Revisiting the T_2 spectrum imaging inverse problem: Bayesian regularized non-negative least squares

Erick Jorge Canales-Rodríguez [a,*], Marco Pizzolato [b,a], Thomas Yu [a,c], Gian Franco Piredda [a,d,e], Tom Hilbert [a,d,e], Joaquim Radua [f,g,h], Tobias Kober [a,d,e], Jean-Philippe Thiran [a,e]

[a] Signal Processing Laboratory (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), EPFL-STI-IEL-LTS5, Station 11, CH-1015, Lausanne, Switzerland
[b] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark
[c] Medical Image Analysis Laboratory, Center for Biomedical Imaging (CIBM), University of Lausanne, Switzerland
[d] Advanced Clinical Imaging Technology, Siemens Healthcare AG, Lausanne, Switzerland
[e] Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[f] Imaging of Mood- and Anxiety-Related Disorders (IMARD) group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), CIBERSAM, Barcelona, Spain
[g] Department of Psychosis Studies, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom
[h] Department of Clinical Neuroscience, Centre for Psychiatric Research and Education, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

ABSTRACT

Multi-echo T_2 magnetic resonance images contain information about the distribution of T_2 relaxation times of compartmentalized water, from which we can estimate relevant brain tissue properties such as the myelin water fraction (MWF). Regularized non-negative least squares (NNLS) is the tool of choice for estimating non-parametric T_2 spectra. However, the estimation is ill-conditioned, sensitive to noise, and highly affected by the employed regularization weight. The purpose of this study is threefold: first, we want to underline that the apparently innocuous use of two alternative parameterizations for solving the inverse problem, which we called the standard and alternative regularization forms, leads to different solutions; second, to assess the performance of both parameterizations; and third, to propose a new Bayesian regularized NNLS method (BayesReg). The performance of BayesReg was compared with that of two conventional approaches (L-curve and Chi-square (X^2) fitting) using both regularization forms. We generated a large dataset of synthetic data, acquired in vivo human brain data in healthy participants for conducting a scan-rescan analysis, and correlated the myelin content derived from histology with the MWF estimated from *ex vivo* data. Results from synthetic data indicate that BayesReg provides accurate MWF estimates, comparable to those from L-curve and X^2, and with better overall stability across a wider signal-to-noise range. Notably, we obtained superior results by using the alternative regularization form. The correlations reported in this study are higher than those reported in previous studies employing the same *ex vivo* and histological data. In human brain data, the estimated maps from L-curve and BayesReg were more reproducible. However, the T_2 spectra produced by BayesReg were less affected by over-smoothing than those from L-curve. These findings suggest that BayesReg is a good alternative for estimating T_2 distributions and MWF maps.

1. Introduction

Multi-component T_2 relaxometry can be used to study brain tissue microstructure. It allows one to obtain relevant information about the water fractions of different tissue constituents and their T_2 relaxation times non-invasively Does (2018). Specifically, this technique provides estimates for the fraction of the water trapped between the myelin layers covering the axons in the brain's white matter (WM) (Piredda et al., 2020). The estimated myelin water fraction (MWF) is highly correlated with myelin content (Alonso-Ortiz et al., 2015) and hence is being used to study brain disorders affecting the WM tissue, like multiple sclerosis (Laule et al., 2008) and schizophrenia (Lang et al., 2014).

Previous studies using Multi-echo T_2 (MET_2) data demonstrated the existence of cellular-compartment-specific T_2 values (Andrews et al., 2005; Deoni et al., 2013; MacKay et al., 2006). For instance, it is believed that the fraction of the measured signal with T_2 in the range of 10–40 ms (in 3T scanner) arises from myelin water. The signal fraction with an intermediate T_2 in the range of 40–200 ms is attributed to the intra- and extra-cellular (IE) water. And the component with the longest

T_2s >1000 ms is assigned to the 'free water' in the cerebrospinal fluid (CSF).

Data from various studies suggest that a high signal-to-noise ratio (SNR) is required for estimating different components in the T_2 distribution (Andrews et al., 2005; Graham et al., 1996). For example, an assessment based on synthetic data found that for a clinically achievable SNR=100 the myelin water component was undetected in 5.6% of the voxels, and the percentage increased to 12.9% for SNR = 50 (Kumar et al., 2012). Similar results have been reported in human brain data, where MWF values equal to zero were estimated in several brain voxels located in myelinated areas of the frontal and lateral projection fibers (Raj et al., 2014). Although the fitting instability of non-parametric estimation algorithms has been improved by using regularized non-negative least squares (NNLS) (Whittall et al., 1997), the estimation is ill-conditioned, sensitive to noise contamination, and strongly depends on the chosen regularization weight.

Recently, we compared several non-parametric regularized NNLS methods across different SNR levels (Canales-Rodríguez et al., 2021) by employing three criteria for choosing the optimal regularization weight (i.e., Chi-square (X^2) residual fitting criterion (Graham et al., 1996; Whittall and MacKay, 1989), L-curve method (Hansen, 1992), and Generalized-cross validation (Golub et al., 1979)). The effect of these criteria was evaluated in combination with three different regularization matrices for promoting smooth solutions: Identity matrix and Laplacian matrices of first- and second-order derivatives. We found that for low SNRs the optimal regularization parameter should be selected by employing the L-curve method Hansen (1992). In contrast, X^2 was the method of choice for higher SNRs. In both cases, the reconstruction algorithms showed superior performance when the regularization matrix was the Identity matrix. However, the performance of these methods was worse outside their optimal SNR ranges. This aspect makes it difficult to provide recommendations for which method to use for datasets with heterogeneous SNR levels.

The main aim of this work is to continue the quest towards finding the optimal methodology for estimating T_2 spectra from MET$_2$ data, with a special focus on improving the quantification of the myelin water fraction. In particular, the purpose of this study is threefold. Firstly, we want to revisit the T_2 spectrum imaging inverse problem to show how the apparently innocuous use of two alternative parameterizations leads to two different solutions: the standard regularization form (Whittall and MacKay, 1989), which promote solutions where the square of the area under the spectrum is minimal, and the alternative form, that promotes solutions minimizing the square of the spectrum intensities. We noted that the alternative regularization form is equivalent to that previously proposed by (Guo et al., 2013). Secondly, as these two regularization strategies have not been compared, we will evaluate them for the first time here. Finally, we propose a new Bayesian regularization technique for determining the optimal regularization parameter. Our proposal is an extension of the seminal Bayesian interpolation framework introduced by (MacKay, 1992), but adapted for physical models where the estimated parameters cannot have negative values. We hypothesize that the introduced Bayesian regularization framework will produce solutions similar to those from state-of-the-art regularization approaches based on L-curve and X^2, but with better overall performance across a wider SNR range, and that the alternative regularization form may provide better performance. To test our hypothesis, a large dataset of synthetic data was generated, and several methods were compared. Moreover, in vivo MET$_2$ data were acquired in healthy volunteers, which were scanned two times to conduct a scan-rescan reproducibility analysis. The performance of the algorithms was evaluated on *ex vivo* multi-echo T2 data by correlating the estimated voxel-wise myelin water fraction and the myelin content derived from histology. To foster reproducible research, the implemented methods were included in our freely distributed multi-component T_2 reconstruction toolbox available at https://github.com/ejcanalesr/multicomponent-T2-toolbox.

2. Methods

2.1. T2 spectrum estimation by regularized NNLS

The measured MET$_2$ signal S for a given echo time (TE) is modeled by (Mackay et al., 1994; Whittall et al., 1997)

$$S(TE) = \int_{T_2^{\min}}^{T_2^{\max}} P(T_2)H(T_2, TE)dT_2$$
$$\approx \sum_{j=1}^{N} P\left(\frac{T_2^j + T_2^{j+1}}{2}\right)H\left(\frac{T_2^j + T_2^{j+1}}{2}, TE\right)\left(T_2^{j+1} - T_2^j\right), \quad (1)$$

where $P(T_2)$ is the unknown distribution (often also called T_2 spectrum) of relaxation times to be estimated, H is the kernel specifying the physical relaxation model for each water pool within the voxel (e.g., Bloch's exponential decay (Mackay et al., 1994) or Extended Phase Graph (EPG) model (Prasloski et al., 2012a)), and $\{T_2^j, j = 1,...N + 1\}$ is the discrete grid of T_2 values used to evaluate the integral numerically. Note that for notational simplicity we used a composite midpoint rule with N subintervals/bins, and that others numerical integration rules are possible without loss of generality.

Eq. (1) can be parameterized in two different ways for estimating the distribution of relaxation times. The conventional approach is to collapse the discrete spectrum intensities and the bin widths into a single parameter, i.e., $w_j = P(\frac{T_2^j + T_2^{j+1}}{2})(T_2^{j+1} - T_2^j)$, in which case the resulting system of linear equations can be written in matrix form as $\mathbf{s} = \mathbf{Hw} + \mathbf{e}$, where $\mathbf{s}$ is the kx1 vector of measured data for k different TEs, $\mathbf{H}$ is the kxN dictionary of synthetic MRI signals for the specified T_2 grid, $\mathbf{e}$ denotes the random noise, and $\mathbf{w} = \{w_j\}$ is the Nx1 vector of weights to be estimated quantifying the area under the curve of the discrete spectrum.

The weights $\hat{\mathbf{w}}$ are commonly estimated through the following regularized non-negative least square (NNLS) problem:

$$\hat{\mathbf{w}}(\lambda) = \arg\min_{\mathbf{w}\geq 0}\|\mathbf{s} - \mathbf{Hw}\|_2^2 + \lambda\|\mathbf{Lw}\|_2^2$$
$$= \arg\min_{\mathbf{w}\geq 0}\left\|\begin{bmatrix}\mathbf{s}\\\mathbf{0}_N\end{bmatrix} - \begin{bmatrix}\mathbf{H}\\\sqrt{\lambda}\mathbf{L}\end{bmatrix}\mathbf{w}\right\|_2^2, \quad (2)$$

where $\mathbf{L}$ is the NxN regularization matrix introduced to stabilize the estimation by injecting prior information about the expected spectrum's shape, e.g., smoothness, and $\mathbf{0}_N$ denotes an Nx1 vector of zeros. The estimated solution $\hat{\mathbf{w}}$ depends on the chosen matrix $\mathbf{L}$ and the regularization parameter λ. The family of solutions defined by these choices can be estimated efficiently by the Lawson-Hanson active set algorithm (Lawson and Hanson, 1995). Note that for $\lambda = 0$ we get the non-regularized NNLS solution.

Eq. (2) is usually solved using $\mathbf{L} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix (Mackay et al., 1994; Whittall et al., 1997; Whittall and MacKay, 1989). This choice promotes solutions where the square of the area under the spectrum is minimal:

$$\|\mathbf{Iw}\|_2^2 = \sum_{j=1}^{N} w_j^2 = \sum_{j=1}^{N}\left[P\left(\frac{T_2^{j+1} + T_2^j}{2}\right)\left(T_2^{j+1} - T_2^j\right)\right]^2. \quad (3)$$

As the T_2 grid is conventionally selected using logarithmically spaced T_2 values, the last elements with bigger bin widths $T_2^{j+1} - T_2^j$ are more penalized than the first elements. Consequently, the resulting spectra are smoother around the portion of the spectra with long T_2 components. In this study, we will refer to this choice as the standard regularization form.

Alternatively, the bin widths $T_2^{j+1} - T_2^j$ can be collapsed with H in Eq. (1), leading to the system of linear eqs. $\mathbf{s} = \mathbf{Kx} + \mathbf{e}$, where $\mathbf{x} = \{x_j = P(\frac{T_2^j + T_2^{j+1}}{2}), j = 1,...N\}$ is the vector of weights to be estimated quantifying the discrete spectrum intensities, $\mathbf{K}=\mathbf{HD}$, and $\mathbf{D}$ is a diagonal

NxN matrix formed by the bin widths, $\mathbf{D}_{j,j} = T_2^{j+1} - T_2^j$. In this case, the weights $\hat{\mathbf{x}}$ are estimated using regularized NNLS as follows:

$$\hat{\mathbf{x}}(\lambda) = \arg\min_{\mathbf{x}\geq 0}\|\mathbf{s} - \mathbf{Kx}\|_2^2 + \lambda\|\mathbf{Lx}\|_2^2. \tag{4}$$

Note that for $\mathbf{L} = \mathbf{I}$, the penalty term only depends on the spectrum intensities but not on the area of the different subintervals:

$$\|\mathbf{Ix}\|_2^2 = \sum_{j=1}^N x_j^2 = \sum_{j=1}^N \left[P\left(\frac{T_2^{j+1}+T_2^j}{2}\right)\right]^2. \tag{5}$$

Notably, Eq. (4) can be rewritten in terms of $\mathbf{H}$ and $\mathbf{w}$ (assuming $\mathbf{L} = \mathbf{I}$ and noting that $\mathbf{Kx} = \mathbf{Hw}$ and $\|\mathbf{Ix}\|_2^2 = \|\mathbf{D}^{-1}\mathbf{w}\|_2^2$) as

$$\hat{\mathbf{w}}(\lambda) = \arg\min_{\mathbf{w}\geq 0}\|\mathbf{s} - \mathbf{Hw}\|_2^2 + \lambda\left\|\mathbf{D}^{-1}\mathbf{w}\right\|_2^2, \tag{6}$$

where $\mathbf{D}^{-1}$ is the inverse of $\mathbf{D}$, i.e., a diagonal matrix with elements $\mathbf{D}_{j,j}^{-1} = (T_2^{j+1} - T_2^j)^{-1}$. Interestingly, this formulation is identical to that proposed in (Guo et al., 2013) based on solving Eq. (2) with $\mathbf{L} = \mathbf{D}^{-1}$. We will call this variant the alternative regularization form due to its less widespread use.

It is important to note how the use of these two parameterization forms leads to two different solutions. Both regularization strategies are compared in this work.

