Check for updates

Variability and reproducibility of multi-echo T_2 relaxometry: Insights from multi-site, multi-session and multi-subject MRI acquisitions

Elda Fischi-Gomez[1,2†], Gabriel Girard[1,3,4†], Philipp J. Koch[5,6,7,8],
Thomas Yu[1,9], Marco Pizzolato[1,10], Julia Brügger[5,6],
Gian Franco Piredda[1,4,9], Tom Hilbert[1,4,9],
Andéol G. Cadic-Melchior[5,6], Elena Beanato[5,6],
Chang-Hyun Park[5,6], Takuya Morishita[5,6],
Maximilian J. Wessel[5,6,11], Simona Schiavi[12,13],
Alessandro Daducci[13], Tobias Kober[1,4,9],
Erick J. Canales-Rodríguez[1*†], Friedhelm C. Hummel[5,6,14†] and
Jean-Philippe Thiran[1,3,4†]

[1]Signal Processing Laboratory 5 (LTS5), École Polytechnique Fédérale de Lausanne (EPFL),
Lausanne, Switzerland, [2]Translational Machine Learning Lab, Department of Radiology, Centre
Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland, [3]CIBM Center for
Biomedical Imaging, Lausanne, Switzerland, [4]Department of Radiology, Centre Hospitalier
Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland, [5]Defitech Chair for Clinical
Neuroengineering, Neuro-X Institute (NIX) and Brain Mind Institute (BMI), École Polytechnique
Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [6]Defitech Chair of Clinical Neuroengineering,
Neuro-X Institute (NIX) and Brain Mind Institute (BMI), École Polytechnique Fédérale de Lausanne
(EPFL Valais), Clinique Romande de Réadaptation, Sion, Switzerland, [7]Department of Neurology,
University of Lübeck, Lübeck, Germany, [8]Center of Brain, Behavior and Metabolism (CBBM),
University of Lübeck, Lübeck, Germany, [9]Advanced Clinical Imaging Technology, Siemens
Healthineers International AG, Lausanne, Switzerland, [10]Department of Applied Mathematics and
Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark, [11]Department of
Neurology, University Hospital and Julius-Maximilians-University, Wuerzburg, Germany,
[12]Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health
(DINOGMI), University of Genoa, Genoa, Italy, [13]Diffusion Imaging and Connectivity Estimation
(DICE) Lab, Department of Computer Science, University of Verona, Verona, Italy, [14]Clinical
Neuroscience, University Hospital of Geneva (HUG), Geneva, Switzerland

Quantitative magnetic resonance imaging (qMRI) can increase the specificity
and sensitivity of conventional weighted MRI to underlying pathology by
comparing meaningful physical or chemical parameters, measured in physical
units, with normative values acquired in a healthy population. This study
focuses on multi-echo T_2 relaxometry, a qMRI technique that probes
the complex tissue microstructure by differentiating compartment-specific
T_2 relaxation times. However, estimation methods are still limited by
their sensitivity to the underlying noise. Moreover, estimating the model's
parameters is challenging because the resulting inverse problem is ill-posed,
requiring advanced numerical regularization techniques. As a result, the
estimates from distinct regularization strategies are different. In this work, we
aimed to investigate the variability and reproducibility of different techniques

for estimating the transverse relaxation time of the intra- and extra-cellular space (T_2^{IE}) in gray (GM) and white matter (WM) tissue in a clinical setting, using a multi-site, multi-session, and multi-run T_2 relaxometry dataset. To this end, we evaluated three different techniques for estimating the T_2 spectra (two regularized non-negative least squares methods and a machine learning approach). Two independent analyses were performed to study the effect of using raw and denoised data. For both the GM and WM regions, and the raw and denoised data, our results suggest that the principal source of variance is the inter-subject variability, showing a higher coefficient of variation (CoV) than those estimated for the inter-site, inter-session, and inter-run, respectively. For all reconstruction methods studied, the CoV ranged between 0.32 and 1.64%. Interestingly, the inter-session variability was close to the inter-scanner variability with no statistical differences, suggesting that T_2^{IE} is a robust parameter that could be employed in multi-site neuroimaging studies. Furthermore, the three tested methods showed consistent results and similar intra-class correlation (ICC), with values superior to 0.7 for most regions. Results from raw data were slightly more reproducible than those from denoised data. The regularized non-negative least squares method based on the L-curve technique produced the best results, with ICC values ranging from 0.72 to 0.92.

1. Introduction

Quantitative magnetic resonance imaging (qMRI) has the potential to increase the specificity and sensitivity of conventional weighted MRI to underlying pathology. This increased sensitivity and specificity has stimulated the use of qMRI methods as potential biomarkers for microstructural integrity of the brain. Pathological processes such as demyelination, edema, iron accumulation, and tissue loss lead to variable and complex changes in tissue microstructure, inducing in turn changes in relaxation times. Therefore, these microstructural features can be inferred by *in-vivo* qMRI at millimeter resolution thanks to biophysical models. The spin-spin transverse relaxation rate T_2 is one of the fundamental tissue-specific MRI contrast sources. In complex tissue, the microstructure can be seen as a combination of different pools of water with different chemical environments (called compartments) which each have their own characteristic T_2. Hence, for complex tissues, multi-component T_2 relaxometry allows probing the tissue microstructure by differentiating compartment-specific T_2 relaxation times (1). Recent advances in multi-component T_2 relaxometry acquisition (2, 3) and reconstruction methods (4) have boosted the use of this technique for the assessment of white matter (WM) integrity in general, but most notably for the determination of the myelin water content (5–11), and the T_2 of the intra- and extra-cellular spaces (T_2^{IE}) (12).

As T_2^{IE} is usually estimated from the T_2 distribution by taking the (geometric) mean in the interval from 40 to 200 ms (at 3T), it is not affected by partial volume contamination from free water (T_2 >200 ms) or myelin water compartments (T_2 <40 ms). This is the main advantage of using T_2^{IE} over the mean intra-voxel T_2 estimated from other qMRI techniques. Interestingly, a recent study revealed a strong correlation between T_2^{IE} and age extending through the whole gray matter (GM), suggesting that this metric is sensitive to microstructural and macro-molecular content changes (13). However, multi-compartment T_2 estimation methods are still sensitive to the underlying noise and the employed regularization technique for solving the resulting inverse problem (14, 15). This limitation is especially relevant for clinically achievable signal-to-noise ratios (SNR), which may affect the reproducibility and stability of the derived scalar maps. Several studies have already focused on the reproducibility and stability of myelin water fraction (MWF) estimates (16–20). However, the variability linked to the intra- and extra-cellular T_2 component has been less studied, particularly in GM regions.

Other than the clinically achievable SNR, variability in qMRI data can be due to different factors such as hardware differences (scanner manufacturer, field strength, etc.), different MR reconstruction methods and acquisition parameters, among others. Even when the same hardware and sequence parameters are employed, high inter- and intra-scanner variability due to local and/or temporal scanner characteristics can occur.

Characterizing this variability is challenging, mainly due to the lack of available data specifically designed to perform this task. In Cai et al. (21), authors tackle this issue with diffusion-weighted MRI data (DWI) using an *ad-hoc* data set (the MASiVar data set). The multi-site, multi-scanner, and multi-subject dataset proposed in that work was used to simultaneously characterize four commonly used diffusion processing techniques [diffusion tensor signal representation, multi-compartment neurite orientation dispersion and density imaging NODDI, microstructure model (22), white-matter bundle segmentation (23), and graph-based connectomics representations (24)].

