

Signal Intensity Normalization of Multispectral Magnetic Resonance Images Obtained with Clinical Sequences

Matthew R. Brier^{a,*}, Biao Xiang^b, Zhuocheng Li^a, Robert T. Naismith^a,
Dmitriy Yablonskiy^b, Anne H. Cross^a, Tammie L. S. Benzinger^b, Abraham Z.
Snyder^b

^a*Department of Neurology, Washington University in St. Louis,*

^b*Department of Radiology, Washington University in St. Louis,*

Abstract

Assessment of intrinsic tissue integrity is commonly accomplished via quantitative relaxometry or other specialized imaging, which requires sequences and analysis procedures not routinely available in clinical settings. We detail an alternative technique for extraction of quantitative tissue biomarkers based on intensity normalization of T1- and T2-weighted images. We develop the theoretical underpinnings of this approach and demonstrate its utility in imaging of multiple sclerosis.

1. Introduction

Magnetic resonance imaging (MRI) is the premier tool for diagnosis and monitoring of numerous neurological conditions [5, 22, 10, 29]. Focal lesions (e.g., multiple sclerosis [MS] plaques, brain tumors, infarcts) are readily detectable both visually and through automated segmentation algorithms [4, 12, 30, 32]. However, additional subtle abnormalities are increasingly recognized as important [23, 1, 21] in MS [3] as well as other conditions [15]. Conventionally, quantification of abnormalities in normal appearing tissue requires specialized sequences or extensive image processing [20, 16, 27], which generally precludes application in routine clinical practice. In this work, we characterize a recently

*Corresponding Author

Email address: brierm@wustl.edu (Matthew R. Brier)

described technique [3] for the estimation of standardized intensity properties in images obtained with clinically available T1-weighted and T2-weighted (FLAIR) sequences.

T1 and T2 are intrinsic properties of tissue that depend on the underlying cellular composition [28, 35]. Reconstructed intensities in T1- and T2-weighted images depend on the intrinsic properties of the tissue being studied interacting with sequence parameters, spatial encoding radiofrequency (RF) field (B1), and the main field (B0) [2]. However, reconstructed image intensities also vary with uncontrolled factors including quality of shim, head size, temperature, etc. These dependencies preclude direct, numerical comparison of T1- and T2-weighted intensities across sessions or individuals. Standardization of T1- and T2-weighted image intensities would facilitate comparison across sessions or individuals and allow for new questions to be asked in legacy or clinical imaging data.

We previously reported results obtained by standardization of T1-weighted (T1w) and T2-weighted (T2w) images using an intensity histogram normalization procedure applied to data from a multinational collaboration (Multiple Sclerosis Partners Advancing Technology and Health Solutions; MS PATHS) [3]. In this procedure, following standard image preprocessing (e.g., spatial alignment, skull stripping), bivariate voxel-intensity histograms are constructed from the T1w and FLAIR (T2w) data. These histograms are affine registered to a standard intensity histogram template, thereby standardizing intensity values across subjects. In our previous work, we obtained signal mean and standard deviation estimates for each image (T1w, FLAIR) and tissue class (cerebrospinal fluid, gray matter, white matter; CSF, GM, WM) by decomposing intensity normalized histograms using Gaussian mixture modeling (GMM). When evaluated in cross-sectional data from patients with MS, these GMM features separated MS patients from controls, distinguished between MS subtypes, and correlated more closely with MS-related disability than more conventional lesion-based imaging measures. Having demonstrated the utility of the intensity normalization approach, we now demonstrate its applicability to longitudinal datasets.

We first describe the intensity normalization algorithm in greater detail and then demonstrate test-retest reliability of derived measures. More specifically, we describe reliability of lesion segmentation as well as intensity-based measures within *a priori* regions of interest.

2. Theory

Objectives of imaging in the present context include distinguishing between tissue classes (e.g., MS lesions vs. normal appearing WM) and quantification of abnormalities in those tissues (e.g., WM in controls vs. patients). It has been established that this can be done on the basis of quantitative determination of T1, T2, and related quantities [20, 35, 31]. But, as discussed above, quantitative measurement of T1 and T2 cannot be reliably done on the basis of T1w and T2w images.

However, in theory, there exists an informational equivalence between T1 and T2 vs. T1w and T2w images can be formally cast in terms of Kullback-Liebler divergence, $I(1:2)$. At the single voxel level, $I(1:2)$ is information in favor of H_1 (e.g., lesion) vs. H_2 (e.g., normal appearing WM)¹ [17]. To further consider this issue theoretically, define $X = \{T1w, FLAIR\}_{norm}$ (normalized intensity measure) vs. $Y = \{T1, T2\}$ (absolute measures). The question, then, is $I(1:2; X)$ comparable to $I(1:2; Y)$? Invariance theory states that $I(1:2; X) \leq I(1:2; Y)$ with equality if there exists a smooth, invertible relation between X and Y . Thus, $I(1:2; X) = I(1:2; Y)$ if there exists some relation, F , such that $Y = F(X)$ and $X = F^{-1}(Y)$. The algebraic proof is remarkably simple when X and Y are multivariate Gaussian distributions and F is a linear transform (see p. 194 of [17]). However, it is not necessary that F be linear (or affine), only that it be non-singular (see Fig. 5 in [31]). This condition presumably applies to the relation between $\{T1w, FLAIR\}_{norm}$ and $\{T1, T2\}$, as it may be assumed that

¹Here we use the original notation in [17]. For clarity, $I(1:2)$ is equivalent to the more modern notation, $D_{KL}(P_1(T1w, FLAIR)|P_2(T1w, FLAIR))$, where P_1 and P_2 represent intensity distributions corresponding to different tissue classes.

reconstructed image intensities depend smoothly on the physical properties of brain tissue and variation in MRI pulse sequence parameters. We empirically demonstrate this principle below (Section 4.1).

We have so far considered discrimination between tissue classes at the single voxel level. We now extend the theory to the whole brain. As was noted above, the information in $\{T1w, FLAIR\}_{norm}$ and $\{T1, T2\}$ is equivalent provided that there exists a smooth relation, F , connecting the two measures. Although it may be assumed that a smooth F exists at every voxel, this relation may not be the same everywhere in the brain. Nevertheless, we here empirically assume a fixed relation between $\{T1w, FLAIR\}_{norm}$ and $\{T1, T2\}$. Our hypothesis is based on the observation that intensity normalization, described below, can mitigate multiple sources of variance including magnetic field inhomogeneities. Accordingly, the experimental question becomes whether this approach supports reliable tissue characterization. Thus, we propose to compensate for uncontrolled factors that influence T1w and T2w image intensities by requiring that $\{T1w, FLAIR\}$ intensity histograms conform to a standard intensity template. At issue, then, is the extent to which normalized intensity histograms of the form described below (Section 3.3) can be used to characterize tissue in place of quantitatively measured T1 and T2.

