



The traveling heads 2.0: Multicenter reproducibility of quantitative imaging methods at 7 Tesla

Maximilian N. Voelker^{a,b,*}, Oliver Kraff^a, Steffen Goerke^c, Frederik B. Laun^d, Jannis Hanspach^d, Kerrin J. Pine^e, Philipp Ehse^f, Moritz Zaiss^{d,g}, Andrzej Liebert^d, Sina Straub^c, Korbinian Eckstein^h, Simon Robinson^h, Armin N. Nagel^d, Maria R. Stefanescuⁱ, Astrid Wollrab^j, Sabrina Klix^k, Jörg Felder^l, Michael Hockⁱ, Dario Bosch^g, Nikolaus Weiskopf^{e,m}, Oliver Speck^{j,n}, Mark E. Ladd^{a,c,o}, Harald H. Quick^{a,b}

^a Erwin L. Hahn Institute for Magnetic Resonance Imaging, University Duisburg-Essen, Essen, Germany

^b High-Field and Hybrid MR Imaging, University Hospital Essen, University Duisburg-Essen, Essen, Germany

^c Division of Medical Physics in Radiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

^d Institute of Radiology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

^e Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

^f German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

^g Max Planck Institute for Biological Cybernetics, Tübingen, Germany

^h High Field MR Center, Department for Biomedical Imaging and Image guided Therapy, Medical University of Vienna, Vienna, Austria

ⁱ Chair of Cellular and Molecular Imaging, Comprehensive Heart Failure Center (CHFC), University Hospital Wuerzburg, Wuerzburg, Germany

^j Otto-von-Guericke-University Magdeburg, Magdeburg, Germany

^k Berlin Ultrahigh Field Facility (B.U.F.F.), Max-Delbrueck-Center for Molecular Medicine, Berlin-Buch, Germany

^l Institute of Neuroscience and Medicine (INM-4), Forschungszentrum Jülich, Jülich, Germany

^m Felix Bloch Institute for Solid State Physics, Faculty of Physics and Earth Sciences, Leipzig University, Leipzig, Germany

ⁿ Leibniz Institute for Neurobiology, Magdeburg, Germany

^o Faculty of Medicine and Faculty of Physics and Astronomy, Heidelberg University, Heidelberg, Germany

ARTICLE INFO

Keywords:

Magnetic resonance imaging
Ultrahigh field
Multicenter
Reproducibility
Quantitative imaging

ABSTRACT

Object: This study evaluates inter-site and intra-site reproducibility at ten different 7 T sites for quantitative brain imaging.

Material and Methods: Two subjects – termed the “traveling heads” – were imaged at ten different 7 T sites with a harmonized quantitative brain MR imaging protocol. In conjunction with the system calibration, MP2RAGE, QSM, CEST and multi-parametric mapping/relaxometry were examined.

Results: Quantitative measurements with MP2RAGE showed very high reproducibility across sites and subjects, and errors were in concordance with previous results and other field strengths. QSM had high inter-site reproducibility for relevant subcortical volumes. CEST imaging revealed systematic differences between the sites, but reproducibility was comparable to results in the literature. Relaxometry had also very high agreement between sites, but due to the high sensitivity, differences caused by different applications of the B1 calibration of the two RF coil types used were observed.

Conclusion: Our results show that quantitative brain imaging can be performed with high reproducibility at 7 T and with similar reliability as found at 3 T for multicenter studies of the supratentorial brain.

1. Introduction

In 2016, a multicenter magnetic resonance imaging (MRI) study at ultra-high magnetic field strength (UHF) was published (Voelker et al., 2016). The study was termed the “traveling heads” in which the same

two participants were imaged with the same imaging protocol at different 7 T sites. The study evaluated inter-site and intra-site reproducibility of 7 T brain imaging using common qualitative, structural pulse sequences, and found a high reproducibility at 7 T, and similar reliability compared to literature values for 3 T. With the latest introduction of 7 T MR systems (FDA clears first 7T magnetic resonance imaging device

* Corresponding author at: Erwin L. Hahn Institute for Magnetic Resonance Imaging, University Duisburg-Essen, Essen, Germany
E-mail address: Maximilian.Voelker@uni-due.de (M.N. Voelker).

2017), which are now available as FDA-approved and CE-labeled medical devices for brain and musculoskeletal applications, a follow-up of the previous study seems timely and shall focus on the reproducibility of state-of-the-art quantitative MR imaging across all three generations of 7 T MR systems.

Quantitative MRI is very challenging, but offers a high diagnostic potential especially at UHF. In conventional weighted imaging, the image contrast depends on multiple factors such as sequence type (e.g. which type of sequence was used for T1-weighted imaging), sequence parameters (e.g. repetition time or echo time) or hardware effects (e.g. B1 field or the type of radiofrequency coil or B0 homogeneity). Quantification of physical imaging parameters promises the inherent advantage that physical tissue properties, e.g. the longitudinal or transversal relaxation time or the tissue susceptibility value are examined that are not based on arbitrary units as in conventional weighted MRI. Therefore, quantification of imaging properties has the potential to increase diagnostic specificity and sensitivity due to its standardized nature but also due to its sensitivity to micro-structural properties of brain tissue such as axons, myelin, iron or water concentration (Weiskopf et al., 2015, Langkammer et al., 2012, MacKay et al., 2006, Harkins et al., 2016).

At UHF, quantitative MR imaging techniques not only benefit of the increased signal-to-noise-ratio (SNR), but also from additional effects such as stronger phase contrast for quantitative susceptibility mapping (QSM) or T1 lengthening and higher spectral dispersion for chemical exchange saturation transfer (CEST) (Pohmann et al., 2016, Ladd et al., 2018). Thus, the concepts of *in vivo* histology at ultra-fine image resolution has been proposed and found clinical use at 7T (Weiskopf et al., 2015, Deistung et al., 2013, Spincemaille et al., 2020, Straub et al., 2020, Louapre et al., 2015).

The higher sensitivity at UHF, however, may also increase the artifact-to-noise ratio (MacKay et al., 2006) and methods that rely on B0 or radio-frequency (RF) field uniformity e.g. some T1-mapping methods or CEST become more challenging due to increased field inhomogeneities and RF wavelength effects, which are more prominent at higher field strength (Uğurbil, 2018). This may influence the image quality and reproducibility for quantitative imaging at UHF.

In this study we addressed this task and examined the reproducibility of state-of-the-art quantitative imaging methods in a multicenter 7 Tesla study. We evaluated the reproducibility of brain volumetry and T1 mapping with the Magnetization Prepared 2 Rapid Acquisition Gradient Echoes (MP2RAGE) approach (Marques et al., 2010), quantitative susceptibility mapping based on multi-echo gradient-echo (GRE) imaging (Li et al., 2014, Wu et al., 2012, Wei et al., 2015), chemical exchange saturation transfer based on the GRE snapshot approach (Zaiss et al., 2018), and relaxometry of T1, T2*, as well as mapping of proton density (PD) based on a multi-parametric mapping technique (Weiskopf et al., 2013, Kirilina et al., 2020). The results were compared to those from the previous study and to reproducibility studies for the specific quantitative imaging methods available in the literature.

2. Material and methods

2.1. Measurement setup

Two male subjects (41 and 36 yrs.) were imaged at ten different UHF sites. All sites operate 7 T whole-body MR systems from the same vendor (Siemens Healthcare GmbH, Erlangen, Germany). The sites cooperate in a user group network called German Ultrahigh Field Imaging (GUFi, www.mr-gufi.de) and are located in Germany and neighboring countries (Austria and the Netherlands). Data were acquired at the following 7 T sites: Berlin, Germany (BER); Bonn, Germany (BON); Erlangen, Germany (ERL); Essen, Germany (ESS); Heidelberg, Germany (HEI); Jülich, Germany (JUL); Leipzig, Germany (LEI); Magdeburg, Germany (MAG); Vienna, Austria (VIE); Würzburg, Germany (WUE). Although all based on the same vendor, the sites operate different generations of 7 T MR systems with different software platforms and differences in basic imag-

ing components that might influence image quality. Five different configurations were identified from combinations of three different magnet versions, three different gradient coil sets, two different RF head coils and two different software versions (Table 1: Sites). Additionally, the latest system version (configuration 5, “Terra”, C5) came with RF power amplifiers with higher maximum output power (11 kW instead of 8 kW) as well as B0 shimming with 3rd-order terms. The oldest available 7 T MR systems within the GUFi network have already been running for up to 15 years, while the most recent systems of the FDA/CE-approved C5 configuration were installed since 2015.

