Medical Image Analysis 69 (2021) 101940

Contents lists available at ScienceDirect

Medical Image Analysis

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

Model-informed machine learning for multi-component T_2 relaxometry

Thomas Yu[a,b], Erick Jorge Canales-Rodríguez[a,d,*], Marco Pizzolato[e,a], Gian Franco Piredda[f,c,a], Tom Hilbert[f,c,a], Elda Fischi-Gomez[a,g], Matthias Weigel[g,h,i], Muhamed Barakovic[g,h], Meritxell Bach Cuadra[b,a,c], Cristina Granziera[g,h], Tobias Kober[f,c,a], Jean-Philippe Thiran[a,c]

[a] Signal Processing Lab 5 (LTS5), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[b] Medical Image Analysis Laboratory, Center for Biomedical Imaging (CIBM), University of Lausanne, Switzerland
[c] Department of Radiology, Lausanne University Hospital and University of Lausanne, Switzerland
[d] FIDMAG Germanes Hospitalàries Research Foundation, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Barcelona, Spain
[e] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark
[f] Advanced Clinical Imaging Technology, Siemens Healthcare AG, Lausanne, Switzerland
[g] Translational Imaging in Neurology Basel, Department of Medicine and Biomedical Engineering, University Hospital Basel and University of Basel, Basel, Switzerland
[h] Neurologic Clinic and Policlinic, Departments of Medicine, University Hospital Basel and University of Basel, Basel, Switzerland
[i] Division of Radiological Physics, Department of Radiology, University Hospital of Basel, Basel, Switzerland

ARTICLE INFO

Article history:
Received 14 July 2020
Revised 9 December 2020
Accepted 10 December 2020
Available online 17 December 2020

Keywords:
Machine learning
T_2 relaxometry
Myelin water imaging

ABSTRACT

Recovering the T_2 distribution from multi-echo T_2 magnetic resonance (MR) signals is challenging but has high potential as it provides biomarkers characterizing the tissue micro-structure, such as the myelin water fraction (MWF). In this work, we propose to combine machine learning and aspects of parametric (fitting from the MRI signal using biophysical models) and non-parametric (model-free fitting of the T_2 distribution from the signal) approaches to T_2 relaxometry in brain tissue by using a multi-layer perceptron (MLP) for the distribution reconstruction. For training our network, we construct an extensive synthetic dataset derived from biophysical models in order to constrain the outputs with *a priori* knowledge of *in vivo* distributions. The proposed approach, called Model-Informed Machine Learning (MIML), takes as input the MR signal and directly outputs the associated T_2 distribution. We evaluate MIML in comparison to a Gaussian Mixture Fitting (parametric) and Regularized Non-Negative Least Squares algorithms (non-parametric) on synthetic data, an *ex vivo* scan, and high-resolution scans of healthy subjects and a subject with Multiple Sclerosis. In synthetic data, MIML provides more accurate and noise-robust distributions. In real data, MWF maps derived from MIML exhibit the greatest conformity to anatomical scans, have the highest correlation to a histological map of myelin volume, and the best unambiguous lesion visualization and localization, with superior contrast between lesions and normal appearing tissue. In whole-brain analysis, MIML is 22 to 4980 times faster than the non-parametric and parametric methods, respectively.

© 2020 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

The spin-spin relaxation rate T_2 is one of the basic tissue-specific, quantitative parameters which can be measured or used to give image contrast in MRI (Haacke et al., 1999). However, while commonly presented as a single number per voxel, tissue heterogeneity and partial volume effects renders it more appropriate to consider distributions of T_2s per voxel rather than a single T_2 value (Menon and Allen, 1991). We distinguish single-component T_2 relaxometry, where each voxel is characterized with a single T_2, from multicomponent T_2 relaxometry, where each voxel is characterized with a T_2 distribution. In general, T_2 distributions are reconstructed

* Corresponding author at: Signal Processing Lab 5 (LTS5), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
E-mail address: erick.canalesrodriguez@epfl.ch (E.J. Canales-Rodríguez).

https://doi.org/10.1016/j.media.2020.101940
1361-8415/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

from multi-echo T_2 MRI signals, which can be acquired, for example, through multi-echo spin echo sequences, where a 90° excitation pulse is followed by a train of 180° refocusing pulses. Given a sequence of n pulses, the signal $\mathbf{s}$ is a vector of n measurements at the corresponding echo times (TE_i). Let $p(T_2)$ and α denote the distribution of T_2s in a voxel and the effective flip angle of the refocusing pulses, respectively. If $\alpha = 180°$ and the voxel is assumed to have a single T_2, then the decay of the signal is exponential, as is implied by the Bloch equations (Bloch, 1946). In practice, inhomogeneities in the transmit field (B_1+) result in an effective refocusing pulse that can vary significantly from 180° and can be spatially heterogeneous (Prasloski et al., 2012a). This leads the resulting signal to deviate from the ideal exponential behavior, which can be modelled using the extended phase graph (EPG) formalism (Hennig, 1988). The EPG formalism considers as parameters α, TE, T_1 and a single T_2. The code used for the EPG simulations in this work is based on work in (Layton et al., 2013). We use the common simplification of fixing $T_1 = 1000$ ms, as the T_1 relaxation time cannot be estimated using the acquisition sequences we examine in this work (Neumann et al., 2014); hence, it is commonly fixed to its mean value in brain tissue. Then the normalized signal follows

$$\mathbf{s}(TE_i) = \int EPG(TE_i, T_1, T_2, \alpha)\, p(T_2) \mathrm{d}T_2. \tag{1}$$

One key application of multi-component T_2 relaxometry is in neuroimaging, where the different parts of the T_2 distribution are assumed to arise from the different anatomical compartments in brain tissue, particularly in white matter. This can be used, for instance, to generate a map of the myelin water fraction (MWF) (Piredda et al., 2020b) such that areas of demyelination corresponding to the effects of neurodegenerative disorders can be identified (MacKay and Laule, 2007). In particular, it is commonly assumed/modelled that the T_2 distribution in white matter contains multiple lobes having well-separated peaks, and that the eventual overlap between the T_2 lobes of myelin and the intra/extra axonal space water pools is minimal (Mackay et al., 1994; Whittall et al., 1997; Vasilescu et al., 1978; Menon and Allen, 1991; Menon et al., 1992).

1.1. Related work

In order to estimate $p(T_2)$ from equation (1), two main approaches are generally used: parametric and non-parametric approaches. Parametric approaches rely on *a priori* information on the T_2 distribution in brain tissue, particularly white matter, in order to fit the parameters of biophysical models to the MRI signal (Raj et al., 2014a; Du et al., 2007; Yu et al., 2019; Chatterjee et al., 2018; Akhondi-Asl et al., 2014; Björk et al., 2016). In these approaches, the MRI signal is modelled as a linear combination of signals from a fixed number of water pools (around 2-3) such as myelin water, the water in the intra-/extra-axonal space, and cerebrospinal fluid:

$$p(T_2) = \sum_{i=1}^{n} v_i F_i(\mathbf{m}_i, T_2) \tag{2}$$

Here n is the number of water pools assumed, and $F_i, \mathbf{m_i}, v_i$ are the probability distribution, parameters of the probability distribution, and volume fraction of the ith water pool. A wide variety of parametric distributions (Delta, Gaussian, Truncated Gaussian, Wald, Gamma, Log-Gaussian, Laplacian) are used to model the T_2 distributions in these pools; however, (Raj et al., 2014a) shows that using these different distributions have negligible differences on the corresponding signal when using the same means and variances; they conclude that due to the ill-posedness of the inverse problem, extracting more than general lobular shapes (characterized by the mean and variance) is extremely difficult if not impossible, even at extremely high signal to noise ratios (SNR). The parameters estimated are the water volume fractions and the parameters of the distributions which are done through optimization (Chatterjee et al., 2018; Björk et al., 2016) or Monte Carlo methods (Prange and Song, 2009; Yu et al., 2019). To stabilize the fitting and to use prior information on the compartments, constraints are enforced on the parameters. For instance, the mean T_2 of myelin water is typically bounded between 10 and 40 ms, and the mean T_2 of CSF is typically assumed to be greater than 1 s. Some works, such as (Chatterjee et al., 2018), go even further and fix the mean or standard deviations of the probability distributions of some compartments to predetermined values. While parametric estimations are generally stable and histologically validated, they are usually computationally expensive and restricted by the biophysical model used; the number of compartments needs to be fixed for each voxel before fitting. Further, we note that the *a priori* information used in the parametric approaches i.e. the assumption of lobular structure, bounds on the parameters of the distribution, etc. comes from historical evidence, where studies used non-parametric methods to estimate the T_2 distributions and assigned lobes in their reconstructions to different water pools (Alonso-Ortiz et al., 2015).

In contrast, non-parametric approaches do not make *a priori* assumptions on the data, such as the number of compartments. This is relevant for studying abnormal brain tissue, where compartments not considered in standard biophysical models might be present (MacKay and Laule, 2007). In addition, they generally require orders of magnitude less computation time than parametric methods. Non-parametric methods discretize equation (1) as a product of a dictionary matrix and a discretized T_2 distribution and solve directly for the discretized T_2 distribution (Whittall et al., 1997; Prasloski et al., 2012a) using non-negative least squares (NNLS) algorithms (Lawson and Hanson, 1995). The T_2 distribution, $p(T_2)$, is recovered by solving an inverse problem (Whittall et al., 1997; Prasloski et al., 2012a). First, given discretized ranges of flip angle (α) values and T_2 values, a dictionary D_α of T_2 decay signals is constructed for each α value through the EPG formalism. D_α is a matrix where the columns are the simulated MRI signals (obtained from the experimental TEs) over a range of T_2 values**.** Given a flip angle α, the corresponding dictionary D_α, and the MRI signal $\mathbf{s}$, the following optimization problem is solved

$$\arg\min_{\mathbf{p}\geq 0} \|D_\alpha \mathbf{p} - \mathbf{s}\|_2^2 + \lambda \Phi(\mathbf{p}) \tag{3}$$

where Φ is a regularization function with parameter λ, and $\mathbf{p}$ is the discretized, un-normalized T_2 distribution to be estimated. The flip angle corresponding to $\mathbf{s}$ is chosen by solving the above problem (with $\lambda = 0$) for multiple values of α and taking the value which corresponds to the least fitting error (Prasloski et al., 2012a). Two standard choices for $\Phi(\mathbf{p})$ (Whittall and MacKay, 1989) are

- $\Phi(\mathbf{p}) = \|\mathbf{p}\|_2^2$, which we refer to as Tikhonov regularization.
- $\Phi(\mathbf{p}) = \|L\mathbf{p}\|_2^2$, where L is a finite difference approximation of the Laplacian operator. We refer to this as Laplacian regularization.

These choices are used in order to promote increased conditioning of the problem and the smoothness of the resulting distribution (Kroeker and Henkelman, 1986). Without regularization, solutions to Eq. (3) are vulnerable to noise and usually produce inaccurate solutions that overfit the signal with e.g. false positive peaks, etc. A common heuristic for selecting λ is to accept λ such that the signal fitting error is approximately 1.02–1.025 times greater than the error from NNLS with no regularization (Laule et al., 2006). However, it is known that regularization can introduce undesirable bias to the reconstructed signals, e.g. over-smoothing. In particular, regularization can contradict the expectation of disparate lobes

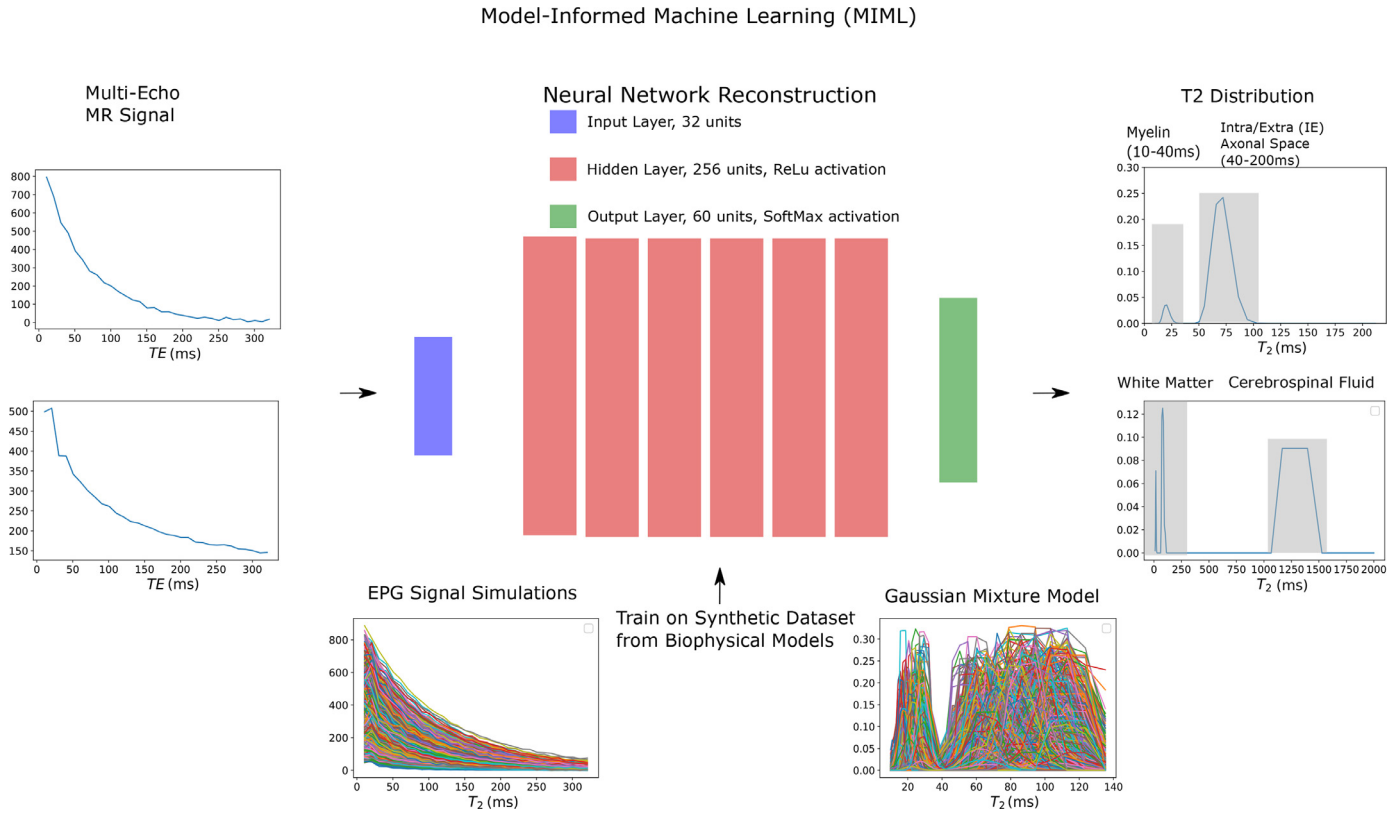

Fig. 1. An overview of our method (MIML) for multicomponent T_2 relaxometry where we learn a mapping from the multi-echo MR signal to the corresponding T_2 distribution. On the left are example MR signals, on the right are the corresponding T_2 distributions: the first distribution is in white matter (WM), where there are assumed to be two lobes: one at a T_2 of around 10–40 ms corresponding to myelin water and one at a T_2 of around 50–120 ms corresponding to the intra and extra axonal spaces. The second distribution includes WM and cerebrospinal fluid (CSF), whose T_2 is commonly assumed to be around 1-2s. Our method consists of training a neural network on a synthetic dataset derived from biophysical models to learn the mapping from signal to distribution. At the bottom, we show a small subset of 1000 simulated signals and corresponding T_2 distributions from our synthetic training dataset.

in the distribution corresponding to disparate tissues in the same voxel (e.g. myelin and intra/extra axonal space water), particularly at lower SNRs. For example, at low SNRs, the myelin water lobe can become completely over-smoothed, for an example see Fig. 1 in the Supplementary Material.

