NeuroImage 207 (2020) 116364

Contents lists available at ScienceDirect

NeuroImage

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

How stable is quantitative MRI? – Assessment of intra- and inter-scanner-model reproducibility using identical acquisition sequences and data analysis programs

René-Maxime Gracien [a,c,*], Michelle Maiworm [a,b,c], Nadine Brüche [c], Manoj Shrestha [c], Ulrike Nöth [c], Elke Hattingen [b], Marlies Wagner [b], Ralf Deichmann [c]

[a] Department of Neurology, Goethe University, Frankfurt/Main, Germany
[b] Department of Neuroradiology, Goethe University, Frankfurt/Main, Germany
[c] Brain Imaging Center, Goethe University, Frankfurt/Main, Germany

ARTICLE INFO

Keywords:
Quantitative MRI
Relaxometry
Proton density
Reproducibility
Scanner models

ABSTRACT

Background: Quantitative MRI (qMRI) techniques allow assessing cerebral tissue properties. However, previous studies on the accuracy of quantitative T1 and T2 mapping reported a scanner model bias of up to 10% for T1 and up to 23% for T2. Such differences would render multi-centre qMRI studies difficult and raise fundamental questions about the general precision of qMRI. A problem in previous studies was that different methods were used for qMRI parameter mapping or for measuring the transmitted radio frequency field B1 which is critical for qMRI techniques requiring corrections for B1 non-uniformities.

Aims: The goal was to assess the intra- and inter-scanner reproducibility of qMRI data at 3 T, using two different scanner models from the same vendor with exactly the same multiparametric acquisition protocol.

Methods: Proton density (PD), T1, T2* and T2 mapping was performed on healthy subjects and on a phantom, performing each measurement twice for each of two scanner models. Although the scanners had different hardware and software versions, identical imaging sequences were used for PD, T1 and T2* mapping, adapting the codes of an existing protocol on the older system line by line to match the software version of the newer scanner. For T2-mapping, the respective manufacturer's sequence was used which depended on the software version. However, system-dependent corrections were carried out in this case. Reproducibility was assessed by average values in regions of interest.

Results: Mean scan-rescan variations were not exceeding 2.14%, with average values of 1.23% and 1.56% for the new and old system, respectively. Inter-scanner model deviations were not exceeding 5.21% with average values of about 2.2–3.8% for PD, 2.5–3.0% for T2*, 1.6–3.1% for T1 and 3.3–5.2% for T2.

Conclusions: Provided that identical acquisition sequences are used, discrepancies between qMRI data acquired with different scanner models are low. The level of systematic differences reported in this work may help to interpret multi-centre data.

1. Introduction

Magnetic resonance imaging (MRI) has become indispensable in diagnostics and research. However, conventional MRI datasets show mixed contrasts which depend on hardware and various protocol parameters (Cercignani et al., 2018). In contrast, quantitative MRI (qMRI) measures tissue parameters such as the T1, T2 and T2* relaxation times and the proton density (PD). These parameter maps can provide information about diffuse or inconspicuous changes in cerebral tissue (Cercignani et al., 2018; Gracien et al., 2016a, 2017; Hattingen et al., 2015). A higher inter-site reproducibility has been observed for qMRI than for conventional MRI techniques (Weiskopf et al., 2013). Calculation of qMRI parameter maps requires corrections for hardware effects (Baudrexel et al., 2009; Volz et al., 2012a, 2012b). Ideally, hardware influences are completely eliminated, so qMRI results obtained on different scanner models are identical.

In the past, several studies assessed the reproducibility of qMRI mapping results across different scanners and/or sites. A study comprising three imaging centres, applying identical T1/T2 mapping

* Corresponding author. Brain Imaging Center, University Hospital, Schleusenweg 2-16, 60528 Frankfurt, Germany.
E-mail address: Rene-Maxime.Gracien@kgu.de (R.-M. Gracien).

https://doi.org/10.1016/j.neuroimage.2019.116364
Received 9 August 2019; Received in revised form 12 November 2019; Accepted 13 November 2019
Available online 15 November 2019

Abbreviations

2D	two-dimensional
3D	three-dimensional
BW	bandwidth
CoV	coefficient of variation
CSF	cerebrospinal fluid
EPI	echo planar imaging
FoV	field of view
GE	gradient echo
GM	gray matter
IR	inversion recovery
MRI	magnetic resonance imaging
MT	magnetization transfer
PD	proton density
PE	phase encoding
PVE	partial volume estimate
qMRI	quantitative MRI
RF	radiofrequency
ROI	region(s) of interest
PV	partial volume
SE	spin echo
SNR	signal-to-noise ratio
TSE	turbo spin echo
VFA	variable flip angle
WM	white matter

protocols on 1.5 T scanners from different vendors, reported an intra-site reproducibility of 6.4% for T1 and 7.9% for T2 (both values referring to the coefficient of variation, CoV) and an inter-site reproducibility of 6.8% for T1 and 8.3% for T2 (Deoni et al., 2008).

In a multi-centre study on Alzheimer's disease, T2 values obtained with 1.5 T scanners from three different vendors were compared, using comparable techniques but with slightly different acquisition parameters, in particular varying the effective TE (Bauer et al., 2010): a significant vendor bias of up to 20–30 ms was observed. Another study compared qMRI values (T1, T2*, PD and magnetization transfer ratio) acquired with the same scanner model at different sites (Weiskopf et al., 2013). The CoV was less than 8%. Furthermore, the inter-site bias (referring to systematic offsets) did not exceed 3.1% for any parameter. In a more recent study (Lee et al., 2019), identical T1 mapping protocols, albeit with different methods for the correction of inhomogeneities of the transmitted radiofrequency (RF) field B1, were implemented on 3 T scanners from two different manufacturers, assessing the intra-scanner and inter-vendor reproducibility of results. While the scan-rescan bias with the same scanner was low (median around 1%), the inter-vendor bias attained values of up to 10%. A further recent study (Buonincontri et al., 2019) assessed the multi-site repeatability of an MR fingerprinting method for T1 and T2 mapping which had been implemented on 1.5 and 3 T systems from the same vendor. The test/retest repeatability was found to be 2–3% for T1 and 5–8% for T2, the inter-site reproducibility was 3–8% for T1 and 8–14% for T2.

The goal of this study was to assess the stability of qMRI values acquired with two different scanner models (both 3 T) from the same vendor, but based on substantially different hardware and software versions, provided the single acquisition techniques are carefully replicated, so exactly the same acquisition protocol can be used. This question, which is essential for example for multi-centre trials, was addressed in the study presented here for four qMRI parameters (PD, T1, T2, T2*). The study was based on phantom and in vivo experiments (healthy subjects). Furthermore, for the phantom experiment, gold standard methods for T1 and T2 mapping were used to provide reference values for comparison.

2. Materials and methods

2.1. Participants and phantom

7 healthy subjects were recruited for the study (3 female, mean age ± SD: 27.9 ± 6.8 years). All participants gave written informed consent before participation. The study was approved by the institutional ethics committee. The investigation followed the principles formulated in the declaration of Helsinki.

The phantom was a spherical glass container (diameter: 16.6 cm) with a homogeneous filling of agarose gel: 1.75% w/v agarose dissolved in 0.9% NaCl solution and doped with gadolinium diethylenetriamine pentaacetic acid (Gd-DTPA, 0.11 mmol/L). It has been shown (Preibisch and Deichmann, 2009a) that this combination yields T1 and T2 values approximately matching the respective relaxation times of brain tissue at 3 T (ca. 900 ms for T1 and 60 ms for T2 in the current work).

2.2. Data acquisition: general

The study comprised in vivo and phantom scans. All data were acquired on two different 3 T MR scanners (Siemens Medical Solutions, Erlangen, Germany): a Trio system installed in late 2003 (software version VA35A) and a Prisma system installed in late 2017 (software version VE11C). On both scanners, RF transmission was performed with the respective scanner's body coil and signal reception with phased-array head coils (Trio: 8 channels, Prisma: 20 channels). For T2 mapping, including a gold standard method for the acquisition of T2 reference values for the phantom, release sequences provided by the manufacturer were used. All other sequences described below were developed in-house, using the manufacturer's software package. Importantly, the functionality of the different sequences was identical on both systems.

Data were processed with custom-built programs written in MatLab (MathWorks, Natick, MA). Furthermore, data processing steps such as brain extraction, segmentation and realignment were performed via the respective functions of the software package FSL, version 5.0.7 (FMRIB, Oxford).

2.3. Data acquisition: protocols

During the set-up of the different protocols, care was taken to minimize secondary effects that are known to reduce the accuracy of measured qMRI parameters, such as insufficient spoiling of transverse magnetization in gradient echo (GE) sequences, inaccurate B1 mapping or the occurrence of secondary echoes in turbo spin echo (TSE) sequences. Thus, protocols were used that have been described in detail in the literature, but differ in spatial coverage and resolution. However, this did not pose a problem for the presented study as the reproducibility for each parameter was assessed separately.

