for less than 15% of the total signal or, for the p_2 map, parameter vectors where the "zeppelin" component accounts for more than 85% of the total signal fraction and $|D_{\Delta;Z}| < 0.4$. The inner legends show the Pearson correlation coefficients (ρ) of the blue points.
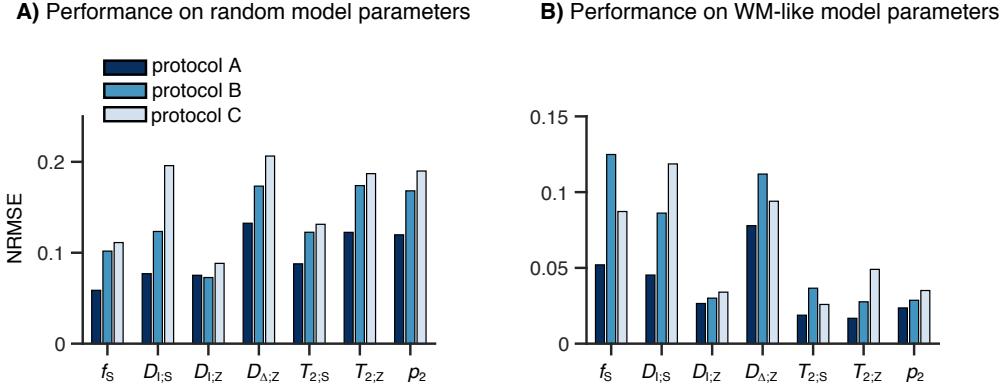
Figure 4. Optimized acquisition protocols result in learning-based parameter estimates with lower errors. The bar plots indicate the normalized root-mean-squared errors (NRMSE) between ground-truth and predicted parameters, for *in silico* datasets generated with different acquisition protocols. Protocol A corresponds to a tensor-valued (τ_E, **B**) protocol optimized for minimal parameter variance (Lampinen et al., 2020), Protocol B is a sub-optimal tensor-valued (τ_E, **B**) protocol where relaxation-diffusion correlations are exclusively established at low *b*-values (Lampinen et al., 2019), and Protocol C is a (τ_E, **B**) protocol limited to linear diffusion encoding ($b_\Delta = 1$) and optimized for parameter precision (Lampinen et al., 2020). Panel **A** shows network performance on parameters sampled from a uniform distribution, and panel **B** shows the performance on *in silico* data based on least-squares fitting results to *in vivo* non-cortical brain tissue data.
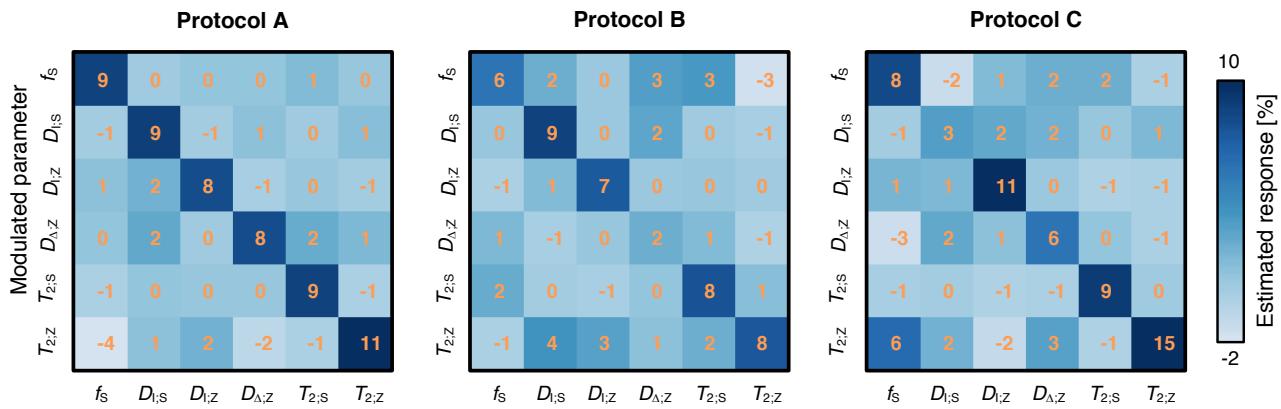
Figure 5. Sensitivity of acquisition protocols to 10% parameter modulations. The matrices display the relation between an induced parameter change and the observed response. When a single parameter on the y-axis is modulated by 10%, the response can be read in all other parameters along the *x*-axis. An ideal network would report a diagonal matrix with the value 10% on the diagonal, and zero otherwise. Protocol A appears sensitive in all parameters, whereas Protocols B and C lack sensitivity to $D_{\Delta;Z}$ and $D_{I;S}$, respectively.