Once the T_2 distribution is recovered, generating parameters of interest such as volume fractions of the water pools in the voxel require either a distribution where distinct lobes can be assigned to distinct compartments (such as in the right side of Fig. 1) or *a priori* information. After examining distributions reconstructed from experimental scans, the different lobes of the distributions (if distinct lobes are present) are assigned to different water pools based on theoretical and experimental grounds (MacKay and Laule, 2007). From the mean and standard deviation of these lobes, bounds are derived for the T_2 values for each water pool. Then water volume fractions for each pool are calculated by integrating the probability distribution between the bounds of the T_2 for each pool. For instance, at 3T the myelin water fraction (MWF) is usually computed as

$$MWF = \frac{\int_{T_2=10\,\mathrm{ms}}^{T_2=40\,\mathrm{ms}} \mathbf{p}(T_2)\mathrm{d}T_2}{\int_{T_2=10\,\mathrm{ms}}^{T_2=2000\,\mathrm{ms}} \mathbf{p}(T_2)\mathrm{d}T_2}, \tag{4}$$

where the bounds 10–40 ms were obtained from the myelin water lobe in NNLS reconstructions in past papers (Alonso-Ortiz et al., 2015).

We note previous studies found that both parametric and non-parametric methods require a high signal-to-noise ratio (SNR) to detect different components in the T_2 distribution (Graham et al., 1996; Andrews et al., 2005; Wiggermann et al., 2020). For a clinically achievable SNR=100, more than 5% of the voxels were incorrectly estimated to have no myelin water component, and the percentage raised to 12% for SNR=50 (Kumar et al., 2012). Similar results were reported in (Raj et al., 2014b), where the myelin water component was not found in human brain regions located in myelinated areas of the frontal and lateral projections fibers. In addition, Wiggermann et al. (2020) found that in synthetic studies, NNLS with Tikhonov Regularization tends to underestimate the true MWF value in the range of 0.3 to 4% at SNR 1000, with the problem worsening at lower SNRs; for reference, the MWF is assumed to be in the range of 0–30% in normal appearing white matter.

Recently, Lee et al.,(Liu et al., 2020) have both proposed to augment non-parametric approaches with machine learning in order to speed up the computation time. As training data, they acquired brain scans in several subjects *in vivo* using a 3D multiple echo gradient and spin echo sequence with 32 echoes (Prasloski et al., 2012b). They then ran regularized NNLS reconstructions on the data and obtained the probability distributions and MWF for each voxel. Liu et al. (2020) trained a multi-layer perceptron (MLP) to take as input the raw data, and output the MWF, using the *in vivo* NNLS reconstructions as ground truth. Lee et al. (2019) trained MLPs to reconstruct the MWF as well as the probability distributions from the raw echo data, using the *in vivo* NNLS reconstructions as ground truth. These approaches have the advantage of reconstructing regularized NNLS solutions for the whole brain in under a minute, a fraction of the time required using the standard NNLS algorithm. However, as their ground truth is the reg-

ularized NNLS solution, their method inherits all the problems of NNLS. Further, by training on data acquired from specific MRI machines using a specific sequence, there is the problem of generalizing to different machines and different sequences. Both would require new acquisitions as well as additional training time.

In summary, parametric methods implicitly regularize and stabilize the problem by using biophysical models and prior knowledge to constrain the space of T_2 distributions. However, the resulting optimization problems to be solved are significantly more costly than those of non-parametric methods, with an additional loss of flexibility due to imposition of the number of compartments and other details of the model. Non-parametric solutions are fast, but also ill-posed and highly susceptible to noise; hence, regularization is necessary, with the concomitant drawbacks of oversmoothing and sparsity of the reconstructed distributions, particularly at clinically achievable SNRs for sequences with high spatial resolution. Further, the extraction of parameters of interest such as the MWF is theoretically based on assuming a lobular structure of the reconstructed distribution, which is often not the case in middling to high levels of noise.

1.2. Contributions

In this paper, we propose a new method for multi-component T_2 relaxometry in brain tissue. In Fig. 1, we show the overview of our proposed method as well as a prototypical T_2 distribution in white matter, composed of the myelin water lobe and the lobe corresponding to the water in the intra/extra axonal space; in addition, we show the corresponding MRI signal. We propose to combine machine learning and aspects of parametric and non-parametric approaches to the reconstruction of T_2 distributions from multi-echo T_2 data. We do this by creating a synthetic dataset derived from biophysical models and training a multi-layer perceptron (MLP) (Rosenblatt, 1958) on this dataset to take as input the MRI signal and directly output the associated T_2 distribution. We call our method Model-Informed Machine Learning (MIML). Our main contributions are as follows:

- Construction of an extensive synthetic dataset that we construct purely from simulations guided by biophysical models, which we use for training the MLP.
- Introduction of a robust loss function for the network to recover the T_2 distribution consisting of a combination of the mean squared error and the Wasserstein-1 Distance (Villani, 2009). We show that training with the Wasserstein distance significantly increases the accuracy of MWF estimates on a realistic, synthetic case, compared to training with solely a mean squared error (MSE) loss function.
- Rigorous and extensive evaluation of our method and previous work in non-parametric and parametric approaches, on synthetic and real datasets (*ex vivo, in vivo*, healthy, pathological). We show that our method outperforms other methods in terms of accuracy, plausibility, and robustness of the reconstructed distributions and MWF maps as well as lesion visualization.

2. Methods

Our method for reconstructing T_2 distributions from MRI data is based on a MLP which is trained to learn a map directly from MRI signals with a 32 echo acquisition scheme to the corresponding T_2 distribution, as is the result in non-parametric methods. To reduce the inherent ill-posedness of this problem, the training is conducted on a synthetic dataset of pairs of MRI signals and T_2 distributions which we constructed using EPG simulations and is informed by biophysical models and realistic values for the parameters of interest, such as the range of T_2s for different water pools, taken from the literature. This implicitly constrains the space of possible T_2 distributions (as in parametric approaches). We show an overview of our method in Fig. 1.

Table 1
The ranges for the possible mean (μ) and the standard deviations (σ) used for the Gaussian, T_2 distributions of the different water pools used in our dataset.

Range of Mean and Standard Deviation for Simulated Water Pools		
Water Pool	Range of Mean T_2 (μ)	Range of Std. of T_2 (σ)
Myelin	15–30 ms	0.1–5 ms
Intra/Extra Axonal Space (IES)	50–120 ms	0.1–12 ms
Gray Matter (GM)	60–300 ms	0.1–12 ms
Pathology	300–1000 ms	0.1–5 ms
CSF	1000–2000 ms	0.1–5 ms

2.1. Synthetic dataset generation

To generate the synthetic T_2 distributions, we start from standard biophysical models for the brain (Whittall et al., 1997). Brain tissue can be roughly subdivided into white matter, grey matter, cerebrospinal fluid, pathological tissue, and combinations of these tissues. The water of these tissues are made up of a combination of different pools of water. We model the T_2 distributions of brain tissue as a mixture of Gaussians, where each Gaussian component corresponds to a different water pool (e.g. myelin water, intra/extra axonal space water).

$$p(T_2) = \sum_i \frac{v_i}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(T_2-\mu_i)^2}{2\sigma_i^2}\right) \tag{5}$$

$$v_i \in [0,1], \sum_i v_i = 1 \tag{6}$$

Here v_i is the volume fraction of the ith water pool, and μ_i, σ_i are the mean and standard deviation of the T_2 distribution of the ith water pool. We justify our choice of modelling using Gaussians by noting that (Raj et al., 2014a) found that modelling the T_2 distributions using a variety of different distributions including the Gaussian distribution had negligible differences in parametric methods; essentially, they found that the ill-posedness of the reconstruction made it extremely difficult to distinguish between different distributions when the mean and standard deviation were fixed. In Table 1, we show the water pools we consider as well as the range of the means and standard deviations for each water pool.

These water pools were chosen based on the commonly used biophysical models for the water pools in brain tissue. In parametric models, white matter is modelled as a combination of the myelin and intra/extra axonal water pools (Raj et al., 2014a; Du et al., 2007; Yu et al., 2019; Chatterjee et al., 2018; Akhondi-Asl et al., 2014). Further, they consider a water pool which accounts for cerebrospinal fluid (CSF). The water in gray matter can be modelled as similar to the IES water pool, with an extended mean T_2. However, brain pathologies can result in T_2 distributions different from those of white matter, gray matter, and CSF; for example, (Laule et al., 2007) found that MS lesions can contain a water pool in the range between that of the IES pool and the CSF pool. We set the mean values in line with those reported in the literature (Mackay et al., 1994; Laule et al., 2007; Wansapura et al., 1999; Alonso-Ortiz et al., 2015). We included an extensive range for the standard deviations, ensuring that our dataset has both sparse, intermediate, and wide T_2 distributions in order not to bias our dataset towards any extreme. However, there can be partial volume effects, where different configurations of brain tissues are contained in a single voxel; for example, water from white matter and CSF could be present in a single voxel. Therefore, in our

dataset, we divide T_2 distributions in the brain into seven cases, each with a characteristic mixture of water pools.

- **White matter (WM)**: 2 Water Pools (Myelin and IES).
- **Cerebrospinal fluid (CSF)**: 1 Water Pool (CSF)
- **Gray matter** : 2 Water Pools (Myelin and GM)
- **Mixture of WM and CSF**: 3 Water Pools (Myelin,IES, and CSF)
- **Mixture of WM and GM**: 3 Water Pools (Myelin,IES, and GM)
- **Mixture of CSF and GM**: 3 Water Pools (GM and CSF)
- **Pathology**: 1 Water Pool (Pathology)

We note that as there is a small quantity of myelin in gray matter, the gray matter pool is composed primarily of the GM component in Table 1 as well as the myelin water component which is constrained to have a random v_i between 0 and 5%. Concretely, suppose we want to generate a random T_2 distribution for the case of a mixture of WM and CSF. This distribution is characterized by the combination of three water pools (myelin, IES, and CSF).

$$p(T_2) = \sum_{i=1}^{3} \frac{v_i}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(T_2-\mu_i)^2}{2\sigma_i^2}\right) \tag{7}$$

Therefore, by randomly selecting v_i from a Dirichlet distribution and uniformly sampling (μ_i, σ_i) for the three pools within the bounds in Table 1, we can generate a random T_2 distribution. Given the T_2 distribution, we use the EPG formalism to simulate the corresponding signal from an acquisition based on acquiring 32 echos with around 10 ms spacing between each echo. In the real data we use for our evaluation, three slightly different echo times are used; the *in vivo* scans of healthy subjects use an echo train of 10.68 ms, 21.36 ms,... 341.76 ms, the *in vivo* scan of the subject with pathology uses an echo train of 10.36 ms, 20.72 ms,..., 331.52 ms, and the *ex vivo* scan uses an echo train of 10 ms, 20 ms,... 320 ms. In the following, we describe our procedure with a single, fixed echo train: for the evaluation, we generated three training datasets, one for each echo train. We note that alternative sequences with different numbers of echoes/different spacings can be accommodated by generating a new dataset. We first construct a family of dictionaries of EPG signals, defined as in the previous section, (D_α), where we vary α from 90 to 180. We use a high resolution T_2 grid (1 ms to 2000 ms with a spacing of 0.1 ms) for the dictionaries. We generate 200,000 T_2 distribution variations per case by sampling (v_i) and (μ_i, σ_i) randomly from flat Dirichlet and uniform distributions, for a total of 1.4 million distributions. We randomly vary the flip angle (α) of the acquisition for each signal between 90 and 180 $^\circ$ so that our method learns to account for different flip angles automatically, rather than having to first estimate the flip angle as in non-parametric methods. Given $v_i, (\mu_i, \sigma_i)$, we numerically approximate the corresponding T_2 distribution on the same T_2 grid as used for the dictionaries (D_α). Let $\mathbf{p_{HR}}$ denote the discretized distribution. Given the pre-constructed dictionary of EPG signals D_α corresponding to the randomly chosen α, the EPG signal corresponding to this distribution, $\mathbf{s_{EPG}}$ is

$$\mathbf{s_{EPG}} = D_\alpha \mathbf{p_{HR}} \tag{8}$$

As noted in the previous section, non-parametric approaches commonly use a much coarser, logarithmically spaced grid of T_2s for the discretization of the distribution. This allows to significantly reduce the computation time. Therefore, to directly compare our approach with non-parametric approaches, we downsample the ground truth distributions from the high resolution T_2 grid to a grid of 60 T_2's logarithmically spaced from 10 ms to 2000 ms, as is commonly used Prasloski et al. (2012b). Denoting this downsampled, ground truth distribution as $\mathbf{p_{DS}}$, the dataset consists of the pairs $(\mathbf{s_{EPG}}, \mathbf{p_{DS}})$. We note that our method does not depend on this downsampling; we use it for a fair comparison with non-parametric approaches. As outlined in the related work, the SNR of the signals is a crucial aspect of the reconstruction and hence the dataset generation. We define SNR with respect to the first echo of the signal sequence. From previous studies (Wiggermann et al., 2020; Mackay et al., 1994), it is known that NNLS methods, perform well in the high-SNR regime (on the order of 1000). However, clinical scans with high spatial resolutions will rarely meet this SNR requirement; in the real scans of healthy subjects we use in our evaluation, we estimate a mean SNR on the order of 100. In order to make our method robust to the realistically low SNR regime, in training we randomly vary the SNRs of the signals between 80 and 200 in order to cover the potential SNR range of the voxels. We use a Rician noise model to add noise to the signals. In our evaluation, we show that training on this SNR range results in robustness to a wide range of SNRs (40-1000) on synthetic data. The data generation for 1.4 million signal/distribution pairs took less than one hour on a cluster using parallelization on 46 threads.

Using the synthetic datasets described, we train a MLP to map the MRI signal to the corresponding T_2 distribution.

2.2. Mapping the MR signal to the T_2 distribution

2.2.1. Architecture

Our network is composed of 6 hidden layers with 256 neurons per layer and an output layer with 60 units, corresponding to the size of the discretization of the distributions we use. The hidden layers use a ReLu function as the activation function, while the output layer uses a SoftMax activation function since the output should be the T_2 distribution. The input to the network is a vector with 32 elements corresponding to the 32 echos of the standard acquisition sequence. We note that we normalize the input by the magnitude of the first echo before feeding it to the network. To select the structure of the network, we trained 12 networks where we varied the number of hidden layers (3-6) and the number of neurons per layer (64,128,256,512). We selected 6 hidden layers and 256 neurons as this configuration had the lowest validation loss at the end of training; however, we note that the validation loss was not significantly different between the configurations.