2.3.1. B1 and B0 mapping

B1 mapping was based on a method described in (Volz et al., 2010). In detail, two two-dimensional (2D) multi-slice GE data sets were acquired: one data set without, the other with magnetization preparation via a 45° RF pulse with a subsequent gradient spoiler, reducing the longitudinal magnetization and thus the signal level. The quotient of both data sets yields the cosine of the preparation angle, and B1 is obtained by comparing the actual angle with the nominal value of 45°. This procedure requires that the signal intensity is proportional to the longitudinal magnetization at the start of the respective scan; therefore, centric phase encoding (PE) was employed. To reduce the influence of T1-dependent saturation effects during image acquisition, the scan duration per slice

was kept short by acquiring slices sequentially rather than interleaved, i.e. collecting all PE data for a particular slice in a single block at a relatively short TR of 11 ms, yielding an acquisition time of 616 ms per slice for 56 PE steps. Furthermore, in a post-processing step, data were corrected for saturation effects. To this end, the reference images and the magnetization-prepared images were first segmented, yielding maps that indicate the probability that a certain pixel belongs to either white matter (WM), gray matter (GM) or cerebrospinal fluid (CSF). Subsequently, both the reference image and the magnetization-prepared image were split up into subimages containing mainly WM, GM or CSF by multiplication of the original images with the respective probability map. For each subimage, k-space lines were obtained via Fourier Transform. A correction for saturation effects was then performed in k-space, assuming the approximate respective T1 value (900, 1400 and 4500 ms for WM, GM and CSF) in the correction algorithm. The corrected k-space data derived from the different subsets were added and transformed back into image space. The final corrected image data were then used to calculate B1 as explained above. A detailed description of the process can be found in the literature (Volz et al., 2010).

The acquisition parameters were: field of view (FoV): 256 × 224 × 160 mm^3, matrix size 64 × 56, 40 sagittal slices with 4 mm thickness, no interslice gap, isotropic spatial resolution: 4 mm, TR/TE/FA = 11 ms/5 ms/11°, readout bandwidth (BW): 260 Hz/pixel, two iterations, duration 1:45 min.

B0 mapping was based on a standard 2D multi-slice dual-echo GE sequence with export of magnitude and phase data, using interleaved slice acquisition, i.e. acquiring data for a certain PE step for all slices before moving on to the next PE step. The acquisition parameters were: identical geometrical parameters (FoV, spatial coverage, matrix size, slice number and thickness, spatial resolution) as for B1 mapping, TR/TE1/TE2/FA = 560 ms/4.89 ms/7.35 ms/60°, BW = 200 Hz/pixel, duration 1:03 min.

2.3.2. T1 mapping

T1 mapping was based on the variable flip angle (VFA) method (Preibisch and Deichmann, 2009a, 2009b; Venkatesan et al., 1998; Wang et al., 1987), acquiring two three-dimensional (3D) RF-spoiled GE data sets with different excitation angles. To increase the signal-to-noise ratio (SNR), two GEs with different phase encoding were acquired per excitation pulse. The acquisition parameters were: FoV and spatial coverage as above, matrix size 256 × 224 × 160, isotropic spatial resolution: 1 mm, TR/TE/FA1/FA2 = 16.4 ms/6.7 ms/4°/24°, BW = 222 Hz/pixel, duration of each rectangular RF pulse = 200 μs, RF spoil increment: 50° (Preibisch and Deichmann, 2009a), spoiler gradient moment: 25 (mT/m)*ms (Baudrexel et al., 2018), total duration for both data sets 9:48 min. To avoid motion artifacts arising from lipids, water selective excitation pulses were used (Howarth et al., 2006) which are composites of two half-angle RF pulses, separated by 1.2 ms. The values of FA1 and FA2 were chosen according to a method described in the literature (Helms et al., 2011) for a target T1 of 1000 ms, thus optimizing the sequence for typical WM and GM T1 values at 3 T. It should be noted that for measuring a broader T1 range, a third scan with a different FA could be added.

2.3.3. T2* correction for PD mapping (in vivo only)

For the derivation of PD maps from the VFA data, a correction of T2* related signal losses is required. For this purpose, two 2D multi-slice GE data sets with different TE were acquired. The acquisition parameters were: matrix size 128 × 112, 80 sagittal slices with 2 mm thickness, no interslice gap, isotropic spatial resolution: 2 mm, TR/TE1/TE2/FA = 1336 ms/4.3 ms/11 ms/50°, BW = 292 Hz/pixel, duration 2:30 min for each data set. In particular, FoV and spatial coverage were chosen as above, thus yielding congruency of the T1 mapping and the T2* correction data. The TE difference of 6.7 ms matches the TE of the VFA scans, so the quotient of the two GE data sets yields the correction factor for compensating T2* related signal losses in the VFA data.

2.3.4. T2* mapping

T2* mapping was based on a method that reduces the impact of subject motion on the data (Nöth et al., 2014). In detail, a 2D multi-slice multi-echo GE sequence was used, acquiring per excitation pulse eight consecutive GEs with increasing TE, induced via readout gradient oscillation. All echoes were sampled under positive readout gradient polarity. The acquisition parameters were: FoV = 240 × 180 mm^2, matrix size = 192 × 144, isotropic in-plane resolution 1.25 mm, 40 axial slices with 2 mm thickness and 1 mm interslice gap, TR/FA = 2400 ms/30°, TE = [10, 16,22,28,34,40,46,52] ms, BW = 299 Hz/pixel, duration: 5:46 min (i.e., 144*TR). The sequence was repeated with 50% and 25% resolution in phase encoding direction, thus sampling only the central parts of k-space. These datasets were used to reduce motion artifacts as described in section 2.6.4. Each of the additional scans was preceded by six dummy scans (total duration 14.4 s) to achieve steady state conditions, yielding durations of 3:07 min and 1:41 min, respectively. Complex image data (modulus and phase) were exported for each data set.

2.3.5. T2 mapping

For T2 mapping, five 2D multi-slice TSE data sets were acquired with different TE. The acquisition parameters were: FoV = 256 × 176 mm^2, matrix size = 256 × 176, isotropic in-plane resolution 1 mm, 50 axial slices with 2.5 mm thickness, no interslice gap, TR/FA_{ex}/FA_{refoc} = 10s/90°/180°, TE = [17, 86, 103, 120, 188] ms, turbo factor = 11, BW = 100 Hz/pixel, echo spacing = 17.1 ms, duration: 2:52 min per data set.

2.3.6. T1 reference mapping (phantom only)

For obtaining reference T1 data on a phantom, 14 inversion recovery (IR) prepared 2D echo planar imaging (EPI) data sets with different inversion times TI were acquired, allowing for full spin relaxation before each scan. The acquisition parameters were: FoV = 192 × 192 mm^2, matrix size = 64 × 64, isotropic in-plane resolution 3 mm, 6 coronal slices with 2 mm thickness and 2 mm interslice gap, TR/TE/FA = 30s/26 ms/90°, TI = [100, 200, 300, 400, 600, 800, 1000, 1200, 1600, 2000, 2600, 3200, 3800, 4500] ms, BW = 2298 Hz/pixel, duration: 30 s per data set. A slice-selective adiabatic inversion pulse was used to optimize spin inversion. Furthermore, the 6 slices were scanned at equidistant intervals of 5 s during the scan duration of 30 s which allows for full spin relaxation before each inversion pulse, given an approximate T1 value of about 1 s. In this way, cross-talk between the slices due to imperfect pulse profiles or magnetization transfer (MT) effects is avoided.

2.3.7. T2 reference mapping (phantom only)

For obtaining reference T2 data on a phantom, three 2D spin echo (SE) EPI data sets with different TE were acquired, allowing for full spin relaxation before each scan. The acquisition parameters were: FoV = 256 × 176 mm^2, matrix size = 128 × 88, isotropic in-plane resolution 2 mm, 11 axial slices with 2.5 mm thickness and 2.5 mm interslice gap, TR/FA = 20s/90°, TE = [50, 90, 120] ms, BW = 1302 Hz/pixel, duration: 20 s per data set.

2.4. Data acquisition: phantom experiment

The phantom was scanned on both MR systems, applying all protocols described above. For assessing the scan-rescan stability, the phantom was scanned twice on each system, performing both runs in a single session, but moving the phantom out of the magnet bore between the runs. As relaxation times are temperature dependent, the phantom was stored inside the respective magnet room during the night preceding the scan to

allow for full thermal accommodation and the temperature of the magnet room was measured before and after each experiment. Furthermore, the Larmor frequency was noted for each experiment as it influences relaxation times (Rooney et al., 2007).

2.5. Data acquisition: in vivo experiment

Each subject was scanned on both MR systems (4 first on the Trio and 3 first on the Prisma scanner), applying the same protocols as for the phantom, apart from T1 and T2 reference mapping. For assessing the scan-rescan stability, each subject was scanned twice on each system on the same day, but subjects were asked to leave the scanner between the runs. For each subject, the time between data acquisition on Trio and Prisma was less than five weeks. The subject's body temperature was measured before each scan using a digital ear thermometer.

2.6. Data analysis

2.6.1. B1 and B0 mapping

B1 maps were derived as described above. It should be noted that B1 values were normalized by attributing a value of 1.0 to areas where actual and nominal excitation angles were identical.

B0 maps (in radians per sec) were derived by unwrapping the phase difference data with the FSL tool PRELUDE and dividing the unwrapped phase differences between the two GE data sets by their TE difference of 2.46 ms.