In this study, we investigated the T_2^{IE} variability and reproducibility in a clinical setting, in both GM and WM, using a multi-site, multi-session and multi-run dataset consisting of 20 healthy adults scanned in two different scanning sites, two separate sessions, and two runs (repetitions) per session and per site, resulting in a total of 160 scans (eight per subject). This unique dataset allows for quantifying the variability of the estimated multi-compartment T_2 spectra, with respect to the subject, the acquisition site, session and run. Moreover, we aimed at evaluating the impact of using different techniques for estimating the T_2 spectra, as well as the effect of the denoising of the data. In particular, three different non-parametric techniques were tested, including two types of regularized non-negative least squares (NNLS) methods based on the Chi-square (5) and L-curve (14) regularization criteria and a recently proposed Model-Informed Machine Learning approach using neural networks trained with synthetic data (25) on the original (raw) data and the same data after denoising.
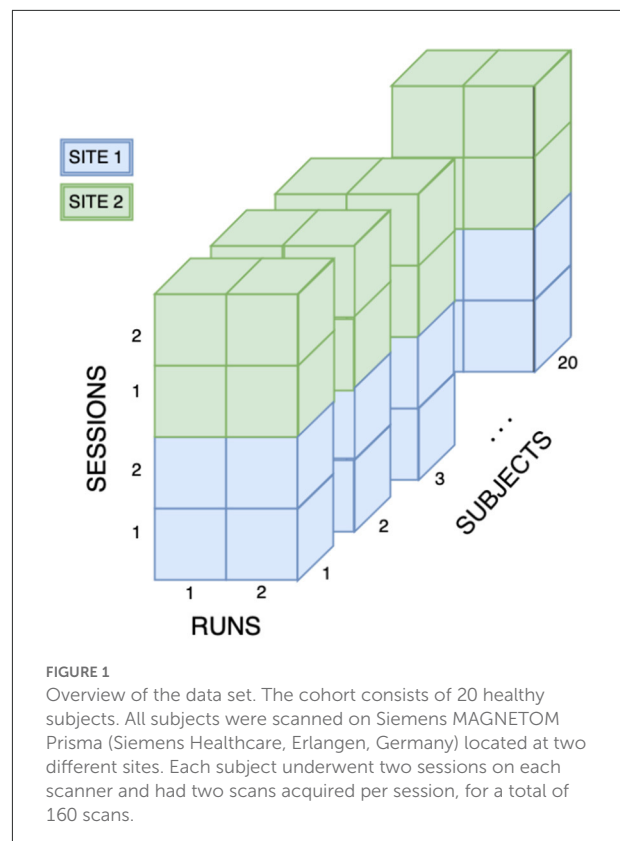
2. Methods

2.1. Human brain MRI data

2.1.1. Population

Twenty healthy subjects (nine female, mean age = 27, standard deviation = ± 3 years, age range = [24–33 years old]) were enrolled in the study. All subjects were right-handed and had no history of psychiatric diseases or any contraindications for performing MRI. Written informed consent was obtained from each participant following the Declaration of Helsinki. The ethics committee of the Canton of Vaud approved the study (Switzerland, project number: 2018-01355).

2.1.2. MR acquisition

The MRI data were collected using a high-resolution 3D multi-echo gradient and spin-echo (GRASE) prototype sequence accelerated through CAIPIRINHA (3) with the following parameters: matrix-size = 144×126×134; voxel-size = 1.6×1.6×1.6 mm³; minimum echo time (TE) = 10.68 ms; Number-of-echoes = 32; Δ(TE) = 10.68 ms; repetition time

FIGURE 1
Overview of the data set. The cohort consists of 20 healthy subjects. All subjects were scanned on Siemens MAGNETOM Prisma (Siemens Healthcare, Erlangen, Germany) located at two different sites. Each subject underwent two sessions on each scanner and had two scans acquired per session, for a total of 160 scans.

(TR) = 1s; prescribed flip angle (FA) = 180°; number-of-slices = 84; acceleration factor = 3 x 2$^{(1)}$; number of averages = 1; acquisition time = 10:30 min. T1-weighted images were also acquired for anatomical localization and definition of regions of interest (ROIs) using a 3D magnetization-prepared rapid gradient-echo (MP-RAGE) sequence with the following parameters: TR = 2,300 ms; Inversion Time (TI) = 7.1 ms; TE = 2.96 ms; FA = 9°; number-of-slices = 192; voxel-size = 1×1×1 mm³, field of view = 256×256 mm².

2.1.3. MRI scanning design

Each subject was scanned in two MRI scanners (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) located at the Geneva University Hospital and Sion Hospital, Switzerland (sites) at two different points in time (sessions). At each session, each subject was scanned twice in two separate runs. Between the runs, subjects exited the scanner and were then repositioned, followed by a new shimming. Eight scans were obtained per subject, for a total of N = 160 scans. Figure 1 shows a schematic representation of the MRI scanning design. The mean elapsed time between intra-site and inter-site repetitions was 16 days (± 10 days) and 29 days (± 17 days), respectively.

2.2. MRI spatial processing

The T1-weighted image was normalized to the MNI standard space by registering them to the MNI152 template image using FSL FLIRT (26). Tissue partial volume estimates were obtained from the T1-weighted image using the FSL BET (27) and FSL FAST (28) methods. The definition of the different regions of interest, in both WM and GM, was performed using the Destrieux atlas available in FreeSurfer (29). Seventy-four regions per hemisphere were obtained for each tissue type. They were further grouped into five main cerebral lobes: parietal, occipital, temporal, the prefrontal part of the frontal lobe and its medial part.

2.3. T_2 estimation

Two different scenarios were tested. In the first approach, the T_2 estimation was performed over the raw data, without any prior denoising (raw data). For the second approach, the multi-echo T_2 MRI data were filtered using a 3D total variation algorithm before fitting, using the *denoise-tv-chambolle* function in the *scikit-image* python toolbox (30) (denoised data). Spatial filtering is effective for decreasing the variability of the estimated maps [see (31–33)]. The noise standard deviation σ for each 3D volume was estimated by employing a robust wavelet-based estimator (34), and each volume was then denoised with a weight of 2σ as described in Canales-Rodríguez et al. (15). In both cases, the T_2 spectra were estimated by using three different methods; two regularized non-negative least squares (NNLS) algorithms (X^2-I, L-curve$-I$) and the Model-Informed Machine Learning (MIML) approach. More details are provided in the next two subsections.

2.3.1. Regularized NNLS

The T_2 spectra were computed in two steps. The refocusing flip angle (FA) value for each voxel was determined first, and then the intra-voxel T_2 spectrum was computed by using the dictionary generated for the estimated FA. Estimating the optimal FA involves various steps. First, different matrices were generated using the Extended Phase Graph (EPG) model (35), 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 to 2,000 ms (36) with $N = 60$ T_2 logarithmically spaced points was employed. Subsequently, we created a smoothed copy of the acquired multi-echo T2 data by using a Gaussian kernel (i.e., FWMH of 4.8 mm) as suggested in Drenthen et al. (20). Afterwards, the non-regularized NNLS algorithm was used to fit the smoothed data for each matrix, and the resulting mean square errors were interpolated using cubic B-splines (35). Next, the optimal FA was selected at the global minima of the resulting interpolated curve. Finally, a new matrix was generated using the estimated FA, which was used then for

estimating the T_2 spectrum (35). Note that the T_2 spectrum and derived maps were computed from the TV-denoised data and not from the Gaussian-filtered copy, which was only employed to get a smooth FA map. To accelerate the estimation, instead of generating a new matrix for each estimated FA, we loaded the optimal kernel 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 kernel was selected by identifying the FA grid point closest to the interpolated FA.