2.2.2. Loss function

Let $(\mathbf{x}, \mathbf{p_x})$ denote the normalized MRI signal and the corresponding T_2 distribution. Let $\Phi(\cdot, \theta)$ denote the multi-layer perceptron function with parameters θ, with $\Phi(\mathbf{x}, \theta)$ the predicted distribution. Given a batch of training samples $(\mathbf{x_i})$ of size n, the cost function we use to train Φ is

$$L(\theta) = \frac{1}{n}\sum_{i}^{n} \lambda \|\mathbf{p_{x_i}} - \Phi(\mathbf{x_i}, \theta)\|_2^2 + W_1(\mathbf{p_{x_i}}, \Phi(\mathbf{x_i}, \theta)) \tag{9}$$

where the first term corresponds to the squared L_2 norm (MSE loss) and the second term corresponds to the Wasserstein-1 distance on probability distributions (Villani, 2009). We set λ to give approximately equal numerical weight to both terms in the loss function. Let u, v denote 1-D probability distributions with cumulative distribution functions U, V. Then the Wasserstein-1 Distance is equivalent to the following formulation (Ramdas et al., 2017)

$$W_1(u, v) = \int_{-\infty}^{\infty} |U(p) - V(p)| \mathrm{d}p \tag{10}$$

In this formulation, the Wasserstein distance can be efficiently computed on GPU using the cumulative sum function. The Wasserstein-1 distance is an appropriate metric to judge reconstruction quality in our application of T_2 distribution recovery as it correctly penalizes deviations from the ground truth distribution in relation to the location of the lobes in contrast to other losses such as MSE or Kullback-Liebler (KL) divergence. In particular, given

two non-overlapping lobes, if the lobes are moved toward each other (but still do not overlap), the Wasserstein Distance will decrease significantly while the MSE and the KL Divergence will not change. An example is presented in Fig 2. in the Supplementary material.

Using the Wasserstein distance helps us to avoid, for example, cases where the location of lobes in the distribution could be arbitrarily placed with a similar loss if other metrics are used. We note that training with either MSE loss or Wasserstein-1 distance exclusively leads to suboptimal results, due to increased Wasserstein-1 distance in the first case and unstable reconstructions in the second case. We find that training with a combination of these results worked optimally; we further show in our evaluation that adding the Wasserstein-1 distance improves the accuracy of MWF estimation in realistic cases in comparison to training exclusively with MSE loss.

2.2.3. Implementation details

We used TensorFlow 2.0 (Abadi et al., 2015) on Python 3.6 Van Rossum et al. (2000) with an Nvidia GTX 2070 laptop GPU for constructing and training the network. For each case, we use 80% of the generated data for training, corresponding to a total of 1,120,000 signal/distribution pairs. We reserve 10% of the dataset as the validation set and the remaining 10% as the test set in our evaluation on synthetic data. We use the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 1e-3 and a batch size of 2000. We trained for 30 epochs, where we stopped the training based on the validation loss oscillating/no longer decreasing. We use the epoch with the lowest validation loss as the final model. This training took approximately 70 seconds to complete, showing the feasibility, given a large database of signals, to retrain models specific to given sequences, etc.

3. Evaluation

We perform reconstructions of the T_2 distributions from synthetic and real data using the following methods:

- Our proposed method, MIML,trained on signals with SNR 80-200 and the appropriate sequence of echoes.
- NNLS with Tikhonov regularization (NNLS-T)
- NNLS with Laplacian regularization (NNLS-L)
- Gaussian Mixture Fitting (GMF)

Both NNLS methods were implemented in-house in Python with full parallelization, and we use a standard method for selecting the regularization parameter (Prasloski et al., 2012a; Laule et al., 2006) by keeping the signal fitting error close to 1.025 times the signal fitting error obtained using NNLS without regularization. GMF is our implementation of a parametric approach similar to that of (Raj et al., 2014a), where we fit a Gaussian mixture model with three compartments (Myelin water, IES water, CSF), extracting the volume fractions, the means/standard deviations of the T_2 of each compartment, and the overall normalization factor. We model as follows:

$$p(T_2) = \sum_{i=1}^{3} v_i \mathcal{N}(\mu_i, \sigma_i, T_2). \tag{11}$$

Then, the corresponding model signal is

$$\mathbf{s}^m(TE_i) = M_0 \int EPG(TE_i, T_1, T_2, \alpha)\, p(T_2) \mathrm{d}T_2. \tag{12}$$

where M_0 is the normalization constant. We calculate this integral numerically using a high resolution grid of T_2s as in the dataset construction. Given the experimental decay signal $\mathbf{s}$, the parameters $\mathbf{x} = ((v_i, \mu_i, \sigma_i), M_0)$ are calculated by solving the following optimization problem:

$$\mathbf{x} = \arg\min_{\mathbf{x}} \|\mathbf{s} - \mathbf{s^m}(\mathbf{x})\|^2 \tag{13}$$

where we constrain the μ_i, σ_i according to the bounds used for generating the dataset for MIML. Finally, v_i are constrained to the interval (0,1), and are normalized before each calculation of the model signal during the optimization. As jointly estimating the flip angle adds significantly to the computation time and contributes to instability, we fix the flip angle in the Gaussian mixture fitting for each voxel to that calculated using a standard NNLS method (Prasloski et al., 2012a). We validated the accuracy of this flip angle estimation by comparing against B1 maps acquired on healthy subjects. We used the least squares optimization function in the Python library Scipy (Virtanen et al., 2020) to fit the signals to the Gaussian model.

3.1. Synthetic data

3.1.1. Test split of synthetic dataset

We show reconstructions on the test split of the synthetic dataset we generated using the acquisition sequence of 10.68 ms, 21.36 ms,... 341.76 ms. We show results over an SNR range from 40 to 1000 (40,80,150,200,400,1000). We compare the methods using the MSE and Wasserstein Distances of the reconstructed distributions with respect to the ground truth distributions.

3.1.2. Realistic synthetic case in WM

MWF mapping is a crucial application of T_2 relaxometry. In order to analyze the robustness and performance of our approach in a realistic case in WM, we show reconstructions on the following model of the distribution in a white matter voxel, with one lobe for myelin water and one lobe for IES water.

$$p(T_2) = v_m * InvGamma(\mu_m, \sigma_m) + v_{IE} * InvGamma(\mu_{IE}, \sigma_{IE}) \tag{14}$$

where we fix the values of the parameters to realistic values in line with those reported in the literature (Alonso-Ortiz et al., 2015; MacKay and Laule, 2007): v_m=0.15, v_{IE}=0.85, $\mu_m = 20$ ms, $\mu_{IE} = 70$ ms, $\sigma_m = 2.5$ ms, $\sigma_{IE} = 6$ ms. We use the inverse Gamma distribution to create the ground truth distribution to test the robustness of our method to changes in the assumed biophysical model. To study robustness to noise, we vary the SNR on the corresponding synthetic MRI signal from 40 to 1000, as in the test split. We generate 1000 realizations of noisy signals per SNR used. Further, we also show numerical results using our method **without** using the Wasserstein Distance in the loss function. We refer to this variant as MIML'. We compare the methods using the MSE, Wasserstein Distance, and estimated MWF of the reconstructed distributions with respect to the ground truth.

3.2. Real data

As there is no ground truth for the T_2 distributions in real data, we evaluate the methods as in the literature by examining the MWF maps/comparing to anatomical scans or correlation to histology, the plausibility of the T_2 distributions, maps of the mean T_2 in the 50–200 ms range, etc. We also report the mean SNR for each dataset, calculated in the same manner as in Wiggermann et al. (2020), where the first echo of the signals is divided by the standard deviations of the residuals from the NNLS-T reconstruction.

3.2.1. Ex vivo data

We show reconstructions from a Multi Echo Spin Echo (MESE) scan from the White Matter Microscopy Database Cohen-Adad et al. (2020) with 32 echoes (starting from 10 ms with 10 ms

spacing), with a TR of 3s and 8-fold averaging, of a single, cervical slice of a dog's spinal cord acquired *ex vivo* with an Agilent 7T animal scanner (Vuong et al., 2017). Five days before scanning, the spinal cord (perfused and post-fixed with paraformaldehyde 4) was extracted and washed in Phosphate-buffered saline (PBS) solution. After scanning, the spinal cord was osmified for two hours, embedded in EMbed 812 Resin, cut using a microtome, and polished. A scanning electron microscope (Low-angle backscattered electron mode) (JEOL 7600F) was used to image an entire slice of the spinal cord at a resolution of 0.26 micrometers per pixel. Using this histology image, we construct a histological map of the fraction of myelin in each voxel using a deep learning segmentation tool called Axon Deepseg (Zaimi et al., 2018). We then register this histological map to the MRI space. The resulting histological map is a map of the fraction of the voxel corresponding to the segmented myelin, not a map of the MWF. However, assuming that the fraction of myelin in a voxel scales with the amount of myelin water, the two maps should be linearly correlated. We conduct a correlation analysis between the histological map and the MWF maps produced from the different methods. The estimated SNR on this slice is 784.

3.2.2. Healthy subjects

We show reconstructions from high-resolution human brain scans acquired from 4 healthy controls using a 3T MRI scanner (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) located at CHUV Hospital (Lausanne, Switzerland), with a standard 64-channel head/neck coil. The dataset was collected using a 3D multi-echo gradient and spin-echo (GRASE) sequence accelerated with CAIPIRINHA (Piredda et al., 2020a) with the following parameters: matrix-size=144 $\times$ 126; voxel-size = 1.6x1.6x1.6mm3; ΔTE/N-echoes/TR = 10.68 ms/32/1 s; prescribed FA =180°; number-of-slices = 84; CAIPIRINHA acceleration factor = 3x2; number of averages = 1; acquisition time=10:30min. Each subject was also scanned using an MPRAGE sequence for whole-brain T_1-weighted imaging (Brant-Zawadzki et al., 1992). To test the repeatability of the reconstructions, the healthy controls were scanned twice over two consecutive scanning sessions (scan-rescan scenario). We compare the MWF maps and the T_2 distributions produced from each method, show the coefficient of variability of the MWF in regions of interest (ROI) in WM, and conduct a study of the reproducibility of each method. The data for these subjects have an estimated mean SNR of 128.

3.2.3. MS subject

We show reconstructions on a high-resolution human brain scan of a patient with relapsing-remitting multiple sclerosis, scanned using a 3T MRI scanner (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) located at the University Hospital of Basel (Basel, Switzerland) with a standard 32-channel head coil. In this case, MET2 data was collected using the previously described GRASE sequence for the healthy subjects, albeit with a starting echo time of 10.36 ms and lower spatial resolution (voxel-size=1.8x1.8x1.8mm3) to accelerate the scan. In addition, a FLAIR (De Coene et al., 1992) scan was acquired. A probabilistic lesion mask was generated by first using a convolutional neural network (CNN) trained to segment WM lesions (La Rosa et al., 2020) on FLAIR images with subsequent manual correction by an expert. The FLAIR image/lesion mask were then registered to the multi-echo T_2 space. We use a threshold of 0.9 to denote a voxel as lesional. We analyze maps of the geometric mean T_2 in the range 50–200 ms and MWF maps to study the MS lesions as in (Levesque et al., 2010). We also compare the correspondence of these maps to the lesion masks. In addition, we compare the T_2 distributions produced from each method in both normal-appearing tissue and the lesions. The estimated SNR of this scan is 112.

4. Results

4.1. Synthetic data

4.1.1. Test split of synthetic dataset

In order to visualize the average performance over the test split, in Fig. 2 we plot the mean distribution over all the ground truth distributions in the test split. In addition, we show the mean reconstructed distributions over the test split from the methods we compare. We also show plots zooming in on the different T_2 regions for better visualization. MIML performs robustly and consistently across the whole SNR range, providing the best conformity to the ground truth distributions over the entire range of T_2s**.** In contrast, NNLS with Tikhonov and Laplacian regularization both require SNR 1000 in order to generate a plausible distribution in the T_2 range $10-50$ms, with SNRs below this resulting in highly over-smoothed distributions. Further, even at high SNRs, both methods have over-smoothing in the T_2 range $50-2000$ms. For the Gaussian mixture fitting, we note that only the cases of WM and WM + CSF correspond to the model used, as it is necessary to fix the number of compartments beforehand. Therefore, the relevant T_2 ranges to examine are $10-120, 1000-2000$ ms. For GMF, high SNRs (200-1000) are required for plausible distributions with respect to the ground truth, with remaining distortions at low T_2 values**.** In Fig. 3 we show boxplots of the MSE and the Wasserstein Distance between the ground truth distributions and the reconstructed distributions from the different methods over the SNR range. As the model used in GMF only applies to WM and WM+CSF, we show the results over the whole test set as well as over just the WM and WM+CSF cases in the test set.

For both MSE and Wasserstein Distance, MIML performs the best with the lowest median error and comparable or smaller interquartile ranges, across the whole SNR range. As expected, all methods improve with increasing SNR. The limitations of the GMF model are clear, as it provides competitive results with the other methods only when restricted to the signals from the WM and WM+CSF cases, due to the need to fix the model/number of compartments beforehand.Overall, MIML, which is trained on signals with SNR 80 to SNR 200, generalizes well to the test set as well as to SNRs outside the range on which it was trained. From the plots of the mean distributions and the boxplots of the error metrics, we can see that MIML performs better in distribution reconstruction than the other methods, parametric and non-parametric, across a wide range of SNRs. In addition, the flexibility of using MIML in comparison to GMF is clear, as MIML does not require fixing the number of compartments**.** However, the test set is generated according to the Gaussian mixture model; further, as we randomly generate the ground truth distributions, not all of the ground truth distributions are realistic, though we note that unrealistic distributions in the training can improve the generalizability of MIML.

