A scalable method to improve gray matter segmentation at ultra high field MRI.

Omer Faruk Gulban[1]◕*, Marian Schneider[1]◕*, Ingo Marquardt[1], Roy A.M. Haast[1,2], Federico De Martino[1,3]

1 Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands.
2 Maastricht Centre for Systems Biology, Maastricht University, Maastricht, The Netherlands
3 Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, USA.

◕These authors contributed equally to this work.
* faruk.gulban@maastrichtuniversity.nl, marian.schneider@maastrichtuniversity.nl

Abstract

High-resolution (functional) magnetic resonance imaging (MRI) at ultra high magnetic fields (7 Tesla and above) enables researchers to study how anatomical and functional properties change within the cortical ribbon, along surfaces and across cortical depths. These studies require an accurate delineation of the gray matter ribbon, which often suffers from inclusion of blood vessels, dura mater and other non-brain tissue. Residual segmentation errors are commonly corrected by browsing the data slice-by-slice and manually changing labels. This task becomes increasingly laborious and prone to error at higher resolutions since both work and error scale with the number of voxels. Here we show that mislabeled, non-brain voxels can be corrected more efficiently and semi-automatically by representing three-dimensional anatomical images using two-dimensional histograms. We propose both a uni-modal (based on first spatial derivative) and multi-modal (based on compositional data analysis) approach to this representation and quantify the benefits in 7 Tesla MRI data of nine volunteers. We present an openly accessible Python implementation of these approaches and demonstrate that editing cortical segmentations using two-dimensional histogram representations aids existing algorithms and yields improved gray matter borders. By making our data and corresponding expert (ground truth) segmentations available, we facilitate future efforts to develop and test segmentation algorithms on this challenging type of data.

Introduction

Magnetic resonance imaging (MRI) has become one of the most important tools to study human brain function and structure in vivo. Moving from high (3 Tesla [T]) to ultra high (7 and 9.4 T) magnetic fields (UHF), together with improvements in acquisition methods, leads to increases in signal and contrast to noise (SNR and CNR, respectively) [1–3]. The increase in SNR can be leveraged to increase the voxels' resolution of both functional and structural images to sub-millimeter scales. Such sub-millimeter spatial resolutions allow for in vivo studies that probe cortical properties at the mesoscale [4–8]. These studies include (i) cortical-depth dependent analyses of

function [9–14] and structure [5, 8, 15], (ii) the mapping of cortical columnar structures [16–21] as well as (iii) sub-millimeter cortical topography [18, 22–24].

Such studies crucially depend on accurate and precise delineations of the gray matter (GM) ribbon both at the inner (white matter; WM) and outer (cerebrospinal; CSF) border. Since the aim of these studies is to investigate how the (functional) MRI (f/MRI) signal varies as a function of small position changes in GM, systematic GM segmentation errors would invalidate the conclusions drawn from these studies. Consider, as an example, an fMRI study conducted with a voxel resolution of 0.8 mm isotropic. Assume an average thickness of human cortex of 2.4 mm and a true signal change at the upper cortical depth level. In this study, the functional resolution would allow a straight piece of cortical ribbon to be divided in three relative cortical depths, each of them one voxel thick. Falsely labeling an additional fourth voxel as GM would make the difference between reporting an fMRI signal change at most superficial (false voxel excluded) or mid-superficial (false voxel included) cortical depth level. This example stresses the importance of accurate and precise GM segmentations.

Obtaining accurate and precise definitions of the GM ribbon, however, is currently a difficult and time-consuming task for sub-millimeter UHF data. The increases in SNR, CNR and resolution attainable in UHF anatomical data as well as analysis [25] and reconstruction [26] strategies specific to UHF reveal several structures outside of the brain that were barely visible on images obtained at conventional field strengths (1.5 and 3 T) and lower resolution ($>$ 1 mm isotropic) [27]. Such structures include the dura mater [28], medium-sized blood vessels in the sulci [29] as well as draining sinuses and connective tissue adjacent to GM [30]. To date, many of the automatic brain segmentation algorithms have been developed and benchmarked on images collected at 1 mm isotropic resolution or lower and at conventional field strengths [31] (but see [32]). If segmentation algorithms do not model these non-brain structures they might falsely label part of these structures as GM. Faced with such segmentation errors, researchers may choose to correct the misclassified voxels manually. However, the increase in resolution leads to an exponential increase in the number of voxels, which renders manual correction a laborious task. Furthermore, manual correction is prone to error and may introduce an observer bias, thereby reducing the reproducibility of subsequent analyses [33]. This currently leaves researchers with the dilemma of accepting the likely erroneous outcome of automatic segmentation algorithms or performing a time-consuming and error-prone manual correction.

CBS tools [32] directly tackle many of the challenges of UHF high-resolution anatomical data by, for example, including pre-processing steps to estimate dura mater and CSF partial voluming. Consequently, these tools provide an improved initial GM segmentation compared to other solutions [32]. However, we show here that in many cases the initial CBS segmentation can still be further improved with the approaches we propose. Furthermore, CBS tools have been optimized for whole-brain data obtained with the MP2RAGE sequence [26]. While the MP2RAGE sequence is commonly used at UHF as a basis for brain tissue class segmentations, we note that many high-resolution studies at UHF also use alternative sequences to define GM [12, 25, 34, 35], some of which offering partial coverage of the brain only [12, 35]. In such cases, alternative approaches that do not depend on particular templates, atlases or other forms of prior information are useful and required.

Here, we show that non-brain voxels misclassified as GM can be corrected using a multi-dimensional transfer function that is specified based on a two-dimensional (2D) histogram representation [36–39] of three-dimensional (3D) MRI brain data. We demonstrate that this transfer function offers an efficient way to single out non-brain tissue voxels. Removing these voxels from GM classifications found by automatic segmentation pipelines improves GM segmentations. This approach addresses the

problems of an entirely manual correction, since it yields a meaningful summary representation of the data that allows to manipulate the data efficiently. As a consequence, it is both more time efficient than manual slice-by-slice correction and it reduces observer bias.

We structured the paper as follows. In section 1, we introduce the technique of specifying transfer functions based on 2D histogram representations of voxel intensity and gradient magnitude. We offer theoretical considerations for why this technique is particularly suited to remove vessels and dura mater voxels in high-resolution MRI data (< 1 mm isotropic voxel size). In section 2, we extend the use of histogram-based transfer functions to multi-modal MRI datasets (e.g. T1 weighted [T1w], Proton Density weighted [PDw], T2* weighted [T2*w]) by considering MRI data in the compositional data analysis framework [40]. We show that this compositional framework yields an intuitive and useful summary representation of multi-modal MRI data which aids the creation of transfer functions. In Section 3 and 4 we validate the suggested methods by evaluating obtained GM segmentation results against expert GM segmentations obtained for nine subjects recorded at 7 T. We demonstrate considerable improvement in segmentation performance metrics for the two methods suggested here. We have implemented the methods described here in a free and open Python software package [41]. Furthermore, we make available the data [42] and processing scripts [43] used to validate the proposed methods.

1 Theory I: Transfer functions and 2D histograms

1.1 Multi-dimensional transfer functions

In the context of MRI data visualization, a transfer function can be understood as a mapping of voxel data to optical properties such as color and opacity. Effective transfer functions make structures of interest in the data visible. This can, for example, be achieved by assigning low opacity values to voxels that make up irrelevant structures and by highlighting desired structures with high opacity and salient color values. Multi-dimensional transfer functions assign renderable properties based on a combination of values [36–38, 44]. This is important in the context of MRI data where features of interest are often difficult to extract based on a single value alone. Considering multiple values, such as the intensity in images acquired using different contrast weighting (e.g. T1, PD and T2*), increases the chances of uniquely isolating a feature and making it visible [44].

In theory, multi-dimensional transfer functions could be used to perform exhaustive tissue-type segmentation of human brain MRI data. In this process, each voxel would be classified as either WM, GM, or CSF by specifying appropriate transfer functions. It has been shown, however, that this approach is less successful than established brain segmentation algorithms [45]. Here, we propose that transfer function-based methods still have a role to play in UHF MRI brain segmentation pipelines because they are well-suited for efficient removal of mislabeled non-brain tissue. We motivate this proposition by considering that brain and non-brain voxels become separable in 2D histogram representations.

1.2 2D histogram representation

2D histogram representations have been shown to greatly facilitate the process of specifying effective, 2D transfer functions [36, 38]. 2D histograms are obtained by taking an n-dimensional dataset and binning its data points along two dimensions. In principle, a 2D histogram can be obtained from any two sets of values. To identify tissue

boundaries a 2D histogram, plotting gradient magnitude against image intensity, has been shown to be particularly useful [36, 38]. The term gradient magnitude here refers to the magnitude of the vector that represents the spatial intensity gradient at every MRI voxel, where the spatial intensity gradient is equal to the first spatial derivative of the image intensity values.

Fig 1 shows how 3D MRI data of a human brain are represented in a 2D histogram (a T1w image was divided by a PDw image [25], multiplied by 500 and brain extracted; images were acquired with 0.7 mm isotropic resolution; for more details see Section 3). The histogram is obtained by plotting gradient magnitude against image intensity. In this representation, different tissue types occupy different regions. CSF voxels are characterized by very low intensity and low gradient magnitude values and therefore occupy the lower-left space of the histogram. GM and WM voxels have medium to high intensities and very low gradient magnitudes. Therefore, these tissue classes form circular regions at the bottom-center of the histogram. Voxels at the GM-WM interface fall within an arc reaching from medium to high intensities, following a low to medium to low gradient magnitude trajectory. Similarly, voxels at the CSF-GM interface span an arc from low to medium intensities. Finally, blood vessels and dura mater are thin structures characterized by very high gradient magnitudes and medium to high intensities. Therefore, these structures occupy up-center and the up-right parts of the 2D histogram.

Since different tissue types occupy different regions in the 2D histogram, each tissue type and boundary can, in principle, be isolated using a 2D transfer function based on image intensity and gradient magnitude. For the purposes of this paper, we focus on the distinction between brain (WM, GM, GM-WM interface) and non-brain (CSF, CSF-GM interface, blood vessels, dura mater) voxels. The intensity-gradient magnitude histogram is particularly suited to distinguish non-brain tissue at high resolutions, because dura mater and vessels do not span more than one or two voxels, given the typical voxel sizes of current high resolution studies (around 0.5 mm - 1 mm isotropic). Therefore, gradient magnitude will be high in the entirety of these structures (see Fig 1B for an example) and the combination of high intensity and high gradient magnitude values renders these structures separable from WM and GM voxels. To capture more extended structures or structures at even higher resolutions (in the future), we recommend the use of multi-modal data to distinguish data points by their compositional characteristics (see Section 2).

1.3 Creating transfer functions

The simplest way to create a transfer function is to explore the data by moving widgets with a specified shape over the 2D histogram representation [38]. For example, Fig 2 shows how a circular sector could be moved on top of the 2D histogram to highlight particular regions. In this case, only MRI voxels whose intensity-gradient magnitude combination falls within the highlighted region of the 2D histogram would be selected. Position and size of the circular sector can then be refined until the desired data has been isolated.

Using such a straightforward process of exploration and refinement [44], however, might yield slightly sub-optimal results. The shape of the widget might not capture the ideal shape given the data or the user might lack the prior knowledge that is required for this task. Alternatively, hierarchical exploration of normalized graph cut decision trees [39] can be used. This graph cut method results in a set of components (i.e. clusters) of the histogram that are mutually exclusive and collectively exhaustive. This allows the user to split and merge clusters in a data-driven and intuitive way that can be aided by the immediate visualization of the resulting segmentation (Fig 3). The method

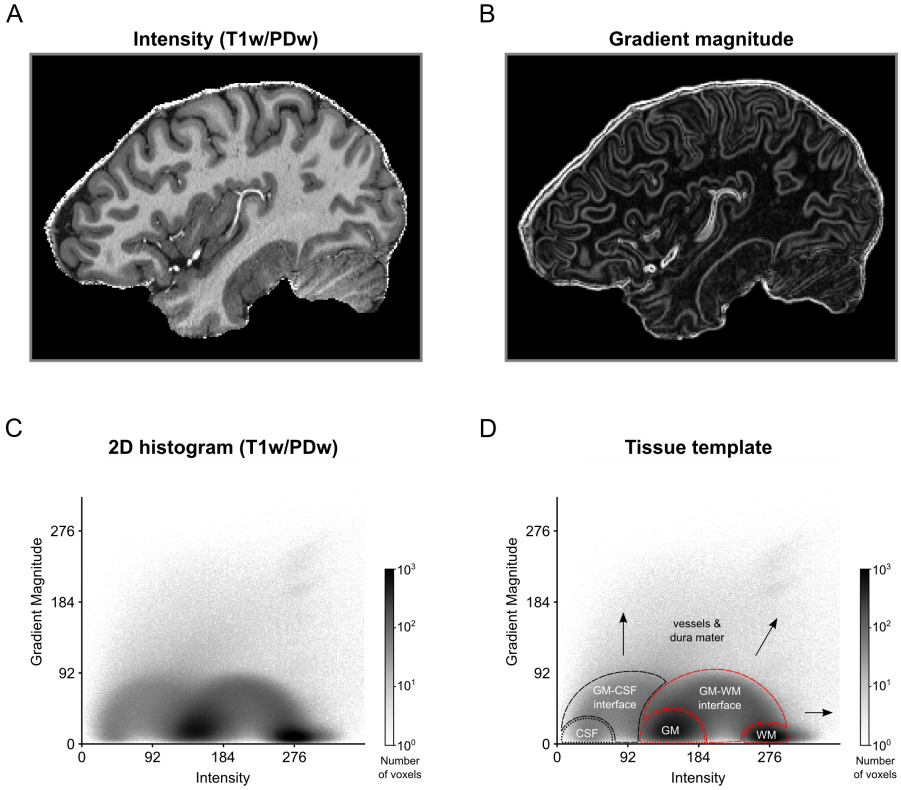


Fig 1. 2D histogram representation for MRI image of a human brain. (A) Intensity and (B) gradient magnitude values of a brain extracted T1w-divided-by-PDw MRI image are represented in a (C) 2D histogram. Darker regions in the histogram indicate that many voxels in the MRI image are characterized by this particular combination of image intensity and gradient magnitude. (D) The 2D histogram displays a characteristic pattern with tissue types occupying particular areas of the histogram. Voxels containing CSF, dura mater or blood vessels (black dashed lines and arrows) cover different regions of the histogram than voxels containing WM and GM (red dashed lines). As a result, brain tissue becomes separable from non-brain tissue.

allows for semi-automatic tissue selection, i.e. the shape of the clusters is data-driven but the decision which clusters to join and which to divide is made by the user.