As the estimated T_2 distribution may depend on the chosen technique for determining the optimal regularization parameter, two approaches were implemented, a regularized NNLS method based on the Chi-square residual fitting criterion (5), and another based on the L-curve technique, as implemented in the multi-component T_2 reconstruction toolbox (14): https://github.com/ejcanalesr/multicomponent-T2-toolbox. Both methods, named $X^2 - I$ and L-curve$-I$, used an identity regularization matrix to promote smooth T_2 distributions.

2.3.2. Model-informed machine learning

The T_2 distributions were obtained by using the MIML approach described in Yu et al. (25). It is based on a multilayer perceptron (MLP) which is trained to learn a map directly from the noisy multi-echo T_2 signals to the corresponding T_2 distributions. The MLP network was composed of 6 hidden layers with 256 neurons per layer and an output layer with 60 units, corresponding to the same resolution we used to solve the regularized NNLS problem described in the previous section. The hidden layers used a *ReLu* function as the activation function, while the output layer used a SoftMax activation function. The input to the network is a vector with 32 elements corresponding to the 32 echos of the acquisition sequence. The loss function is composed of two penalty terms, a squared L2 norm term and the Wasserstein distance on probability distributions.

This technique was implemented using TensorFlow 2.0 (37) on Python 3.6 (38) with an *Nvidia GTX 2070* GPU. The MLP network was trained using a total of 1,120,000 signal/distribution pairs. The optimization was carried out by using the Adam optimizer (39) with a learning rate of $5e^{-4}$, a batch size of 2,000, and 30 epochs. For more details, see Yu et al. (25). The trained MLP network was used to predict the T_2 spectra from the measured multi-echo T_2 data. The trained model and code are available at the following website: https://github.com/thomas-yu-epfl/Model_Informed_Machine_Learning.

2.4. Statistical analysis and evaluation metrics

In order to investigate all possible sources of variability in the (semi)-quantitative T_2^{IE} maps in a clinical setting, four

different effects were studied for both WM and GM. They included inter-run, inter-session, inter-scanner (i.e. inter-site) and inter-subject effects. The N = 160 scans were grouped in different subsets: (i) inter-run variability was computed using scans acquired within the same session of the same subject on the same scanner; (ii) for the inter-session variability, we used the scans acquired between different sessions of the same subject on the same scanner; (iii) inter-scanner variability included scans acquired between different sessions of the same subject on different scanners and (iv) finally, inter-subject was assessed averaging the scans of different subjects acquired in different sessions on the same scanner.

For each of the four effects studied, the variability was evaluated by computing variability within each of the aforementioned groups and then visualizing the distribution across groups on the inter-run, inter-session, inter-scanner and inter-subject levels. The variability was computed by means of the coefficient of variation (CoV), defined for each group as the standard deviation (SD) of the scalar metrics in each group, divided by the mean of the group, times 100. Intuitively, CoV is computed as the proportion of the average scalar measurement attributable to variability. As such, higher CoV indicates higher variability. The inter-run, inter-session, inter-scanner and inter-subject variability was independently computed on the T_2^{IE} maps obtained for each of the reconstruction methods used ($X^2 - I$, L-curve$-I$, and MIML). For all reconstruction methods, the variability due to session, site and subject effects were compared pairwise using a Wilcoxon rank-sum test with Bonferroni correction.

In addition, the reproducibility of each reconstruction method (defined as the agreement of multiple assessments of the same subjects) was computed *via* the Intraclass Correlation Coefficient (ICC). While ICC is generally used to determine if subjects can be rated reliably by different raters, it can also be used for test-retest (repeated measures of the same subject).

The definition of the intraclass correlation is mainly based on analysis of variance or random effects models. While several ICC estimators have been proposed, most of the estimators can be defined in terms of the random effects model

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}, \qquad (1)$$

where Y_{ij} is the ith observation in the jth group. The terms μ, α_j and ϵ_{ij} are the unobserved overall mean, the unobserved random effect shared by all values in group j and the unobserved noise, respectively. Using a random effect framework, the population ICC can be defined as

$$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}. \qquad (2)$$

The selection of the ICC statistics needs to be chosen carefully, as different statistics can produce different results for the same data. In this work, and following Koo and Li (40), we focus