2.6.2. T1 mapping

T1 maps were calculated as described in the literature (Venkatesan et al., 1998; Wang et al., 1987), including a correction for insufficient spoiling of transverse magnetization according to (Baudrexel et al., 2018). In detail, T1 was derived by plotting S(FA)/sin(FA) versus S(FA)/tan(FA) where FA is the excitation angle and S(FA) is the signal in the respective GE data set. This plot shows a linear dependence with the slope given by exp(-TR/T1) from which T1 can be obtained. Before performing the plot, the following three correction procedures were carried out for the FA maps: (1) The local excitation angles were corrected for B1 non-uniformities. (2) The effect of B0 distortions on the water selective excitation pulses was assessed numerically via concatenation of rotation matrices describing spin excitation and spin dephasing during the 1.2 ms delay of the composite pulse. (3) A correction was performed which addresses the problem that due to the effects of insufficient spoiling in GE data, actual signal levels differ from the values predicted by the Ernst equation, yielding erroneous T1 values as this equation forms the basis of the VFA concept. The method derives modified excitation angles FA′ that represent the angles for which the Ernst equation yields the measured signal levels. Thus, insertion of FA′, rather than FA, in the VFA algorithm outlined above yields correct T1 values. The effective angles FA′ are derived from the actual values FA via FA' = C*FA where the correction factor C is calculated from FA (in deg) and TR (in ms) according to:

$$C = \sum_{k,l=0}^{k+l\leq 5} P_{k,l} \cdot FA^k \cdot TR^l$$

The required matrix elements $P_{k,l}$ are dimensionless numbers that can be obtained for different RF spoiler increments from the literature (Baudrexel et al., 2018). Here, the values for an RF spoiler increment of 50° were chosen.

2.6.3. PD mapping (in vivo only)

The evaluation of PD maps was based on a method described in the literature (Volz et al., 2012a). In detail, the VFA data set acquired with the lower excitation angle was corrected for any T1 and B1 bias and for T2* related signal losses, as described in section 2.3.3 above. The corrected data represent the product of the PD map, the receiver coil profile and an unknown scaling factor. The receiver coil profile bias was removed via bias field correction and the PD map was scaled to a value of 100 percent units (pu) in CSF.

2.6.4. T2* mapping

T2* mapping was performed as described in (Baudrexel et al., 2009) and (Nöth et al., 2014). In detail, to reduce the influence of motion artifacts, the complex image data of the three data sets were first converted into k-space data via Fourier transform. As acquisitions had been performed with full, half and quarter spatial resolution in PE direction at constant FoV, k-space lines belonging to the central 50% of k-space had been acquired at least two times, with three acquisitions for the central 25%. Consequently, these lines were available from different acquisitions, so lines affected by subject motion in one of these acquisitions can be suppressed or removed without compromising the entirety of k-space data. To achieve this goal, a weighted average of these data was obtained separately for each TE, choosing the weighting factors individually for each line in a way which suppresses data affected by subject motion. The averaged k-space data were converted back into complex image space, B0 maps were derived from the phase data and modulus images were corrected for signal losses induced by macroscopic B0 distortions. For the corrected data, the TE dependence of the signal amplitude was fitted mono-exponentially, yielding maps of T2*.

2.6.5. T2 mapping

For deriving T2 values, the TE dependence of signal amplitudes in the acquired TSE data sets was fitted mono-exponentially. However, it should be noted that this procedure yields apparent T2 values T2(app) which can differ considerably from the true values (McPhee and Wilman, 2015). This is due to the occurrence of stimulated and indirect echoes in TSE data which distort the signal decay curve in cases where the excitation and refocusing pulses differ from the ideal values of 90° and 180°, respectively, mainly due to B1 inhomogeneities and in particular to variations of the effective angles across the excited slice. Corrections of this effect were performed as described in (Nöth et al., 2017). In detail, corrected T2 values were derived from T2(app) according to:

$$T2 = K1(B1) + K2(B1) * T2(app)$$

with:

$$K1(B1) = P1(1)\cdot(B1\text{-}1)^2 + P1(2)\cdot(B1\text{-}1) + P1(3)$$

$$K2(B1) = P2(1)\cdot(B1\text{-}1)^2 + P2(2)\cdot(B1\text{-}1) + P2(3)$$

As in the study presented here a manufacturer's release TSE sequence with unknown RF pulses was used, the parameters P1(i) and P2(i) were derived for both scanners in a separate phantom experiment according to (Nöth et al., 2017), yielding:

Trio Scanner:

$$P1(1) = 45.6,\ P1(2) = 1.80,\ P1(3) = 2.25$$

$$P2(1) = -2.30,\ P2(2) = -0.235,\ P2(3) = 0.765$$

Prisma Scanner:

$$P1(1) = 45.7,\ P1(2) = 1.72,\ P1(3) = 2.85$$

$$P2(1) = -2.07,\ P2(2) = -0.189,\ P2(3) = 0.680$$

2.6.6. T1 reference mapping (phantom only)

T1 reference values were obtained by mono-exponential fitting of the TI dependence of signal amplitudes in the IR-EPI data. In general, a 3-parameter fit is chosen for this purpose. It has been shown (Kellman and Hansen, 2014; Weingärtner et al., 2015) that using a 2-parameter fit may provide a significant gain of precision. However, there is a risk of reduced accuracy as a 2-parameter fit is sensitive to the inversion or

saturation efficiency (depending on the sequence type) of the magnetization preparation pulses employed (Kellman and Hansen, 2014; Weingärtner et al., 2015). To address this issue, a 2-parameter fit was chosen with inclusion of the inversion efficiency as obtained via pulse simulations. Although an adiabatic inversion pulse was chosen which yields full spin inversion in the slice centre over a range of B1 values, the inversion efficiency may be reduced due to imperfect slice profiles. To assess this effect, both the excitation and the inversion profiles (i.e. the transverse magnetization M_t after excitation and the longitudinal magnetization M_z after inversion, respectively) were calculated numerically by solving the Bloch equations with a Runge-Kutta algorithm. Subsequently, M_z was averaged, using M_t as weighting factor, yielding a value of −0.964 which indicates a slightly reduced inversion efficiency. To account for this effect, the 2-parameter fit was performed according to:

$$S(TI) = S_0 \cdot [1–1.964 \cdot \exp(-TI/T1)]$$

2.6.7. T2 reference mapping (phantom only)

T2 reference values were obtained by mono-exponential fitting of the TE dependence of signal amplitudes in the SE-EPI data. It should be noted that no distortion correction was performed on the EPI data sets that were acquired for T1 and T2 reference mapping, as image distortions due to static magnetic field inhomogeneities are independent of the choice of TI and TE, thus having no effect on the T1 or T2 values in the resulting maps. Spatial distortions in the maps are not problematic since a homogeneous phantom was investigated.

2.6.8. Synthetic anatomies

Pseudo PD maps were obtained from T1 data as described previously in the literature (Volz et al., 2012b). Synthetic T1/PD-weighted MP-RAGE anatomies for tissue segmentation in healthy subjects were derived from these pseudo PD maps and the T1 maps as explained previously (Gracien et al., 2019; Nöth et al., 2015), assuming the following (virtual) acquisition parameters: TI/TR = 900 ms/1900 ms, FA = 9°, isotropic resolution: 1 mm, FoV: 256 × 224 × 160 mm^3. The synthetic MP-RAGE data, rather than the original source images, were chosen for segmentation as they are already corrected for any bias imposed by B1 variations or inhomogeneities of the receive RF coil.

2.6.9. Analysis of phantom data

For each scanner and each run, values from the different parameter maps were collected across all pixels inside a phantom mask (derived by thresholding the respective source images, thus excluding background noise). To reduce the influence of outliers, the mean value of the parameter of interest and its standard deviation were determined across the phantom mask. Subsequently, only those pixels were considered for further analysis for which the respective parameter differed by not more than three standard deviations from the mean. This limit was chosen empirically, since for all measurements it resulted in the exclusion of spurious outliers only. For the T2* mapping data (and, to a certain degree, the T1 reference data), these outliers were located in areas of strong B0 distortions (in the phantom neck close to the air-gel interface). For the SE-EPI based T2 reference data, outliers occurred in the centre of the phantom where B1 was increased due to the refocusing pulse of 180°, yielding reduced SNR. Subsequently, the mean value and standard deviation of the parameter of interest were calculated across the remaining pixels.

2.6.10. Analysis of in vivo data

For comparison of qMRI values between scanners and time-points in vivo, WM and deep GM segmentations were performed similarly as described previously (Gracien et al., 2016a) for the data acquired on healthy subjects: To this aim, all datasets were first skull-stripped using FSL BET (Smith, 2002). Then, WM maps were obtained from the synthetic anatomies with FAST (Zhang et al., 2001). Furthermore, deep GM regions were identified with FIRST (Patenaude et al., 2011). WM maps were modified by removing pixels inside deep GM masks and the cerebellum and voxels with a partial volume (PV) estimate less than 1. The deep GM masks with the most robust segmentation results (thalamus, putamen, caudate nucleus and globus pallidus) were eroded with a 3 × 3 × 3 mm^3 box kernel (twice for the thalamus) to remove the outermost voxels at the edge of the respective masks and, accordingly, minimize PV effects. Combined deep GM masks were obtained from these masks. ROIs in WM and deep GM were chosen because of their low susceptibility to PV effects.

The WM and combined deep GM masks were coregistered to the T2 and T2* maps by applying the coregistration matrices between the anatomies and the respective maps using nearest neighbor interpolation. The T1 and PD maps were already in alignment with the synthetic anatomies and, accordingly, with the masks. The WM and combined deep GM masks were applied to the qMRI maps to determine mean values in WM/deep GM.

Furthermore, for cortical analysis, cortical PV estimate (PVE) maps were derived from the GM PVE maps identified with FAST by removing voxels with a PVE < 0.95 and pixels in WM, deep GM, cerebellum and ventricle masks (Gracien et al., 2016b). Cortical mean values were determined for the parameter maps with a 1 mm isotropic resolution (T1 and PD) which allows for an accurate analysis of the thin cortical layer because of the low susceptibility to PV effects.