2 Theory II: Multi-modal MRI data analysis

2.1 Compositional analysis for MRI data

More than one MRI contrast is often available and a combination of different contrasts can be useful in distinguishing different tissue types by differentially highlighting unique intrinsic properties. Two images with different contrast weighting can be combined using, for example, a ratio image [25, 46, 47]. This approach is beneficial for two reasons: 1) it reduces image biases as all acquisitions are affected by the same sensitivity profile of the receive elements in the radio frequency coil, and 2) if the images carry opposing

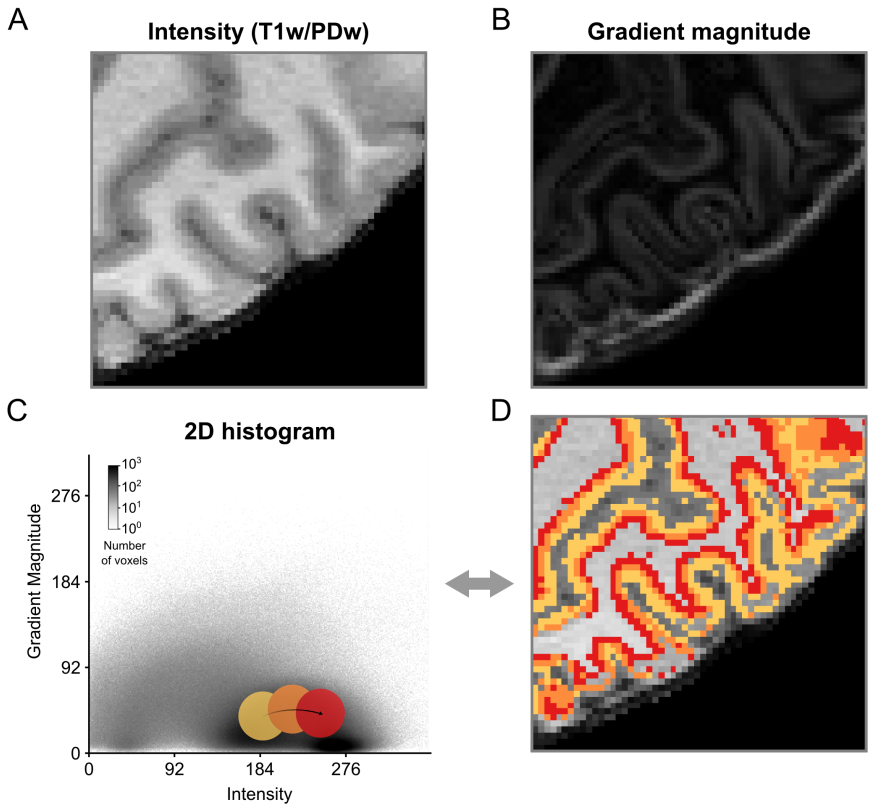


Fig 2. Creation of 2D transfer functions with pre-defined shapes. (A) Intensity and (B) gradient magnitude values of of a brain extracted T1w-divided-by-PDw MRI image are represented in a 2D histogram. By moving widgets of pre-defined shape, e.g. a circle, over the (C) 2D histogram and (D) concurrent visualization of selected voxels on a 2D slice of brain, positions of different tissue types in the 2D histogram can be probed and transfer functions can be created. In this example, the different probe positions (yellow, orange and red circles) appear to contain different aspects of GM.

contrast for the tissues of interest, the ratio increases contrast and benefits the delineation of the structures (tissue) of interest.

The ratio image approach, however, is limited to pairs of images. To operate on the relative information of more than two images, we propose to use the barycentric coordinate system [48]. In the barycentric coordinate system, coordinates of a point represent a simplex whose center of mass is determined by the weights at its vertices (the term *n-simplex* in geometry is the generalized form of the triangle [49]; for example the 0-simplex is a point, the 1-simplex is a line segment, the 2-simplex is a triangle, the 3-simplex is a tetrahedron and so on). In other words, points in the barycentric coordinate system represent compositions of non-negative fractions whose sum of components gives a constant value. When a set of measurements is represented in the barycentric coordinate system (that is when the sampling space is an n-simplex), the compositional data (CoDa) analysis methods [40, 50] can be used. The compositional data analysis offers a set of principled operations taking the geometry of the simplex sample space into account. Although MRI magnitude images are not strictly compositional (since they do not represent percentages), they still fulfill the technical

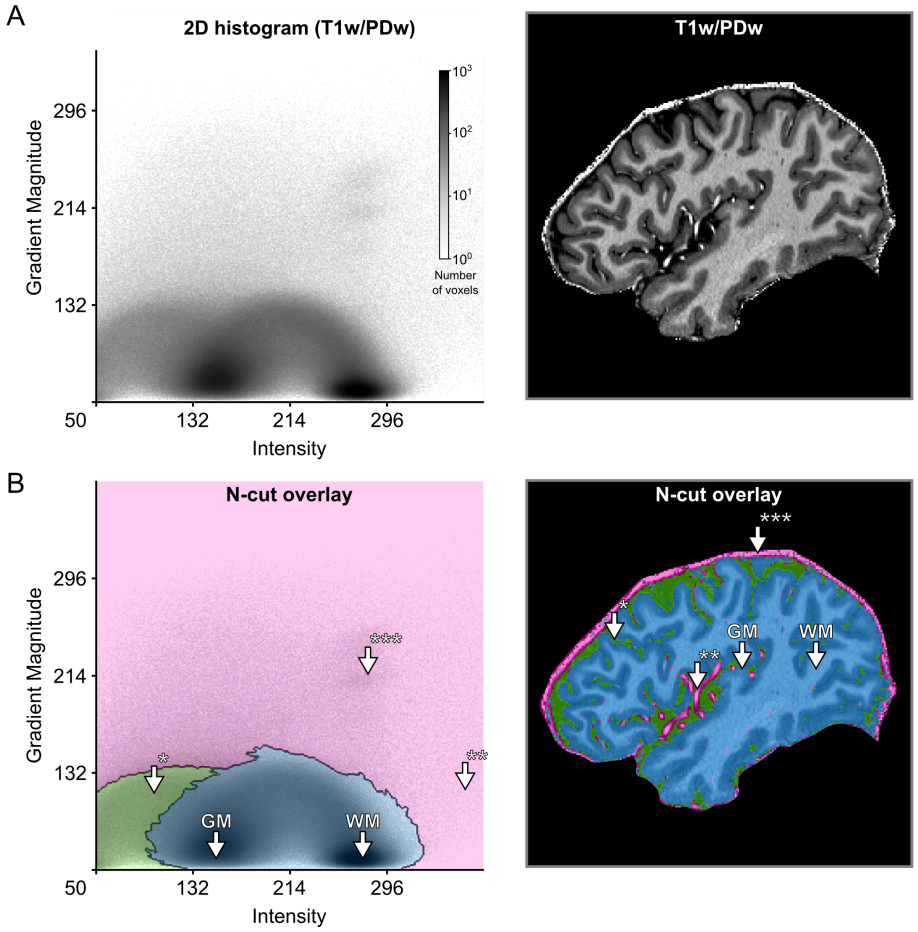


Fig 3. Creation of 2D transfer functions with data-driven shapes. (A) The user starts with the 2D histogram representation of image intensity and gradient magnitude (left side) and concurrent visualization of the original brain data (right side). The user can then interact with and select data in the 2D histogram to specify transfer functions. In this example, this was done with the help of a normalized graph cut decision tree. (B) The interaction with the 2D histogram results in data-driven shapes of selected areas, here shaded in pink, green and blue (left side). Voxels selected by those areas are highlighted in corresponding colors against the backdrop of the original brain data (right side). The visualization reveals that the area of the 2D histogram shaded in blue selects brain voxels, while the areas shaded in green and pink select CSF* and blood vessel voxels**/dura mater***, respectively.

requirements for treating each voxel as a composition. We therefore propose to process them in the CoDa framework, which allows us to exploit the principled operations of that framework. Although the general framework for CoDa analysis and its fundamental operations have already been rigorously documented in [40], for completeness we provide a step-by-step illustration of how multi-modal MRI data with three image contrasts (here T1w, PDw and T2*w magnitude images) can be processed under the compositional data analysis framework to acquire a useful representation of different tissue types.

Let multi-modal MRI data, V_{cMRI}, be defined as:

$$V_{cMRI} = [v_i, v_{i+1}, ..., v_{i+n}],\ \ v_i = [v_{iT1w}, v_{iPDw}, v_{iT2^*w}] \tag{1}$$

where v_i is the composition of voxel i. Let the composition v_i in this example consist of T1w, PDw and T2*w measurements.

The first step in compositional MRI data analysis is to convert the data components from Cartesian coordinates in real space (R^3) to barycentric coordinates in simplex space (S^3), using a voxel-wise *closure* operation:

$$\begin{aligned} X &= C(V_{cMRI}) = [C(v_i), C(v_{i+1}), ..., C(v_{i+n})], \\ C(v_i) &= k\left[\frac{v_{iT1w}}{I(v_i)}, \frac{v_{iPDw}}{I(v_i)}, \frac{v_{iT2^*w}}{I(v_i)}\right] \end{aligned} \tag{2}$$

where k is an arbitrary constant (here selected as 1) and $I(v_i)$ is the total amount of voxel intensity in a specific voxel:

$$I(v_i) = v_{iT1w} + v_{iPDw} + v_{iT2^*w} \tag{3}$$

As the next step, the barycentric coordinates of compositions (X) are centered by finding the *sample center* and *perturbing* each composition with the inverse of the sample center:

$$\begin{aligned} \mathsf{cen}(X) &= \hat{\mathbf{g}} = C[g_{T1w}, g_{PDw}, g_{T2^*w}], \\ \hat{X} &= X \oplus \hat{\mathbf{g}}^{\mathbf{-1}} \end{aligned} \tag{4}$$

where $\hat{\mathbf{g}}$ stands for the component-wise geometric mean across all voxels

$$\hat{g}_i = \left(\prod_{i=1}^{n} x_{ij}\right)^{1/n}, j = [T1w, PDw, T2^*w] \tag{5}$$

and $\oplus$ denotes the perturbation operation defined in multi-dimensional simplex space (S^D), which can be considered as an analogue of addition in real space:

$$x \oplus y = C[x_1y_1, x_2y_2, ..., x_Dy_D] \in S^D \tag{6}$$

In the 3D case of (T1w, PDw, T2*w) we obtain:

$$x \oplus y = C[x_{T1w}y_1, x_{PDw}y_2, x_{T2*w}y_3] \in S^3 \tag{7}$$

After centering, the data is standardized:

$$\hat{\hat{X}} = \hat{X} \odot \mathsf{totvar}[X]^{-1/2} \tag{8}$$

where the total variance is computed by

$$\mathsf{totvar}[X] = \frac{1}{n}\sum_{i=1}^{n} d_a^2(x_i, \hat{\mathbf{g}}) \tag{9}$$

and the $\odot$ symbol stands for the power operation defined in simplex space, which can be considered as an analogue of multiplication in real space:

$$x \odot a = C[x_1^a, x_2^a, ..., x_D^a] \tag{10}$$

After standardization, the barycentric coordinates are transformed from the 3D simplex space (S^3) to 2D real space (R^2) by using the isometric logratio (ilr) transformation [51]:

$$ilr(x_i) = \ln(x_i) \cdot H \tag{11}$$

where H indicates a Helmert matrix [52] of 3 rows and 2 columns. The purpose of the transformation is to visualize the distribution in a 2D histogram.

Note that the closure operation described in eq. (2) implies scale invariance. If the receive (and in some cases transmit) field (B1) inhomogeneities for MRI data are similar across modalities and assumed to be having a multiplicative effect on the measured signal, applying closure will mitigate inhomogeneities by canceling out the common multiplicative term (ie. bias field) in each image modality. For instance, assume two voxels contain the same tissue type but have dissimilar intensities due to a multiplicative effect. If before the closure operation voxel 1 has an intensity of 100 in all recorded modalities and voxel 2 has an intensity of 500 in all modalities, then after the closure operation both voxels will have the same compositional description, which would be desired. It should be noted that if B1 inhomogeneities differ significantly across modalities, the closure operation will yield inaccurate compositional descriptions. In this case, we recommend to use bias field correction algorithms before using the compositional data analysis framework. A practical example for this case is that for magnetization-prepared rapid acquisition gradient-echo (MPRAGE) sequences, the transmit field in T1w image is effected by an inversion pulse which is not present in PDw and T2*w images. In such cases, individual image bias field correction is recommended.