on the ICC for scan-rescan reliability study using a "two-way" random effects (i.e., considering both subjects and "raters" as random effects) a single repeated measurement model with "consistency" as the agreement term. ICC values range between 0 and 1, with a value of ICC close to 1 indicating high similarity between rater's scores (or repeated measurements) and ICC close to 0 showing low similarity of the values. According to Koo and Li (40), an ICC of less than 0.50 indicates poor reliability, while an ICC ranging from [0.5, 0.75] indicates a moderate one. Good reliability is indicated by an ICC value of 0.75 and higher, with excellent reliability defined as an ICC greater than 0.9.
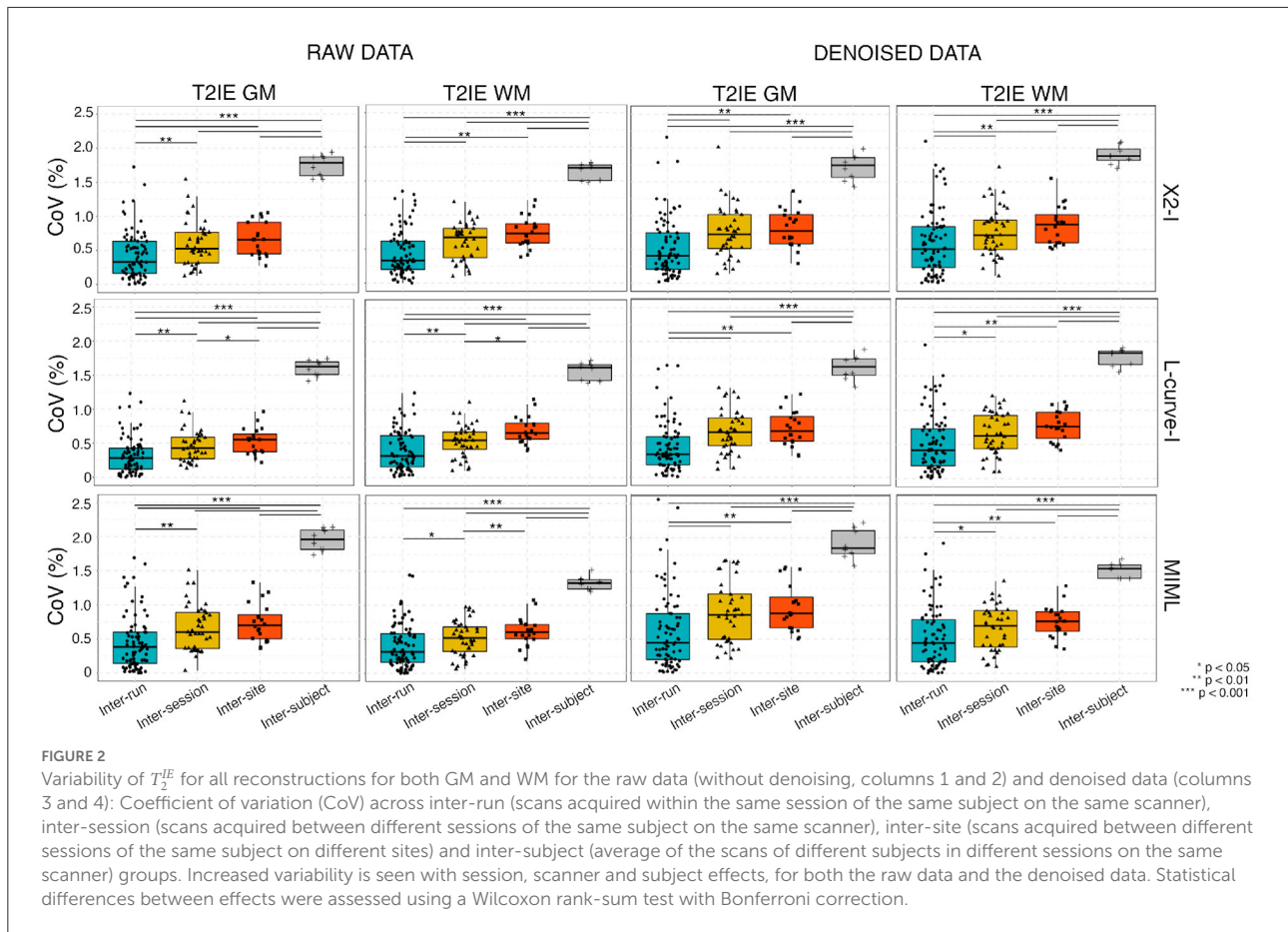
For each reconstruction method, the ICC was computed by arranging the MRI scans in a matrix where rows represented subjects, and the columns represented each one of the repeated measurements. In order to assess the reproducibility of the MR measurements in specific brain regions and determine whether one region is more reproducible than others, several matrices were computed, where each cell of the matrix stored the mean T_2^{IE} value within the ROI for each scan and reconstruction method. The ROIs under study comprised the whole WM and GM tissues as well as the prefrontal, frontal, parietal, temporal and occipital in WM and GM, independently. The ICC was computed in R Studio (Version 1.2.5042) using the *icc* function from the *irr* package https://cran.r-project.org/web/packages/irr/irr.pdf. Only the *single raters absolute* ICC was used. The confidence interval was set to 0.095.

3. Results

3.1. Inter-run, inter-session, inter-scanner and inter-subject variability

The inter-run, inter-session, inter-scanner and inter-subject variability for each reconstruction method in WM and GM tissue independently for both the raw data and the denoised data is shown in Figure 2 (from top to bottom, $X^2 - I$, L-curve$-I$, and MIML methods).

For all reconstruction methods under study in both the raw and the denoised cases, the mean CoV ranged between 0.32 and 1.64%. Overall, results from the raw and denoised data were similar. The highest T_2^{IE} variability was found for the inter-subject effect, which was higher than the CoV induced by different sessions and sites for all the reconstruction methods in both GM and WM. This increased variability speaks in favor of the specificity of this metric for characterizing brain tissues. Moreover, our results demonstrate that, as expected, the variability consistently increases when changing session and site, with inter-run being the least variable followed by inter-session, and then inter-scanner and inter-subject for all the estimation techniques. Interestingly, the inter-session variability was close to the inter-scanner variability in both WM and GM. The pairwise comparison of the 4 effects show statistical significant

FIGURE 2
Variability of T_2^{IE} for all reconstructions for both GM and WM for the raw data (without denoising, columns 1 and 2) and denoised data (columns 3 and 4): Coefficient of variation (CoV) across inter-run (scans acquired within the same session of the same subject on the same scanner), inter-session (scans acquired between different sessions of the same subject on the same scanner), inter-site (scans acquired between different sessions of the same subject on different sites) and inter-subject (average of the scans of different subjects in different sessions on the same scanner) groups. Increased variability is seen with session, scanner and subject effects, for both the raw data and the denoised data. Statistical differences between effects were assessed using a Wilcoxon rank-sum test with Bonferroni correction.

pairwise differences for all effects besides between inter-session and inter-site effects (see Figure 2).

3.2. Variability in brain lobes

The voxelwise regional variability for each reconstruction method is represented in Figure 3, which shows heat maps of T_2^{IE} mean (first panel) and T_2^{IE} standard deviation (second panel) across the whole dataset. While the raw and denoised datasets appeared qualitatively similar, more anatomical details can be seen in the mean maps estimated from raw data. Overall, the mean T_2^{IE} was higher in GM, and some WM regions, including the corticospinal tract and the optic radiations.

Table 1 show the mean and standard deviation values of the $T2^{IE}$ in the different brain regions for both WM and GM for all reconstruction methods under analysis. All three reconstruction methods showed consistent results, although the MIML method displayed a higher mean T_2^{IE} values. The values obtained with the NNLS methods were almost identical, with only slight differences mostly related to the smoothness of the solution. On the other hand, the variability (measured by

the standard deviation) is higher in frontal regions for all the reconstruction techniques.

3.3. Reproducibility of each reconstruction method

We investigated the reproducibility of each reconstruction method by means of the ICC. The result of the voxel-wise ICC in the whole brain GM and WM, as well as different brain regions, for both the raw and denoised data is shown in Figure 4.