3. Materials and Methods

3.1. Study Design, Participants, and Image Acquisition

This study makes use of two datasets. The first dataset included $N_1 = 7$ healthy young adults with both quantitative T1 imaging as well as clinical-type T1w images. This dataset was used to demonstrate the theoretical informational equivalence between weighted and absolute intensity histograms. The second dataset included $N_2 = 30$ subjects with MS who were repeatedly studied using clinically available, high resolution T1w (MP-RAGE) and FLAIR (T2w) sequences. This dataset was used to demonstrate the test-retest reliability of intensity normalized measures. Each of these datasets are small compared to

existing datasets but benefit from containing specific sequences (e.g., quantitative T1 imaging) or are densely sampled (e.g., 4 imaging sessions within the span of a week) and are thus suitable for technique development.

3.1.1. Quantitative vs. Weighted Comparison Cohort

Seven healthy controls (2 male, 5 female, age range 22–27 years) were imaged for the comparison of the informational content of intensity normalized clinical images vs. absolute T1 estimation. All participants were neurologically normal, able to undergo MRI, and provided informed consent. Imaging was conducted on a 3T Prisma scanner (Siemens, Erlangen, Germany) equipped with a 32-channel phased-array head coil. T1w images were acquired with an MP-RAGE sequence (0.8mm isotropic voxels, TR 2.4s, TE 2.22ms, inversion time 1s, $\alpha = 8^\circ$). For quantitative T1 mapping, voxels were 1.3mm isotropic, acquired using three-dimensional multi-gradient-echo sequence with five flip angles ($\alpha = 5^\circ, 10^\circ, 20^\circ, 40^\circ, 60^\circ$) and three gradient echo times ($TE = 3ms, 7ms, 11ms$) for each α . The sequence incorporated a generalized auto-calibrating, partially parallel acquisition algorithm [13] with an acceleration factor of 2, and 24 auto-calibrating lines in each phase encoding direction. T1 maps were generated by fitting the combined data to the Ernst equation [9] with $B1$ correction [36] that accounted for the imperfect radio frequency spoiling and transverse magnetization relaxation.

3.1.2. Test-Retest Reliability Cohort

Thirty MS patients (17 female, age ranged 21–55 years, EDSS [18] 0–6) were enrolled in the reliability sub-study of MS PATHS [24]. MS PATHS is a natural history study funded by Biogen in which participating centers contribute standardized imaging and behavioral data from patients with MS. For the reliability sub-study, each participant underwent four MRI sessions on two scanners over two days. Scanning sessions on the same day were 1–6 hours apart and scanning days were 2–7 days apart. Thus, each participant contributed 4 repeated MP-RAGE (T1w; field of view 256mm \times 256mm, 1mm isotropic voxels, TR 2.3s, TE

2.98ms, TI 900ms) and 4 repeated FLAIR (T2w; field of view 256mm \times 256mm, 1mm isotropic voxels, TR 5s, TE 393ms, TI 1800ms) images of the same brain². All imaging was acquired on Siemens 3T scanners; model of scanner varied with acquisition site.

3.2. Image Pre-processing

T1w and FLAIR images from each session were 9-parameter affine coregistered (rigid body + cardinal axis scaling) and the data obtained in all four sessions were mutually coregistered using in-house software³. Following composition of transforms, all images were resampled (one step) in a standard atlas space (1mm³ voxels). Images then underwent brain extraction (BET) [33] and bias field correction (FAST) [37], implemented in the FSL toolbox [14]. Intensity values in each weighted image were scaled (one multiplicative constant per image) to obtain a voxel intensity distribution mode value of 1,000 over the whole-brain; this scaling accelerates subsequent intensity normalization.

3.3. Image Intensity Histogram Normalization

The goal of image intensity histogram normalization is the standardization of bivariate intensity (T1w and FLAIR) distributions to match a standard template [3]. Prior to normalization, histogram features are grossly similar across individuals but vary in extent and skew owing to factors of non-interest (e.g., proximity of the head to the receiver coils) in addition to biologically relevant factors (e.g., disease status). To accomplish normalization, each individual's data were 6-parameter affine transformed to match the template. Histogram similarity was evaluated as the mean-squared error between the template and transformed individual histogram. Optimization was achieved by regular step gradient descent [34] implemented in MATLAB 2020a (Mathworks, Natick,

²Assuming that MS disease pathophysiology did not progress measurably over 7 days, given no clinical relapses in the cohort.

³<http://4dfp.readthedocs.io>

MA). Let A be the affine transform that minimizes the difference between each individual's data and the template. This matrix has the form

$$A = \begin{bmatrix} sc_{T1} & sh_{T1} & tr_{T1} \\ sh_{FL} & sc_{FL} & tr_{FL} \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where sc , sh , and tr represent scaling, shearing, and translation along either the T1w or FLAIR axis. Let $X = \{T1w, FLAIR\}$ represent an individual's bivariate intensity histogram prior to normalization. Multiplication of X by optimized A

$$X_{norm} = A \cdot X \quad (2)$$

yields the normalized bivariate intensity histogram. Voxelwise application of A to the bivariate data generates a T1w and FLAIR images that have been intensity normalized to a standard template.

3.4. Image Tissue Scores and Gradients

Tissue classes are represented as distributions around centroids in bivariate intensity space. Tissue class centroids were defined *a priori* by manual segmentation of a single representative image. For each voxel, a continuous score $u_i \in [0, 1]$ was assigned for each tissue class i such that $\sum_i u_i = 1$. The algebra regarding the calculation of u_i follows.

For any multispectral image voxel, the standardized squared distance from the i -th tissue class centroid \bar{f}_{ij} is given as

$$d_i^2 = \sum_j \left[\frac{f_j - \bar{f}_{ij}}{r_j} \right]^2 \quad (3)$$

where j indexes contrasts (e.g., T1w, FLAIR) and r_j is the intensity range containing non-artifactual intensity values of contrast j . Raw membership in class i varies as $1/d_i^2$ and the normalized membership class as

$$u_i = \frac{1/d_i^2}{\sum_k 1/d_k^2} \quad (4)$$

where k is a dummy summation index. Tissue class membership can be defined by identifying the class i with the largest u_i ('winner-take-all' [WTA] segmentation).

Tissue class gradients have multiple uses in the analysis of multi-spectral images. Let $D_j \equiv \partial/\partial f_j$. It follows,

$$D_j d_i^2 = 2 \sum_j \left[\frac{f_j - \bar{f}_{ij}}{r_j^2} \right] \quad (5)$$

$$D_j(1/d_i^2) = -(d_i^2)^{-2} D_j d_i^2 \quad (6)$$

$$D_j u_i = \frac{D_j(1/d_i^2) - (1/d_i^2)(D_j g)/g}{g \cdot r_j} \quad (7)$$

where $g = \sum_k (1/d_k^2)$ and $D_j g = D_j \sum_k (1/d_k^2)$. Application of the chain rule yields tissue class membership gradients in image space

$$\nabla u_i = \left[\frac{\partial u_i}{\partial f_j} \frac{\partial f_i}{\partial x} \right] \hat{i} + \left[\frac{\partial u_i}{\partial f_j} \frac{\partial f_i}{\partial y} \right] \hat{j} + \left[\frac{\partial u_i}{\partial f_j} \frac{\partial f_i}{\partial z} \right] \hat{k} \quad (8)$$

where $\{\hat{i}, \hat{j}, \hat{k}\}$ are the (x, y, z) cardinal unit vectors in image space.