2.2 2D histogram representation and creation of transfer functions

Fig 4 shows how three different 3D MRI contrast images of a human brain (T1w, PDw and T2*w brain extracted images; 0.7 mm isotropic resolution; for more details see Section 3) can be represented in a 2D histogram. The 2D histogram is obtained by taking the three MRI contrast images as an input and performing the operations of the CoDa analysis framework described above. In particular, applying the ilr transformation to the barycentric coordinates allows the three images to be represented along two dimensions. Different tissue types have different compositional characteristics and therefore occupy different regions in the resulting 2D histogram. WM and GM voxels are separated in two distinct clusters which mainly differ along the T1w axis. CSF voxels occupy the lower left corner of the histogram, which represents a combination of low T1w with high PDw and T2*w values. CSF voxels still differ from WM and GM voxels mainly along the T1w axis. In contrast, vessel and dura mater voxels differ from WM, GM and CSF voxels also along the PDw and T2*w axes, which makes these voxels to be spread out in the direction orthogonal to the T1w axis. To see how a combination of two MP2RAGE images (UNI, INV2) and one T2* image estimated from a multi-echo 3D gradient recalled echo (GRE) sequence are represented in a 2D histogram, please see S1 Fig.

The dimensionality reduction accomplished by the ilr transformation allows to specify 2D transfer functions even though the input consists of three channels. Fig 5 shows how normalized graph cuts can be used on 2D histogram representation of ilr coordinates to create transfer functions. The resulting transfer functions highlight specific clusters that readily separate brain tissue from non-brain tissue.

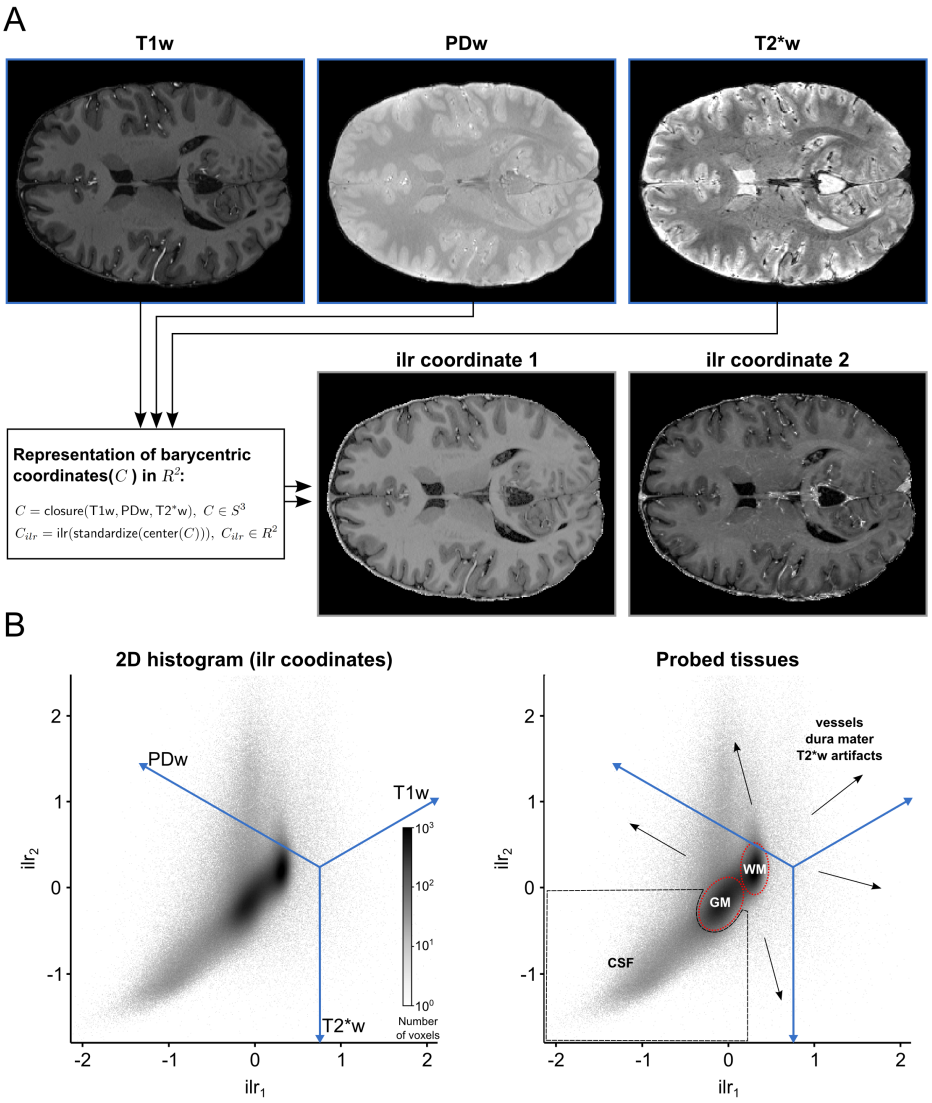


Fig 4. 2D histogram representation of three 3D MRI contrast images. (A) Each voxel is considered as a 3 part composition in 3D real space. The barycentric coordinates of each composition which reside in 3D simplex space are represented in 2D real space after using a isometric log-ratio (ilr) transformation. (B) The ilr coordinates are used to create 2D histograms representing all voxels in the images. The blue lines are the embedded 3D real space primary axes. It should be noted that in this case the ilr coordinates are not easily interpretable by themselves but they are useful to visualize the barycentric coordinates which are interpretable via the embedded real space primary axes. Darker regions in the histogram indicate that many voxels are characterized by this particular scale invariant combination of the image contrasts. In this representation, brain tissue (WM and GM, red dashed lines) becomes separable from non-brain tissue (black dashed lines and arrows).

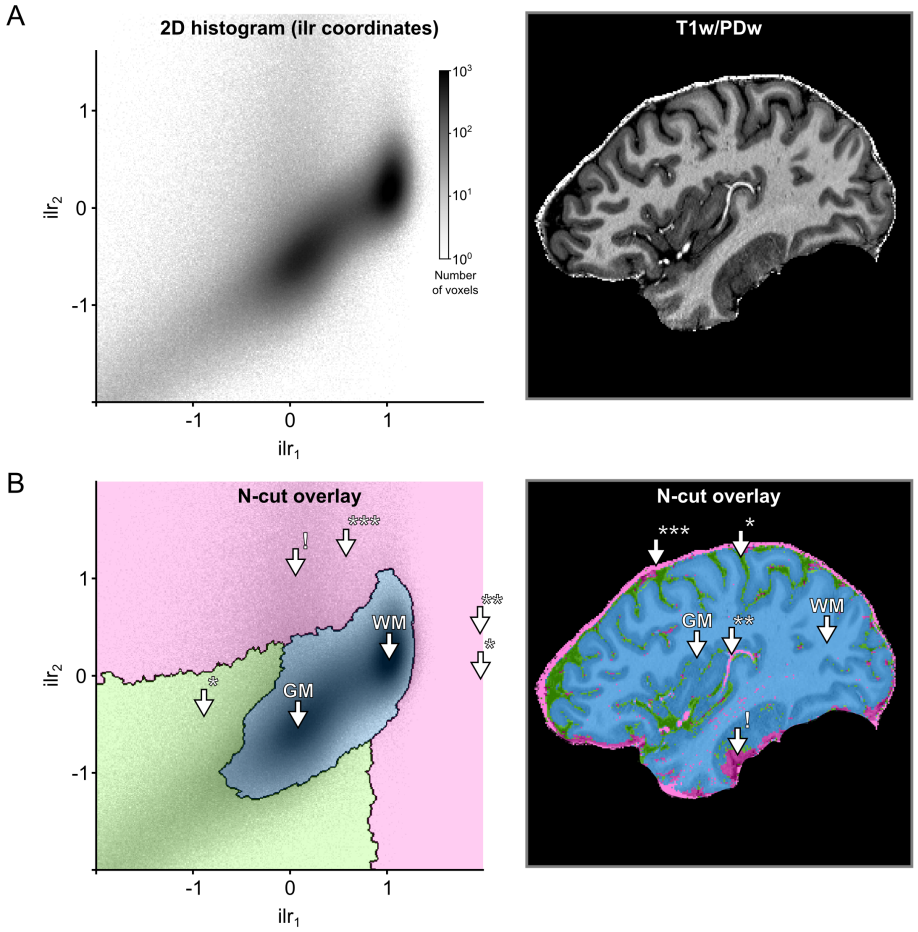


Fig 5. Creation of transfer functions using ilr coordinates. (A) The user starts with the 2D histogram representation of ilr coordinates 1 and 2 (left side) and concurrent visualization of the original brain data (right side). The user can then interact with and select data as described in Fig3. (B) The interaction with the 2D histogram results in data-driven shapes, here shaded in pink, green and blue (left side). Voxels selected by those areas are highlighted in corresponding colors against the backdrop of the original brain data (right side). The visualization reveals that the area of the 2D histogram shaded in blue selects brain voxels, while the areas shaded in green and pink select CSF* and blood vessel voxels**/dura mater***, respectively. The arrow with exclamation mark (!) indicates an area affected by T2*w image artifacts.

3 Validation methods

3.1 Validation data set overview

In order to validate the methods proposed above, we created two validation data sets based on the acquisition of high-resolution 7 T data of nine subjects and corresponding manually-guided expert segmentations of GM. In particular, we created two validation sets based of on two of the most common acquisition sequences. For five subjects, we collected MPRAGE T1w, PDw, and T2* data (we refer to this data set as the MPRAGE data set below). For four different subjects, we collected MP2RAGE data, to obtain unbiased (uni) images, and multi-echo 3D GRE data, to obtain T2* maps (we refer to this data set as the MP2RAGE data set below). The MPRAGE data set can be downloaded from [42]. The MPR2AGE data can be obtained by contacting the corresponding authors.

3.1.1 Ethics statement

The experimental procedures were approved by the ethics committee of the Faculty for Psychology and Neuroscience (MPRAGE data set) or the Medical Ethical Committee at Faculty of Health, Medicine and Life Sciences (MP2RAGE data set) at Maastricht University, and were performed in accordance with the approved guidelines and the Declaration of Helsinki. Informed consent was obtained from each participant before conducting the experiments.

3.1.2 MRI acquisition parameters

All images were acquired on a Siemens 7 T whole body scanner (Siemens Medical Solutions, Erlangen, Germany) using a head radio-frequency (RF) coil (Nova Medical, Wilmington, MA, USA; single transmit, 32 receive channels). In all acquisitions, we used dielectric pads [53].

For n=5 subjects (age range 24-30, 2 females, no medical condition), the MPRAGE data set consisted of: a T1w image using a 3D MPRAGE sequence (repetition time [TR] = 3100 ms; time to inversion [TI] = 1500 ms [adiabatic non-selective inversion pulse]; time echo [TE] = 2.42 ms; flip angle = 5°; generalized auto-calibrating partially parallel acquisitions [GRAPPA] = 3 [54]; field of view [FOV] = 224 × 224 mm2; matrix size = 320 × 320; 256 slices; 0.7 mm isotropic voxels; pixel bandwidth = 182 Hz/pixel; first phase encode direction anterior to posterior; second phase encode direction left to right), a PDw image (0.7 mm isotropic) with the same MPRAGE as for the T1w image but without the inversion pulse (TR = 1380 ms; TE = 2.42 ms; flip angle = 5°; GRAPPA = 3; FOV = 224 × 224 mm; matrix size = 320 × 320; 256 slices; 0.7 mm iso. voxels; pixel bandwidth = 182 Hz/pixel; first phase encode direction anterior to posterior; second phase encode direction left to right), and a T2*w anatomical image using a modified MPRAGE sequence that allows freely setting the TE (TR = 4910 ms; TE = 16 ms; flip angle = 5°; GRAPPA= 3; FOV = 224 × 224 mm; matrix size = 320 × 320; 256 slices; 0.7 mm iso. voxels; pixel bandwidth = 473 Hz/pixel; first phase encode direction anterior to posterior; second phase encode direction left to right).

For n=4 subjects (age range 24-58, 2 female, no medical condition) the MP2RAGE data set consisted of: 3D MP2RAGE data (TR = 5000 ms; TI1/TI2 = 900/2750 ms; TE = 2.46 ms; FA1/FA2 = 5°/3°; FOV = 224×224 mm2; matrix size = 320 × 320; slices = 240; 0.7 mm iso voxels) [26]. For the same subjects, T2*w images were obtained with a multi-echo 3D GRE sequence (TR = 33 ms; TE1/TE2/TE3/TE4 = 2.53/7.03/12.55/20.35 ms; FA1 = 11°; FOV = 224 x 159 mm2; matrix = 320 x 227; slices = 208; 0.7 mm iso voxels). More details on the MP2RAGE data acquisition and the T2* estimation can be found in [55].

3.1.3 Manually-guided expert segmentations

For every subject, we established "ground truth" GM classifications via manually-guided expert segmentations. All segmentations were created manually by the same expert (OFG), using ITK-SNAP [56] and a graphics tablet (Intuos Art; Wacom Co. Ltd; Kazo, Saitama, Japan). To avoid resulting tissue type classification to be ragged, the expert followed a particular processing sequence. The brain was first traversed in a single direction (e.g. sagittally) and the ground truth was established slice-by-slice. Subsequently, the brain was traversed in the two other directions (e.g. axially, then coronally). This sequence was repeated several times across several regions until the GM segmentation of the whole brain was considered of good quality.