4.1.2. Realistic synthetic case in WM

In Fig. 4, we plot the ground truth distribution and the mean reconstructed distributions from each method over the SNR range. In addition, we show boxplots of the MSE and Wasserstein Distance between the ground truth distribution and the reconstructed distributions from the different methods over the SNR range. Further, we show a boxplot of the error in MWF estimation. MIML performs robustly and consistently, on average, across the whole SNR range**.** However, the reconstructed distributions resolve a more spread out myelin water lobe than in the ground truth, even at SNR 1000; this could be due to training on significantly lower SNRs or the model mismatch. At SNRs below 400, NNLS-T and NNLS-L are unable to resolve a myelin water lobe due to over-smoothing as well as a displaced IES lobe; GMF resolves the myelin water lobe,

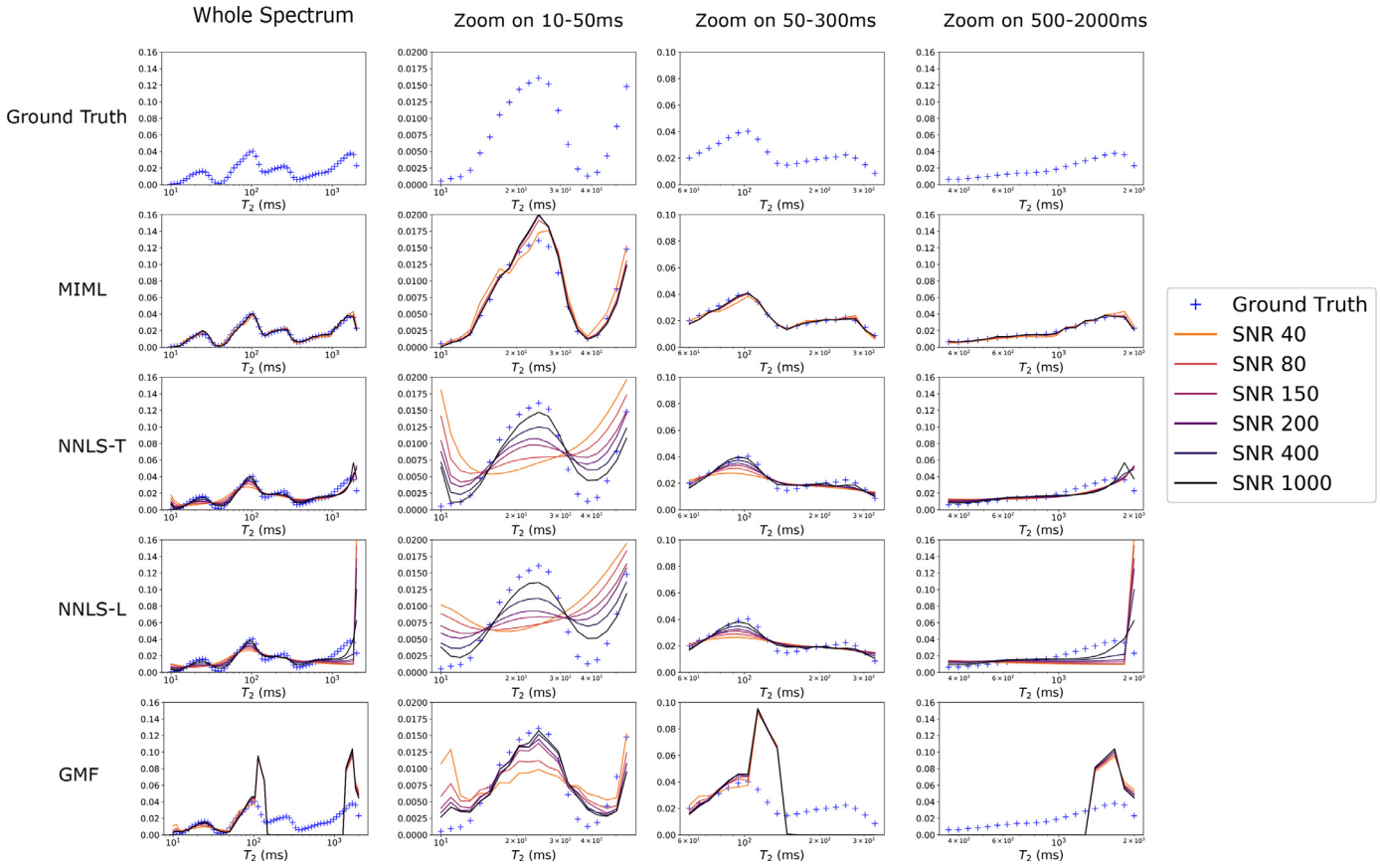

Fig. 2. Plots of the mean distribution over all of the ground truth distributions in the test split of our synthetic dataset, as well as the mean reconstructed distribution over the test split from each method. We zoom in on the T_2 ranges 10–50 ms, 50–400 ms, 400–2000 ms to show the average performance in the different cases (WM, CSF, etc.). Our method produces the most robust and accurate reconstructions with respect to changing SNR and the ground truth distributions respectively. All other methods require high SNRs (400–1000) for plausible distributions that, however, still retain distortions, particularly in the range of T_2s associated with myelin water (10–50 ms). We note that the poor performance of GMF outside the T_2 range 10–120 ms, and 1000–2000 ms is due to model mismatch; GMF is valid only for the WM and WM+CSF cases. We use a logarithmic scale for the T_2 axis.

but with a significantly displaced mean. At SNR 1000, NNLS-T and NNLS-L are able to resolve the myelin water lobe accurately, albeit still with a small distortion at $T_2 = 10$; GMF is able to accurately capture the myelin water lobe at SNR 1000, albeit with a displaced IE lobe.

With regard to MSE and Wasserstein Distance, MIML performs the best, with the lowest median error and comparable or smaller interquartile range across the whole SNR range. As expected, MIML' performs similarly to MIML with respect to MSE and significantly worse with respect to Wasserstein Distance, as it is only trained with the MSE loss. With regard to the estimated MWF (obtained by summing from T_2 bounds of 10–40 ms), we see that MIML performs the best in the SNR range 80-400, with median errors closest to zero, and comparable or smaller interquartile ranges; we remind that the ground truth MWF was 0.15. At SNR 40, all methods either significantly over or underestimate the MWF, while at SNR 1000, MIML and NNLS-L provide comparable median errors. However, we note that NNLS-L has a significantly higher standard interquartile range than MIML at SNR 1000. Further, the results are consistent with results in (Wiggermann et al., 2020) that the NNLS methods tend to underestimate the MWF. MIML' provides mediocre performance, generally underestimating the MWF value. Comparing the performance of MIML and MIML', we can see that using the Wasserstein Distance in the loss function during training significantly improves the performance of our method in terms of MWF estimation in a realistic case as well as the Wasserstein Distance of reconstructed distributions to the ground truth. Finally, in Fig. 5, we show the reconstructed distributions and the mean distribution for each method for SNRs of 200 and 1000. Although the mean distribution from NNLS-T and NNLS-L corresponds well to the ground truth distribution at SNR 1000, the reconstructions of NNLS-T and NNLS-L are highly sensitive to added noise, with huge variability in the reconstructed distributions, particularly in the myelin lobe**.** In contrast, MIML, and to a lesser extent, GMF, are much more robust to the noise, showing little variability in the reconstructed distributions. Overall, MIML performs accurately and robustly across the SNR range with respect to the MSE, Wasserstein Distance, and the MWF value, showing the robustness to changing the assumed Gaussian model for the distribution as well as the applicability in a realistic case. Other methods perform comparably, on average, at high SNR values (SNR 1000), as expected.

From the results on the synthetic data, we conclude that MIML, even trained on a limited range of SNRs, is able to robustly and accurately reconstruct T_2 distributions over a wide range of SNR values. Overall, MIML outperforms all other methods in terms of MSE and Wasserstein Distance with respect to the ground truth. Furthermore, from the realistic case, MIML is the most accurate overall method for MWF estimation, showing the applicability to MWF estimation. In addition, we can see the robustness to changes in the assumed model for the T_2 distributions, and the importance of including the Wasserstein Distance in the loss function of MIML. Finally, from examining all the reconstructed distributions and the mean reconstructed distribution, MIML is the most robust to noise,

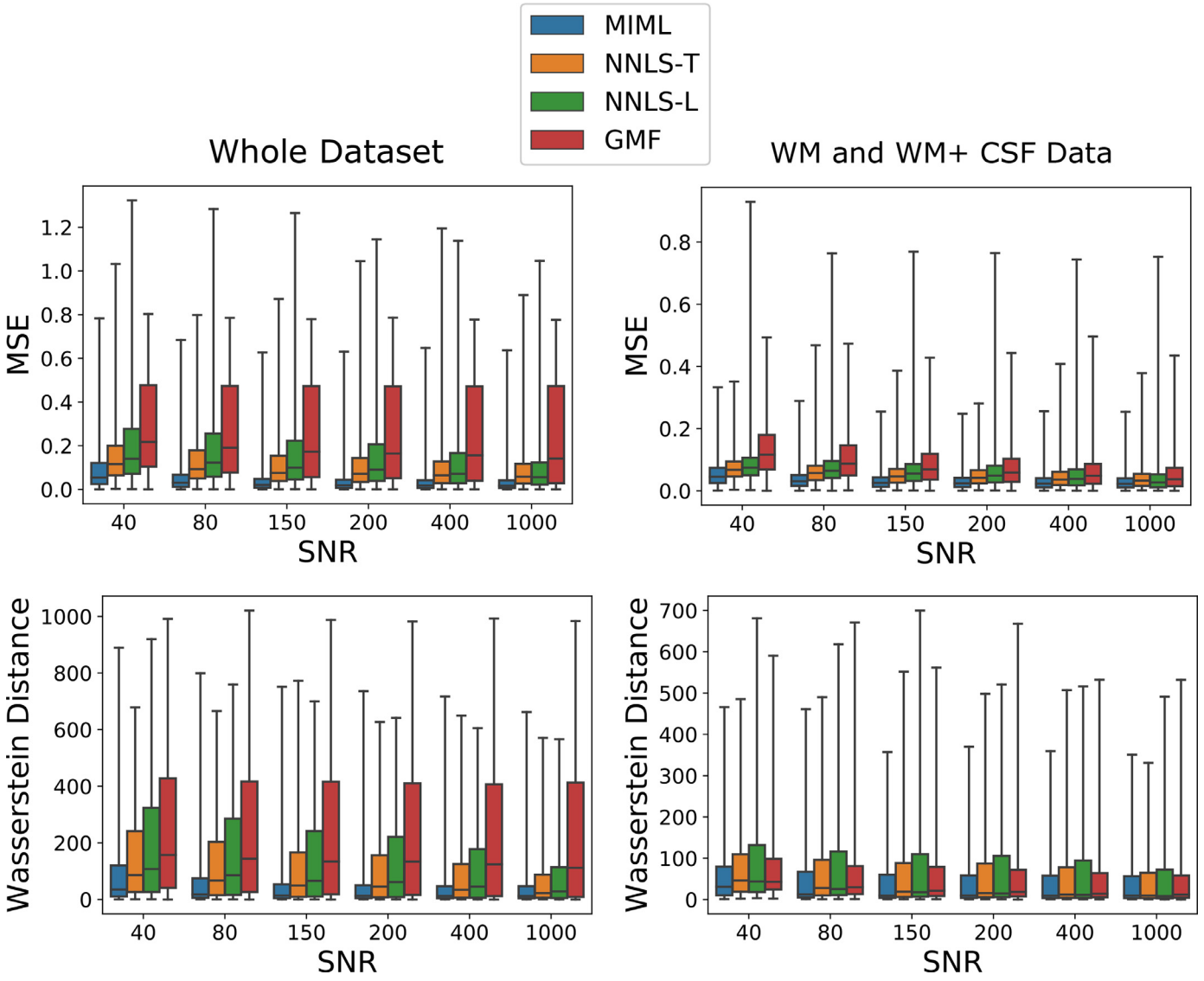

Fig. 3. Boxplots of the MSE and Wasserstein Distance between the ground truth distributions in the test set of our synthetic dataset and the corresponding, reconstructed distributions from each method over a range of different SNRs. We show results over both the whole test set, as well as results restricted to the WM and WM+CSF cases, where the GMF model is valid. In general, MIML provides the most accurate and robust reconstructions, with the lowest median errors as well as lower or comparable interquartile ranges. As expected, the performance of GMF becomes comparable to other methods when we restrict to only the WM and WM+CSF cases, where the GMF model is valid.

while the non-parametric methods show high variability and sensitivity to noise even at SNR 1000**.** In the next section, we show results on real data from *in vivo* and *ex vivo* scans, considering both healthy and pathological cases.

4.2. Real data

4.2.1. Ex vivo Data

We note that in Equation (4), the MWF is obtained by summing from $T_2 = 10$ms to $T_2 = 40$ms. This formula, commonly used for acquisitions at 3T, in theory should be adjusted for higher field strengths due to the shortening of T_2s (Kolind et al., 2009; Laule et al., 2008). We note that these limits historically derive from assignment of the different lobes in T_2 distributions to different water pools e.g. myelin, IE space, etc.(Alonso-Ortiz et al., 2015). For instance, in (Mackay et al., 1994), the authors use NNLS-T on their data (acquired at 1.5T) and found two large T_2 lobes, one in the range of 10–50 ms and the other in the range of 70–100 ms; they then assigned these to myelin water and the IES water respectively. In the following, we restrict our analysis to the white matter, and we will show two versions of MWF maps, with accompanying correlations to histology obtained as follows:

- Fixed Limits: Following (Mackay et al., 1994), we fix the limits of summation for each method by taking the limits of the myelin water lobe in the mean T_2 distribution from using NNLS-T. This corresponds to bounds of 10–35 ms.
- Tailored Limits: For each method, we set the limits of summation from the limits of the low T_2 lobe in the mean T_2 distribution from that method. For MIML and NNLS-L this corresponds to bounds of 10–38 ms and 10–32 ms respectively.