Boundaries between tissue compartments correspond to loci at which two class memberships are dominant and approximately equal. In the archetypical case, $u_1 = u_2 = 0.5$, taking classes 1 and 2 as examples. In general, the tissue class boundary corresponds to $u_1 \approx u_2$, with $u_1 + u_2 < 1$, because of finite representation of other tissue classes ($\sum_i u_i = 1$). Figure 1 illustrates a toy case in which the tissue class boundary normal points exactly in the x -direction. The boundary occurs where $|u_2(du_1/dx) - u_1(du_2/dx)|$ is maximal. In 3-D, this expression generalizes to the locus where $|u_2 \nabla u_1 - u_1 \nabla u_2|$ is maximal. A rationale for this expression may be seen in the equality

$$u_2 \nabla u_1 - u_1 \nabla u_2 = (u_1^2 + u_2^2) \nabla \arctan(u_1/u_2) \quad (9)$$

The term $\nabla \arctan(u_1/u_2)$ attains a maximal value when $u_1 = u_2$. Thus, a boundary corresponds to a locus at which $u_1 \approx u_2$ and $(u_1^2 + u_2^2)$ is substantial.

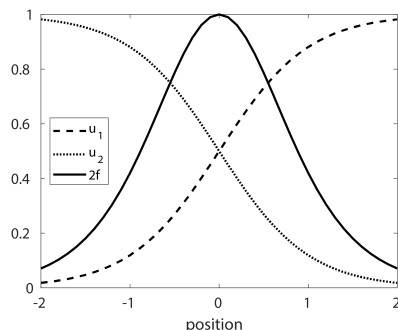


Figure 1: Toy case illustrating the tissue class boundary condition. Let $u_1 = e^x / (e^x + e^{-x})$ and $u_2 = e^{-x} / (e^x + e^{-x})$. Trivially, $u_1 + u_2 = 1$ (in this case). Evaluating $|u_2 (du_1/dx) - u_1 (du_2/dx)|$, we obtain $2 / (e^x + e^{-x})^2 = f$. The algebra in this toy case has been simplified by the use of logistic functions. However, Equation 9 applies wherever there exist relatively sharp transitions between low vs. high tissue class values.

4. Results

4.1. Demonstration of Informational Equivalence Between Weighted and Absolute Intensity Distributions

Section 2 established the theoretical equivalence of two multivariate histograms provided that they are related via a non-singular (invertible) transformation. Here, we empirically demonstrate the existence of such a transform between reconstructed intensities in a T1w image and quantitative T1. Figure 2 shows the T1w and T1 histograms of a representative subject. Note inversion of the horizontal axes ($T1 \propto 1/T1w$). Nevertheless, comparable features are evident: two prominent peaks corresponding to GM and WM and a long tail at small values (in T1w) or large values (in T1) representing CSF. Thus, it is intuitively plausible that the histograms contain similar information.

We demonstrate that T1w and T1 values can be made equivalent by histogram matching. Let C_{T1} and C_{T1w} be the cumulative density function of the image intensity histograms corresponding to the T1 and T1w images. The transformation between the two histograms therefore is

$$F(C_{T1}) = \arg \min |F(C_{T1}) - C_{T1w}|. \quad (10)$$

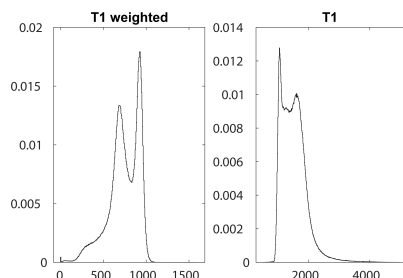


Figure 2: T1w and T1 intensity histograms in a representative subject. Image intensities were extracted from voxels within the brain mask and displayed as histograms. Note inverted order of histogram features: In the T1w image, tissue classes are ordered CSF < GM < WM. In contrast, measured T1 values are ordered WM < GM < CSF.

The resulting transformed histograms are shown as cumulative density functions in Figure 3. The T1w function (black solid line) and the transformed T1 function (green dashed line) are nearly identical. Note that the mapping (F) is different in each subject. We evaluated error as $\epsilon^2 = \sum (C_{T1w} - \hat{C}_{T1})^2$ where \hat{C}_{T1} is the cumulative density function with F applied. Mean ϵ^2 across subjects was $\bar{\epsilon}^2 = 0.0101$. This result demonstrates the existence of a mapping between the intensity histograms of T1w and absolute T1 images. Figure 3 demonstrates 1-D histogram matching. In principle, 2-D histogram matching could be similarly done *in normal participants*, provided that each had the same proportion of tissue classes. However, this is not possible in MS patients as the proportion of tissue classes, especially lesion, varies greatly. In the following, we report results obtained by affine transformation of 2-D histograms.

4.2. Intensity Normalization

Figure 4 illustrates 2-D histogram matching by affine transformation (Eq. 2) in two example subjects. The first panel shows the template intensity histogram. The top row shows the template and individual subject histograms overlaid. Following intensity normalization (bottom row), the overlap between the individual subject histograms and the template is markedly improved. Alignment is determined by minimizing the mean squared error between the template and

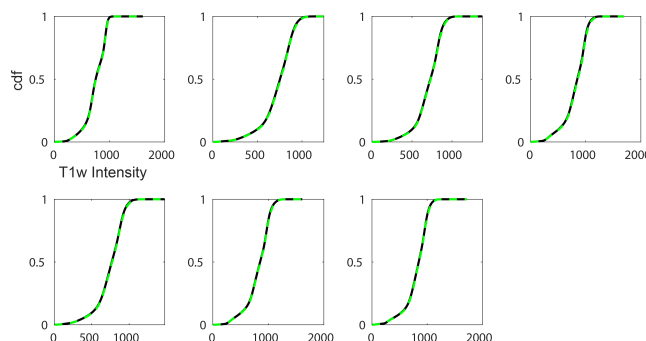


Figure 3: Informational equivalence between T1 and T1w images. Superimposed, individual subject ($N_1 = 7$), T1w cumulative density (black) and transformed T1 cumulative density functions (dashed green). Note near identity of the two cumulative density functions. Thus, there exists a [generally different] F for each subject.

the aligned image.

4.3. Within-Subject Reliability of Intensity Normalization

Intensity histogram normalization generates T1w and FLAIR images with standardized intensities. Figure 5 addresses within-subject and cross-subject comparison of normalized intensity histograms. It may be assumed that there was no interval change over sessions in these subjects. The left two panels show normalized T1w (left) and FLAIR (middle) image intensity histograms for a representative subject superimposed over the four sessions. Each scanning session is represented by a different color. The salient feature of these histograms is their striking consistency.