3.2 Software implementation

We implemented the creation of transfer function based on 2D histograms in an open source Python package called Segmentator [41], which is built upon several other scientific packages such as Numpy [57], Scipy [58], Matplotlib [59] and Nibabel [60]. Segmentator allows for selection of data points in a 2D histogram (for example gradient magnitude over intensity) and concurrent visualization of selected brain voxels on a 2D slice. Data points can be selected using a circular sector widget with variable reflex angle and radius. Alternatively, data selection can be performed using the normalized graph cut (n-cut) method (i.e. spectral clustering) as described above. The n-cut algorithm from Scikit-image [61] was modified to export an augmented output which provides step-wise access to independent branches of the decision tree and employed in Segmentator (the modification is available at: https://github.com/ofgulban/scikit-image/tree/ncut-rag-options).

The package provides several options to calculate the gradient magnitude image. All the 2D histogram analyses described in this paper were based on gradient magnitude images that were computed as the Euclidean norm of the first spatial derivative estimated using a $3 \times 3 \times 3$ Scharr kernel [62, 63]. Subsequently, transfer functions were specified using the normalized graph cut algorithm and user intervention for the selection of the non-brain tissue transfer functions. Processing data for a single subject took about 10 minutes on average. The Segmentator package is openly and freely accessible at https://github.com/ofgulban/segmentator. All the operations of the CoDa analysis described above have been implemented as a separate open source Python package [64] freely accessible at https://github.com/ofgulban/tetrahydra. This package uses Numpy [57] and Scipy [58].

3.3 Segmentation procedure

For both validation data sets, we followed similar procedures, with modifications where necessary to accommodate for differences in the sequences' output. Our goal was to obtain initial GM segmentations from established segmentation algorithms and to quantify the improvement in segmentation accuracy that can be obtained when using the methods described here as post-processing steps. To establish the initial GM segmentations we used FSL FAST [65] and the SPM 12 "unified segmentation" algorithm [66] for the MPRAGE data set and FSL FAST and CBS tools [32] for the MP2RAGE data set. SPM and CBS tools have been developed and benchmarked on MPRAGE and MP2RAGE images respectively. FSL FAST is suited to process either type, so we used it for both data sets. We then quantified the impact of the following additional post-processing steps: (i) using uni-modal input and transfer functions based on 2D histogram representations of intensity and gradient magnitude (see Section 1) or (ii) using multi-modal input and the compositional data analysis framework (see Section

2). These two procedures will be referred to below as the gradient magnitude (GraMag) and the compositional data analysis (CoDa) method, respectively. Both methods resulted in masks that could be used to further refine the initial GM segmentation, e.g. by removing blood vessels and dura mater that were falsely labeled as GM initially. In total, we thus used 2 (MPRAGE and MP2RAGE data set) x 2 (GraMag and CoDa) = 4 analysis procedures. All four procedures are summarized in flow chart diagrams (S2 Fig, S3 Fig, S4 Fig, S5 Fig). Furthermore, in an effort to make our analyses fully reproducible, we made the Python and bash scripts used for pipeline processing openly available at [43].

For the MPRAGE dataset, we first computed ratio images (T1w divided by PDw) [25] to reduce inhomogeneities. Ratio images were input to either FSL FAST or SPM 12. FSL FAST was used with default values. The FAST algorithm requires an initial brain extraction procedure that we performed using FSL BET [67]. Additionally, we masked the images to exclude cerebellum and brain stem tissue with masks that were created manually for all subjects (below referred to as "nosub mask"). In SPM 12 we used default settings with one exception. We set the number of Gaussians to be modeled to 3 for GM and 2 for WM (default values are 1 and 1). As part of their standard segmentation routine, both FSL FAST and SPM 12 perform initial inhomogeneity correction. We output and inspected the bias corrected images to ensure that the algorithms had converged on plausible solutions. We specified for the FSL FAST algorithm to output hard segmentation labels. Since SPM 12 outputs probabilities for six tissue classes, we transformed this soft output to hard segmentation labels by assigning each voxel to the tissue class with the highest posterior probability. Since the SPM segmentation algorithm works best with unmasked images, we applied the nosub mask only to the resulting SPM GM segmentations, not to the input data. The resulting GM segmentations from FSL and SPM were saved for later evaluation.

For the GraMag method (S2 Fig) we proceeded with bias-corrected ratio images from either SPM or FSL. Since the GraMag method works best with brain extracted images, we combined SPM's WM and GM segmentation outcomes to form a brain mask and performed brain extraction of the ratio images from SPM. After brain extraction, we also excluded cerebellum and brain stem tissue using the nosub mask. FSL's bias-corrected ratio images images did not require masking as the brain extraction (and cerebellum removal) was already performed before segmentation. We then used the 2D histogram representation of intensity and gradient magnitude together with the hierarchical exploration of normalized graph cut decision trees (as described in Section 1) to create transfer functions. The resulting transfer function was used to separate brain from non-brain tissue voxels. Non-brain tissue voxels were removed from GM if they were included in the initial FSL and SPM segmentations.

For the CoDa method (S3 Fig) we followed a similar procedure, except that we started from three separate images - the bias-corrected T1w, PDw and T2*w images. Again, these images were brain extracted and cerebellum and brain stem tissue were removed using the nosub mask. These images were transformed into barycentric coordinates, using the closure operator (as outlined in in Section 2). In this case, there were three barycentric coordinates per voxel constrained to a 2-simplex vector space structure. The triplets of barycentric coordinates were mapped to 2D real-space using the *ilr* transformation. We could therefore proceed with the 2D histogram representation using the first and the second real-space coordinates of the compositions and the hierarchical exploration of normalized graph cut decision trees in this 2D space to separate brain from non-brain tissue voxels. Non-brain tissue voxels were again removed from GM if included in the initial segmentations (of SPM and FSL).

For the MP2RAGE dataset, the T1 map, T1w (uni) and second inversion image from the MP2RAGE sequence were input to CBS tools [32]. Only the

brain-extracted [67] uni image was input to FSL FAST, since this resulted in higher performance than inputting all three images. Both FSL FAST and CBS tools were run with default settings. Note that the default settings for CBS tools include removal of non brain tissue by estimating dura mater and CSF partial voluming. The resulting GM segmentations from FSL and CBS were saved for later evaluation. For the GraMag method (S4 Fig), we proceeded with the FSL FAST bias-corrected, brain-extracted and nosub masked uni image and proceeded as for the MPRAGE data set to obtain a secondary brain mask. For the CoDa method (S5 Fig), we used FSL FAST bias-corrected, brain-extracted and nosub masked uni, second inversion and T2* images but otherwise proceeded as for the MPRAGE data set.

We observed susceptibility artifacts in some regions of the brain (mostly inferior frontal lobe) in the T2*w images. These artifacts make the affected regions noisy and reduce the effectiveness of using T2*w images in the CoDa method. To quantify the effect of these artifacts, we created masks for the artifact-affected regions and ran all our analyses both with and without the artifact regions included. Results shown in the paper were obtained with the affected regions excluded. Results with the affected regions included are shown in the Supplementary Materials.

3.4 Quantification

The segmentation procedures resulted in three different GM segmentations for each data set and initial segmentation algorithm (SPM or CBS and FSL FAST): (i) an initial segmentation without any further changes, (ii) after correction using the GraMag method and (iii) after correction using the CoDa method. To compare segmentation quality among these three outcomes, we calculated the Dice coefficient (DICE) and the Average Hausdorff Distance (AVHD) using the openly available EvaluateSegmentation Tool (2016; VISCERAL, http://www.visceral.eu).

The DICE is an overlap-based metric and it is the most popular choices for validating volume segmentations [68]. We included it here as a familiar reference for the reader. However, overlap-based metrics like the DICE are not recommended for validating segmentation boundaries against the ground truth, as is our aim here, since they are relatively insensitive to boundary errors. In contrast, the AVHD is a distance-based metric and is sensitive to boundary errors [68]. We therefore consider the AVHD to be a more suitable metric for our purposes and we based our conclusions on the comparisons made with the AVHD.

Given that the AVHD quantifies the similarity of two boundaries, we first extracted WM-GM and GM-CSF boundaries from the ground truth segmentations and the six different GM segmentations before calculating the AVHD. Here, an AVHD of 0 indicates a perfect match between the segmentation and ground truth boundaries, while values >0 indicate a mismatch. In this case, the value represents the average number of voxels by which the two boundaries deviate from one another. For example, an AVHD of 1 indicates that the segmentation boundary, on average, deviates of one voxel from the ground truth.

4 Validation results

Visual inspection revealed that applying the GraMag method to the MPRAGE data set excluded most of the vessels and dura mater voxels and resulted in a more plausible GM matter definition. The CoDa method equally removed most of the vessels and dura mater voxels. Additionally, the CoDa method excluded structures like the sagittal sinus from the GM definition (see Fig 6 and Fig 7).

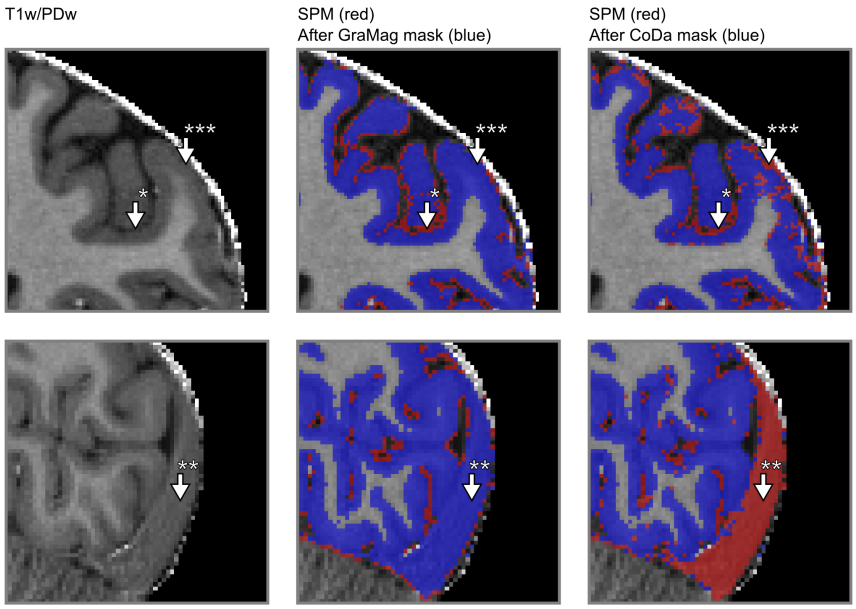


Fig 6. Comparison of GM segmentation results for MPRAGE data. GM segmentation results are shown for one representative subject on a transverse (upper row) and a sagittal slice (lower row) of the brain before and after applying the GraMag and CoDa methods. The original image that is input to the segmentation is shown on the left. The original GM segmentation obtained from SPM 12 is shown in red (middle and right column). GM segmentations after additional polishing with brain mask obtained with either the GraMag (middle column) or the CoDa method (right column) are overlaid in blue. Additional masking removes blood vessels, CSF (arrow *) and most of dura mater (arrow †) voxels from the SPM GM definition. Because of its unique compositional properties, connective tissue from the sagittal sinus can be captured and excluded using the CoDa method (arrow **). An area badly affected by the CoDa mask is also indicated with arrow ***.

Table 1 compares segmentation performance before and after applying GraMag and CoDa methods to the initial GM segmentations of the MPRAGE dataset. The GraMag method led to an improvement of GM segmentations in all subjects, independently of whether the initial segmentation was done by SPM 12 or FSL FAST. On average, the AVHD decreased from 0.729 ±0.087 (mean ±standard deviation across subjects) to 0.547 ±0.051 for SPM 12 and from 0.531 ±0.109 to 0.472 ±0.089 for FSL FAST. The GraMag method also increased the DICE coefficient in all subjects individually and, on average, from 0.861 ±0.020 to 0.882 ±0.016 for SPM 12 and from 0.892 ±0.027 to 0.900 ±0.089 for FSL FAST. The CoDa method equally yielded improved segmentation performance. Compared to the initial segmentation, the AVHD decreased in all subjects and, on average, to 0.542 ±0.054 for SPM 12 and to 0.479 ±0.033 for FSL FAST. The DICE coefficent increased for SPM 12 segmentations to 0.871 ±0.021 but it decreased for FSL FAST segmentations to 0.879 ±0.013. All these results were obtained after exclusion of areas affected by artifacts in the T2s image. For results obtained without the artifact masks, please see S1 Table.

Table 2 compares segmentation performances before and after applying the GraMag and CoDa methods to the initial GM segmentations of the MP2RAGE dataset. The

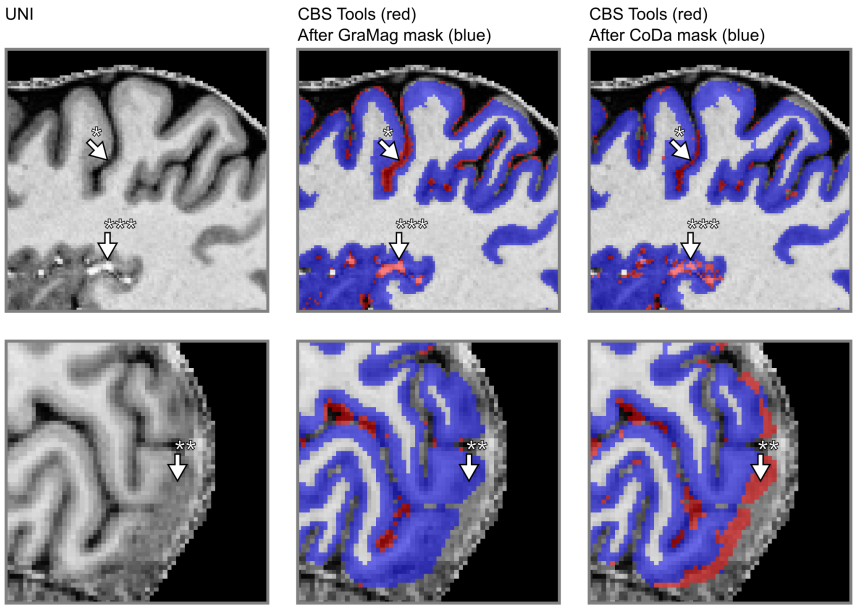


Fig 7. Comparison of GM segmentation results for MP2RAGE data. Same conventions as in Fig 6) but with initial segmentation results obtained with CBS tools instead of SPM 12.