The conventional method in T_2 relaxometry for estimating λ is based on the X^2 residual fitting criterion (Graham et al., 1996; Whittall and MacKay, 1989), which can be rewritten as the following optimization problem (Canales-Rodríguez et al., 2021):

$$\hat{\lambda} = \arg\min_{\lambda\geq 0}\left|1 - \frac{\|\mathbf{s} - \mathbf{H}\hat{\mathbf{w}}(\lambda)\|_2^2}{c\|\mathbf{s} - \mathbf{H}\hat{\mathbf{w}}(\lambda=0)\|_2^2}\right|, \tag{7}$$

where $\hat{\mathbf{w}}(\lambda = 0)$ is the solution of Eq. (2) for $\lambda=0$ and c is a constant conventionally fixed at $c = 1.02$ (Laule et al., 2006). This method provides robust fits in the presence of noise (Whittall and MacKay, 1989) because it reduces the overfitting by accepting a regularized solution whose mean squared error is 2% higher than the non-regularized one (MacKay et al., 2006).

Another well-known regularization approach is the L-curve method Hansen (1992), which plots the log-log graph of the curve given by $(\log\|\mathbf{s} - \mathbf{H}\hat{\mathbf{w}}(\lambda)\|_2^2, \quad \log\|\mathbf{L}\hat{\mathbf{w}}(\lambda)\|_2^2)$ for many values of λ. The point located at the corner of the resulting "L-curve" is chosen as the optimal λ. This method was applied in previous multi-component T_2 relaxometry studies, e.g., see (Kumar et al., 2012) and (Canales-Rodríguez et al., 2021).

As algorithms based on the Bayesian theory usually provide competitive performance, a new Bayesian method for the optimal selection of λ is introduced in this work.

2.2. Bayesian regularization

Using the standard framework of likelihood, prior and posterior as proposed in MacKay (1992), we define the following probabilities.

Likelihood: The noise is modeled as a zero-mean Gaussian distribution with standard deviation σ. Hence, the probability of the data given the parameters $\mathbf{H}$, $\mathbf{w}$, and σ is

$$P(\mathbf{s}|\mathbf{H}, \mathbf{w}, \beta) = \frac{1}{Z_d(\beta)}\exp\left(-\frac{\beta}{2}\|\mathbf{s} - \mathbf{Hw}\|_2^2\right), \tag{8}$$

where $\beta = 1/\sigma^2$ and $Z_d(\beta) = (2\pi/\beta)^{k/2}$.

Prior: To obtain a solution equivalent to Eq. (2), the prior must be a zero-mean truncated Gaussian distribution with standard deviation σ_P:

$$P(\mathbf{w}|\mathbf{L}, \alpha) = \frac{1}{Z_w(\alpha)}\exp\left(-\frac{\alpha}{2}\|\mathbf{Lw}\|_2^2\right)u(\mathbf{w}), \tag{9}$$

where $\alpha = 1/\sigma_P^2$, $u(\mathbf{w})$ is the indicator function guaranteeing the non-negativity of $\mathbf{w}$,

$$u(\mathbf{w}) = \begin{cases} 1, if\ \mathbf{w} \geq 0 \\ 0, otherwise, \end{cases} \tag{10}$$

and the normalization constant is determined as

$$Z_w(\alpha) = \int_0^\infty \exp\left(-\frac{\alpha}{2}\|\mathbf{Lw}\|_2^2\right)d\mathbf{w} \tag{11}$$

Posterior: According to the Bayes theorem, the posterior distribution of $\mathbf{w}$ given all the other parameters is given by MacKay (1992)

$$P(\mathbf{w}|\mathbf{s}, \mathbf{H}, \beta, \mathbf{L}, \alpha) = \frac{P(\mathbf{s}|\mathbf{w}, \mathbf{H}, \beta)P(\mathbf{w}|\mathbf{L}, \alpha)}{P(\mathbf{s}|\mathbf{H}, \beta, \mathbf{L}, \alpha)}. \tag{12}$$

By plugging Eqs. (8) and (9) into Eq. (12) we get

$$P(\mathbf{w}|\mathbf{s}, \mathbf{H}, \beta, \mathbf{L}, \alpha) = \frac{u(\mathbf{w})}{Z_P(\beta, \alpha)}\exp\left(-\frac{\beta}{2}\|\mathbf{s} - \mathbf{Hw}\|_2^2 - \frac{\alpha}{2}\|\mathbf{Lw}\|_2^2\right), \tag{13}$$

where

$$\frac{1}{Z_P(\beta, \alpha)} = \frac{1}{P(\mathbf{s}|\mathbf{H}, \beta, \mathbf{L}, \alpha)Z_d(\beta)Z_w(\alpha)}. \tag{14}$$

This normalization factor can be estimated in practice as

$$Z_P(\beta, \alpha) = \int_0^\infty \exp\left(-\frac{\beta}{2}\|\mathbf{s} - \mathbf{Hw}\|_2^2 - \frac{\alpha}{2}\|\mathbf{Lw}\|_2^2\right)d\mathbf{w}. \tag{15}$$

Notably, the vector $\mathbf{w}$ that maximizes the posterior probability (i.e., Eq. (13)) is the solution of Eq. (2) for $\lambda = \alpha/\beta$. Therefore, the Bayesian formulation replicates the regularized NNLS solution and allows us to better interpret the regularization parameter in terms of α and β. This equivalence indicates that the conventional reconstruction given by Eq. (2) implicitly assumes a Gaussian noise model and a zero-mean truncated Gaussian prior. Thus, it promotes solutions with many elements equal to zero and positive smooth lobes.

2.3. Bayesian choice of α and β

The Bayesian theory provides a solid framework to infer the values of α and β from the data $\mathbf{s}$ and the specified model ($\mathbf{H}, \mathbf{L}$) as those maximizing the posterior probability distribution MacKay (1992)

$$P(\alpha, \beta|\mathbf{s}, \mathbf{H}, \mathbf{L}) = \frac{P(\mathbf{s}|\mathbf{H}, \beta, \mathbf{L}, \alpha)P(\alpha, \beta|\mathbf{H}, \mathbf{L})}{P(\mathbf{s}|\mathbf{H}, \mathbf{L})}. \tag{16}$$

If it is unknown what value α and β should have, a flat uninformative prior is usually assumed for $P(\alpha, \beta|\mathbf{H}, \mathbf{L})$. Moreover, the normalization term $P(\mathbf{s}|\mathbf{H}, \mathbf{L})$ does not depend on α and β. Therefore, the optimal values of α and β can be estimated as those maximizing $P(\mathbf{s}|\mathbf{H}, \beta, \mathbf{L}, \alpha)$, a term referred to as the *evidence* MacKay (1992).

Interestingly, the evidence was introduced in the previous section and is determined by

$$P(\mathbf{s}|\mathbf{H}, \beta, \mathbf{L}, \alpha) = \frac{Z_P(\beta, \alpha)}{Z_d(\beta)Z_w(\alpha)}. \tag{17}$$

We could not find available expressions in the literature for the normalization constants appearing in Eq. (17) for the case where the integration variable $\mathbf{w}$ vary from $[0,\infty]$. Therefore, in this study, we estimated these terms. A step-by-step detailed derivation is presented in the Supplementary Material and here we summarize the main results. After some algebraic manipulations, we found

$$Z_w(\alpha) = \frac{1}{\det \mathbf{L}}\left(\frac{\pi}{2\alpha}\right)^{N/2}, \tag{18}$$

and

$$Z_P(\beta, \alpha) = \frac{\exp(-M(\hat{\mathbf{w}}))}{\det \mathbf{U}}\left(\frac{\pi}{2}\right)^{N/2}\prod_{j=1}^N\left[1 + erf(\frac{1}{\sqrt{2}}(\mathbf{U}\hat{\mathbf{w}})_j)\right], \tag{19}$$

where $M(\hat{\mathbf{w}}) = -(\frac{\beta}{2})\|\mathbf{s} - \mathbf{H}\hat{\mathbf{w}}(\lambda)\|_2^2 - (\frac{\alpha}{2})\|\mathbf{L}\hat{\mathbf{w}}(\lambda)\|_2^2$, $\mathbf{U}$ is the upper triangular matrix obtained after applying the Cholesky decomposition to matrix $\mathbf{A} = \beta\mathbf{H}^T\mathbf{H} + \alpha\mathbf{L}^T\mathbf{L}$ (i.e., $\mathbf{A}=\mathbf{U}^T\mathbf{U}$), det denotes the determinant, and erf is the error function.

Substituting Eqs. (18), (19), and $Z_d(\beta)$ as defined in Eq. (8) into Eq. (17) we obtain

$$P(\mathbf{s}|\mathbf{H}, \beta, \mathbf{L}, \alpha) = \frac{\exp\left(-M(\hat{\mathbf{w}})\right)}{\det \mathbf{U}} \left(\frac{\pi}{2}\right)^{N/2} \left(\frac{\pi}{2\alpha}\right)^{-N/2} \left(\frac{2\pi}{\beta}\right)^{-k/2}$$

$$\times \det \mathbf{L} \prod_{j=1}^{N} \left[1 + erf\left(\frac{1}{\sqrt{2}}(\mathbf{U}\hat{\mathbf{w}})_j\right)\right]. \tag{20}$$

The optimal values of α and β are those minimizing $J = -\log P(\mathbf{s}|\mathbf{H}, \beta, \mathbf{L}, \alpha)$:

$$J = M(\hat{\mathbf{w}}) + \log(\det \mathbf{U}) - \sum_{j=1}^{N} \log\left[1 + erf\left(\frac{1}{\sqrt{2}}(\mathbf{U}\hat{\mathbf{w}})_j\right)\right]$$

$$- \frac{N}{2}\log\left(\frac{\pi}{2}\right) + \frac{k}{2}\log(2\pi) - \frac{k}{2}\log(\beta) + \frac{N}{2}\log(\pi)$$

$$- \frac{N}{2}\log(2\alpha) - \log(\det \mathbf{L}). \tag{21}$$

In our work, a faster evaluation of Eq. (21) was implemented by assuming that a good estimate for β (i.e., the noise) is available, in which case Eq. (21) is minimized only for α. Concretely, β was estimated as

$$\hat{\beta} = \frac{1}{\hat{\sigma}^2}, \hat{\sigma}^2 = \frac{(\mathbf{s} - \mathbf{H}\hat{\mathbf{w}}(\lambda=0))^T(\mathbf{s} - \mathbf{H}\hat{\mathbf{w}}(\lambda=0))}{k-p}, \tag{22}$$

where $\hat{\mathbf{w}}(\lambda=0)$ is the sparse solution provided by the non-regularized NNLS and p is the number of positive elements in the solution. In a preliminary evaluation, we verified that this approximation provides an acceptable estimate for the underlying noise level. Alternatively, the evaluation of Eq. (22) can be avoided if an independent estimate for the noise level is available.

2.4. Metrics derived from the T_2 spectrum

Although the primary aim of this work is on determining MWF, we also computed three additional metrics derived from the T_2 spectrum which are of potential interest for clinical applications, including the total water content (TWC), IE water fraction (IEWF), and T_2 of the IE water (T_2^{IE}). For this, cutoff T_2 values that separate the spectrum into regions corresponding to distinct tissue compartments were defined. The cutoff values for the myelin water T_2^m and IE water T_2^i were fixed to standard values previously reported (MacKay and Laule, 2016): $T_2^m = 40ms$ and $T_2^i = 200ms$.

MWF was calculated as the area under the curve for T_2 times smaller than the myelin water cutoff T_2^m, normalized by TWC:

$$MWF = \frac{1}{TWC} \sum_{j=1}^{m} \hat{\mathbf{w}}_j, \tag{23}$$

where m is the index of the predefined T_2 grid that matches the defined cutoff T_2^m and TWC (i.e., the predicted signal for TE = 0) was estimated by calculating the total area under the T_2 spectrum (Meyers et al., 2017)

$$TWC = \sum_{j=1}^{N} \hat{\mathbf{w}}_j, \tag{24}$$

Similarly, IEWF was estimated as the area under the curve in the range $T_2^{m+1} - T_2^i$, normalized by TWC

$$IEWF = \frac{1}{TWC} \sum_{j=m+1}^{i} \hat{\mathbf{w}}_j, \tag{25}$$

where i is the index of the T_2 grid corresponding to the cutoff T_2^i for the IE compartment and T_2^{m+1} is the T_2 time for the grid point with index $m+1$. Finally, T_2^{IE} is computed as the geometric mean of the spectrum in the range $T_2^{m+1} - T_2^i$ Bjarnason (2011):

$$T_2^{IE} = \exp\left(\sum_{j=m+1}^{i} \hat{\mathbf{w}}_j \log(T_2^j) / \sum_{j=m+1}^{i} \hat{\mathbf{w}}_j\right). \tag{26}$$

2.5. Implemented algorithms

We compared the conventional X^2 residual fitting and L-curve criteria with the proposed Bayesian Regularization approach, named BayesReg. Specifically, Eqs. (7) and (21) were minimized using Brent's

Table 1
Seven algorithms were evaluated in this study. All the algorithms are based on solving Eq. (2) for different choices of $\mathbf{L}$ and λ.

Method	Matrix ($\mathbf{L}$)	Regularization parameter (λ)
NNLS	N.A.	N.A.
X^2-I	$\mathbf{L} = \mathbf{I}$	Eq. (7), $c = 1.02$
L-curve-I	$\mathbf{L} = \mathbf{I}$	L-curve (Ref. (Castellanos et al., 2002))
BayesReg-I	$\mathbf{L} = \mathbf{I}$	Eqs. (21) and (22)
X^2-D^{-1}	$\mathbf{L} = \mathbf{D}^{-1}$	Eq. (7), $c = 1.02$
L-curve-D^{-1}	$\mathbf{L} = \mathbf{D}^{-1}$	L-curve (Ref. (Castellanos et al., 2002))
BayesReg-D^{-1}	$\mathbf{L} = \mathbf{D}^{-1}$	Eqs. (21) and (22)

Table 2
Parameters used to generate the synthetic T_2 spectra. $x \sim U(a,b)$ denotes a random sample from a uniform distribution on the interval $[a, b]$; T_2 time is in milliseconds; std: standard deviation.