The relative differences between the second and first run for both scanners and the relative differences between Prisma and Trio of the mean values of both runs were calculated for each subject, for each region (WM, deep GM, cortex) and for all investigated qMRI parameters. Average values of both runs were chosen to reduce the effect of potential scan-rescan variations on the comparison between scanner models. Signed and unsigned relative differences were averaged across the group. Statistical comparisons were performed via t-tests. P-values below 0.05 were considered significant.

Additionally, histograms were generated for each parameter (T1, PD, T2, T2*), each subject, each scanner model and each run, including only voxels with a GM or WM PVE equal to or exceeding 0.95. The histograms were scaled to a maximum peak amplitude of 1.

To visualize the spatial distribution of qMRI parameter differences, all qMRI maps were coregistered to the datasets acquired during the first Trio scan for each subject. Afterwards, mean values and standard deviations were determined within combined WM/GM-masks (identified with FAST and eroded with an 8 mm box kernel to reduce PV effects with CSF) and averaged across the group for each parameter. For the following analysis steps, only qMRI values that deviated by less than two standard deviations from the mean values, obtained as described above, were included. These limits were chosen empirically to reduce PV effects with CSF. For visualization of focal changes, relative difference maps between the second and the first run for either scanner (denoting the respective scan-rescan deviation) and between the first runs from both scanners (denoting the inter-scanner deviation) were obtained for all non-zero voxels. Histograms of these difference maps were created, scaled in a way that the histograms had a peak amplitude of 1 and averaged across subjects.

3. Results

3.1. Phantom experiment

The temperature in the magnet rooms was 18 °C for the Trio scanner and 21 °C for the Prisma scanner and remained constant during the experiments, i.e. in each case, temperatures were identical before the first and after the second run. The Larmor frequencies were 123.2542 MHz on the Trio scanner and 123.2603 MHz on the Prisma scanner.

As shown in Table 1, T1 and T2 values in the phantom experiment

Table 1
qMRI values obtained in the phantom experiment (Mean ± standard deviation, in ms).

	Trio, run 1	Trio, run 2	Prisma, run 1	Prisma, run 2
T1 reference	895.6 ± 16.9	904.7 ± 18.0	919.7 ± 11.6	923.5 ± 11.4
T1	894.4 ± 47.7	907.6 ± 49.7	923.4 ± 38.7	935.3 ± 39.1
T2*	65.1 ± 3.8	64.4 ± 3.4	62.0 ± 2.6	61.2 ± 2.9
T2 reference	64.3 ± 1.5	63.5 ± 1.5	62.2 ± 1.3	61.7 ± 1.4
T2	64.1 ± 2.2	63.7 ± 2.2	62.7 ± 5.4	62.0 ± 5.3
T2 (app)	81.2 ± 2.4	80.3 ± 2.5	88.1 ± 7.9	87.1 ± 7.7

closely match the reference values (absolute values of relative deviations ≤ 1.27%). The scan-rescan stability in-vitro is high for both scanners (absolute values of relative deviations ≤ 1.47%). The relative deviations between scanner models (averaged qMRI values across both runs) were slightly higher (T1: 3.10%, T2*: −4.99%, T2: −2.46%). In the case of the uncorrected T2 (app), relatively large differences between the scanners were observed (8.14%), stressing the need to correct TSE based T2 values with scanner-specific parameters (Nöth et al., 2017).

3.2. In-vivo experiment

The body temperature was slightly higher before the second run than before the first run and the temperatures were virtually identical for both scanners (Trio, run 1: 36.6 ± 0.7 °C, run 2: 37.1 ± 0.3 °C; Prisma, run 1: 36.5 ± 0.4 °C, run 2: 37.1 ± 0.3 °C).

Fig. 1 presents typical results of qMRI mapping (PD, T1, T2* and T2 maps, single slice for a representative subject, comprising caudate nucleus, putamen and thalamus). Fig. 2 shows the results of the deep GM, cortex and WM segmentation for the same representative subject (slice position similar to Fig. 1, data acquired with Prisma).

As demonstrated in Table 2, the scan-rescan stability in-vivo was high for both scanners. For the Trio, the maximum deviation of a single parameter (given as the unsigned relative difference, averaged across subjects) was 2.14% (the value refers to T2* in WM), for the Prisma 2.10% (T1 in deep GM). Comparison of these scan-rescan deviations between Trio and Prisma across all parameters and regions yielded higher values for Trio (1.56 ± 0.29%, Prisma: 1.23 ± 0.40%, paired *t*-test: $p = 0.02$), indicating an increased stability of the Prisma scanner.

Table 3 presents the relative differences between both scanners across the group of healthy subjects as a measure for in-vivo inter-scanner-model reproducibility. For the uncorrected T2(app) values, large differences were observed, in particular in WM (≤7.83%). After correction, these differences were reduced to 3.29% in WM. Furthermore, increased values for Prisma were found for T1 and PD and decreased values for T2* and T2 (deviation up to 5.21%), indicating a systematic bias.

The deviations between scanner-models (averages of unsigned relative differences across the group excluding T2(app): ≤5.21%) were larger than scan-rescan-deviations (≤2.14%). However, in single cases scan-rescan deviations up to 4.44% were observed (c.f. maximum in Table 2).

Fig. 3 shows representative relative difference maps between scan and rescan (left column, demonstrated for Prisma) and between the scanner models (right column), with hot colors indicating increases for the second scan (left) or for Prisma (right). Data are shown in T1 space (slice position 30 mm superior to Fig. 1). Strong deviations could be observed for T2(app) between scanner models, but not between scan and rescan. For the other parameters, differences are relatively low in most regions but in some areas slightly higher for the comparison between scanner models than between scan and rescan.

The Supplementary Fig. 1 shows for each parameter, i.e. for PD, T1, T2*, T2 and T2(app), histograms of the voxel-wise relative parameter differences, averaged across the group. The blue/red histogram refers to the Prisma/Trio scan-rescan deviations, respectively, and the green line to the inter-scanner-model differences.

Fig. 4 shows for all subjects (rows) histograms of T1 values (left column) and PD values (right column). In each case, the blue/red histogram refers to the Prisma/Trio scanner, and the solid/dashed line type refers to the first/second run. Fig. 5 shows the respective results for T2 (left column) and T2* (right column). Average parameter values

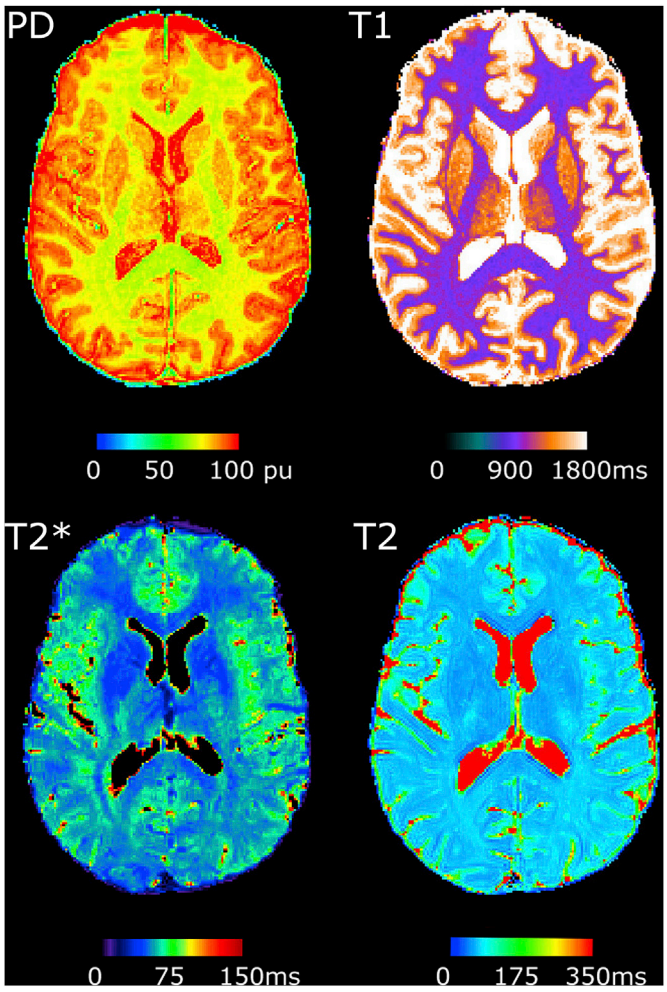

Fig. 1. The results of qMRI mapping: PD, T1, T2* and T2 maps for a representative subject.

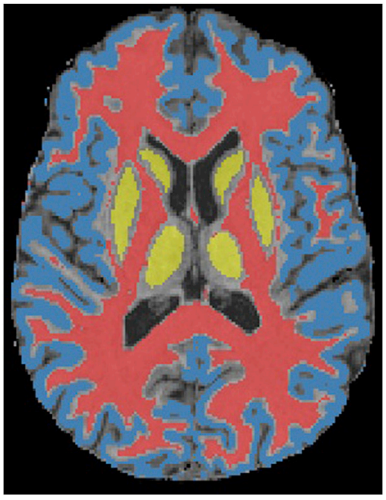

Fig. 2. The results of the deep GM, cortex and WM segmentation for the same subject as in Fig. 1 (same slice position as in Fig. 1). Red areas indicate the white matter, blue areas the cortex and yellow areas the deep gray matter segmentation.

Table 2
In-vivo scan-rescan stability. Minimum, maximum, mean, standard deviation (SD) and mean/SD of the unsigned (i.e. absolute (abs.)) values of the relative differences between the second and the first run across the group of healthy subjects. All values are given in percent. The last lines for each scanner demonstrate the averages of the values above (in bold). WM = white matter, GM = gray matter.