In Table 2, we show the spatial Pearson correlations (with accompanying p-values) between the MWF maps for each method and the histology map. In both cases, the MWF map from MIML has the highest correlation to the histology map**.** Only the correlation of NNLS-L changes between the two cases, increasing when using the fixed bounds. In Fig 6, we show the MWF maps corresponding to each case for the bounds, the histology map, and the

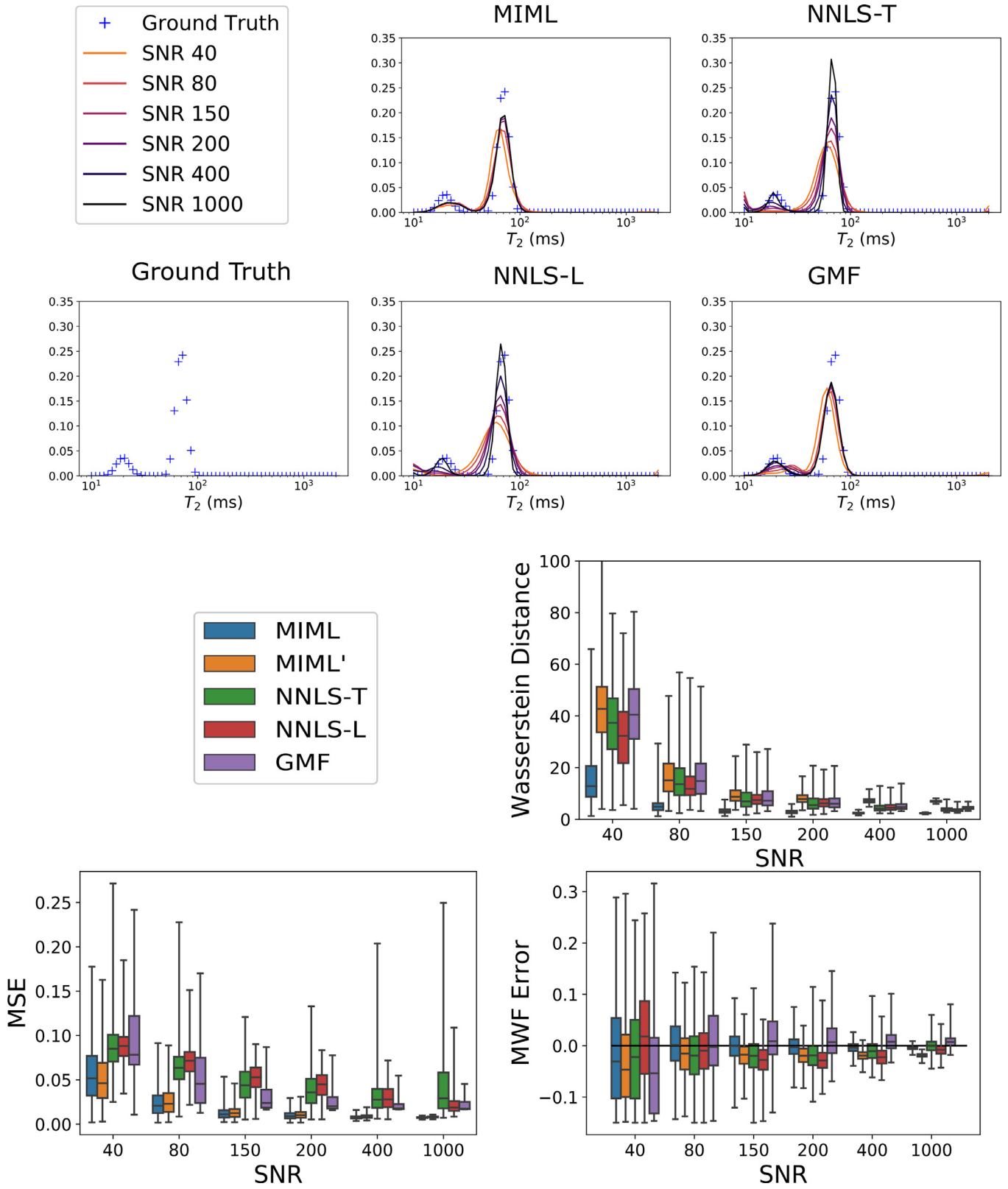

Fig. 4. Mean reconstructed distributions (ground truth and from each method) over a range of SNRs as well as boxplots of the MSE and Wasserstein distance between the ground truth and reconstructed distributions from the results on the realistic, synthetic case. MIML produces the most robust reconstructions with respect to changing SNR, albeit with a consistently over-smoothed myelin water lobe. However, the other methods require high SNRs (1000) to resolve a myelin water lobe (still with distortions) close to the ground truth lobe as well as correct placement of the IE lobe. With regard to MSE and Wasserstein Distance, MIML performs the best, with the lowest median error and comparable or smaller interquartile range across the whole SNR range. With regard to MWF error (the ground truth MWF value is 0.15), MIML performs the best in the SNR range 80-400, with median errors closest to zero, and comparable or smaller interquartile ranges. MIML' performs similarly to MIML with respect to the MSE and significantly worse with respect to the Wasserstein Distance and MWF Error, showing the importance of using the Wasserstein Distance in the training of our method.

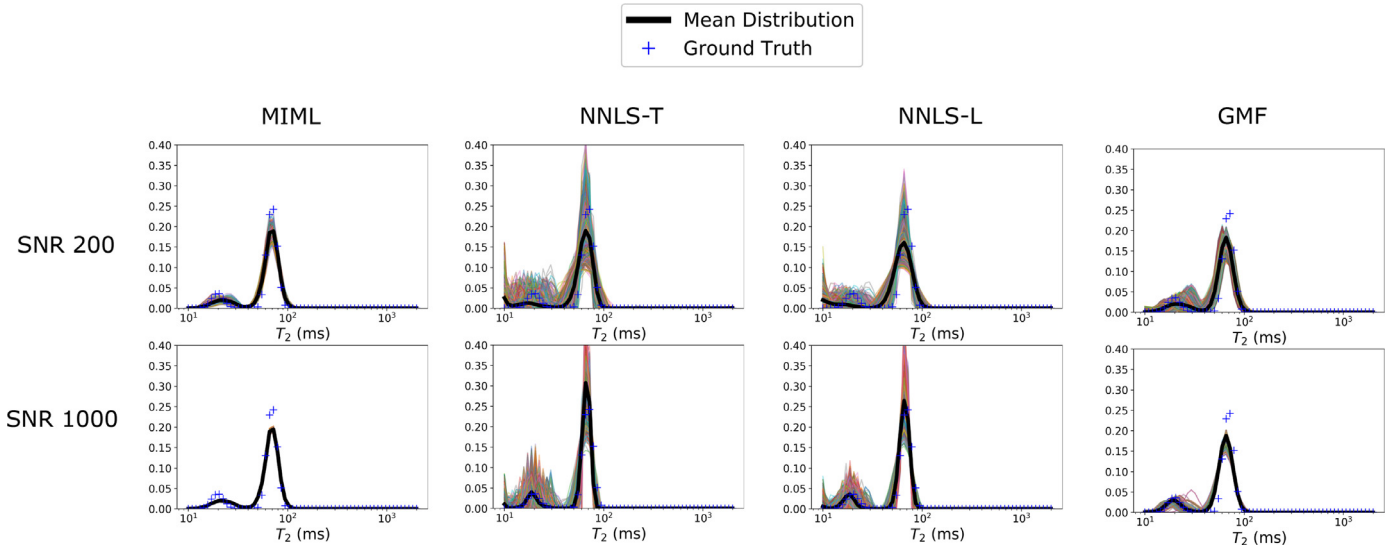

Fig. 5. Distribution reconstructions from different noise realizations for SNRs 200 and 1000 on the realistic, synthetic case. The ground truth distribution is shown with blue crosses, and all the corresponding reconstructed distributions from different noise realizations are shown in color. The mean reconstructed distribution is shown in black Note that the reconstructions of NNLS-T and NNLS-L are highly sensitive to noise, even at SNR 1000, with large variability in the reconstructed distributions; this is in contrast to the stability and robustness of the reconstructions from MIML and, to a lesser extent, GMF. Note that for SNR 1000, MIML predicts virtually the same distribution for all the noisy signals. We use a logarithmic scale for the T_2 axis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Table of the spatial Pearson correlations (with p-values) between the MWF maps constructed from each method and the histology map of the myelin in a white matter mask. In bold are the highest correlations. We note that the histology map is not a map of the myelin water fraction, but a map of the fraction of pixels in the histology which correspond to the myelin tissue. In either case of fixed or tailored bounds, the MWF map from MIML has the highest spatial correlation to the histology.

Pearson Correlation of MWF Maps to Histology				
	MIML	NNLS-T	NNLS-L	GMF
Tailored Bounds	(**0.54**,5.63E-81)	(0.44,8.58E-53)	(0.45,5.51E-55)	(0.39,1.43E-39)
Fixed Bounds	(**0.54**,2.04E-81)	(0.44,8.58E-53)	(0.49,1.13E-64)	(0.39,1.43E-39)

reconstructed distributions for each method. MIML predicts higher values for the MWF than the other methods, particularly the NNLS methods. The MIML MWF map is smoother/less noisy than the other methods and corresponds better to the histology map. We can see from the mean distributions that all methods are, on average, able to recover the myelin water and IES water lobe in similar locations; however, the NNLS methods, in particular NNLS-L, produce more implausible, over-smoothed lobes in comparison to MIML and GMF. Examining the reconstructed distributions, the influence of the model priors in MIML and GMF is evident, with clear separation between the myelin lobe and the IES lobe, while the NNLS methods produce distributions which are spread more uniformly across the T_2 axis**.** We note that the small number of distributions with lobes in the range 200–1000 ms and the lobes in the range 1000 ms–2000 ms can be attributed to the gray matter around the spine as well as CSF.

For all methods, the MWF values are significantly higher than those of the *in vivo* 3T scans we show later**.** However, this could be attributed to the differences resulting from the fact that *ex vivo* scan is of chemically treated spinal cord at 7T while the *in vivo* scans are of human brain at 3T.

4.2.2. Healthy subjects

In Fig. 7 we show the MWF maps in axial, coronal, and sagittal slices for two healthy subjects with corresponding, registered MPRAGE images for comparison. In MPRAGE images, WM is hyperintense; hence, we treat the MPRAGE as a very rough proxy for the MWF map since MWF values are highest in the WM. Although the MWF maps are fairly similar, the MWF map of MIML most accurately and smoothly conforms to the MPRAGE image**.** The NNLS methods exhibit higher distortions, e.g. in the ventricles of subject 1, and difficulty in recovering the MWF in the frontal region of the brain. GMF produces maps comparable to the NNLS methods, albeit, looking noisier. We note that all methods exhibit lower MWFs in the frontal part of the brain as compared to other regions, which may stem from effects due to the gradient echo acquisition (Alonso-Ortiz et al., 2017). In Fig. 8, we show the reconstructed distributions over the WM voxels in the axial slices. Only MIML produces a mean WM distribution with two distinct, well-separated lobes corresponding to myelin water and the IES water as is expected from previous studies. Further, the peaks of the myelin water lobe and the IES water lobe correspond to the range expected at 3T. The NNLS methods recover the IES water lobe in line with expectations, but over-smooth the distribution in the region corresponding to myelin water, as was seen in the results on the synthetic data, with an implausible myelin water peak at 10 ms. GMF also recovers the IE lobe in line with expectations, but produces a dispersed lobe in the myelin region. From the reconstructed distributions, we can again see the influence of the model priors on MIML and GMF, with the NNLS methods producing much more variable distributions. We note that the small component in the range 1000 ms–2000 ms for each method can be attributed to partial volume effects with the CSF.

In order to compare the MWF maps on regions of interest, and to conduct the scan-rescan analysis we did the following: in a first step, all the estimated MWF images for the 4 subjects were registered to the 'ICBM-DTI-81' white-matter tract labels atlas (Oishi et al., 2008; Mori et al., 2008) using the non-linear

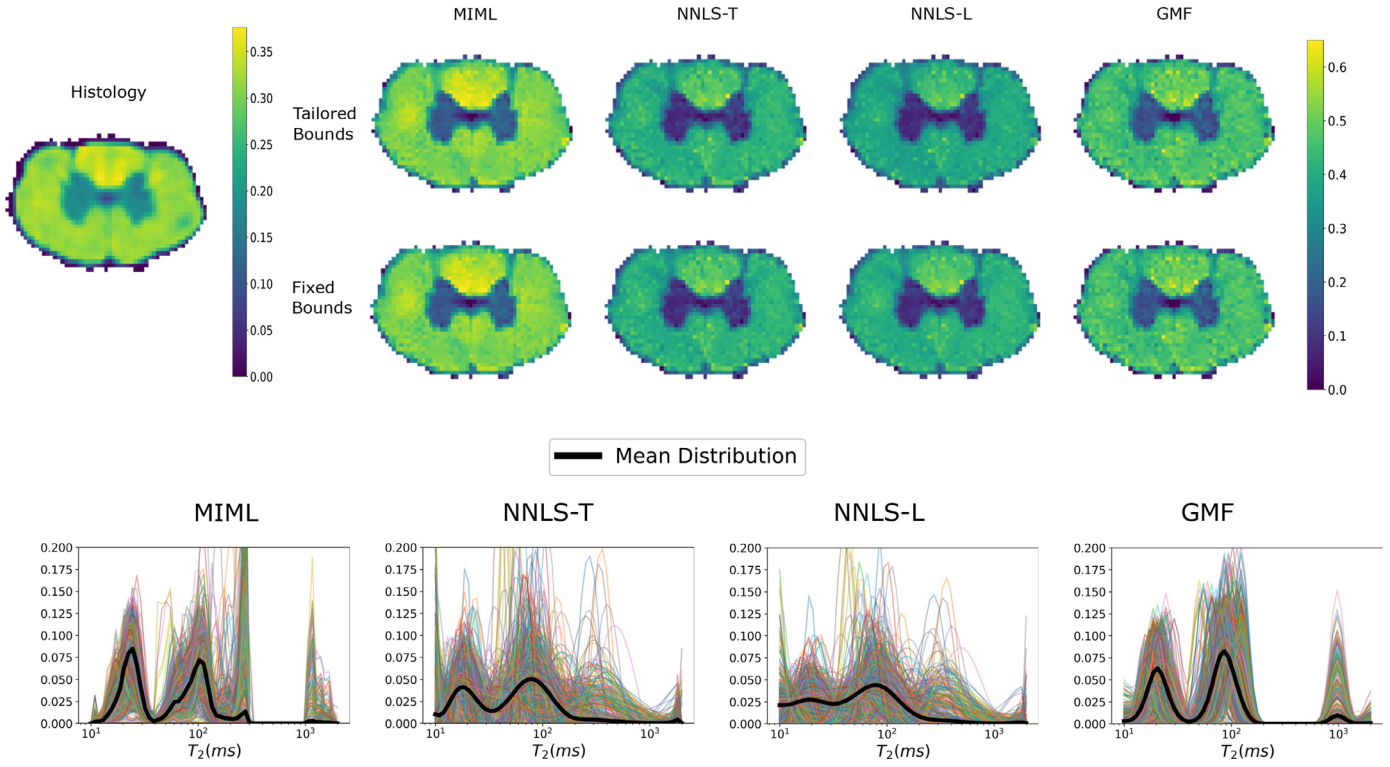

Fig. 6. MWF maps from each method, the histology map, and the reconstructed distributions for each method on the ex-vivo data. We can see that the MIML MWF map is smoother than those of other maps and corresponds better to the histology map. The mean distribution is shown in black, and all the reconstructed distributions in color. All methods can recover the myelin water and IES water lobe in similar locations; however, the NNLS methods produce more smooth lobes in comparison to those of MIML and GMF. The effect of the model prior on MIML and GMF is clear, with unambiguous separation between the myelin lobe and the IES lobe. We emphasize that the histology map is a map of the fraction of the voxel which is occupied by myelin, not the MWF. We use a logarithmic scale for the T_2 axis.

registration 'BSplineSyN' algorithm included in the ANTs software (https://github.com/ANTsX/ANTs). After visually inspecting the images, we removed small ROIs affected by registration errors and kept 44 tract labels showing a good anatomical agreement between the atlas and subject native spaces. Finally, the mean MWF value and the coefficient of variation of the MWF for each region of interest (ROI) was calculated for the scan and rescan maps from each method. A list of the ROIs can be found in Table 1 of the Supplementary material. In Fig. 9, we show boxplots of the mean MWF and the standard deviation of the MWF over the WM ROIs for each subject. We note that MIML results in a larger mean MWF across all subjects than the non-parametric methods; this can be explained by the underestimation of MWF by the non-parametric methods as is shown in the results on the realistic synthetic data as well as in (Wiggermann et al., 2020). However, the standard deviations of the MWF from MIML are generally comparable to that of the other methods, similar to that of NNLS-T and slightly higher than that of NNLS-L. In particular, the magnitude of the increase in the mean MWF using MIML as compared to the other methods is larger than the increase in the standard deviation of the MWF. This indicates that the smoothness of the MIML MWF map is comparable to that of other methods.