GraMag method decreased AVHD for every subject and, on average, from 0.413 ±0.088 to 0.347 ±0.083 for CBS tools and from 0.988 ±0.062 to 0.790 ±0.088 for FSL FAST. It decreased the DICE coefficient, on average, from 0.907 ±to 0.898 ±for CBS tools and from 0.831 ±to 0.806 ±for FSL FAST. The CoDa method decreased the AVDH, on average, to 0.379 ±0.082 for CBS tools and to 0.624 ±0.069 for FSL FAST. It also increased the DICE coefficient to 0.850 ±0.030 for FSL FAST but decreased it to 0.895 ±0.024 for CBS tools. For results for the MP2RAGE data obtained without the artifact masks, please see S2 Table.

5 Discussion

Functional and anatomical MRI studies at the mesoscale (< 1 mm isotropic) require accurate and precise definitions of the GM ribbon. Creating such definitions is currently a challenging task since sub-millimeter UHF data bring non-brain structures like blood vessels and dura mater into sharper focus. As a result, segmentation algorithms that have been benchmarked at lower resolution data might falsely label part of these structures as GM. Here we presented two methods (GraMag and CoDa) to correct such mislabeled non-brain voxels efficiently and semi-automatically. The two methods are based on theoretical expectations of how 3D brain data is to be represented in 2D histograms. These expectations imply that brain and non-brain tissue should become separable in 2D histogram representations that are either based on gradient magnitude and intensity or on compositional dimensions. We validated these expectations by implementing the suggested methods in an openly available software package and by quantifying their benefit using a new high-resolution validation data set. The suggested methods improved the initial GM segmentations in general. However, we found some

Table 1. Segmentation performance scores MPRAGE data set. The table shows the DICE (larger is better) and AVHD (less is better) for the initial SPM 12 and FSL FAST GM segmentations as well as after additional polishing, using either the gradient magnitude or the compositional data method.

		SPM		FAST	
		DICE[a]	AVHD[a]	DICE[a]	AVHD[a]
S02					
	Init	0.8608	0.7274	0.8870	0.5659
	Init + GraMag	0.8771	0.5466	0.8914	0.5079
	Init + CoDa	0.8620	0.5524	0.8715	0.5040
S03					
	Init	0.8409	0.8496	0.8782	0.5835
	Init + GraMag	0.8677	0.6084	0.8918	0.5043
	Init + CoDa.	0.8534	0.6131	0.8724	0.5038
S05					
	Init	0.8687	0.6320	0.8858	0.5545
	Init + GraMag	0.8902	0.4675	0.8947	0.4826
	Init + CoDa	0.8768	0.4873	0.8928	0.4690
S06					
	Init	0.8446	0.7716	0.8712	0.6098
	Init + GraMag	0.8694	0.5612	0.8778	0.5464
	Init + CoDa.	0.8563	0.5680	0.8670	0.4929
S07					
	Init	0.8914	0.6630	0.9402	0.3395
	Init + GraMag	0.9048	0.5495	0.9432	0.3186
	Init + CoDa	0.9045	0.4882	0.8933	0.4254

DICE, DICE Coefficient; AVHD, Average Hausdorff Distance; Init, initial segmentation; GraMag, gradient magnitude method; CoDa, compositional data method.
[a] After masking of areas affected by artifact in the T2s image.

differences in the degree of improvement with respect to (i) the two presented methods, the (ii) type of data and (iii) the algorithm used for initial segmentation.

We will discuss these three influences in turn. First, the two methods differ in their prerequisites and their segmentation improvement. The GraMag method only requires uni-modal input such as T1w/PDw or MP2RAGE uni images, while the CoDa method requires multi-modal input of images with different contrast weightings. This makes the GraMag method the method of choice when only a single input image is available. In accordance with our theoretical expectation, the GraMag method identified and removed blood vessels and dura mater tissue. If multi-dimensional input is available, even bigger improvements might be obtained with the CoDa method. Notably, in contrast to the GraMag method, the CoDa method can additionally capture and remove connected tissue of the sagittal sinus. This tissue is usually falsely labeled as GM because of similar intensity values and spatial proximity. It then requires tedious manual removal. How well the CoDa method performs, however, critically depends on the quality of all the input images and the specific combination of contrasts. Performance can be affected by low quality on a single input image, as was the case here with T2* images due to susceptibility artifacts. Furthermore, performance will depend on the specific choice of contrasts and whether these contrasts maximize the compositional difference between brain and non-brain tissue.

Second, we found that the improvements were slightly larger and more consistent

Table 2. Segmentation performance scores MP2RAGE data set. The table shows the DICE (larger is better) and AVHD (less is better) for the initial CBS tools and FSL FAST GM segmentations as well as after additional masking, using either the gradient magnitude or the compositional data method.

		CBS		FAST	
		DICE[a]	AVHD[a]	DICE[a]	AVHD[a]
S001					
	Init	0.9313	0.2979	0.8491	1.0524
	Init + GraMag	0.9283	0.2657	0.8247	0.8302
	Init + CoDa	0.9247	0.2825	0.8751	0.5634
S013					
	Init	0.8849	0.5025	0.7856	1.0266
	Init + GraMag	0.8780	0.4386	0.7652	0.8951
	Init + CoDa	0.8720	0.4607	0.8083	0.7222
S014					
	Init	0.8934	0.4553	0.8396	0.9210
	Init + GraMag	0.8872	0.3954	0.8136	0.7217
	Init + CoDa	0.8800	0.4320	0.8508	0.6164
S019					
	Init	0.9169	0.3950	0.8491	0.9520
	Init + GraMag	0.8989	0.2888	0.8219	0.7121
	Init + CoDa	0.9051	0.3411	0.8670	0.5928

DICE, DICE Coefficient; AVHD, Average Hausdorff Distance; Init, initial segmentation; GraMag, gradient magnitude method; CoDa, compositional data method.
[a] After masking of areas affected by artifact in the T2s image.

across subjects for the MPRAGE than for the MP2RAGE data set. This might be explained by the fact that the MPRAGE data conformed more to our theoretical expectations than the MP2RAGE data set. Especially, we found GM values in the MP2RAGE uni image to be less focused on one particular area of the 2D histogram (S6 Fig) than the MPRAGE division image. This might result from differences in myelination level across cortical areas and depth [47, 69, 70], which the MP2RAGE uni image might pick up more than MPRAGE division image [71].

Third, we observed that the performance of the initial segmentation algorithm had an influence on how much we could further improve the GM segmentation. If performance of the initial segmentation algorithm was already relatively high, the improvement obtained with our methods tended to be smaller. Differences in initial segmentation performance might be explained by whether the algorithm has been benchmarked on this particular type of data. We assume FSL FAST and CBS tools to have been benchmarked on MPRAGE and MP2RAGE data respectively, which would explain their relative high performance for these data types.

Importantly, our goal here was to aid established segmentation pipelines to deal with UHF sub-millimeter resolution data, not to replace those pipelines. Instead, the methods presented here should be considered as an alternative to manual slice-by-slice polishing of segmentations. Manually correcting the labels for even a minority of voxels is very time-consuming and can quickly become unreliable. In contrast, our methods greatly reduce the time required for manual polishing because they offer an efficient 2D summary and are more reliable because they are semi-automatic. We estimated that for whole brain cortical ribbon segmentation on average 7.5 hours of manual work could be saved by using our methods (for more details on this estimation see S1 Appendix). We

expect our methods to be most useful for UHF sub-millimeter resolution data since we expect vessels and dura mater to be less visible in images obtained with typical acquisition parameters at conventional field strength. This is because acquisitions at field strength of 3 T and below have less CNR and processing strategies [25, 26], that usually highlight non-brain tissue at UHF, are absent.

Moreover, we introduced the compositional data analysis framework to the neuroimaging community. Here, we used this framework to combine MRI acquisitions with three different image contrasts in order to derive improved tissue type segmentations. It should be emphasized, however, that the framework can be employed more generally. MRI can provide a multitude of informative images that weight tissue properties to generate the image contrast. The compositional data framework is ideally suited for the analysis and visualization of multiple images as it provides a principled way to combine any number of images. An envisioned future application of the compositional framework to MRI data is to use it to single out targeted cortical or subcortical structures based on their compositional properties. For an example of identifying subcortical structures see S1 Fig. For discussion of the broader implications of the application of compositional data analysis to images in general please see [72].

Our theoretical expectations implied that the methods presented here require high-resolution data (<1 mm). This requirement was unfortunately not met by most available segmentation validation data sets. Simulated phantom ("BrainWeb") data [73] are available at 1 mm and thus fell short of the resolution required for our purposes. Although an updated data set ("updated BrainWeb", [74, 75]) is available at higher resolution, the simulations in this data set were based on initial 3T MRI acquisitions. As a consequence, the updated BrainWeb data revealed considerably less bright vessel and dura mater voxels than 7 T data usually does and was not suitable to validate our methods.

These considerations led us to create our own high-resolution segmentation validation data sets for which we established the "ground truth" via manually-guided expert segmentation. While expert segmentations have well-known drawbacks [33, 76], they also have important advantages to alternative methods of establishing the ground truth, such as simulated phantom data. In particular, creating a validation data set based on empirical data and expert segmentations allowed us to benchmark our methods under conditions where our theoretical assumptions were met and the expected distribution of image intensities in 7 T brain data was respected. Being aware of the problems with expert segmentations, we alleviate concerns about the quality of our expert judgment and consequently the validity of the results presented here by making the data sets and corresponding ground truth segmentations as well as our precessing scripts available. This will allow other researchers to come up with their own judgment of the quality of the ground truth segmentation and validation data. In case changes to the ground truth are suggested and implemented, quantification could be re-run easily.

The 2D histogram method presented here is, in principle, capable of generating its own exhaustive tissue-type classifications, i.e. it does not necessarily depend on established segmentation pipelines to derive GM and WM labels. While we expect the 2D histogram method to give no advantage over established segmentation algorithms under standard conditions, the histogram method will compare well in cases where standard algorithms fail. Importantly, the 2D histogram method used here does not assume the data to conform to any atlas or template shape. Therefore, it is suitable also for acquisitions with only partial coverage (surface coils) or for specific populations (e.g. infants).

Using histogram-based methods would be more attractive if the process of specifying transfer functions was fully automatic. We note that there is no principled obstacle to doing this. Indeed, information-theoretic measures have been suggested [39] that would

make the normalized graph-cut application fully automatic, given the specification of an appropriate stopping criterion. The transfer functions (i.e. the circles and arcs applied to our 2D histograms) that we observed for the different brain tissue types were stable across subjects. This would allow to define probabilistic templates in the histogram space and transform the methods proposed here to a fully automatic exhaustive tissue-type classifications.

By making our validation data set(s) publicly available, we hope to inspire further algorithmic testing and development. There is currently a lack of validation data for the performance of tissue-type classification of MRI data acquired at ultra-high fields with sub-millimeter resolution. By publishing our data, our code and our work flow, we invite fellow scientists to benefit from our work but also to further contribute to it. The neuroimaging community can use our data to test the performance of entirely new methods or modifications to existing segmentation algorithms. Contributions could be made in the form of additional high-resolution data, more ground truth segmentations and algorithmic improvement. Anticipating such algorithmic improvements, we envision a future where segmentation of volumetric images will become gradually less laborious despite increasing resolution and volume of the data.

Acknowledgments

We would like to thank Thomas Emmerling for his guidance and advise on software implementation and Rainer Goebel for his encouragement and valuable comments. This work was financed by the Netherlands Organisation for Scientific Research (NWO). The authors O.F.G. and F.D.M. as well as data acquisition for the MPRAGE data set were supported by NWO VIDI grant 864-13-012. Author M.S. was supported by NWO research talent grant 406-14-108. Author I.M. was supported by NWO research talent grant 406-14-085. Author R.H. and acquisition of the MP2RAGE was funded by Technology Foundation STW (grant 12724).

References

1. Vaughan JT, Garwood M, Collins CM, Liu W, Delabarre L, Adriany G, et al. 7T vs. 4T: RF power, homogeneity, and signal-to-noise comparison in head images. Magnetic Resonance in Medicine. 2001;46(1):24–30. doi:10.1002/mrm.1156.

2. Duyn JH. The future of ultra-high field MRI and fMRI for study of the human brain; 2012.

3. Ugurbil K. Magnetic resonance imaging at ultrahigh fields. IEEE Transactions on Biomedical Engineering. 2014;61(5):1364–1379. doi:10.1109/TBME.2014.2313619.

4. Martino FD, Yacoub E, Kemper V, Moerel M, Uludag K, Weerd PD, et al. The impact of ultra-high field MRI on cognitive and computational neuroimaging. NeuroImage. 2017;doi:https://doi.org/10.1016/j.neuroimage.2017.03.060.

