NeuroImage 245 (2021) 118709

Contents lists available at ScienceDirect

NeuroImage

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

Multisite reproducibility and test-retest reliability of the T1w/T2w-ratio: A comparison of processing methods

Stener Nerland[a,b,*], Kjetil N. Jørgensen[a,b], Wibeke Nordhøy[c], Ivan I. Maximov[d,e,f], Robin A.B. Bugge[c], Lars T. Westlye[d,e], Ole A. Andreassen[b,d], Oliver M. Geier[c], Ingrid Agartz[a,b,g]

[a] Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo 0319, Norway
[b] NORMENT, Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[c] Department of Diagnostic Physics, Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway
[d] NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
[e] Department of Psychology, University of Oslo, Oslo, Norway
[f] Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway
[g] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Keywords:
Magnetic resonance imaging
T1w/T2w-ratio
Myelin mapping
Bias field correction
Intensity normalization
Outlier correction
Partial volume correction

ABSTRACT

Background: The ratio of T1-weighted (T1w) and T2-weighted (T2w) magnetic resonance imaging (MRI) images is often used as a proxy measure of cortical myelin. However, the T1w/T2w-ratio is based on signal intensities that are inherently non-quantitative and known to be affected by extrinsic factors. To account for this a variety of processing methods have been proposed, but a systematic evaluation of their efficacy is lacking. Given the dependence of the T1w/T2w-ratio on scanner hardware and T1w and T2w protocols, it is important to ensure that processing pipelines perform well also across different sites.

Methods: We assessed a variety of processing methods for computing cortical T1w/T2w-ratio maps, including correction methods for nonlinear field inhomogeneities, local outliers, and partial volume effects as well as intensity normalisation. These were implemented in 33 processing pipelines which were applied to four test-retest datasets, with a total of 170 pairs of T1w and T2w images acquired on four different MRI scanners. We assessed processing pipelines across datasets in terms of their reproducibility of expected regional distributions of cortical myelin, lateral intensity biases, and test-retest reliability regionally and across the cortex. Regional distributions were compared both qualitatively with histology and quantitatively with two reference datasets, YA-BC and YA-B1+, from the Human Connectome Project.

Results: Reproducibility of raw T1w/T2w-ratio distributions was overall high with the exception of one dataset. For this dataset, Spearman rank correlations increased from 0.27 to 0.70 after N3 bias correction relative to the YA-BC reference and from -0.04 to 0.66 after N4ITK bias correction relative to the YA-B1+ reference. Partial volume and outlier corrections had only marginal effects on the reproducibility of T1w/T2w-ratio maps and test-retest reliability. Before intensity normalisation, we found large coefficients of variation (CVs) and low intraclass correlation coefficients (ICCs), with total whole-cortex CV of 10.13% and whole-cortex ICC of 0.58 for the raw T1w/T2w-ratio. Intensity normalisation with WhiteStripe, RAVEL, and Z-Score improved total whole-cortex CVs to 5.91%, 5.68%, and 5.19% respectively, whereas Z-Score and Least Squares improved whole-cortex ICCs to 0.96 and 0.97 respectively.

Conclusions: In the presence of large intensity nonuniformities, bias field correction is necessary to achieve acceptable correspondence with known distributions of cortical myelin, but it can be detrimental in datasets with less intensity inhomogeneity. Intensity normalisation can improve test-retest reliability and inter-subject comparability. However, both bias field correction and intensity normalisation methods vary greatly in their efficacy and may affect the interpretation of results. The choice of T1w/T2w-ratio processing method must therefore be informed by both scanner and acquisition protocol as well as the given study objective. Our results highlight limitations of the T1w/T2w-ratio, but also suggest concrete ways to enhance its usefulness in future studies.

* Corresponding author.
E-mail address: stener.nerland@medisin.uio.no (S. Nerland).

https://doi.org/10.1016/j.neuroimage.2021.118709.
Received 9 June 2021; Received in revised form 29 October 2021; Accepted 2 November 2021
Available online 27 November 2021.

1. Introduction

Recently, the study of the myelin content of the cerebral cortex has undergone a revival, driven by advances in magnetic resonance imaging (MRI) techniques, a heightened appreciation of the importance of cortical myelin for brain function, and its relevance to a variety of brain disorders. Quantitative relaxometry methods such as R1 (1/T1) and R2* (1/T2*) mapping are considered particularly well-suited for in vivo cortical myelin mapping but are not yet commonly used in clinical studies (Edwards et al., 2018; Uddin et al., 2019). The ratio of T1-weighted (T1w) and T2-weighted (T2w) MRI images has been proposed as a proxy measure of cortical myelin, with the advantage of only requiring conventional MRI sequences. This use of the T1w/T2w-ratio is based on several observations regarding the MRI signal in the cerebral cortex. First, the T1w and T2w image contrast between grey and white matter is mainly determined by myelin (Koenig, 1991; Koenig et al., 1990). Second, the T1 signal can be used to identify and delineate myeloarchitectonic regions in the cerebral cortex (Clark et al., 1992; Eickhoff et al., 2004; Walters et al., 2003). Third, the T2 signal is roughly anticorrelated with myelin, and finally shared field inhomogeneities in the T1w and T2w images are reduced by taking the ratio (Glasser and Van Essen, 2011). In the cerebral cortex, high correlations between R1, a quantitative measure associated with myelin content, and the T1w/T2w-ratio have been reported (Shams et al., 2019), but the microstructural correlates of the T1w/T2w-ratio in white matter are less clear (Uddin et al., 2018).

The use of the T1w/T2w-ratio in neuroimaging research can be placed within three broad categories. First, its regional *intra-subject* distribution has been used to characterise the myeloarchitecture of the cortex, utilising its close spatial correspondence with histological measures of cortical myelin content (Glasser and Van Essen, 2011; Nieuwenhuys and Broere, 2017). For example, the T1w/T2w-ratio is one of the two main structural measures underlying the multi-modal parcellation of the cerebral cortex in Glasser et al. (2016) and its relationship with gradients of cyto- and myeloarchitecture, regional gene expression, and synaptic density are topics of ongoing research (Valk et al., 2020; Wang, 2020). Secondly, the T1w/T2w-ratio is used to characterise *inter-subject* differences in cortical myelin content. In healthy individuals, it has been used for brain age prediction (Rokicki et al., 2020), to characterise cross-sectional age trajectories of cortical myelination (Grydeland et al., 2019), and to study the relationship between cortical myelin and cognitive performance (Grydeland et al., 2013; Norbom et al., 2020; Tzourio-Mazoyer et al., 2019). It has also been studied in clinical conditions with known or hypothesised disturbance of cortical myelin where T1w and T2w images, but not more direct measures of myelin, are commonly acquired. Such conditions include multiple sclerosis (Nakamura et al., 2017; Preziosa et al., 2021; Righart et al., 2017), Huntington's disease (Rowley et al., 2018), Alzheimer's disease (Pelkmans et al., 2019), schizophrenia (Iwatani et al., 2015) and bipolar disorders (Ishida et al., 2017). Finally, the T1w/T2w-ratio has been used in whole-brain analyses (Ganzetti et al., 2014), for example to enhance contrast for brain tissue segmentation (Misaki et al., 2015) and to investigate tissue alterations in multiple sclerosis (Beer et al., 2016; Cooper et al., 2019).

The T1w/T2w-ratio is not, however, a direct measurement of cortical myelin as it is based on intensities expressed in arbitrary units influenced by non-biological variation (Fortin et al., 2016; Shinohara et al., 2014). Residual field inhomogeneities, head placement within the receive coil, image intensity scaling factors, and pulse sequence-specific acquisition parameters are factors that can affect the ratio. Receive coil (B1-) inhomogeneities are routinely corrected for during image acquisition using methods such as Phased array Uniformity Enhancement (PURE) on GE systems, Constant Level AppeaRance (CLEAR) on Philips systems, and Prescan Normalize on Siemens systems. In principle, computing the T1w/T2w-ratio should cancel residual intensity inhomogeneity shared between the T1w and T2w images while enhancing contrast related to myelin content (Glasser and Van Essen, 2011). However, other sources of intensity inhomogeneity, notably bias in the transmit field (B1+), may manifest differently in T1w and T2w images and will therefore not be cancelled in the ratio. It is often assumed that B1+ inhomogeneity profiles are similar across the T1w and T2w images, but this does not hold in general (Glasser and Van Essen, 2011; Sereno et al., 2013). It may therefore be necessary to perform post-hoc bias field correction to minimise residual field inhomogeneities. While dedicated sequences for estimating the B1+ bias field exist, the standard approaches are post-hoc image filters such as the nonparametric nonuniform intensity normalization (N3) algorithm (de Boor, 1972) and its more recent variant N4ITK (Tustison et al., 2010).

Since the T1w/T2w-ratio is formed from two separate image acquisitions, head movement between each pulse sequence may affect the results. Images must therefore be spatially aligned before the ratio is calculated, usually by rigid transformation of the T2w image to the T1w image (Glasser et al., 2013; Glasser and Van Essen, 2011). This step requires robust registration methods and considerable developmental efforts have been directed towards that goal (Greve and Fischl, 2009). However, head movement or poor image coregistration may still cause spurious T1w/T2w-ratio values in isolated voxels. Likewise, the sampling of voxels along reconstructed surfaces is sensitive to errors in the surface placement and this may similarly lead to spurious T1w/T2w-ratio values in isolated surface vertices. For these reasons, outlier correction methods both on the voxel and vertex level have previously been recommended (Glasser and Van Essen, 2011).

A further limitation is the resolution of the T1w and T2w images, typically 1 mm isotropic in clinical studies. Low resolution can lead to greater partial volume effects, whereby intensities within grey matter are contaminated by intensities in cerebrospinal fluid or white matter. To address this issue, Shafee et al. (2015) developed a method for performing partial volume correction (PVC) and found regionally dependent differences between the PVC corrected and uncorrected T1w/T2w-ratio maps, where greater differences were seen in sulci compared to gyri. Depending on the specific application of the T1w/T2w-ratio some of these technical limitations may be more important than others. For example, in vivo parcellation of myeloarchitecture requires reproducible regional distributions of T1w/T2w-values, particularly in boundary regions (Glasser and Van Essen, 2011). In this context, it is less important that T1w/T2w-ratio values are directly comparable across subjects. In contrast, the *precision* of measurements is vital to studies performing *inter-subject* comparisons or investigating associations with biological variables. This is commonly estimated by examining test-retest reliability. A variety of intensity normalisation methods have been proposed to improve the inter-subject comparability of the intensities of T1w and T2w images (Ghassemi et al., 2015; Nyúl and Udupa, 1999; Shah et al., 2011; Shinohara et al., 2014). These methods can be used to correct the T1w/T2w-ratio by applying them to the T1w and T2w images separately before taking the ratio.

In the present study, we implemented several previously proposed methods for computing the T1w/T2w-ratio, including bias field correction and intensity normalisation of the T1w and T2w images prior to taking the ratio, and applied these to four test-retest datasets acquired on four different scanners. These processing pipelines included corrections for field inhomogeneity, partial volume effects and surface outliers, as well as intensity normalisation methods which aim to improve inter-subject comparability of image intensities. We evaluated each pipeline based on two distinct set of criteria. First, correspondence with myelin staining studies is important when using the T1w/T2w-ratio to perform myeloarchitectonic parcellation and depends on good *intra-subject* reproducibility of regional distributions. We therefore assessed the correspondence of T1w/T2w-ratio maps qualitatively with cortical myelin maps from histological staining studies (Nieuwenhuys and Broere, 2017) and quantitatively using rank correlations with in vivo data from Glasser et al. (2016) and from a recent preprint from Glasser et al. (2021). Laterality indices (LIs) were used as a proxy for spatially dependent field bias which may lead to hemispheric

Table 1
T1w and T2w sequences and parameters for each dataset.

	T1w sequence				T2w sequence			
Dataset	NOR-MR750	NOR-Premier	DONDERS	HCP	NOR-MR750	NOR-Premier	DONDERS	HCP
Vendor	General Electric	General Electric	Siemens	Siemens	General Electric	General Electric	Siemens	Siemens
Scanner	Discovery MR750	SIGNA Premier	Magnetom Prisma	Connectom Skyra	Discovery MR750	SIGNA Premier	Magnetom Prisma	Connectom Skyra
Head coil	32-channel	48-channel	32-channel	32-channel	32-channel	48-channel	32-channel	32-channel
Sequence	BRAVO	MPRAGE	MPRAGE	MPRAGE	CUBE	CUBE	SPACE	SPACE
Resolution	1 mm iso	0.8 mm iso	0.8 mm iso	0.7 mm iso	1 mm iso	0.8 mm iso	0.8 mm iso	0.7 mm iso
TA (mm:ss)	04:43	06:22	06:02	07:40	04:23	05:21	06:19	08:24
TR (ms)	8.16	2356	2200	2400	2500	3202	3200	3200
TE (ms)	3.18	3.13	2.64	2.14	71.68	90.6	569	565
TI (ms)	450	950	1100	1000	–	–	–	–
FA (deg)	12	8	11	8	90	Variable	Variable	Variable
BW (Hz/px)	244	238	170	210	488	555	579	744
iPAT	2	2	3	2	2 × 2	2 × 2	3	2
B1- calibration	PURE	PURE	Prescan Normalise	No	PURE	PURE	Prescan Normalise	No

intensity biases. Second, acceptable test-retest reliability is important when using the T1w/T2w-ratio for *inter-subject* comparisons. We therefore evaluated the test-retest reliability of the T1w/T2w-ratio via coefficients of variation (CVs) and intraclass correlation coefficients (ICCs). This gives an estimate of the precision which is an important constraint on the detectible effects when using the measure (Matheson, 2019).

2. Materials and methods

2.1. Participants

Participants were included from four test-retest datasets acquired on two 3T GE scanners (GE Medical Systems, Milwaukee, USA) and two 3T Siemens scanners (Siemens, Erlangen, Germany). There was an overlap of eight participants for the NOR-MR750 and NOR-Premier datasets. For these participants, the scan-rescan interval between the first session on in the NOR-MR750 and the first session on the NOR-Premier dataset ranged from 9.8 to 10 months. Scanner model, sequence types and acquisition parameters for each dataset is given in Table 1.

2.1.1. Dataset 1 - The Norwegian centre for mental disorders research - general electric discovery MR750 (NOR-MR750)

This dataset was collected as part of a reliability study conducted at the Norwegian Centre for Mental Disorders Research (NORMENT). Nine healthy participants (mean age = 35.76 years; range = [26.31–59.70]; 55% male) were recruited internally and scanned on a 3T GE Discovery MR750 equipped with a 32-channel receive coil in two sessions with a two-week interval between sessions. This interval was included to account for the influence of intra-subject variability related to day-to-day variation. Each participant was scanned at the same time of day across sessions. In each session two pairs of T1w and T2w MRI images were acquired, with repositioning between each pair, resulting in four T1w and four T2w images for each participant per session.

2.1.2. Dataset 2 - The Norwegian centre for mental disorders research - general electric signa premier (NOR-Premier)

This dataset was acquired as a continuation of the NOR-MR750 test-retest dataset after a major scanner upgrade to 3T GE SIGNA Premier equipped with a 48-channel head coil on the same magnet. The dataset consisted of nine healthy participants (mean age = 35.80 years; range = [25.24–60.52]; 55% male) of which eight participants overlapped with the NOR-MR750 dataset. As with the NOR-MR750 dataset, a pair of test-retest acquisitions were made in two separate sessions with a two week interval between sessions. The interval between the first session of the NOR-MR750 dataset and the first session of the NOR-Premier dataset was 9.6 months on average.

2.1.3. Dataset 3 - the Donders centre for cognitive neuroimaging - siemens magnetom prisma (Donders)

Seventeen healthy participants (mean age = 24.7 ± 2.8 years) were enroled in a test-retest study at the Donders Centre for Cognitive Neuroimaging, Radboud University, Nijmegen, Netherlands (https://www.ru.nl/donders/). Participants were scanned on a 3T Siemens Magnetom Prisma equipped with a 32-channel head coil. Each participant was scanned twice with repositioning between each pair of T1w and T2w acquisitions and randomised head fixation either with cushions only or cushions and a chin-rest.

2.1.4. Dataset 4 - human connectome project - siemens connectom Skyra - (HCP)

A subset of 34 participants (mean age = 30.70 years; range = [22–35]; 29% male) with a scan-rescan interval of less than six months (mean = 3.85 months; range = [2–5]) were selected from test-retest acquisitions included as part of the 1200 Subjects release (S1200 Release, February 2017) by the WU-Minn Human Connectome Project (HCP) (Van Essen et al., 2013). Participants were recruited as part of the HCP Young Adult study and were scanned twice on a customised 3T Siemens Connectom Skyra scanner using a standard 32-channel receive head coil.

2.1.5. Ethics statement

The study was carried out in accordance with the Helsinki Declaration. Participants provided written informed consent at each site and agreed to participation in scientific studies. Each site had ethical approval from their local ethics committees. Data was handled according to guidelines set forth by the Norwegian Data Protection Authority in compliance with GDPR regulations.