In Tables 3 and 4, we show the results of our scan-rescan analysis over all four healthy subjects; we show a table of the mean and standard deviation of the absolute difference between the mean MWF values of the scan and rescan in the specified ROIs as well as a table of the Pearson correlation and linear regression coefficients between the mean MWF values of the scan and rescan in the specified ROIs. We can see that in general, GMF provides the smallest mean differences and highest Pearson correlations. In particular, it is difficult to rank MIML and the NNLS methods as they perform better/worse on different subjects. We note that GMF's superior reproducibility may stem from the lower flexibility in the fitting of the MWF, as compared to MIML and the NNLS methods. However, overall, the reproducibility of the methods is quite similar.

4.2.3. MS subject

In Fig. 10, we show the maps of the geometric mean T_2 in the IE range of 50–200 ms as well as the MWF maps in an axial slice of a subject with MS. In addition, in Fig. 3 of the Supplementary Material, we zoom in on the lesions for better visualization. For the mean T_2 maps, in all methods, all except one of the lesions can be clearly seen as hyperintensities i.e. with increased mean IE T_2. Further, the maps are similar across the methods, with the main differences residing in the ventricles**.** Visualizing the lesions is far more difficult with MWF maps than with the mean T_2 maps, as the MWF maps are much noisier independent of the applied method. However, as with the healthy subjects, the MIML MWF map in both slices most smoothly and accurately conforms to the WM and the cortices, with the other methods exhibiting more variability and missing patches in the WM and worse delineation of the cortices; this occurs particularly in the frontal region. All three lesions can be seen on the MIML MWF map with minimal ambiguity; in particular, in lesions 1 and 3, we can clearly delineate the lesions from very close, adjacent structures. Concerning the NNLS methods, it appears that Lesion 1 is exaggerated in size and mixed with the adjacent structure, making it difficult to delineate the lesion as the dark region is extended far beyond the lesion region on the FLAIR image. In addition, due to poor contrast between the normal-appearing tissue and lesion tissue/noise, it is difficult to identify Lesion 2 unambiguously with the NNLS methods. As with Lesion 1, Lesion 3 can be seen but is connected to the adjacent grey matter, making localization problematic. Further, we can see that the MWF in the lesion is comparable to the

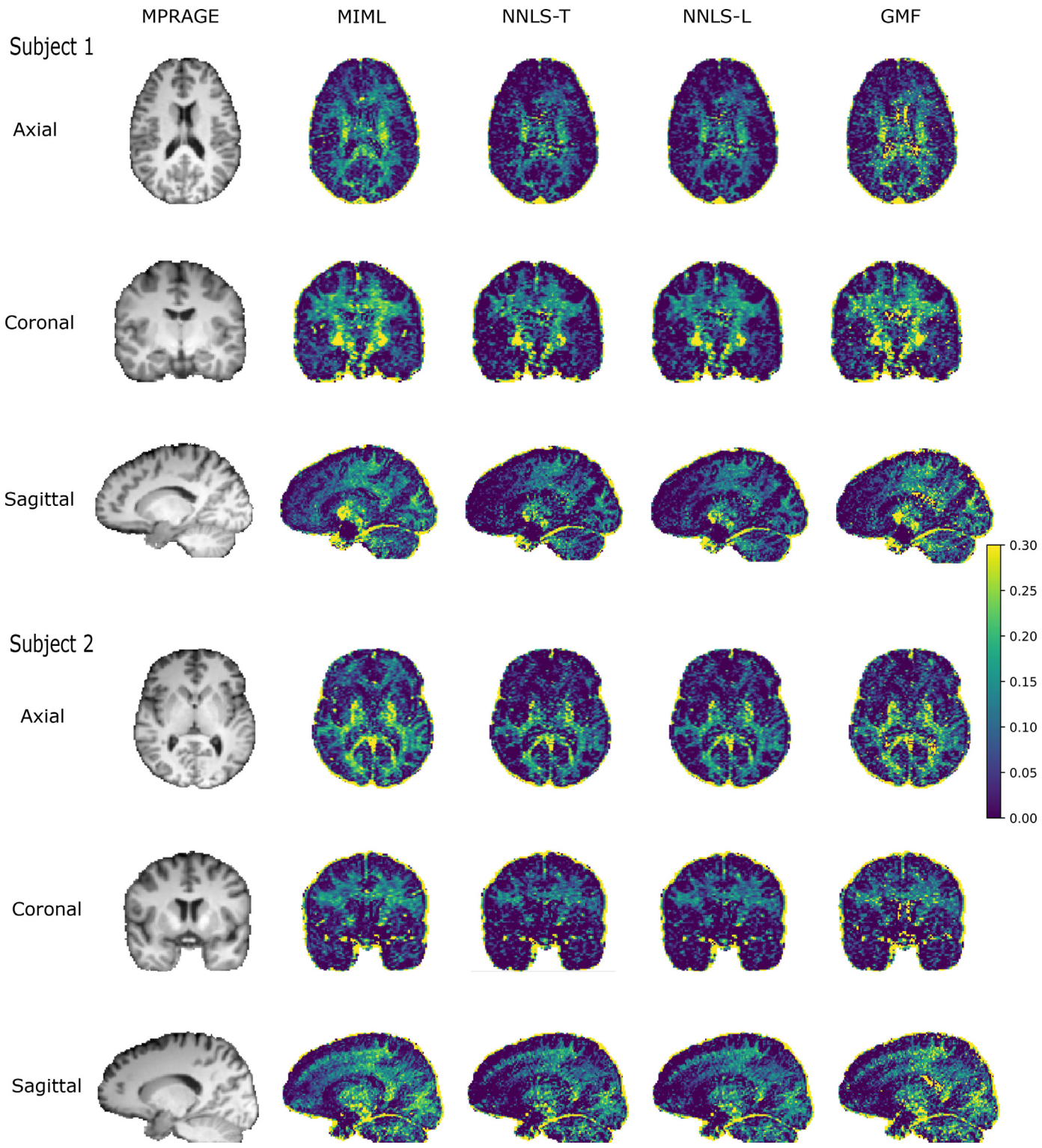

Fig. 7. Example MWF maps produced from each method, in the axial, coronal, and sagittal planes of two healthy subjects. On the left, we show the corresponding MPRAGE slice. Compared to the MPRAGE (where WM is hyper-intense), we can see that MIML most accurately and smoothly reproduces the extent of white matter, which is consistent with WM having relatively high MWF values. Particularly, the NNLS methods struggle in MWF recovery in the frontal part of the brain. GMF produces comparable to better MWF recovery than the NNLS methods, but with a noisier map. In addition, MIML has the least distortion in the ventricles.

MWF of the normal-appearing, contralateral brain region, due to the poor MWF reconstruction. The GMF MWF map resembles the MIML MWF map albeit noisier/ with greater variability.

In addition to the mean T_2 and MWF maps, in Fig. 4 of the Supplementary Material, we compare the T_2 distributions in the lesion masks to the T_2 distributions in the normal appearing, contralateral regions. As in the healthy subjects, MIML consistently produces a mean distribution with two distinct, well-separated lobes corresponding to myelin water and the IES water as is expected from previous studies**.** Further, the peaks of

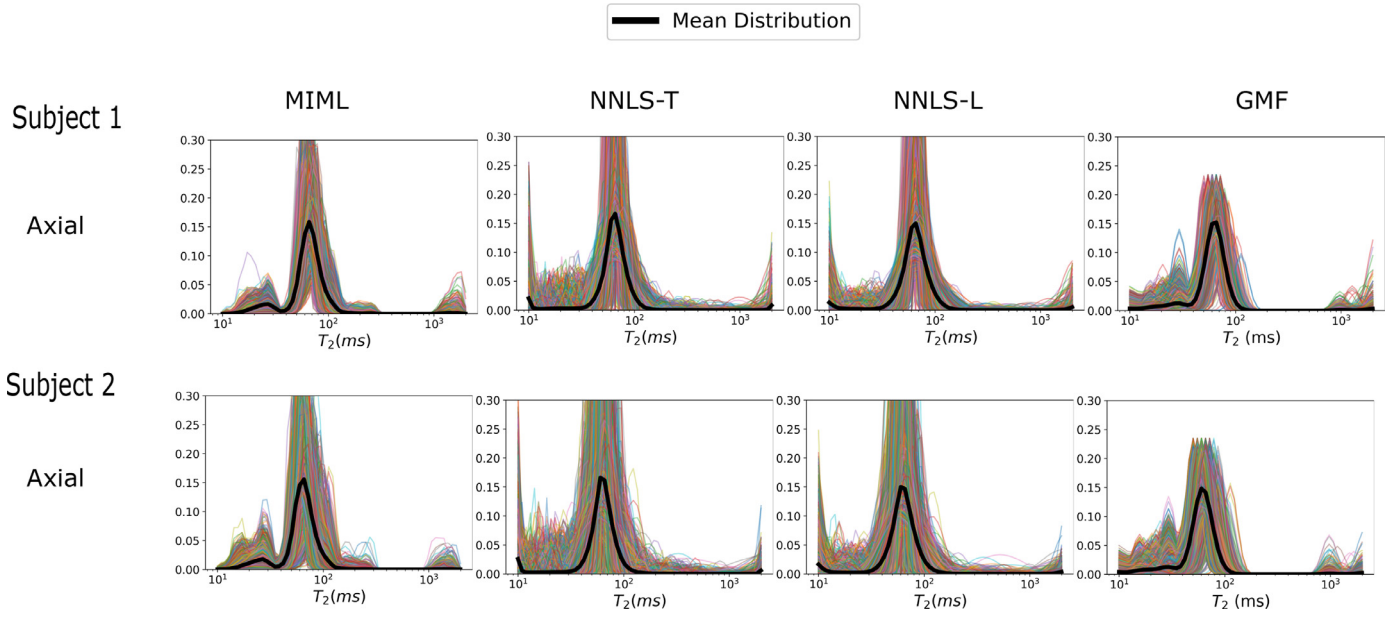

Fig. 8. Reconstructed distributions (in color) in the WM voxels of the axial slices of two healthy subjects for each method. The mean distribution is shown in black. We note that only MIML produces a mean WM distribution with two distinct, well-separated lobes, and the myelin water peak in line with expectations at 3T. The NNLS methods and GMF recover the IE lobe well, but the myelin lobe is either irregular or appears at an extremely low T_2. Further, we see that NNLS-T/L produces a much more variable set distributions in contrast to those from MIML and GMF which are constrained by model priors. We use a logarithmic scale for the T_2 axis.

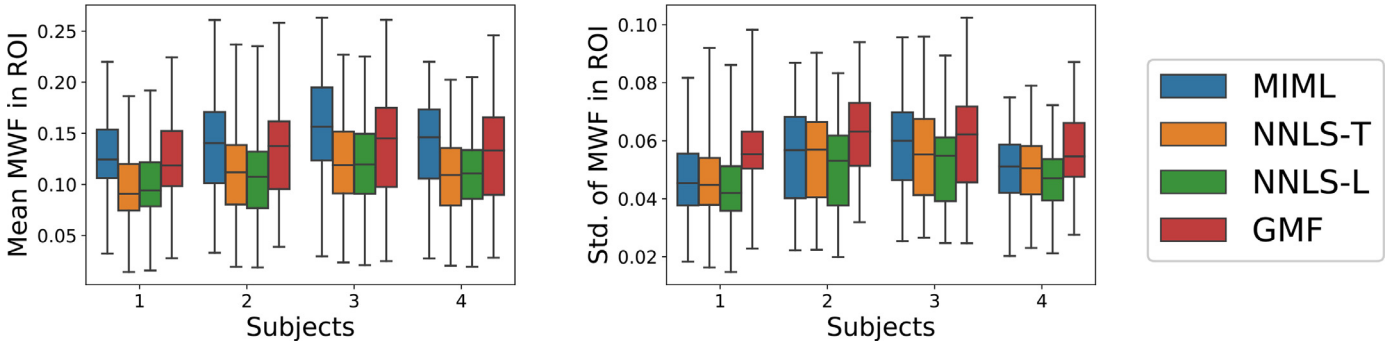

Fig. 9. Boxplots of the mean MWF (left) and the standard deviation of the MWF (right) for each method over all the WM ROIs for each subject in the cohort of healthy subjects. We see that MIML produces a larger mean MWF across all subjects than the other methods. This is likely due to the underestimation of MWF by the non-parametric methods as is shown in the results on the realistic synthetic data as well as in (Wiggermann et al., 2020). The standard deviations of the MWF from MIML are generally comparable to that of the other methods, similar to that of NNLS-T and slightly higher than that of NNLS-L. We note that the magnitude of the increase in the mean MWF using MIML as compared to the other methods is larger than the increase in the standard deviation of the MWF. This indicates that the smoothness of the MIML MWF map is comparable to that of other methods.

Table 3
Table of the mean and standard deviation of the absolute difference between the mean MWF values of the scan and rescan in white matter ROIs for each method and for each healthy subject. In bold are the lowest values per subject. Overall, GMF has the smallest mean differences for 3/4 subjects with standard deviations comparable to those of other methods. The performance of MIML and the NNLS methods are overall quite similar; while MIML has the smallest mean difference on Subject 2, NNLS methods have smaller mean differences in Subjects 3 and 4.