18
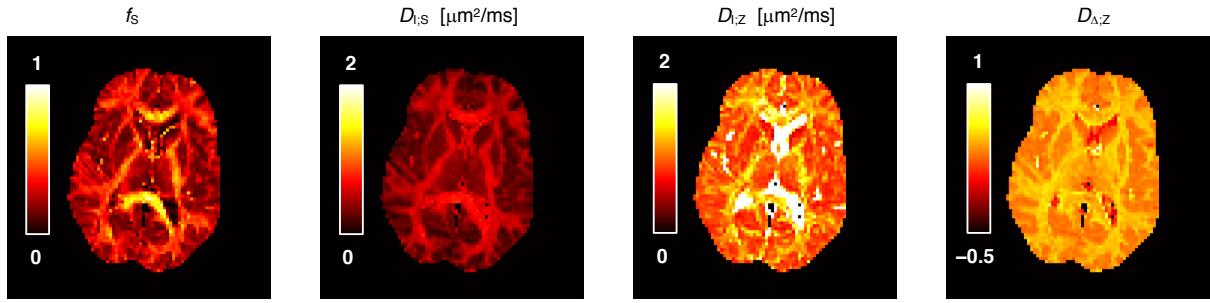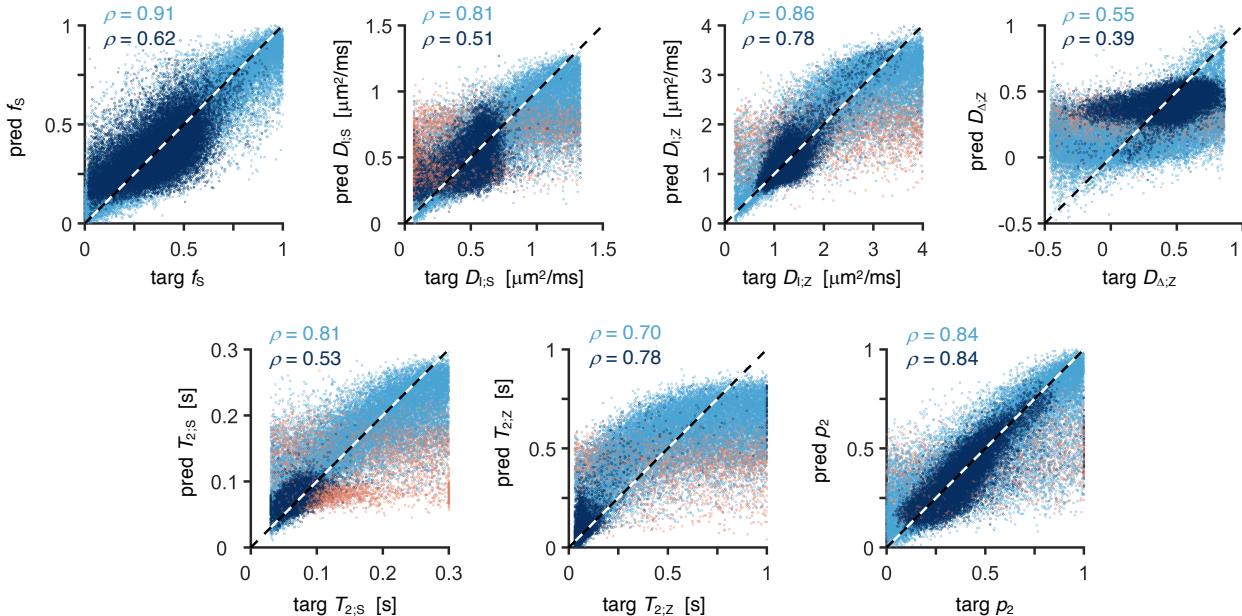
Figure 6. Neural network fitting of a rotationally invariant (RotInv) model results in plausible maps but poor target-estimate correlations. **A)** Maps of microstructural diffusion parameters – f_S, $D_{I;S}$, $D_{I;Z}$ and $D_{\Delta;Z}$ – obtained from fitting a RotInv network to rotationally invariant *in vivo* brain data. The RotInv network was trained using a fraction of $f_{\text{vivo}} = 0.35$ between rotationally invariant $\boldsymbol{m}_{\text{brain}}$ and $\boldsymbol{m}_{\text{unif}}$ training parameter vectors. **B)** Correlations between network-based parameter estimates and ground-truth parameter targets. The colour-coding and legends follow the same convention as **Figure 3**.

REFERENCES

Alexander, D.C., 2008. A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features. Magnetic Resonance in Medicine 60, 439-448.

Alexander, D.C., Dyrby, T.B., Nilsson, M., Zhang, H., 2019. Imaging brain microstructure with diffusion MRI: practicality and applications. NMR in Biomedicine 32, e3841.

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

Barbieri, S., Gurney-Champion, O.J., Klaassen, R., Thoeny, H.C., 2020. Deep learning how to fit an intravoxel incoherent motion model to diffusion-weighted MRI. Magnetic Resonance in Medicine 83, 312-321.

Bates, A.P., Daducci, A., Sadeghi, P., Caruyer, E., 2020. A 4D Basis and Sampling Scheme for the Tensor Encoded Multi-Dimensional Diffusion MRI Signal. IEEE Signal Processing Letters 27, 790-794.

Bertleff, M., Domsch, S., Weingärtner, S., Zapp, J., O'Brien, K., Barth, M., Schad, L.R., 2017. Diffusion parameter mapping with the combined intravoxel incoherent motion and kurtosis model using artificial neural networks at 3 T. NMR in Biomedicine 30, e3833.

Coelho, S., Pozo, J.M., Jespersen, S.N., Frangi, A.F., 2019. Optimal Experimental Design for Biophysical Modelling in Multidimensional Diffusion MRI. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer International Publishing, Cham, pp. 617-625.

Conturo, T.E., McKinstry, R.C., Akbudak, E., Robinson, B.H., 1996. Encoding of anisotropic diffusion with tetrahedral gradients: A general mathematical diffusion formalism and experimental results. Magnetic Resonance in Medicine 35, 399-412.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems 2, 303-314.

de Almeida Martins, J.P., Tax, C.M.W., Szczepankiewicz, F., Jones, D.K., Westin, C.F., Topgaard, D., 2020. Transferring principles of solid-state and Laplace NMR to the field of in vivo brain MRI. Magn. Reson. 1, 27-43.

de Almeida Martins, J.P., Topgaard, D., 2018. Multidimensional correlation of nuclear relaxation rates and diffusion tensors for model-free investigations of heterogeneous anisotropic porous materials. Sci Rep 8, 2488.

Driscoll, J.R., Healy, D.M., 1994. Computing Fourier Transforms and Convolutions on the 2-Sphere. Advances in Applied Mathematics 15, 202-250.

Eriksson, S., Lasic, S., Nilsson, M., Westin, C.F., Topgaard, D., 2015. NMR diffusion-encoding with axial symmetry and variable anisotropy: Distinguishing between prolate and oblate microscopic diffusion tensors with unknown orientation distribution. J Chem Phys 142, 104201.

Fadnavis, S., Batson, J., Garyfallidis, E., 2020. Patch2Self: Denoising Diffusion MRI with Self-Supervised Learning. arXiv preprint arXiv:2011.01355.

Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. international conference on machine learning, pp. 1050-1059.

Golkov, V., Dosovitskiy, A., Sperl, J.I., Menzel, M.I., Czisch, M., Sämann, P., Brox, T., Cremers, D., 2016. q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. IEEE Transactions on Medical Imaging 35, 1344-1351.

Grussu, F., Battiston, M., Palombo, M., Schneider, T., Wheeler-Kingshott, C.A.G., Alexander, D.C., 2020a. Deep learning model fitting for diffusion-relaxometry: a comparative study. bioRxiv.