For most GM and WM regions, the ICC of T_2^{IE} was higher when the T_2 spectra were estimated using regularized NNLS methods (a few exceptions are the prefrontal in GM and WM, and the parietal in GM). When comparing the two NNLS methods under study, the one based on the L-curve$-I$ criterion performed better. It produced the highest ICC scores for almost all the GM and WM regions, with values ranging from 0.72 to 0.92. Concerning the reproducibility of the reconstruction methods in specific brain regions, all three methods behaved similarly. Overall, the ICC in the whole GM and WM was moderate to good for the NNLS methods (ICC > 0.72 for L-curve$-I$, and > 0.70 for $X^2 - I$), while the MIML method
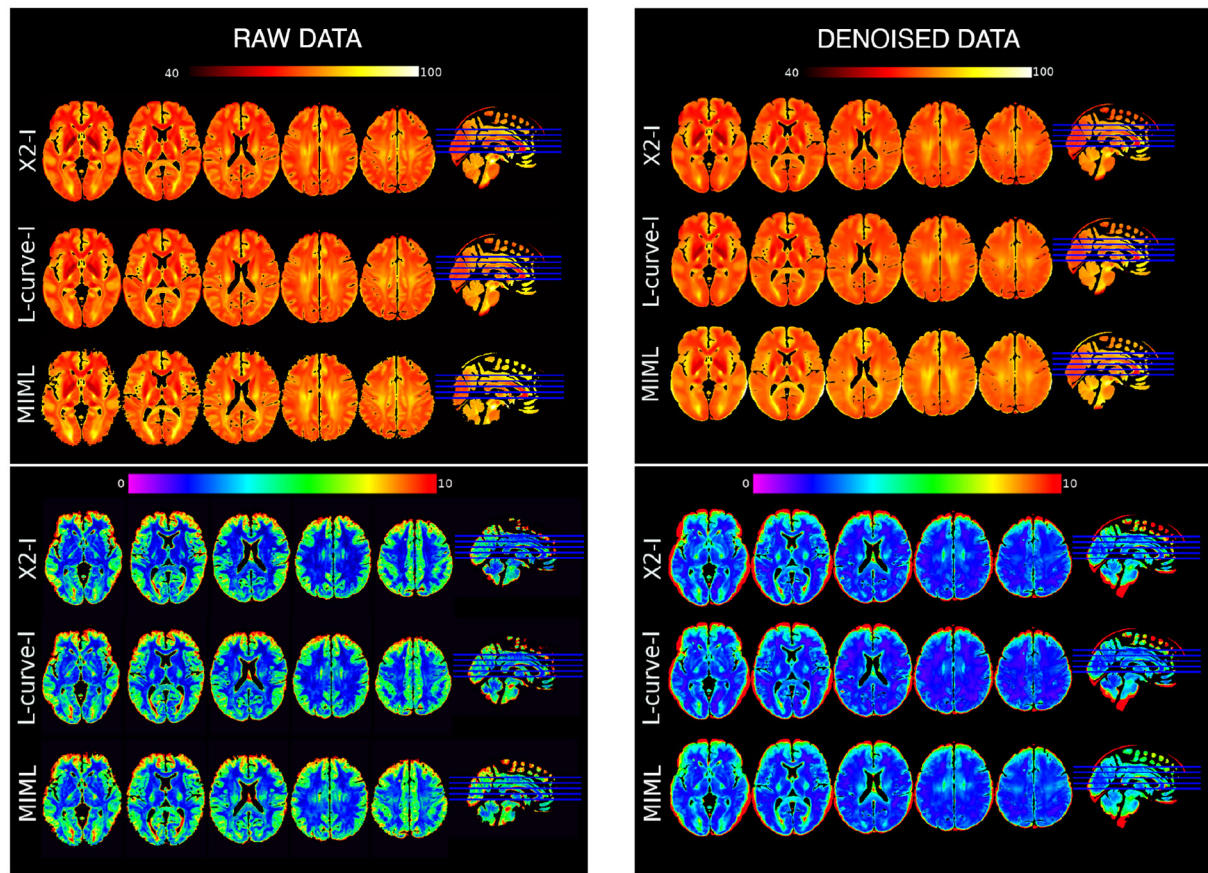
FIGURE 3
Whole-brain voxelwise $T2^{IE}$ mean (upper panels, hot colormap) and standard deviation (lower panels, cold colormap) maps for the three reconstruction methods: $X^2 - I$, L-curve$-I$, and MIML. Left column, raw data. Right column, denoised data.

TABLE 1 T_2^{IE} mean (standard deviation) in WM and GM regions for all reconstruction methods under analysis.

| ROI	RAW DATA			DENOISED DATA		
	L-CURVE-I	**X2-I**	**MIML**	**L-CURVE-I**	**X2-I**	**MIML**
GM (whole brain)	68.09087 (1.074768)	67.31724 (1.154775)	69.34736 (1.33019)	69.19943 (1.105125)	68.55026 (1.1633760)	71.31032 (1.340239)
Prefrontal GM	72.54601 (7.8579)	72.23892 (8.124845)	76.43832 (6.546387)	72.88112 (7.50065)	72.43613 (7.8527090)	76.21879 (6.583347)
Frontal GM	71.46121 (4.971109)	71.47943 (4.942458)	76.81224 (4.137405)	73.01721 (4.285368)	72.49305 (4.433513)	76.45853 (3.986504)
Parietal GM	72.50995 (3.400083)	72.38689 (3.376828)	76.1375 (2.455073)	73.00612 (2.770335)	72.77265 (2.849852)	76.00545 (2.688274)
Temporal GM	73.57243 (3.604324)	73.39499 (3.891249)	79.01297 (2.923089)	73.91503 (3.68813)	73.30519 (4.158284)	79.24136 (3.283316)
Occipital GM	70.42572 (1.982738)	69.51394 (2.066751)	72.34303 (2.929528)	70.01067 (2.027722)	69.21856 (2.129861)	72.06157 (3.041027)
WM (whole brain)	71.68393 (1.123686)	71.5595 (1.165196)	76.91091 (1.005409)	72.69097 (1.272473)	72.28397 (1.350981)	76.87399 (1.147712)
Prefrontal WM	65.8683 (2.772873)	64.89867 (2.933406)	67.18295 (3.023913)	67.31196 (2.875424)	66.51591 (3.110145)	69.46465 (3.057553)
Frontal WM	68.67319 (2.961073)	67.7981 (3.022375)	69.8067 (3.290552)	70.14292 (2.638769)	69.12846 (2.707075)	71.52526 (3.074072)
Parietal WM	68.75922 (2.078946)	67.91948 (2.005582)	69.85067 (2.294526)	69.71293 (1.960121)	69.11882 (1.95721)	72.01084 (2.35551)
Temporal WM	68.1875 (2.561752)	67.58988 (2.742603)	70.06063 (2.503632)	69.64208 (2.399448)	69.23569 (2.538462)	72.58441 (2.308901)
Occipital WM	69.34293 (2.868029)	68.95491 (3.168218)	71.22032 (3.485063)	69.38705 (2.224087)	69.26719 (2.5307)	72.14998 (2.709813)
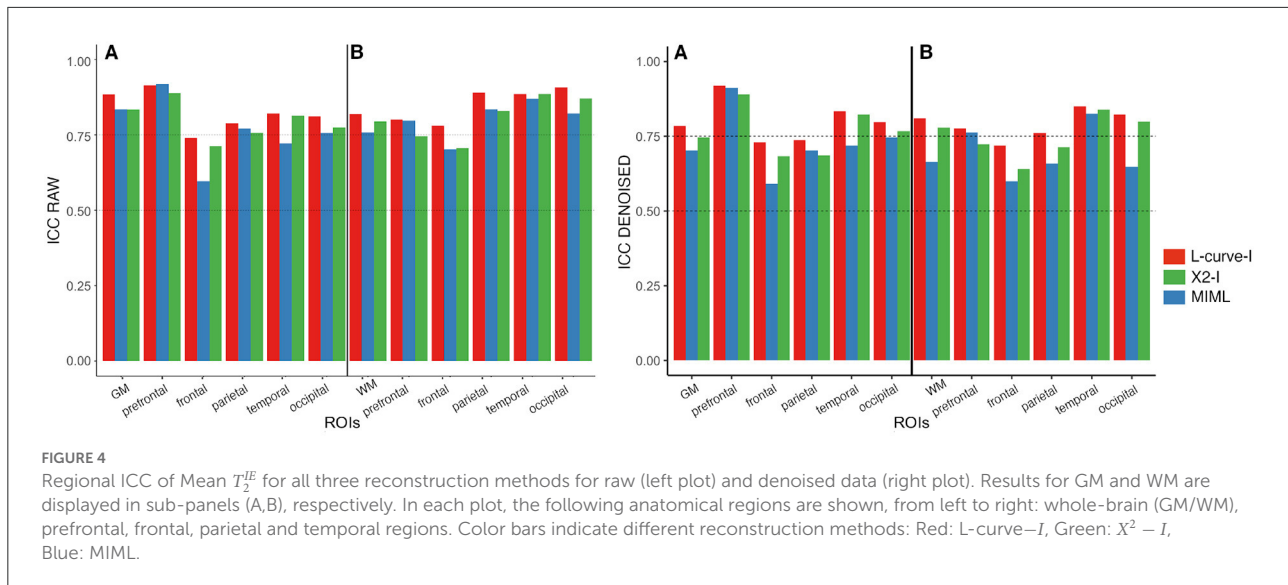
FIGURE 4
Regional ICC of Mean T_2^{IE} for all three reconstruction methods for raw (left plot) and denoised data (right plot). Results for GM and WM are displayed in sub-panels (A,B), respectively. In each plot, the following anatomical regions are shown, from left to right: whole-brain (GM/WM), prefrontal, frontal, parietal and temporal regions. Color bars indicate different reconstruction methods: Red: L-curve$-I$, Green: $X^2 - I$, Blue: MIML.