Mean and Standard Deviation of MWF Differences between Scan and Rescan WM ROIs				
	MIML	NNLS-T	NNLS-L	GMF
Subject 1	(0.0067,0.0388)	(0.0093,0.0353)	(0.0094,**0.0305**)	(**0.0055**,0.0357)
Subject 2	(**0.0004**,0.0448)	(0.0012,0.0473)	(0.0010,0.0418)	(0.0044,**0.0454**)
Subject 3	(0.0191,0.0726)	(0.0134,0.0755)	(0.0176,0.0723)	(**0.0108**,**0.0712**)
Subject 4	(0.0104,0.0573)	(0.0107,0.0543)	(0.0103,**0.0504**)	(**0.0065**,0.0561)

the myelin water lobe and the IES water lobe correspond to the range expected at 3T. The NNLS methods produce over-smoothed myelin water lobes with peaks occurring at implausibly low T_2 values. The IES water lobes are generally plausible. GMF produces more plausible myelin water lobes than those of the NNLS methods, but the lobes in the contralateral tissue are more variable, with the estimated mean and standard deviation of the myelin lobes varying significantly over the 3 regions of contralateral tissue. MIML reconstructs a diminished myelin water lobe in the lesions as compared to the normal-appearing tissue, reflecting lower MWF; this is in line with expectations of MS as a demyelinating disorder. In contrast, the distributions from the NNLS methods in Lesions 2/3 exhibit larger myelin water lobes in lesion tissue as compared to normal-appearing tissue, indica-

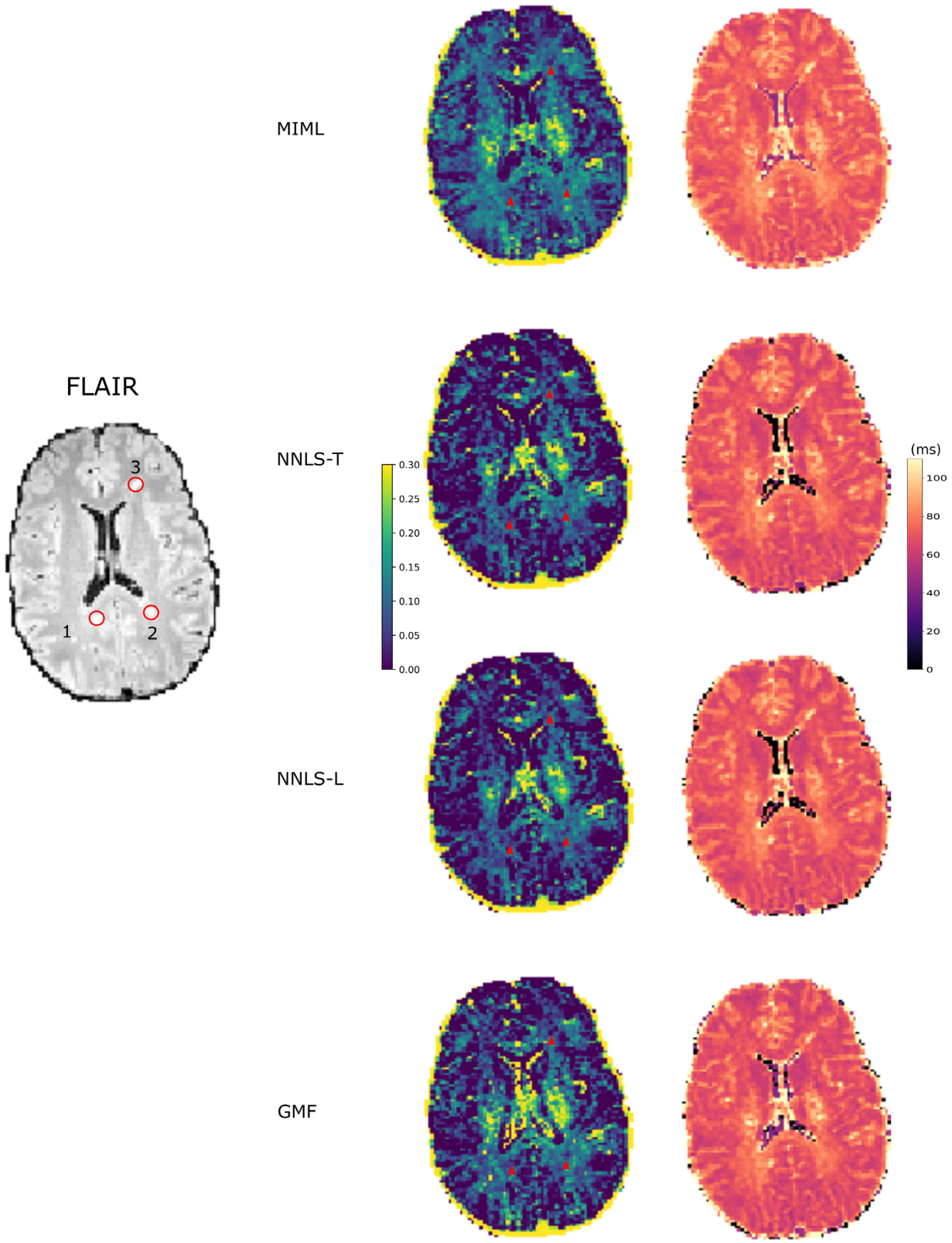

Fig. 10. Anatomical FLAIR map where the white matter is hypointense (first column), maps of the MWF (second column), and maps of the geometric mean T_2 in the range corresponding to the IE space (50–200 ms) (third column) for an axial slice in a subject with MS. We show the MS lesions on the FLAIR map marked in red and labeled numerically. Regarding the mean T_2 maps, we can see that the all lesions but Lesion 2 can be seen as hyperintensities, with the maps very similar across all methods. Regarding the MWF maps, as in the healthy subjects, MIML most smoothly and accurately reconstructs the WM, with the other methods exhibiting more noisy maps with missing patches. MIML provides the best lesion visualization due to better contrast between normal appearing tissue and lesions and a more smooth MWF map; in particular, lesions can clearly be delineated from close, adjacent structures in contrast to the NNLS methods (see Lesion 1, 3). See Fig. 3 in the Suppl. material for a closer look/analysis of the MWF maps compared to the lesions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Table of the spatial Pearson correlation and the linear regression coefficients (slope and intercept) between the mean MWF values of the scan and rescan in white matter ROIs for each method and each healthy subject. In bold are the highest Pearson correlations per subject. All Pearson correlations have p values less than 0.01. Overall, all methods perform quite similarly. However, GMF has the best correlations between scans (by a small margin), with MIML and the NNLS methods performing similarly.

Pearson Correlation and Linear Regression Coefficients between Scan and Rescan WM ROIs				
	MIML	NNLS-T	NNLS-L	GMF
Subject 1	(0.92, 0.89,0.0069)	(0.91,0.887,0.0016)	(**0.93**,0.90,0.0004)	(**0.93**,0.85,0.0130)
Subject 2	(0.90,1.00,0.0002)	(0.87,0.98,0.0008)	(0.89,1.02,-0.0034)	(**0.91**,0.99,0.0045)
Subject 3	(**0.77**,0.74,0.0546)	(0.70,0.65,0.0527)	(0.72,0.69,0.0520)	(**0.77**,0.76,0.0433)
Subject 4	(0.87,0.87,0.0087)	(0.87,0.97,-0.0064)	(**0.89**,0.97,-0.0076)	(**0.89**,0.96,0.0001)

Table 5
Table of the average computation time for whole brain reconstructions for each method on the healthy subjects; all reconstructions were done using the same computer, with 16 threads. MIML is orders of magnitude faster than the other methods.

Average Computation for Whole Brain				
	MIML	NNLS-T	NNLS-L	GMF
Time	34s	752s	701.2s	159382.4s

tive of the poor MWF reconstruction in the normal-appearing tissue.

In conclusion, all methods perform similarly in detecting lesions from the mean T_2. However, MIML improves upon the NNLS methods and GMF in detecting lesions from MWF maps, by providing better contrast between lesions and normal appearing tissues, clearer delineation of lesions from adjacent structures, and smoother, more plausible reconstructions overall in the WM. The comparison of the myelin water lobes of lesion and normal appearing tissue from MIML is consistent with the demyelinating nature of MS in contrast to that from the NNLS methods. Therefore, the performance of MIML meets or exceeds the performance of the other methods when used on a pathological case.

From our results on real data, we see that MIML generalizes to different machines, different magnetic field strengths, and different sequences since it is trained on a model of the signal decay which is agnostic to these differences; MIML's performance on the real data shows its potential for multi-component T_2 relaxometry at clinically achievable SNRs in high resolution scans.

4.3. Computation time

Here we provide a brief overview of the computational cost of the different methods. For consistent comparison, we used one computer using Ubuntu 18.04 with an Intel Xeon CPU E5-1650v4 running at 3.6 GHz with 12 available threads to run parallelized whole-brain reconstructions on four of the healthy subjects (matrix size 144x126x84) using MIML, NNLS-T, NNLS-L, and GMF; we recorded the time to completion and show the average computation time for each method in Table 5. We can see that MIML is 1 to 4 orders of magnitude faster than the other methods.

5. Discussion

Overall, from our evaluation on synthetic data, an *ex vivo* scan and *in vivo* scans (healthy and pathological), we conclude that MIML provides fast, noise-robust, and plausible reconstructions of T_2 distributions, with potential for use in myelin water fraction mapping. We attribute the performance of our method to the blending of the advantages of machine learning, parametric, and non-parametric methods. We note that our approach is essentially using machine learning to solve the inverse problem of parametric approaches, albeit expressing the solution non-parametrically. We view our approach as an extension of the recent progress in using machine learning to solve inverse problems in many domains (Adler and Öktem, 2017). By using machine learning, our method is much faster than standard parametric or non-parametric approaches. By training on solely simulated data, our approach does not require expensive, *in vivo* acquisitions for training data, nor the need for multiple scans to adapt to different machines or sequences. Further, by simulating random flip angles in the dataset, our method is able to automatically account for the flip angle, in contrast to non-parametric methods which need to estimate the flip angle before fitting. Altogether, this allows for noise-robust reconstruction by training the network on simulated signals with an SNR range and noise model corresponding to those from clinical scans. By generating the simulations guided by biophysical models, we can simultaneously retain stability in the reconstruction by constraining the space of T_2 distributions while not being restricted to a specific number of water pools at inference time. Further, the produced distributions are implicitly constrained to have a plausible, lobular structure (as in parametric approaches), which makes the interpretation of parameters of interest such as the MWF consistent with past studies, in contrast to potential irregular distributions from non-parametric methods. The trained MIML model and code for generating the synthetic data and training the model will be available at the following website: https://github.com/thomas-yu-epfl/Model_Informed_Machine_Learning.

However, our current approach has several limitations. First, while we attempted to be as comprehensive as possible in the simulated dataset, advances in biophysical modelling make it possible that there are additional relevant water pools to be estimated. For example, the Gaussian Mixture model we use assumes the symmetry of the mixture distributions, which may not be true in real distributions; in the case of skew, ground truth distributions, our method can result in a biased reconstruction. Second, while we fixed the Rician noise model for the training signals, with a fixed SNR range of 80-200, we note that in some sequences, more complex noise models such as the non-central chi distribution (Aja-Fernández and Vegas-Sánchez-Ferrero, 2016) with different SNR ranges may also be appropriate. Third, we only consider 32-echo sequences in this work. Fourth, we use a fixed, logarithmic T_2 discretization consisting of 60 points from 10 ms to 2000 ms for both our method and the NNLS methods. However, finer or coarser discretizations could also have been used. Finally, there may be relevant physical effects such as magnetization transfer (Sled, 2018; Malik et al., 2018) which, if modelled in the dataset, could improve the reconstructions. However, we highlight the flexibility and modularity of our approach for accounting for these limitations. Additional water pools can be easily added to the training dataset. The noise model and SNR range used in training can be swapped out for different noise models and SNRs. A sequence with a different number of echoes can be accommodated by reconstructing the dataset with the required number of echos and retraining the network. Different T_2 discretizations would simply require downsampling of the high resolution T_2 distributions in our dataset to match the new discretization, with subsequent retraining of the

network. More advanced physical modelling can be added to the generation of new datasets. As the training of the network is quite fast (70s on a laptop GPU), the bottleneck for addressing these limitations is the dataset generation (1 h on 46 CPU threads). However, while we generated our dataset on CPU, GPU acceleration of the EPG formalism can potentially speedup dataset generation significantly (Wang et al., 2020).

As for future work: in this paper, we did not study the impact of denoising the data on the reconstruction performance of the methods compared. This is first because in our overview of the literature, we found that presenting results on denoised data is not typical unless the subject of the paper is denoising. Second, the type of denoising, setting of denoising parameters, and accounting for potential biases due to denoising all require careful justification and study, which we felt was out of the scope of this paper, which introduces a proof of concept. However, we note that in the MS data, particularly for the NNLS methods, ostensibly normal appearing regions of the brain had unusually low MWF values, sometimes less than that predicted for the lesion. These areas of unusually low MWF values could also be seen in the scans of healthy subjects. These may be due to, in part, instability/ill-posedness in the estimation due to comparatively low SNRs in the *in vivo* scans; the *in vivo* scans we used have fairly high resolution (1.6-1.8mm) and are isotropic, while typical scans in the literature generally use much thicker slices ($\geq$ 2mm) along the axial direction (Alonso-Ortiz et al., 2015; Prasloski et al., 2012a). We note that both distributions and MWF maps from the NNLS methods were more plausible in the *ex vivo* scan, where the SNR was much higher. This is consistent with the observations in (Wiggermann et al., 2020) concerning the noise dependence of NNLS methods. Future studies will be conducted to study the impact of denoising algorithms such as PCA denoising (Does et al., 2019), or the NESMA filter (Bouhrara et al., 2018) on MIML as well as other methods, and any effect this has on their comparison.

Our method, as well as the other methods compared to in this work, reconstruct the T_2 distribution in each voxel separately. However, there are parametric and non-parametric approaches to T_2 relaxometry which use spatial regularization (El-Hajj et al., 2020; Hwang and Du, 2009; Kumar et al., 2018). These approaches assume that voxels spatially close to each other should also have similar reconstructions; hence, they perform reconstructions on groups of adjacent voxels simultaneously, with constraints that limit the variation of the reconstructions over the group. In addition, another approach estimates over groups of voxels by assuming the joint sparsity of the distributions in a region of interest (Nagtegaal et al., 2020). In future work, we will study how regularization/simultaneous fitting over regions of interest can be incorporated into our machine-learning framework as well as its effects on distribution reconstruction.

In this paper, we tested our method on two types of sequences: a multi-echo spin echo sequence and a 3D gradient and spin echo sequence. While in principle our approach is agnostic to the sequence used, in the future we will further validate our method on data from other sequences such as the T_2 prepared gradient echo sequences (Nguyen et al., 2012).

Finally, we note that using more advanced neural networks such as Long short term memory (LSTM) networks (Hochreiter and Schmidhuber, 1997), which are suitable for time series data, may offer improved reconstructions as well as potentially eliminating the need for fixed size inputs. In addition, while our synthetic dataset generation is based solely on the most common cases for biophysical modelling, we will investigate how to improve dataset generation i.e. the number of pools, the maximum number of pools present per signal, etc. in order to optimize the generalization capabilities of the network while minimizing the ill-posedness of the reconstruction.

6. Conclusion

In this work, we presented Model-Informed Machine Learning (MIML), an approach for estimating T_2 distributions from MRI signals using a neural network trained on synthetic data derived from biophysical models. Through our evaluations on synthetic data, an *ex vivo* scan, as well as healthy and pathological *in vivo* data, we show that MIML provides more robust, accurate, and plausible T_2 distributions than standard parametric and non-parametric methods across a wide range of SNRs. We show that MWF maps derived from MIML show the highest conformity to anatomical scans, have the greatest correlation to a histological map of myelin volume, and improve upon the lesion visualization capabilities of other methods, with better contrast between lesions and normal-appearing tissue as well as clearer delineation between lesions and close adjacent structures. The code for generating the datasets and training the network will be made available at https://github.com/thomas-yu-epfl/Model_Informed_Machine_Learning.

Declaration of Competing Interest

The authors declare that they do not have any financial or non-financial conflict of interests

CRediT authorship contribution statement

Thomas Yu: Conceptualization, Methodology, Software, Writing - original draft. **Erick Jorge Canales-Rodríguez:** Conceptualization, Methodology, Software, Writing - original draft, Supervision. **Marco Pizzolato:** Conceptualization, Methodology, Writing - original draft. **Gian Franco Piredda:** Investigation, Resources, Writing - review & editing. **Tom Hilbert:** Investigation, Resources, Writing - review & editing. **Elda Fischi-Gomez:** Investigation, Resources, Writing - review & editing. **Matthias Weigel:** Investigation, Resources, Writing - review & editing. **Muhamed Barakovic:** Investigation, Resources, Writing - review & editing. **Meritxell Bach Cuadra:** Writing - review & editing, Resources, Supervision, Funding acquisition. **Cristina Granziera:** Investigation, Resources, Writing - review & editing. **Tobias Kober:** Investigation, Resources, Writing - review & editing. **Jean-Philippe Thiran:** Writing - review & editing, Resources, Supervision, Funding acquisition.