Grussu, F., Blumberg, S.B., Battiston, M., Kakkar, L.S., Lin, H., Ianuş, A., Schneider, T., Singh, S., Bourne, R., Punwani, S., Atkinson, D., Gandini Wheeler-Kingshott, C.A.M., Panagiotaki, E., Mertzanidou, T., Alexander, D.C., 2020b. "Select and retrieve via direct upsampling" network (SARDU-Net): a data-driven, model-free, deep learning approach for quantitative MRI protocol design. bioRxiv, 2020.2005.2026.116491.

Gyori, N.G., Clark, C.A., Dragonu, I., Alexander, D.C., Kaden, E., 2019. In-vivo neural soma imaging using b-tensor encoding and deep learning. 27th Annual Meeting of the ISMRM, Montreal, Canada.

Healy, D.M., Hendriks, H., Kim, P.T., 1998. Spherical Deconvolution. Journal of Multivariate Analysis 67, 1-22.

Hill, I., Palombo, M., Santin, M., Branzoli, F., Philippe, A.-C., Wassermann, D., Aigrot, M.-S., Stankoff, B., Baron-Van Evercooren, A., Felfli, M., Langui, D., Zhang, H., Lehericy, S., Petiet, A., Alexander, D.C., Ciccarelli, O., Drobnjak, I., 2021. Machine learning based white matter models with permeability: An experimental study in cuprizone treated in-vivo mouse model of axonal demyelination. NeuroImage 224, 117425.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359-366.

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2020. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods.

Jelescu, I.O., Veraart, J., Fieremans, E., Novikov, D.S., 2016. Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue. NMR in Biomedicine 29, 33-47.

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

Jespersen, S.N., Lundell, H., Sønderby, C.K., Dyrby, T.B., 2013. Orientationally invariant metrics of apparent compartment eccentricity from double pulsed field gradient diffusion experiments. NMR in Biomedicine 26, 1647-1662.

Kaandorp, M., Barbieri, S., Klaassen, R., van Laarhoven, H.W., Crezee, H., While, P.T., Nederveen, A.J., Gurney-Champion, O.J., 2020. Improved unsupervised physics-informed deep learning for intravoxel-incoherent motion modeling and evaluation in pancreatic cancer patients. arXiv preprint arXiv:2011.01689.

Kaden, E., Kruggel, F., Alexander, D.C., 2016. Quantitative mapping of the per-axon diffusion coefficients in brain white matter. Magnetic Resonance in Medicine 75, 1752-1763.

Kellner, E., Dhital, B., Kiselev, V.G., Reisert, M., 2016. Gibbs-ringing artifact removal based on local subvoxel-shifts. Magnetic resonance in medicine 76, 1574-1581.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2009. Elastix: a toolbox for intensity-based medical image registration. IEEE transactions on medical imaging 29, 196-205.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2016. Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv preprint arXiv:1612.01474.

Lampinen, B., Szczepankiewicz, F., Mårtensson, J., van Westen, D., Hansson, O., Westin, C.-F., Nilsson, M., 2020. Towards unconstrained compartment modeling in white matter using diffusion-relaxation MRI with tensor-valued diffusion encoding. Magnetic Resonance in Medicine 84, 1605–1623.

Lampinen, B., Szczepankiewicz, F., Noven, M., van Westen, D., Hansson, O., Englund, E., Martensson, J., Westin, C.F., Nilsson, M., 2019. Searching for the neurite density with diffusion MRI: Challenges for biophysical modeling. Hum Brain Mapp 40, 2529-2545.

Lasič, S., Szczepankiewicz, F., Eriksson, S., Nilsson, M., Topgaard, D., 2014. Microanisotropy imaging: quantification of microscopic diffusion anisotropy and orientational order parameter by diffusion MRI with magic-angle spinning of the q-vector. Frontiers in Physics 2, 11.

LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.-R., 2012. Efficient BackProp. In: Montavon, G., Orr, G.B., Müller, K.-R. (Eds.), Neural Networks: Tricks of the Trade: Second Edition. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 9-48.

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

Nilsson, M., Szczepankiewicz, F., Lampinen, B., Ahlgren, A., de Almeida Martins, J.P., Lasic, S., Westin, C.-F., Topgaard, D., 2018. An open-source framework for analysis of multidimensional diffusion MRI data implemented in MATLAB. 26th Annual Meeting of the ISMRM, Paris, France.

Nilsson, M., Szczepankiewicz, F., van Westen, D., Hansson, O., 2015. Extrapolation-Based References Improve Motion and Eddy-Current Correction of High B-Value DWI Data: Application in Parkinson's Disease Dementia. PLoS One 10, e0141825.

Nilsson, M., van Westen, D., Ståhlberg, F., Sundgren, P.C., Lätt, J., 2013. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. Magn. Reson. Mater. Phy. 26, 345-370.

Novikov, D.S., Fieremans, E., Jespersen, S.N., Kiselev, V.G., 2019. Quantifying brain microstructure with diffusion MRI: Theory and parameter estimation. NMR Biomed 32, e3998.

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 174, 518-538.

Palombo, M., Ianus, A., Guerreri, M., Nunes, D., Alexander, D.C., Shemesh, N., Zhang, H., 2020. SANDI: A compartment-based model for non-invasive apparent soma and neurite imaging by diffusion MRI. NeuroImage 215, 116835.

Pizzolato, M., Palombo, M., Bonet-Carne, E., Tax, C.M., Grussu, F., Ianus, A., Bogusz, F., Pieciak, T., Ning, L., Larochelle, H., 2020. Acquiring and Predicting Multidimensional Diffusion (MUDI) Data: An Open Challenge. Computational Diffusion MRI. Springer, pp. 195-208.

Qin, Y., Liu, Z., Liu, C., Li, Y., Zeng, X., Ye, C., 2021. Super-Resolved q-Space deep learning with uncertainty quantification. Medical Image Analysis 67, 101885.

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.

Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magnetic Resonance in Medicine 67, 1210-1224.

Sjölund, J., Szczepankiewicz, F., Nilsson, M., Topgaard, D., Westin, C.-F., Knutsson, H., 2015. Constrained optimization of gradient waveforms for generalized diffusion encoding. Journal of Magnetic Resonance 261, 157-168.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, S208-S219.

Song, Y.-Q., Xiao, L., 2020. Optimization of multidimensional MR data acquisition for relaxation and diffusion. NMR in Biomedicine n/a, e4238.