We quantified histogram similarity using the two-sample Kolmogorov–Smirnov (K-S) test statistic, $KS_{nm} = \sup_x |C_{1,n}(x) - C_{2,m}(x)|$, where $C_{1,n}$ and $C_{2,m}$ are the two probability distributions. Small values of KS indicate similar distributions. We calculated the K-S statistic between all histogram pairs ($4 \times 30 = 120$) within a contrast mechanism (T1w, FLAIR), distinguishing between within-subject vs. between-subject comparisons. The distributions of these K-S values are shown in the right panel of Figure 5. Note small K-S statistics representing within-subject comparisons for both T1w and FLAIR images. These results

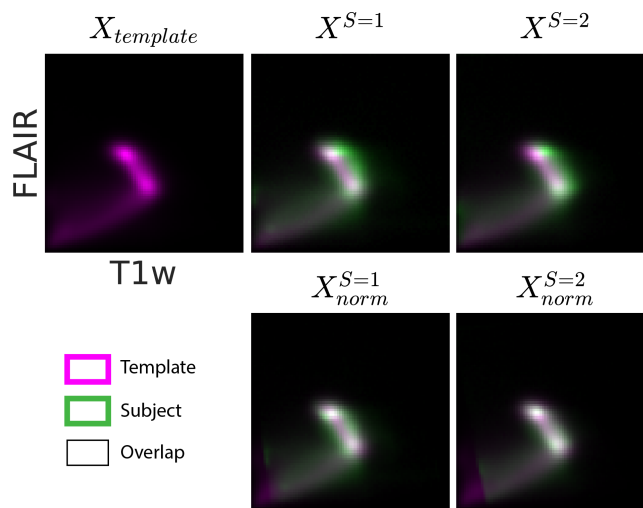


Figure 4: Bivariate intensity histogram normalization improves the match between single-subject data and the template. The top left panel shows the template bivariate intensity histogram, $X_{template}$, in purple. The horizontal and vertical axes index T1w and FLAIR (T2w) intensities, respectively. The two panels in the top row show the template histogram with "raw" single subject histograms overlaid in green. The union overlap of green and purple is displayed as white. The bottom rows show the results of intensity normalization using Equation 2. Objective function error before and after affine transformation was $24.7 \rightarrow 11.5$ and $30.8 \rightarrow 10.7$, respectively, for the two subjects.

provide further empirical evidence that the intensity normalization procedure is reliable. Generally larger cross-subject K-S values reflect individual differences.

We next evaluated test-retest reliability tissue class scores within GM, WM and lesion, determined by WTA classification of intensity-normalized images. Mean (μ) and within-tissue class standard deviation (σ) signal intensities were calculated for T1w and FLAIR images, for all sessions and subjects. Reliability was quantified using intra-class correlation (ICC) for T1w μ , FLAIR μ , T1w σ , and FLAIR σ separately. This calculation was performed with sessions as repeated measures and subjects as the ensemble average. Resulting ICC values are shown in Table 1. Large ICC values indicate reliable tissue score mean and variability estimates.

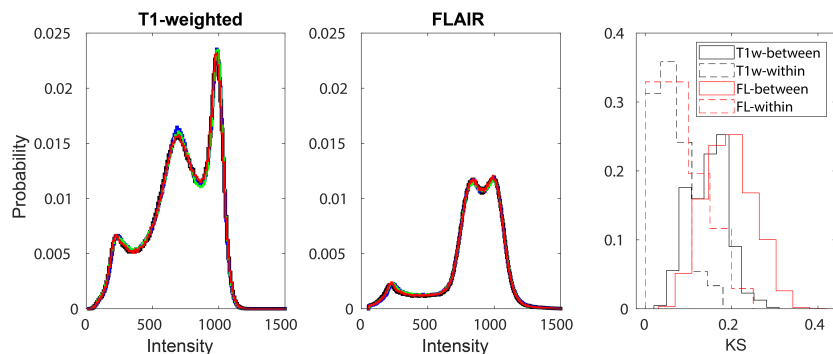


Figure 5: Normalized image intensity histograms are reliable across imaging sessions. Intensity normalized image intensity histograms from a single representative subject are shown for the T1-weighted and FLAIR images. The four test-retest sessions are shown in different colors. The critical feature is nearly perfect overlap across sessions. The rightmost panel shows the distribution of K-S statistics for within-subject and between-subject histogram comparisons for T1-weighted and FLAIR.

	T1w μ	FLAIR μ	T1 σ	FLAIR σ
WM	0.81	0.80	0.89	0.93
GM	0.89	0.90	0.95	0.80
Lesion	0.89	0.95	0.75	0.86

Table 1: ICC representing the reproducibility of tissue characteristic estimates over sessions, averaged across subjects.

We repeated the tissue class score ICC analysis in large, subcortical structures segmented in the T1w image using the FIRST tool implemented in FSL [26]. Note generally high ICC values (> 0.7) in these structures (Table 2).

4.4. Between-Subject Reliability of Intensity Normalization

We next assess across-subject reliability of 2-D histogram normalization in the reproducibility cohort. This assessment is complicated by substantial variation in the degree of MS severity. Hence, some variability of intensity normalized histograms is expected. Nevertheless, major features of these histograms should be similar. The first column of Figure 6 shows the mean T1w and FLAIR intensity normalized histogram (black line) with the single subject histograms

	T1 μ	FLAIR μ	T1 σ	FLAIR σ	Volume (mm ³)
Left Thalamus	0.92	0.74	0.97	0.90	7959.5
Left Caudate	0.85	0.89	0.93	0.90	3621.6
Left Putamen	0.87	0.90	0.76	0.90	5128.9
Left Pallidum	0.92	0.92	0.93	0.89	1864.6
Right Thalamus	0.91	0.67	0.97	0.78	7841.6
Right Caudate	0.84	0.91	0.93	0.89	3809.3
Right Putamen	0.90	0.89	0.85	0.87	5271.4
Right Pallidum	0.91	0.92	0.95	0.90	1900.2

Table 2: ICC for image intensity parameters within subjects across sessions. Mean structure volume is also shown.

(green lines) superimposed. To further clarify the reliability across subjects, the second and third columns of Figure 6 show the peaks corresponding to the GM and WM locus. Peak locations vary across subjects but there is a strong central tendency. The box plots in the final column show how the peak location varies across subjects. This display demonstrates that histogram normalization produces comparable intensity histograms across subjects.

4.5. Tissue-Score and Gradient-Based Lesion Segmentation

Image segmentation can be accomplished by labeling each voxel according to the tissue class, i , with the greatest value of u_i (WTA approach). Although this is possible for all tissue classes, we focus here on lesion segmentation. Figure 7 shows the "raw" WTA segmentation result in a representative MS patient. Note the presence of voxels surrounding lesions, misclassified as GM (Figure 7). This misclassification is attributable to image noise and volume averaging interacting with proximity of the GM centroid to the trajectory connecting the centroids corresponding to lesion and WM (Figure 7). We describe a novel approach to solving this problem incorporating tissue score gradients, ∇u_i .