2.2. MRI preprocessing

All scans were processed in FreeSurfer (version 7.1.0; https://surfer.nmr.mgh.harvard.edu/) with both T1w and T2w images used as input and the *-T2pial* flag enabled. Briefly, FreeSurfer segments cortical and subcortical tissue compartments, and creates surface meshes representing the boundaries between grey and white matter (WM) and grey matter (GM) and non-brain. For more details on the FreeSurfer processing steps see Fischl (2012). For the HCP dataset, the PreFreeSurfer pipeline had already been used to create an undistorted structural image for the subject in native space, perform the initial brain extraction, coregister T1w and T2w images, perform bias field correction using the square of the T1w and T2w images, and transform the data to MNI space. The PreFreeSurfer pipeline was applied to the HCP dataset for all pipelines due to technical considerations particular to this dataset, including the application of gradient nonlinearity distortion in images acquired with oblique slices relative to the scanner

coordinate system and the lack of on-scanner B1- correction, following the recommendations in Glasser et al. (2013). All images and surface reconstructions were quality controlled by trained research assistants and manual editing was performed following standard FreeSurfer guidelines in the event of surface reconstruction errors. A total number of 170 T1w/T2w pairs were processed, composed of nine participants in the NOR-MR750 dataset and nine participants in the NOR-Premier dataset each scanned four times (scan-rescan in two sessions), 15 participants in the Donders dataset (after three participants were excluded) and 34 participants in the HCP dataset each scanned twice (scan-rescan on the same day and across two sessions respectively).

For the HCP Minimal Processing Pipeline (HCP-MPP), FreeSurfer v6.0.0 and the PreFreeSurfer, FreeSurfer and PostFreeSurfer pipelines were used. The Connectome Workbench was used to map each myelin map computed with these pipelines, as well as the previously published HCP Multi-Modal Parcellation (HCP-MMP; Glasser et al., 2016) scheme, from the *FS_LR* surface to the standard FreeSurfer *fsaverage* surfaces as described on the HCP Wiki (https://wiki.humanconnectome.org/download/attachments/63078513/Resampling-FreeSurfer-HCP.pdf).

For all pipelines, T2w images were rigidly coregistered to T1w images using *bbregister* in FreeSurfer with FSL initialisation (Greve and Fischl, 2009). The resolution of the datasets included 0.7 mm for the HCP dataset, 0.8 mm for the Donders and NOR-Premier dataset, and 1 mm isotropic voxel resolutions for the NOR-MR750 dataset. In order to facilitate the comparison between the different datasets, and to ensure that the results would be generalisable to the more commonly used 1 mm isotropic voxel resolution in clinical settings, we conformed all images to 1 mm prior to FreeSurfer and T1w/T2w-ratio processing. Follow-up analyses were conducted to assess the effect of downsampling (Section 2.4.8). Throughout, *fslmaths* and *fslstats* from FSL (v6.0.4; https://fsl.fmrib.ox.ac.uk/fsl/) were used for image operations and preprocessing was done on the TSD (Service for Sensitive Data) computing cluster at the University of Oslo.

2.3. Calculation of T1w/T2w-ratio maps

We first implemented corrections for partial voluming, surface outliers and bias field in nine different pipelines using local and non-linear correction methods for computing the T1w/T2w-ratio. Here, the goal was to assess pipelines on their ability to reproduce the expected regional distributions of the *intra-subject* T1w/T2w-ratio. These processing pipelines included the raw ratio ($T1w/T2w_{Raw}$), N3 or N4ITK bias field correction only ($T1w/T2w_{N3}$ & $T1w/T2w_{N4}$), partial volume correction only ($T1w/T2w_{PVC}$), surface outlier only ($T1w/T2w_{OC}$), N3 or N4ITK bias field correction with partial volume and outlier correction ($T1w/T2w_{N3\text{-}All}$ & $T1w/T2w_{N4\text{-}All}$), and finally the HCP Minimal Processing Pipeline with ($T1w/T2w_{HCP\text{-}BC}$) and without ($T1w/T2w_{HCP\text{-}Raw}$) template-based bias correction (Section 2.3.6). We qualitatively assessed the correspondence of T1w/T2w-ratio maps with histological myelin maps (Section 2.4.2) and computed Spearman rank correlations with the YA-BC and the YA-B1+ reference datasets (Sections 2.3.7 and 2.4.3), and used the Laterality Index (Section 2.4.6) as a measure of the removal of lateral intensity bias.

Secondly, in order to assess and improve the test-retest reliability of the T1w/T2w-ratio, i.e. the consistency between repeated measurements, we evaluated T1w/T2w-ratio maps after applying eight different intensity normalisation methods (Section 2.3.5). These intensity normalisation methods were applied to the $T1w/T2w_{Raw,}$ $T1w/T2w_{N3\text{-}All}$, and the $T1w/T2w_{N4\text{-}All}$ pipelines. The main outcomes of interest here were coefficients of variation (Section 2.4.4 and 2.4.7), a standardised measure of dispersion, and intraclass correlation coefficients (Section 2.4.5 and 2.4.7) as a measure of the agreement between groups of measurements. It was assumed that true intra-subject scan-rescan differences are small, and smaller for scan-rescan pairs closer in time, relative to inter-subject variation which is ideally still present after intensity normalisation. In total, we evaluated $9 + 8 \times 3 = 33$ separate pipelines. An overview of the processing pipelines implemented in this study is shown in Fig. 1.

2.3.1. Surface projection

For all processing methods, T1w/T2w-ratio intensities in grey matter were sampled along equivolumetric surfaces at distances of 10% to 80% of cortical thickness from the white-grey surface created with the Surface Tools toolbox (https://github.com/kwagstyl/surface_tools). This toolbox reconstructs intermediary surfaces with improved correspondence with cortical laminae compared to standard sampling methods (Waehnert et al., 2014). The values of vertices in the medial wall were set to zero for all figures and analyses.

2.3.2. Bias field correction

For bias field correction, we used both N3 bias correction (Sled et al., 1998) and the more recent N4ITK algorithm (Tustison et al., 2010). Both bias field correction methods were used with a shrink factor of 1 and masked with the whole-brain segmentation created in the standard FreeSurfer pipeline.

2.3.3. Partial volume correction

Partial volume correction was performed as described in Shafee (2015) using the implementation included in FreeSurfer. This method first calculates the proportion of voxels in an upsampled MR image falling within the cortical ribbon, i.e. between the reconstructed pial and white-grey surfaces, and then inverts systems of linear equations to recover the “true” intensity values adjusted for partial volume effects.

2.3.4. Surface outlier correction

Surface-based outlier correction was performed as in Glasser et al. (2011). This algorithm calculates, for each vertex, the mean and standard deviation (SD) of normal-appearing vertex values (only values within 3 SDs of the total mean T1w/T2w-ratio) in a neighbourhood of 10 steps along the surface and replaces values, in vertices exceeding this mean by more than 2 SDs, with a Gaussian-weighted mean of its neighbours. The algorithm was implemented in MATLAB (The Mathworks, Inc., Massachusetts, USA) using the *geodesic* toolbox (Wang, 2021) for computing distances between vertices.

2.3.5. Intensity normalisation

To assess the effect of intensity normalisation, we applied eight different normalisation methods to T1w/T2w-ratio maps from the $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, and $T1w/T2w_{N4\text{-}All}$ processing methods which were compared to the unnormalised $T1w/T2w_{Raw}$, and $T1w/T2w_{HCP\text{-}Raw}$ processing methods. All intensity normalisation methods were based on the implementations in the Intensity Normalization toolbox (https://github.com/jcreinhold/intensity-normalization; Reinhold et al., 2019) with more detailed documentation available here: https://intensity-normalization.readthedocs.io/. Intensity normalisation was performed on T1w and T2w images separately.

The eight normalisation methods used in this study were: 1) Z-score (ZS), which subtracts the mean intensity of the skull-stripped image and divides by the standard deviation of the same region, 2) Fuzzy C-Means (FCM; Chen and Giger, 2004), which determines a three-class segmentation using fuzzy C-means clustering, and then normalises the image to a standard value using the mean of a specified tissue class, in our case GM, 3) Gaussian Mixture Model (GMM), which fits a gaussian mixture to the skull-stripped image and translates the WM mean to a standard value, 4) Kernel Density Estimation (KDE), which uses kernel density estimation to find the peak of WM intensities and translates this peak to a standard value, 5) Nyúl and Udupa (N&U; Nyúl and Udupa, 1999), which uses piecewise linear histogram matching to estimate a transformation that standardises the histograms of a population of images, 6) WhiteStripe (WS; Shinohara et al., 2014), which identifies normal-appearing WM voxels, subtracts the mean of this region from the image and finally divides the image by its standard deviation, 7) Removal of Artificial

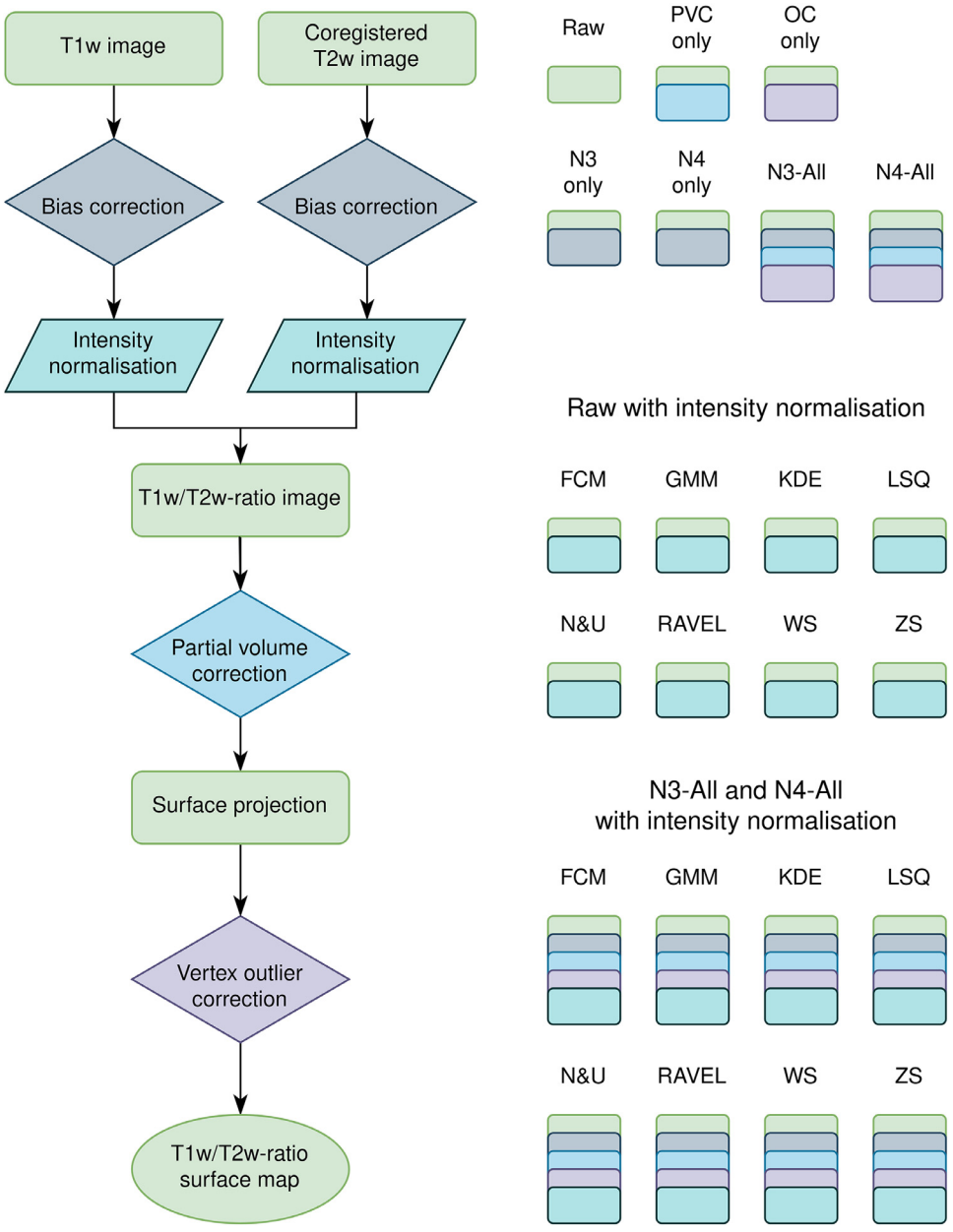

Fig. 1. Overview of the processing methods implemented in this study. Each box and colour on the left corresponds to a step in the pipeline. Collections of boxes on the right correspond to a pipeline. On the top are the pipelines $T1w/T2w_{Raw}$, $T1w/T2w_{PVC}$, $T1w/T2w_{OC}$, $T1w/T2w_{N3}$, $T1w/T2w_{N4}$, $T1w/T2w_{N3\text{-}All}$ and $T1w/T2w_{N4\text{-}All}$ which were primarily assessed on their ability to reproduce the expected T1w/T2w-ratio distributions. On the bottom are the pipelines $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, and $T1w/T2w_{N4\text{-}All}$ with each intensity normalisation method. $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ were computed in the separate HCP Minimal Processing Pipeline and are not depicted.

Voxel Effect by Linear regression (RAVEL; Fortin et al., 2016), which first uses WhiteStripe and then estimates the unwanted non-biological variation from a control region which is then regressed out for each voxel in the image, and finally 8) Least Squares (LSQ), which minimises the squared distance between CSF, GM and WM for a population of images. Of these intensity normalisation methods, N&U, RAVEL, and LSQ are population-based methods that are applied to a set of images, while the other methods are used on each image separately.

2.3.6. HCP-MPP T1w/T2w-ratio maps

The T1w/T2w-ratio maps of the HCP-MPP are calculated in the PostFreeSurfer pipeline. Briefly, in this pipeline T1w/T2w-ratio values are projected to a midthickness surface as described in Glasser & Van Essen (2011) and individual surfaces are mapped to the population-average Conte69 surface. The vertex-wise difference between individual $T1w/T2w_{HCP\text{-}Raw}$ maps and the average Conte69 map is then smoothed with a large kernel and subtracted from the $T1w/T2w_{HCP\text{-}Raw}$ maps to create the $T1w/T2w_{HCP\text{-}BC}$ maps. The goal of this procedure is to replace spurious low frequency variation in individual maps with values from the more robust group map. The resulting $T1w/T2w_{HCP\text{-}BC}$ maps therefore closely match the Conte69 average, having undergone both a template-based bias field correction as well as a nonlinear intensity normalisation. This is in contrast to the linear intensity normalisation methods described in Section 2.3.5 and results in a regional shift towards the Conte69 template precluding direct inter-subject comparisons of T1w/T2w-ratio values. See Glasser & Van Essen (2011) and Glasser et al. (2013) for more details.

2.3.7. Reference T1w/T2w-ratio datasets

As reference T1w/T2w-ratio datasets, we used two publicly available T1w/T2w-ratio datasets from the HCP Young Adult (HCP-YA) project acquired using the HCP MRI protocol (Uğurbil et al., 2013; Van Essen et al., 2013). The first reference dataset, YA-BC, consists of T1w/T2w-ratio maps from 210 participants (aged 22–35 years) and is available through the HCP-500 data release. These maps were computed using the $T1w/T2w_{HCP\text{-}BC}$ pipeline and were used, together with thickness and fMRI, in the creation of the HCP-MMP parcellation in Glasser et al. (2016). The dataset is available online at:

https://balsa.wustl.edu/WDpX. The second reference dataset, YA-B1+, consists of T1w/T2w-ratio maps from 1042 participants (aged 22–35 years) and is available through the HCP-1200 data release. These maps were computed with the $T1w/T2w_{HCP\text{-}Raw}$ pipeline with the addition of a new correction procedure, where a separate actual flip angle (AFI) sequence was used to adjust the T1w/T2w-ratio maps for residual B1+ field effects instead of the template-adjustment method used in the $T1w/T2w_{HCP\text{-}BC}$. For details on this procedure see the preprint by Glasser et al. (2021).

For both reference datasets, mean T1w/T2w-ratio values for each region of interest (ROI) in the HCP-MMP were extracted directly from the native *FS_LR* surface using the Connectome Workbench. For the YA-B1+ dataset, we extracted data corresponding to the group correction method described in the recent preprint by Glasser et al. (2021). This dataset is available at: https://balsa.wustl.edu/study/show/mDBP0.

2.4. Data analysis

Analyses aimed to answer two main questions: 1) What is the reproducibility of regional T1w/T2w-ratio distributions compared to the expected distribution from the literature on cortical myelin and can local and nonlinear correction factors improve it? 2) What is the test-retest reliability of the T1w/T2w-ratio and can intensity normalisation methods be used to improve it?