Szczepankiewicz, F., Sjolund, J., Stahlberg, F., Latt, J., Nilsson, M., 2019a. Tensor-valued diffusion encoding for diffusional variance decomposition (DIVIDE): Technical feasibility in clinical MRI systems. PLoS One 14, e0214238.

Szczepankiewicz, F., Westin, C.-F., Nilsson, M., 2021. Gradient waveform design for tensor-valued encoding in diffusion MRI. Journal of Neuroscience Methods 348, 109007.

Szczepankiewicz, F., Westin, C.F., Nilsson, M., 2019b. Maxwell-compensated design of asymmetric gradient waveforms for tensor-valued diffusion encoding. Magnetic Resonance Medicine 82, 1424-1437.

Tanno, R., Worrall, D.E., Kaden, E., Ghosh, A., Grussu, F., Bizzi, A., Sotiropoulos, S.N., Criminisi, A., Alexander, D.C., 2021. Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI. NeuroImage 225, 117366.

Tian, Q., Bilgic, B., Fan, Q., Liao, C., Ngamsombat, C., Hu, Y., Witzel, T., Setsompop, K., Polimeni, J.R., Huang, S.Y., 2020. DeepDTI: High-fidelity six-direction diffusion tensor imaging using deep learning. NeuroImage 219, 117017.

Topgaard, D., 2017. Multidimensional diffusion MRI. J Magn Reson 275, 98-113.

Veraart, J., Novikov, D.S., Fieremans, E., 2018. TE dependent Diffusion Imaging (TEdDI) distinguishes between compartmental T2 relaxation times. Neuroimage 182, 360-369.

Wang, H., Zheng, R., Dai, F., Wang, Q., Wang, C., 2019. High-field mr diffusion-weighted image denoising using a joint denoising convolutional neural network. Journal of Magnetic Resonance Imaging 50, 1937-1947.

Westin, C.-F., Knutsson, H., Pasternak, O., Szczepankiewicz, F., Özarslan, E., van Westen, D., Mattisson, C., Bogren, M., O'donnell, L.J., Kubicki, M., 2016. Q-space trajectory imaging for multidimensional diffusion MRI of the human brain. NeuroImage 135, 345-362.

Ye, C., 2017. Estimation of Tissue Microstructure Using a Deep Network Inspired by a Sparse Reconstruction Framework. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T., Shen, D. (Eds.), Information Processing in Medical Imaging. Springer International Publishing, Cham, pp. 466-477.

	in silico			*in vivo*
	target	prediction	(prediction − target)	prediction

f_S

$D_{I;S}$ [μm²/ms]

$D_{I;Z}$ [μm²/ms]

$D_{\Delta;Z}$

$T_{2;S}$ [s]

$T_{2;Z}$ [s]