Parameters	Fast component	Slow component
Water fraction	$MWF \sim U(0.05, 0.25)$	$1 - MWF$
T_2 mean value	$T_2^M \sim U(15, 35)$	$T_2^{IE} \sim U(60, 90)$
T_2 std	$std(T_2^M) \sim U(1, 3)$	$std(T_2^{IE}) \sim U(6, 12)$
Flip angle	$FA \sim U(90°, 180°)$	
Signal to noise ratio	Simulation #1: $SNR \sim U(50, 100)$	
	Simulation #2: $SNR \sim U(100, 200)$	
	Simulation #3: $SNR \sim U(200, 400)$	
	Simulation #4: $SNR \sim U(400, 1000)$	

method (Brent, 1971) included in the *SciPy* python package (i.e., *fminbound* function) (Virtanen et al., 2020). The L-curve was evaluated for 50 logarithmically spaced regularization values ranging from 1e-8 to 10 and the optimal λ was selected as described in (Castellanos et al., 2002). It is important to note that each regularization criterion was implemented for both the standard and alternative regularization forms (i.e., $\mathbf{L} = \mathbf{I}$ and $\mathbf{L} = \mathbf{D}^{-1}$), resulting in seven estimation algorithms, which were named based on the resulting combinations: X^2-I, L-curve-I, BayesReg-I, X^2-D^{-1}, L-curve-D^{-1}, BayesReg-D^{-1}, and the non-regularized NNLS. For more details see Table 1.

2.6. Experimental data

2.6.1. Numerical simulations and evaluation metrics

The implemented algorithms were compared across a range of simulated T_2 spectra and SNRs. In particular, several T_2 spectra consisting of a mixture of two Gaussians with different mean values and standard deviations were generated, along with their corresponding synthetic signals. Various MWFs, T_2s, flip angles (FAs), and SNRs were considered to simulate the heterogeneity observed in the WM of real brain data. For more details see Table 2.

Each simulated T_2 spectrum was discretized into 1000 equidistant points in the range of 1–300 ms. The EPG model was used to create the signals for each T_2 and TEs. The simulated signal $S_d(TE)$ was calculated by taking the sum of all signals with different T_2, weighted by the intensities of the T_2 discrete spectrum. The resulting signals were contaminated with MRI Rician noise as follows:

$$S(TE) = \sqrt{(S_d(TE) + \varepsilon_1)^2 + \varepsilon_2^2} \tag{27}$$

where ε_1 and ε_2 are two different noise realizations from a Gaussian distribution with zero-mean value and standard deviation σ, which was adjusted for each predefined SNR according to the relationship $SNR = S_d(TE_{min})/\sigma$, where TE_{min} denotes the minimum TE employed in the in vivo human brain acquisition sequence (TE_{min}=10.68 ms). Four independent datasets were generated using the following SNR ranges: 1) 50–100, 2) 100–200, 3) 200–400 and 4) 400–1000. Each dataset consisted of $N = 10,000$ independent voxels for a total of 40,000 simulated voxels.

To facilitate the comparison of the evaluated algorithms, the estimated MWFs were compared with the ground-truth values employing six different metrics: the Mean Absolute Error (MAE), Root-Mean-

Square Error (RMSE), Uncertainty with a 95% confidence level (U95) Gueymard (2014), Pearson correlation coefficient (R); moreover, RMSE was decomposed into its two contributing terms: the centered RMSE (i.e., RMSE after removing the mean of both samples) and the Mean Bias Error (MBE) Taylor (2001). Additionally, two statistics were employed to compare the estimated and ground-truth T_2 distributions: the Mean Jensen-Shannon Distance (MJSD-S) (Endres and Schindelin, 2003) and the Mean Wasserstein Distance (MWD-S) (Vaserstein, 1969) between the estimated and actual spectra. Additionally, we created Cumming estimation plots (Cumming, 2014; Ho et al., 2019) summarizing the statistical comparison among all methods for two of these metrics: 1) the absolute error (AE, defined as the absolute value of the pairwise difference between the ground-truth and estimated MWF for each simulated voxel), and 2) the Jensen-Shannon Distance (JSD) between the ground-truth and the estimated T_2 distribution. We then compared all methods against the best method, i.e., the method that produced the smallest mean value for each metric under study (i.e., AE or JSD) and each SNR level. However, given the enormous sample size ($N = 10,000$ for each SNR), all methods would yield statistically significantly worse metrics than the best method, despite most differences would be minuscule and meaningless in a practical sense (Lin et al., 2013). Therefore, to provide useful statistical tests, we instead assessed whether the metrics of the methods were >10% worse than the metrics of the best method. We used one-tailed t-tests (after verifying the normality of the data distribution) corrected for multiple comparisons using the Bonferroni method (Bland and Altman, 1995).

2.6.2. Human brain data and repeatability metrics

Human brain MET_2 data were acquired from ten healthy volunteers (five men, age range = 23–33 years old) at 3T (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) using a standard 64-channel head/neck coil. The data were collected using a prototype high-resolution 3D multi-echo gradient and spin-echo (GRASE) sequence accelerated with CAIPIRINHA (Piredda et al., 2021). Each subject was scanned twice, over two consecutive scanning sessions using the following parameters: matrix-size = $144 \times 126 \times 134$; voxel-size = $1.6 \times 1.6 \times 1.6$ mm^3; minimum TE = 10.68 ms; N-echoes = 32; ΔTE = 10.68 ms; TR = 1 s; prescribed FA = 180^0; number-of-slices = 84; acceleration factor = 3×2; number of averages = 1; acquisition time = 10:30 min. Two subjects were excluded from the study due to motion artifacts. Figure S1 in the supplementary material shows the raw data for one representative subject for various TEs.

The repeatability of the reconstructions was evaluated through a scan-rescan analysis in which the MWFs computed from the two scans were compared. In a first step, the MWF images from both scanning sessions were non-linearly registered to the ICBM-DTI-81 white-matter tract labels atlas (Mori et al., 2008; Oishi et al., 2008) using the ANTs software (Avants et al., 2008). For each subject, a single deformation field was estimated by registering the MET_2 data to the reference T2w image included in the ICBM-DTI-81 atlas. The registration was performed by identifying the 3D volume from the 4D MET_2 data having a contrast similar to that of the T2w image (i.e., $n = 10$, TE = 106.8 ms). The estimated deformation field was applied then to the MWF maps estimated by all methods. 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, for each region of interest (ROI), the mean MWF was estimated. The linear dependence and agreement between the scan-rescan estimates were evaluated by computing the slope and intercept of the linear regression line, Pearson's correlation, and a Bland-Altman analysis from which the standard deviation of the difference and MBE were assessed.

Additionally, we investigated how much the use of different cutoffs T_2^m can influence the estimated MWF values. Specifically, we computed the mean MWF in the WM for the following three cutoffs, T_2^m=37.5, 40, and 42.5 ms. The relative MWF change (estimated as the difference between the MWFs for the longest and shortest cutoffs, 42.5 ms and 37.5 ms, divided by the MWF from the standard cutoff, 40 ms) was used as a measure of robustness against miscalibration (or non-optimality) of the standard cutoff T_2^m.

The study was approved by the local ethics committee and written informed consent was obtained from the participants.

2.6.3. Ex vivo data and histology

Multi-echo spin-echo T2 data previously acquired from a cervical dog spinal cord and the myelin volume fraction (MVF) obtained from histology were employed as described in (Vuong et al., 2017). These datasets are available at the White Matter Microscopy Database (Cohen-Adad et al., 2018; Stikov et al., 2015). The acquisition was carried out on an Agilent 7T animal scanner equipped with 600 mT/m gradients using the following acquisition parameters: field-of-view = $9 \times 9 \times 2$ mm^3; matrix-size = 64×64; voxel-size = $0.14 \times 0.14 \times 2$ mm^3; minimum TE = 10 ms; ΔTE/N-echoes/TR = 10 ms/32/3 s; number-of-slices = 1; number-of-averages = 8. The total acquisition time was 25 min.

The sample (post-fixed paraformaldehyde 4%) was extracted and washed in phosphate-buffered saline (PBS) solution five days before the scanning session. After the MRI acquisition, the spinal cord specimen was osmified (2% OsO4 for 2 h), embedded in EMbed 812 Resin, cut using a microtome, and polished. A scanning electron microscope (JEOL 7600F, Low-angle backscattered electron mode) was used to image an entire slice of the spinal cord at a resolution of 0.26 μm/pixel. MVF at each pixel was determined by segmenting the slice using the axon and myelin segmentation tool 'AxonDeepSeg' (https://github.com/neuropoly/axondeepseg) based on convolutional neuronal networks (Zaimi et al., 2018).

2.7. Preprocessing and estimation

The human brain MRI and the *ex vivo* datasets were filtered using a 3D total variation algorithm before fitting (i.e., using the *denoise_tv_chambolle* function in the *scikit-image* python toolbox). Previous studies showed that spatial filtering is effective for decreasing the variability of the estimated MWF maps, e.g., see (Bouhrara et al., 2018; Does et al., 2019; Jones et al., 2003). In detail, the noise standard deviation σ for each human brain 3D volume and 2D spinal cord slice was estimated by employing a robust wavelet-based estimator (Donoho and Johnstone, 1994) (i.e., using the *estimate_sigma* function in *scikit-image*) and each volume/slice was then denoised with a weight = 2σ, respectively. This denoising weight was found empirically for our data. In Fig. S2 of the supplementary material the estimated maps resulting from different denoising algorithms are presented.

T_2 spectra were estimated in two steps, as conventionally done. First, the refocusing FA value for each voxel was determined. Then, the intra-voxel T_2 spectrum was computed by using the dictionary **H** generated for the estimated FA and the inversion was carried out by employing the algorithms described in Table 1.

Estimating the optimal FA involves various steps. First, different matrices **H** were generated using the EPG model (Prasloski et al., 2012a), each one corresponding to a fixed refocusing FA value selected from a discrete set of 15 equally spaced values between 90° and 180° A fixed T_2 range from 10 ms to 2000 ms (Prasloski et al., 2012b) with $N = 60$ T_2 logarithmically spaced points was employed for the human brain data, and a shorter T_2 range from 10 ms to 1000 ms was used for the *ex vivo* data. We verified that the water fractions estimated for higher T_2s were negligible in this case. It is important to note that T_2 times of the *ex vivo* and in vivo samples might not be equal due to several factors, including the study of different animal species, anatomical regions, cell types, chemical environments, temperature (Birkl et al., 2016) (i.e., dog spinal cord vs human brain, living vs non-living tissue with chemical fixation), and the use of MRI scanners with different field strengths (i.e., 7T vs 3T). Subsequently, we created a smoothed copy of the acquired MET_2 data by using a Gaussian kernel (i.e., FWMH of 4.8 mm) as sug-

Table 3

Statistical indicators for the evaluated models for different SNR ranges: 50–100, 100–200, 200–400, and 400–1000. Two sets of metrics were considered, one to characterize the estimated myelin water fraction (MWF), and another for the estimated spectrum. Bold values show the two most accurate models regarding each indicator.

SNR=50–100	Quality metrics for MWF						Quality metrics for the whole distribution	
	MAE	RMSE	cRMSE	MBE	U95	R	MJSD-S	MWD-S
NNLS	0.075	0.0963	0.0923	−0.0272	0.2614	0.4354	0.5973	0.0148
X^2-I	0.0588	0.0731	0.0638	−0.0355	0.1901	0.621	0.4076	0.0099
L-curve-I	0.0547	0.0668	0.0651	**−0.015**	0.1829	0.6538	0.4465	0.0129
BayesReg-I	0.0573	0.0703	0.0621	−0.033	0.184	0.656	0.4303	0.0114
X^2-D^{-1}	0.0533	0.067	0.06	−0.0298	0.1762	0.6284	**0.3652**	**0.0082**
L-curve-D^{-1}	**0.0449**	**0.0556**	**0.0537**	**−0.0144**	**0.1514**	**0.6624**	0.397	0.0109
BayesReg-D^{-1}	**0.05**	**0.0621**	**0.0564**	−0.0261	**0.1645**	**0.6731**	0.376	**0.0097**

SNR=100–200	Quality metrics for MWF						Quality metrics for the whole distribution	
	MAE	RMSE	cRMSE	MBE	U95	R	MJSD-S	MWD-S
NNLS	0.0577	0.0768	0.0759	**−0.0117**	0.2115	0.5759	0.5648	0.0143
X^2-I	0.0484	0.0611	0.0475	−0.0384	0.1516	0.7415	0.3627	0.0083
L-curve-I	0.0519	0.0622	0.0476	−0.04	0.1535	0.7645	0.4119	0.0109
BayesReg-I	0.0519	0.0634	0.0454	−0.0442	0.1528	0.7613	0.3879	0.0096
X^2-D^{-1}	**0.0425**	0.0543	0.0457	−0.0293	0.1391	0.7604	**0.3246**	**0.007**
L-curve-D^{-1}	**0.0419**	**0.0508**	**0.0435**	−0.0263	**0.131**	**0.7766**	0.3717	0.0097
BayesReg-D^{-1}	0.0431	**0.0539**	**0.0433**	−0.0321	0.1355	**0.7881**	**0.3342**	**0.0079**

SNR=200–400	Quality metrics for MWF						Quality metrics for the whole distribution	
	MAE	RMSE	cRMSE	MBE	U95	R	MJSD-S	MWD-S
NNLS	0.0491	0.0671	0.067	**−0.0008**	0.1859	0.6562	0.5413	0.0137
X^2-I	0.0399	0.0516	0.0394	−0.0333	0.1272	0.8081	0.3159	0.0069
L-curve-I	0.0498	0.0593	0.0368	−0.0464	0.1368	0.8245	0.3734	0.0092
BayesReg-I	0.0455	0.0559	**0.0357**	−0.043	0.1301	0.828	0.3388	0.008
X^2-D^{-1}	**0.0346**	**0.0452**	0.0382	**−0.0241**	**0.1159**	0.8275	**0.2876**	**0.006**
L-curve-D^{-1}	0.0392	0.0479	0.0362	−0.0314	0.1177	**0.8426**	0.3378	0.0083
BayesReg-D^{-1}	**0.0362**	**0.0464**	**0.035**	−0.0305	**0.114**	**0.8476**	**0.2963**	**0.0066**