2.2. Imaging protocol and image processing

2.2.1. System calibration

The imaging protocol (Inline Supplementary Table 1) started with system calibration. B1 was mapped and adjusted with the Dual Refocusing Echo Acquisition Mode (DREAM) B1 mapping technique (Ehse et al., 2019, Nehrke and Börnert, 2012). This sequence is routinely not available at the systems and was shared via core competence partnerships (C2P) between the sites. The version we used has an integrated calculation of a so-called “transmitter reference voltage map” that represents the required input power to reach the desired flip angle in each voxel. The input power was calibrated for a manually chosen region of interest (ROI) drawn centrally with a circular extension to the lateral brain borders and averaged over 3 central axial slices oriented along the connection between the lateral ventricles and the 3rd ventricle. After the adjustment, B1 was measured again with DREAM to evaluate the reproducibility of the adjustment between sites. The measured flip angle relative to the nominal flip angle was used as relative B1 and analyzed in B1 histograms.

B0 was shimmed iteratively using automatic shimming routines provided by the vendor. The shimming volume was chosen manually as a cuboid that enclosed the longest brain axis oriented along the AC-PC line. For the C5 sites, 3rd-order shimming was enabled, while at the other sites 2nd-order shimming was used. The results of the B0 calibration were mapped with a GRE B0 field mapping sequence, which was repeated at the end of the imaging protocol (Inline Supplementary Table 1) to check for B0 drift. Standard deviation and histograms were calculated to compare B0 homogeneity between sites.

2.2.2. MP2RAGE

Volumetric measurements were based on image data acquired with the MP2RAGE (Marques et al., 2010) technique. MP2RAGE is a widely used technique in 7 T brain imaging as it provides T1-weighted contrast with reduced transmit field inhomogeneity without bias of T2* weighting, PD weighting or the reception field. The version we used was available as a works-in-progress (WIP) version provided by the vendor at the C1–C4 sites and as standard product sequence at the C5 sites (Inline Supplementary Table 1). The “uniform” images and T1 maps were additionally corrected for transmit B1 inhomogeneity (Marques and Gruetter, 2013). Therefore, the mean of both DREAM contrasts was used for co-registration to the MP2RAGE scan. Subsequently, the B1 maps were used for B1 correction of the MP2RAGE data, but also the MP2RAGE generated masks were used for analysis of the B1 maps in different brain sub-volumes. To create these analysis masks, the subsequent image-processing pipeline consisted of the following steps. Bias correction was realized with the N4 function implemented in Advanced Normalization Tools (ANTs, v2.1.0) (Tustison et al., 2010). Brain extraction was done by combining the results of the method implemented in ANTs (Avants et al., 2011) and the FSL brain extraction tool (FMRIB Software Library v6.0, The University of Oxford) (Smith, 2002), and the results were refined with filters for dura and arteries similar to (Haast et al., 2018) as implemented in CBS High-Res Brain Processing Tools (CBS Tools v.3.1.0, Max Planck Institute for Human and Cognitive Brain Sciences) (Bazin et al., 2014). Volumes of interest (VOIs) for analysis were defined that include different tissue types (FSL-FAST

Table 1

Sites. Five different hard- and software configurations were identified at the GUFU sites. All 7T MR systems are from the same vendor. Configurations 1-4 use magnets manufactured by Agilent Technologies (Santa Clara, CA, USA); Configuration 5 by Siemens (Siemens Healthcare GmbH, Erlangen, Germany). The radiofrequency power amplifier (RPFA) was more powerful at the C5 sites. All sites are equipped with a commercially available RF head coil (Nova Medical, Inc., Wilmington, MA, USA) that was used for the measurements, whereby the current version has 32 and the older version 24 receive channels and a slightly different transmit component.

Configuration type	Magnet type	Gradient coil	RPFA	RF Coil	Software version	Location (Scans)
1 (PS, AS95, 24ch)	Passively shielded (PS)	38 mT/m (“AS95”)	8 kW	24 ch	VB	Berlin (1) Heidelberg (1)
2 (PS, AS95)	Passively shielded (PS)	38 mT/m (“AS95”)	8 kW	32 ch	VB	Essen (2)
3 (PS, SC72)	Passively shielded (PS)	70 mT/m (“SC72”)	8 kW	32 ch	VB	Magdeburg (1) Leipzig (1) Vienna (1)
4 (AS, SC72)	Actively shielded (AS)	70 mT/m (“SC72”)	8 kW	32 ch	VB	Bonn (1)
5 “Terra”	Actively shielded	80 mT/m	11 kW	32 ch	VE	Erlangen (3) Würzburg (2) Jülich (2, Subj.2: 1)

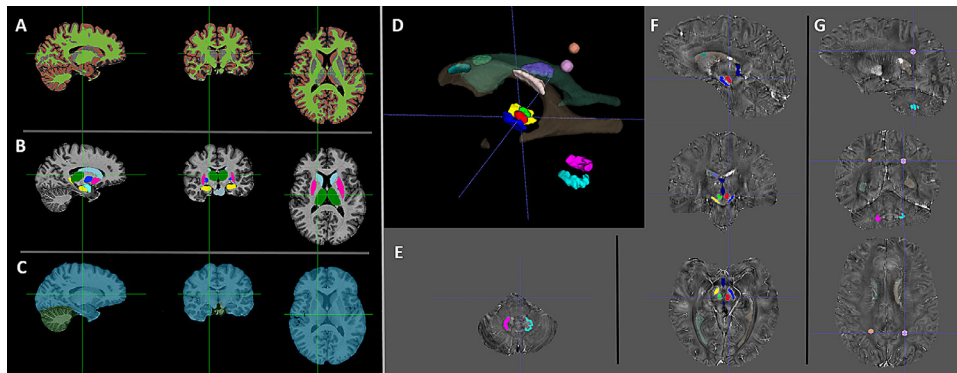


Fig. 1. VOIs. VOIs used for reproducibility comparisons. Based on T1-weighted imaging, tissue types (A) and subcortical VOIs (B) have been segmented with FSL tools for each subject at each site. Whole-brain masks were derived from brain extraction and subdivided (C) to compare the cranial part without the influence of inferior parts or subcortical areas (blue = supratentorial brain mask), where B1 and B0 inhomogeneities are most prominent. Semi-automatic segmentation was used for delineation of additional VOIs for QSM analysis (D-G). Red nucleus (green/red, D,F), substantia nigra (yellow/blue D,F), dentate nuclei (magenta/cyan, D,E) as well as reference VOIs in WM (orange/pink D,G) and lateral ventricles

(light-green/blue-green, beige/light-blue, D,F) were segmented for each subject at one site. Reference volumes in the lateral ventricles (D,F) were chosen in a frontal cranial region with homogeneous susceptibility and in a medial region (between frontal and dorsal part). WM VOI (G) for QSM was chosen as 10-voxel-diameter (7 mm) ball centrally in the biggest WM fiber bundle visible supracallosal in the parietal lobe.

(Zhang et al., 2001)) and different subcortical structures (FSL-FIRST (Patenaude et al., 2011)). The subcortical segmentation with the FSL “run_first_all” script was implemented after gradient distortion correction (GDC) (Jovicich et al., 2006) and warping of the uniform images into Montreal Neurological Institute (MNI) image space was performed with ANTs. GDC was realized with the “gradunwarp” script recommended in the Human Connectome Project pipeline (Glasser et al., 2013) using the vendor provided correction matrices for each gradient coil type. The Jacobian intensity correction was not used to preserve the quantitative data measured by the individual scanners. Warping was realized with ANTs to increase the reproducibility (Feng et al., 2017). Different additional brain masks were defined originating from warps of the MNI template masks and the FIRST segmentations to discriminate between whole-brain and supratentorial structures (i.e. without the inferior parts such as cerebellum, brain stem etc.) (Fig. 1: VOIs). The reproducibility of measured (MP2RAGE) T1 maps and the size of the different VOIs were analyzed in the B1- and GDC-corrected images and compared between sites and rescans of the subjects. The coefficient of variation (CoV = standard deviation/mean) was used as measure for inter-site reproducibility. For intra-site data the relative span of the mean VOI values was calculated for the sites with 2 or 3 measurements and the maximum was used as a measure for the highest repetition error.