p_2
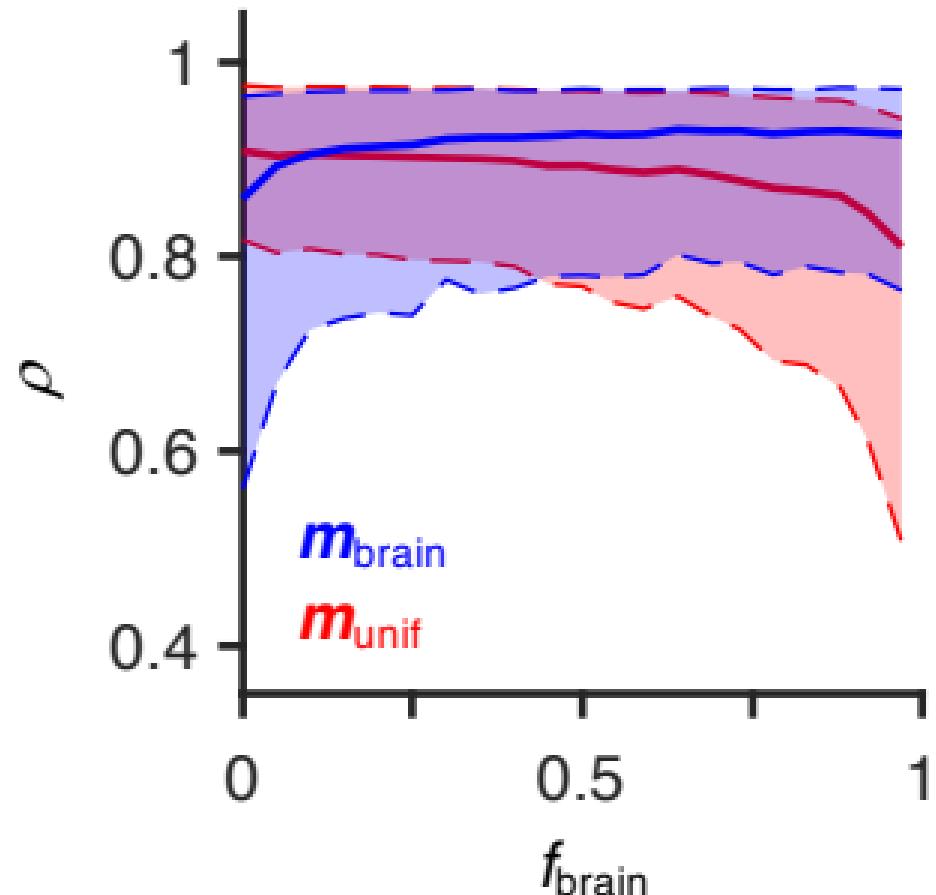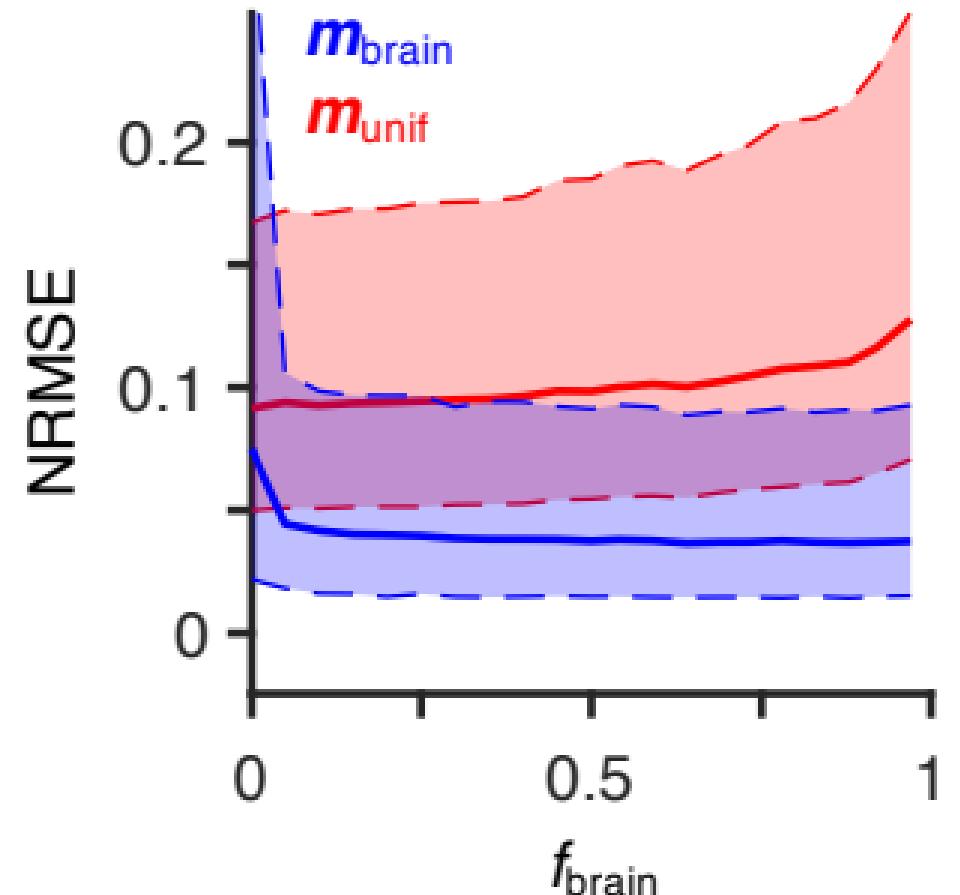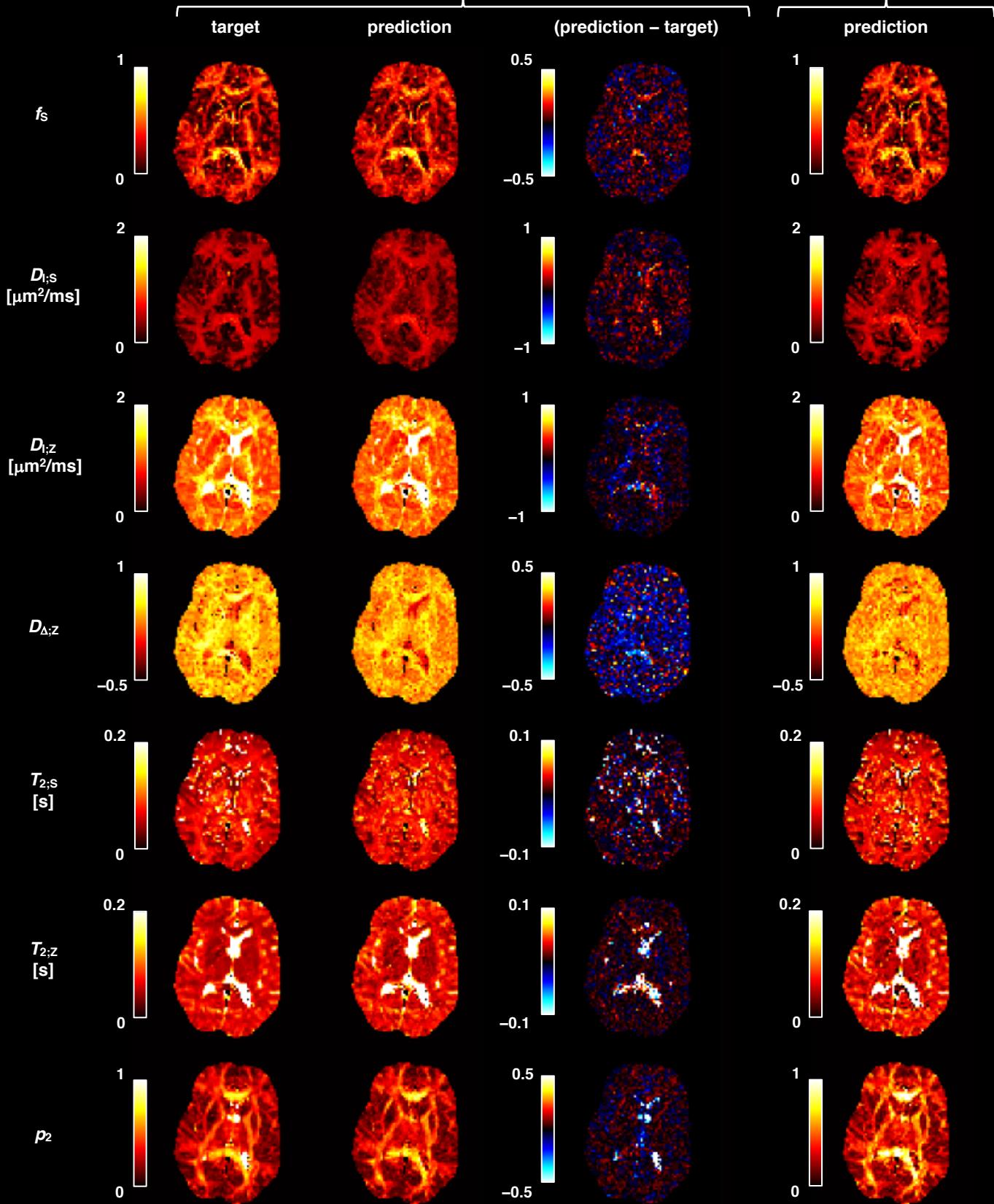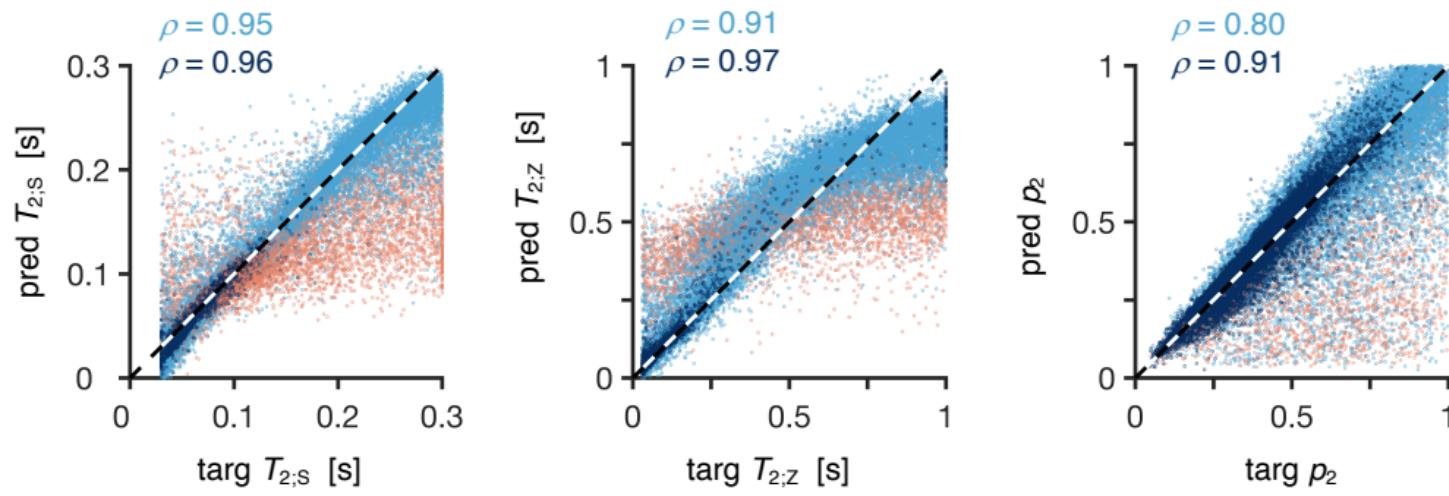