SNR=400–1000	Quality metrics for MWF						Quality metrics for the whole distribution	
	MAE	RMSE	cRMSE	MBE	U95	R	MJSD-S	MWD-S
NNLS	0.0407	0.056	0.0542	**0.0139**	0.1527	0.7453	0.515	0.0134
X^2-I	0.0291	0.0395	0.0324	−0.0227	0.1001	0.8648	0.2557	0.0055
L-curve-I	0.0413	0.0506	0.0304	−0.0405	0.1158	0.8664	0.305	0.0072
BayesReg-I	0.0324	0.0426	**0.0295**	−0.0307	0.1015	**0.8778**	0.2631	0.006
X^2-D^{-1}	**0.0256**	**0.0348**	0.0331	**−0.0109**	**0.0942**	0.8727	**0.2509**	**0.005**
L-curve-D^{-1}	0.0341	0.0437	0.0317	−0.0301	0.1058	0.8718	0.2956	0.0068
BayesReg-D^{-1}	**0.0265**	**0.0362**	**0.0299**	−0.0205	**0.0921**	**0.8863**	**0.2499**	**0.0053**

gested in (Drenthen et al., 2019). Afterward, the non-regularized NNLS algorithm was used to fit the smoothed data for each matrix **H**, and the resulting mean square errors were interpolated using cubic B-splines (Prasloski et al., 2012a). Next, the optimal FA was selected at the global minima of the resulting interpolated curve. Finally, a new matrix **H** was generated using the estimated FA, which was used then for estimating the T_2 spectrum (Prasloski et al., 2012a). It is important to note that the T_2 spectrum and derived maps were computed from the TV-denoised data, and not from the Gaussian-filtered copy, thus preserving the high spatial resolution of our acquisitions. To accelerate the estimation, instead of generating a new matrix **H** for each estimated FA, we loaded the optimal **H** from a set of precomputed matrices which were created on a high-resolution grid of FA values from 90° to 180° with a step-size of 0.33° The optimal **H** was selected by identifying the FA grid point closest to the interpolated FA.
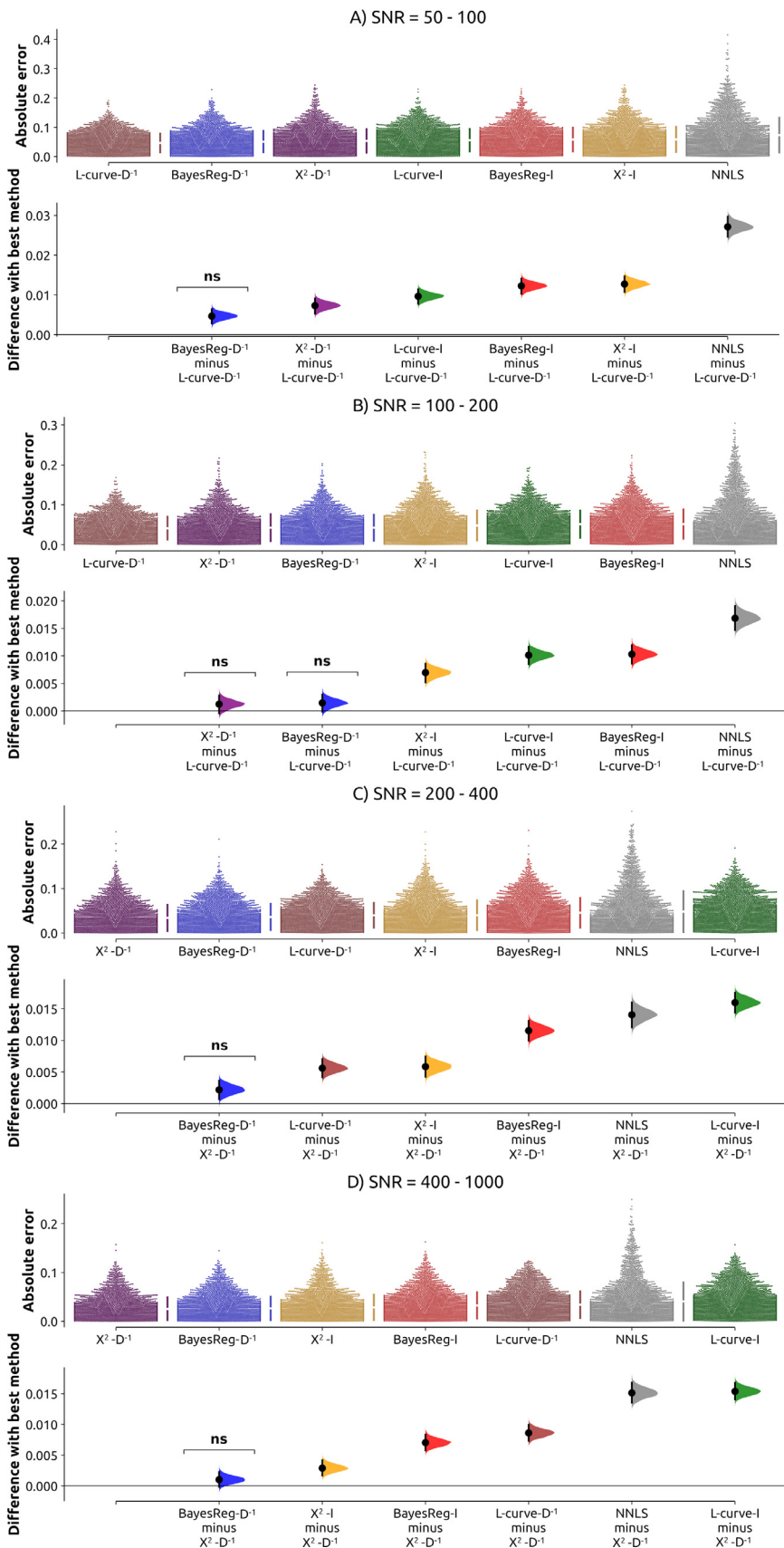
3. Results

3.1. Numerical simulations

Table 3 shows the evaluation metrics for the four SNR ranges considered. For the two lowest SNR ranges (i.e., 50–100 and 100–200) L-

curve-D^{-1} was the optimal method for estimating MWF, followed by BayesReg-D^{-1}. On the other hand, X^2-D^{-1} and BayesReg-D^{-1} produced T_2 distributions closer to the ground-truth spectrums. For the two highest SNR ranges (i.e., 200–400 and 400–1000), X^2-D^{-1} and BayesReg-D^{-1} agree best with the ground-truth, both in terms of MWF and the whole T_2 distribution. Of the poorer performing methods, NNLS has the lowest correlations, and the highest MAEs and RMSEs, demonstrating how important regularization is. Overall, the implementations using the alternative regularization form (i.e., $L = D^{-1}$) showed better metrics than those using the standard form (i.e., $L = I$). In order to summarize how stable each method was, we computed the number of times (n) each method was among the two best methods for each one of the eight metrics considered for all the SNR ranges (see Table 3). The methods showing more stable reconstructions were BayesReg-D^{-1} (i.e., it was selected $n = 27$ times), followed by X^2-D^{-1} ($n = 17$) and L-curve-D^{-1} ($n = 13$).

Results of the quantitative statistical comparison among models are shown in Fig. 1 and the supplementary Figure S3. Fig. 1 depicts Cumming estimation plots comparing all methods in terms of their absolute errors. For the SNR ranges 50–100 and 100–200, L-curve-D^{-1} was the reference method, while X^2-D^{-1} was the reference at the lowest noise levels (i.e., SNR=200–400 and 400–1000). BayesReg-D^{-1} was not significantly different from the best method at each SNR at the predefined

Fig. 1. Cumming estimation plot for the absolute error metric. Results from each SNR range are shown in a different panel. The upper figure of each panel depicts the absolute errors from the simulated voxels (for each method) as a swarmplot, which orders each point to display the underlying distribution. For each method, summary measurements (mean ± standard deviation) are drawn as gapped lines. The accompanying figure below shows the differences among each method and the reference method, selected as the best method for each SNR. All methods were ordered and displayed from left to right based on their performance. The filled curves on the differences axis show the distributions of the mean inter-group differences, where the 95% confidence intervals are illustrated by black vertical lines. Methods that were not statistically significant from the reference method at the predefined tolerance level (i.e., 10%) are signaled (i.e., ns). Conversely, methods that produced errors significantly higher than the tolerance level (i.e., indicating a relative increase >10% compared to the reference method) were not annotated. The statistical significance level was defined for p-values<0.05, corrected for multiple comparisons using the Bonferroni method.
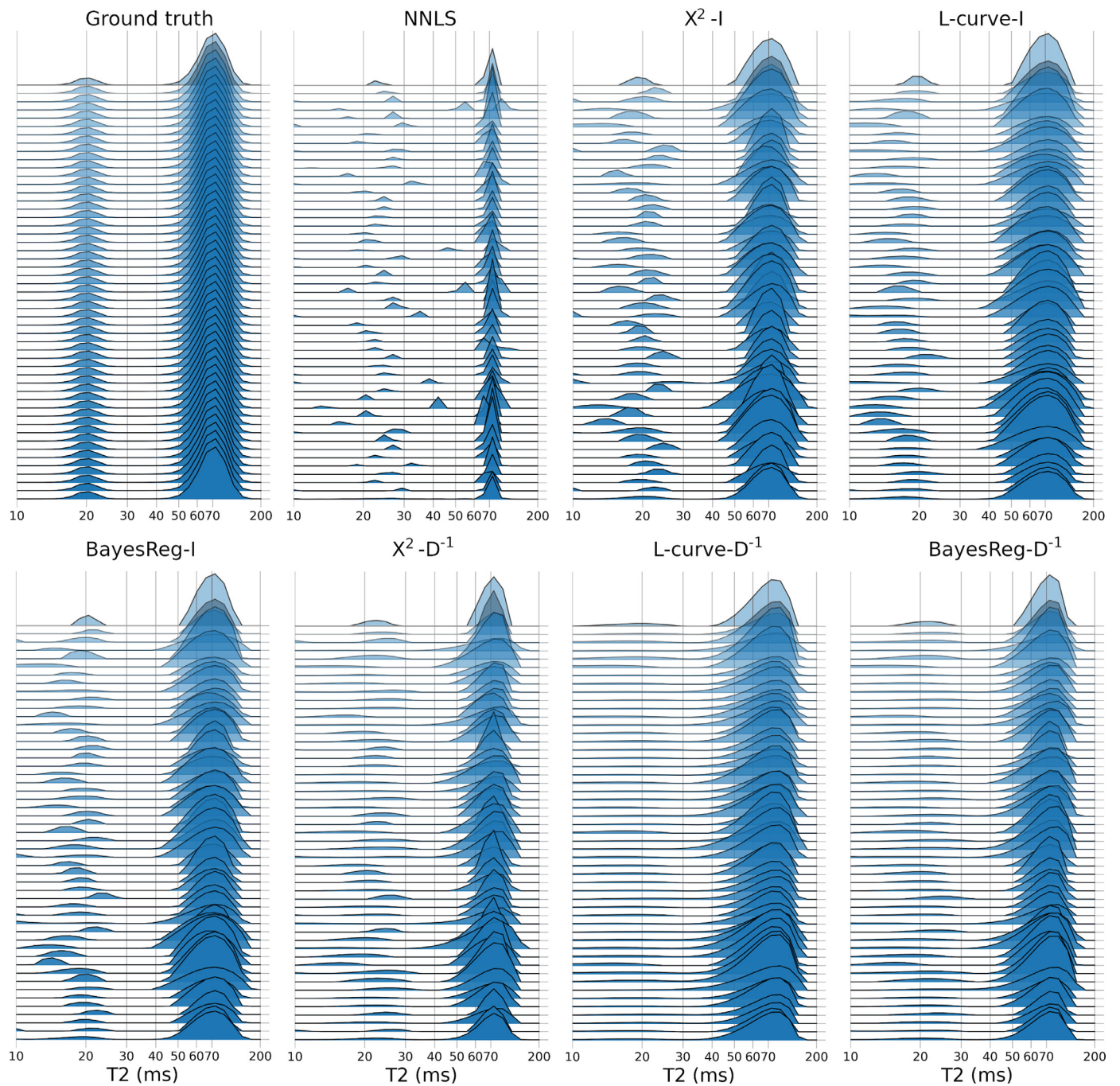
Fig. 2. T_2 distributions estimated by the seven evaluated algorithms for $N = 50$ noisy simulated voxels with random SNRs in the range of 400–1000 and random FAs (see Table 2). The same ground-truth T_2 distribution was simulated for each voxel as a mixture of two Gaussians. The mean value and standard deviation of the dominant Gaussian (simulating the IE water compartment) with a signal fraction of 0.9 were 70 ms and 9 ms, respectively. The non-dominant Gaussian with a signal fraction of 0.1 (simulating the myelin water compartment) had a mean value of 20 ms and a standard deviation of 2 ms. The x-axis of each panel shows the T_2 time in logarithm scale in the range of 10–200 ms, and the y-axis encodes the intensity of the distributions.

tolerance level (i.e., 10%). All the other methods produced absolute errors larger than 10% compared to the reference, except X^2-D^{-1} for SNR=100–200. Results for the Jensen-Shannon Distance are depicted in Figure S3, where X^2-D^{-1} and BayesReg-D^{-1} were the reference methods.

Fig. 2 shows an example of the ground-truth and estimated T_2 distributions for 50 voxels. These signals were corrupted with different noise levels in the range SNR=400–1000. Similar results are reported in Figures S4, S5, and S6 of the supplementary material for the SNR ranges 50–100, 100–200, and 200–400, respectively. Note that for the highest

level of noise (i.e., SNR = 50–100, Fig. S4), it was extremely difficult to accurately recover the simulated T_2 distributions. The parameter showing more uncertainty was the location of the non-dominant lobe (i.e., myelin water): while NNLS recovered it at almost random locations, the regularized algorithms estimated it around the minimum possible T_2. (i.e., 10 ms). This larger uncertainty is expected to increase the MWF estimation error and is in line with results in Table 3, where the MAE for the SNR range 50–100 is around two times higher than for SNR=400–1000.

Table 4

Statistical indicators of the scan-rescan analysis for the evaluated methods in real data. The following quality metrics were estimated for each one of the six subjects: Pearson's Correlation (C), Slope (S) and Intercept (I) of the regression line, Mean Bias Error (MBE), and Standard Deviation of the difference (STD) for all ROIs. The table reports the mean value for all ROIs of all subjects. Bold values show the two most reproducible methods for each indicator.