Due to the different required transformation steps, the tissue segmentation masks were thresholded by a margin of 0.9 and the subcortical segmentation masks by 0.6 to reduce partial volume effects. Therefore, we included boundary voxels that had an overlap of at least 90% to the segmented tissue type after the final transformation and 60% for subcortical VOIs respectively. The same VOI masks were used for analysis of all image data, and images were co-registered to align the masks between different measurements using ANTs.

Inter-site co-registration for the graphical analysis of the results was realized between the denoised, bias and distortion corrected uniform MP2RAGE contrasts. The images were masked with their individual brain mask before co-registration to use the brain boundaries as edge features. Mutual information was used to calculate the cost function in the iterative registration approach in ANTs with parameters as in the “antsRegistrationSyN” script modified to a higher convergence threshold (10^{-9} instead of 10^{-6}) for the optimization. A rigid co-registration step was used to analyze the volumetric overlap of the distortion corrected images. With the same parameters an additional affine co-registration step was calculated based on the rigid co-registered data to optimize the co-registration overlap.

2.2.3. QSM

QSM has become very popular over the last years (Haacke et al., 2015) and benefits from the larger phase differences at UHF and therefore higher achievable contrast-to-noise ratio (Ladd et al., 2018, Deistung et al., 2013, Spincemaille et al., 2020). We acquired the QSM data with a multi-echo gradient echo (ME-GRE) sequence (Inline Supplementary Table 1) that uses the ASPIRE coil combination technique (“A Simple Phase Image Reconstruction for multi-Echo data”) (Eckstein et al., 2018) and monopolar readout. The sequence that integrated this approach is routinely not available at the systems and was shared via C2P (Siemens Healthcare GmbH) between the sites. To generate susceptibility maps of each echo time, phase maps were unwrapped using a Laplacian-based phase unwrapping algorithm (Li et al., 2014, Schofield and Zhu, 2003). Brain masks were generated at one site for each subject and each echo of the ME-GRE using the HD-BET algorithm (Isensee et al., 2019). An affine co-registration (ANTs) of the first echo magnitude images was calculated between all sites and the reference

site to transfer these masks to all sites' image space. Thus, it was attempted to include the same brain volume for each subject at each site. Subsequently, the background field was removed using Sophisticated Harmonic Artifact Reduction for Phase data (V-SHARP) with varying spherical kernel sizes up to 12 mm (Wu et al., 2012). Finally, susceptibility maps were calculated by using the STreaking Artifact Reduction for QSM (STAR-QSM) (Wei et al., 2015) algorithm. An echo-time-averaged susceptibility map was created by using the squared echo time and the squared signal magnitude as a weight (Chen et al., 2018). Algorithms from the STISuite Toolbox (v.3.0, University of California, Berkeley) and the MEDI Toolbox (01/15/2020, Cornell MRI Research Lab, New York) were used for the QSM reconstruction pipeline. For analysis of the QSM maps, the masks generated via the corresponding MP2RAGE analysis via affine co-registration were used. A single image erosion (3-voxel-square kernel) was applied to these analysis masks to account for partial volume due to the additional transformation steps. But also a semi-automatic segmentation of QSM-relevant VOIs was performed with ITK-SNAP (v.3.8.0) (Yushkevich et al., 2006) at the C2 reference site for both subjects and the resulting masks (Fig. 1: VOIs.). These VOIs were transferred and applied in the original imaging space of the QSM maps at each site. The semi-automatically segmented VOIs were used to compare the data to the results of other groups and may be used as reference for inter-subject normalization.

2.2.4. CEST

At UHF, CEST imaging (Ward et al., 2000) gains SNR and the dispersion of the peaks in the Z-spectra is greater, but the acquisition may suffer from less homogeneous B1 excitation (Ladd et al., 2018, Kim et al., 2015). CEST MRI was implemented according to a previously optimized acquisition protocol at 7 T (Zaiss et al., 2015, Zaiss et al., 2015) in combination with recent improvements to enhance the reproducibility (Katrin, 2020). The protocol comprises a low-power pre-saturation (reconstructed mean B1 = 0.7 μ T, flip angle equivalent), a 5-pool Lorentzian-fit analysis (i.e. rNOE, amide, amine, MT and water (Windschuh et al., 2015)), and a correction for B0 and B1 inhomogeneities using the field maps from the WASABI measurement (Windschuh et al., 2015, Schuenke et al., 2017) included in the CEST protocol. In contrast to the previous protocol, the image readout was extended to three dimensions using the snapshot-CEST approach (Zaiss et al., 2018) with adapted imaging parameters (Inline Supplementary Table 1). In addition, Z-spectra were de-noised using a principal component analysis (i.e. 12 preserved components) (Breitling et al., 2019). The final CEST contrast for each pool was calculated by the Lorentzian difference (LD) metric (Jones et al., 2013, Zaiss et al., 2014) using the individual label and reference Z-spectrum of the multi-pool Lorentzian fit analysis at -3.5, +3.5 and -2.0 ppm, respectively (Zaiss et al., 2014). The amine pool at +2.2 ppm was not evaluated individually but was used to stabilize the amide contrast. For the reference Z-spectrum of the amide signal, the Lorentzian was combined with the one dedicated to the amine pool to avoid instabilities from the overlap of the two spectrally neighboring pools. The sequence that realized this protocol is routinely not available at the systems and was shared via C2P between the sites. All subsequent image processing steps were performed offline from the scanner. An additional calibration scan was necessary due to the hardware constraints of the systems. This calibration was measured with a phantom and used the same sequence modified to measure the difference of the flip angle with and without the application of the pre-saturation pulse chain. Therefore, the influence of thermal heat on the output power was assessed to correct for systematic differences between the different systems (Supplementary data). For analysis all CEST datasets were co-registered to the corresponding MP2RAGE scan, and the GM and WM VOIs as well as the caudate VOI that had significant overlap of more than 50% with the common CEST slab, which was oriented cranial adjacent to the AC-PC line, were used for reproducibility comparisons.

2.2.5. Relaxometry

Quantitative maps of the T1 and T2* relaxation rates ($R1 = 1/T1$, $R2^* = 1/T2^*$) as well as PD were acquired using a multi-parametric mapping technique (Weiskopf et al., 2013, Kirilina et al., 2020) further adapted for 7 T and for increased isotropic resolution (Inline Supplementary Table 1). The measurement of the basic MRI tissue parameters is particularly important in quantitative imaging and may, in combination with biophysical modeling (Kirilina et al., 2020, Henkelman et al., 2001), enable the concept of *in vivo* histology using MRI (hMRI, (Weiskopf et al., 2015)). The protocol consisted of two 3D ME-GRE scans with T1 and PD weightings (measurement of magnetization transfer saturation was omitted due to scan time concerns). The sequence, which was both RF- and gradient-spoiled, is routinely not available at the systems and was shared via research agreements between the sites. Maps were calculated with the hMRI toolbox (v.0.2.2., <http://hmri.info>, (Tabelow et al., 2018)) within the SPM 12 framework (<http://www.fil.ion.ucl.ac.uk/spm/>) in Matlab (R2017b, The Mathworks, Inc., Natick, MA; USA). The calculation of the parameters was based on the Ernst equation (Ernst and Anderson, 1966, Helms et al., 2008, Helms et al., 2008), where corrections for B1 transmit and receive field bias were implemented. R1 and PD in percentage units (p.u.) were estimated from a solution of the Ernst equation, using approximations for small repetition time TR (Tabelow et al., 2018, Helms et al., 2008). To combine the multiple echoes, PD- and T1-weighted signals were extrapolated to TE = 0. The B1 transmit field correction is realized with the correction field f_t measured with the DREAM B1 mapping sequence (one scan aligned to the MPM slab), the effective local flip angle is calculated as:

$$\alpha_{cor} = f_t \alpha_{nom} \text{ and } T_{1app} \sim T_1 f_t^2$$

Corrections for imperfect spoiling were applied (Preibisch and Deichmann, 2009). R2* was calculated across all echoes of the multi-contrast data by weighted ordinary least squares fitting.