			Min	Max	Mean ± SD	Mean ± SD abs.
Trio	PD	WM	−3.05	3.31	−0.26 ± 2.00	1.46 ± 1.27
		deep GM	−2.86	2.16	−0.28 ± 1.67	1.29 ± 0.97
		cortex	−2.13	3.24	−0.24 ± 1.80	1.35 ± 1.09
	T1	WM	−4.53	0.90	−1.27 ± 1.91	1.66 ± 1.52
		deep GM	−4.20	2.21	−1.36 ± 2.18	1.99 ± 1.51
		cortex	−5.14	0.48	−1.66 ± 1.94	1.87 ± 1.70
	T2*	WM	−2.86	4.28	−0.11 ± 2.77	2.14 ± 1.54
		deep GM	−4.57	0.29	−1.24 ± 1.64	1.32 ± 1.56
	T2	WM	−1.40	3.66	0.53 ± 1.96	1.35 ± 1.42
		deep GM	−1.22	3.59	0.68 ± 1.98	1.56 ± 1.27
	T2 (app)	WM	−1.22	3.01	0.30 ± 1.74	1.30 ± 1.08
		deep GM	−2.22	3.29	−0.10 ± 1.87	1.47 ± 0.99
Mean Trio			**−2.95**	**2.54**	**−0.42 ± 0.78**	**1.56 ± 0.29**
Prisma	PD	WM	−1.10	2.75	0.40 ± 1.25	0.91 ± 0.88
		deep GM	−1.09	3.68	0.76 ± 1.57	1.31 ± 1.07
		cortex	−0.69	2.59	0.35 ± 1.14	0.78 ± 0.85
	T1	WM	−2.72	2.81	−0.04 ± 1.75	1.25 ± 1.12
		deep GM	−3.65	4.44	0.18 ± 2.68	2.10 ± 1.45
		cortex	−2.68	2.58	−0.33 ± 1.71	1.31 ± 1.01
	T2*	WM	−1.38	2.14	0.30 ± 1.16	0.88 ± 0.75
		deep GM	−1.24	0.38	−0.56 ± 0.66	0.67 ± 0.52
	T2	WM	−4.87	2.69	−0.44 ± 2.41	1.57 ± 1.78
		deep GM	−2.95	3.10	−0.04 ± 2.10	1.63 ± 1.14
	T2 (app)	WM	−4.22	2.50	0.00 ± 2.06	1.22 ± 1.59
		deep GM	−0.83	2.84	0.75 ± 1.37	1.18 ± 0.96
Mean Prisma			**−2.28**	**2.71**	**0.11 ± 0.43**	**1.23 ± 0.40**

Table 3
In-vivo inter-scanner-model reproducibility. Minimum, maximum, mean, standard deviation (SD) and mean/SD of the unsigned (i.e. absolute (abs.)) values of the relative differences between both scanners across the group of healthy subjects. Positive values indicate increases for Prisma (except for the last column). All values are given in percent. WM = white matter, GM = gray matter.

		Min	Max	Mean ± SD	Mean ± SD abs.
PD	WM	−0.12	5.67	2.52 ± 2.12	2.56 ± 2.07
	deep GM	−3.13	3.20	0.09 ± 2.53	2.21 ± 0.83
	cortex	1.44	6.41	3.77 ± 1.87	3.77 ± 1.87
T1	WM	1.04	7.01	3.12 ± 1.93	3.12 ± 1.93
	deep GM	−0.12	4.96	2.52 ± 1.90	2.56 ± 1.84
	cortex	−0.42	4.98	1.49 ± 1.80	1.61 ± 1.67
T2*	WM	−7.91	−1.00	−3.01 ± 2.32	3.01 ± 2.32
	deep GM	−5.50	0.37	−2.41 ± 2.12	2.51 ± 1.97
T2	WM	−4.24	−2.11	−3.29 ± 0.83	3.29 ± 0.83
	deep GM	−8.15	−2.79	−5.21 ± 1.97	5.21 ± 1.97
T2 app	WM	7.12	8.96	7.83 ± 0.61	7.83 ± 0.61
	deep GM	3.01	5.99	4.53 ± 1.05	4.53 ± 1.05

(averaged across subjects, scanners, runs and the respective WM or GM tissue mask) were: T1 = 913 ± 36 ms (WM) and 1601 ± 43 ms (GM), T2 = 69.7 ± 1.4 ms (WM) and 91.6 ± 4.5 ms (GM), T2* = 53.4 ± 1.4 ms (WM) and 59.5 ± 1.2 ms (GM), PD = 70.0 ± 2.5 pu (WM) and 83.1 ± 2.8 pu (GM). These results are in good agreement with values reported in the literature (PD, T1: (Abbas et al., 2014); T2: (Petrovic et al., 2015); T2*: (Baudrexel et al., 2009; Péran et al., 2007)).

4. Discussion

The presented study evaluates intra- and inter-scanner model reproducibility for multiple qMRI parameters, comparing two scanner models from the same vendor, but with different hard- and software versions, using exactly the same acquisition protocol and analyses.

The intra-scanner variability was low, not exceeding 2.14% for any scanner or parameter, both in vivo and in vitro. The in vivo deviations between scanner models manufactured by the same vendor did not exceed 5.21% for PD, T1, T2* and T2. This is in line with previously reported findings (Lee et al., 2019), where intra-scanner variabilities were low (1.4%), whereas substantial deviations between scanner models up to 10.0% were reported. However, it should be noted that in the latter study models from different vendors were compared. Furthermore, while the reported values in the present work are based on average values in ROIs, calculation of differences in the previous work by Lee at al. was performed voxel-wise before reading average differences within ROIs. While Lee et al. observed widespread focal T1 deviations between scanner vendors by up to 20–30% (Lee et al., 2019), in the presented investigation focal T1 values showed relatively minor deviations in most regions (cf. Fig. 3). Still, in some regions, focal T1 differences equal to or exceeding ± 15% could be observed. A potential reason for this discrepancy in results may be that in this previous study the correction for B1 inhomogeneities was performed with different methods, whereas in the present study, the same acquisition protocol was transferred from Trio to Prisma, yielding identical functionality for the older and the newer software versions. Likely, these efforts explain the high reproducibility in the presented study which is similar to values reported in a previous study which compared qMRI values acquired with the same scanner model at different sites (Weiskopf et al., 2013). The higher scan-rescan stability observed for Prisma might be explained by the 14 years of difference between the installations of both scanners as the newer Prisma scanner can be expected to use advanced hardware components.

It should be noted that the relative inter-scanner differences as stated in Table 3 are for most comparisons larger than the respective intra-scanner differences in Table 2. The reason is that differences between the scanners arise both from the intra-scanner stabilities and a systematic bias between the scanners.

In vitro, the presented T1 and T2 values closely matched the reference values (deviations ≤ 1.27%). In contrast, a previous investigation observed higher T1 values with the VFA method than with the reference method (Stikov et al., 2015). The authors discussed that incomplete spoiling of transverse magnetization and incorrect flip angle calibration might be potential sources of this discrepancy (Stikov et al., 2015). In the

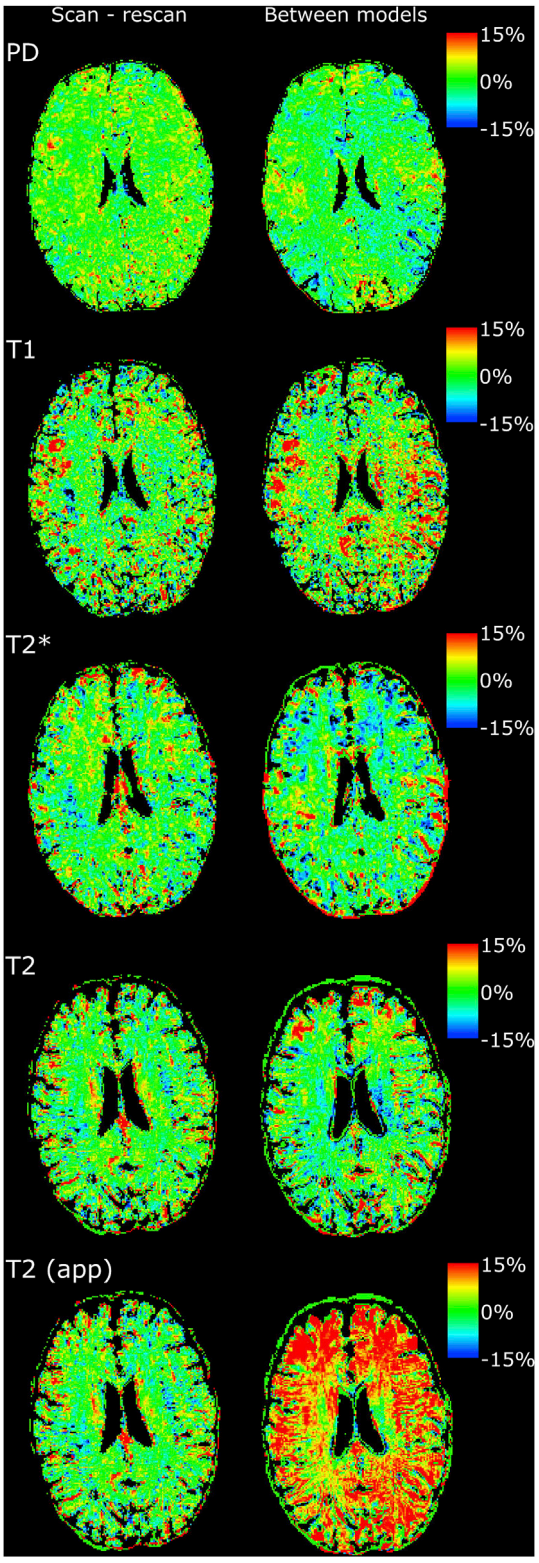

Fig. 3. Relative difference maps between scan and rescan (left column, demonstrated for Prisma) and between the scanner models (right column) for a representative subject (slice position 30 mm superior than in Fig. 1).