Method	C	S	I	MBE	STD
NNLS	0.807	0.727	0.019	0.017	0.030
X^2-I	0.864	0.821	0.013	0.008	0.025
L-curve-I	**0.892**	**0.860**	0.013	0.009	0.02283
BayesReg-I	**0.884**	0.849	**0.0124**	**0.007**	0.024
X^2-D^{-1}	0.865	0.820	0.013	0.009	0.024
L-curve-D^{-1}	0.862	**0.857**	0.013	0.0079	**0.021**
BayesReg-D^{-1}	0.883	0.855	**0.0117**	**0.0076**	**0.0228**

3.2. Human brain data

Results of the scan-rescan analysis using in vivo MET$_2$ data are shown in Table 4, and Figs. 3 and 4. Table 4 lists the average metrics estimated for all methods from the studied subjects. While L-curve-I and BayesReg-I produced the highest mean linear correlation, BayesReg-I and BayesReg-D^{-1} had the lowest mean bias error and intercept. Both algorithms using L-curve provided mean slopes that were the closest to one. L-curve-D^{-1} showed the smallest standard deviation of the scan-rescan differences, followed by BayesReg-D^{-1}. Once again, NNLS was the worst method for all the evaluated metrics. Overall, none of the regularization forms (i.e., $\mathbf{L} = \mathbf{I}$ or $\mathbf{L} = \mathbf{D}^{-1}$) was superior to the other. BayesReg and L-curve (i.e., using both the standard and the alternative regularization forms) were more reproducible: BayesReg was among the two best methods for each one of the five evaluated metrics $n = 6$ times, and L-curve $n = 4$ times.

Fig. 3 depicts brain images of four descriptors derived from the T$_2$ spectra for one representative subject. Relevant differences among the images resulting from different methods can be noted. The frontal part of the brain is the region where the estimation of MWF is more difficult for NNLS, where no myelin water was detected in several voxels. This shortcoming is partially overcome by using any of the regularized methods (X^2, L-curve, or BayesReg). Overall, all methods using the alternative regularization form showed a clearer contrast between the white and gray matter tissue in frontal brain regions of the MWF and IEWF images, in comparison to methods using the standard form. In contrast, we do not perceive differences in the TWC maps.

Additional slices of the estimated myelin maps are illustrated in Fig. 4 (panel A), which also shows the relative difference in MWF between each method and X^2-I (panel B), which was taken as a reference here due to its widespread use. As can be noted, the two methods showing larger relative differences are NNLS (the blue color in white matter indicates an underestimation in MWF in relation to X^2-I) and L-curve-D^{-1} (the red color in gray matter and some white matter regions in the frontal lobe indicates an overestimation in MWF in comparison to X^2-I). X^2-I and BayesReg-I are the two most similar methods, and the resulting difference maps from L-curve-I and BayesReg-D^{-1} are in between those from BayesReg-I and L-curve-D^{-1}.

Fig. 5 shows the T$_2$ spectra estimated for one representative slice of one subject. The spectra computed with NNLS and X^2 are sharper than those based on BayesReg and L-curve, the latter producing the smoothest estimates. The individual spectra estimated by X^2 have multiple lobes with well-separated peaks as reported by (Whittall et al., 1997), while those estimated by L-curve are blurred like those reported by (Guo et al., 2013). When comparing methods using the standard regularization form versus the alternative form, the impact of the different priors on the estimated spectra can be noted. While the standard form tends to produce spectra with sharper shapes in the portion of the spectrum correspond-
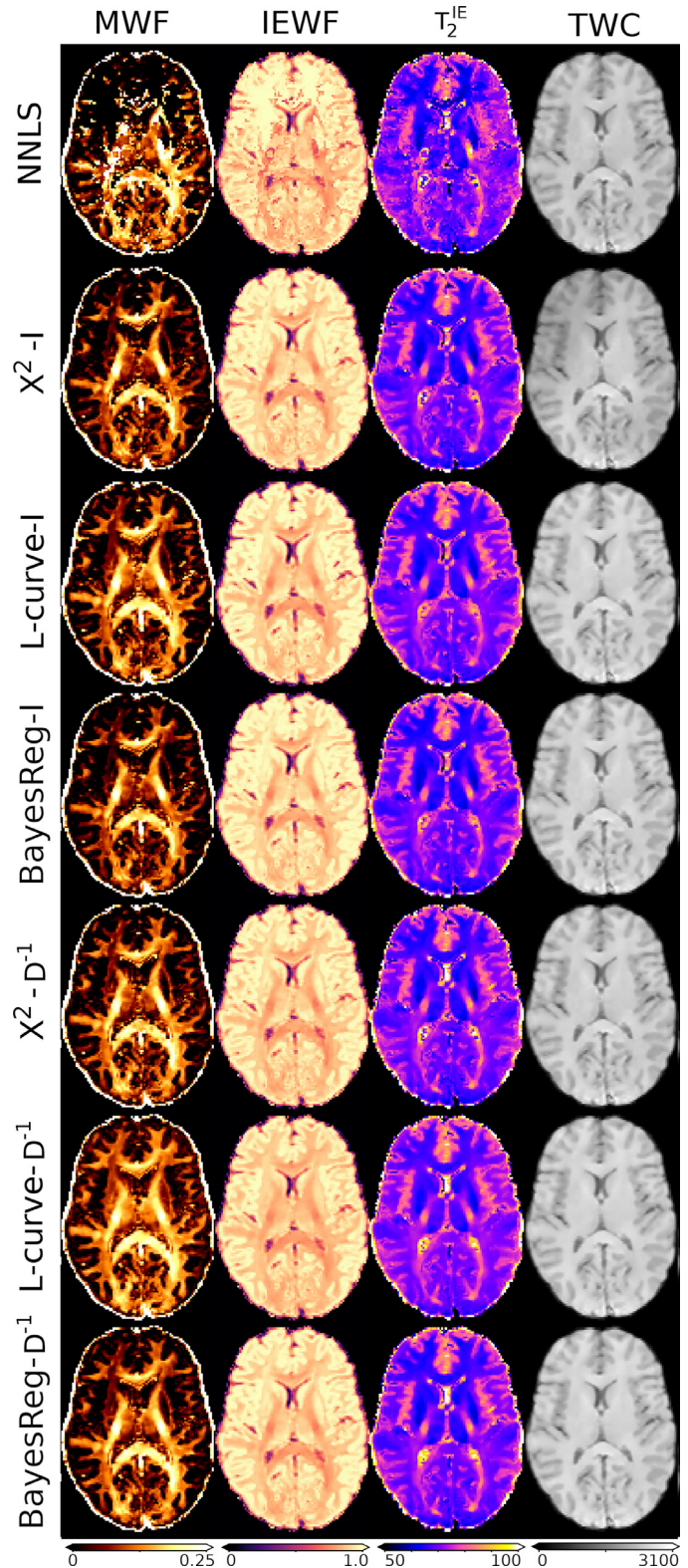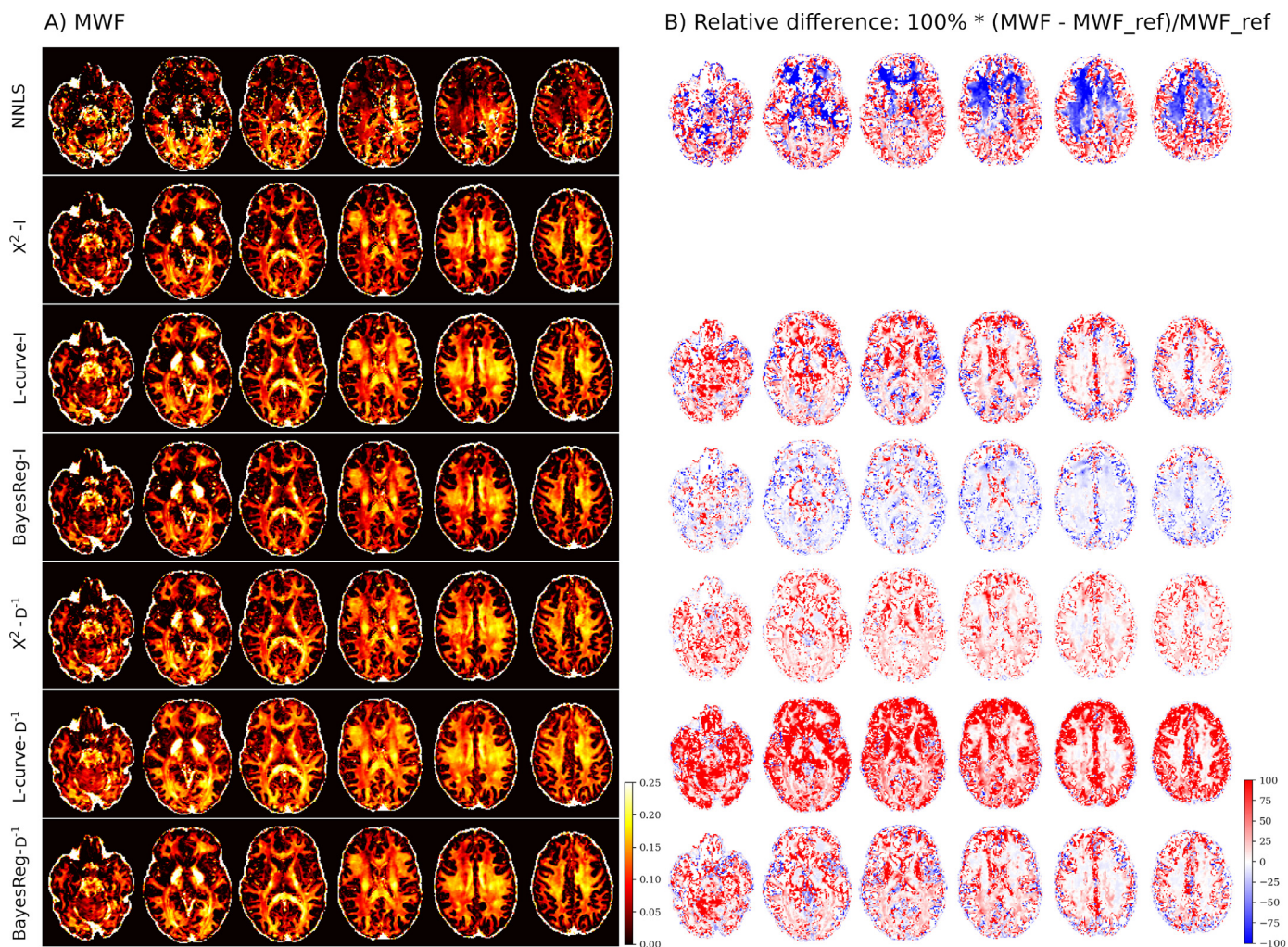

Fig. 3. Metrics derived from the estimated T$_2$ spectra in brain data, including the myelin water fraction (MWF), intra- and extra-cellular water fraction (IEWF), its mean relaxation time T_2^{IE} (in milliseconds), and the total water content (TWC, arbitrary units). The rows show the estimated maps for the seven evaluated algorithms in one representative axial slice.

Fig. 4. Panel A) shows the myelin water fraction (MWF) maps estimated by the methods evaluated in this study for a representative subject and six brain slices. The relative difference between each MWF map and a reference map (MWF_ref) is shown in panel B. To simplify the visualization, the same reference map was used for all cases: we selected the MWF estimated by the X^2-I method. According to the used color scale, red, white, and blue colors correspond to voxels where the MWF values are higher, equal, or lower than those from the reference map, respectively.

Table 5

Mean myelin water fraction in the white matter across subjects estimated for three different myelin T_2 cutoffs: 37.5, 40, and 42.5 ms. The last column reports the relative change between the cutoffs 42.5 ms and 37.5 ms in relation to the default cutoff, 40 ms. Bold values show the two methods with the smallest relative change.

Method	MWF_1 ($T_2^m = 37.5 ms$)	MWF_2 ($T_2^m = 40 ms$)	MWF_3 ($T_2^m = 42.5 ms$)	Relative change: $100\%*(MWF_3 - MWF_1)/MWF_2$
NNLS	0.0886	0.0949	0.1015	**13.54**
X^2-I	0.0703	0.0850	0.1029	37.81
L-curve-I	0.0763	0.1018	0.1314	53.78
BayesReg-I	0.0664	0.0845	0.1069	47.14
X^2-D^{-1}	0.0796	0.0925	0.1076	**29.86**
L-curve-D^{-1}	0.0936	0.1164	0.1420	41.24
BayesReg-D^{-1}	0.0780	0.0942	0.1132	36.91

ing to the myelin water (i.e., $T_2 < 40$ ms) and smoother segments for longer T_2s, the alternative form produces smoother spectra around the myelin water region and sharper segments in the tail of the distribution.

Results showing how the use of different T_2^m cutoffs influenced the estimated MWF values are reported in Table 5. NNLS and X^2-D^{-1} were the methods with the smallest relative MWF change: 14% and 30% respectively, while L-curve-I had the worst performance, about 54%. The methods using the alternative regularization form showed reduced relative changes in comparison to the standard form.