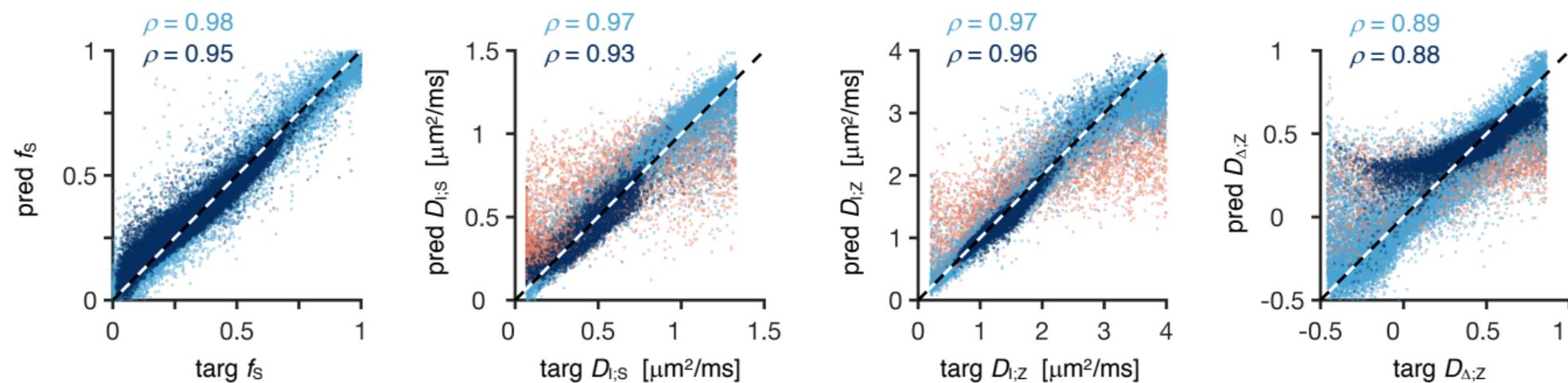
A) Performance on random model parameters
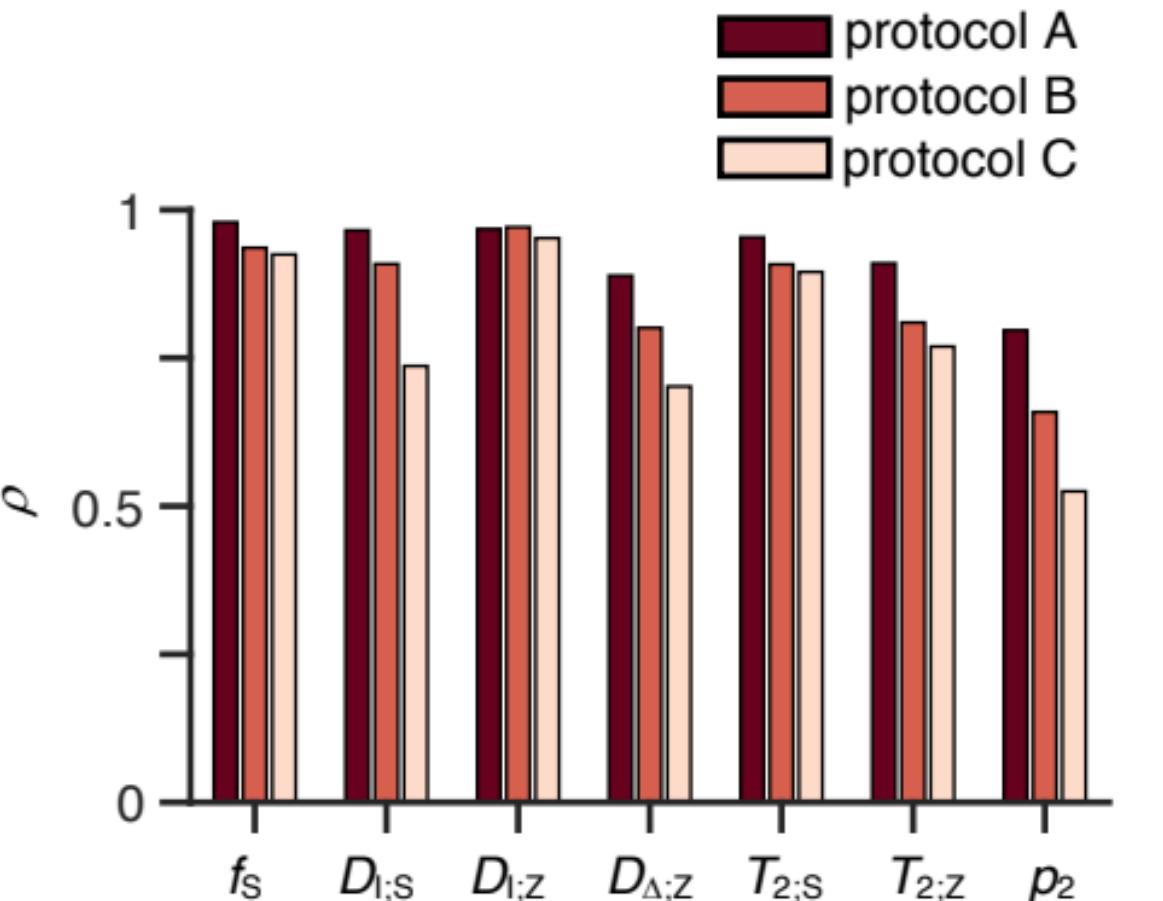
B) Performance on WM-like model parameters