achieved an acceptable ICC (ICC > 0.6). For all methods, the highest scores were obtained in the prefrontal GM region and the temporal and occipital GM/WM regions. The lowest ICC scores were reported in the frontal GM and WM regions. Overall, ICC scores from raw data were slightly higher than those from the denoised data.

4. Discussion

In this study, we have evaluated the reproducibility of the intra- and extra-cellular T_2 relaxation time estimated by three non-parametric reconstruction techniques using a unique multi-echo T_2 MRI dataset acquired in different subjects, scanning sites, and in separate sessions and runs. Moreover, two scenarios were tested with respect to the T2 estimation: first, estimating from the raw data (without any processing) and second, estimating from the data after denoising. An important contribution of our study is the analysis of reproducibility and stability in both GM and WM tissue types.

As the T_2^{IE} is less affected by partial volume effects (PVE) than the voxelwise mean T_2 estimated by other standard quantitative MRI techniques, we hypothesized that it could be a helpful imaging biomarker and an alternative or complement to the myelin water fraction. PVE appear especially close to the brain cortex, where voxels may contain a mixture of tissue and free water (e.g., GM/CSF) due to the highly convoluted geometry of the cortical mantle. As a result, the voxelwise mean T_2 will be the average of T_2^{IE} and the T_2 of free water, weighted by their relative volumes. Thus, the voxelwise mean T_2 may depend on the voxel location. This issue does not affect the T_2^{IE} estimated in this study since it is computed from the T_2 spectrum by discarding the portion corresponding to the myelin water and free water. Consequently, the intra- and

extra-cellular T_2 relaxation time may provide more specific information about potential abnormalities within the intra- and extra-cellular tissue compartments.

Interestingly, our results point to relatively higher stability of the T_2^{IE} in the GM compared to WM. Hence, our findings suggest that the T_2^{IE} could be a relevant metric for studying pathological conditions in the GM. A plausible explanation for this finding might be related to the complexity of the distribution of T_2 times in these brain regions. While the T_2 spectrum in GM commonly has a single lobe (or two very well separated lobes, when a portion of the voxel contains free water), in WM, there is a dominant lobe encoding the information from the intra- and extra-axonal water and a non-dominant lobe representing the myelin water. Estimating the myelin water lobe is difficult because (1) the myelin water signal decays very fast, and therefore, few data points contain relevant information about this water pool, and (2) the relative volume fraction of the myelin water is smaller than that of the intra- and extra-axonal water. Hence, its quantification is more affected by the underlying noise. Any error in estimating the right location and amplitude of the myelin water could increase the uncertainty in assessing the dominant lobe, i.e., T_2^{IE} in the WM. In contrast, the T_2^{IE} in GM is less affected by this issue. Moreover, the exchange between the intra- and extra-cellular water in GM is more significant, which tends to homogenize the spectrum, further reducing its complexity/variance.

4.1. Estimated variability in comparison to previous studies

Two primary metrics were employed to characterize the performance of the evaluated methods, the coefficient of

variation (CoV) and the intra-class correlation (ICC). The range of values for the CoV resulting from the evaluated methods (i.e., in the range 0.32 and 1.64%) are smaller than those obtained in similar neuroimaging studies using other scalar metrics obtained from diffusion MRI data (in WM), and are concordant with previous studies using multi-echo T_2 relaxometry data. For example, the median CoV values reported in the MASiVar study [21] for the fractional anisotropy and mean diffusivity varied from 3.34 to 11.95%, and from 1.37 to 5.12%, respectively [21]. On the other hand, an average within-subject coefficient of variation (CoV) of 5.9% for the MWF metric was reported in Drenthen et al. [20]. In contrast, Lee et al. [19] reported a mean inter-site MWF CoV across participants of 2.77% in the global WM. The only two previous studies that evaluated the T_2^{IE} metric found a mean longitudinal CoV of 4% using 1.5T data and ROIs in WM and subcortical structures [11] and intra-site and inter-site CoVs of 0.51 and 0.31% in WM [17], respectively.

It is important to note that although other previous studies assessed the reproducibility of the parameters estimated from multi-echo T_2 data, they mainly focused on characterizing the myelin water fraction [11, 16–20]. The only two studies that analyzed the T_2^{IE} time [11, 20] exclusively considered WM regions and subcortical structures. Therefore, to the best of our knowledge, this is the first study evaluating the reproducibility of T_2^{IE} in the cortical GM.

4.2. Local microstructure sensitivity

The whole-brain voxelwise mean and standard deviation maps displayed in Figure 3 show a rich anatomical contrast provided by the T_2^{IE}. It suggests that this parameter is sensitive to the local microstructure and chemical properties. Notably, higher T_2^{IE} values are located in GM and WM tracks with potentially higher axon calibers (e.g., the corticospinal tract and the optic radiations) [41], and smaller values are found in subcortical structures (e.g., putamen, pallidum) which may be affected by iron deposition, which is well-know to alter the magnetic field homogeneity and to decrease the T_2. Mean T_2^{IE} maps with similar image contrasts were previously reported in a multi-parametric atlas throughout the adult life span [42].

4.3. Estimation methods for T_2^{IE}

Results from the ICC metric indicate that the regularized NNLS method based on the L-curve−I is the best method for estimating T_2^{IE}, in terms of reproducibility. Interestingly, a previous study reported a superior performance of the MIML algorithm to estimate the MWF in WM Yu et al. However, as the performance of both methods for estimating T_2^{IE} was not compared, it is not possible to know if our findings are agreement or disagreement with those reported in Yu et al.

[25]. We noticed that a potential bias in the evaluation could emerge as the MIML algorithm was trained for noisy data and, as such, results from denoised data may be sup-optimal. Nevertheless, we can discard this issue because our evaluation was conducted for both raw and denoised data, and results from both datasets showed a similar pattern. In fact, results from raw data were slightly more reproducible than those from denoised data, at a ROI level. Regarding the variability in different brain regions, the lowest ICC was found in frontal lobe regions and the highest in parietal regions, where the ICC strongly depended on the reconstruction method, being the L-curve−I the one that performed the best, and MIML the worst. The lower ICC found in frontal regions could be explained by a higher level of local noise, MRI-related artifacts, and residual errors in estimating the flip angle due to magnetic field inhomogeneity.