3.3. Ex vivo data

Results from the validation study using histological and *ex vivo* MRI data from a dog spinal cord are illustrated in Fig. 6 and Table 6, which reports the Pearson's Correlation coefficient between the estimated MWF values and the myelin content derived from histology in WM, GM, and both tissue types, respectively. The corresponding correlation plots are shown in Figure S7 in the supplementary material. It is important to note that the predefined regularization parameters used
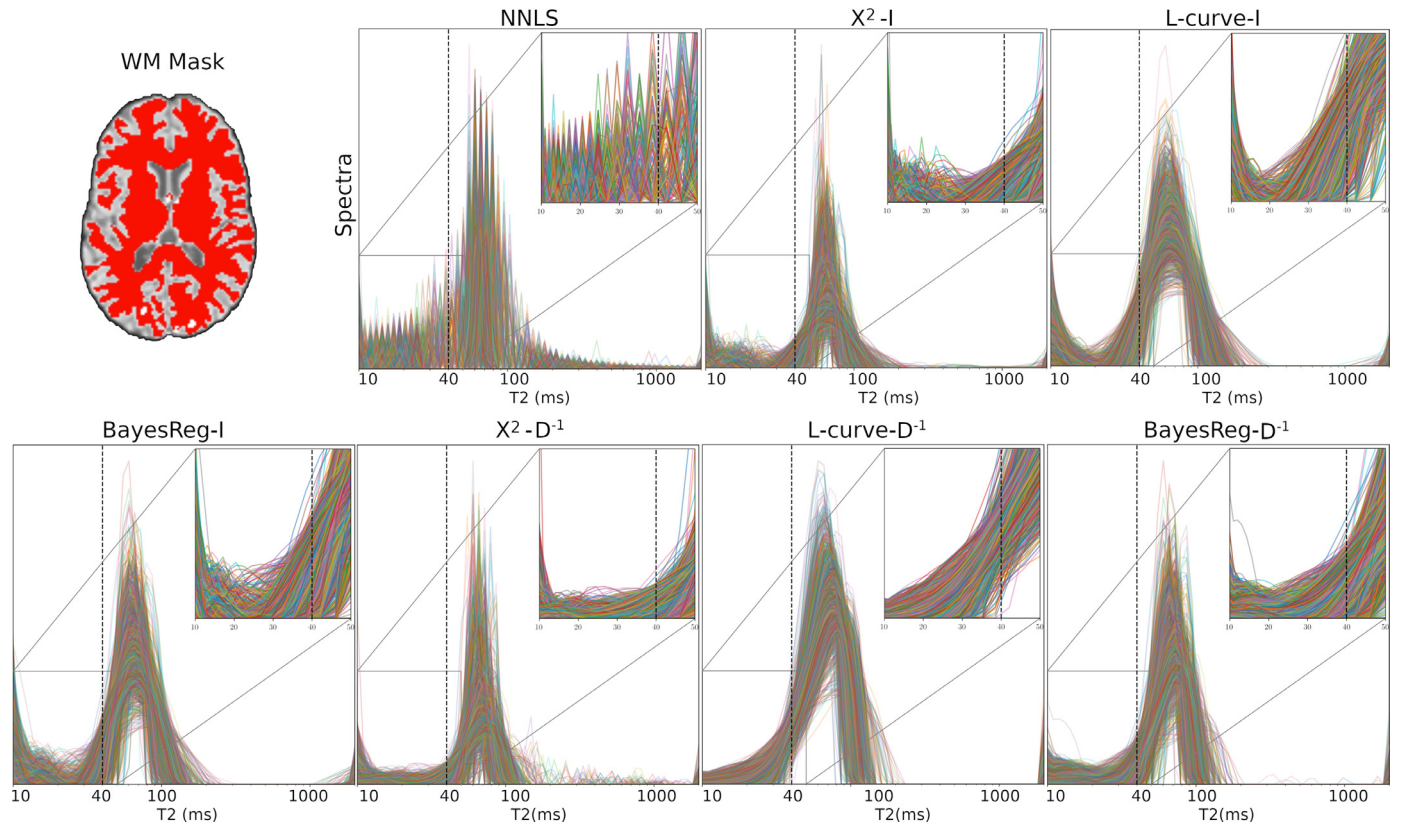
Fig. 5. T_2 spectra estimated by the seven evaluated algorithms from the real MRI data for one representative subject. Each panel shows the T_2 spectra estimated for all the voxels (∼3000) within a white matter (WM) mask (in red). The x-axis depicts the T_2 times in logarithmic scale in the range of 10–2000 ms and the y-axis quantifies the intensity of the spectra. To facilitate the visualization of the spectrum's portion corresponding to the myelin water, the T_2 range 10–50 ms is zoomed. The vertical dashed black line located at $T_2 = 40$ ms indicates the T_2 cutoff used to estimate MWF. The black spectrum is the average T_2 spectrum obtained after computing the point-by-point mean intensity for all spectra.

Table 6

Pearson's Correlation among the estimated myelin water fraction maps and the myelin content. Bold values show the most correlated methods for each evaluated mask delineating the white matter (WM), gray matter (GM), and both tissue types (WM+GM).

Method	WM	GM	WM+GM
NNLS	0.516	0.848	0.933
X^2-I	**0.553**	0.858	**0.938**
L-curve-I	0.502	0.866	0.934
BayesReg-I	0.536	0.861	0.937
X^2-D^{-1}	0.54	**0.869**	**0.938**
L-curve-D^{-1}	0.526	0.868	**0.938**
BayesReg-D^{-1}	0.53	**0.869**	**0.938**

for X^2 (i.e., $k = 1.02$) and L-curve (i.e., $\lambda = [1e\text{-}8, 10]$) were suboptimal in this data, producing over-smoothed spectra. As a result, all metrics reported for both methods here were obtained by modifying these criteria for improving the performance. Specifically, we used $k = 1.01$ for X^2 and the interval $\lambda = [1e\text{-}8, 1e\text{-}2]$ for L-curve. Moreover, the myelin water T_2^m cutoff was reduced from 40 ms to 35 ms, as a better separation between the myelin and the intra/extra-cellular water lobes was observed for this value, for all methods.

The methods exhibiting the highest correlations with histology were X^2-I in WM and BayesReg-D^{-1} and X^2-D^{-1} in GM. When both tissues were analyzed together, the three methods using the alternative regularization form and X^2-I had the highest correlations. The estimated FAs were in a narrow interval from 150^0 to 155^0, suggesting that the B1

field was homogenous across the slice in this sample. MWF values reported from all methods are in the range 0–0.35 in GM, and 0.35–0.5 in WM (see Fig. 6), which are much higher than those estimated in human brains.

4. Discussion

In this work, we evaluated two alternative regularization forms for solving the T_2 spectrum imaging inverse problem. Moreover, we developed a new Bayesian regularized non-negative least squares method 'BayesReg' for selecting the optimal regularization parameter. BayesReg is based on an established Bayesian formulation MacKay (1992), which we extended for the case of non-negative variables, a requirement that the estimated spectrum must satisfy to have physical meaning.

Using synthetic data, three methods for determining the optimal regularization parameter were compared (X^2, L-curve, and BayesReg) for various SNR ranges, encompassing the noise levels expected in real brain data. While L-curve-D^{-1} showed the best performance to estimate MWF for high levels of noise, X^2-D^{-1} outperformed the former for low noise levels. Notably, BayesReg-D^{-1} showed competitive results across all SNR levels for all the evaluated metrics, including the metrics related to MWF as well as those metrics comparing the simulated and estimated T_2 distributions. When compared results from the two evaluated regularization matrices, methods using the alternative form $\mathbf{L} = \mathbf{D}^{-1}$ had an overall superior performance. These results extend and update our previous findings (Canales-Rodríguez et al., 2021) where we reported a superior performance for L-curve-I and X^2-I in comparison to two other regularization Laplacian matrices (i.e., $\mathbf{L}_1$ and $\mathbf{L}_2$). For more details see Table 3, Fig. 1, and the supplementary Figure S3. A plausible explanation for un-
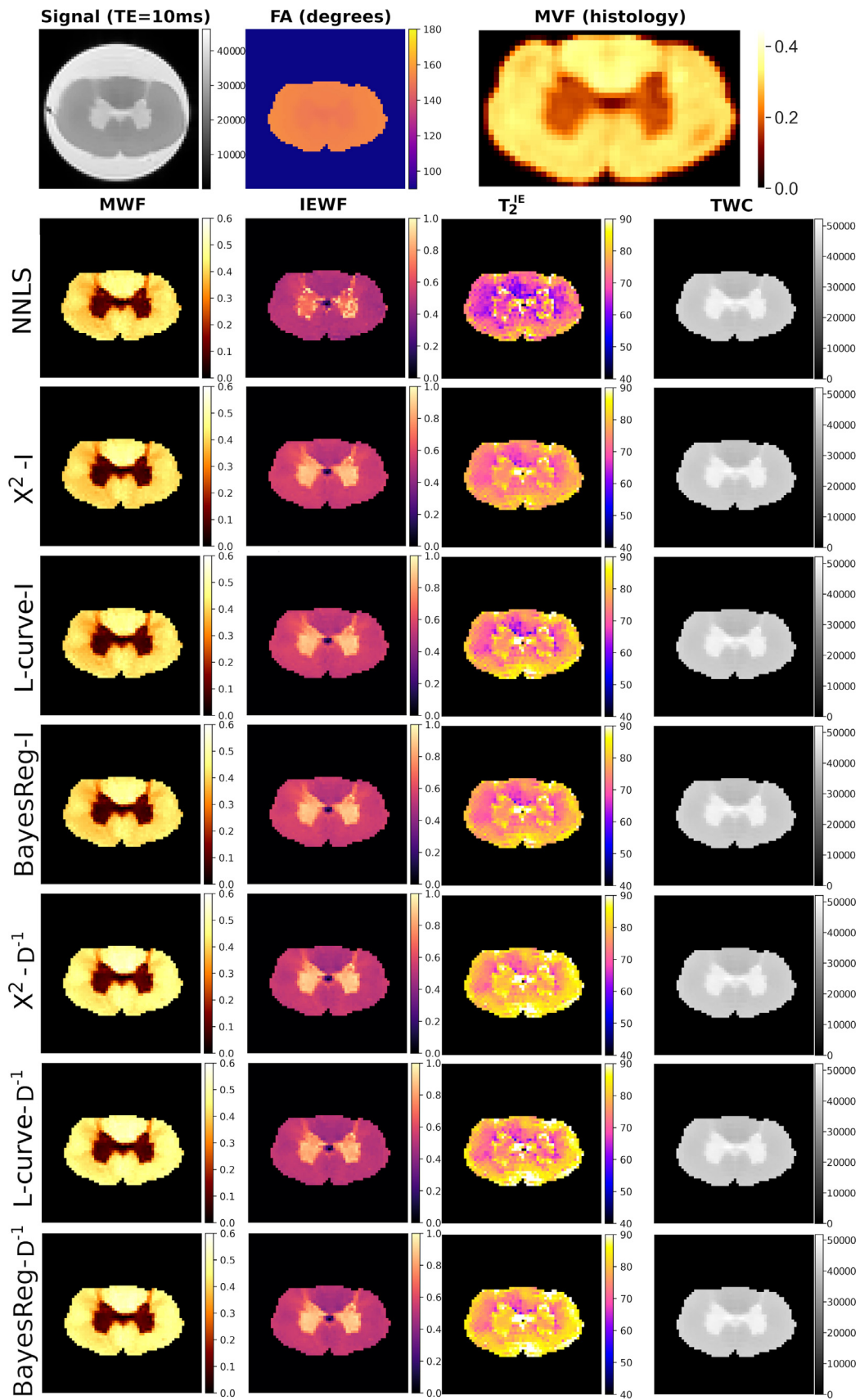
Fig. 6. Metrics derived from the estimated T_2 spectra in the *ex vivo* spinal cord data, including the myelin water fraction (MWF), intra- and extra-cellular water fraction (IEWF), its mean relaxation time T_2^{IE} (in milliseconds), and the total water content (TWC, arbitrary units). The upper row shows the first echo image (TE=10 ms), the estimated flip angle (FA, in degrees), and the myelin content (myelin volume fraction, MVF) from the histological analysis, respectively, and the rows below show the estimated maps for the seven evaluated algorithms.

derstanding the success of the alternative regularization form is related to the fact that, in contrast to the standard form, the elements of the discrete T_2 distribution are not penalized based on their bin widths. As a result, the estimated spectra (in comparison with those estimated by the standard form) are smoother around the portion with short T_2 components and sharper around the long T_2 components. The increased regularization exerted by the alternative form over the spectrum's segment corresponding to the myelin water helps to reduce the high variability in locating this non-dominant component (e.g., see Figures S5 and S6), and consequently, the uncertainty of the estimated MWF values.

As the results from in vivo human brain data cannot be validated against ground truth, we focused on characterizing the reproducibility of the estimates through a scan-rescan analysis. Our findings suggest that BayesReg and L-curve (i.e., using both the standard and the alternative regularization forms) are more reproducible. These results are expected if we consider that the SNR of our real data ranged between 50 and 250, and indeed, these two methods showed superior performance in the synthetic evaluation for the SNR range 50–200. The visual inspection of the estimated MWF maps revealed that all the regularized methods depicted a better contrast between white and gray matter tissues and were less affected than NNLS in detecting the myelin water in frontal brain regions. Moreover, the maps resulting from the alternative regularization form showed a slightly superior image contrast. We noted that in a few subjects, there was an asymmetrical pattern of MWF values in the frontal lobe (e.g., see Fig. 3). This was also reported in the previous study where we introduced the MET_2 sequence employed in this work (Piredda et al., 2021). The frontal lobe may be the most problematic white matter region for MWF quantification because, in general, it shows higher B1 inhomogeneities (for more details see the supplementary Figure S8), and any residual error in determining the correct flip angle could affect the estimation. Although we did our best to minimize flip angles errors by using a 3D sequence, residual errors across the whole brain can exist for 3D sequences, even after EPG modeling. Future work should focus on exploring different algorithms for this task and on comparing the estimated flip angle maps with acquired B1 inhomogeneity maps. In previous studies, smaller MWF values in the frontal lobe have been reported, e.g., (Lee et al., 2020). A reduced myelin volume may increase the water exchange between the myelin and IE compartments, which may shift the T_2 lobes corresponding to these water components, respectively. The closer these lobes are together, the more difficult it is to estimate them. We have verified that in some frontal lobe voxels, a single T_2 lobe was recovered, partially spreading over the myelin and IE T_2 segments. This result may be in line with other studies: for example, (Lee et al., 2020) found an MWF close to zero in the frontal lobe when the myelin cutoff T_2 value was reduced from 40 ms to 30 ms, indicating that the myelin water lobe is shifted to the right side of the predefined interval, and could be partially merged with the IE water lobe. On the other hand, the high acceleration factor (i.e., k-space undersampling) required to collect our high-resolution data in a clinically acceptable time, produced ghosting artifacts in the frontal lobe that may have affected the myelin quantification in this region (Piredda et al., 2021). Recent work suggests that MWF measured using a GRASE sequence with TR on the order of 1 s may experience biases due to fiber-orientation dependence effects and the different T_1 relaxation times of the different water compartments. These effects could impact the in vivo results presented here. According to that study, a longer TR would reduce the estimated 'apparent' MWF (Birkl et al., 2021).