Figure 8 illustrates one FLAIR slice containing a prominent MS lesion. The lesion as well as a lesion-free region of brain are shown in close-up. Tissue class

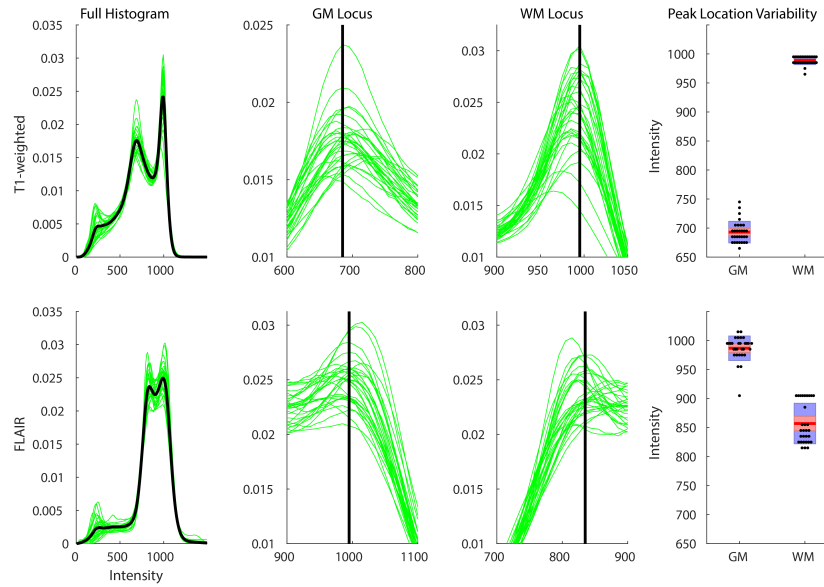


Figure 6: Intensity Normalization Across Subjects. The top and bottom rows show T1w data and FLAIR data, respectively. The first column shows the mean (across subjects) intensity histogram (black lines) with individual subjects superimposed (green lines). The second and third columns are restricted to the region around the GM and WM loci. Vertical black lines indicate locus center-of-mass. The fourth column shows the distribution of peak location across subjects.

gradients (∇u_i) are represented as quiver plots. Note prominent true WM and true lesion gradients as well prominent factitious GM gradients surrounding the lesion. Note absence of lesion gradients in normal appearing WM and presence of lesion gradients inside the lesion boundary. The key feature that enables correction of GM misclassification at the lesion boundary is $\nabla u_{GM} \approx -\nabla u_{Les}$; hence, $\nabla u_{GM} \cdot \nabla u_{Les}$ is strongly negative. Identification of this feature enables correction of the misclassification (see Fig. 9).

Figure 9 shows a FLAIR slice with several prominent lesions shown in close-up. The second column shows voxels classified as GM, demonstrating misclassification surrounding lesions. This misclassification was corrected using an empirical rule: GM voxels within a radius of 2mm of any lesion voxel, in which $\nabla u_{GM} \cdot \nabla u_{Les} < 0.1$, were assigned to the next most highly scored tissue

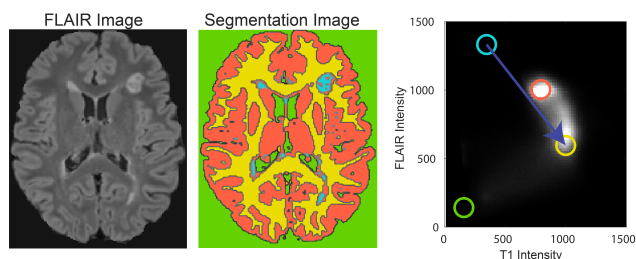


Figure 7: Tissue-Score Based Segmentation. The left panel shows a FLAIR image from a representative MS patient. The WTA segmentation, shown in the middle panel, plausibly corresponds to the FLAIR image. CSF is green; WM is yellow; GM is orange; lesion is blue. Note the presence of voxels misclassified as GM surrounding lesions. A heuristic explanation for this misclassification is illustrated in the right panel: The GM centroid (orange circle) is close to the trajectory (dark blue arrow) connecting the lesion centroid (light blue circle) and the WM centroid (yellow circle).

class. This procedure greatly reduced the factitious ring of misclassified GM surrounding lesions (Figure 9).

Finally, we examined the reliability of lesion segmentation using the presently described lesion segmentation technique (Section 4.5). Measured total lesion volume in the brain and the measured individual lesion volume should not change across the four imaging sessions. To avoid spurious identification of lesion voxels at the brain edge (e.g., as illustrated in Fig. 8) this analysis was restricted voxels identified as WM by FAST segmentation of our standard MP-RAGE atlas template [14]. Figure 10 shows the correlation between total lesion burden pairwise for all session combinations within a subject for all subjects. In total, there are $30 \cdot {}_4C_2$ points represented. The agreement across imaging sessions is very good ($r = 0.97$). Similarly, we identified spatially coincident lesions in each session and compared the segmented volume across sessions. This correlation also is very strong ($r = 0.99$). Thus, our lesion segmentation algorithm produces stable lesion extent estimates. Lastly, we compared the segmentation results derived from our technique and an existing algorithm in the literature [6] for all $30 \cdot 4$ imaging sessions. The correlation in recovered total lesion volume was high ($r = 0.94$) and the correlation between individual lesion volumes was

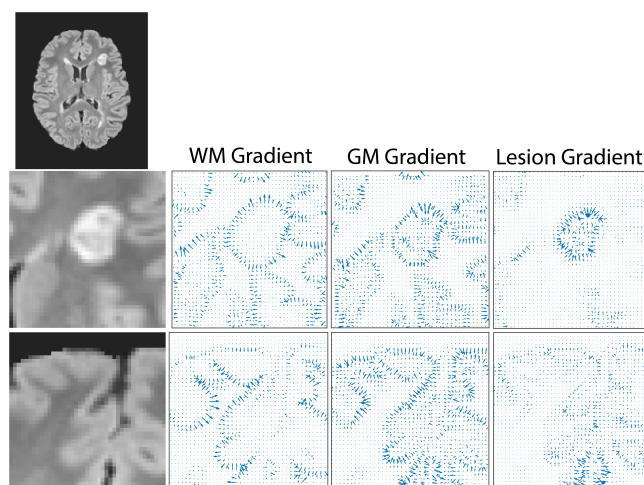


Figure 8: Tissue Class Gradients. The top panel shows one slice of a FLAIR image from a representative MS patient. Two regions are shown in close up in lower rows. The first region contains a stereotypical lesion; the second region is lesion-free. Tissue score gradients corresponding to WM, GM, and lesion are shown as ‘quiver’ plots in which length encodes gradient magnitude. Arrowheads point towards increasing tissue class score. Note true WM gradients and lesion gradients as well as factitious GM gradients surrounding the lesion.

similarly high ($r = 0.95$). One note is that the slope of the total lesion volume scatter plot (Figure 10, left-lower panel) is not 1, indicating small differences in lesion boundary sensitivity and specificity. Overall, this supports the reliability and validity of the present approach.