4.4. Limitations

This study has certain limitations. First, all the analyses are based on the same multi-echo T_2 MRI sequence. Future studies should compare additional acquisition parameters and data collected from different MRI vendors (e.g., Siemens, Philips, GE). Second, all the participants were young healthy subjects. While this allowed us to reduce age-related differences, additional analyses comparing the sensitivity of the employed reconstruction methods to detect age-related differences were not possible. Third, future studies could evaluate the sensitivity of T_2^{IE} to detect abnormalities in patients, as well as the comparison of the different metrics derived from multi-component T_2 analyzes [see [43]]. Fourth, while the implemented EPG framework allowed us to study the variation in the MRI signals as a function of flip angle error, other factors including magnetization transfer effects and exchange [44], and the effect of diffusion due to internal gradients [45] caused by magnetic susceptibility differences at the fluid-tissue interfaces, were not considered. However, this is a standard assumption in multi-echo T_2 relaxometry and its generalization would imply the development and validation of new models, which is beyond the scope of this work. Five, despite the fact that T_2^{IE} is relatively immune to partial volume effects, we noticed image gradients around the boundary between GM and CSF. Nevertheless, as these gradients were not present in the individual images, we conclude that they were introduced by the normalization process, e.g., due to minor registration inconsistencies at the group level, and the linear interpolation (i.e., smoothing) done by the registration algorithm at the subject level. Finally, although we evaluated three reconstruction methods, additional algorithms could be analyzed, including the recently proposed Bayesian NNLS method [14] and the spatially regularized technique proposed by Kumar et al. [46].

5. Conclusion

We have acquired a unique multi-echo T2 MRI dataset to characterize the variability and reproducibility of the intra- and extra-cellular T2 relaxation time (T_2^{IE}). Moreover, we compared the estimates from three different reconstruction methods, including two classical algorithms based on regularized non-negative least squares and a novel machine learning approach trained with synthetic data. The analysis was conducted by using raw and denoised data, separately. We found that the smallest source of variance is the run (i.e., inter-run), followed by inter-session, inter-scanner, and inter-subject effects, respectively. Notably, there were no statistical differences between the inter-session and inter-scanner effects for any of the evaluated reconstruction techniques, suggesting that the acquisition sequence and employed methodology may be used in multi-site neuroimaging studies. Results from raw data were slightly more reproducible than those from denoised data. To the best of our knowledge, this is the first work reporting the variability and reproducibility of T_2^{IE} across the cortical mantle, globally and in different brain lobes. Interestingly, the variability in the GM was smaller than that in the WM. Therefore, T_2^{IE} could be a helpful imaging biomarker to characterize microstructure and molecular abnormalities in a range of pathological conditions in both GM and WM tissue types. Finally, the non-negative least squares method based on the L-curve technique produced the lowest intra-class correlation; therefore, it may be the preferred method for estimating the T_2^{IE} time.

Data availability statement

The datasets presented in this article are not readily available because requires a formal data sharing agreement. Requests to access the datasets should be directed to friedhelm.hummel@epfl.ch.

Ethics statement

The studies involving human participants were reviewed and approved by Vaud Cantonal Ethics Committee (Switzerland) Project number: 2018-01355. The patients/participants provided their written informed consent to participate in this study.

Author contributions

EF-G, GG, PK, MP, JB, GP, TH, AD, TK, EC-R, FH, and J-PT formulated the research goals and aims. EF-G, GG, MP, GP, TH, TY, SS, AD, TK, and EC-R design the methodology and analyzed the data. PK, JB, AC-M, EB, C-HP, TM, and MW collected the data. EF-G, GG, and EC-R wrote the original draft. EF-G, GG, PK, MP, GP, TH, TY, TK, EC-R, and FH reviewed and edited the manuscript. FH and J-PT acquired the financial support to perform the study. All authors contributed to the article and approved the submitted version.

Conflict of interest

Authors GP, TH, TY, and TK are employees of Siemens Healthineers International AG, Switzerland.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Mackay A, Whittall K, Adler J, Li D, Paty D, Graeb D. *In vivo* visualization of myelin water in brain by magnetic resonance. *Magn Reson Med*. (1994) 31:673–7. doi: 10.1002/mrm.1910310614

2. Dvorak AV, Wiggermann V, Gilbert G, Vavasour IM, MacMillan EL, Barlow L, et al. Multi-spin echo T2 relaxation imaging with compressed sensing (METRICS) for rapid myelin water imaging. *Magn Reson Med*. (2020) 84:1264–79. doi: 10.1002/mrm.28199

3. Piredda GF, Hilbert T, Canales-Rodríguez EJ, Pizzolato M, von Deuster C, Meuli R, et al. Fast and high-resolution myelin water imaging: accelerating multi-echo GRASE with CAIPIRINHA. *Magn Reson Med*. (2021) 85:209–22. doi: 10.1002/mrm.28427

4. Alonso-Ortiz E, Levesque IR, Pike GB. MRI-based myelin water imaging: a technical review. *Magn Reson Med*. (2015) 73:70–81. doi: 10.1002/mrm.25198

5. Whittall KP, MacKay AL. Quantitative interpretation of NMR relaxation data. *J Magn Reson*. (1989) 84:134–52. doi: 10.1016/0022-2364(89)90011-5

6. Borich MR, MacKay AL, Vavasour IM, Rauscher A, Boyd LA. Evaluation of white matter myelin water fraction in chronic stroke. *Neuroimage*. (2013) 2:569. doi: 10.1016/j.nicl.2013.04.006

7. MacKay AL, Laule C. Magnetic resonance of myelin water: an *in vivo* marker for myelin. *Brain Plasticity*. (2016) 2:71–91. doi: 10.3233/BPL-160033

8. Lakhani B, Hayward KS, Boyd LA. Hemispheric asymmetry in myelin after stroke is related to motor impairment and function. *Neuroimage Clin*. (2017) 14:344–53. doi: 10.1016/j.nicl.2017.01.009

9. Faizy TD, Thaler C, Broocks G, Flottmann F, Leischner H, Kniep H, et al. The myelin water fraction serves as a marker for age-related myelin alterations in the cerebral white matter-a multiparametric MRI aging study. *Front Neurosci*. (2020) 14:136. doi: 10.3389/fnins.2020.00136

10. Piredda GF, Hilbert T, Thiran JP, Kober T. Probing myelin content of the human brain with MRI: A review. *Magn Reson Med*. (2021) 85:627–52. doi: 10.1002/mrm.28509

11. Levesque IR, Chia CLL, Pike GB. Reproducibility of *in vivo* magnetic resonance imaging-based measurement of myelin water. *J Magn Reson Imaging*. (2010) 32:60–8. doi: 10.1002/jmri.22170

12. Laule C, Vavasour IM, Mädler B, Kolind SH, Sirrs SM, Brief EE, et al. MR evidence of long T2 water in pathological white matter. *J Magn Reson Imaging*. (2007) 26:1117–21. doi: 10.1002/jmri.21132

13. Canales-Rodríguez EJ, Alonso-Lana S, Verdolini N, Sarró S, Feria I, Montoro I, et al. Age- and gender-related differences in brain tissue microstructure revealed by multi-component T2 relaxometry. *Neurobiol Aging*. (2021) 106:68–79. doi: 10.1016/j.neurobiolaging.2021.06.002