5. Kemper VG, Martino FD, Emmerling TC, Yacoub E, Goebel R. High resolution data analysis strategies for mesoscale human functional MRI at 7 and 9.4T. NeuroImage. 2017;doi:https://doi.org/10.1016/j.neuroimage.2017.03.058.

6. Dumoulin SO, Fracasso A, van der Zwaag W, Siero JCW, Petridou N. Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function. NeuroImage. 2017;doi:https://doi.org/10.1016/j.neuroimage.2017.01.028.

7. Polimeni JR, Renvall V, Zaretskaya N, Fischl B. Analysis strategies for high-resolution UHF-fMRI data. NeuroImage. 2017;doi:https://doi.org/10.1016/j.neuroimage.2017.04.053.

8. Trampel R, Bazin PL, Pine K, Weiskopf N. In-vivo magnetic resonance imaging (MRI) of laminae in the human cortex. NeuroImage. 2017;(September):1–9. doi:10.1016/j.neuroimage.2017.09.037.

9. Polimeni JR, Fischl B, Greve DN, Wald LL. Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. NeuroImage. 2010;52(4):1334–1346. doi:10.1016/j.neuroimage.2010.05.005.

10. Koopmans PJ, Barth M, Orzada S, Norris DG. Multi-echo fMRI of the cortical laminae in humans at 7 T. NeuroImage. 2011;56(3):1276–1285. doi:10.1016/j.neuroimage.2011.02.042.

11. De Martino F, Zimmermann J, Muckli L, Ugurbil K, Yacoub E, Goebel R. Cortical Depth Dependent Functional Responses in Humans at 7T: Improved Specificity with 3D GRASE. PLoS ONE. 2013;8(3):e60514. doi:10.1371/journal.pone.0060514.

12. Huber L, Goense J, Kennerley AJ, Trampel R, Guidi M, Reimer E, et al. Cortical lamina-dependent blood volume changes in human brain at 7 T. NeuroImage. 2015;107:23–33. doi:10.1016/j.neuroimage.2014.11.046.

13. Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, et al. Contextual Feedback to Superficial Layers of V1. Current Biology. 2015;25(20):2690–2695. doi:10.1016/j.cub.2015.08.057.

14. Kok P, Bains LJ, Van Mourik T, Norris DG, De Lange FP. Selective activation of the deep layers of the human primary visual cortex by top-down feedback. Current Biology. 2016;26(3):371–376. doi:10.1016/j.cub.2015.12.038.

15. Waehnert MD, Dinse J, Weiss M, Streicher MN, Waehnert P, Geyer S, et al. Anatomically motivated modeling of cortical laminae. NeuroImage. 2014;93:210–220. doi:10.1016/j.neuroimage.2013.03.078.

16. Yacoub E, Harel N, Ugurbil K. High-field fMRI unveils orientation columns in humans. Proceedings of the National Academy of Sciences. 2008;105(30):10607–10612. doi:10.1073/pnas.0804110105.

17. Zimmermann J, Goebel R, de Martino F, van de Moortele PF, Feinberg D, Adriany G, et al. Mapping the organization of axis of motion selective features in human area mt using high-field fmri. PLoS ONE. 2011;6(12):1–10. doi:10.1371/journal.pone.0028716.

18. De Martino F, Moerel M, Ugurbil K, Goebel R, Yacoub E, Formisano E. Frequency preference and attention effects across cortical depths in the human primary auditory cortex. Proceedings of the National Academy of Sciences. 2015;112(52):16036–16041. doi:10.1073/pnas.1507552112.

19. Goncalves NR, Ban H, Sanchez-Panchuelo RM, Francis ST, Schluppeck D, Welchman AE. 7 Tesla fMRI Reveals Systematic Functional Organization for Binocular Disparity in Dorsal Visual Cortex. Journal of Neuroscience. 2015;35(7):3056–3072. doi:10.1523/JNEUROSCI.3047-14.2015.

20. Nasr S, Polimeni JR, Tootell RBH. Interdigitated Color- and Disparity-Selective Columns within Human Visual Cortical Areas V2 and V3. Journal of Neuroscience. 2016;36(6). doi:10.1523/JNEUROSCI.3518-15.2016.

21. Tootell RBH, Nasr S. Columnar Segregation of Magnocellular and Parvocellular Streams in Human Extrastriate Cortex. The Journal of Neuroscience. 2017; p. 0690–17. doi:10.1523/JNEUROSCI.0690-17.2017.

22. Harvey BM, Klein BP, Petridou N, Dumoulin SO. Topographic Representation of Numerosity in the Human Parietal Cortex. Science. 2013;341(6150):1123–1126. doi:10.1126/science.1239052.

23. Harvey BM, Fracasso A, Petridou N, Dumoulin SO. Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(44):13525–30. doi:10.1073/pnas.1515414112.

24. Fracasso A, Petridou N, Dumoulin SO. Systematic variation of population receptive field properties across cortical depth in human visual cortex. NeuroImage. 2016;139:427–438. doi:10.1016/j.neuroimage.2016.06.048.

25. Van de Moortele PF, Auerbach EJ, Olman C, Yacoub E, Uğurbil K, Moeller S. T1 weighted brain images at 7 Tesla unbiased for Proton Density, T2* contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. NeuroImage. 2009;46(2):432–446. doi:10.1016/j.neuroimage.2009.02.009.

26. Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele PF, Gruetter R. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. NeuroImage. 2010;49(2):1271–1281. doi:10.1016/j.neuroimage.2009.10.002.

27. Viviani R, Pracht ED, Brenner D, Beschoner P, Stingl JC, Stöcker T. Multimodal MEMPRAGE, FLAIR, and R2* segmentation to resolve dura and vessels from cortical gray matter. Frontiers in Neuroscience. 2017;11(MAY):1–13. doi:10.3389/fnins.2017.00258.

28. van der Kouwe AJW, Benner T, Salat DH, Fischl B. Brain morphometry with multiecho MPRAGE. NeuroImage. 2008;40(2):559–569. doi:10.1016/j.neuroimage.2007.12.025.

29. Viviani R. A Digital Atlas of Middle to Large Brain Vessels and Their Relation to Cortical and Subcortical Structures. Frontiers in Neuroanatomy. 2016;10(February):12. doi:10.3389/FNANA.2016.00012.

30. Helms G, Kallenberg K, Dechent P. Contrast-driven approach to intracranial segmentation using a combination of T2- and T1-weighted 3D MRI data sets. Journal of Magnetic Resonance Imaging. 2006;24(4):790–795. doi:10.1002/jmri.20692.

31. Helms G. Segmentation of human brain using structural MRI. Magnetic Resonance Materials in Physics, Biology and Medicine. 2016;29(2):111–124. doi:10.1007/s10334-015-0518-z.

32. Bazin PL, Weiss M, Dinse J, Sch??fer A, Trampel R, Turner R. A computational framework for ultra-high resolution cortical segmentation at 7 Tesla. NeuroImage. 2014;93:201–209. doi:10.1016/j.neuroimage.2013.03.077.

33. Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. Computational and mathematical methods in medicine. 2015;2015:450341. doi:10.1155/2015/450341.

34. Renvall V, Witzel T, Wald LL, Polimeni JR. Automatic cortical surface reconstruction of high-resolution T1 echo planar imaging data. NeuroImage. 2016;134:338–354. doi:10.1016/j.neuroimage.2016.04.004.

35. Kashyap S, Ivanov D, Havlicek M, Poser BA, Uludağ K. Impact of acquisition and analysis strategies on cortical depth-dependent fMRI; 2017.

36. Kindlmann G, Durkin JW. Semi-automatic generation of transfer functions for direct volume rendering. In: Proceedings of the 1998 IEEE symposium on Volume visualization - VVS '98. New York, New York, USA: ACM Press; 1998. p. 79–86.

37. Kniss J, Kindlmann G, Hansen C. Interactive volume rendering using multi-dimensional transfer functions and direct manipulation widgets. In: Proceedings Visualization, 2001. VIS '01. IEEE; 2001. p. 255–562.

38. Kniss J, Kindlmann G, Hansen CD. Multidimensional transfer functions for volume rendering. Visualization Handbook. 2005; p. 189–209. doi:10.1016/B978-012387582-2/50011-3.

39. Ip CY, Varshney A, Jaja J. Hierarchical exploration of volumes using multilevel segmentation of the intensity-gradient histograms. IEEE Transactions on Visualization and Computer Graphics. 2012;18(12):2355–2363. doi:10.1109/TVCG.2012.231.

40. Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R. Modelling and Analysis of Compositional Data. Chichester, UK: John Wiley & Sons, Ltd; 2015.

41. Gulban OF, Schneider M. Segmentator v1.3.0; 2017. Available from: `https://doi.org/10.5281/zenodo.999487`.

42. Gulban OF, Schneider M, Marquardt I, Haast RAM, Martino FD. Dataset: A scalable method to improve gray matter segmentation at ultra high field MRI.; 2017. Available from: `https://doi.org/10.5281/zenodo.1117859`.

43. Schneider M, Gulban OF. Segmentator paper processing scripts; 2017. Available from: `https://doi.org/10.5281/zenodo.1098294`.

44. Kniss J, Kindlmann G, Hansen C. Multidimensional transfer functions for interactive volume rendering. IEEE Transactions on Visualization and Computer Graphics. 2002;8(3):270–285. doi:10.1109/TVCG.2002.1021579.

45. Ljung P, Krüger J, Groller E, Hadwiger M, Hansen CD, Ynnerman A. State of the Art in Transfer Functions for Direct Volume Rendering. Computer Graphics Forum. 2016;35(3):669–691. doi:10.1111/cgf.12934.

46. Glasser MF, Van Essen DC. Mapping Human Cortical Areas In Vivo Based on Myelin Content as Revealed by T1- and T2-Weighted MRI. Journal of Neuroscience. 2011;31(32):11597–11616. doi:10.1523/JNEUROSCI.2180-11.2011.

47. De Martino F, Moerel M, Xu J, Van De Moortele PF, Ugurbil K, Goebel R, et al. High-resolution mapping of myeloarchitecture in vivo: Localization of auditory areas in the human brain. Cerebral Cortex. 2015;25(10):3394–3405. doi:10.1093/cercor/bhu150.

48. Fauvel J, Wilson R, Flood R. Möbius and his Band: Mathematics and astronomy in nineteenth-century Germany. Oxford, England: Oxford University Press; 1993.

49. Munkres JR. Elements of Algebraic Topology. 1st ed. Addison-Wesley; 1984.

50. Aitchison J. The Statistical Analysis of Compositional Data. Journal of the Royal Statistical Society. 1982;44(2):139–177.

51. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. Isometric Logratio Transformations for Compositional Data Analysis. Mathematical Geology. 2003;35(3):279–300. doi:10.1023/A:1023818214614.

52. Lancaster HO. The Helmert Matrices. The American Mathematical Monthly. 1965;72(1):4. doi:10.2307/2312989.

53. Teeuwisse WM, Brink WM, Webb AG. Quantitative assessment of the effects of high-permittivity pads in 7 tesla MRI of the brain. Magnetic resonance in medicine. 2012;67(5):1285–1293.

54. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). Magnetic Resonance in Medicine. 2002;47(6):1202–1210. doi:10.1002/mrm.10171.

55. Haast RAM, Ivanov D, Formisano E, Uludağ K. Reproducibility and Reliability of Quantitative and Weighted T1 and T2(∗) Mapping for Myelin-Based Cortical Parcellation at 7 Tesla. Frontiers in neuroanatomy. 2016;10:112. doi:10.3389/fnana.2016.00112.

56. Yushkevich PA, Piven J, Cody Hazlett H, Gimpel Smith R, Ho S, Gee JC, et al. User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. Neuroimage. 2006;31(3):1116–1128.

57. Van Der Walt S, Colbert SC, Varoquaux G. The NumPy array: a structure for efficient numerical computation. Computing in Science & Engineering. 2011;13(2):22–30.

58. Jones E, Oliphant T, Peterson P, et al.. SciPy: Open source scientific tools for Python; 2001–. Available from: `http://www.scipy.org/`.

59. Hunter JD. Matplotlib: A 2D graphics environment. Computing In Science & Engineering. 2007;9(3):90–95. doi:10.1109/MCSE.2007.55.

60. Brett M, Hanke M, Côté MA, Markiewicz C, Ghosh S, Wassermann D, et al.. nipy/nibabel: 2.2.0; 2017. Available from: `https://doi.org/10.5281/zenodo.1011207`.

61. 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;2:e453. doi:10.7717/peerj.453.

62. Scharr H. Optimale operatoren in der digitalen bildverarbeitung; 2000.

63. Jähne B, Haußecker H. Handbook of computer vision and applications. vol. 2. Elsevier; 2000.

64. Gulban OF. ofgulban/tetrahydra: Tetrahydra v0.2.1; 2017. Available from: https://doi.org/10.5281/zenodo.1007952.

65. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging. 2001;20(1):45–57. doi:10.1109/42.906424.

66. Ashburner J, Friston KJ. Unified segmentation. NeuroImage. 2005;26(3):839–851. doi:10.1016/j.neuroimage.2005.02.018.

67. Smith SM. Fast robust automated brain extraction. Human Brain Mapping. 2002;17(3):143–155. doi:10.1002/hbm.10062.

68. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Medical Imaging. 2015;15(1):29. doi:10.1186/s12880-015-0068-x.

69. Sereno MI, Lutti A, Weiskopf N, Dick F. Mapping the human cortical surface by combining quantitative T1 with retinotopy. Cerebral Cortex. 2013;23(9):2261–2268. doi:10.1093/cercor/bhs213.