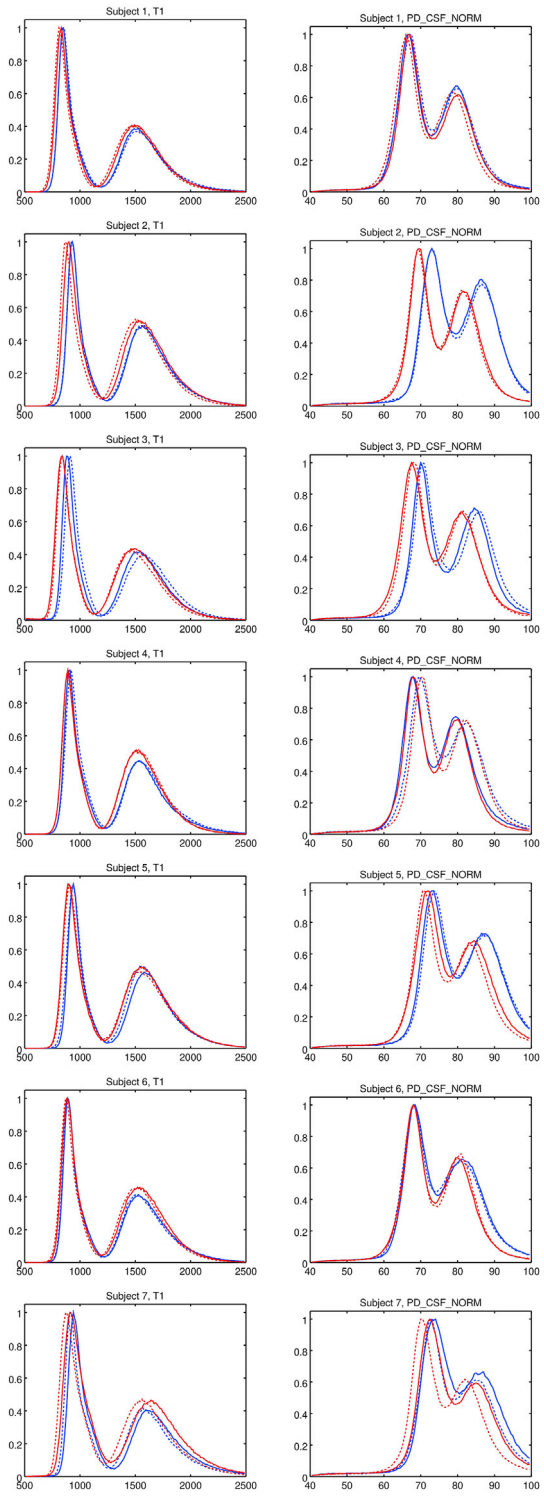

Fig. 4. Histograms for all subjects (rows) of T1 values (left column) and PD values normalized to 100 percent units in CSF (PD_CSF_NORM, right column). In each case, the blue/red histogram refers to the Prisma/Trio scanner, and the solid/dashed line type refers to the first/second run.

presented study, correction for the incomplete spoiling of transverse magnetization was performed as described previously (Baudrexel et al., 2018), yielding a marked reduction of deviations.

Phantom T1 values obtained with both the VFA and the reference methods were approximately 25 ms/3% higher for Prisma. This finding cannot be explained by the Larmor frequencies, which were virtually identical, but potentially by the temperature difference of 3 K between the scanner rooms. Previous studies observed that T1 increases by approximately 2.5% per Kelvin (Nelson and Tung, 1987; Schwarzbauer et al., 1995). Accordingly, a difference of 7.5% between Trio and Prisma

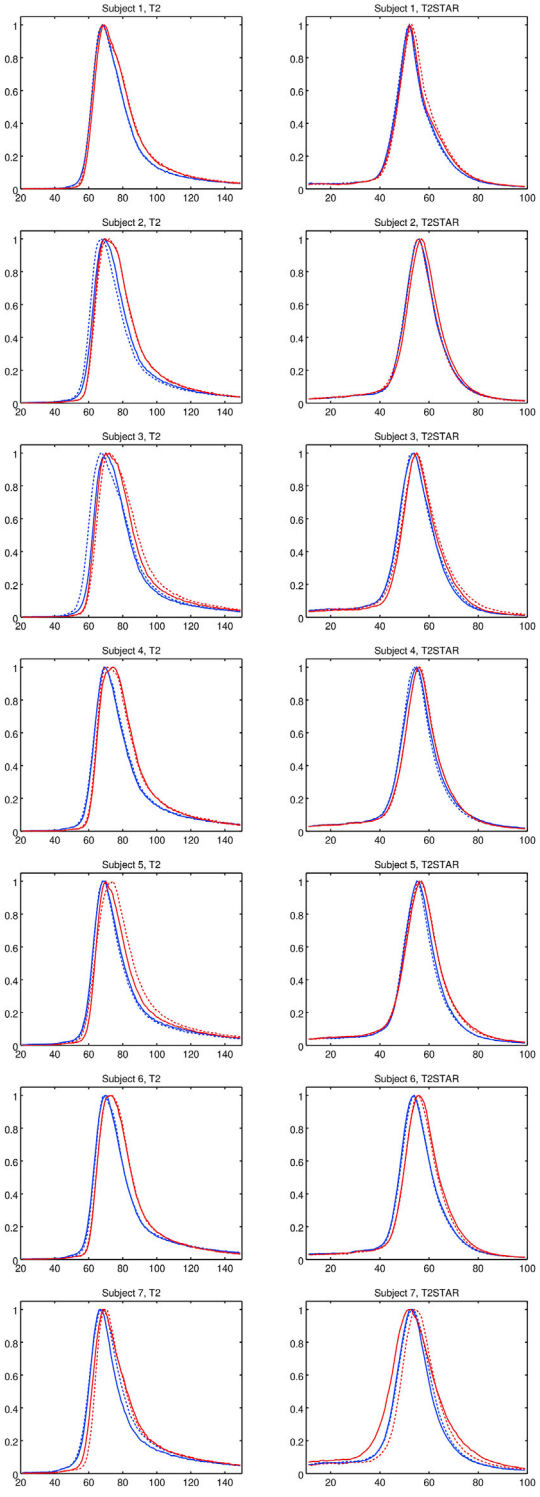

Fig. 5. Histograms for all subjects (rows) of T2 values (left column) and T2* values (right column). In each case, the blue/red histogram refers to the Prisma/Trio scanner, and the solid/dashed line type refers to the first/second run.

could be expected in the present study. However, the mentioned investigations reported values observed at different field strengths (Nelson and Tung, 1987; Schwarzbauer et al., 1995). Additionally, the temperature dependence of T1 is material specific (Nelson and Tung, 1987). Although similar material was used for the phantoms in the present study and in the previous investigation by Schwarzbauer et al. (1995), the phantoms were not identical. Furthermore, in the current study, the temperature was measured in the scanner room, but not inside the magnet bore where gradient heating may have caused different temperature levels. Still, in vivo results imply that differences between the scanners might not be explained by the temperature only. In particular, this view is supported by the finding of similar T1 differences but virtually identical body temperatures in vivo.

In the case of the in vitro experiment, the question arises if gradient or RF heating might be the origin of the slightly higher T1 reference values in the second run at Trio (T1 difference of approximately 10 ms), especially because the first run was the first measurement in the morning and the scanner was still cold. As mentioned above, the room temperature did not change over the duration of both runs. Still, as the temperature was not measured directly inside the magnet bore, local heating cannot be wholly ruled out. For the Prisma, T1 reference values remained constant. A possible reason might be that the gradient cooling is more effective and/or that the room temperature was higher, so the effects of local gradient heating would be less noticeable.

Although the inter-scanner deviation of PD values according to Table 3 was on average about 2.8%, there are larger deviations up to 6.41% (listed maximum value in Table 3) for single subjects (e.g. subject 2 in Fig. 4, right column). A potential reason may be the normalization of PD maps to 100 pu in CSF which can be problematic, in particular for small ventricles. As an alternative, it has been proposed to scale PD maps to 69 pu in WM (Weiskopf et al., 2013), since homogeneous WM regions are in general larger than the ventricles. The Supplementary Fig. 2 shows PD histograms for all subjects on the basis of PD maps that were scaled in this way, clearly reducing the variation between runs and scanners. However, scaling to a fixed value in WM may introduce a bias if pathologies or physiological changes affect the PD in a large area of WM (Weiskopf et al., 2013). Alternatively, it has been proposed to use an external reference for scaling (Lorio et al., 2019).

When comparing qMRI data obtained on different MR systems, it has to be taken into account that T1 depends crucially on the static magnetic field strength B0. According to the literature (Rooney et al., 2007), this dependence can be approximately expressed as $T1 = 0.583*(B0)^{0.382}$ in WM and $T1 = 0.857*(B0)^{0.376}$ in GM, where T1 is given in seconds and B0 in Tesla. In the study presented here, the difference between Larmor frequencies amounted to about 50 ppm, the effect of which is negligible. However, a Larmor frequency of 123.6 MHz as measured for the Prisma scanner corresponds to a B0 of about 2.9 T. Thus, comparison with data acquired on a system with a B0 of exactly 3.0 T would yield a systematic T1 difference of 1.3% in WM and GM. This effect has to be considered when interpreting multi-centric quantitative T1 data.