2.4.1. Data quality control

All images were visually inspected prior to analysis to ensure acceptable image quality without large image artefacts such as severe ringing due to movement, misregistration between the T1w and T2w images, or inconsistencies in image intensity scale or acquisition parameters. This resulted in the exclusion of the three first subjects in the Donders dataset due to one missing T2w image, one image with ringing artefacts, and one image with an intensity range suggestive of acquisition or reconstruction differences compared to the other images. No other data was excluded.

2.4.2. Qualitative assessment of cortical maps

To qualitatively assess the regional correspondence with known cortical myelin distributions, we created mean T1w/T2w-ratio maps for each of the non-intensity normalised processing methods. The correspondence of these mean maps was assessed visually against cortical myelin maps based on histological staining studies, which we considered to be the gold standard. These maps are based on extensive work spanning the period from 1910 to 1970, later composed to form a map of the human cortex (Nieuwenhuys and Broere, 2017). We also compared the mean T1w/T2w-ratio maps to previously published T1w/T2w-ratio maps of the YA-BC dataset (Glasser et al., 2016) and of the YA-B1+ where a new adjustment method to account for B1+ field effects has been applied as described in the recent preprint by Glasser et al. (2021). Based on this literature, we located regions known to exhibit high, intermediate, or low cortical myelination. We then assessed the T1w/T2w-ratio maps and reported where the maps diverged from expectations. Briefly, we expected high T1w/T2w-ratio values in primary sensory and motor regions as well as early association cortices, and lower T1w/T2w-ratio values in higher association cortices and in frontal regions.

The regions that were expected to exhibit high levels of cortical myelin were: 1) the motor-somatosensory strip in the central sulcus on the lateral side and extending into the medial 'notch' in the paracentral lobule, 2) the visual cortex in the occipital lobe extending into the temporal and parietal lobes, and 3) early auditory areas in the Sylvian fissure and the upper segment of the temporal lobe. Regions where we expected intermediate cortical myelin levels were: 1) the inferior parietal cortex, and 2) most of the temporal, prefrontal, cingulate and medial and superior parietal cortices. The lowest myelin levels were expected in: 1) the anterior insula, 2) temporal pole, 3) medial prefrontal cortex, and 4) portions of the anterior cingulate cortex. Consensus was reached by comparing notes on each T1w/T2w-ratio map. For assessments of the gradients, we examined if the transitions from high-to-intermediate, high-to-low and intermediate-to-low myelination matched expectations as described above. See Nieuwenhuys & Broere (2017) and Glasser & Van Essen (2011) for further information on the distribution of myelin in the human cerebral cortex.

2.4.3. ROI-wise correlations with reference datasets

To quantify the correspondence with expected myelin distributions, we extracted, for each dataset and processing method, mean T1w/T2w-ratio values in 176 ROIs of the HCP Multi-Modal Parcellation (HCP-MMP), where 4 ROIs (H_ROI, EC_ROI, PreS_ROI and RSC_ROI) were excluded for each hemisphere due to their overlapping with the medial wall yielding spurious T1w/T2w-ratio values. The mean values were calculated across all the T1w/T2w-ratio maps in that dataset, i.e. each test-retest pair in the HCP and Donders datasets and test-retest pair for each of the two sessions in the NOR-MR750 and NOR-Premier datasets. They were then used to compute Spearman rank correlations, using the *cor.test* function from the *stats* package in R, with mean T1w/T2w-ratio values from two reference T1w/T2w-ratio datasets, the YA-BC and YA-B1+ (Section 2.3.7). These datasets were used as references given their acquisition parameters optimised for T1w/T2w-ratio mapping and high correspondence with histology.

2.4.4. Coefficients of variation

To quantify test-retest reliability, we calculated percentage CVs by first calculating vertex-wise CVs which were combined for each participant using the root-mean-square (RMS) formula in each HCP-MMP ROI and for the whole cortex. The ROI and whole-cortex CVs were then combined across individuals to form summary ROI and whole-cortex CVs for each processing method and pipeline. To compute the vertex-wise CVs for each participant we used the formula

$$CV_{Vertex,i,j} = \frac{sd(x_{i,j})}{mean(x_{i,j})}$$

where $x_{i,j}$ is the set of T1w/T2w-ratio values for the i^{th} subject in the j^{th} vertex. For each subject, these were combined for each ROI in the HCP-MMP atlas and for the whole cortical surface to form individual regional coefficients of variation ($CV_{ROI,i}$) and whole-cortex coefficients of variation ($CV_{Cortex,i}$) using the RMS formula as follows

$$CV_{ROI,i} = \sqrt{\sum_{j=1}^{N_{ROI}} \frac{CV_{Vertex,i,j}^2}{N_{ROI}}}$$

where the sum is over the N_{ROI} vertices in each ROI and the whole cortex. To combine ROI and whole-cortex coefficients of variation across all participants, we used the subject-wise CVs calculated above in the RMS formula

$$CV_{ROI} = \sqrt{\sum_{i=1}^{N_{Subjects}} \frac{CV_{ROI,i}^2}{N_{Subjects}}}$$

where the sum is taken over all the subjects in each dataset. Here the RMS is used rather than the arithmetic mean, since the mean is a biased estimator and will tend to underestimate the true test-retest variability (Glüer et al., 1995). Lower CVs indicate higher measurement precision.

2.4.5. Intraclass correlation coefficients

To assess global test-retest variability of the T1w/T2w-ratio distributions, we calculated intraclass correlation coefficients (ICCs) of the median of the T1w/T2w-ratio values (ICC_{Median}), using the *icc* function from the *irr* package in R. We used the two-way mixed single score ICC, denoted ICC(A,1) in the classification by McGraw & Wong (1996). The median was used as a more robust descriptor of the global T1w/T2w-ratio distribution. Higher ICCs represent better absolute agreement between test and retest.

2.4.6. Laterality indices

To assess how correction methods, in particular bias field correction, affected lateral bias in T1w/T2w-ratio values, we computed average interhemispheric T1w/T2w-ratio differences using the percentage laterality index (LI) for each of the 180 regions of the HCP-MMP atlas with the mean laterality index

$$LI = \frac{200}{180}\sum_{j=1}^{180}\frac{ROI_{j,lh} - ROI_{j,rh}}{ROI_{j,lh} + ROI_{j,rh}}$$

Assuming that inter-hemispheric distributions of cortical myelin is not systematically skewed, higher laterality indices indicate the presence of field bias and successful bias field correction should result in a lower mean laterality index.

2.4.7. Visualisation of results

To visualise regional distributions of CV_{ROI} and ICC_{ROI} we created surface maps based on the HCP-MMP atlas using the *fsbrain* toolbox in R (https://github.com/dfsp-spirit/fsbrain). For comparison of the test-retest reliability and laterality indices of each processing method and visualisation of outlier subjects, we created box plots of the whole-cortex CVs, median ROI-wise ICCs and mean laterality indices across ROIs for each participant grouped by processing method and dataset. To visualise the test-retest reliability and intra-subject variation compared to inter-subject variation we created ridgeline plots of the T1w/T2w-ratio distributions for each processing method using the *ggridges* toolbox in R (https://wilkelab.org/ggridges/) which are presented in the Supplementary Materials.

2.4.8. Follow-up analyses

Three of the datasets included in this study were comprised of high-resolution images that were downsampled to 1 mm isotropic to ensure that results were generalisable to commonly available datasets. Past studies have suggested that the use of high-resolution data is beneficial for T1w/T2w-ratio mapping (Glasser and Van Essen, 2011). We therefore performed follow-up analyses for the NOR-Premier and the HCP datasets where we assessed the effect of downsampling by running all analyses described above for $T1w/T2w_{Raw\text{-}Hires}$, $T1w/T2w_{HCP\text{-}Raw\text{-}Hires}$, and T1w/T2wHCP-BC-Hires where we used high-resolution data. The Donders dataset was not included in these analyses, due to the incompatibility of the *-cw256* flag, which was necessary in this dataset due to defacing, and the *-hires* flag of the high-resolution *recon-all* pipeline.

In our main analyses, we transformed output from the $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ pipelines from the native *FS_LR* surface of the HCP-MPP to the native surface of FreeSurfer, *fsaverage*. This transformation could be a source of error, and we therefore tested its effects by extracting data directly from *FS_LR* space for the $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ pipelines using Connectome Workbench and computing ROI-wise correlations with the reference datasets.

3. Results

3.1. Qualitative assessment of T1w/T2w-ratio maps

Mean maps for $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, $T1w/T2w_{N4\text{-}All}$, $T1w/T2w_{HCP\text{-}Raw}$, and $T1w/T2w_{HCP\text{-}BC}$ for each dataset are presented in Fig. 2. Intensity normalisation and corrections for outliers or partial volume effects did not appreciably influence the T1w/T2w-ratio maps and these are therefore not depicted or discussed in the qualitative assessment. See Supplementary Figures 1–5 for ridgeline plots of the T1w/T2w-ratio distributions for each processing method.

3.1.1. NOR-MR750

As expected from histology, high $T1w/T2w_{Raw}$ values were seen in the lateral motor cortex and in the auditory cortex, but the visual cortex had relatively low values despite this region being known to be highly myelinated. Some frontal medial regions had relatively high T1w/T2w-ratio values despite these regions being known to be lightly myelinated. N3 bias correction improved the correspondence with cortical myelin maps, with higher values in the medial occipital lobe but with no clearly delineated notch in the paracentral lobule. N4ITK bias correction led to an improved correspondence with histology on the medial side, but parietal and superior frontal regions with spuriously high values were observed with unexpectedly high values around Brodmann areas 44 and 45. $T1w/T2w_{HCP\text{-}Raw}$ was similar to $T1w/T2w_{Raw}$, with additional smoothing inherent to the HCP Minimal Processing Pipeline. The best correspondence with cortical myelin maps was seen with $T1w/T2w_{HCP\text{-}BC}$, where high values were found in the auditory, motor and visual cortices as expected and a clearly delineated notch was present in the paracentral lobule.

3.1.2. NOR-Premier

$T1w/T2w_{Raw}$ closely matched histological cortical myelin maps with high values in the auditory, motor and visual cortices and low values in frontal, temporal and parietal regions. N3 bias correction only had a marginal effect on the T1w/T2w-ratio maps, but N4ITK bias correction introduced spuriously high values especially in frontal and parietal regions where cortical myelin content is known to be low. $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ gave similar results to $T1w/T2w_{Raw}$ and $T1w/T2w_{N3\text{-}All}$ with some additional smoothing.

3.1.3. Donders

$T1w/T2w_{Raw}$ values were high in the auditory, visual and motor cortices as expected, but lightly myelinated frontal and parietal regions contained spuriously high T1w/T2w-ratio values. N3 bias correction improved the visual correspondence with histology, but N4ITK bias correction seemed to further exacerbate the poor correspondence in regions with spuriously high T1w/T2w-ratio values. $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ were similar, but $T1w/T2w_{HCP\text{-}BC}$ had a better contrast between high and low myelinated regions, with relatively lower T1w/T2w-ratio values in frontal parietal regions in particular, and thus gave the best results visually.

3.1.4. HCP

Regional distributions of $T1w/T2w_{Raw}$ closely matched histology, with the exception of some frontal regions with spuriously high values especially in the left hemisphere. N3 bias correction had only minor effects on the T1w/T2w-ratio, but N4ITK bias correction seemed to reduce the presence of spuriously high values in frontal regions. $T1w/T2w_{HCP\text{-}Raw}$ showed excellent correspondence with the expected cortical myelin distribution, but unexpectedly high values were seen in a frontal parietal region. This region was attenuated in $T1w/T2w_{HCP\text{-}BC}$, which gave the best results.

3.2. ROI-wise correlations with reference datasets

See Table 2 for Spearman rank correlations between the YA-BC and YA-B1+ reference datasets for each of the non-intensity normalised processing methods, and Tables 3–5 for Spearman rank correlations after intensity normalisation applied to $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, and $T1w/T2w_{N4\text{-}All}$ respectively.

3.2.1. NOR-MR750

We found low correspondence with the YA-BC reference for $T1w/T2w_{Raw}$ [ρ = 0.27], $T1w/T2w_{OC}$ [ρ = 0.27], $T1w/T2w_{PVC}$ [ρ = 0.26], and $T1w/T2w_{HCP\text{-}Raw}$ [ρ = 0.23]. This was greatly improved for $T1w/T2w_{N3}$ [ρ = 0.70] and $T1w/T2w_{N3\text{-}All}$ [ρ = 0.70], and moderately improved for $T1w/T2w_{N4}$ [ρ = 0.53] and $T1w/T2w_{N4\text{-}All}$ [ρ = 0.57]. The best results were seen for $T1w/T2w_{HCP\text{-}BC}$ [ρ = 0.92].

We found very low correspondence with the YA-B1+ reference, for $T1w/T2w_{Raw}$ [ρ = −0.05], $T1w/T2w_{OC}$ [ρ = −0.04], $T1w/T2w_{PVC}$ [ρ

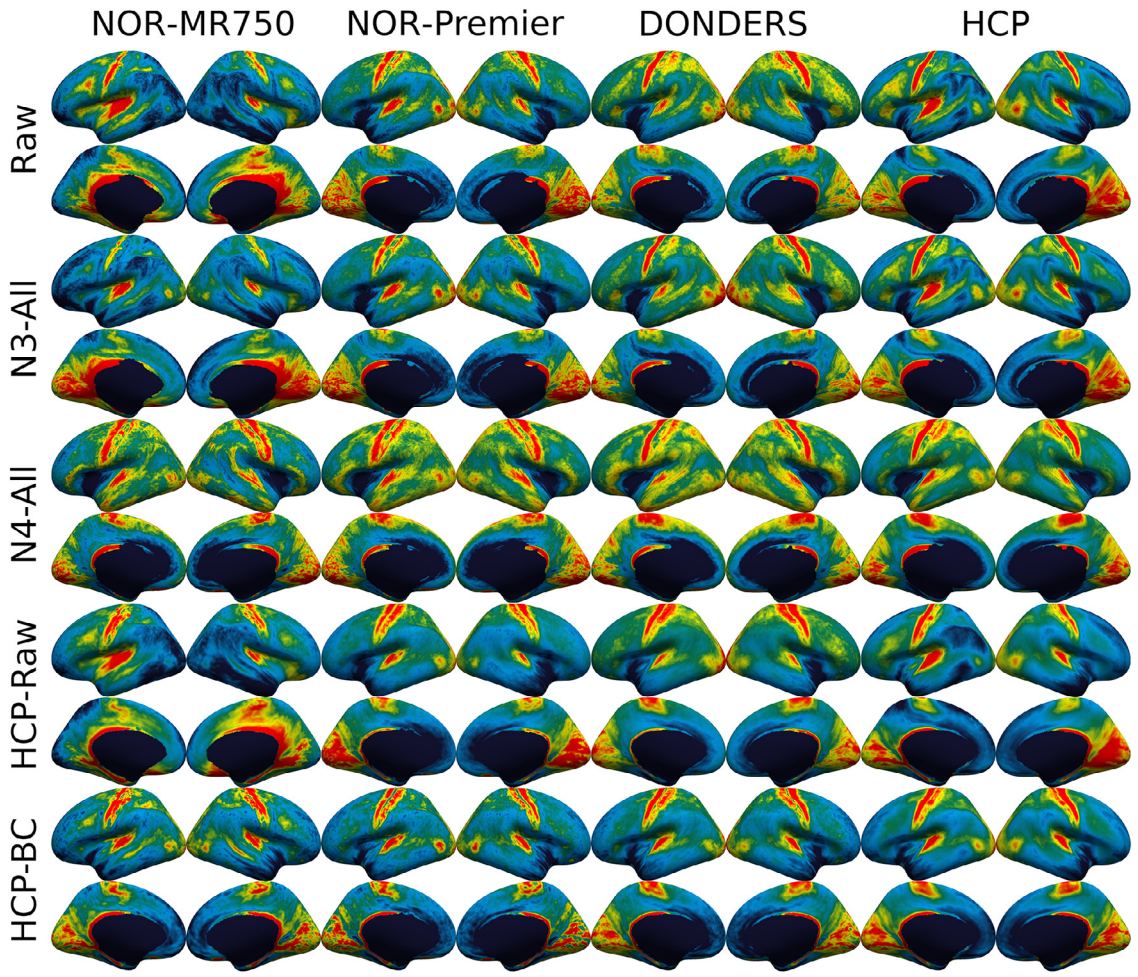

Fig. 2. Mean T1w/T2w-ratio maps in each dataset for a selection of calculation methods. Partial volume correction and surface outlier correction yielded similar maps as the raw processing method. Intensity normalisation after bias correction yielded qualitatively identical maps to those before intensity normalisation and are therefore not depicted. Colours were chosen by distributing them at five equally spaced points from the 3rd percentile to the 96th percentile, and were set to black, blue, green, yellow and red respectively.

= −0.05], and T1w/T2w$_{HCP\text{-}Raw}$ [ρ = −0.04]. There was a moderate improvement for T1w/T2w$_{N3}$ [ρ = 0.36] and T1w/T2w$_{N3\text{-}All}$ [ρ = 0.36]. The highest correlations were seen for T1w/T2w$_{N4}$ [ρ = 0.66], T1w/T2w$_{N4\text{-}All}$ [ρ =0.70], and T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.77].