5. Discussion

Extraction of voxelwise biomarkers of pathology using quantitative relaxometry has been well established [28, 20]. Here, we present an alternative technique which uses intensity normalization of data acquired using standard FDA-approved sequences. Basic information theoretic considerations predict that this is possible; we demonstrate this principle using histogram matching (Figure 3). We show that the present approach, based on affine transformation of bispectral histograms (Figure 4), generates reproducible, quantitative biomarkers of tissue properties. Additionally, these voxelwise measurements facilitate

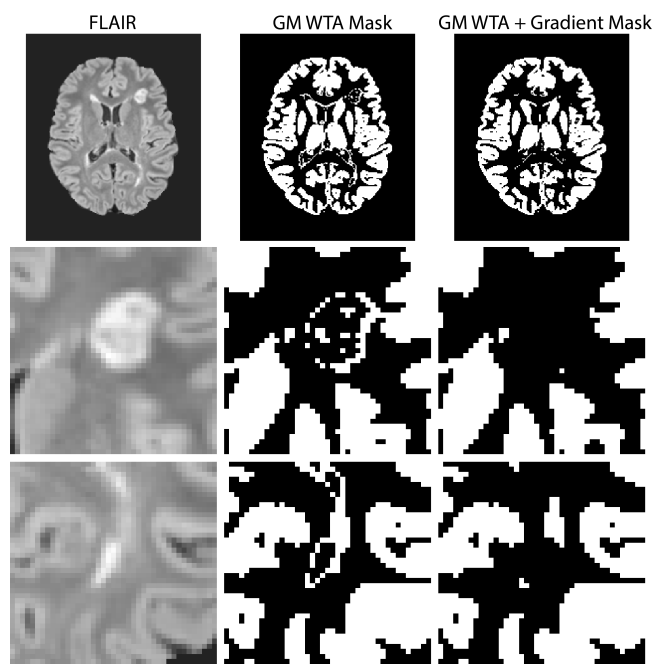


Figure 9: Correction of factitiously labeled GM voxels using tissue score gradients. The left columns shows one slice of a stereotypical FLAIR image from a MS patient. Two regions are shown in close-up in the second and third rows. The second column shows voxels classified as GM prior to correction of misclassification. The third column shows voxels classified as GM following gradient-based correction at lesion boundaries. Note substantial mitigation of misclassification.

tissue segmentation with minimal dependence on template priors. The present technique generates continuous tissue class membership scores. Spatial gradients of these scores are useful for refining tissue segmentation and potentially could be useful in longitudinal studies of subtle changes in lesion boundaries and within-lesion heterogeneity.

We began by claiming that two histograms A and B are informationally equivalent if there exists a transform $A = F(B)$ [17]. F need not be linear (or affine) and only must be smooth and invertible. When applied to T1w images with a known corresponding T1 image, F can be attained by histogram matching. However, when applied to images absent a known underlying distri-

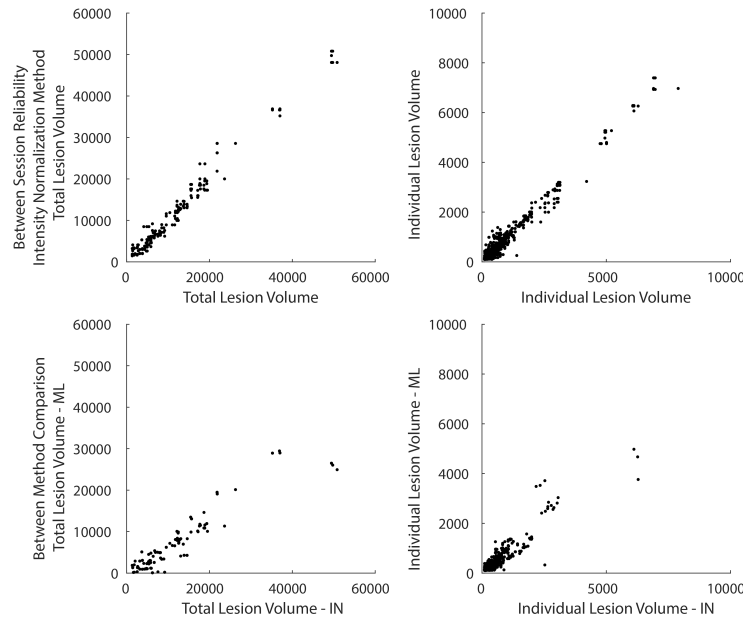


Figure 10: Segmented lesion volume is reliable across scanning sessions. The upper-left panel shows the correlation between total lesion burden across session pairs within subject, compiled over all subjects using the intensity normalization approach. The upper-right panel shows the correlation between individual lesion volumes across session pairs (within subject) compiled over all subjects using the intensity normalization approach. In both cases, the correlation is very strong, indicating reliable lesion segmentation using the procedure illustrated in Figure 9. The bottom panels display the same information comparing the intensity normalization approach with an existing approach in the literature [6]. The correlation between the two methods is high indicating similar lesion segmentation. IN - Intensity normalization approach, ML - machine learning approach described in [6]

bution, direct histogram matching is not feasible given the unknown target. To circumvent this limitation, we utilize affine registration of intensity histograms to a template derived from normal control data. Affine registration allows for the preserved representation of biologically relevant features in the registered data. Such features include discrete lesions or tissue atrophy in addition to changes in intrinsic tissue intensity. As implemented here, affine registration of bispectral intensity histograms reduces the contribution of factors of no interest (e.g., differences in scanner characteristics, head coil differences) while allowing

for the measurement of biologically meaningful signal differences between images. Intensity normalized images exist on an arbitrary but constant intensity scale which we use in place of quantitative relaxometry [20]. Affine registration of bispectral histograms does not account for variability of F across voxels or subjects, which makes the utility of the present method an empirical question. The results shown in Figures 5, 6, and 8 – 10 suggest that the present method has utility in the extraction of quantitative tissue biomarkers as well as lesion segmentation.

Intensity normalization yields reproducible images. Intensity normalized histograms are nearly indistinguishable across sessions within subjects but significantly differ across subjects (Figure 5). We have previously reported the use of tissue mean intensity and variability as a biomarker of MS severity [3]. In the present results these quantities exhibit ICC values above 0.75 on the basis of WTA segmentation (NAWM, GM, Lesion; Figure 9; Table 1). Similarly, using an external subcortical tissue segmentation algorithm, repeatability of tissue mean intensity and variability remained high (ICC greater than 0.67, Table 2). Therefore, intensity normalization yields reliable intensity maps that are stable in a test-retest scheme.