For analysis, the first echo of the PD-weighted scan was co-registered to the second inversion contrast of the MP2RAGE measurement and the resultant matrices were used to transfer the VOI masks to the relaxometry image space. The transformed analysis masks were thresholded by a margin of 0.99 to exclude most of the partial volume, which effectively resulted in a mild erosion.

2.3. Datasets

Including rescans at selected sites, Subject 1 was imaged 15 times and Subject 2 was imaged 14 times at 10 different sites, leading to a total of 29 datasets (Inline Supplementary Table 2). The image data we acquired can be accessed online via zenodo (DOI: [10.5281/zenodo.4117947](https://doi.org/10.5281/zenodo.4117947)). The first measurement at the C2 site was chosen as reference for image co-registration of each subject. All were evaluated and further compared to the datasets obtained from the previous traveling heads study performed in 2016 (Voelker et al., 2016). The current study included the same two subjects as the previous experiment, which allowed to compare the data individually and quantitatively to the previous results. The total measurement time for each dataset was 60 min per subject.

2.4. Exceptions

Exceptions occurred for various measurements. At the C1 sites with the 24-channel version of the RF head coil, the required input power to reach the nominal flip angle was measured to be higher, which led to SAR problems. As the power constraints were more restrictive for this RF coil, the maximum applicable input power was set to a fixed value (350 V transmitter reference voltage), for which most of the measurements would run without further modifications instead of using the calculated necessary input power. Therefore, the applied B1 was smaller for all

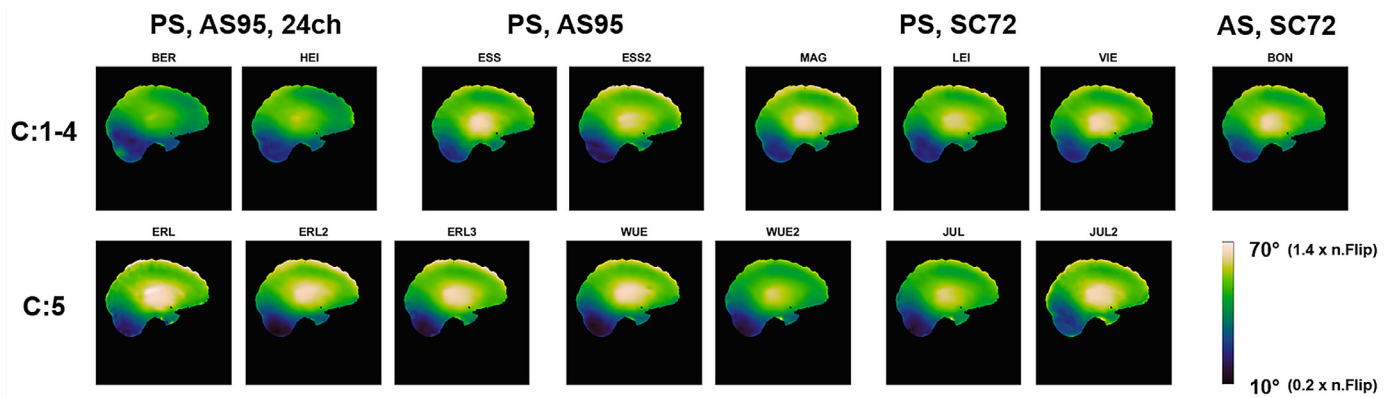


Fig. 2. B1. B1 field maps were obtained after transmitter calibration to yield a nominal flip angle of 50°. All maps show Subject 1 including rescans and were co-registered to the first measurement at the only C2 site (ESS). Note that the C1 sites with the 24-channel RF coil were limited to a fixed value of transmit power (350V transmitter reference voltage). Since this coil type was less power-efficient, SAR constraints of other measurements in the imaging protocol limited the maximum applicable input power.

measurements at these C1 sites. Modifications were necessary for the B0 mapping (four slices less in BER, six in HEI) and for the T1-weighted ME-GRE scan, that used a nominal FA of 20° (BER) or 19° (HEI) instead of 21°.

The B1 calibration for one measurement at one C5 site (JUL2) was accidentally measured with lower nominal flip angle (35° instead of 50°), which led to different SNR of the measurement and slight differences in measurement accuracy. The B1 mapping had also some minor differences between the VE software version at the C5 sites and VB version at the other sites, and different acceleration methods were used (VE with CAIPIRINHA (Breuer et al., 2005), VB with GRAPPA (Griswold et al., 2002) and identical acceleration factors).

For the B0 shimming, one exception occurred during the first scan at a C5 site (Erlangen), where inadvertently a different shimming program was used that led to degraded results. Consequently, the data were not used for analysis.

The MP2RAGE sequence uses a sequence optimizer, which minimizes TE and that cannot be switched off. Therefore, slight differences in TE (2.93–3.06 ms) had to be used at the sites with different gradient coils. For the MP2RAGE measurements in Leipzig and Vienna (two C3 sites), the inversion pulse (HS4) was not loaded correctly probably due to conflicts with other versions of that sequence installed on the systems. Hence, MP2RAGE data are not available for these sites.

The additional CEST calibration for duty cycle could not be done at one C3 site (LEI), as the scanner had been replaced in the meantime.

The ME-GRE sequence, which was shared via C2P for the relaxometry measurements, was not applied successfully at the C5 sites with VE software due to various sequence failures that arose from software adaptation. Consequently, the reproducibility analysis was restricted to C1-C4 sites.

3. Results

3.1. System calibration

The transmitter voltages of the 7 T systems that used the 32-channel RF coil (C2-C5) were successfully calibrated with DREAM B1 mapping. A mean relative B1 of 92% was measured for the whole brain (WB) volume, and 96% in the supratentorial brain (SB). Variations were small between sites ($\pm 5\%$ at maximum) for each subject individually and between subjects at all sites ($< 2\%$ variation of the inter-site average), where the individual maximum inter-subject difference at the same site was found to be 10%. The transmit power at the C1 sites with the 24-channel RF coil was fixed to 350 V leading to lower B1 values (84% / 86% in WB / SB, Fig. 2: B1, Inline Supplementary Table 3) and narrower

B1 distributions at those sites (Fig. 3: Histogram, Inline Supplementary Table 3).

In contrast, the B0 shimming performed slightly differently between the five hardware configurations of the sites (Fig. 4: B0). The 3rd-order shimming routines of the C5 sites allowed higher field homogeneity than the 2nd-order shimming at the C1-C4 sites, where the mean full width at half maximum (FWHM) of the peak of the whole-brain B0 distribution was about 1.7 times broader than at the C5 sites for both subjects (Fig. 3: Histogram, Inline Supplementary Table 3). Furthermore, the C1-C2 sites slightly outperformed the C3-C4 sites for both subjects with a 39% and 32% broader FWHM at the C3-C4 sites for Subject 1 and 2, respectively. However, using the standard deviation (SD) of the B0 maps as the criterion for homogeneity, the differences between the sites were much smaller with similar tendencies slightly favoring the C1-C2 sites. The mean SD for both subjects was equivalent between the C5 and the C1-C2 sites, while the C3-C4 sites had about 5% higher SD of the whole-brain B0 maps.

3.2. MP2RAGE

MP2RAGE showed very high agreement in image contrast, measured T1 values and volumetric measurements (Inline Supplementary Table 4, Inline Supplementary Figure 1). The gray matter/white matter (GM/WM) contrast in the T1-weighted uniform images was measured to be -0.29 (Michelson definition (Michelson, 1927)) for the supratentorial brain of both subjects. Comparing the reproducibility of the T1-values determined in corresponding T1 maps, an inter-site CoV of less than 2% was found for all extracted sub-volumes in the brain of both subjects. Exceptions could only be found for the most inferior regions, i.e. brain stem and cerebellum, where B1 and B0 homogeneity is highly reduced (Fig. 2: B1, Fig. 4: B0) leading to a CoV of 6.8% as the worst case for one subject. The average inter-site CoV of the 14 subcortical nuclei segmented with FSL FIRST (excluding the brainstem and cerebellum due the low B1 in the inferior region and consequent lower reliability) of both subjects was three times smaller if calculated only between the C5 sites (0.5%) compared to a group of the older generation C1-C4 sites (1.5%), which showed slightly higher variation than the overall average of 1.3%. Indeed 27/28 of the C5 VOIs had higher reproducibility than the respective VOIs of the C1-C4 group. The maximum intra-site repetition errors for all rescans were found between 0.3% and 3% for the single subcortical VOIs with an average of 1.3% for subject 1 and 1.9% for subject 2.