Intensity normalisation did not strongly affect rank correlations with the YA-BC and the YA-B1+ reference datasets (see Tables 3-5).

3.2.2. NOR-Premier

In the NOR-Premier dataset, we found high correlations with the YA-BC reference for T1w/T2w$_{Raw}$ [ρ = 0.89], T1w/T2w$_{OC}$ [ρ = 0.89], T1w/T2w$_{PVC}$ [ρ = 0.89], T1w/T2w$_{N3}$ [ρ = 0.86], T1w/T2w$_{N3\text{-}All}$ [ρ = 0.86], T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.90], and T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.92]. Lower correlations were seen with T1w/T2w$_{N4}$ [ρ = 0.56] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.56].

We found high correlations with the YA-B1+ reference for T1w/T2w$_{Raw}$ [ρ = 0.86], T1w/T2w$_{OC}$ [ρ = 0.87], T1w/T2w$_{PVC}$ [ρ = 0.86], T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.88], and T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.90]. Correlations were marginally lower for T1w/T2w$_{N3}$ [ρ = 0.81] and T1w/T2w$_{N3\text{-}All}$ [ρ = 0.82]. The lowest correlations were observed for T1w/T2w$_{N4}$ [ρ = 0.73] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.73].

Intensity normalisation did not greatly affect correlations with the reference datasets, with the exception of ZS which decreased the correlations with the YA-B1+ reference to ρ = 0.77 when applied to T1w/T2w$_{Raw}$.

3.2.3. Donders

In the Donders dataset, high correlations with the YA-BC reference were seen for T1w/T2w$_{Raw}$ [ρ = 0.80], T1w/T2w$_{OC}$ [ρ = 0.80], T1w/T2w$_{PVC}$ [ρ = 0.80], T1w/T2w$_{N3}$ [ρ = 0.76], T1w/T2w$_{N3\text{-}All}$ [ρ = 0.76], and T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.79]. Lower correlations were seen with T1w/T2w$_{N4}$ [ρ = 0.31] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.46]. The best results were seen with T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.96].

We found high correlations with the YA-B1+ reference for T1w/T2w$_{Raw}$ [ρ = 0.96], T1w/T2w$_{OC}$ [ρ = 0.96], T1w/T2w$_{PVC}$ [ρ = 0.96], T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.91], and T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.97]. Correlations were lower for T1w/T2w$_{N3}$ [ρ = 0.86] and T1w/T2w$_{N3\text{-}All}$ [ρ = 0.87]. The lowest correlations were observed for T1w/T2w$_{N4}$ [ρ = 0.44] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.63].

Intensity normalisation did not greatly affect correlations with the reference datasets, with the exception of RAVEL and GMM which decreased the correlations with the YA-BC reference toρ = 0.68 and ρ = 0.18 respectively when applied to T1w/T2w$_{Raw}$.

3.2.4. HCP

In the HCP dataset, high correlations with the YA-BC reference were seen for T1w/T2w$_{Raw}$ [ρ = 0.90], T1w/T2w$_{OC}$ [ρ = 0.89], T1w/T2w$_{PVC}$ [ρ = 0.90], T1w/T2w$_{N3}$ [ρ = 0.92], T1w/T2w$_{N3\text{-}All}$ [ρ = 0.91], T1w/T2w$_{HCP\text{-}Raw}$ [ρ = 0.93], and T1w/T2w$_{HCP\text{-}BC}$ [ρ = 0.97]. The lowest correlations were seen for T1w/T2w$_{N4}$ [ρ = 0.71] and T1w/T2w$_{N4\text{-}All}$ [ρ = 0.71].

Table 2
Summary statistics for processing methods without intensity normalisation, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		Raw	PVC only	OC only	N3 only	N3-All	N4 only	N4-All	HCP-Raw	HCP-BC
Mean T1w/T2w-ratio	NOR-MR750	1.37	1.36	1.37	1.44	1.43	2.75	1.49	2.02	1.30
	NOR-Premier	1.97	1.96	1.97	2.13	2.11	1.98	2.17	3.20	1.29
	DONDERS	1.38	1.37	1.38	1.48	1.47	2.53	1.57	2.55	1.29
	HCP	1.25	1.24	1.24	1.30	1.29	1.53	1.52	1.81	1.30
	Total (Mean)	1.49	1.48	1.49	1.59	1.57	2.20	1.68	2.39	1.30
Cofficient of variation (%)	NOR-MR750	12.30	12.33	12.00	12.14	11.84	14.32	13.96	6.93	6.61
	NOR-Premier	8.82	8.84	8.45	8.69	8.33	9.25	8.88	6.23	8.48
	DONDERS	8.56	8.55	8.18	8.05	7.65	9.42	9.05	6.59	6.73
	HCP	10.41	10.45	10.11	10.17	9.88	10.70	10.43	4.88	4.32
	Total (RMS)	10.13	10.15	9.80	9.89	9.56	11.11	10.78	6.20	6.70
Intraclass correlation coefficient	NOR-MR750	0.13	0.13	0.13	0.17	0.17	0.26	0.26	0.31	0.41
	NOR-Premier	0.90	0.90	0.90	0.93	0.93	0.93	0.93	0.59	0.54
	DONDERS	0.57	0.57	0.57	0.61	0.62	0.27	0.27	0.14	0.67
	HCP	0.73	0.73	0.73	0.73	0.73	0.74	0.74	0.69	0.55
	Total (Mean)	0.58	0.58	0.58	0.61	0.61	0.55	0.55	0.43	0.54
Laterality Index (%)	NOR-MR750	8.80	8.83	8.82	−0.20	−0.13	0.90	1.01	9.59	−0.03
	NOR-Premier	1.93	1.91	1.92	0.03	0.00	0.11	0.05	1.92	−0.10
	DONDERS	0.36	0.34	0.41	0.07	0.09	−0.50	−0.43	0.65	−0.11
	HCP	−0.14	−0.13	0.02	−0.41	−0.25	−0.27	−0.11	0.14	−0.22
	Total (RMS)	4.51	4.52	4.52	0.23	0.15	0.54	0.55	4.90	0.13
Correlation with YA-BC	NOR-MR750	0.27	0.26	0.27	0.70	0.70	0.53	0.57	0.23	0.92
	NOR-Premier	0.89	0.89	0.89	0.86	0.86	0.56	0.56	0.90	0.92
	DONDERS	0.80	0.80	0.80	0.76	0.76	0.31	0.46	0.79	0.96
	HCP	0.90	0.90	0.89	0.92	0.91	0.71	0.71	0.93	0.97
	Total (Mean)	0.71	0.71	0.71	0.81	0.81	0.53	0.58	0.71	0.94
Correlation with YA-B1+	NOR-MR750	−0.05	−0.05	−0.04	0.36	0.36	0.66	0.70	−0.04	0.77
	NOR-Premier	0.86	0.86	0.87	0.81	0.82	0.73	0.73	0.90	0.88
	DONDERS	0.96	0.96	0.96	0.86	0.87	0.44	0.63	0.97	0.91
	HCP	0.72	0.72	0.72	0.79	0.80	0.80	0.82	0.77	0.88
	Total (Mean)	0.62	0.62	0.63	0.71	0.71	0.66	0.72	0.65	0.86

We found high correlations with the YA-B1+ reference for $T1w/T2w_{N3\text{-}All}$ [$\rho = 0.77$], $T1w/T2w_{N3}$ [$\rho = 0.79$], $T1w/T2w_{N3\text{-}All}$ [$\rho = 0.80$], $T1w/T2w_{N4}$ [$\rho = 0.80$], $T1w/T2w_{N4\text{-}All}$ [$\rho = 0.82$], and $T1w/T2w_{HCP\text{-}BC}$ [$\rho = 0.88$]. The lowest correlations were observed for $T1w/T2w_{Raw}$ [$\rho = 0.72$], $T1w/T2w_{OC}$ [$\rho = 0.72$], and $T1w/T2w_{PVC}$ [$\rho = 0.72$].

Intensity normalisation did not strongly affect rank correlations with the YA-BC reference (see Tables 3-5).

3.3. Whole-cortex coefficients of variation

See Table 2 for whole-cortex coefficients of variation, CV_{Cortex}, for each processing method without intensity normalisation grouped by dataset, and Tables 3-5 for CV_{Cortex} after intensity normalisation for $T1w/T2w_{Raw}$, $T1w/T2w_{N3\text{-}All}$, and $T1w/T2w_{N4\text{-}All}$ respectively. See Fig. 4 for box plots of subject-wise CV_{Cortex} for each processing method.

3.3.1. NOR-MR750

Total whole-cortex CVs were high in the NOR-MR750 dataset for all the non-normalised processing methods, $T1w/T2w_{Raw}$ [$CV_{Cortex} = 12.30\%$], $T1w/T2w_{PVC}$ [$CV_{Cortex} = 12.33\%$], $T1w/T2w_{OC}$ [$CV_{Cortex} = 12.00\%$], $T1w/T2w_{N3}$ [$CV_{Cortex} = 12.14\%$], and $T1w/T2w_{N3\text{-}All}$ [$CV_{Cortex} = 11.84\%$]. Whole-cortex CVs were higher with N4ITK bias correction; $T1w/T2w_{N4}$ [$CV_{Cortex} = 14.32\%$] and $T1w/T2w_{N4\text{-}All}$ [$CV_{Cortex} = 13.96\%$]. The lowest CVs were seen with $T1w/T2w_{HCP\text{-}Raw}$ [$CV_{Cortex} = 6.93\%$] and $T1w/T2w_{HCP\text{-}BC}$ [$CV_{Cortex} = 6.61\%$].

3.3.2. NOR-Premier

For the NOR-Premier dataset, whole-cortex CVs were lower than in the NOR-MR750 and HCP datasets and similar to those of the Donders dataset, with $T1w/T2w_{Raw}$ [$CV_{Cortex} = 8.82\%$], $T1w/T2w_{PVC}$ [$CV_{Cortex} = 8.84\%$], $T1w/T2w_{OC}$ [$CV_{Cortex} = 8.45\%$], $T1w/T2w_{N3}$ [$CV_{Cortex} = 8.69\%$], $T1w/T2w_{N3\text{-}All}$ [$CV_{Cortex} = 8.33\%$], $T1w/T2w_{N4\text{-}All}$ [$CV_{Cortex} = 8.88\%$], and $T1w/T2w_{HCP\text{-}BC}$ [$CV_{Cortex} = 8.48\%$]. Whole-cortex CVs were the highest with N4ITK bias correction only; $T1w/T2w_{N4}$ [$CV_{Cortex} = 9.25\%$] and the lowest with $T1w/T2w_{HCP\text{-}Raw}$ [$CV_{Cortex} = 6.23\%$].

3.3.3. Donders

For the Donders dataset the total whole-cortex CVs were overall lower than in the NOR-MR750 and HCP datasets, with $T1w/T2w_{Raw}$ [$CV_{Cortex} = 8.56\%$], $T1w/T2w_{PVC}$ [$CV_{Cortex} = 8.55\%$], $T1w/T2w_{OC}$ [$CV_{Cortex} = 8.18\%$], $T1w/T2w_{N3}$ [$CV_{Cortex} = 8.05\%$], and $T1w/T2w_{N3\text{-}All}$ [$CV_{Cortex} = 7.65\%$]. Whole-cortex CVs were the highest with N4ITK bias correction; $T1w/T2w_{N4}$ [$CV_{Cortex} = 9.42\%$] and $T1w/T2w_{N4\text{-}All}$ [$CV_{Cortex} = 9.05\%$]. The lowest CVs were seen with $T1w/T2w_{HCP\text{-}Raw}$ [$CV_{Cortex} = 6.59\%$] and $T1w/T2w_{HCP\text{-}BC}$ [$CV_{Cortex} = 6.73\%$].

3.3.4. HCP

In the HCP dataset, total whole-cortex CVs were similar across most of the non-intensity normalised processing methods, with $T1w/T2w_{Raw}$ [$CV_{Cortex} = 10.41\%$], $T1w/T2w_{PVC}$ [$CV_{Cortex} = 10.45\%$], $T1w/T2w_{OC}$ [$CV_{Cortex} = 10.11\%$], $T1w/T2w_{N3}$ [$CV_{Cortex} = 10.17\%$], $T1w/T2w_{N3\text{-}All}$ [$CV_{Cortex} = 9.88\%$], $T1w/T2w_{N4}$ [$CV_{Cortex} = 10.70\%$], and $T1w/T2w_{N4\text{-}All}$ [$CV_{Cortex} = 10.43\%$]. Similar to the other datasets, the lowest CVs were seen with $T1w/T2w_{HCP\text{-}Raw}$ [$CV_{Cortex} = 4.88\%$] and $T1w/T2w_{HCP\text{-}BC}$ [$CV_{Cortex} = 4.32\%$].

3.3.5. Intensity normalisation

For intensity normalisation with the raw processing method, the greatest improvement of CV_{Cortex} across datasets compared to

Table 3
Summary statistics for $T1w/T2w_{Raw}$ after intensity normalisation, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for the median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		Raw - FCM	Raw - GMM	Raw - KDE	Raw - LSQ	Raw - N&U	Raw - RAVEL	Raw - WS	Raw - ZS
Mean T1w/T2w-ratio	NOR-MR750	0.53	0.58	0.57	0.94	0.69	0.84	0.84	0.97
	NOR-Premier	0.57	0.62	0.59	0.91	2.04	0.84	0.84	0.97
	DONDERS	0.59	0.60	0.69	0.79	1.41	0.87	0.86	0.97
	HCP	0.49	0.52	0.52	0.83	1.27	0.80	0.80	0.96
	Total (Mean)	0.55	0.58	0.59	0.87	1.35	0.84	0.83	0.97
Cofficient of variation (%)	NOR-MR750	8.32	8.29	8.48	8.35	17.83	5.94	6.04	5.61
	NOR-Premier	7.90	7.88	7.96	8.02	9.89	5.96	5.91	5.59
	DONDERS	7.47	7.80	10.64	7.43	13.21	5.40	5.50	4.52
	HCP	7.28	7.21	10.25	7.36	14.65	5.37	6.18	4.96
	Total (RMS)	7.75	7.81	9.40	7.80	14.18	5.68	5.91	5.19
Intraclass correlation coefficient	NOR-MR750	0.94	0.96	0.87	0.99	0.78	0.51	0.66	0.98
	NOR-Premier	0.91	0.93	0.89	0.99	0.97	0.76	0.69	0.97
	DONDERS	0.77	0.64	0.78	0.96	0.51	0.82	0.83	0.93
	HCP	0.86	0.88	0.53	0.96	0.85	0.51	0.40	0.93
	Total (Mean)	0.87	0.86	0.77	0.97	0.78	0.65	0.65	0.96
Laterality Index (%)	NOR-MR750	8.80	8.80	8.80	8.80	8.78	2.28	2.28	1.27
	NOR-Premier	1.93	1.93	1.93	1.93	1.92	0.64	0.64	0.36
	DONDERS	0.36	0.36	0.36	0.36	0.29	0.23	0.32	0.32
	HCP	−0.14	−0.07	−0.14	−0.14	−0.15	−0.01	0.01	−0.01
	Total (RMS)	4.51	4.51	4.51	4.51	4.49	1.19	1.20	0.68
Correlation with YA-BC	NOR-MR750	0.28	0.28	0.27	0.28	0.25	0.27	0.29	0.27
	NOR-Premier	0.88	0.88	0.88	0.89	0.89	0.86	0.90	0.87
	DONDERS	0.80	0.18	0.80	0.80	0.77	0.68	0.83	0.88
	HCP	0.90	0.90	0.90	0.90	0.89	0.88	0.89	0.88
	Total (Mean)	0.71	0.56	0.71	0.72	0.70	0.67	0.73	0.73
Correlation with YA-B1+	NOR-MR750	−0.04	−0.04	−0.04	−0.03	−0.06	−0.04	−0.02	−0.05
	NOR-Premier	0.88	0.88	0.88	0.87	0.87	0.88	0.88	0.77
	DONDERS	0.96	0.96	0.96	0.96	0.95	0.87	0.96	0.92
	HCP	0.72	0.72	0.72	0.72	0.71	0.71	0.73	0.72
	Total (Mean)	0.63	0.63	0.63	0.63	0.62	0.61	0.64	0.59

$T1w/T2w_{Raw}$ [CV_{Cortex} = 10.13%] was seen after intensity normalisation with WhiteStripe [CV_{Cortex} = 5.91%], RAVEL [CV_{Cortex} = 5.68%], and Z-Score [CV_{Cortex} = 5.19%]. For the other five intensity normalisation methods, some resulted in improvements while others made CVs worse with CV_{Cortex} ranging from 7.80% to 14.18% after intensity normalisation.