- protocol A
- protocol B
- protocol C

f_{S} $D_{\mathrm{I;S}}$ $D_{\mathrm{I;Z}}$ $D_{\Delta;\mathrm{Z}}$ $T_{2;\mathrm{S}}$ $T_{2;\mathrm{Z}}$ p_2

ρ
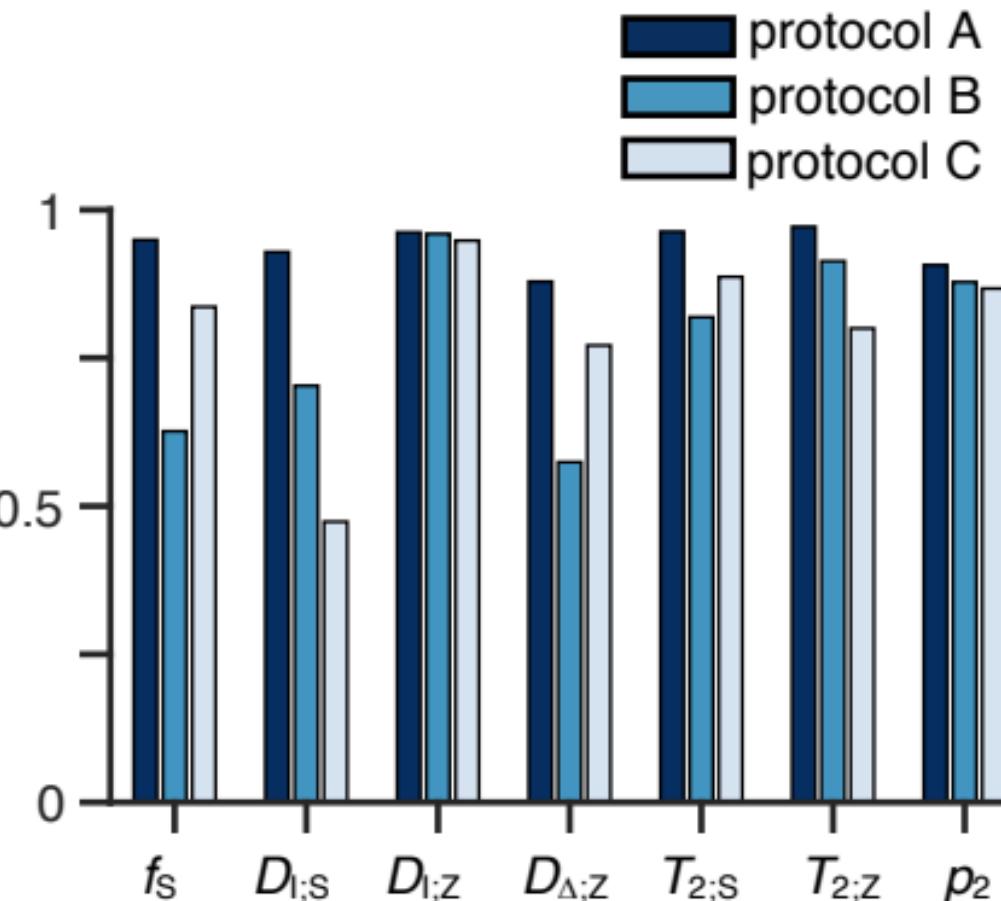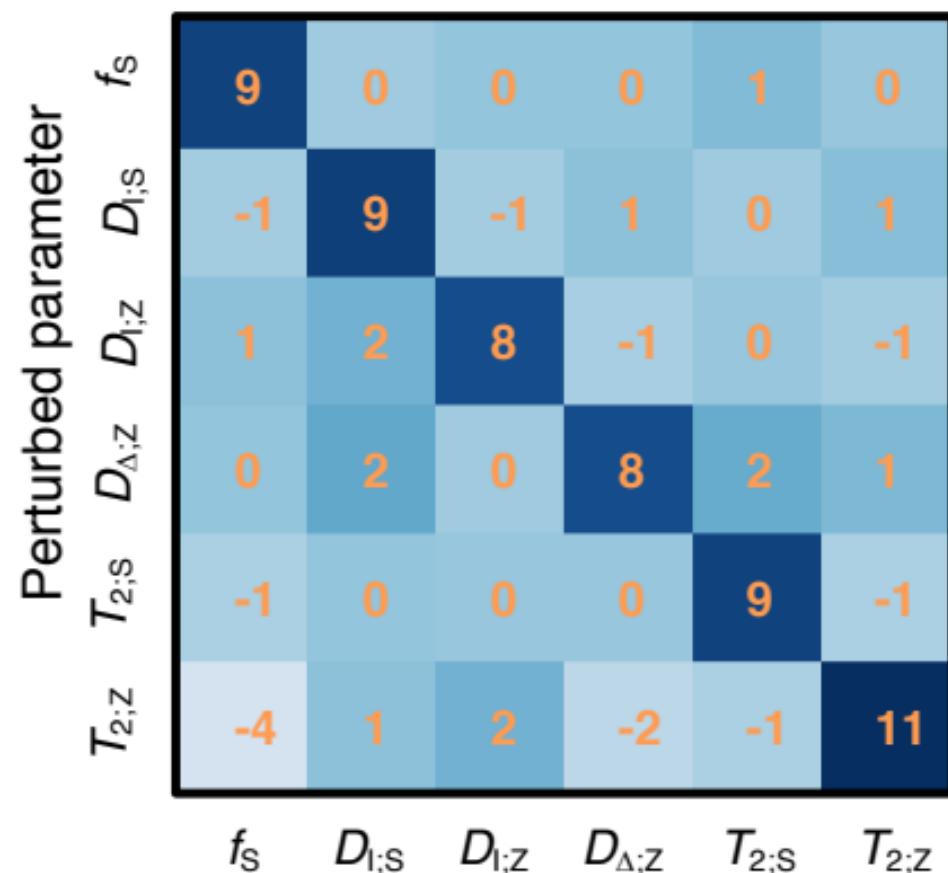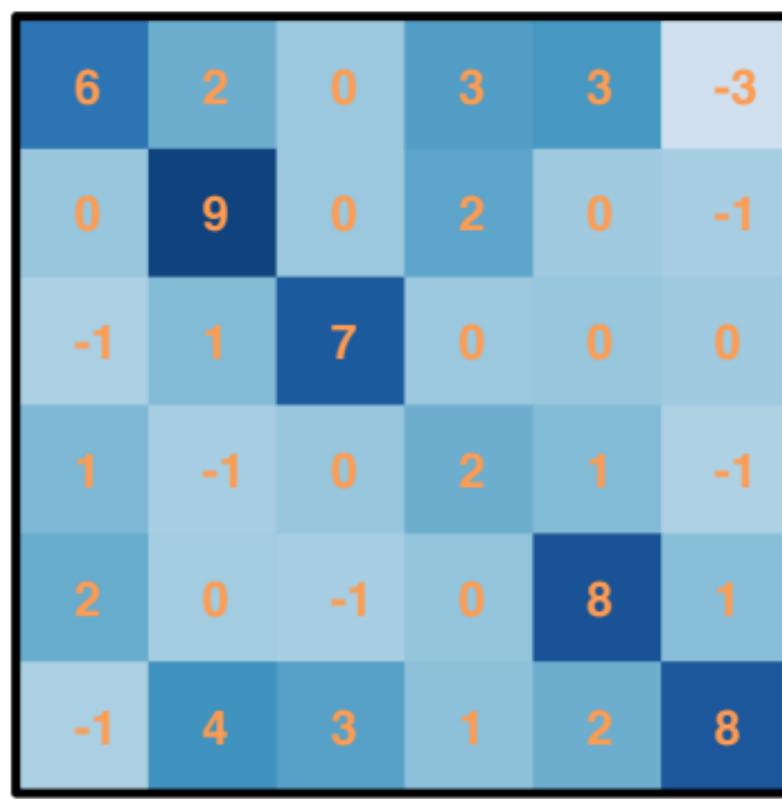
Protocol A, Protocol B, Protocol C
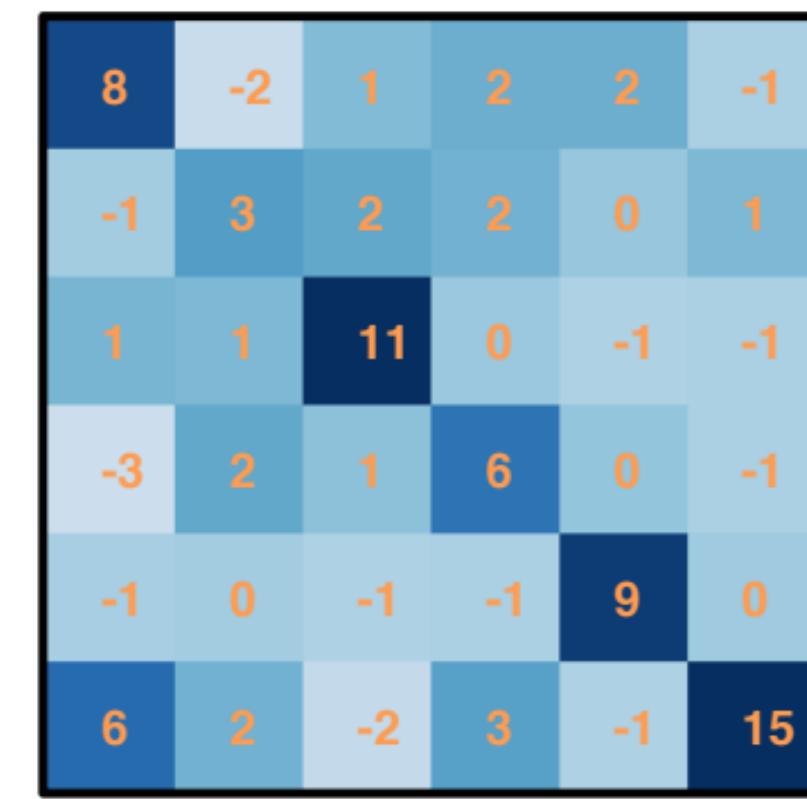
A) Selected experimental parameter maps