A further potential bias in T1 mapping is the occurrence of MT effects. Although it has been suggested that MT effects are rather a problem in 2D multi-slice sequences, due to the irradiation of off-resonance RF pulses (Stikov et al., 2015), recent studies have shown that on-resonance saturation effects in the bound proton pool are not to be neglected (Teixeira et al., 2019). As a consequence, even when using 3D sequences, MT effects may lead to systematic errors in VFA based T1 mapping, in dependence on the B1 amplitude of the irradiated RF pulses (Teixeira et al., 2019). The resulting bias may be positive or negative, depending on RF pulse length and shape, with a relatively small bias for a pulse duration of 500 μs (Al-Abasse and Helms, 2016). For the study presented here, the total duration of the fat-insensitive RF excitation (consisting of two rectangular RF pulses) was 400 μs, so a certain MT related bias can be expected. This is in line with the findings of a previous in vivo study (Preibisch and Deichmann, 2009b) where T1 values obtained with the identical VFA sequence and similar parameters (TR = 15.9 ms, flip angles of 4° and 25°) were about 5% larger than reference values obtained with IR-EPI. Still, when acquiring multi-centric quantitative T1 data, it is of vital importance that RF pulses of identical shapes and durations are used, so the MT induced bias is identical across the centres.

For PD, T1 and T2* mapping, identical imaging sequences developed in-house were used, adapting the respective codes from the older to the

newer system. Only for T2 mapping, a manufacturer sequence was used to acquire TSE data, but the resulting T2 values were corrected with scanner-specific parameters (Nöth et al., 2017). As a result, the corrected T2 values closely matched the reference values, whereas uncorrected T2(app) values were systematically too high. Furthermore, the uncorrected values differed between scanner models in vitro and in vivo (Fig. 3, last row), whereas there was no such discrepancy for the corrected values. Accordingly, in multi-centre studies such correction is strongly advisable.

An important point meriting consideration when setting up a multi-centre study employing qMRI methods is the fact that most qMRI techniques are based on the assumption of mono-exponential longitudinal or transverse relaxation behaviour. While this is true for a homogeneous phantom, it is only an approximation for brain tissue. As an example, transverse relaxation in brain tissue shows a multi-exponential behaviour (Bjarnason et al., 2005; Fischer et al., 1990; Laule et al., 2007; MacKay et al., 1994), so mono-exponential fitting rather yields an "effective" T2 which depends on the choice of TE values (Nöth et al., 2017). For this reason, identical TE values were chosen for both scanners in this study. Although effective parameter values do not provide accurate information on the multi-exponential characteristics of relaxation and mono-exponential fitting is rather a pragmatic simplification (Woermann et al., 2001), the comparison of effective qMRI parameters is still useful as it helps to establish a consistent range of normal values, thus aiding the differentiation of normal from abnormal tissue (Boulby and Rugg-Gunn, 2003; Nöth et al., 2017) in a clinical environment. However, to ensure the comparability of results in multi-centre studies, comparable acquisition techniques should be used and differences between acquisition parameters (such as the TE values chosen for T2 mapping or the TR value and excitation angles in VFA-based T1 mapping) should be kept to a minimum.

T2* is always lower than T2, due to the effect of magnetic field distortions. However, in the phantom experiment, T2* values closely matched T2. This is most likely due to the relatively high magnetic field homogeneity of the homogeneous phantom and the correction for macroscopic B0 distortions in the T2* maps. Furthermore, relatively long TE values up to 188 ms were chosen in the T2 mapping protocol which allows for the measurement of a broad range of in vivo T2 values but may reduce the accuracy for a homogeneous phantom with a transverse relaxation time of about 60 ms.

The presented investigation assesses the reproducibility of qMRI methods between two scanner models of the same vendor with different software versions. However, it should be noted that a comparison between different qMRI protocols or different software versions on the same scanner or between scanner models from different vendors was not a part of this study. Furthermore, the sample size of the presented study is limited. The investigation of the interaction of different effects on qMRI reproducibility in a larger cohort with a broader age range might be a promising approach for future qMRI studies.

5. Conclusions

The presented data suggest that qMRI parameter deviations between different scanner models from the same vendor are low, provided identical sequences are used for all scanners. Consequently, multi-centre qMRI studies with different scanner models can be expected to yield comparable data if either the investigated biological effects exceed the observed systematic differences between scanners or if the comparison of data acquired on different scanner models considers additional phantom or healthy subject data to correct for scanner-dependent deviations.

Author contribution

René-Maxime Gracien: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing - original draft; Writing - review & editing.

Michelle Maiworm: Data curation; Investigation; Validation; Writing - review & editing.

Nadine Brüche: Data curation; Investigation; Validation; Writing - review & editing.

Manoj Shrestha: Investigation; Software; Validation; Writing - review & editing.

Ulrike Nöth: Investigation; Software; Validation; Writing - review & editing.

Elke Hattingen: Conceptualization; Investigation; Validation; Writing - review & editing.

Marlies Wagner: Conceptualization; Investigation; Funding acquisition; Project administration; Supervision; Validation; Writing - review & editing.

Ralf Deichmann: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing - original draft; Writing - review & editing.

Declaration of competing interest

The authors report no conflicts of interest relevant to this study.

Dr R Deichmann received compensation as a Consultant for MR scanner procurement by the Wellcome Trust Centre for Neuroimaging, UCL, London, UK.

Acknowledgements

This work was supported by the Clinician Scientists program at Goethe University and by the State of Hesse with a LOEWE-Grant to the CePTER-Consortium (http://www.uni-frankfurt.de/67689811). The sponsors did not influence the study design or the collection, analysis or interpretation of data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116364.

References

Abbas, Z., Gras, V., Möllenhoff, K., Keil, F., Oros-Peusquens, A.-M., Shah, N.J., 2014. Analysis of proton-density bias corrections based on T1 measurement for robust quantification of water content in the brain at 3 Tesla. Magn. Reson. Med. 72 (6), 1735–1745. https://doi.org/10.1002/mrm.25086.

Al-Abasse, Y., Helms, G., 2016. Influence of pulse length and shape on variable flip angle T1 mapping of the human brain. In: Proc. Of the 24th Annual Meeting of the Intl. Soc. Mag. Reson. Med. Program Number 696.

Baudrexel, S., Nöth, U., Schüre, J.-R., Deichmann, R., 2018. T1 mapping with the variable flip angle technique: a simple correction for insufficient spoiling of transverse magnetization. Magn. Reson. Med. 79 (6), 3082–3092. https://doi.org/10.1002/mrm.26979.

Baudrexel, S., Volz, S., Preibisch, C., Klein, J.C., Steinmetz, H., Hilker, R., Deichmann, R., 2009. Rapid single-scan T2*-mapping using exponential excitation pulses and image-based correction for linear background gradients. Magn. Reson. Med. 62 (1), 263–268. https://doi.org/10.1002/mrm.21971.

Bauer, C.M., Jara, H., Killiany, R., 2010. Whole brain quantitative T2 MRI across multiple scanners with dual echo FSE: applications to AD, MCI, and normal aging. Neuroimage 52 (2), 508–514. https://doi.org/10.1016/j.neuroimage.2010.04.255.

Bjarnason, T.A., Vavasour, I.M., Chia, C.L.L., MacKay, A.L., 2005. Characterization of the NMR behavior of white matter in bovine brain. Magn. Reson. Med. 54 (5), 1072–1081. https://doi.org/10.1002/mrm.20680.

Boulby, A.P., Rugg-Gunn, F.J., 2003. T2: the transverse relaxation time. In: Tofts, P. (Ed.), Quantitative MRI of the Brain: Measuring Changes Caused by Disease. John Wiley and Sons Ltd, Chichester.

Buonincontri, G., Biagi, L., Retico, A., Cecchi, P., Cosottini, M., Gallagher, F.A., Gómez, P.A., Graves, M.J., McLean, M.A., Riemer, F., Schulte, R.F., Tosetti, M., Zaccagna, F., Kaggie, J.D., 2019. Multi-site repeatability and reproducibility of MR fingerprinting of the healthy brain at 1.5 and 3.0 T. Neuroimage 195, 362–372. https://doi.org/10.1016/j.neuroimage.2019.03.047.

Cercignani, M., Dowell, N.G., Tofts, P. (Eds.), 2018. Quantitative MRI of the Brain: Principles of Physical Measurement. CRC Press Taylor & Francis Group, Boca Raton FL.

Deoni, S.C.L., Williams, S.C.R., Jezzard, P., Suckling, J., Murphy, D.G.M., Jones, D.K., et al., 2008. Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5 T. NeuroImage 40 (2), 662–671. https://doi.org/10.1016/j.neuroimage.2007.11.052.

Fischer, H.W., Rinck, P.A., van Haverbeke, Y., Muller, R.N., 1990. Nuclear relaxation of human brain gray and white matter: analysis of field dependence and implications for MRI. Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med. 16 (2), 317–334.

Gracien, R.M., Jurcoane, A., Wagner, M., Reitz, S.C., Mayer, C., Volz, S., Hof, S.M., Fleischer, V., Droby, A., Steinmetz, H., Groppa, S., Hattingen, E., Deichmann, R., Klein, J.C., 2016a. Multimodal quantitative MRI assessment of cortical damage in relapsing-remitting multiple sclerosis. J. Magn. Reson. Imag. JMRI 44 (6), 1600–1607. https://doi.org/10.1002/jmri.25297.