The best results for the N3-All processing method after intensity normalisation as compared with $T1w/T2w_{N3\text{-}All}$ [CV_{Cortex} = 9.56%] were seen with WhiteStripe [CV_{Cortex} = 5.42%], RAVEL [CV_{Cortex} = 5.46%], and Z-Score [CV_{Cortex} = 5.17%]. Similarly, the best results for the N4-All processing method after intensity normalisation as compared with $T1w/T2w_{N4\text{-}All}$ [CV_{Cortex} = 10.78%] were seen with WhiteStripe [CV_{Cortex} = 5.45%], RAVEL [CV_{Cortex} = 5.57%], and Z-Score [CV_{Cortex} = 5.17%].

3.4. Regional coefficients of variation maps

See Fig. 3 for CV_{ROI} for a selection of processing methods. Regional maps of CV_{ROI} for the HCP-MMP atlas, showed that the regions with the greatest CV_{ROI} tended to be located in the motor cortex across all datasets and in the occipital lobe for the NOR-Premier, Donders, and the HCP dataset. N3 bias correction lowered CV_{ROI} in the occipital lobes in the NOR-Premier and the HCP datasets. Occipital lobe CV_{ROI} remained high for $T1w/T2w_{HCP\text{-}BC}$ especially for the NOR-Premier and HCP datasets. For the Donders dataset, some frontal regions also showed relatively high CV_{ROI} particularly with $T1w/T2w_{HCP\text{-}Raw}$. Some regions in the rostral anterior cingulate showed high CV_{ROI} across multiple datasets and processing methods. This was likely caused by unstable parcellation performance in this region due to the proximity to the medial wall. See Supplementary Figure 6–9 for CV_{ROI} maps for $T1w/T2w_{N3\text{-}All}$ and $T1w/T2w_{N4\text{-}All}$ after intensity normalisation.

3.5. Intraclass correlation coefficients

Intraclass correlation coefficients for whole-cortex median T1w/T2w-ratio values were in general low for the NOR-MR750 dataset without intensity normalisation, ranging from 0.13 for $T1w/T2w_{Raw}$ to 0.41 for $T1w/T2w_{HCP\text{-}BC}$. For the NOR-Premier dataset, ICC_{Median} was high for all pipelines with a range of 0.90 to 0.93 except for $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ with an ICC_{Median} of 0.59 and 0.54 respectively. For the Donders dataset, ICC_{Median} ranged from 0.67 for $T1w/T2w_{HCP\text{-}BC}$ to 0.14 for $T1w/T2w_{HCP\text{-}Raw}$. For the HCP dataset, ICC_{Median} ranged from 0.55 for $T1w/T2w_{HCP\text{-}BC}$ to 0.74 for $T1w/T2w_{N4}$ and $T1w/T2w_{N4\text{-}All}$.

After intensity normalisation with both the Raw, N3-All, and N4-All processing methods, ICC_{Median} improved the most with Least-Squares and Z-Score normalisation. For N3-All with Nyúl & Udupa normalisation, ICC_{Median} was also high for NOR-Premier [ICC_{Median} = 0.98], NOR-MR750 [ICC_{Median} = 0.86] and HCP [ICC_{Median} = 0.89], but lower for Donders [ICC_{Median} = 0.49]. For N4-All with Nyúl & Udupa normalisation, ICC_{Median} was high for all four datasets. See Tables 2-5 for median ICCs for each dataset and processing method and Fig. 5 for box plots of ICCs of mean T1w/T2w-ratio values in each ROI of the HCP-MMP atlas. See Supplementary Figures 10–14 for regional ICCs of mean T1w/T2w-ratio values in each ROI of the HCP-MMP atlas.

3.6. Laterality indices

Laterality indices showed large left-right differences for the raw processing method for the NOR-MR750 dataset with $T1w/T2w_{Raw}$ [LI = 8.80%] compared to the NOR-Premier dataset with $T1w/T2w_{Raw}$ [LI = 1.93%], the Donders dataset with $T1w/T2w_{Raw}$ [LI = 0.36%], and the HCP dataset with $T1w/T2w_{Raw}$ [LI = −0.14%], indicating

Table 4
Summary statistics for $T1w/T2w_{N3\text{-}All}$ after intensity normalisation, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		N3-All - FCM	N3-All - GMM	N3-All - KDE	N3-All - LSQ	N3-All - N&U	N3-All - RAVEL	N3-All - WS	N3-All - ZS
Mean T1w/T2w-ratio	NOR-MR750	0.54	0.59	0.57	0.94	1.24	0.84	0.85	0.97
	NOR-Premier	0.57	0.62	0.60	0.88	1.96	0.85	0.85	0.97
	DONDERS	0.58	0.60	0.71	0.80	1.43	0.88	0.87	0.96
	HCP	0.49	0.52	0.52	0.81	1.26	0.81	0.81	0.96
	Total (Mean)	0.54	0.58	0.60	0.86	1.47	0.85	0.84	0.97
Cofficient of variation (%)	NOR-MR750	7.61	7.60	7.68	7.66	15.42	5.83	5.78	5.59
	NOR-Premier	7.30	7.30	7.40	7.54	8.71	5.91	5.74	5.58
	DONDERS	6.53	6.74	9.63	6.74	13.66	4.92	4.91	4.50
	HCP	6.53	6.46	6.52	6.55	12.22	5.10	5.19	4.95
	Total (RMS)	7.01	7.04	7.89	7.14	12.74	5.46	5.42	5.17
Intraclass correlation coefficient	NOR-MR750	0.95	0.97	0.93	0.99	0.86	0.65	0.78	0.99
	NOR-Premier	0.94	0.97	0.92	0.99	0.98	0.85	0.88	0.99
	DONDERS	0.75	0.70	0.79	0.95	0.49	0.84	0.84	0.96
	HCP	0.88	0.90	0.99	0.98	0.89	0.94	0.93	0.96
	Total (Mean)	0.88	0.89	0.91	0.98	0.80	0.82	0.86	0.97
Laterality Index (%)	NOR-MR750	−0.13	−0.13	−0.13	−0.13	−0.10	−0.17	−0.15	−0.02
	NOR-Premier	0.00	0.00	0.00	0.00	0.02	−0.06	0.00	0.01
	DONDERS	0.09	0.09	0.09	0.09	0.11	0.05	−0.05	−0.04
	HCP	−0.25	−0.24	−0.25	−0.25	−0.26	−0.10	−0.08	−0.05
	Total (RMS)	0.15	0.14	0.15	0.15	0.15	0.11	0.09	0.03
Correlation with YA-BC	NOR-MR750	0.70	0.70	0.70	0.71	0.69	0.71	0.70	0.69
	NOR-Premier	0.85	0.85	0.85	0.86	0.85	0.74	0.84	0.85
	DONDERS	0.76	0.76	0.76	0.76	0.75	0.60	0.78	0.77
	HCP	0.91	0.91	0.91	0.91	0.91	0.90	0.91	0.91
	Total (Mean)	0.81	0.81	0.81	0.81	0.80	0.74	0.81	0.80
Correlation with YA-B1+	NOR-MR750	0.37	0.37	0.37	0.37	0.35	0.38	0.36	0.35
	NOR-Premier	0.82	0.82	0.82	0.82	0.81	0.74	0.82	0.76
	DONDERS	0.87	0.87	0.87	0.87	0.86	0.72	0.85	0.85
	HCP	0.80	0.80	0.80	0.80	0.80	0.78	0.81	0.80
	Total (Mean)	0.71	0.71	0.71	0.72	0.71	0.65	0.71	0.69

large interhemispheric intensity bias in the NOR-MR750 dataset. Similarly, large laterality indices were seen in the NOR-MR750 dataset for $T1w/T2w_{PVC}$, $T1w/T2w_{OC}$, and $T1w/T2w_{HCP\text{-}Raw}$. Bias field correction had marginal effects on the laterality indices of the NOR-Premier, Donders, and HCP datasets, but reduced those of the NOR-MR750 dataset with $T1w/T2w_{N3}$ [LI = −0.20%], $T1w/T2w_{N3\text{-}All}$ [LI = −0.13%], $T1w/T2w_{N4}$ [LI = 0.90%], and $T1w/T2w_{N4\text{-}All}$ [LI = 1.01%]. Similarly, $T1w/T2w_{HCP\text{-}BC}$ decreased the laterality index to −0.03% for the NOR-MR750 dataset.

After intensity normalisation in the NOR-MR750 dataset using the Raw processing method, laterality indices decrased with RAVEL [LI = 2.28%], WhiteStripe [LI = 2.28%] and Z-Score [LI = 1.27%]. Otherwise laterality indices were only marginally affected by intensity normalisation using the Raw and N3-All processing methods. See Fig. 6 for box plots of per-subject laterality indices for processing method grouped by dataset.

3.7. Follow-up analyses

3.7.1. Analyses on high-resolution data

Spearman rank correlations with the reference datasets, ICC_{Median}, and laterality indices were similar to those observed in the main analyses on downsampled data for both the NOR-Premier and HCP datasets. Whole-cortex coefficients of variation, CV_{Cortex} were lower across datasets for $T1w/T2w_{Raw\text{-}Hires}$ [CV_{Cortex} = 7.02%] and $T1w/T2w_{HCP\text{-}Raw\text{-}Hires}$ [CV_{Cortex} = 4.71%], but higher for T1w/T2wHCP-BC-Hires [CV_{Cortex} = 9.98%]. The high CV_{Cortex} for T1w/T2wHCP-BC-Hires was driven by the NOR-Premier dataset which had a CV_{Cortex} of 13.60%. See Supplementary Table 1 for details.

3.7.2. Extraction directly from FS_LR space

Extracting ROI-wise T1w/T2w-ratio data directly from *FS_LR* space for $T1w/T2w_{HCP\text{-}Raw}$ and $T1w/T2w_{HCP\text{-}BC}$ had only marginal effects on the Spearman rank correlations with the reference datasets, with coefficients differing by less than 0.01 for all datasets. See Supplementary Table 2 for the results for each processing method and dataset.

4. Discussion

Two key findings emerged from our investigation. First, the performance of the T1w/T2w-ratio in faithfully reproducing myeloarchitectonic maps is highly variable across sites and processing pipelines but can in some cases be improved with bias correction. For example, we found that for the NOR-MR750 dataset, for which reproducibility was initially low, the correlation with the YA-BC reference improved from 0.27 for $T1w/T2w_{Raw}$ to 0.70 after data-driven N3 bias correction and to 0.92 with the template-adjusted HCP-BC pipeline. Similarly, in this dataset we found an increase in the correlation with the B1+ corrected YA-B1+ dataset from −0.05 to 0.66 after data-driven N4ITK bias correction and to 0.77 with the template-adjusted HCP-BC pipeline. The second main finding was that although the test-retest reliability of the raw T1w/T2w-ratio is poor, large improvements to reliability were achieved with the use of some intensity normalisation methods, whereas other methods resulted in lower test-retest reliability.

The poor reproducibility of regional distributions seen for some datasets and processing pipelines can be an obstacle to for the myeloarchitectonic parcellation of the cerebral cortex if not sufficiently accounted for, for example using N4ITK bias correction or the template-based correction method in the HCP-MPP pipeline. This is likely caused by B1+ field inhomogeneities that are not cancelled when taking the ra-

Table 5
Summary statistics for $T1w/T2w_{N4\text{-}All}$ after intensity normalisation, including mean T1w/T2w-ratio values, percentage coefficients of variation, intraclass correlation coefficients for median T1w/T2w-ratio value, whole-cortex percentage laterality indices averaged across individuals, and Spearman rank correlations with the YA-BC and YA-B1+ reference datasets.

		N4-All -FCM	N4-All -GMM	N4-All -KDE	N4-All - LSQ	N4-All - N&U	N4-All - RAVEL	N4-All - WS	N4-All - ZS
Mean T1w/T2w-ratio	NOR-MR750	0.60	0.66	0.64	1.09	3.08	0.85	0.85	0.97
	NOR-Premier	0.60	0.67	0.65	0.91	2.26	0.84	0.85	0.97
	DONDERS	0.60	0.62	0.77	0.82	1.67	0.88	0.87	0.96
	HCP	0.52	0.54	0.57	0.86	1.60	0.81	0.81	0.96
	Total (Mean)	0.58	0.62	0.65	0.92	2.15	0.85	0.85	0.97
Cofficient of variation (%)	NOR-MR750	7.59	7.60	7.72	7.77	15.73	6.20	5.85	5.59
	NOR-Premier	7.23	7.32	7.55	7.95	9.11	5.92	5.79	5.58
	DONDERS	6.26	6.52	7.09	6.59	10.88	4.90	4.75	4.50
	HCP	6.40	6.49	8.54	6.71	10.80	5.17	5.35	4.95
	Total (RMS)	6.89	7.00	7.74	7.28	11.89	5.57	5.45	5.17
Intraclass correlation coefficient	NOR-MR750	0.91	0.88	0.69	0.98	0.94	0.71	0.81	0.99
	NOR-Premier	0.94	0.92	0.78	0.98	0.97	0.76	0.86	0.99
	DONDERS	0.89	0.80	0.87	0.95	0.93	0.83	0.90	0.97
	HCP	0.89	0.86	0.85	0.96	0.96	0.83	0.86	0.96
	Total (Mean)	0.91	0.86	0.80	0.97	0.95	0.78	0.86	0.98
Laterality Index (%)	NOR-MR750	1.03	1.03	1.03	1.03	1.35	0.15	0.23	0.17
	NOR-Premier	0.07	0.07	0.07	0.07	0.09	0.00	0.04	0.03
	DONDERS	−0.38	−0.38	−0.38	−0.38	−0.81	−0.15	−0.11	−0.09
	HCP	−0.12	−0.11	−0.12	−0.12	−0.16	−0.12	−0.05	−0.04
	Total (RMS)	0.55	0.55	0.55	0.55	0.79	0.12	0.13	0.10
Correlation with YA-BC	NOR-MR750	0.58	0.58	0.58	0.58	0.49	0.46	0.56	0.61
	NOR-Premier	0.57	0.57	0.57	0.57	0.56	0.54	0.58	0.62
	DONDERS	0.47	0.47	0.47	0.47	0.46	0.35	0.51	0.49
	HCP	0.71	0.71	0.71	0.71	0.69	0.72	0.72	0.72
	Total (Mean)	0.58	0.58	0.58	0.58	0.55	0.52	0.59	0.61
Correlation with YA-B1+	NOR-MR750	0.70	0.70	0.70	0.71	0.65	0.60	0.69	0.69
	NOR-Premier	0.73	0.73	0.73	0.74	0.72	0.70	0.74	0.74
	DONDERS	0.63	0.63	0.63	0.63	0.60	0.50	0.66	0.65
	HCP	0.82	0.82	0.82	0.82	0.78	0.77	0.82	0.82
	Total (Mean)	0.72	0.72	0.72	0.72	0.69	0.64	0.73	0.72

tio. In these cases, N4ITK bias correction improves the correspondence with the B1+ adjusted reference dataset. Based on this, we propose that studies on the T1w/T2w-ratio should as a first step calculate correlations between the data and well-validated cortical myelin maps from the literature. We also recommend the use of on-scanner B1- field correction and dedicated sequences to calculate B1+ field maps in order to account for both receive and transmit field inhomogeneities. In the event that field maps are not available, our findings suggest that residual field inhomogeneities can be attenuated with data-driven bias correction where N4ITK bias correction is associated with the highest correspondence to B1+ adjusted reference data. However, care must be taken since bias correction can also reduce correspondence with the expected regional distributions depending on the intensity inhomogeneity profile of the dataset.

Low test-retest reliability is an obstacle for studies investigating group differences in the T1w/T2w-ratio or its associations with variables of interest. Some intensity normalisation methods improved test-retest reliability considerably, but others had a marginal effect and some even had the paradoxical effect of lowering reliability. Our results suggest that the best performance is given by the WhiteStripe and Z-score intensity normalisation methods.

4.1. The role of data acquisition: scanner and sequence

The best results were seen with the MPRAGE pulse sequence (NOR-Premier, Donders, HCP), independently of scanner vendor. This could indicate that MPRAGE is better suited than the BRAVO pulse sequence for cortical T1w/T2w-ratio mapping. Even so, by means of bias field correction and intensity normalisation, it was possible to achieve improved reproducibility of myeloarchitectonic distributions and test-retest reliability with BRAVO (NOR-MR750). As such, researchers using a different T1-weighted sequence than MPRAGE may achieve good T1w/T2w-ratio results with the use of appropriate post-hoc corrections in the processing pipeline.

4.2. Bias field correction

Given the susceptibility of the T1w/T2w-ratio to nonlinear field inhomogeneities, particularly those associated with the B1+ field, (Glasser et al., 2021; Glasser and Van Essen, 2011), we expected bias field correction to be one of the most influential correction factors. In line with this expectation, we found a major improvement in the correspondence between the NOR-MR750 dataset and the YA-BC dataset after N3 bias correction and a similar improvement in the correlation with the B1+ corrected YA-B1+ dataset after N4ITK bias correction. As the same type of on-scanner B1- field correction was performed in the NOR-MR750 dataset as for the NOR-Premier dataset, we ascribe these improvements to the reduction of field inhomogeneities caused by the B1+ field. However, we only saw minor improvements in the HCP dataset and in the NOR-Premier and Donders datasets the correspondence worsened with bias correction. This may indicate less initial field inhomogeneity due to scanner hardware differences or more effective field inhomogeneity correction at the image reconstruction stage in these datasets.