The variation of the examined size of the sub-volumes revealed high inter-site reproducibility, with CoV of less than 2.2% if the specific sub-volume had a size of at least 2 cm³. For smaller regions such as, for ex-

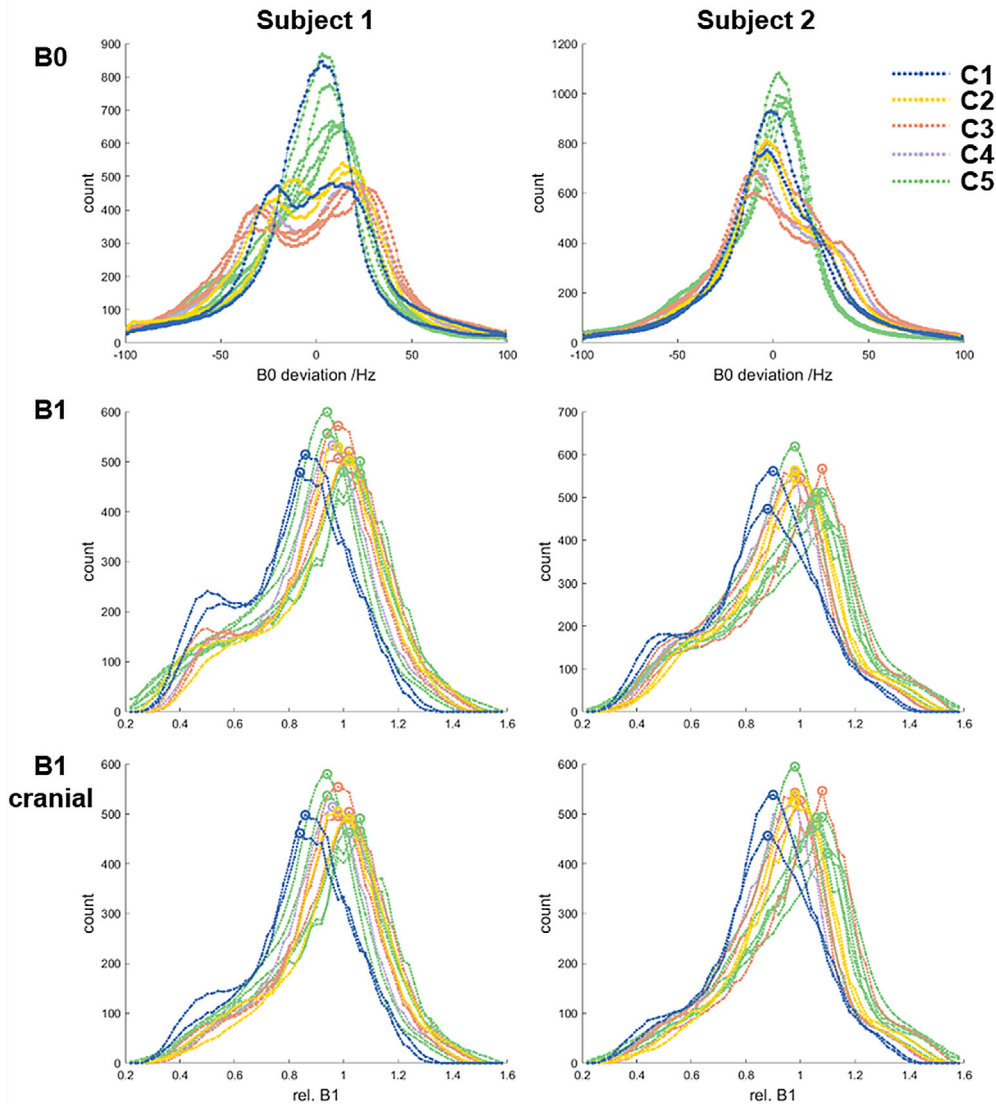


Fig. 3. Histogram. Histograms of the field maps. B0 and B1 histograms were plotted from data of the whole-brain masks after multiplying in common space to measure in the overlapping region only. Plots were denoised in Matlab by a moving average low-pass filter (9 values kernel) for better depiction of the differences. Subject 1 had a $1.3 \times$ broader B0 field distribution than Subject 2. The 3rd-order shimming at the C5 sites reduced the average peak width of the distribution by 36%. The C1 sites showed a 26% smaller bandwidth than the C2-C4 sites. However, the B1 distribution revealed that the B1 maps at C1 were measured with lower input power, which moved the peaks of the distributions to about 15% lower relative B1 and narrowed the B1 distribution slightly (maximum count marked with dots). By masking out the inferior brain regions, the cranial B1 distributions became (tighter and) more centered towards the nominal flip angle (relative B1 = 1) due to the low transmit variation of the RF coils in this region.

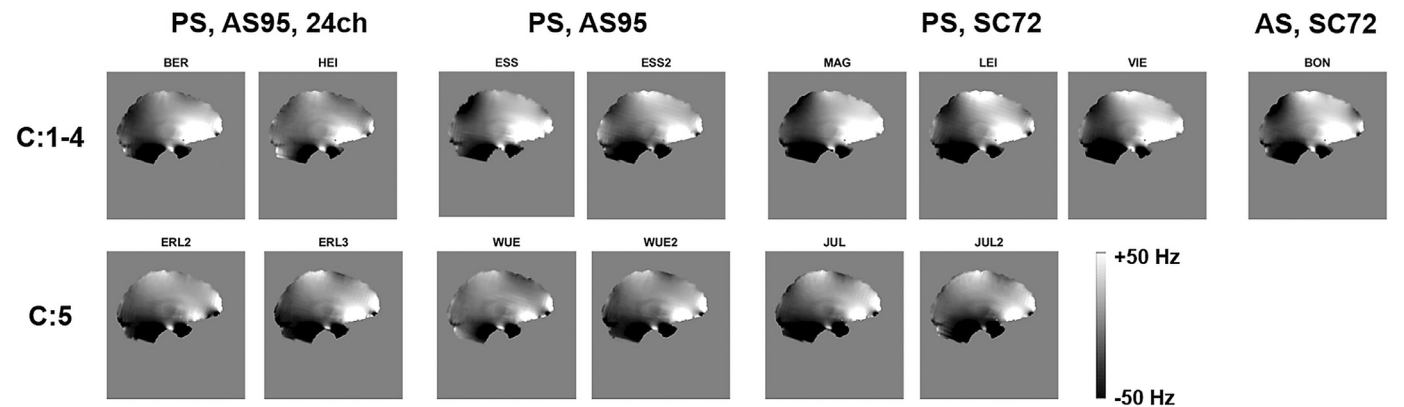


Fig. 4. B0. B0 field maps. B0 shimming was conducted with 2nd-order shimming at C1-C4 sites and 3rd-order shimming at the C5 “Terra” sites. The 3rd-order shimming yielded better homogeneity compared to the other 32-channel coil sites for Subject 1 shown here. The measurements at the C1 sites with the 24-channel coil, especially HEI, had more homogeneous field distributions than the C2-C4 sites. The same co-registered excerpt from the Subject 1 dataset is shown as for Fig. 2: B1. (excluding ERL1, see exceptions in the text).