f_S $D_{i;S}$ [µm²/ms] $D_{i;Z}$ [µm²/ms] $D_{\Delta;Z}$
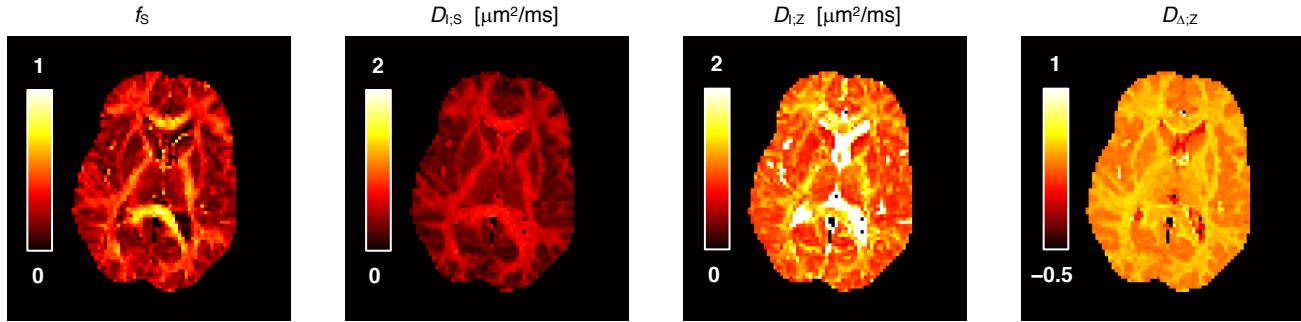
B) Target-Prediction correlation plots