Acknowledgments

This project is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie project TRABIT (agreement No 765148 to TY) and by the Swiss National Science Foundation (SNSF, Ambizione grant PZ00P2_185814/1 to EJC-R). It has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 754462 (to MP) and the Strategic Focal Area "Personalized Healthcare and Related Technologies (PHRT)" of the ETH domain (grant 2018-425 to EFG), as well as the Centre for Biomedical Imaging (CIBM) of the University of Lausanne, the Swiss Federal Institute of Technology Lausanne, the Lausanne University Hospital (to MBC and J-PT), and the University of Geneva and Geneva University Hospital (to MBC). In addition, this project was supported by Swiss National Funds PZ00P3_154508, PZ00P3_131914 and PP00P3_176984 (CG,MW,MB).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I.,

Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. URL: http://tensorflow.org/. Software available from tensorflow.org.

Adler, J., Öktem, O., 2017. Solving ill-posed inverse problems using iterative deep neural networks. Inverse Probl. 33 (12), 124007.

Aja-Fernández, S., Vegas-Sánchez-Ferrero, G., 2016. Statistical Analysis of Noise in MRI. Switzerland: Springer International Publishing.

Akhondi-Asl, A., Afacan, O., Mulkern, R.V., Warfield, S.K., 2014. T2-relaxometry for myelin water fraction extraction using wald distribution and extended phase graph. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 145–152.

Alonso-Ortiz, E., Levesque, I.R., Paquin, R., Pike, G.B., 2017. Field inhomogeneity correction for gradient echo myelin water fraction imaging. Magn. Reson. Med. 78 (1), 49–57.

Alonso-Ortiz, E., Levesque, I.R., Pike, G.B., 2015. MRI-Based myelin water imaging: a technical review. Magn. Reson. Med. 73 (1), 70–81.

Andrews, T., Lancaster, J.L., Dodd, S.J., Contreras-Sesvold, C., Fox, P.T., 2005. Testing the three-pool white matter model adapted for use with T2 relaxometry. Magn. Reson. Med. 54 (2), 449–454.

Björk, M., Zachariah, D., Kullberg, J., Stoica, P., 2016. A multicomponent t2 relaxometry algorithm for myelin water imaging of the brain. Magn. Reson. Med. 75 (1), 390–402.

Bloch, F., 1946. Nuclear induction. Phys. Rev. 70 (7–8), 460.

Bouhrara, M., Reiter, D.A., Maring, M.C., Bonny, J.-M., Spencer, R.G., 2018. Use of the NESMA filter to improve myelin water fraction mapping with brain MRI. J. Neuroimaging 28 (6), 640–649.

Brant-Zawadzki, M., Gillan, G.D., Nitz, W.R., 1992. MP RAGE: A three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain. Radiology 182 (3), 769–775.

Chatterjee, S., Commowick, O., Afacan, O., Warfield, S.K., Barillot, C., 2018. Multi--compartment model of brain tissues from T2 relaxometry MRI using gamma distribution. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 141–144.

Cohen-Adad, J., Does, M., DUVAL, T., Dyrby, T. B., Fieremans, E., Foias, A., Nami, H., Sepehrband, F., Stikov, N., Zaimi, A., et al., 2020. White matter microscopy database. 10.17605/OSF.IO/YP4QG

De Coene, B., Hajnal, J.V., Gatehouse, P., Longmore, D.B., White, S.J., Oatridge, A., Pennock, J., Young, I., Bydder, G., 1992. MR of the brain using fluid-attenuated inversion recovery (FLAIR) pulse sequences.. Am. J. Neuroradiol. 13 (6), 1555–1564.

Does, M.D., Olesen, J.L., Harkins, K.D., Serradas-Duarte, T., Gochberg, D.F., Jespersen, S.N., Shemesh, N., 2019. Evaluation of principal component analysis image denoising on multi-exponential MRI relaxometry. Magn. Reson. Med. 81 (6), 3503–3514.

Du, Y.P., Chu, R., Hwang, D., Brown, M.S., Kleinschmidt-DeMasters, B.K., Singel, D., Simon, J.H., 2007. Fast multislice mapping of the myelin water fraction using multicompartment analysis of T decay at 3T: a preliminary postmortem study. Magn. Reson. Med. 58 (5), 865–870.

El-Hajj, C., Moussaoui, S., Collewet, G., Musse, M., 2020. Multi-exponential transverse relaxation times estimation from magnetic resonance images under rician noise and spatial regularization. IEEE Trans. Image Process..

Graham, S.J., Stanchev, P.L., Bronskill, M.J., 1996. Criteria for analysis of multicomponent tissue T2 relaxation data. Magn. Reson. Med. 35 (3), 370–378.

Haacke, E.M., Brown, R.W., Thompson, M.R., Venkatesan, R., Thomphson, M., Venkatesan, M., Haacke, M., Brown, W., Thompson, M., 1999. Magnetic resonance imaging: physical principles and sequence design.

Hennig, J., 1988. Multiecho imaging sequences with low refocusing flip angles. J. Magn. Reson. (1969) 78 (3), 397–407.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hwang, D., Du, Y.P., 2009. Improved myelin water quantification using spatially regularized non-negative least squares algorithm. J. Magn. Reson. Imaging 30 (1), 203–208.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv:1412.6980.

Kolind, S.H., Mädler, B., Fischer, S., Li, D.K., MacKay, A.L., 2009. Myelin water imaging: implementation and development at 3.0 T and comparison to 1.5 T measurements. Magn. Reson. Med. 62 (1), 106–115.

Kroeker, R.M., Henkelman, R.M., 1986. Analysis of biological NMR relaxation data with continuous distributions of relaxation times. J. Magn. Reson. (1969) 69 (2), 218–235.

Kumar, D., Hariharan, H., Faizy, T.D., Borchert, P., Siemonsen, S., Fiehler, J., Reddy, R., Sedlacik, J., 2018. Using 3D spatial correlations to improve the noise robustness of multi component analysis of 3D multi echo quantitative T2 relaxometry data. NeuroImage 178, 583–601.

Kumar, D., Nguyen, T.D., Gauthier, S.A., Raj, A., 2012. Bayesian algorithm using spatial priors for multiexponential T2 relaxometry from multiecho spin echo MRI. Magn. Reson. Med. 68 (5), 1536–1543.

La Rosa, F., Abdulkadir, A., Fartaria, M.J., Rahmanzadeh, R., Lu, P.-J., Galbusera, R., Barakovic, M., Thiran, J.-P., Granziera, C., Cuadra, M.B., 2020. Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE. NeuroImage 102335.

Laule, C., Kozlowski, P., Leung, E., Li, D.K., MacKay, A.L., Moore, G.W., 2008. Myelin water imaging of multiple sclerosis at 7 T: correlations with histopathology. Neuroimage 40 (4), 1575–1580.

Laule, C., Leung, E., Li, D.K., Traboulsee, A., Paty, D., MacKay, A., Moore, G.R., 2006. Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology. Multiple Sclerosis J. 12 (6), 747–753.

Laule, C., Vavasour, I.M., Kolind, S.H., Traboulsee, A.L., Moore, G., Li, D.K., MacKay, A.L., 2007. Long T2 water in multiple sclerosis: what else can we learn from multi-echo T2 relaxation? J. Neurol. 254 (11), 1579–1587.

Lawson, C.L., Hanson, R.J., 1995. Solving Least Squares Problems, vol. 15. Siam.

Layton, K.J., Morelande, M., Wright, D., Farrell, P.M., Moran, B., Johnston, L.A., 2013. Modelling and estimation of multicomponent t_{-} {2} distributions. IEEE Trans. Med. Imaging 32 (8), 1423–1434.

Lee, J., Lee, D., Choi, J. Y., Shin, D., Shin, H.-G., Lee, J., 2019. Artificial neural network for myelin water imaging. Magn. Reson. Med..

Levesque, I.R., Giacomini, P.S., Narayanan, S., Ribeiro, L.T., Sled, J.G., Arnold, D.L., Pike, G.B., 2010. Quantitative magnetization transfer and myelin water imaging of the evolution of acute multiple sclerosis lesions. Magn. Reson. Med. 63 (3), 633–640.

Liu, H., Xiang, Q.-S., Tam, R., Dvorak, A.V., MacKay, A.L., Kolind, S.H., Traboulsee, A., Vavasour, I.M., Li, D.K., Kramer, J.K., Laule, C., 2020. Myelin water imaging data analysis in less than one minute. NeuroImage 210, 116551.

Mackay, A., Whittall, K., Adler, J., Li, D., Paty, D., Graeb, D., 1994. In vivo visualization of myelin water in brain by magnetic resonance. Magn. Reson. Med. 31 (6), 673–677.

MacKay, A.L., Laule, C., 2007. Myelin water imaging. eMagRes.

Malik, S.J., Teixeira, R.P.A., Hajnal, J.V., 2018. Extended phase graph formalism for systems with magnetization transfer and exchange. Magn. Reson. Med. 80 (2), 767–779.

Menon, R., Allen, P., 1991. Application of continuous relaxation time distributions to the fitting of data from model systmes and excised tissue. Magn. Reson. Med. 20 (2), 214–227.

Menon, R., Rusinko, M., Allen, P., 1992. Proton relaxation studies of water compartmentalization in a model neurological system. Magn. Reson. Med. 28 (2), 264–274.

Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A.V., Mahmood, A., Woods, R., et al., 2008. Stereotaxic white matter atlas based on diffusion tensor imaging in an icbm template. Neuroimage 40 (2), 570–582.

Nagtegaal, M., Koken, P., Amthor, T., de Bresser, J., Mädler, B., Vos, F., Doneva, M., 2020. Myelin water imaging from multi-echo T2 MR relaxometry data using a joint sparsity constraint. NeuroImage 117014.

Neumann, D., Blaimer, M., Jakob, P.M., Breuer, F.A., 2014. Simple recipe for accurate T 2 quantification with multi spin-echo acquisitions. Magn. Reson. Mater. Phys. Biol. Med. 27 (6), 567–577.

Nguyen, T.D., Wisnieff, C., Cooper, M.A., Kumar, D., Raj, A., Spincemaille, P., Wang, Y., Vartanian, T., Gauthier, S.A., 2012. T2 prep three-dimensional spiral imaging with efficient whole brain coverage for myelin water quantification at 1.5 Tesla. Magn. Reson. Med. 67 (3), 614–621.

Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., Akhter, K., Hua, K., Woods, R., Toga, A.W., et al., 2008. Human brain white matter atlas: identification and assignment of common anatomical structures in superficial white matter. Neuroimage 43 (3), 447–457.

Piredda, G.F., Hilbert, T., Canales-Rodrìgez, E.J., Pizzolato, M., Meuli, R., Pfeuffer, J., Thiran, J.-P., Kober, T., 2020. Fast and high-resolution myelin water imaging: accelerating multi-echo GRASE with CAIPIRINHA. Magn. Reson. Med..

Piredda, G.F., Hilbert, T., Thiran, J.-P., Kober, T., 2020. Probing myelin content of the human brain with MRI: a review. Magn. Reson. Med..

Prange, M., Song, Y.-Q., 2009. Quantifying uncertainty in NMR T2 spectra using Monte Carlo inversion. J. Magn. Reson. 196 (1), 54–60.

Prasloski, T., Mädler, B., Xiang, Q.-S., MacKay, A., Jones, C., 2012. Applications of stimulated echo correction to multicomponent T2 analysis. Magn. Reson. Med. 67 (6), 1803–1814.

Prasloski, T., Rauscher, A., MacKay, A.L., Hodgson, M., Vavasour, I.M., Laule, C., Mädler, B., 2012. Rapid whole cerebrum myelin water imaging using a 3D GRASE sequence. Neuroimage 63 (1), 533–539.

Raj, A., Pandya, S., Shen, X., LoCastro, E., Nguyen, T.D., Gauthier, S.A., 2014. Multi-compartment T2 relaxometry using a spatially constrained multi-Gaussian model. PLoS One 9 (6).

Raj, A., Pandya, S., Shen, X., LoCastro, E., Nguyen, T.D., Gauthier, S.A., 2014. Multi-compartment T2 relaxometry using a spatially constrained multi-Gaussian model. PLoS One 9 (6), e98391.

Ramdas, A., Trillos, N.G., Cuturi, M., 2017. On wasserstein two-sample testing and related families of nonparametric tests. Entropy 19 (2), 47.

Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain.. Psychol. Rev. 65 (6), 386.

Sled, J.G., 2018. Modelling and interpretation of magnetization transfer imaging in the brain. Neuroimage 182, 128–135.

Van Rossum, G., et al., 2000. Python reference manual.

Vasilescu, V., Katona, E., Simplaceanu, V., Demco, D., 1978. Water compartments in the myelinated nerve. III. Pulsed NMR result. Experientia 34 (11), 1443–1444.

Villani, C., 2009. The wasserstein distances. In: Optimal Transport. Springer, pp. 93–111.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E.W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Contributors, S., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. doi:10.1038/s41592-019-0686-2.

Vuong, M.-T., Duval, T., Cohen-Adad, J., Stikov, N., 2017. On the precision of myelin imaging: characterizing ex vivo dog spinal cord., p. 3760.
Wang, D., Ostenson, J., Smith, D.S., 2020. snapMRF: GPU-accelerated magnetic resonance fingerprinting dictionary generation and matching using extended phase graphs. Magn. Reson. Imaging 66, 248–256.
Wansapura, J.P., Holland, S.K., Dunn, R.S., Ball Jr, W.S., 1999. NMR Relaxation times in the human brain at 3.0 Tesla. J. Magn. Reson. Imaging 9 (4), 531–538.
Whittall, K.P., MacKay, A.L., 1989. Quantitative interpretation of NMR relaxation data. J. Magn. Reson. (1969) 84 (1), 134–152.
Whittall, K.P., Mackay, A.L., Graeb, D.A., Nugent, R.A., Li, D.K., Paty, D.W., 1997. In vivo measurement of T2 distributions and water contents in normal human brain. Magn. Reson. Med. 37 (1), 34–43.
Wiggermann, V., Vavasour, I.M., Kolind, S., MacKay, A.L., Helms, G., Rauscher, A., 2020. Non-negative least squares computation for in vivo myelin mapping using simulated multi-echo spin-echo T2 decay data. NMR Biomed. e4277.
Yu, T., Pizzolato, M., Canales-Rodríguez, E.J., Thiran, J.-P., 2019. Robust T 2 relaxometry with hamiltonian MCMC for myelin water fraction estimation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 1813–1817.
Zaimi, A., Wabartha, M., Herman, V., Antonsanti, P.-L., Perone, C.S., Cohen-Adad, J., 2018. AxonDeepSeg: automatic axon and myelin segmentation from microscopy data using convolutional neural networks. Sci. Rep. 8 (1), 1–11.