Gracien, R.-M., Nürnberger, L., Hok, P., Hof, S.-M., Reitz, S.C., Rüb, U., Steinmetz, H., Hilker-Roggendorf, R., Klein, J.C., Deichmann, R., Baudrexel, S., 2017. Evaluation of brain ageing: a quantitative longitudinal MRI study over 7 years. Eur. Radiol. 27 (4), 1568–1576. https://doi.org/10.1007/s00330-016-4485-1.

Gracien, R.-M., Reitz, S.C., Hof, S.M., Fleischer, V., Zimmermann, H., Droby, A., Steinmetz, H., Zipp, F., Deichmann, R., Klein, J.C., 2016b. Changes and variability of proton density and T1 relaxation times in early multiple sclerosis: MRI markers of neuronal damage in the cerebral cortex. Eur. Radiol. 26 (8), 2578–2586. https://doi.org/10.1007/s00330-015-4072-x.

Gracien, R.-M., van Wijnen, A., Maiworm, M., Petrov, F., Merkel, N., Paule, E., Steinmetz, H., Knake, S., Rosenow, F., Wagner, M., Deichmann, R., 2019. Improved synthetic T1-weighted images for cerebral tissue segmentation in neurological diseases. Magn. Reson. Imag. 61, 158–166. https://doi.org/10.1016/j.mri.2019.05.013.

Hattingen, E., Jurcoane, A., Nelles, M., Müller, A., Nöth, U., Mädler, B., Mürtz, P., Deichmann, R., Schild, H.H., 2015. Quantitative MR imaging of brain tissue and brain pathologies. Clin. Neuroradiol. 25 (Suppl. 2), 219–224. https://doi.org/10.1007/s00062-015-0433-8.

Helms, G., Dathe, H., Weiskopf, N., Dechent, P., 2011. Identification of signal bias in the variable flip angle method by linear display of the algebraic Ernst equation. Magn. Reson. Med. 66 (3), 669–677. https://doi.org/10.1002/mrm.22849.

Howarth, C., Hutton, C., Deichmann, R., 2006. Improvement of the image quality of T1-weighted anatomical brain scans. Neuroimage 29 (3), 930–937. https://doi.org/10.1016/j.neuroimage.2005.08.004.

Kellman, P., Hansen, M.S., 2014. T1-mapping in the heart: accuracy and precision. J. Cardiovasc. Magn. Reson. 16 (1), 2. https://doi.org/10.1186/1532-429X-16-2.

Laule, C., Vavasour, I.M., Kolind, S.H., Traboulsee, A.L., Moore, G.R.W., Li, D.K.B., MacKay, A.L., 2007. Long T2 water in multiple sclerosis: what else can we learn from multi-echo T2 relaxation? J. Neurol. 254 (11), 1579–1587. https://doi.org/10.1007/s00415-007-0595-7.

Lee, Y., Callaghan, M.F., Acosta-Cabronero, J., Lutti, A., Nagy, Z., 2019. Establishing intra- and inter-vendor reproducibility of T1 relaxation time measurements with 3T MRI. Magn. Reson. Med. 81 (1), 454–465. https://doi.org/10.1002/mrm.27421.

Lorio, S., Tierney, T.M., McDowell, A., Arthurs, O.J., Lutti, A., Weiskopf, N., Carmichael, D.W., 2019. Flexible proton density (PD) mapping using multi-contrast variable flip angle (VFA) data. Neuroimage 186, 464–475. https://doi.org/10.1016/j.neuroimage.2018.11.023.

MacKay, A., Whittall, K., Adler, J., Li, D., Paty, D., Graeb, D., 1994. In vivo visualization of myelin water in brain by magnetic resonance. Magn. Reson. Med. 31 (6), 673–677.

McPhee, K.C., Wilman, A.H., 2015. T2 quantification from only proton density and T2-weighted MRI by modelling actual refocusing angles. Neuroimage 118, 642–650. https://doi.org/10.1016/j.neuroimage.2015.05.079.

Nelson, T.R., Tung, S.M., 1987. Temperature dependence of proton relaxation times in vitro. Magn. Reson. Imag. 5 (3), 189–199.

Nöth, U., Hattingen, E., Bähr, O., Tichy, J., Deichmann, R., 2015. Improved visibility of brain tumors in synthetic MP-RAGE anatomies with pure T1 weighting. NMR Biomed. 28 (7), 818–830. https://doi.org/10.1002/nbm.3324.

Nöth, U., Shrestha, M., Schure, J.-R., Deichmann, R., 2017. Quantitative in vivo T2 mapping using fast spin echo techniques - a linear correction procedure. Neuroimage 157, 476–485. https://doi.org/10.1016/j.neuroimage.2017.06.017.

Nöth, U., Volz, S., Hattingen, E., Deichmann, R., 2014. An improved method for retrospective motion correction in quantitative T2* mapping. Neuroimage 92, 106–119. https://doi.org/10.1016/j.neuroimage.2014.01.050.

Patenaude, B., Smith, S.M., Kennedy, D.N., Jenkinson, M., 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 56 (3), 907–922. https://doi.org/10.1016/j.neuroimage.2011.02.046.

Péran, P., Hagberg, G., Luccichenti, G., Cherubini, A., Brainovich, V., Celsis, P., Caltagirone, C., Sabatini, U., 2007. Voxel-based analysis of R2* maps in the healthy human brain. J. Magn. Reson. Imaging : JMRI 26 (6), 1413–1420. https://doi.org/10.1002/jmri.21204.

Petrovic, A., Scheurer, E., Stollberger, R., et al., 2015. Closed-form solution for T2 mapping with nonideal refocusing of slice selective CPMG sequences. Magnetic resonance in medicine 73 (2), 818–827. https://doi.org/10.1002/mrm.25170, 818–827.

Preibisch, C., Deichmann, R., 2009a. Influence of RF spoiling on the stability and accuracy of T1 mapping based on spoiled FLASH with varying flip angles. Magn. Reson. Med. 61 (1), 125–135. https://doi.org/10.1002/mrm.21776.

Preibisch, C., Deichmann, R., 2009b. T1 mapping using spoiled FLASH-EPI hybrid sequences and varying flip angles. Magn. Reson. Med. 62 (1), 240–246. https://doi.org/10.1002/mrm.21969.

Rooney, W.D., Johnson, G., Li, X., Cohen, E.R., Kim, S.-G., Ugurbil, K., Springer, C.S., 2007. Magnetic field and tissue dependencies of human brain longitudinal 1H2O relaxation in vivo. Magn. Reson. Med. 57 (2), 308–318. https://doi.org/10.1002/mrm.21122.

Schwarzbauer, C., Zange, J., Adolf, H., Deichmann, R., Noth, U., Haase, A., 1995. Fast measurement of temperature distributions by rapid T1 mapping. J. Magn. Reson., Ser. B 106 (2), 178–180. https://doi.org/10.1006/jmrb.1995.1029.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155. https://doi.org/10.1002/hbm.10062.

Stikov, N., Boudreau, M., Levesque, I.R., Tardif, C.L., Barral, J.K., Pike, G.B., 2015. On the accuracy of T1 mapping: searching for common ground. Magn. Reson. Med. 73 (2), 514–522. https://doi.org/10.1002/mrm.25135.

Teixeira, R.P.A.G., Malik, S.J., Hajnal, J.V., 2019. Fast quantitative MRI using controlled saturation magnetization transfer. Magn. Reson. Med. 81 (2), 907–920. https://doi.org/10.1002/mrm.27442.

Venkatesan, R., Lin, W., Haacke, E.M., 1998. Accurate determination of spin-density andT1 in the presence of RF-field inhomogeneities and flip-angle miscalibration. Magn. Reson. Med. 40 (4), 592–602. https://doi.org/10.1002/mrm.1910400412.

Volz, S., Nöth, U., Deichmann, R., 2012a. Correction of systematic errors in quantitative proton density mapping. Magn. Reson. Med. 68 (1), 74–85. https://doi.org/10.1002/mrm.23206.

Volz, S., Nöth, U., Jurcoane, A., Ziemann, U., Hattingen, E., Deichmann, R., 2012b. Quantitative proton density mapping: correcting the receiver sensitivity bias via pseudo proton densities. Neuroimage 63 (1), 540–552. https://doi.org/10.1016/j.neuroimage.2012.06.076.

Volz, S., Nöth, U., Rotarska-Jagiela, A., Deichmann, R., 2010. A fast B1-mapping method for the correction and normalization of magnetization transfer ratio maps at 3 T. Neuroimage 49 (4), 3015–3026. https://doi.org/10.1016/j.neuroimage.2009.11.054.

Wang, H.Z., Riederer, S.J., Lee, J.N., 1987. Optimizing the precision in T1 relaxation estimation using limited flip angles. Magn. Reson. Med. 5 (5), 399–416. https://doi.org/10.1002/mrm.1910050502.

Weingärtner, S., Roujol, S., Akçakaya, M., Basha, T.A., Nezafat, R., 2015. Free-breathing multislice native myocardial T1 mapping using the slice-interleaved T1 (STONE) sequence. Magn. Reson. Med. 74 (1), 115–124. https://doi.org/10.1002/mrm.25387.

Weiskopf, N., Suckling, J., Williams, G., Correia, M.M., Inkster, B., Tait, R., Ooi, C., Bullmore, E.T., Lutti, A., 2013. Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. Front. Neurosci. https://doi.org/10.3389/fnins.2013.00095.

Woermann, F.G., Steiner, H., Barker, G.J., Bartlett, P.A., Elger, C.E., Duncan, J.S., Symms, M.R., 2001. A fast FLAIR dual-echo technique for hippocampal T2 relaxometry: first experiences in patients with temporal lobe epilepsy. J. Magn. Reson. Imaging : JMRI 13 (4), 547–552.

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