Image segmentation, particularly WM lesion (WML) segmentation, is a well developed technique [4, 12, 30, 32]. Algorithms for automatic tissue segmentation use a variety of approaches including probabilistic mapping and machine learning (for review, see [19, 12]). Tissue segmentation tools (e.g., FAST, FreeSurfer [11, 14]) as well as many WML segmentation tools (e.g., using machine learning algorithms [4, 19]) use both tissue intensity and anatomical priors. Here, we demonstrate tissue segmentation with minimal dependence on anatomical priors. The use of a simple WM prior was needed to constrain the identification of WMLs owing to large factitious lesion scores at the brain edge. Similarly, tissue score gradients were used to improve lesion segmentation as shown in Figure 9. Thus, in addition to extraction of intensity-based biomarkers, intensity normalization has value in image segmentation.

Tissue score gradients describe the transition from low to high tissue class

scores. Considered in pairs, these gradients describe the transition from one tissue class to another. In this way, tissue transitions that are unlikely can be identified and heuristically corrected. For example, in the present work, the transition from Lesion to GM near a lesion cluster is unlikely. In the present application, consideration of tissue score gradients facilitated correction of factitious labeling of GMs around lesions. We note that juxtacortical and GM lesions do occur in MS.

Lesion identification *per se* is only one application of image intensity gradients. In both MS and cerebral small vessel disease, WMLs are dynamic entities capable of expansion or contraction [8, 7] or histological conversion e.g., into black holes [25]. Consideration of tissue gradients potentially allows for the sub-voxel determination of lesion boundaries as described in Figure 1. In this way, given the established reliability of intensity normalization, the boundary of lesions can be assessed longitudinally. Similarly, the existence of tissue gradients within structures (e.g., lesions) may indicate differential tissue damage and may represent an approach to longitudinal monitoring more specific than atrophy. Potential applications of tissue gradients will be assessed in future work.

Acknowledgments

MS PATHS is funded by Biogen. Dr. Brier was supported by 2R25NS090978-06. Dr. Xiang was supported by the National MS Society FG-1908-34882. Dr. Cross was supported by the Manny & Rosalyn Rosenthal – Dr John L. Trotter MS Center Chair in Neuroimmunology. Dr. Snyder was supported by P30NS098577 and P50HD103525. Dr. Benzinger was supported by 1R01AG054567-01A1, P01AG003991, and U19AG03243808.

References

- [1] BARANOVICOVA, E., MLYNARIK, V., KANTOROVA, E., HNILICOVA, P., AND DOBROTA, D. Quantitative evaluation of cerebral white matter in

- patients with multiple sclerosis using multicomponent T2 mapping. *Neurological Research* 38, 5 (2016), 389–396.
- [2] BERNSTEIN, M., KING, K., AND ZHOU, X. *Handbook of MRI Pulse Sequences*. Elsevier, 2004.
- [3] BRIER, M. R., SNYDER, A. Z., TANENBAUM, A., RUDICK, R. A., FISHER, E., JONES, S., SHIMONY, J. S., CROSS, A. H., BENZINGER, T. L., AND NAISMITH, R. T. Quantitative signal properties from standardized MRIs correlate with multiple sclerosis disability. *Annals of Clinical and Translational Neurology* 8, 5 (2021), 1096–1109.
- [4] COMMOWICK, O., ISTACE, A., KAIN, M., LAURENT, B., LERAY, F., SIMON, M., POP, S. C., GIRARD, P., AMÉLI, R., FERRÉ, J. C., KERBRAT, A., TOURDIAS, T., CERVENANSKY, F., GLATARD, T., BEAUMONT, J., DOYLE, S., FORBES, F., KNIGHT, J., KHADEMI, A., MAHBOD, A., WANG, C., MCKINLEY, R., WAGNER, F., MUSCHELLI, J., SWEENEY, E., ROURA, E., LLADÓ, X., SANTOS, M. M., SANTOS, W. P., SILVA-FILHO, A. G., TOMAS-FERNANDEZ, X., URIEN, H., BLOCH, I., VALVERDE, S., CABEZAS, M., VERA-OLMOS, F. J., MALPICA, N., GUTTMANN, C., VUKUSIC, S., EDAN, G., DOJAT, M., STYNER, M., WARFIELD, S. K., COTTON, F., AND BARILLOT, C. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. *Scientific Reports* 8, 1 (2018), 1–17.
- [5] DEBETTE, S., SCHILLING, S., DUPERRON, M. G., LARSSON, S. C., AND MARKUS, H. S. Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. *JAMA Neurology* 76, 1 (jan 2019), 81–94.
- [6] DORENT, R., BOOTH, T., LI, W., SUDRE, C. H., KAFIABADI, S., CARDOSO, J., OURSELIN, S., AND VERCAUTEREN, T. Learning joint segmentation of tissues and brain lesions from task-specific hetero-modal domain-shifted datasets. *Medical Image Analysis* 67 (jan 2021), 101862.

- [7] ELLIOTT, C., BELACHEW, S., WOLINSKY, J. S., HAUSER, S. L., KAPPOS, L., BARKHOF, F., BERNASCONI, C., FECKER, J., MODEL, F., WEI, W., AND ARNOLD, D. L. Chronic white matter lesion activity predicts clinical progression in primary progressive multiple sclerosis. *Brain* *142*, 9 (sep 2019), 2787–2799.
- [8] ELLIOTT, C., WOLINSKY, J. S., HAUSER, S. L., KAPPOS, L., BARKHOF, F., BERNASCONI, C., WEI, W., BELACHEW, S., AND ARNOLD, D. L. Slowly expanding/evolving lesions as a magnetic resonance imaging marker of chronic active multiple sclerosis lesions. *Multiple Sclerosis Journal* *25*, 14 (dec 2019), 1915–1925.
- [9] ERNST, R. R., AND ANDERSON, W. A. Application of fourier transform spectroscopy to magnetic resonance. *Review of Scientific Instruments* *37*, 1 (dec 1966), 93–102.
- [10] FILIPPI, M., AND AGOSTA, F. Imaging biomarkers in multiple sclerosis. *Journal of Magnetic Resonance Imaging* *31*, 4 (apr 2010), 770–788.
- [11] FISCHL, B. FreeSurfer, aug 2012.
- [12] GARCIA-LORENZO, D., FRANCIS, S., NARAYANAN, S., ARNOLD, D. L., COLLINS, D. L., GARCÍA-LORENZO, D., FRANCIS, S., NARAYANAN, S., ARNOLD, D. L., AND COLLINS, D. L. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical image analysis* *17*, 1 (jan 2013), 1–18.
- [13] GRISWOLD, M. A., JAKOB, P. M., HEIDEMANN, R. M., NITTKA, M., JELLUS, V., WANG, J., KIEFER, B., AND HAASE, A. Generalized Auto-calibrating Partially Parallel Acquisitions (GRAPPA). *Magnetic Resonance in Medicine* *47*, 6 (jun 2002), 1202–1210.
- [14] JENKINSON, M., BECKMANN, C. F., BEHRENS, T. E., WOOLRICH, M. W., AND SMITH, S. M. FSL. *NeuroImage* *62*, 2 (aug 2012), 782–790.