14. Canales-Rodríguez EJ, Pizzolato M, Yu T, Piredda GF, Hilbert T, Radua J, et al. Revisiting the T2 spectrum imaging inverse problem: bayesian regularized non-negative least squares. *Neuroimage*. (2021) 244:118582. doi: 10.1016/j.neuroimage.2021.118582

15. Canales-Rodríguez EJ, Pizzolato M, Piredda GF, Hilbert T, Kunz N, Pot C, et al. Comparison of non-parametric T2 relaxometry methods for myelin water quantification. *Med Image Anal*. (2021) 69:101959. doi: 10.1016/j.media.2021.101959

16. Meyers SM, Laule C, Vavasour IM, Kolind SH, Mädler B, Tam R, et al. Reproducibility of myelin water fraction analysis: a comparison of region of interest and voxel-based analysis methods. *Magn Reson Imaging*. (2009) 27:1096–103. doi: 10.1016/j.mri.2009.02.001

17. Meyers SM, Vavasour IM, Mädler B, Harris T, Fu E, Li DKB, et al. Multicenter measurements of myelin water fraction and geometric mean T2: Intra- and intersite reproducibility. *J Magn Reson Imaging*. (2013) 38:1445–53. doi: 10.1002/jmri.24106

18. Nguyen TD, Deh K, Monohan E, Pandya S, Spincemaille P, Raj A, et al. Feasibility and reproducibility of whole brain myelin water mapping in 4 minutes using fast acquisition with spiral trajectory and adiabatic T2prep (FAST-T2) at 3T. *Magn Reson Med*. (2016) 76:456–65. doi: 10.1002/mrm.25877

19. Lee LE, Ljungberg E, Shin D, Figley CR, Vavasour IM, Rauscher A, et al. Inter-Vendor reproducibility of myelin water imaging using a 3D gradient and spin echo sequence. *Front Neurosci*. (2018) 12:854. doi: 10.3389/fnins.2018.00854

20. Drenthen GS, Backes WH, Aldenkamp AP, Jansen JFA. Applicability and reproducibility of 2D multi-slice GRASE myelin water fraction

with varying acquisition acceleration. *Neuroimage*. (2019) 195:333–9. doi: 10.1016/j.neuroimage.2019.04.011

21. Cai LY, Yang Q, Kanakaraj P, Nath V, Newton AT, Edmonson HA, et al. MASiVar: multisite, multiscanner, and multisubject acquisitions for studying variability in diffusion weighted MRI. *Magn Reson Med*. (2021) 86:3304–20. doi: 10.1002/mrm.28926

22. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: practical *in vivo* neurite orientation dispersion and density imaging of the human brain. *Neuroimage*. (2012) 61:1000–16. doi: 10.1016/j.neuroimage.2012.03.072

23. Garyfallidis E, Côté M, Rheault F, Sidhu J, Hau J, Petit L, et al. Recognition of white matter bundles using local and global streamline-based registration and clustering. *Neuroimage*. (2018) 170:283–95. doi: 10.1016/j.neuroimage.2017.07.015

24. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*. (2010) 53:1059–69. doi: 10.1016/j.neuroimage.2009.10.003

25. Yu T, Canales-Rodríguez EJ, Pizzolato M, Piredda GF, Hilbert T, Fischi-Gomez E, et al. Model-informed machine learning for multi-component T2 relaxometry. *Med Image Anal*. (2021) 69:101940. doi: 10.1016/j.media.2020.101940

26. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. (2002) 17:825–41. doi: 10.1006/nimg.2002.1132

27. Smith S. Fast robust automated brain extraction. *Hum Brain Mapp*. (2002) 17:143–55. doi: 10.1002/hbm.10062

28. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging*. (2001) 20:45–57. doi: 10.1109/42.906424

29. Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*. (2010) 53:1–15. doi: 10.1016/j.neuroimage.2010.06.010

30. Van Der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, et al. Scikit-image: image processing in python. *PeerJ*. (2014) 2014:e453. doi: 10.7717/peerj.453

31. Bouhrara M, Reiter DA, Bergeron CM, Zukley LM, Ferrucci L, Resnick SM, et al. Evidence of demyelination in mild cognitive impairment and dementia using a direct and specific MRI measure of myelin content. *Alzheimers Dementia*. (2018) 14:998. doi: 10.1016/j.jalz.2018.03.007

32. Jones CK, Whittall KP, MacKay AL. Robust myelin water quantification: averaging vs. spatial filtering. *Magn Reson Med*. (2003) 50:206–9. doi: 10.1002/mrm.10492

33. Does MD, Olesen JL, Harkins KD, Serradas-Duarte T, Gochberg DF, Jespersen SN, et al. Evaluation of principal component analysis image denoising on multi-exponential MRI relaxometry. *Magn Reson Med*. (2019) 81:3503–14. doi: 10.1002/mrm.27658

34. Donoho DL, Johnstone IM. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. (1994) 81:425. doi: 10.1093/biomet/81.3.425

35. Prasloski T, Mädler B, Xiang QS, MacKay A, Jones C. Applications of stimulated echo correction to multicomponent T2 analysis. *Magn Reson Med*. (2012) 67:1803–14. doi: 10.1002/mrm.23157

36. Prasloski T, Rauscher A, MacKay AL, Hodgson M, Vavasour IM, Laule C, et al. Rapid whole cerebrum myelin water imaging using a 3D GRASE sequence. *Neuroimage*. (2012) 63:533–9. doi: 10.1016/j.neuroimage.2012.06.064

37. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467* (2016) doi: 10.48550/arXiv.1603.04467

38. vanRossum G. *Python Reference Manual. Department of Computer Science [CS]*, (R 9525) (1995).

39. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv [Preprint]*.(2014). arXiv: 1412.6980. Available online at: https://arxiv.org/pdf/1412.6980.pdf

40. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012

41. Veraart J, Raven EP, Edwards LJ, Weiskopf N, Jones DK. The variability of MR axon radii estimates in the human white matter. *Hum Brain Mapp*. (2021) 42:2201–13. doi: 10.1002/hbm.25359

42. Dvorak AV, Swift-LaPointe T, Vavasour IM, Lee LE, Abel S, Russell-Schulz B, et al. An atlas for human brain myelin content throughout the adult life span. *Sci Rep*. (2021) 11:1–13. doi: 10.1038/s41598-020-79540-3

43. Canales-Rodríguez EJ, Pomarol-Clotet E, Radua J, Sarró S, Alonso-Lana S, Del Mar Bonnín C, et al. Structural abnormalities in bipolar euthymia: a multicontrast molecular diffusion imaging study. *Biol Psychiatry*. (2014) 76:239–48. doi: 10.1016/j.biopsych.2013.09.027

44. Malik SJ, Teixeira RPAG, Hajnal JV. Extended phase graph formalism for systems with magnetization transfer and exchange. *Magn Reson Med*. (2018) 80:767–79. doi: 10.1002/mrm.27040

45. Weigel M, Schwenk S, Kiselev VG, Scheffler K, Hennig J. Extended phase graphs with anisotropic diffusion. *J Magn Reson*. (2010) 205:276–85. doi: 10.1016/j.jmr.2010.05.011

46. Kumar D, Siemonsen S, Heesen C, Fiehler J, Sedlacik J. 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*. (2016) 43:800–17. doi: 10.1002/jmri.25078