The visual assessment of the estimated T_2 spectra in human brain data showed interesting results. X^2 produced spectra with the clearest multi-lobular structure whilst those forecasted by L-curve were much smoother. BayesReg produced spectra with smoothness in-between X^2 and L-curve. It seems that for L-curve, a portion of the area under the spectra assigned to the myelin water is taken from the left side of the lobe assigned to the IE water compartment, and vice-versa, i.e., a fraction of the signal assigned to the myelin is transferred to the IE water pool. The blurring introduced by L-curve in our real data is not nec-

essarily a drawback if one is mainly interested in determining MWF. According to our synthetic results, the area under the curve within the T_2 cutoff defining MWF is relatively preserved, despite the blurring introduced by L-curve. Moreover, the estimated MWF values in our real data are within the range reported in the literature. Nevertheless, we don't fully recommend using L-curve if the main goal of the research is to estimate the whole spectrum. Our analysis evaluating how much the estimated MWF is affected by changing the myelin T_2 cutoff revealed that the method producing smoothest spectra, L-curve-I, is the most affected, while the method producing sparsest solutions, NNLS, is the least affected. The greater stability and reproducibility of the methods promoting smooth solutions is paid for by a greater dependence on the cutoff parameter. Therefore, methods like X^2-D^{-1} and BayesReg-D^{-1}, showing an intermediate behavior between NNLS and L-curve-I, could be better alternatives.

Results from the *ex vivo* multi-echo T_2 data revealed a high correlation among the histology estimates and the MWF maps calculated from all the evaluated methods, including NNLS (despite the resulting maps in Fig. 6 are noisier), suggesting that for very low noise levels all methods provide similar estimates (see the supplementary Figure S7). Interestingly, the correlations estimated in this study in WM regions are higher than those reported in previous studies employing the same MRI and histological data, e.g. (Vuong et al., 2017; Yu et al., 2019), and are similar to those reported for our recently proposed model-informed machine learning approach (Yu et al., 2021), suggesting that our implementations are competitive. X^2-I exhibited the highest correlation with histology in WM, while BayesReg-D^{-1} and X^2-D^{-1} achieved the highest correlations in GM. If we consider that the estimated MWF values in the GM of the dog spinal cord are in the same range as the MWF reported in the WM of human brains (i.e., 0.05–0.3), then our *ex vivo* GM results may provide a better indicator of the expected performance in human brain data.

This work has some limitations. First, the scan-rescan analysis is based on ten subjects acquired on the same scanner, of which two were discarded due to strong motion artifacts. Future studies should be conducted on larger populations, using different scanners, and employing different acquisition sequences. This is a requirement to verify if our findings can be extrapolated to other imaging sequences using different acquisition parameters with different noise levels and k-space sampling approaches, e.g., (Dvorak et al., 2020). From a methodological point of view, our formulation may be generalized to consider other penalty terms (i.e., prior distributions) based on the $l1$-norm (Canales-Rodríguez et al., 2019) promoting sparse spectra, e.g., (Song et al., 2020; Zimmermann et al., 2019), as well as Likelihood distributions based on non-Gaussian noise models, e.g., (Canales-Rodríguez et al., 2015). Nevertheless, for the relatively high SNRs associated with multi-echo T_2 data (i.e., >50), the Gaussian noise model could provide an adequate approximation, and we don't expect a significant improvement by considering Rician or Non-central Chi noise models. Indeed, in a preliminary implementation of BayesReg we obtained similar results for Gaussian and Rician likelihood functions (results not shown), but the computation time associated with the latter was significantly longer because the estimation required the evaluation of special functions and it cannot be conducted utilizing the NNLS solver (Canales-Rodríguez et al., 2015).

Additional work should be done to identify the best denoising method. In a preliminary evaluation, we visually inspected the output of various algorithms and identified TV as an appropriate method for our data. Notably, the Marchenko-Pastur PCA (MP-PCA) denoising method (Cordero-Grande et al., 2019; Does et al., 2019; Veraart et al., 2016b, 2016a), which is a state-of-the-art technique commonly employed for improving the quality of diffusion MRI and multi-echo T_2 data, did not work properly (see the supplementary Fig. S2). This may be related to violations of the assumptions behind this method on our data. Because of the employed acceleration factor and multi-coil k-space reconstruction technique, the noise is correlated in space and time in our acquisitions (i.e., it is not independent in adjacent voxels, and for a given voxel, it is

not independent at different echo times). Furthermore, it does not follow a Gaussian distribution as is assumed by MP-PCA. A similar result was reported in (Does et al., 2019), where the automatic criterion for selecting the number of principal components based on the MP distribution failed in real multi-channel data. It was beyond the scope of this work comparing different denoising algorithms, and our findings may not be extrapolated to other datasets acquired using different MRI sequences.

The correlation analysis in *ex vivo* data is based on comparing MWF and MVF. Although these metrics are correlated, they do not portray the same information. Moreover, the slice thickness of the histological data is thinner than that of the MRI data. Therefore, both datasets do not characterize the same tissue volume, although significant differences in adjacent regions along the spinal cord are not expected. Additionally, despite the fact that histological data is considered as a reference value, different myelin segmentation algorithms may produce slightly different results. Further discrepancies may arise due to the manipulation steps (i.e., chemical fixation and slicing) required for analyzing the sample. Finally, as the estimated MWF values are much higher than those reported in human brain data, and the noise level is much lower due to the averaging of several repetitions, *ex vivo* results may correspond to a best-case scenario. Hence, our findings should not be extrapolated to results from real data and should not be taken as a definite ranking among the evaluated methods.

To establish the best possible MWF estimation methodology, the implemented methods should be compared with other alternative implementations, like those using Monte Carlo (Yu et al., 2019) and multi-voxel fitting (Kumar et al., 2018, 2016). It is important to note that our study is limited to voxelwise regularization algorithms and that we did not implement spatially regularized approaches, like those introduced in (Kumar et al., 2018, 2016). Consequently, our results should be taken as an indicator of reconstruction quality, and not as a definitive ranking. Future studies shall be carried out to generalize our implementations by considering 3D spatial correlations. In this line, it should be investigated which approach is more effective, whether to remove the noise from the data by taking into account the spatial correlations (i.e., like in TV) or whether to constrain the estimation by taking into account the spatial information (i.e., as in (Kumar et al., 2018, 2016)). Given the significant improvements achieved by machine learning algorithms in several research domains, the proposed approaches should be compared with recent techniques based on neural networks, e.g., see (Lee et al., 2020; Yu et al., 2021).

It is important to note that our findings in synthetic data may not be fully extrapolated to results from real data because the EPG model used to construct the kernel-dictionary of synthetic signals only considered the flip angle errors, but other potential confounding factors that modulate the measured signals were not taken into account, including the effect of diffusion (Weigel, 2015; Weigel et al., 2010) due to internal gradients caused by magnetic susceptibility differences at the tissue-fluid interfaces, water exchange due to cell permeability, magnetization transfer effects (Malik et al., 2018), and anisotropic T_2 relaxation (Gil et al., 2016). Future studies should investigate to which extent these phenomena affect the myelin water quantification, and which biophysical model is more appropriate. Another factor affecting the estimation is the selection of proper cutoff T_2 values for separating the myelin and IE water pools. The optimal cutoff may vary for different brain regions with different microstructure properties and in abnormal/damaged tissue. Ideally, the cutoff values may be individually selected for each voxel by using information about the location, width, and distance between the T_2 lobes. However, such an approach would be compromised by the high variability in the location of the myelin water lobe.

The total computation time (i.e., including all the involved steps, from the preprocessing (i.e., TV denoising and brain extraction) to the estimation of flip angles, T_2 distributions, derived metrics (e.g., myelin map), and creation of quality control plots) for a whole-brain image on a personal laptop (Intel Core i7, 7th Gen, 8 CPUs, 2.7 GHz, 16 GB Ram) with parallel processing for NNLS/X^2/L-curve/BayesReg

was 8 min/23 min/1:02 h/4:00 h, respectively. The increased computation time for BayesReg is related to the matrix operations involved in Eq. (21), which are re-estimated within the iterative optimization process for different values of the regularization parameter. This makes the computation of MWF maps at the scanner not possible during the course of a standard examination. Therefore, we currently recommend its use for neuroimaging studies where the MWF maps are not used for an immediate clinical evaluation. There are some strategies for accelerating the BayesReg estimation process, e.g., by using a computer equipped with more CPUs or by reducing the number of T_2 grid-points from $N = 60$ to 40, as is done in many studies. We verified that the latter option alone decreased the BayesReg estimation time by half without reducing its performance. Other strategies shall be explored in the future, such as defining a discrete grid of regularization parameters (i.e., as in L-curve) and pre-computing the more costly matrix operations for this grid, using a more optimized programming language (Doucette et al., 2020), utilizing a GPU, e.g.,(Blas et al., 2016), and training a neural network for predicting the solutions provided by BayesReg from the measured data, e.g., see (Lee et al., 2020)."

Other well-known criteria for estimating the regularization parameter, like the Akaike Information Criterion (AIC) and the Bayesian information criterion (BIC), were not evaluated. In a preliminary implementation of these techniques, we noted that these criteria produced inferior results (results not shown), and consequently we did not include them in this report. Additionally, a third level for the Bayesian inference framework MacKay (1992) could be implemented for comparing the solutions obtained from different models and prior distributions. Further understanding of the capabilities of the implemented methods may be attained by identifying microstructure abnormalities in patients (e.g., see (Canales-Rodríguez et al., 2013)) and age- and gender-related differences in healthy controls (Canales-Rodríguez, 2021). As the proposed BayesReg algorithm is general, it may be useful in other applications involving the estimation of non-negative variables, like in spherical deconvolution of diffusion MRI data (Canales-Rodríguez et al., 2019). All the algorithms evaluated in this work were included in our multi-component T_2 reconstruction toolbox available at https://github.com/ejcanalesr/multicomponent-T2-toolbox.

Declaration of Competing Interest

Credit authorship contribution statement

Erick Jorge Canales-Rodríguez: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Supervision, Writing – original draft, Writing – review & editing. **Marco Pizzolato:** Investigation, Data curation, Software, Writing – review & editing. **Thomas Yu:** Investigation, Data curation, Software, Writing – review & editing. **Gian Franco Piredda:** Investigation, Data curation, Software, Writing – review & editing. **Tom Hilbert:** Investigation, Data curation, Software, Writing – review & editing. **Joaquim Radua:** Investigation, Formal analysis, Writing – review & editing. **Tobias Kober:** Funding acquisition, Resources, Investigation, Writing – review & editing. **Jean-Philippe Thiran:** Resources, Investigation, Project administration, Supervision, Writing – review & editing.

Acknowledgements

Data code availability statement

All the reconstruction techniques evaluated in this study were packaged in a freely distributed python toolbox available at https://github.com/ejcanalesr/multicomponent-T2-toolbox, which aims at fostering reproducible research. Moreover, the code will include a representative multi-echo T_2 MRI data acquired in a healthy volunteer.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118582.

References

Alonso-Ortiz, E., Levesque, I.R., Pike, G.B., 2015. MRI-based myelin water imaging: a technical review. Magn. Reson. Med. 73, 70–81. doi:10.1002/mrm.25198.

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 T_2 relaxometry. Magn. Reson. Med. 54, 449–454. doi:10.1002/mrm.20599.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41 doi:S1361-8415(07)00060-6 [pii]10.1016/j.media.2007.06.004.

Birkl, C., Doucette, J., Fan, M., Hernández-Torres, E., Rauscher, A., 2021. Myelin water imaging depends on white matter fiber orientation in the human brain. Magn. Reson. Med. 85, 2221–2231. doi:10.1002/MRM.28543.

Birkl, C., Langkammer, C., Golob-Schwarzl, N., Leoni, M., Haybaeck, J., Goessler, W., Fazekas, F., Ropele, S., 2016. Effects of formalin fixation and temperature on MR relaxation times in the human brain. NMR Biomed. 29, 458–465. doi:10.1002/nbm.3477.

Bjarnason, T.A., 2011. Proof that gmT2 is the reciprocal of gmR2. Concepts Magn. Reson. Part A 38A, 128–131. doi:10.1002/cmr.a.20216.

Bland, j.M., Altman, D.G., 1995. Multiple significance tests: the Bonferroni method. BMJ 310, 170. doi:10.1136/bmj.310.6973.170.

Blas, J.G., Dolz, M.F., Daniel Garcia, J., Carretero, J., Daducci, A., Aleman, Y., Canales-Rodriguez, E.J., 2016. Porting Matlab applications to high-performance C++ codes: CPU/GPU-accelerated spherical deconvolution of diffusion MRI data. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 630–643. doi:10.1007/978-3-319-49583-5_49.

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, 640–649. doi:10.1111/jon.12537.

Brent, R.P., 1971. An algorithm with guaranteed convergence for finding a zero of a function. Comput. J. 14, 422–425. doi:10.1093/comjnl/14.4.422.

Canales-Rodríguez, E.J.Silvia Alonso-Lana, Norma Verdolini, Salvador Sarró, Isabel Feria, Irene Montoro, Beatriz Garcia-Ruiz, Esther Jimenez, Cristina Varo, Auria Albacete, Isabel Argila-Plaza, Anna Lluch, C. Mar Bonnin, Elisabet Vilella, Eduard Vieta, Edith Pomarol-Clotet, Raymond Salvador, 2021. Age- and gender-related differences in brain tissue microstructure revealed by multi-component T_2 relaxometry. Neurobiology of Aging 106, 68–79. doi:10.1016/j.neurobiolaging.2021.06.002.

Canales-Rodríguez, E.J., Daducci, A., Sotiropoulos, S.N., Caruyer, E., Aja-Fernández, S., Radua, J., Mendizabal, J.M.Y., Iturria-Medina, Y., Melie-García, L., Alemán-Gómez, Y., Thiran, J.-.P., Sarró, S., Pomarol-Clotet, E., Salvador, R., 2015. Spherical deconvolution of multichannel diffusion MRI data with non-Gaussian noise models and spatial regularization. PLoS ONE 10, e0138910. doi:10.1371/journal.pone.0138910.

Canales-Rodríguez, E.J., Legarreta, J.H., Pizzolato, M., Rensonnet, G., Girard, G., Patino, J.R.-., Barakovic, M., Romascano, D., Alemán-Gómez, Y., Radua, J., Pomarol-Clotet, E., Salvador, R., Thiran, J.-P.P., Daducci, A., 2019. Sparse wars: a survey and comparative study of spherical deconvolution algorithms for diffusion MRI. Neuroimage 184, 140–160. doi:10.1016/j.neuroimage.2018.08.071.