We found a greater improvement in the correlation between the NOR-MR750 dataset and the YA-BC reference after bias correction with N3 rather than N4ITK. For correlations with the new B1+ adjusted YA-B1+ reference, however, we found a greater improvement with N4ITK bias correction in the NOR-MR750 dataset. We observed moderately increased correlations between the HCP dataset and the YA-B1+ reference after both N3 and N4ITK bias correction, but for the other datasets both N3 and N4ITK bias correction lowered correlations. Given that the

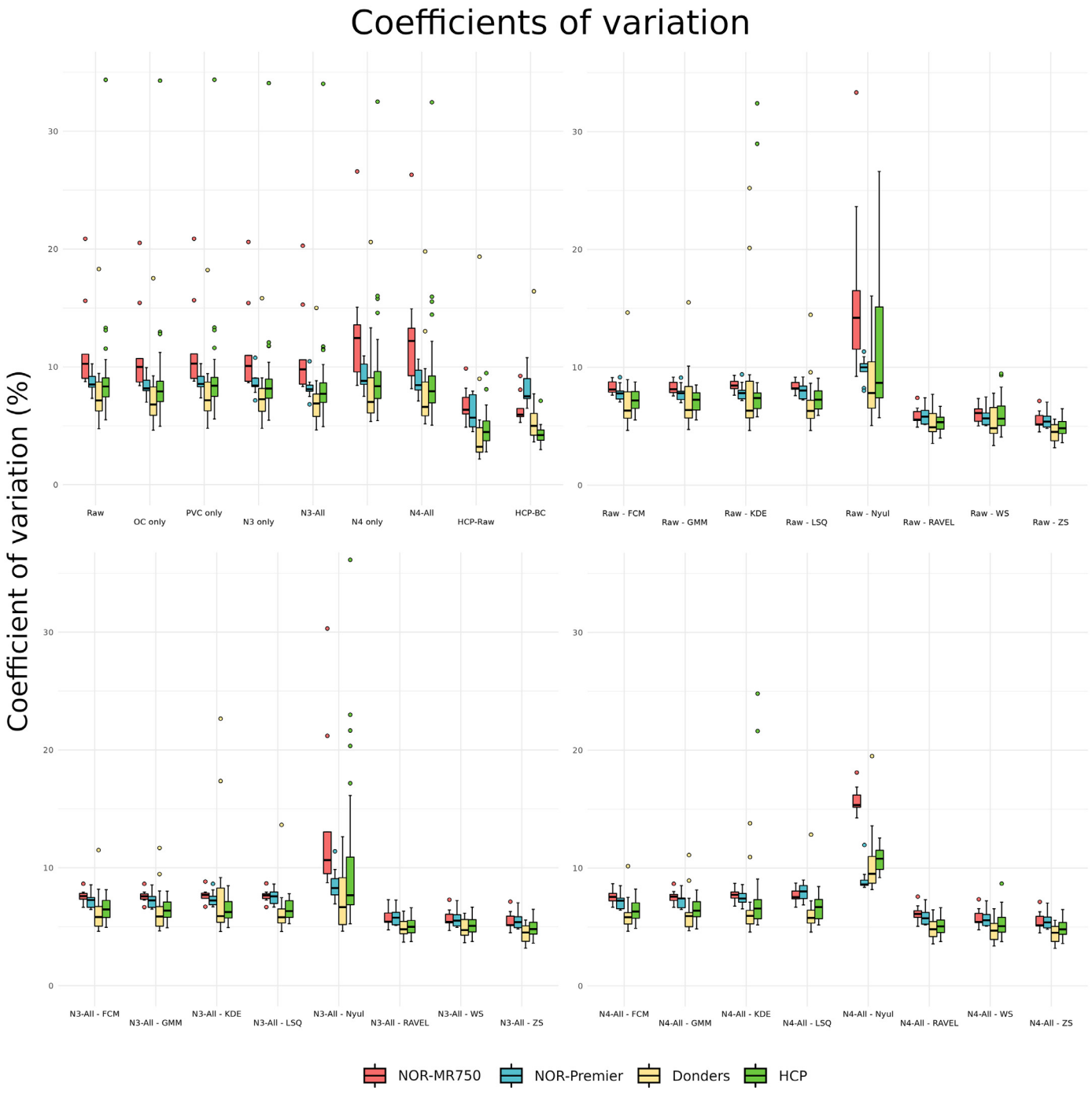

Fig. 3. CVs in each ROI of the HCP-MMP atlas for a selection of calculation methods without intensity normalisation.

YA-B1+ reference has been corrected for intensity inhomogeneities attributable to the B1+ field, we conclude that N4ITK bias correction provides the best data-driven bias correction approach of the two when residual field bias is present in the data. However, we caution against the use of either N3 or N4ITK bias correction in datasets with less intensity inhomogeneity, since bias correction may in this case worsen reproducibility. It is important that researchers seeking to use the T1w/T2w-ratio first compute correlations with well-validated datasets in order to assess the presence of intensity inhomogeneity in the data before deciding on the bias correction strategy.

The greatest reduction of the correlations with the YA-BC and YA-B1+ references in the NOR-Premier and Donders datasets was seen with N4ITK bias correction. The principal difference between the N3 and N4ITK algorithms is the B-spline smoothing strategy and the iterative optimisation approach. Whereas the N3 algorithm estimates the *total* bias field in each iteration, the N4ITK algorithm estimates the *residual* bias field from the corrected image from the previous iteration. While this has been considered to give better convergence properties for the algorithm, it may also be more aggressive than N3 bias correction. This could explain the tendency of the N4ITK algorithm to worsen correspondence with the reference datasets, especially in regions of the cortex where low T1w/T2w-ratio values were expected. It is also possible that the choice of N4ITK parameters negatively affected these results, but a thorough exploration of the N4ITK parameter space was outside the scope of the present study. Given the importance of bias field correction to the reproducibility of the NOR-MR750 dataset, and its paradoxical

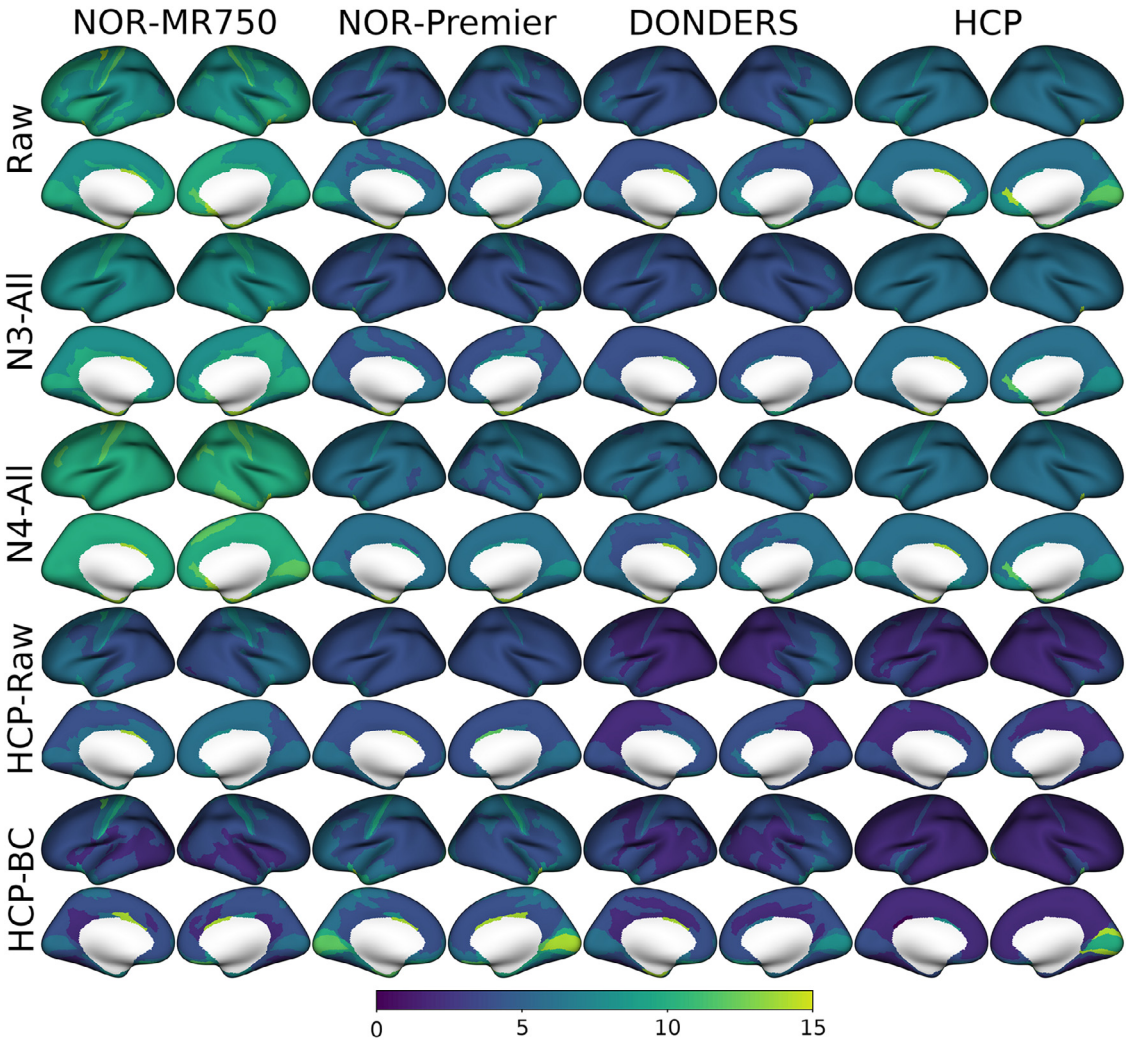

Fig. 4. Box plots of whole-cortex CVs for each participant grouped by processing method and dataset. Four outlier values (CVs > 35%) from two participants were not depicted for visualisation reasons. Their values were 38% and 46% for T1w/T2w$_{Raw}$ and 43% and 36% for T1w/T2w$_{N3\text{-}All}$ both after N&U intensity normalisation.

results of lowering correspondence with the expected T1w/T2w-ratio distributions in the NOR-Premier and Donders datasets, we encourage researchers to investigate these questions in more detail in future studies.

While T1w/T2w$_{HCP\text{-}Raw}$ neither improved the large laterality indices of the NOR-MR750 dataset nor the low correlations with the YA-BC or the YA-B1+ reference datasets, the highest correlations were seen with the T1w/T2w$_{HCP\text{-}BC}$ pipeline. The poor performance of the T1w/T2w$_{HCP\text{-}Raw}$ is likely due to the lack of bias field correction in this pipeline, with the result that residual field inhomogeneities, particularly those caused by the B1+ field, remain in the T1w/T2w$_{HCP\text{-}Raw}$ maps and consequently lead to poor correspondence with the reference datasets. In contrast, the HCP-BC pipeline employs a combined bias correction and intensity normalisation method using the smoothed difference between individual T1w/T2w-ratio maps and the population-average Conte69 template. This results in a non-linear correction of the T1w/T2w$_{HCP\text{-}Raw}$ maps where low-frequency local deviations from the Conte69 template are removed directly. Importantly, the YA-BC reference dataset was also created using the HCP-BC pipeline and it is therefore not surprising that the highest correlations to the YA-BC reference was seen with the HCP-BC pipeline. These factors complicate the comparison with the other bias correction methods, since both N3 and N4ITK bias correction are data-driven methods invoking few assumptions about the expected distributions of the final maps.

4.3. Test-retest reliability before intensity normalisation

We found overall low test-retest reliability of the T1w/T2w-ratio across all datasets and for most of the processing pipelines. The lowest whole-cortex CVs were achieved with T1w/T2w$_{HCP\text{-}Raw}$, while global reliability as measured with ICC$_{Median}$ remained poor. We expected lower test-retest reliability with longer scan-rescan interval. This was in line with the lower test-retest reliability of the HCP dataset (CV$_{Cortex}$ = 10.41% for T1w/T2w$_{Raw}$) which had a mean scan-rescan interval of 3.85 months compared to the Donders dataset which used similar acquisition parameters but acquired scan-rescan pairs on the same day and showed higher test-retest reliability (CV$_{Cortex}$ = 8.56% for T1w/T2w$_{Raw}$). However, we also found low test-retest reliability (12.30% for T1w/T2w$_{Raw}$) with the NOR-MR750 dataset where CV$_{Cortex}$ was calculated on the basis of same-day scan-rescan pairs in two sessions with only a two week scan-rescan interval.

Test-retest reliability did not improve after bias field correction and, even decreased after N4ITK bias correction in three of the datasets. This might indicate a drift of intensities during bias field correction, which could also be reflected in the increased means of T1w/T2w$_{N4}$ and T1w/T2w$_{N4\text{-}All}$ (Table 2) for two of the datasets. We also saw a broad range of ICC$_{Median}$, a measure of global test-retest agreement in the T1w/T2w-ratio maps, with values from 0.13 for the NOR-MR750 dataset to 0.90 for the NOR-Premier dataset. This variability may

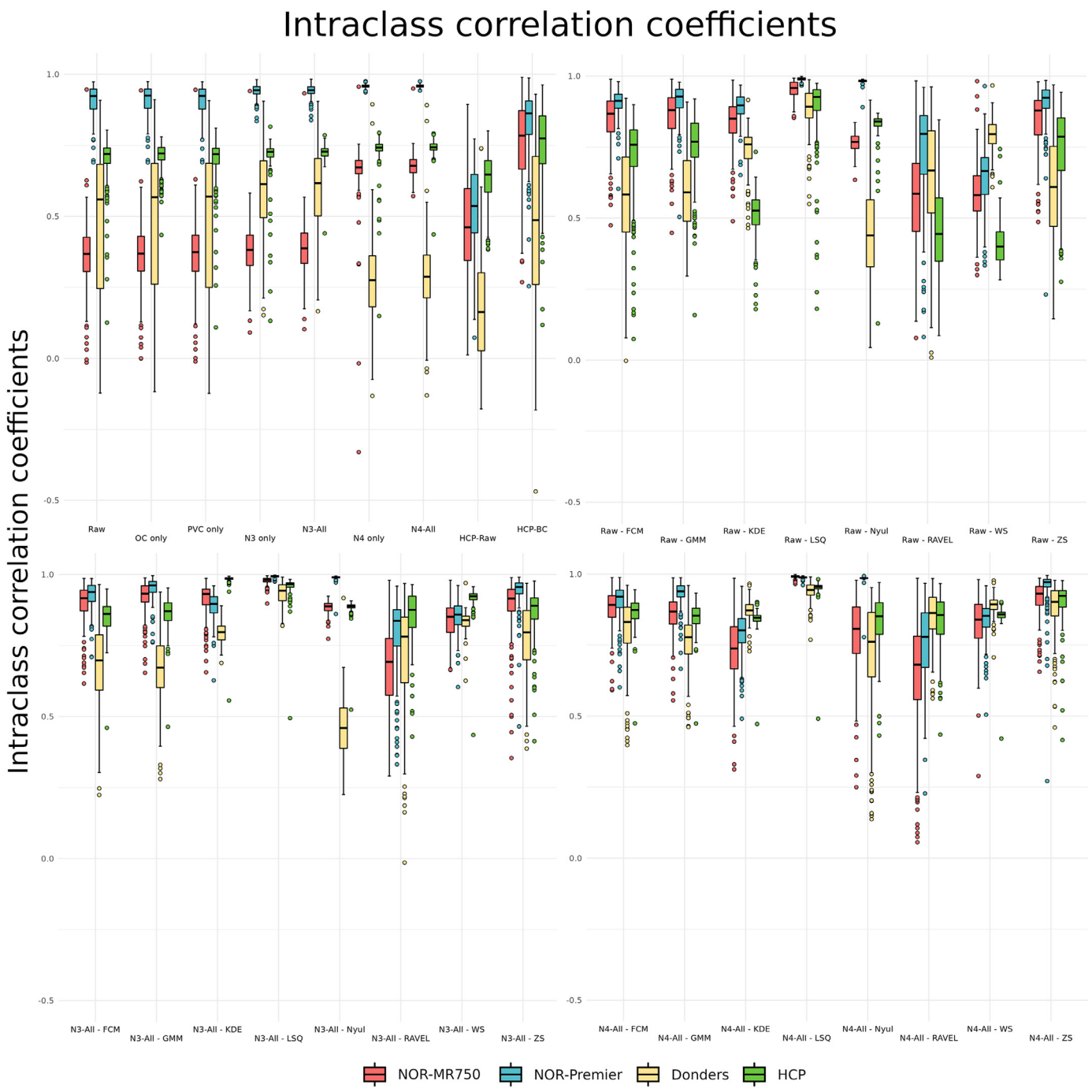

Fig. 5. Box plots of ICCs for each ROI grouped by processing method and dataset.

indicate scanner differences in individual transmit and receive gain settings determined during prescan. This may lead to global intensity drift that make *inter-subject* comparisons without intensity normalisation difficult if not impossible. It is noteworthy that some pipelines, such as T1w/T2w$_{\text{HCP-Raw}}$ decreased ICC$_{\text{Median}}$ in three of the datasets.