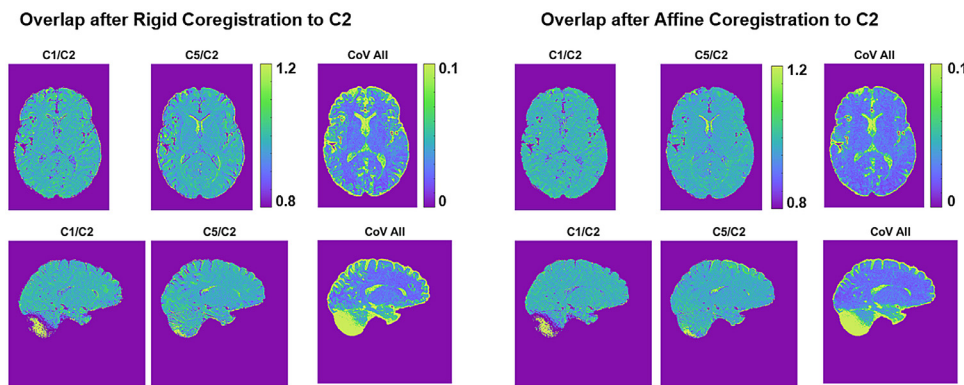


Fig. 5. Co-registration performance. The mean overlap of the co-registered subject 1 datasets of configuration C1 and C5 is shown normalized to the C2 dataset as relative deviation of the co-registered MP2RAGE T1 maps. Rigid co-registration with ANTs was not sufficient which can be found as high contrast areas with structural visibility in the difference maps especially at the WM-GM-CSF borders of the cerebral cortex. This misalignment degrades the graphically calculated CoV maps. An additional affine co-registration step was necessary to reduce the influence of misalignment to the CoV. The co-registration showed better overlap for both rigid as well as the affine computations while configurations using the same gradient coil type had been

co-registered (C1/C2). The hardware configuration of C1 used the same gradient coil type but different RF coil than C2, while the C5- sites used the same RF coil but a different gradient coil type with higher gradient strengths but lower linearity.

ample, right amygdala or right globus pallidus, as well as for the most inferior brain parts, the segmentation was found to be less reproducible resulting in a variation of the measured size of up to 10% (Inline Supplementary Table 4). Small absolute errors in delineation of the specific region contribute much to the relative coefficient of variation here. Inter-site CoVs of subcortical volume measurements did not show differences between subgroups of the different site configurations. For the big subcortical VOIs ($>2 \text{ cm}^3$) the maximum intra-site repetition errors of the rescans were measured between 0.6% and 5.1% with an average of 3.0% for subject 1 and 2.5% for subject 2. For smaller VOIs e.g. the right globus pallidus worst case volume deviations of up to 36% could be identified out of the small dataset of 2-3 scans per site.

Rigid co-registration was unable to align image data of all configurations correctly to the C2 reference site. Affine co-registration showed much better overlap to the reference especially those configurations that had installed the different type of gradient coil (C3-C5) (Fig. 5: Co-registration performance, Inline Supplementary Figure 2). Consequently the affine co-registration matrices were used to align all graphical representations (all figures) of the measurement results to the common MP2RAGE reference image measured at C2. The quantification itself was done in the native image space for each measurement, with co-registration necessary only between volumetry/MP2RAGE and the respective method (exception is the manual segmentation of the additional QSM VOIs, which was co-registered via affine MP2RAGE matrices).

3.3. QSM

The QSM maps also showed a very high agreement between sites (Fig. 6: QSM, Inline Supplementary Table 5). Small regions of higher variation were identified near air cavities and vessels, e.g. near the paranasal sinuses (Fig. 9: Reproducibility). The quantification of distinct subcortical gray matter VOIs yielded a slightly higher variation between sites than the MP2RAGE analysis. The most reproducible results between sites were found for the dentate nuclei, where the CoV ranged between 2-4% (Inline Supplementary Table 5). The three reference volumes in the lateral ventricles and the occipital WM had similar inter-site variations as the three manually segmented VOIs (red nucleus, substantia nigra, dentate nuclei) and the three VOIs derived from FSL FAST segmentation of the MP2RAGE images (caudate nuclei, globus pallidus, putamen). For Subject 1, the inter-site variations were slightly higher than for Subject 2, with a maximum CoV reaching 13% at the left WM reference volume or 11% at the right substantia nigra, whereas for Subject 2 the CoV was lower than 10% in all measured sub-volumes. Comparing the measured susceptibility between subjects, the maps showed about 1.25 times ($R^2=0.97$) higher susceptibility values over all VOIs for Subject 2 than for Subject 1. Systematic differences in reproducibility between site configurations or rescans could not be identified.

3.4. CEST

The initial CEST analysis for the acquired rNOE maps revealed a bias between the data measured at the C5 sites and the data measured at the older generation C1-C4 sites, whereby the rNOE contrast at the C5 sites exceeded that at the other sites by a factor of 9.5% in GM and WM tissues. Potential thermal effects of the RF power amplifier were investigated and found to lead to different flip angles played out between the sites by the individual RFPAs (Supplementary data). The long pre-saturation phase of the sequence ($140 \times 15\text{-ms}$ Gaussian shaped pulses were applied with a delay of 10 ms per RF block) caused very high thermal stress to the RFPAs. Therefore, deviations of the flip angle after pre-saturation were observed, leading to an altered effective pre-saturation RF amplitude. The mean deviation in B1 measured at the sites was 6.3%, while the best site had about 3.3% and the highest deviation was found to be 11.2%. After applying these correction factors to the data, the maps showed good agreement between all sites (Fig. 7: CEST, Inline Supplementary Table 6), and the bias between the C1-C4 and C5 sites was reduced to 1.4% in GM and WM tissues. The mean rNOE contrast differed for GM (0.22) and WM (0.25) areas, and the differences between subjects were minimal (<0.01). The inter-site reproducibility was between 2.4% and 3% for GM and WM tissues and between 3.8% and 4.7% for the caudate nuclei, the only subcortical VOI that had significant overlap with the CEST imaging slab. The inter-site reproducibility of the C5 configuration was between 1.1% and 1.6% for GM and WM, while the group of the configurations C1-C4 with more heterogeneous hardware setup had slightly lower inter-site reproducibility (2.9–3.9%).

The contrast of the Amide peak did also show good agreement between the sites, but the reproducibility was slightly lower. A residual configuration bias with 7.5% less signal measured at the C5 sites in the GM and WM VOIs was found in the final maps. This reduced the measured inter-site reproducibility between all sites. The inter-site CoVs of the C5 configuration data alone measured in the GM and WM VOIs was measured between 1.0% and 3.1% compared to the CoVs found between all sites (4.4–5.7%).

The MT contrast was also calculated and showed high reproducibility between the sites (Inline Supplementary Figure 3, Inline Supplementary Table 6). A small configuration bias with about 7% higher signal measured at the C5 configuration was found. Therefore, the inter-site reproducibility was reduced from CoVs between 1.6% and 2.7% in the GM and WM VOIs of the C5- configuration data alone to 3.4–5.5% if calculated between all sites.

3.5. Relaxometry

Multi-parametric mapping with the ME-GRE method showed very good agreement between sites (Fig. 8: Relaxometry, Inline Supplemen-

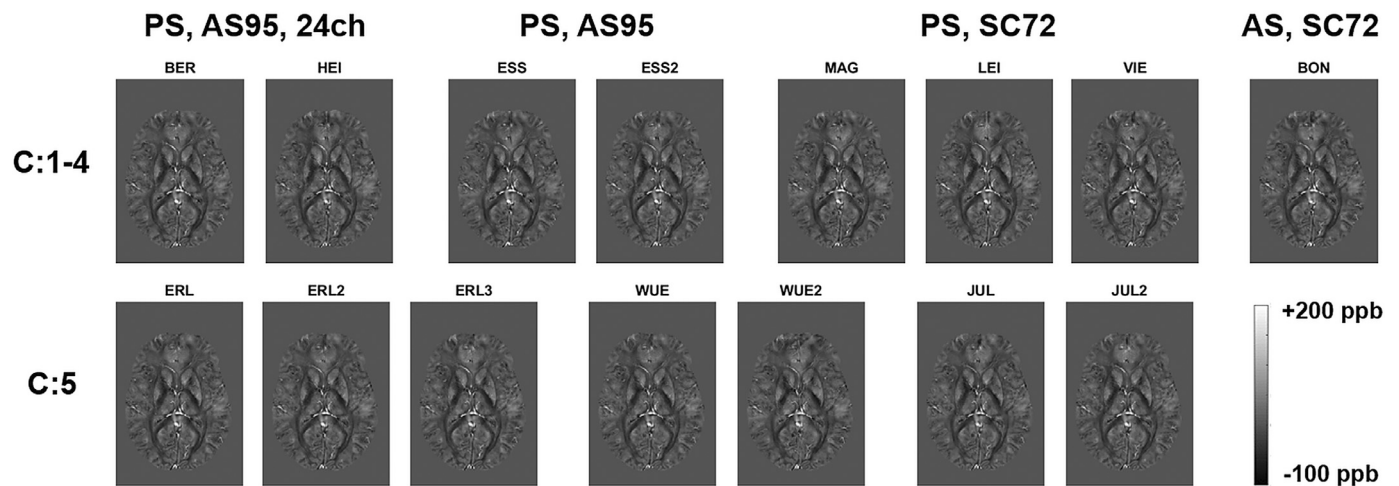


Fig. 6. QSM. QSM maps show very similar contrast, but small artifacts are visible, for example in median regions like the corpus callosum. Small differences may be introduced near cavities or vessels where the measured susceptibility is particularly sensitive to possible differences due to B0 shimming or subject positioning. The Subject 1 dataset is shown including all rescans aligned to the MP2RAGE images ([Inline Supplementary Figure 1](#)) via co-registration.