- [15] KHAN, W., KHLIF, M. S., MITO, R., DHOLLANDER, T., AND BRODTMANN, A. Investigating the microstructural properties of normal-appearing white matter (NAWM) preceding conversion to white matter hyperintensities (WMHs) in stroke survivors. *NeuroImage* 232 (may 2021), 117839.
- [16] KLISTORNER, A., WANG, C., YIANNIKAS, C., PARRATT, J., DWYER, M., BARTON, J., GRAHAM, S. L., YOU, Y., LIU, S., AND BARNETT, M. H. Evidence of progressive tissue loss in the core of chronic MS lesions: A longitudinal DTI study. *NeuroImage: Clinical* 17, December 2017 (2018), 1028–1035.
- [17] KULLBACK, S. *Information Theory and Statistics*. John Wiley and Sons, Ltd, 1959.
- [18] KURTZKE, J. F. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology* 33, 11 (nov 1983), 1444–1452.
- [19] LLADÓ, X., OLIVER, A., CABEZAS, M., FREIXENET, J., VILANOVA, J. C., QUILES, A., VALLS, L., RAMIÓ-TORRENT, L., AND ROVIRA, L. Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Information Sciences* 186, 1 (mar 2012), 164–185.
- [20] MACKAY, A. L., VAVASOUR, I. M., RAUSCHER, A., KOLIND, S. H., MÄDLER, B., MOORE, G. R., TRABOULSEE, A. L., LI, D. K., AND LAULE, C. MR Relaxation in Multiple Sclerosis. *Neuroimaging Clinics of North America* 19, 1 (2009), 1–26.
- [21] MANIEGA, S. M., VALDÉS HERNÁNDEZ, M. C., CLAYDEN, J. D., ROYLE, N. A., MURRAY, C., MORRIS, Z., ARIBISALA, B. S., GOW, A. J., STARR, J. M., BASTIN, M. E., DEARY, I. J., AND WARDLAW, J. M. White matter hyperintensities and normal-appearing white matter integrity in the aging brain. *Neurobiology of Aging* 36, 2 (feb 2015), 909–918.

- [22] MANTO, M. *The Neuroimaging of Brain Diseases. Structural and Functional Advances*. Springer, 2018.
- [23] MOLL, N. M., RIETSCH, A. M., THOMAS, S., RANSOHOFF, A. J., LEE, J. C., FOX, R., CHANG, A., RANSOHOFF, R. M., AND FISHER, E. Multiple sclerosis normal-appearing white matter: Pathology-imaging correlations. *Annals of Neurology* 70, 5 (2011), 764–773.
- [24] MOWRY, E. M., BERMEL, R. A., WILLIAMS, J. R., BENZINGER, T. L., DE MOOR, C., FISHER, E., HERSH, C. M., HYLAND, M. H., IZBUDAK, I., JONES, S. E., KIESEIER, B. C., KITZLER, H. H., KRUPP, L., LUI, Y. W., MONTALBAN, X., NAISMITH, R. T., NICHOLAS, J. A., PELLEGRINI, F., ROVIRA, A., SCHULZE, M., TACKENBERG, B., TINTORE, M., TIVARUS, M. E., ZIEMSEN, T., AND RUDICK, R. A. Harnessing Real-World Data to Inform Decision-Making: Multiple Sclerosis Partners Advancing Technology and Health Solutions (MS PATHS). *Frontiers in Neurology* 11 (aug 2020), 632.
- [25] NAISMITH, R. T., AND CROSS, A. H. Multiple sclerosis and black holes: Connecting the pixels. *Archives of Neurology* 62, 11 (nov 2005), 1666–1668.
- [26] PATENAUDE, B., SMITH, S. M., KENNEDY, D. N., AND JENKINSON, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* 56, 3 (jun 2011), 907–922.
- [27] PIKE, G. B., DE STEFANO, N., NARAYANAN, S., WORSLEY, K. J., PELLETIER, D., FRANCIS, G. S., ANTEL, J. P., AND ARNOLD, D. L. Multiple sclerosis: Magnetization transfer MR imaging of white matter before lesion appearance on T2-weighted images. *Radiology* 215, 3 (jun 2000), 824–830.
- [28] PLEWES, D. B., AND KUCHARCZYK, W. Physics of MRI: A primer. *Journal of Magnetic Resonance Imaging* 35, 5 (may 2012), 1038–1054.
- [29] POLMAN, C. H., REINGOLD, S. C., EDAN, G., FILIPPI, M., HARTUNG, H.-P., KAPPOS, L., LUBLIN, F. D., METZ, L. M., MCFARLAND, H. F.,

- O'CONNOR, P. W., SANDBERG-WOLLHEIM, M., THOMPSON, A. J., WEINSHENKER, B. G., AND WOLINSKY, J. S. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Annals of Neurology* 58, 6 (dec 2005), 840–846.
- [30] ROURA, E., OLIVER, A., CABEZAS, M., VALVERDE, S., PARETO, D., VILANOVA, J. C., RAMIÓ-TORRENTÀ, L., ROVIRA, À., AND LLADÓ, X. A toolbox for multiple sclerosis lesion segmentation. *Neuroradiology* 57, 10 (2015), 1031–1043.
- [31] SATI, P., CROSS, A. H., LUO, J., HILDEBOLT, C. F., AND YABLONSKIY, D. A. In vivo quantitative evaluation of brain tissue damage in multiple sclerosis using gradient echo plural contrast imaging technique. *NeuroImage* 51, 3 (jul 2010), 1089–1097.
- [32] SCHMIDT, P., GASER, C., ARSIC, M., BUCK, D., FÖRSCHLER, A., BERTHELE, A., HOSHI, M., ILG, R., SCHMID, V. J., ZIMMER, C., HEMMER, B., AND MÜHLAU, M. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage* 59, 4 (2012), 3774–3783.
- [33] SMITH, S. M. Fast robust automated brain extraction. *Human Brain Mapping* 17, 3 (nov 2002), 143–155.
- [34] VAN DER BOM, I. M. J., KLEIN, S., STARING, M., HOMAN, R., BARTELS, L. W., AND PLUIM, J. P. W. Evaluation of optimization methods for intensity-based 2D-3D registration in x-ray guided interventions. In *Medical Imaging 2011: Image Processing* (mar 2011), vol. 7962, SPIE, p. 796223.
- [35] WEN, J., GOYAL, M. S., ASTAFIEV, S. V., RAICHLE, M. E., AND YABLONSKIY, D. A. Genetically defined cellular correlates of the baseline brain MRI signal. *Proceedings of the National Academy of Sciences* 115, 41 (2018), E9727–E9736.

- [36] WEN, J., SUKSTANSKII, A. L., AND YABLONSKIY, D. A. Phase-sensitive B1 mapping: Effects of relaxation and RF spoiling. *Magnetic Resonance in Medicine* 80, 1 (jul 2018), 101–111.
- [37] ZHANG, Y., BRADY, M., AND SMITH, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* 20, 1 (jan 2001), 45–57.