4.4. Test-retest reliability after intensity normalisation

There were major differences between intensity normalisation methods. Some methods, in particular WhiteStripe and Z-score, improved the test-retest reliability considerably relative to T1w/T2w$_{\text{Raw}}$. The similarity between WhiteStripe and RAVEL is likely due to WhiteStripe being performed as part of the RAVEL procedure. Despite improvements in the total whole-cortex CVs with RAVEL, ICC$_{\text{Median}}$ was low for the NOR-MR750 dataset at 0.71 for T1w/T2w$_{\text{N4-All}}$, whereas more consistent results were found with WhiteStripe normalisation. We observed the best numerical results with Z-Score normalisation, possibly the most straightforward method, yielding a total whole-cortex CV of 5.17% for the T1w/T2w$_{\text{N4-All}}$ pipeline. Some intensity normalisation procedures yielded poor results for test-retest reliability, in particular one previously proposed method, N&U (Nyúl and Udupa, 1999), led to total CVs across datasets of 14.18% when applied to T1w/T2w$_{\text{Raw}}$ which was higher than the CV of 10.13% for T1w/T2w$_{\text{Raw}}$.

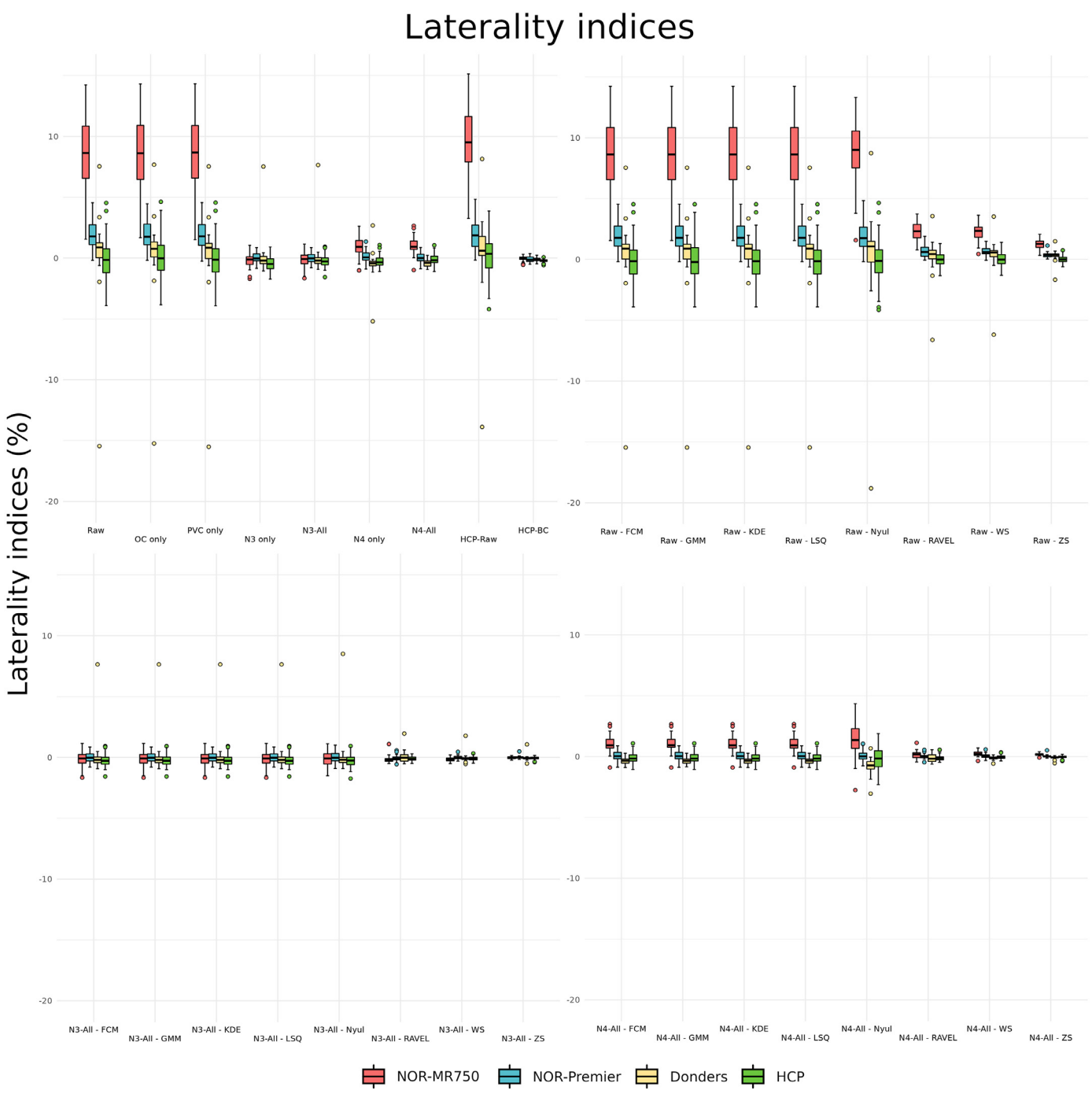

Fig. 6. Box plots of mean laterality indices for each participant grouped by processing method and dataset.

4.5. Outlier and partial volume corrections

Partial volume correction had only marginal effects on the reproducibility and reliability of T1w/T2w-ratio maps. Nevertheless, in large datasets the variable of interest may exert small but systematic partial volume effects, as in the ageing brain where the intensity contrast used to separate cortex and non-cortex is known to be affected. In such cases, partial voluming can become more influential (Shafee et al., 2015). Given these considerations, we decided to still include partial volume correction in the reliability tests. It is worth pointing out that the outlier correction method used in this study was used in the early versions of the HCP-MPP, but was later removed when it was found that submillimetre resolution alleviated the need for it due to fewer artefactual vertex-wise values caused by misregistration of the T1w and T2w images and surface reconstruction errors (Glasser et al., 2013). This is in line with our findings in the analyses on high-resolution data. Nonetheless, we considered outlier correction to still be relevant to evaluate in our study, given that we conformed images to 1 mm isotropic resolution to ensure generalisability of our findings to clinical studies where 1 mm resolution is still the norm. Furthermore, the unstable numerics of the T1w/T2w-ratio can yield extreme values whenever T2w image intensities are close to zero. Given that the vertex-wise method used in this study is resource intensive it may be advantageous to base outlier correction on individual T1w/T2w-ratio histograms in large datasets. Since we conformed images to 1 mm isotropic resolution, this correction may have been avoidable if high resolution images had been used.

Furthermore, since the vertex-wise method used in this study is resource intensive it may be advantageous to base outlier correction on individual T1w/T2w-ratio histograms in large datasets.

4.6. Impact of correction methods on the interpretation of results

Taken together, our findings point to the benefit of optimised T1w/T2w-ratio processing depending on the specific properties of the input data, but there are potential drawbacks, particularly for the interpretation of results of clinical analyses. Ideally, correction methods reduce both intensity bias and noise in the measurements, while retaining as much biological variation as possible. However, most intensity normalisation methods use signal variation in control regions to adjust variation in the region under study (e.g. the cerebral cortex) which introduces a dependence on intensity values in the control region. In our study, the two intensity normalisation methods with the best performance, WhiteStripe and Z-score normalisation, use variation in normal-appearing WM and the whole brain respectively, which may cause confounding by these control regions when they are used to adjust the intensities in the rest of the image.

As a practical example, consider a study of group differences in cortical T1w/T2w-ratio values in individuals who have a clinical condition where white matter myelination is also affected, for example multiple sclerosis, where normal-appearing WM has also been shown to be affected (Beer et al., 2016; Cooper et al., 2019; Granziera et al., 2021). In such cases confounding is likely and should be taken into account when the intensity normalisation method and control region is chosen. Otherwise, dependence on the control region may introduce spurious group differences in the cerebral cortex or obfuscate true effects. It is important to note here that it may not be possible to completely circumvent such limitations when employing relative non-dimensional measures such as weighted MRI intensities. This has been taken as an argument for the use of quantitative MRI techniques (Edwards et al., 2018; Weiskopf et al., 2021). Still, when such data are not available, our results suggest that intensity normalisation can improve test-retest reliability and facilitate inter-subject comparisons with conventional T1w and T2w sequences, with the caveat that researchers should examine their assumptions about the independence of their biological variables of interest with respect to control regions in order to substantiate their results.

Regarding bias field correction, it is less straightforward to predict how this affects the interpretation of results. For the N3 and N4ITK bias correction methods, low spatial frequencies are identified and removed from the image. While it is possible to constrain global intensity shifts, the dependence between local intensities within the cerebral cortex and other brain tissue is harder to assess and it is possible that individual variation in the region of interest is removed or that spurious variation is introduced with these correction methods. This issue might be addressed through the development of bias field correction methods that adjust intensities in the cerebral cortex only on the basis of the *shared* estimated field between the cerebral cortex and white matter. We encourage researchers looking to acquire data for studies on the T1w/T2w-ratio to perform on-scanner B1- field correction and to obtain dedicated B1+ field maps, for example using Actual Flip angle Imaging (Yarnykh, 2007). In the recent preprint from Glasser et al. (2021), such B1+ field maps were used to attenuate the effects of B1+ field inhomogeneity on the T1w/T2w-ratio. Interestingly, they also introduce a pseudo-transmit field correction method which relies on more commonly available spin echo and gradient echo sequences. Our results indicate that in situations where such field maps are not available, the N4ITK algorithm provides an effective alternative which can greatly improve the reproducibility of the regional distribution of the T1w/T2w-ratio values as expected from the myeloarchitectonic literature.

4.7. Impact of image resolution on results

We found that whole-cortex CVs improved when high-resolution data was used. This is in line with previous findings that higher resolution images are advantageous for T1w/T2w-ratio mapping (Glasser and Essen, 2011). This improvement is likely a result of the more accurate surface reconstruction afforded by the higher resolution data. We therefore advise that high-resolution data is used, insofar as it is possible, in studies on the T1w/T2w-ratio. However, we acknowledge that practical constraints may prevent the acquisition of high-resolution structural MRI. Importantly, longer acquisition times may increase the influence of head movement, which for clinical studies, where patients often tend to move more in the scanner than controls, can be a major confound (Reuter et al., 2015). For the mean T1w/T2w-ratios, ICCs, laterality indices, and correspondence with the reference T1w/T2w-ratio datasets, the effect of resolution was marginal. This suggests that while test-retest reliability may be improved with higher resolution data reproducibility and global T1w/T2w-ratio shifts are mainly driven by properties of the data other than resolution.

4.8. Strengths and limitations

Strengths of the present study include the use of test-retest datasets acquired on four different scanners across scanner models and vendors with a large total number of scan-rescan pairs in a test-retest context. The datasets that were included in this study are highly suited to address questions of reproducibility and reliability. We implemented a large variety of previously proposed processing methods within a standardised framework which allowed for the harmonised processing of each dataset. One such standardisation procedure was to downsample high resolution datasets in order to ensure the generalisability of our findings to typically available datasets with the more commonly used voxel resolution of 1 mm isotropic. We also included the HCP Minimal Processing Pipeline as a reference processing pipeline, which is widely used and considered state-of-the-art. Finally, each processing method was compared directly with two reference datasets in order to quantify the correspondence between T1w/T2w-ratio maps. This facilitates the generalisability of results and allows for direct comparison across processing pipelines.

As reviewed in the introduction, T1w and T2w image intensities are inherently non-quantitative as voxel intensity values do not represent direct and dimensionful measurements of biophysical properties. In this context, segmentation-based methods may be particularly powerful in providing measures that may be less sensitive to spurious inter-individual variation (Rowley et al., 2015; Viviani et al., 2017). In the present study, the focus was on the cortical T1w/T2w-ratio as an independent measure with the main focus on myeloarchitectonic parcellation and direct inter-subject comparisons. As such, questions regarding its use in segmentation-based approaches and correspondence with other measures of cortical myelin such as T1 relaxometry were not investigated. Future studies should investigate segmentation-based approaches and whether different T1w/T2w-ratio processing methods affect its validity for cortical myelin mapping through correlations with quantitative myelin measures. Notably, Shams et al. (2019) found high correlations between R1 maps and the T1w/T2w-ratio with the greatest deviations observed in regions where the B1+ field deviated the most from its nominal value. For segmentation and cortical reconstruction, we used FreeSurfer and the HCP-MPP, also based on FreeSurfer, as these are the most commonly used software suites for surface-based analyses and the cortical T1w/T2w-ratio. It is possible that other cortical segmentation methods would yield different results, but a systematic comparison of such methods was outside the scope of our study.

While the present study focused on the cortical T1w/T2w-ratio, an interesting question is how the different processing pipelines presented affect the whole-brain T1w/T2w-ratio, though the T1w/T2w-ratio in non-cortical regions may present its own set of challenges (Arshad et al., 2017; Hagiwara et al., 2018; Uddin et al., 2018, 2019). Another limitation of this study is the long scan-rescan interval of the HCP test-retest dataset (mean = 3.85 months; range = [2–5]). This was partially addressed by excluding those with scan-rescan interval greater than 6

months, but the scan-rescan interval remained greater than those of the other datasets. Finally, the datasets included in this study were composed of healthy individuals and it is possible that the outcomes might be different when applying these methods in clinical studies.

Conclusion

We recommend that future studies using the T1w/T2w-ratio for myeloarchitectonic parcellation assess the reproducibility of cortical myelin distributions by direct comparison with datasets that correspond closely with the expected myeloarchitectonic distributions, such as the YA-B1+ dataset, or atlases based on histological cortical myelin maps. For researchers planning to acquire data for studies on the T1w/T2w-ratio, we recommend that they carefully assess the scan-rescan stability of the head coil, perform on-scanner B1- field correction, and acquire scans suitable for estimating the B1+ field to correct for transmit field inhomogeneities. However, in the presence of field inhomogeneities obscuring the expected regional distributions of the T1w/T2w-ratio, bias field correction should be used and its performance quantified. Our findings suggest improved performance with the N4ITK algorithm, but this depends on the inhomogeneity profile of the specific dataset, in particular that of the B1+ field. We also found that high resolution (< 1 mm) data performed better than data downsampled to 1 mm resolution, and we recommend that images used for T1w/T2w-ratio mapping are acquired at high resolution whenever possible. We demonstrated that the test-retest reliability of the raw T1w/T2w-ratio is poor, which reduces the ability to test group differences or associations with clinical variables. Intensity normalisation methods can be used to improve reliability; the caveat being that choice of method may also affect the interpretation of results.

Data availability statement

The Donders dataset was previously used for a comparison of the T1w/T2w-ratio with MP2RAGE-based R1 mapping (Shams et al., 2019) and is available at the Donders Institute for Brain, Cognition and Behaviour repository (https://data.donders.ru.nl/). The HCP dataset is available from the ConnectomeDB (https://db.humanconnectome.org/). For the NOR-MR750 and the NOR-Premier datasets, access to the T1w and T2w images can be arranged by contacting the corresponding author and filling out a standard data sharing agreement, including a study description, given that the necessary ethics permissions have been secured.

Declaration of Competing Interest

OAA has received a speaker honorarium from Lundbeck and is a consultant for HealthLytix. Otherwise, no other declarations of interest.

Credit authorship contribution statement

Stener Nerland: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Funding acquisition. **Kjetil N. Jørgensen:** Conceptualization, Methodology, Writing – original draft. **Wibeke Nordhøy:** Investigation, Writing – original draft, Supervision. **Ivan I. Maximov:** Writing – original draft, Supervision. **Robin A.B. Bugge:** Investigation, Writing – original draft. **Lars T. Westlye:** Writing – original draft. **Ole A. Andreassen:** Writing – original draft, Resources. **Oliver M. Geier:** Writing – original draft, Supervision, Project administration. **Ingrid Agartz:** Resources, Writing – original draft, Supervision, Project administration, Funding acquisition.

Acknowledgements

This work was funded by the South-Eastern Norway Regional Health Authority; Grant No. 2017-097 and 2019-104, and the Research Council of Norway; Grant No. 223273. The Donders dataset was provided by The Donders Institute for Brain, Cognition and Behaviour (https://www.ru.nl/donders/) supported by the European FP7 program - Advanced Brain Imaging with MRI (ABRIM) and additional unprocessed data was kindly provided by Zahra Shams. The HCP dataset was provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell center for Systems Neuroscience at Washington University.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118709.

References

Arshad, M., Stanley, J.A., Raz, N., 2017. Test–retest reliability and concurrent validity of in vivo myelin content indices: myelin water fraction and calibrated T1w/T2w image ratio. Hum. Brain Mapp. 38 (4), 1780–1790. doi:10.1002/hbm.23481.