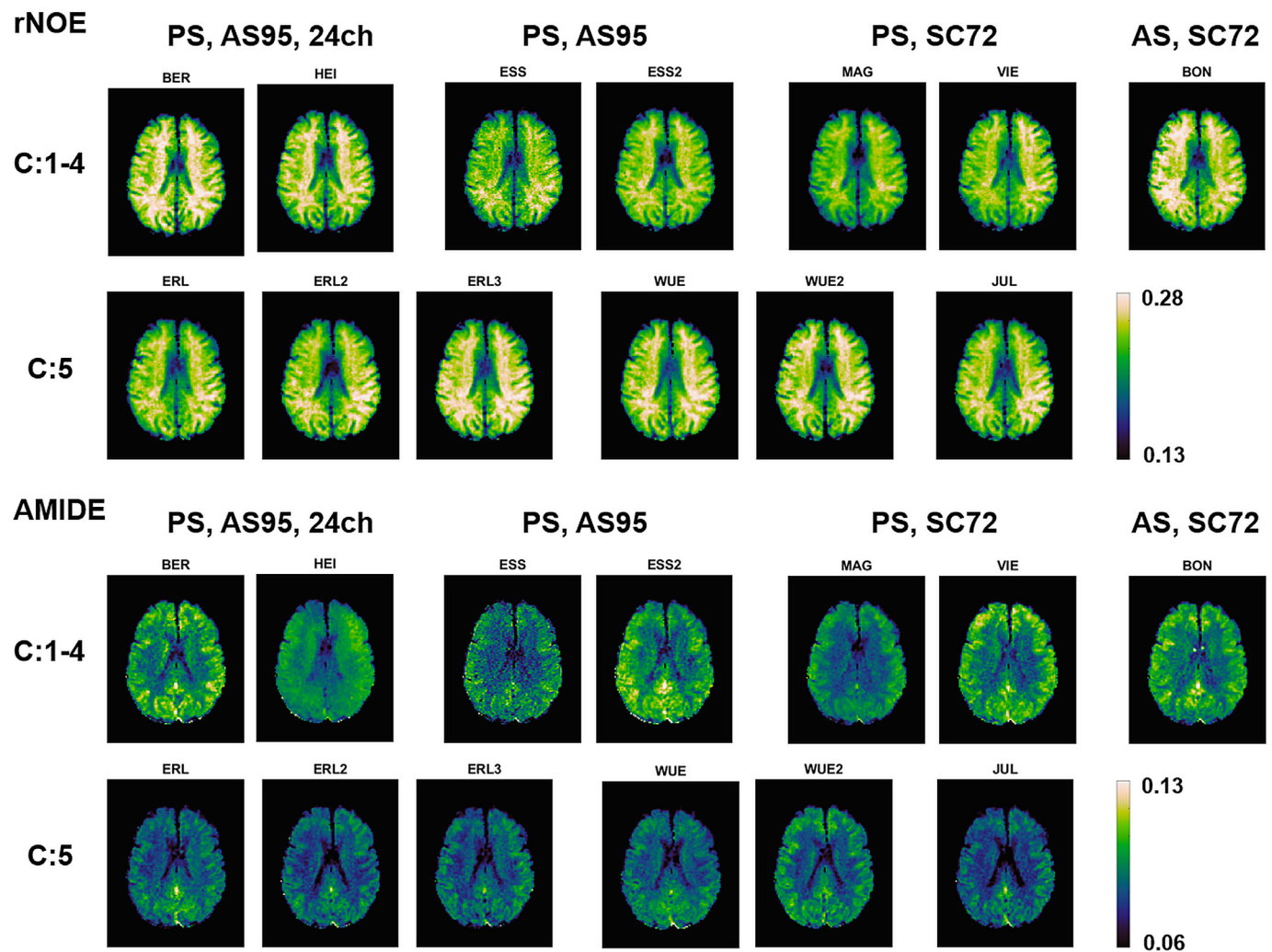


Fig. 7. CEST. CEST rNOE maps showing high agreement between the different sites. For the C5 sites (“Terra”), a small gradient with higher CEST contrast in the dorsal area is visible. Systematic hardware differences in the RFPA behavior under CEST duty cycle were calibrated, but small dynamic differences or modifications related to the software/sequence implementation might still exist and cause the observed contrast enhancements. The Subject 1 dataset is shown including all rescans in AC-PC orientation co-registered to the first scan at C2 (ESS).

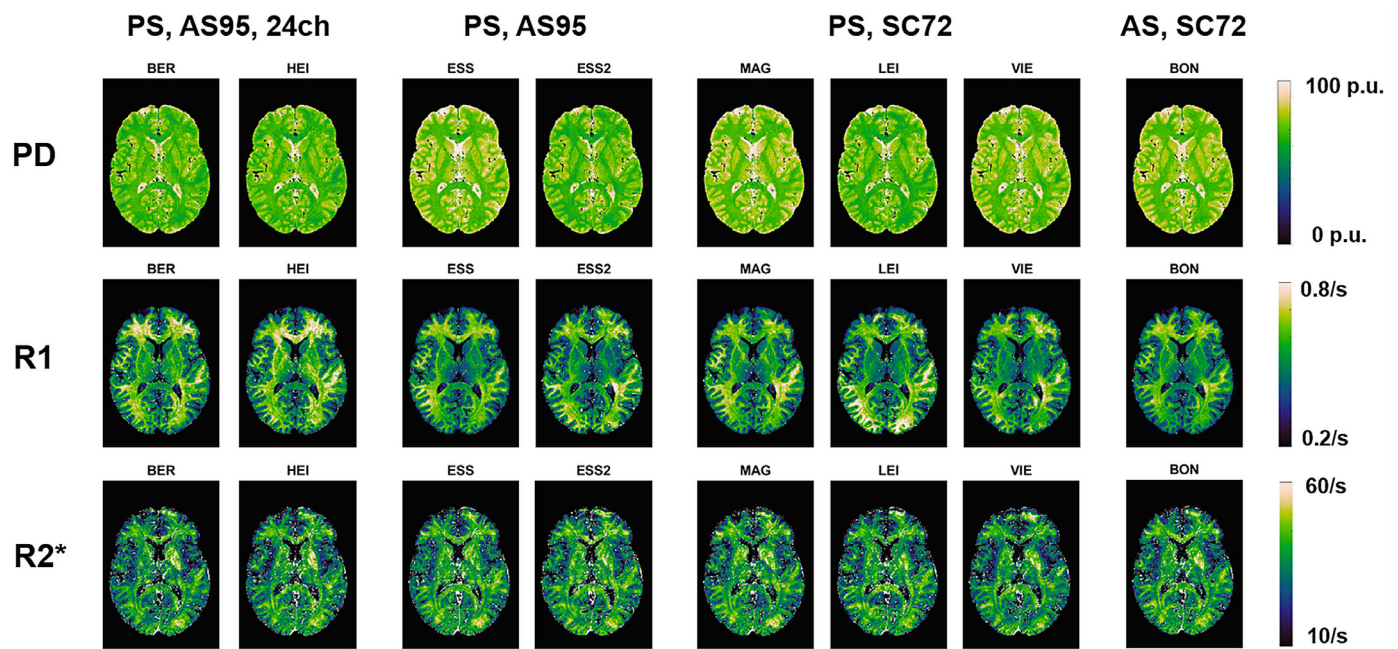


Fig. 8. Relaxometry. R1, R2* and PD maps calculated from ME-GRE images. The PD maps and central brain parts in the R1 maps at the sites with the 24-channel RF coil showed slightly more homogeneous values than at the C2-C4 sites with the 32-channel RF coil. Note that at the 24-channel coil sites the power calibration was limited to a fixed value and mean B1 was about 10% lower. At the C2-C4 sites the higher transmit power and the resultant higher flip angles of the B1 mapping used for corrections of the relaxometry data possibly could not be measured as accurately and this might explain these differences. The Subject 1 dataset is shown including all rescans demonstrating the data aligned to the MP2RAGE (Inline Supplementary Table 4) and the QSM (Fig. 6: QSM.) data excerpt via co-registration.