Canales-Rodríguez, E.J., Pizzolato, M., Piredda, G.F., Hilbert, T., Kunz, N., Pot, C., Yu, T., Salvador, R., Pomarol-Clotet, E., Kober, T., Thiran, J.-.P., Daducci, A., 2021. Comparison of non-parametric T2 relaxometry methods for myelin water quantification. Med. Image Anal. 101959 doi:10.1016/j.media.2021.101959.

Canales-Rodríguez, E.J., Pomarol-Clotet, E., Radua, J., Sarró, S., Alonso-Lana, S., Del Mar Bonnín, C., Goikolea, J.M., Maristany, T., García-Álvarez, R., Vieta, E., McKenna, P., Salvador, R., 2013. Structural abnormalities in bipolar euthymia: a multicontrast molecular diffusion imaging study. Biol. Psychiatry 76, 239–248. doi:10.1016/j.biopsych.2013.09.027.

Castellanos, J.L., Gómez, S., Guerra, V., 2002. The triangle method for finding the corner of the L-curve. Appl. Numer. Math. doi:10.1016/S0168-9274(01)00179-9.

Cohen-Adad, J., Does, M., DUVAL, T., Dyrby, T.B., Fieremans, E., Foias, A., Sepehrband, F., Stikov, N., Zaimi, A., Nami, H., 2018. White Matter Microscopy Database [WWW Document]. doi:10.17605/OSF.IO/YP4QG.

Cordero-Grande, L., Christiaens, D., Hutter, J., Price, A.N., Hajnal, J.V., 2019. Complex diffusion-weighted image estimation via matrix recovery under general noise models. Neuroimage 200, 391–404. doi:10.1016/j.neuroimage.2019.06.039.

Cumming, G., 2014. The new statistics: why and how. Psychol. Sci. 25, 7–29. doi:10.1177/0956797613504966.

Deoni, S.C.L.L., Matthews, L., Kolind, S.H., 2013. One component? Two components? Three? the effect of including a nonexchanging "free" water component in multicomponent driven equilibrium single pulse observation of T_1 and T_2. Magn. Reson. Med. 70, 147–154. doi:10.1002/mrm.24429.

Does, M.D., 2018. Inferring brain tissue composition and microstructure via MR relaxometry. Neuroimage 182, 136–148. doi:10.1016/J.NEUROIMAGE.2017.12.087.

Does, M.D., Olesen, J.L., Harkins, K.D., Serradas-Duarte, T., Gochberg, D.F., Jespersen, S.N., Shemesh, N., 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. doi:10.1002/mrm.27658, mrm.27658.

Donoho, D.L., Johnstone, J.M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455. doi:10.1093/biomet/81.3.425.

Doucette, J., Kames, C., Rauscher, A., 2020. DECAES – Decomposition and component analysis of exponential signals. Z. Med. Phys. doi:10.1016/j.zemedi.2020.04.001.

Drenthen, G.S., Backes, W.H., Aldenkamp, A.P., Jansen, J.F.A., 2019. Applicability and reproducibility of 2D multi-slice GRASE myelin water fraction with varying acquisition acceleration. Neuroimage 195, 333–339. doi:10.1016/j.neuroimage.2019.04.011.

Dvorak, A.V., Wiggermann, V., Gilbert, G., Vasavour, I.M., MacMillan, E.L., Barlow, L., Wiley, N., Kozlowski, P., MacKay, A.L., Rauscher, A., Kolind, S.H., 2020. Multi-spin echo T_2 relaxation imaging with compressed sensing (METRICS) for rapid myelin water imaging. Magn. Reson. Med. 84, 1264–1279. doi:10.1002/mrm.28199.

Endres, D.M., Schindelin, J.E., 2003. A new metric for probability distributions. IEEE Trans. Inf. Theory. doi:10.1109/TIT.2003.813506.

Gil, R., Khabipova, D., Zwiers, M., Hilbert, T., Kober, T., Marques, J.P., 2016. An in vivo study of the orientation-dependent and independent components of transverse relaxation rates in white matter. NMR Biomed 29, 1780–1790. doi:10.1002/nbm.3616.

Golub, G.H., Heath, M., Wahba, G., 1979. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21, 215. doi:10.2307/1268518.

Graham, S.J., Stanchev, P.L., Bronskill, M.J., 1996. Criteria for analysis of multicomponent tissue T_2 relaxation data. Magn. Reson. Med. 35, 370–378. doi:10.1002/mrm.1910350315.

Gueymard, C.A., 2014. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: towards a better bankability of solar projects. Renew. Sustain. Energy Rev. doi:10.1016/j.rser.2014.07.117.

Guo, J., Ji, Q., Reddick, W.E., 2013. Multi-slice myelin water imaging for practical clinical applications at 3.0 T. Magn. Reson. Med. doi:10.1002/mrm.24527.

Hansen, P.C., 1992. Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev doi:10.1137/1034115.

Ho, J., Tumkaya, T., Aryal, S., Choi, H., Claridge-Chang, A., 2019. Moving beyond P values: data analysis with estimation graphics. Nat. Methods. doi:10.1038/s41592-019-0470-3.

Jones, C.C.K., Whittall, K.P.K., MacKay, A.L., 2003. Robust myelin water quantification: averaging vs. spatial filtering. Magn. Reson. Med. 50. doi:10.1002/mrm.10492.

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. doi:10.1016/j.neuroimage.2018.05.026.

Kumar, D., Nguyen, T.D.T., Gauthier, S.S.A., Raj, A., 2012. Bayesian algorithm using spatial priors for multiexponential T_2 relaxometry from multiecho spin echo MRI. Magn. Reson. Med. 68, 1536–1543. doi:10.1002/mrm.24170.

Kumar, D., Siemonsen, S., Heesen, C., Fiehler, J., Sedlacik, J., 2016. Noise robust spatially regularized myelin water fraction mapping with the intrinsic B1-error correction based on the linearized version of the extended phase graph model. J. Magn. Reson. Imaging 43, 800–817. doi:10.1002/jmri.25078.

Lang, D.J.M., Yip, E., Mackay, A.L., Thornton, A.E., Vila-Rodriguez, F., Macewan, G.W., Kopala, L.C., Smith, G.N., Laule, C., Macrae, C.B., Honer, W.G., 2014. 48 echo T2 myelin imaging of white matter in first-episode schizophrenia: evidence for aberrant myelination. NeuroImage Clin 6, 408–414. doi:10.1016/j.nicl.2014.10.006.

Laule, C., Kozlowski, P., Leung, E., Li, D.K.B., MacKay, A.L., Moore, G.R.W., 2008. Myelin water imaging of multiple sclerosis at 7 T: correlations with histopathology. Neuroimage 40, 1575–1580. doi:10.1016/j.neuroimage.2007.12.008.

Laule, C., Leung, E., Li, D.K.B., Traboulsee, A.L., Paty, D.W., MacKay, A.L., Moore, G.R.W., 2006. Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology. Mult. Scler. 12, 747–753. doi:10.1177/1352458506070928.

Lawson, C.L., Hanson, R.J., 1995. Solving least squares problems, solving least squares problems. doi:10.1137/1.9781611971217.

Lee, Jieun, Lee, D., Choi, J.Y., Shin, D., Shin, H.G., Lee, Jongho, 2020. Artificial neural network for myelin water imaging. Magn. Reson. Med. 83, 1875–1883. doi:10.1002/mrm.28038.

Lin, M., Lucas, H.C., Shmueli, G., 2013. Too big to fail: large samples and the p-value problem. Inf. Syst. Res. 24, 906–917. doi:10.1287/isre.2013.0480.

MacKay, A., Laule, C., Vavasour, I., Bjarnason, T., Kolind, S., Mädler, B., 2006. Insights into brain microstructure from the T_2 distribution. Magn. Reson. Imaging. doi:10.1016/j.mri.2005.12.037.

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, 673–677. doi:10.1002/mrm.1910310614.

MacKay, A.L., Laule, C., 2016. Magnetic resonance of myelin water: an in vivo marker for myelin. Brain Plast 2, 71–91. doi:10.3233/BPL-160033.

MacKay, D.J.C., 1992. Bayesian interpolation. Neural Comput doi:10.1162/neco.1992.4.3.415.

Malik, S.J., Teixeira, R.P.A.G., Hajnal, J.V., 2018. Extended phase graph formalism for systems with magnetization transfer and exchange. Magn. Reson. Med. 80, 767–779. doi:10.1002/mrm.27040.

Meyers, S.M., Kolind, S.H., MacKay, A.L., 2017. Simultaneous measurement of total water content and myelin water fraction in brain at 3 T using a T_2 relaxation based method. Magn. Reson. Imaging 37, 187–194. doi:10.1016/j.mri.2016.12.001.

Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A.V., Mahmood, A., Woods, R., Toga, A.W., Pike, G.B., Neto, P.R., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P., Mazziotta, J., 2008. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. Neuroimage doi:10.1016/j.neuroimage.2007.12.035.

Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., Akhter, K., Hua, K., Woods, R., Toga, A.W., Pike, G.B., Rosa-Neto, P., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P.C.M., Mazziotta, J., Mori, S., 2008. Human brain white matter atlas: identification and assignment of common anatomical structures in superficial white matter. Neuroimage doi:10.1016/j.neuroimage.2008.07.009.

Piredda, G.F., Hilbert, T., Canales-Rodríguez, E.J., Pizzolato, M., von Deuster, C., Meuli, R., Pfeuffer, J., Daducci, A., Thiran, J.-P.P., Kober, T., 2021. Fast and high-resolution myelin water imaging: accelerating multi-echo GRASE with CAIPIRINHA. Magn. Reson. Med. 85, 209–222. doi:10.1002/mrm.28427.

Piredda, G.F., Hilbert, T., Thiran, J., Kober, T., 2020. Probing myelin content of the human brain with MRI: a review. Magn. Reson. Med. doi:10.1002/mrm.28509, mrm.28509.

Prasloski, T., Mädler, B., Xiang, Q.S., MacKay, A., Jones, C., 2012a. Applications of stimulated echo correction to multicomponent T_2 analysis. Magn. Reson. Med. doi:10.1002/mrm.23157.

Prasloski, T., Rauscher, A., MacKay, A.L., Hodgson, M., Vavasour, I.M., Laule, C., Mädler, B., 2012b. Rapid whole cerebrum myelin water imaging using a 3D GRASE sequence. Neuroimage doi:10.1016/j.neuroimage.2012.06.064.

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, e98391. doi:10.1371/journal.pone.0098391.

Song, J.E., Shin, J., Lee, H.J., Lee, H.J., Moon, W.J., Kim, D.H., 2020. Blind source separation for myelin water fraction mapping using multi-echo gradient echo imaging. IEEE Trans. Med. Imaging 39, 2235–2245. doi:10.1109/TMI.2020.2967068.

Stikov, N., Campbell, J.S.W., Stroh, T., Lavelée, A., Frey, S., Novek, J., Nuara, S., Ho, M.K., Bedell, B.J., Dougherty, R.F., Leppert, I.R., Boudreau, M., Narayanan, S., Duval, T., Cohen-Adad, J., Picard, P.A., Gasecka, A., Côté, D., Pike, G.B., 2015. In vivo histology of the myelin g-ratio with magnetic resonance imaging. Neuroimage 118, 397–405. doi:10.1016/j.neuroimage.2015.05.023.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. doi:10.1029/2000JD900719.

Vaserstein, L.N., 1969. Markov processes over denumerable products of spaces, describing large systems of automata. Probl. Peredachi Inf..

Veraart, J., Fieremans, E., Novikov, D.S., 2016a. Diffusion MRI noise mapping using random matrix theory. Magn. Reson. Med. 76, 1582–1593. doi:10.1002/mrm.26059.

Veraart, J., Novikov, D.S., Christiaens, D., Ades-aron, B., Sijbers, J., Fieremans, E., 2016b. Denoising of diffusion MRI using random matrix theory. Neuroimage 142, 394–406. doi:10.1016/j.neuroimage.2016.08.016.

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., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, 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., Vijaykumar, A., Bardelli, A.Pietro, Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods. 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 with MRI and histology. In: Proceedings of the 25th Annual Meeting of ISMRM, p. 3760.

Weigel, M., 2015. Extended phase graphs: dephasing, RF pulses, and echoes - pure and simple. J. Magn. Reson. Imaging. doi:10.1002/jmri.24619.

Weigel, M., Schwenk, S., Kiselev, V.G., Scheffler, K., Hennig, J., 2010. Extended phase graphs with anisotropic diffusion. J. Magn. Reson. 205, 276–285. doi:10.1016/j.jmr.2010.05.011.

Whittall, K.P., MacKay, A.L., 1989. Quantitative interpretation of NMR relaxation data. J. Magn. Reson. doi:10.1016/0022-2364(89)90011-5.

Whittall, K.P., MacKay, A.L., Graeb, D.A., Nugent, R.A., Li, D.K.B.B., Paty, D.W., 1997. In vivo measurement of T_2 distributions and water contents in normal human brain. Magn. Reson. Med. 37, 34–43. doi:10.1002/mrm.1910370107.

Yu, T., Canales-Rodríguez, E.J., Pizzolato, M., Piredda, G.F., Hilbert, T., Fischi-Gomez, E., Weigel, M., Barakovic, M., Bach Cuadra, M., Granziera, C., Kober, T., Thiran, J.P., 2021. Model-informed machine learning for multi-component T_2 relaxometry. Med. Image Anal. 69, 101940. doi:10.1016/j.media.2020.101940.

Yu, T., Pizzolato, M., Canales-Rodríguez, E.J., Thiran, J.-P.J.P.J.-P, 2019. Robust T2 relaxometry with hamiltonian MCMC for myelin water fraction estimation. In: Proceedings - International Symposium on Biomedical Imaging doi:10.1109/ISBI.2019.8759446.

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. doi:10.1038/s41598-018-22181-4.

Zimmermann, M., Oros-Peusquens, A.M., Iordanishvili, E., Shin, S., Yun, S.D., Abbas, Z., Shah, N.J., 2019. Multi-exponential relaxometry using ℓ1-regularized iterative NNLS (MERLIN) with application to myelin water fraction imaging. IEEE Trans. Med. Imaging. doi:10.1109/TMI.2019.2910386.