Beer, A., Biberacher, V., Schmidt, P., Righart, R., Buck, D., Berthele, A., Kirschke, J., Zimmer, C., Hemmer, B., Mühlau, M., 2016. Tissue damage within normal appearing white matter in early multiple sclerosis: assessment by the ratio of T1- and T2-weighted MR image intensity. J. Neurol. 263 (8), 1495–1502. doi:10.1007/s00415-016-8156-6.

Chen, W., Giger, M.L., 2004. A fuzzy c-means (FCM) based algorithm for intensity inhomogeneity correction and segmentation of MR images. In: 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821), 2, pp. 1307–1310. doi:10.1109/ISBI.2004.1398786.

Clark, V.P., Courchesne, E., Grafe, M., 1992. In vivo myeloarchitectonic analysis of human striate and extrastriate cortex using magnetic resonance imaging. Cereb. Cortex 2 (5), 417–424. doi:10.1093/cercor/2.5.417.

Cooper, G., Finke, C., Chien, C., Brandt, A.U., Asseyer, S., Ruprecht, K., Bellmann-Strobl, J., Paul, F., Scheel, M., 2019. Standardization of T1w/T2w ratio improves detection of tissue damage in multiple sclerosis. Front. Neurol. 10. doi:10.3389/fneur.2019.00334.

de Boor, C., 1972. On calculating with B-splines. J. Approx. Theory 6 (1), 50–62. doi:10.1016/0021-9045(72)90080-9.

Edwards, L.J., Kirilina, E., Mohammadi, S., Weiskopf, N., 2018. Microstructural imaging of human neocortex in vivo. Neuroimage 182, 184–206. doi:10.1016/j.neuroimage.2018.02.055.

Eickhoff, S., Walters, N.B., Schleicher, A., Kril, J., Egan, G.F., Zilles, K., Watson, J.D.G., Amunts, K., 2004. High-resolution MRI reflects myeloarchitecture and cytoarchitecture of human cerebral cortex. Hum. Brain Mapp. 24 (3), 206–215. doi:10.1002/hbm.20082.

Fortin, J.-P., Sweeney, E.M., Muschelli, J., Crainiceanu, C.M., Shinohara, R.T., 2016. Removing inter-subject technical variability in magnetic resonance imaging studies. Neuroimage 132, 198–212. doi:10.1016/j.neuroimage.2016.02.036.

Ganzetti, M., Wenderoth, N., Mantini, D., 2014. Whole brain myelin mapping using T1- and T2-weighted MR imaging data. Front. Hum. Neurosci. 8. doi:10.3389/fnhum.2014.00671.

Ghassemi, R., Brown, R., Narayanan, S., Banwell, B., Nakamura, K., Arnold, D.L., 2015. Normalization of white matter intensity on T1-weighted images of patients with acquired central nervous system demyelination. J. Neuroimaging 25 (2), 184–190. doi:10.1111/jon.12129.

Glasser, M.F., Coalson, T.S., Harms, M.P., Baum, G.L., Autio, J.A., Auerbach, E.J., Xu, J., Greve, D.N., Yacoub, E., Essen, D.C.V., Bock, N.A., & Hayashi, T. (2021). *Transmit field bias correction of T1w/T2w Myelin maps* (p. 2021.08.08.455570). https://doi.org/10.1101/2021.08.08.455570

Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D.C., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536 (7615), 171–178. doi:10.1038/nature18933.

Glasser, M.F., Essen, D.C.V., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. J. Neurosci. 31 (32), 11597–11616. doi:10.1523/JNEUROSCI.2180-11.2011.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal pre-processing pipelines for the Human Connectome Project. Neuroimage 80, 105–124. doi:10.1016/j.neuroimage.2013.04.127.

Glüer, C.-.C., Blake, G., Lu, Y., Blunt1, B.A., Jergas1, M., Genant1, H.K., 1995. Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. Osteoporos. Int. 5 (4), 262–270. doi:10.1007/BF01774016.

Granziera, C., Wuerfel, J., Barkhof, F., Calabrese, M., De Stefano, N., Enzinger, C., Evangelou, N., Filippi, M., Geurts, J.J.G., Reich, D.S., Rocca, M.A., Ropele, S., Rovira, À., Sati, P., Toosy, A.T., Vrenken, H., Gandini Wheeler-Kingshott, C.A.M., Kappos, L.the MAGNIMS Study Group, 2021. Quantitative magnetic resonance imaging towards clinical application in multiple sclerosis. Brain 144 (5), 1296–1311. doi:10.1093/brain/awab029.

Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48 (1), 63–72. doi:10.1016/j.neuroimage.2009.06.060.

Grydeland, H., Vértes, P.E., Váša, F., Romero-Garcia, R., Whitaker, K., Alexander-Bloch, A.F., Bjørnerud, A., Patel, A.X., Sederevičius, D., Tamnes, C.K., Westlye, L.T., White, S.R., Walhovd, K.B., Fjell, A.M., Bullmore, E.T., 2019. Waves of maturation and senescence in micro-structural MRI markers of human cortical myelination over the lifespan. Cereb. Cortex 29 (3), 1369–1381. doi:10.1093/cercor/bhy330.

Grydeland, H., Walhovd, K.B., Tamnes, C.K., Westlye, L.T., Fjell, A.M., 2013. Intracortical myelin links with performance variability across the human lifespan: results from T1- and T2-weighted MRI myelin mapping and diffusion tensor imaging. J. Neurosci. 33 (47), 18618–18630. doi:10.1523/JNEUROSCI.2811-13.2013.

Hagiwara, A., Hori, M., Kamagata, K., Warntjes, M., Matsuyoshi, D., Nakazawa, M., Ueda, R., Andica, C., Koshino, S., Maekawa, T., Irie, R., Takamura, T., Kumamaru, K.K., Abe, O., Aoki, S., 2018. Myelin measurement: comparison between simultaneous tissue relaxometry, magnetization transfer saturation index, and T1w/T2w ratio methods. Sci. Rep. 8 (1), 10554. doi:10.1038/s41598-018-28852-6.

Ishida, T., Donishi, T., Iwatani, J., Yamada, S., Takahashi, S., Ukai, S., Shinosaki, K., Terada, M., Kaneoke, Y., 2017. Elucidating the aberrant brain regions in bipolar disorder using T1-weighted/T2-weighted magnetic resonance ratio images. Psychiatry Res.: Neuroimaging 263, 76–84. doi:10.1016/j.pscychresns.2017.03.006.

Iwatani, J., Ishida, T., Donishi, T., Ukai, S., Shinosaki, K., Terada, M., Kaneoke, Y., 2015. Use of T1-weighted/T2-weighted magnetic resonance ratio images to elucidate changes in the schizophrenic brain. Brain Behav. 5 (10). doi:10.1002/brb3.399.

Koenig, S.H., 1991. Cholesterol of myelin is the determinant of gray-white contrast in MRI of brain. Magn. Reson. Med. 20 (2), 285–291. doi:10.1002/mrm.1910200210.

Koenig, S.H., Brown, R.D., Spiller, M., Lundbom, N., 1990. Relaxometry of brain: why white matter appears bright in MRI. Magn. Reson. Med. 14 (3), 482–495. doi:10.1002/mrm.1910140306.

Matheson, G.J., 2019. We need to talk about reliability: making better use of test-retest studies for study design and interpretation. PeerJ 7. doi:10.7717/peerj.6918.

McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychol. Methods 1 (1), 30–46. doi:10.1037/1082-989X.1.1.30.

Misaki, M., Savitz, J., Zotev, V., Phillips, R., Yuan, H., Young, K.D., Drevets, W.C., Bodurka, J., 2015. Contrast enhancement by combining T1- and T2-weighted structural brain MR images. Magn. Reson. Med. 74 (6), 1609–1620. doi:10.1002/mrm.25560.

Nakamura, K., Chen, J.T., Ontaneda, D., Fox, R.J., Trapp, B.D., 2017. T1-/T2-weighted ratio differs in demyelinated cortex in multiple sclerosis. Ann. Neurol. 82 (4), 635–639. doi:10.1002/ana.25019.

Nieuwenhuys, R., Broere, C.A.J., 2017. A map of the human neocortex showing the estimated overall myelin content of the individual architectonic areas based on the studies of Adolf Hopf. Brain Struct. Funct. 222 (1), 465–480. doi:10.1007/s00429-016-1228-7.

Norbom, L.B., Rokicki, J., Alnæs, D., Kaufmann, T., Doan, N.T., Andreassen, O.A., Westlye, L.T., Tamnes, C.K., 2020. Maturation of cortical microstructure and cognitive development in childhood and adolescence: a T1w/T2w ratio MRI study. Hum. Brain Mapp. 41 (16), 4676–4690. doi:10.1002/hbm.25149.

Nyúl, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale. Magn. Reson. Med. 42 (6), 1072–1081 10.1002/(SICI)1522-2594(199912)42:6<1072::AID–MRM11>3.0.CO;2-M.

Pelkmans, W., Dicks, E., Barkhof, F., Vrenken, H., Scheltens, P., van der Flier, W.M., Tijms, B.M., 2019. Gray matter T1-w/T2-w ratios are higher in Alzheimer's disease. Hum. Brain Mapp. 40 (13), 3900–3909. doi:10.1002/hbm.24638.

Preziosa, P., Bouman, P.M., Kiljan, S., Steenwijk, M.D., Meani, A., Pouwels, P.J., Rocca, M.A., Filippi, M., Geurts, J.J.G., Jonkman, L.E., 2021. Neurite density explains cortical T1-weighted/T2-weighted ratio in multiple sclerosis. J. Neurol., Neurosurg. Psychiatry 92 (7), 790–792. doi:10.1136/jnnp-2020-324391.

Reinhold, J.C., Dewey, B.E., Carass, A., Prince, J.L., 2019. Evaluating the impact of intensity normalization on MR image synthesis. In: Proceedings of SPIE–the International Society for Optical Engineering, p. 10949. doi:10.1117/12.2513089.

Reuter, M., Tisdall, M.D., Qureshi, A., Buckner, R.L., van der Kouwe, A.J.W., Fischl, B., 2015. Head motion during MRI acquisition reduces gray matter volume and thickness estimates. Neuroimage 107, 107–115. doi:10.1016/j.neuroimage.2014.12.006.

Righart, R., Biberacher, V., Jonkman, L.E., Klaver, R., Schmidt, P., Buck, D., Berthele, A., Kirschke, J.S., Zimmer, C., Hemmer, B., Geurts, J.J.G., Mühlau, M., 2017. Cortical pathology in multiple sclerosis detected by the T1/T2-weighted ratio from routine magnetic resonance imaging. Ann. Neurol. 82 (4), 519–529. doi:10.1002/ana.25020.

Rokicki, J., Wolfers, T., Nordhøy, W., Tesli, N., Quintana, D.S., Alnæs, D., Richard, G., de Lange, A.G., Lund, M.J., Norbom, L., Agartz, I., Melle, I., Nærland, T., Selbæk, G., Persson, K., Nordvik, J.E., Schwarz, E., Andreassen, O.A., Kaufmann, T., Westlye, L.T., 2020. Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. Hum. Brain Mapp. 42 (6), 1714–1726. doi:10.1002/hbm.25323.

Rowley, C.D., Bazin, P.-L., Tardif, C.L., Sehmbi, M., Hashim, E., Zaharieva, N., Minuzzi, L., Frey, B.N., Bock, N.A., 2015. Assessing intracortical myelin in the living human brain using myelinated cortical thickness. Front. Neurosci. 9, 396. doi:10.3389/fnins.2015.00396.

Rowley, C.D., Tabrizi, S.J., Scahill, R.I., Leavitt, B.R., Roos, R.A.C., Durr, A., Bock, N.A., 2018. Altered intracortical T1-weighted/T2-weighted ratio signal in Huntington's disease. Front. Neurosci. 12. doi:10.3389/fnins.2018.00805.

Sereno, M.I., Lutti, A., Weiskopf, N., Dick, F., 2013. Mapping the human cortical surface by combining quantitative T1 with retinotopy. Cereb. Cortex (New York, NY) 23 (9), 2261–2268. doi:10.1093/cercor/bhs213.

Shafee, R., Buckner, R.L., Fischl, B., 2015. Gray matter myelination of 1555 human brains using partial volume corrected MRI images. Neuroimage 105, 473–485. doi:10.1016/j.neuroimage.2014.10.054.

Shah, M., Xiao, Y., Subbanna, N., Francis, S., Arnold, D.L., Collins, D.L., Arbel, T., 2011. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. Med. Image Anal. 15 (2), 267–282. doi:10.1016/j.media.2010.12.003.

Shams, Z., Norris, D.G., Marques, J.P., 2019. A comparison of in vivo MRI based cortical myelin mapping using T1w/T2w and R1 mapping at 3T. PLoS ONE 14 (7). doi:10.1371/journal.pone.0218089.

Shinohara, R.T., Sweeney, E.M., Goldsmith, J., Shiee, N., Mateen, F.J., Calabresi, P.A., Jarso, S., Pham, D.L., Reich, D.S., Crainiceanu, C.M., 2014. Statistical normalization techniques for magnetic resonance imaging. NeuroImage : Clin. 6, 9–19. doi:10.1016/j.nicl.2014.08.008.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17 (1), 87–97. doi:10.1109/42.668698.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29 (6), 1310–1320. doi:10.1109/TMI.2010.2046908.

Tzourio-Mazoyer, N., Maingault, S., Panzieri, J., Pepe, A., Crivello, F., Mazoyer, B., 2019. Intracortical myelination of Heschl's gyrus and the planum temporale varies with heschl's duplication pattern and rhyming performance: an investigation of 440 healthy volunteers. Cereb. Cortex 29 (5), 2072–2083. doi:10.1093/cercor/bhy088.

Uddin, M.N., Figley, T.D., Marrie, R.A., Figley, C.R., 2018. Can T1w/T2w ratio be used as a myelin-specific measure in subcortical structures? Comparisons between FSE-based T1w/T2w ratios, GRASE-based T1w/T2w ratios and multi-echo GRASE-based myelin water fractions. NMR Biomed. 31 (3), e3868. doi:10.1002/nbm.3868.

Uddin, M.N., Figley, T.D., Solar, K.G., Shatil, A.S., Figley, C.R., 2019. Comparisons between multi-component myelin water fraction, T1w/T2w ratio, and diffusion tensor imaging measures in healthy human brain structures. Sci. Rep. 9 (1), 2500. doi:10.1038/s41598-019-39199-x.

Uğurbil, K., Xu, J., Auerbach, E.J., Moeller, S., Vu, A., Duarte-Carvajalino, J.M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P.F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D.A., Smith, S.M., … Yacoub, E., 2013. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. Neuroimage 80, 80–104. doi:10.1016/j.neuroimage.2013.05.012.

Valk, S.L., Xu, T., Margulies, D.S., Masouleh, S.K., Paquola, C., Goulas, A., Kochunov, P., Smallwood, J., Yeo, B.T.T., Bernhardt, B.C., Eickhoff, S.B., 2020. Shaping brain structure: genetic and phylogenetic axes of macroscale organization of cortical thickness. Sci. Adv. 6 (39). doi:10.1126/sciadv.abb3417.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn human connectome project: an overview. Neuroimage 80, 62–79. doi:10.1016/j.neuroimage.2013.05.041.

Viviani, R., Stöcker, T., Stingl, J.C., 2017. Multimodal FLAIR/MPRAGE segmentation of cerebral cortex and cortical myelin. Neuroimage 152, 130–141. doi:10.1016/j.neuroimage.2017.02.054.

Waehnert, M.D., Dinse, J., Weiss, M., Streicher, M.N., Waehnert, P., Geyer, S., Turner, R., Bazin, P.-L., 2014. Anatomically motivated modeling of cortical laminae. Neuroimage 93, 210–220. doi:10.1016/j.neuroimage.2013.03.078.

Walters, N.B., Egan, G.F., Kril, J.J., Kean, M., Waley, P., Jenkinson, M., Watson, J.D.G., 2003. In vivo identification of human cortical areas using high-resolution MRI: an approach to cerebral structure–function correlation. In: Proceedings of the National Academy of Sciences of the United States of America, 100, pp. 2981–2986. doi:10.1073/pnas.0437896100.

Wang, X.-J., 2020. Macroscopic gradients of synaptic excitation and inhibition in the neocortex. Nat. Rev. Neurosci. 21 (3), 169–178. doi:10.1038/s41583-020-0262-x.

Weiskopf, N., Edwards, L.J., Helms, G., Mohammadi, S., Kirilina, E., 2021. Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. Nat. Rev. Phys. 3 (8), 570–588. doi:10.1038/s42254-021-00326-1.

Yarnykh, V.L., 2007. Actual flip-angle imaging in the pulsed steady state: a method for rapid three-dimensional mapping of the transmitted radiofrequency field. Magn. Reson. Med. 57 (1), 192–200. doi:10.1002/mrm.21120.

Wang, L. (2021). geodesic (https://www.mathworks.com/matlabcentral/fileexchange/6522-geodesic), MATLAB central file exchange. Retrieved February 13, 2021.