70. Dick F, Taylor Tierney A, Lutti A, Josephs O, Sereno MI, Weiskopf N. In Vivo Functional and Myeloarchitectonic Mapping of Human Primary Auditory Areas. Journal of Neuroscience. 2012;32(46):16095–16105. doi:10.1523/JNEUROSCI.1712-12.2012.

71. Marques JP, Gruetter R. New Developments and Applications of the MP2RAGE Sequence - Focusing the Contrast and High Spatial Resolution R1 Mapping. PLoS ONE. 2013;8(7). doi:10.1371/journal.pone.0069294.

72. Gulban OF. The relation between color spaces and compositional data analysis demonstrated with magnetic resonance image processing applications; 2017. Available from: https://arxiv.org/abs/1705.03457.

73. Collins DL, Zijdenbos aP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, et al. Design and construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging. 1998;17(3):463–468. doi:10.1109/42.712135.

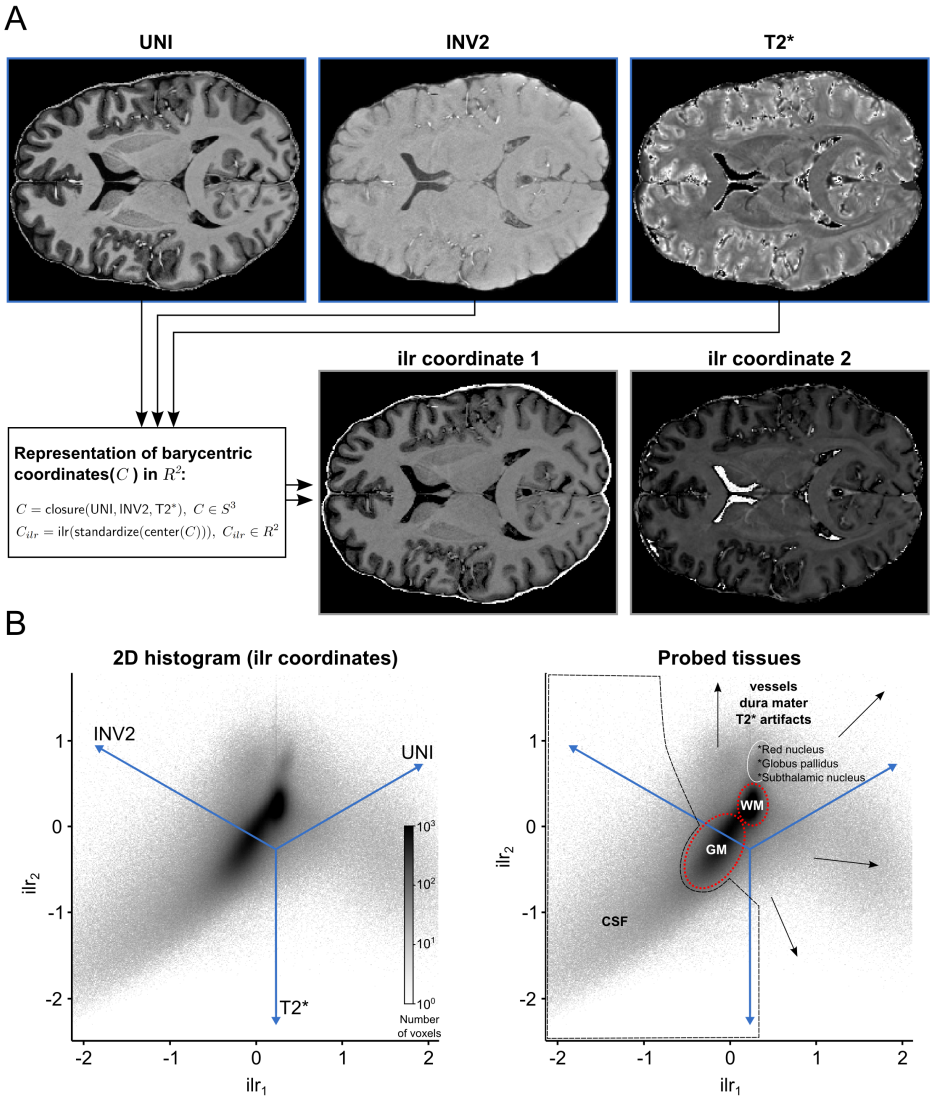
74. Aubert-Broche B, Evans AC, Collins L. A new improved version of the realistic digital brain phantom. NeuroImage. 2006;32(1):138 – 145. doi:https://doi.org/10.1016/j.neuroimage.2006.03.052.

75. Aubert-Broche B, Griffin M, Pike GB, Evans AC, Collins DL. Twenty New Digital Brain Phantoms for Creation of Validation Image Data Bases. IEEE Trans Med Imaging. 2006;25(11):1410–1416.

76. Valverde S, Oliver A, Cabezas M, Roura E, Lladó X. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. Journal of Magnetic Resonance Imaging. 2015;41(1):93–101. doi:10.1002/jmri.24517.

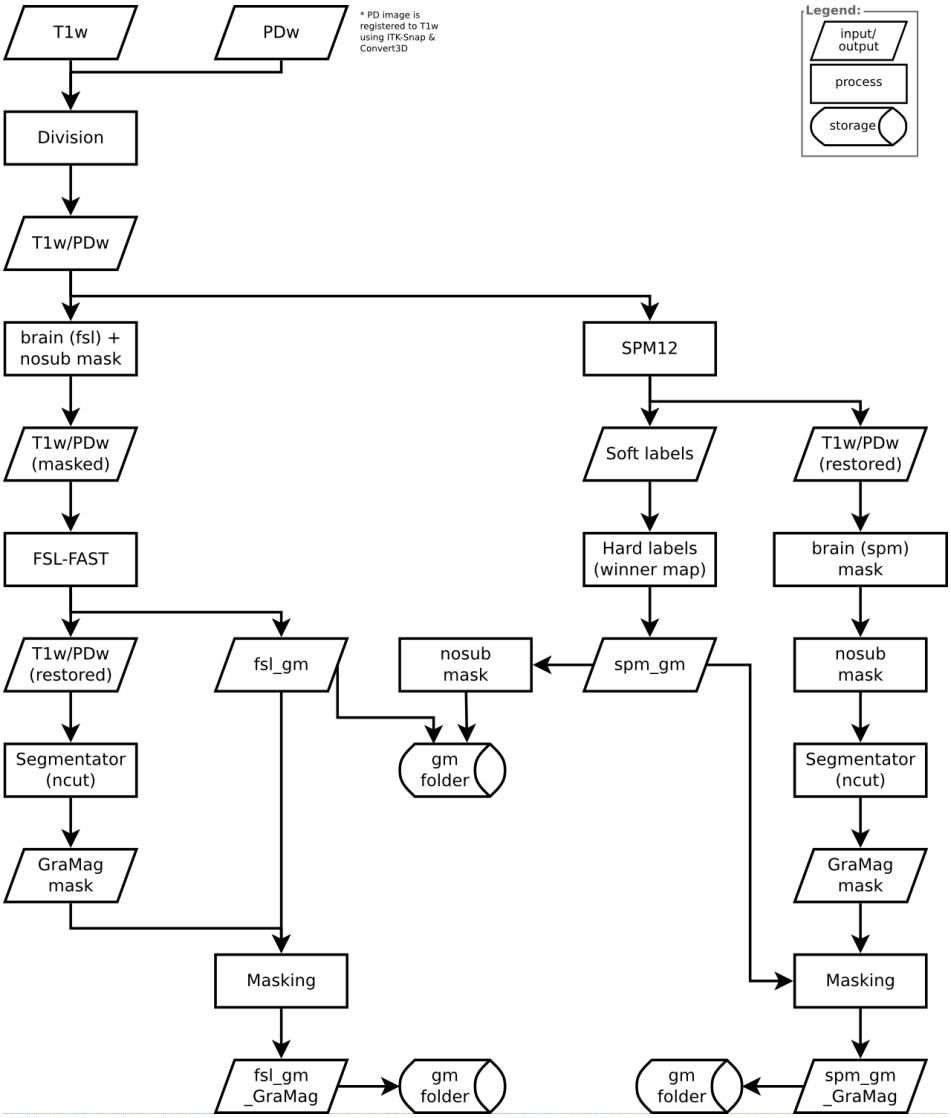
Supporting information

S1 Appendix. Time benefit estimation. To estimate the time saved by substituting manual correction with Segmentator, we did the following calculation: Assuming that the objective of manually correcting mislabeled GM voxels is to reduce the number of false positives and to increase the number of true positives, we computed the number of false and true positives both before and after Segmentator intervention. We computed the difference in true positives and false positives before and after the application of segmentator and subtracted the true positive difference from the false positive difference. The resulting number is assumed to indicate the number of voxels which would have to be subtracted in the process of manually correcting all the voxels without using Segmentator. For the MRI data presented here, this number amounted to around 200000 voxels to be corrected (out of a total number of 1100000 voxels in the cortical ribbon). On average 35000 out of those 200000 voxels could be corrected using Segmentator. Assuming that a trained operator can manually correct one voxel per second, on average, this amounts to 7.5 hours of manual work that can be substituted with 10 minutes of Segmentator usage for the whole brain GM ribbon segmentation at 0.7 mm isotropic resolution. The time that can potentially be saved by using Segmentator will scale with the total number of GM voxels - it will be higher for high resolution acquisition (more voxels) and lower at low resolution (less voxels). The script used for our estimation is available at [43].

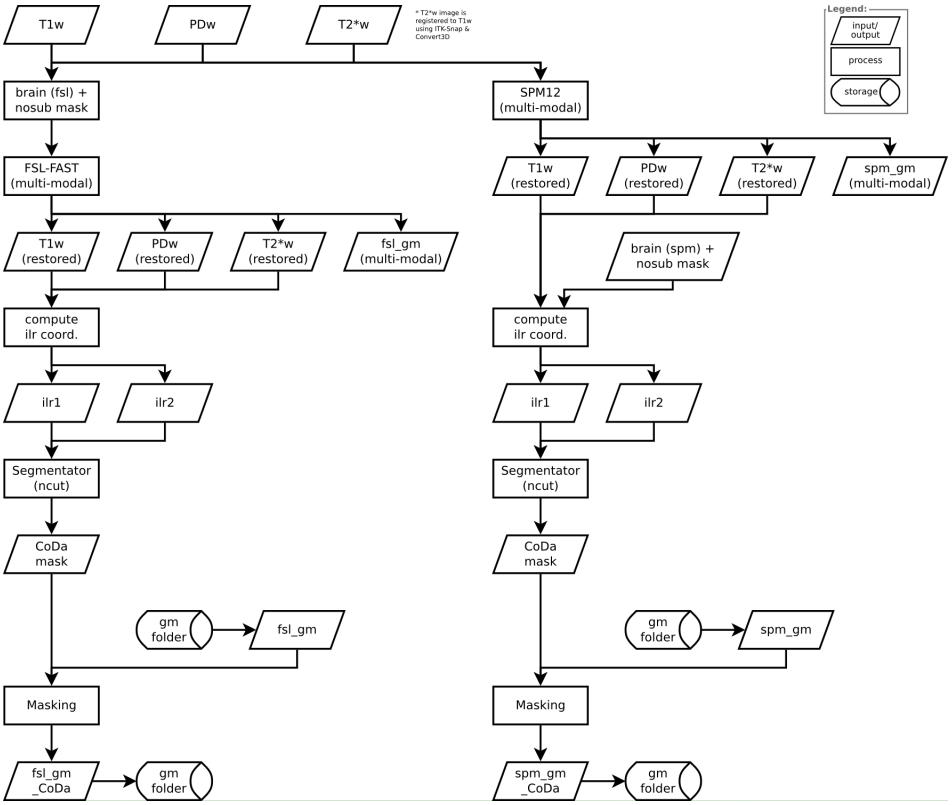


S1 Fig. 2D histogram representation of three 3D MRI contrast images. (A) Each voxel is considered as a 3 part composition. The barycentric coordinates of each composition which reside in 3D simplex space are represented in 2D real space after using a isometric log-ratio (ilr) transformation. (B) The ilr coordinates are used to create 2D histograms representing all voxels in the images. The blue lines are the embedded 3D real space primary axes (note that the input image units were initially normalized to have similar dynamic ranges to account for the large scale difference between T2* and MP2RAGE images). In this case, the ilr coordinates are not easily interpretable by themselves but they are useful to visualize the barycentric coordinates which are interpretable via the embedded real space axes. Darker regions in the histogram indicate that many voxels are characterized by this particular scale invariant combination of the image contrasts. In this representation, brain tissue (WM and GM, red dashed lines) becomes separable from non-brain tissue (black dashed lines and arrows). If desired, subcortical structures like the red nucleus, the globus pallidus and the subthalamic nucleus can additionally be identified (white circle).

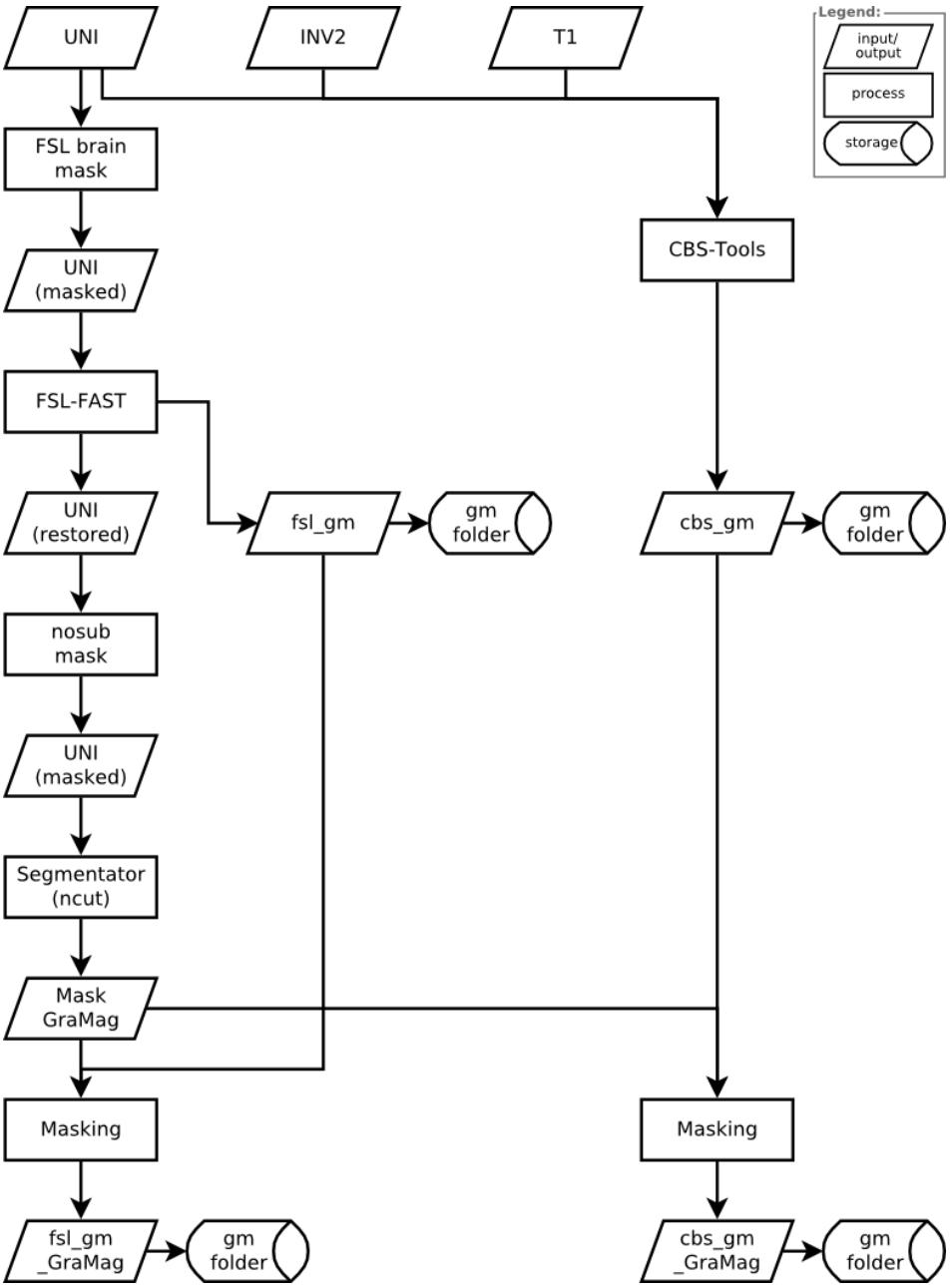
862



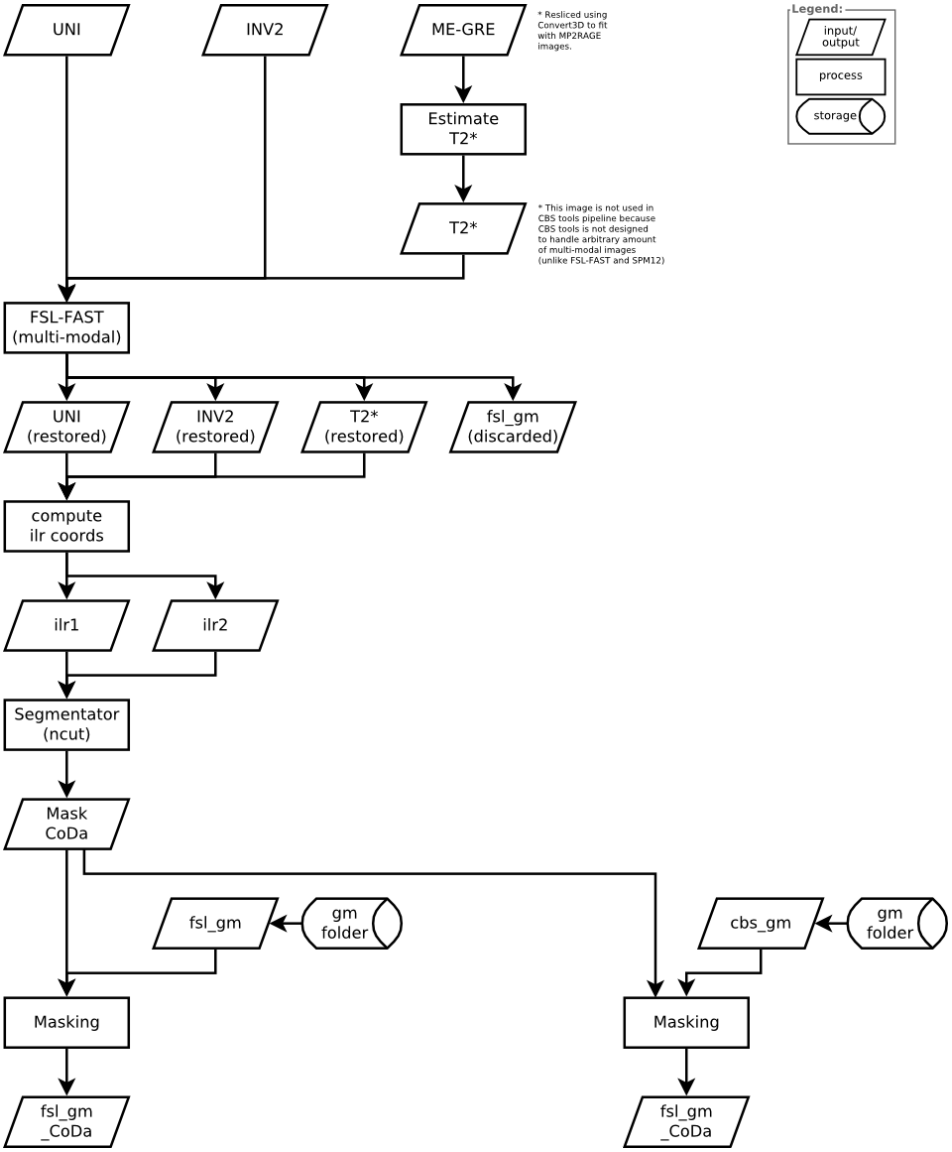
S2 Fig. Flowchart diagram MPRAGE GraMag pipeline. This diagram provides a detailed overview of all the inputs, processing steps and outputs for MPRAGE GraMag pipeline. Rectangular shapes represent processing steps, rhombic shapes represent input or outputs and cylindrical shapes represent input or output locations.



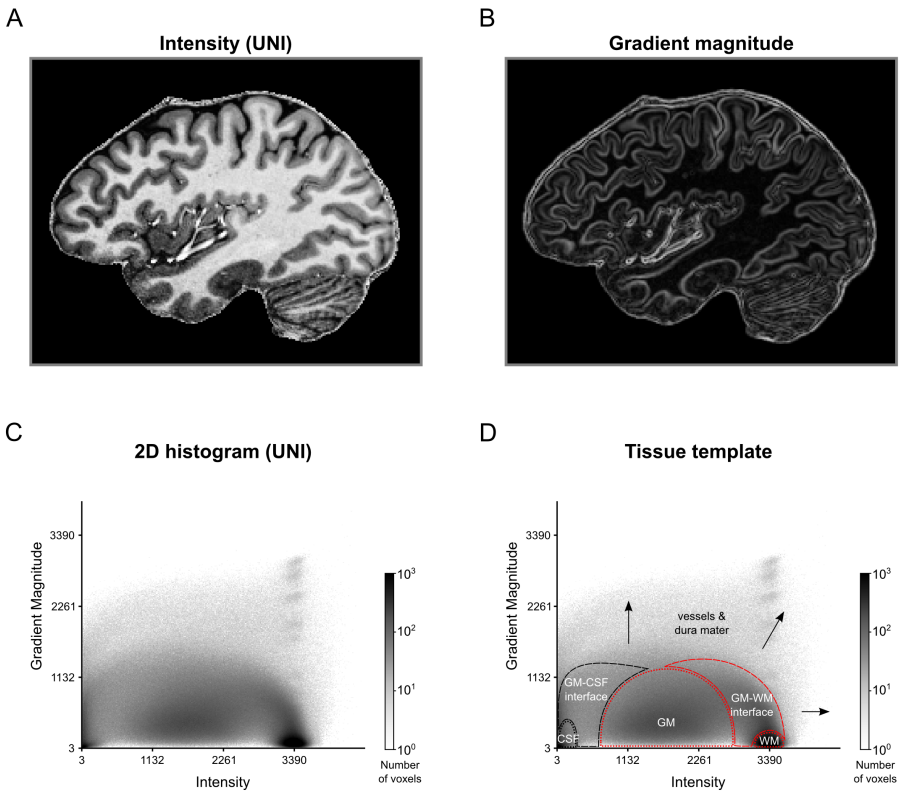
S3 Fig. Flowchart diagram MPRAGE CoDa pipeline. This diagram provides a detailed overview of all the inputs, processing steps and outputs for MPRAGE CoDa pipeline. Rectangular shapes represent processing steps, rhombic shapes represent input or outputs and cylindrical shapes represent input or output locations.



S4 Fig. Flowchart diagram MP2RAGE GraMag pipeline. This diagram provides a detailed overview of all the inputs, processing steps and outputs for MP2RAGE GraMag pipeline. Rectangular shapes represent processing steps, rhombic shapes represent input or outputs and cylindrical shapes represent input or output locations.



S5 Fig. Flowchart diagram MP2RAGE CoDa pipeline. This diagram provides a detailed overview of all the inputs, processing steps and outputs for MP2RAGE CoDa pipeline. Rectangular shapes represent processing steps, rhombic shapes represent input or outputs and cylindrical shapes represent input or output locations.



S6 Fig. 2D Histogram Representation for MRI Image of a Human Brain. The intensity (A) and gradient magnitude (B) values of a T1w-divided-by-PDw MRI image (MP2RAGE, 0.7 mm isotropic resolution) are represented in a 2D histogram (C). Darker regions in the histogram indicate that many voxels are characterized by this particular combination of image intensity and gradient magnitude. The 2D histogram displays a characteristic pattern with tissue types occupying particular areas of the histogram (D). Voxels containing CSF, dura mater or blood vessels (black dashed lines and arrows) cover different regions of the histogram than voxels containing WM and GM (red dashed lines). As a result, brain tissue becomes separable from non-brain tissue.

S1 Table. Segmentation performance MPRAGE data without artifact masking. The table shows the DICE (larger is better) and AVHD (less is better) for the initial SPM 12 and FSL FAST GM segmentations as well as after additional polishing, using either the gradient magnitude or the compositional data method.

		SPM		FAST	
		DICE	AVHD	DICE	AVHD
S02					
	Init	0.8721	0.7531	0.8936	0.5374
	Init + GraMag	0.8562	0.5620	0.8903	0.4895
	Init + CoDa	0.8392	0.6053	0.8431	0.5571
S03					
	Init	0.8638	0.8760	0.8907	0.5739
	Init + GraMag	0.8374	0.6236	0.8781	0.5038
	Init + CoDa	0.8325	0.6616	0.8388	0.5684
S05					
	Init	0.8864	0.6480	0.8969	0.5229
	Init + GraMag	0.8661	0.4800	0.8909	0.4679
	Init + CoDa	0.8562	0.5506	0.8641	0.5275
S06					
	Init	0.8661	0.7858	0.8753	0.6134
	Init + GraMag	0.8429	0.5768	0.8688	0.5525
	Init + CoDa	0.8386	0.6171	0.8446	0.5491
S07					
	Init	0.8974	0.6904	0.9330	0.3610
	Init + GraMag	0.8856	0.5845	0.9311	0.3465
	Init + CoDa	0.8818	0.5638	0.8528	0.5224

DICE, DICE Coefficient; AVHD, Average Hausdorff Distance; GraMag, gradient magnitude method; CoDa, compositional data method.

S2 Table. Segmentation performance MP2RAGE data without artifact masking. The table shows the DICE (larger is better) and AVHD (less is better) for the initial CBS tools and FSL FAST GM segmentations as well as after additional masking, using either the gradient magnitude or the compositional data method.

		CBS		FAST	
		DICE	AVHD	DICE	AVHD
S001					
	Init	0.9145	0.3339	0.8285	1.0439
	Init + GraMag	0.9179	0.3057	0.8092	0.8446
	Init + CoDa	0.8914	0.3672	0.8428	0.6286
S013					
	Init	0.8679	0.5519	0.7546	1.0556
	Init + GraMag	0.8644	0.4854	0.7358	0.9355
	Init + CoDa	0.8475	0.5091	0.7730	0.7613
S014					
	Init	0.8714	0.5068	0.8113	0.9850
	Init + GraMag	0.8717	0.4485	0.7865	0.8001
	Init + CoDa	0.8525	0.4955	0.8203	0.6723
S019					
	Init	0.9000	0.4195	0.8310	0.9678
	Init + GraMag	0.8927	0.3215	0.8079	0.7454
	Init + CoDa	0.8914	0.3716	0.8513	0.6114

DICE, DICE Coefficient; AVHD, Average Hausdorff Distance; GraMag, gradient magnitude method; CoDa, compositional data method.