tary Table 7). The proton density maps provided the best inter-site reproducibility of all multi-parametric maps. CoV of the PD data of Subject 2 was as low as for the T1 measured with MP2RAGE, while it was slightly higher (1–2%) in Subject 1 for most VOIs (Inline Supplementary Table 7). The T1 and the T2* mapping with the ME-GRE approach both had higher inter-site variations than the PD data. For T1 the highest inter-site variations were found centrally, e.g. thalamus (left/right: Subject 1: 9/6%, Subject 2: 7/6%) and at inferior brain regions, with Subject 2 data being more reproducible (1–2%) than the VOIs compared in the Subject 1 dataset. For T2*, the highest variations were found in the nucleus accumbens (left/right: Subject 1: 11/12%, Subject 2: 8/6%) and amygdala (left/right: Subject 1: 11/10%, Subject 2: 8/9%), but reproducibility over all supratentorial VOIs was only slightly less compared to the T1 measurements (5.5/6.5% T1/T2* for Subject 1, 4.2/4.8% for Subject 2) (Fig. 9: Reproducibility). Comparing the inter-subject difference in the supratentorial VOIs, a mean difference of less than 5% was measured for all three parametric maps with a maximum of 13% for single VOIs. However, analyzing the differences between the different site configurations, deviations between the C1 configuration with the 24-channel RF coil and the C2-C4 configuration, where the 32-channel RF coil was used, could be identified. For the PD maps, the C1 values were slightly lower with a mean of 7% in the SB volumes of Subject 1 and 2% for Subject 2. For the T1 maps, the differences between the configurations showed a clear gradient toward the center of the brain, where VOIs of the C2-C4 sites had up to 17.5% (left thalamus) higher values (mean of SB VOIs for both subjects: 8%). For T2*, the C1 configuration measured slightly higher values than the C2-C4 configuration (mean of SB VOIs: Subject 1: 5%, Subject 2: 3%). The main difference between measurements taken at the C1 sites compared to the other sites was the lower input power used due to harder SAR restrictions for the 24-channel RF coil, which led to power limitations of the measurement protocol.

Comparing the T1 values between the ME-GRE method and the MP2RAGE approach the average T1 relaxation was measured about 35% higher with the ME-GRE method. The supratentorial GM showed 33%

higher T1 while the respective WM VOI had 41% higher mean T1 for both subjects with consistent relation between subjects (5% mean difference).

4. Discussion

Quantitative imaging is challenging at UHF MRI as these methods are very sensitive but not yet standardized. In this study we successfully measured a comprehensive state-of-the-art quantitative neuroimaging protocol in a multi-center UHF experiment with many scans and only few failures. We extracted quantitative data and compared the reproducibility systematically between the different MR systems, between subjects, and between individual brain regions.

4.1. System calibration

Individual UHF systems, even from the same vendor, have differences in hardware and software that might potentially influence quantification of image data. The system calibration is central to measuring or correcting those differences and to achieve reproducible results in multi-site experiments. At the C5 sites, the vendor recommends automatic adjustments for B1 and B0 based on the examined body region and RF coil setup. As those methods were not available at the older C1-C4 sites, we used methods established at the C1-C4 sites at all sites including C5 to avoid bias in the multi-site reproducibility due to different calibration methods.

B1 mapping is very important to calibrate the measurements, especially at UHF, where the shorter wavelength leads to less homogeneous transmit fields (Vaughan et al., 2001). We used the DREAM B1 mapping technique to calibrate the necessary input power and map the adjusted flip angle. This method is especially beneficial because of the fast acquisition that can be achieved (Ehse et al., 2019; Nehrke et al., 2015), but there is currently no gold standard for B1 mapping at UHF. To calibrate the transmit amplitude, we averaged circular ROIs in three consecutive axial slices located centrally at the slices where the lateral ventricles and the 3rd ventricle are visible, since these are the structures that one

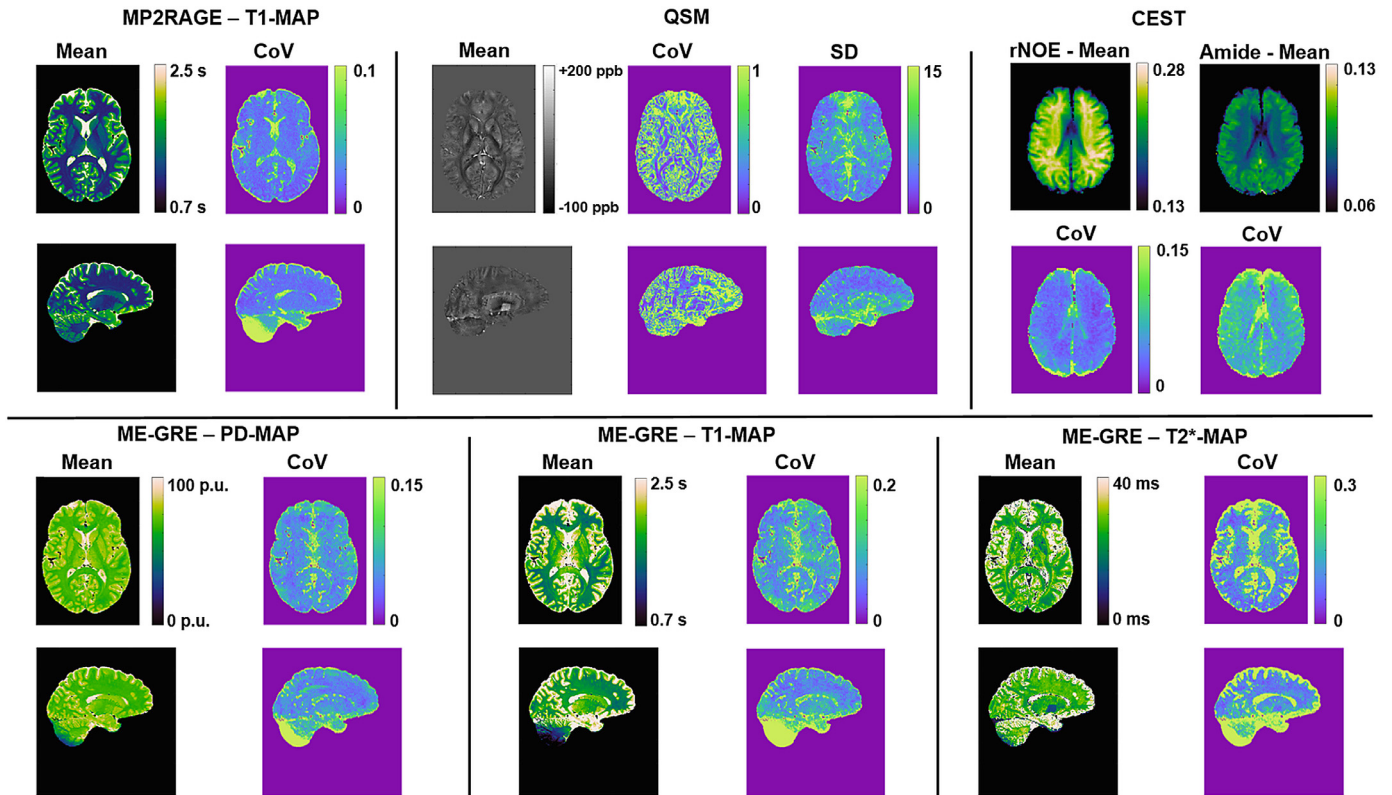


Fig. 9. Reproducibility. Reproducibility of quantitative imaging methods. The mean and the coefficient of variation (CoV) for all examined imaging methods are shown for the Subject 1 dataset. Note that the scale of the CoV maps has been adjusted to provide similar image contrast between methods. The T1 mapping with the MP2RAGE approach proved highly reproducible between different sites. The reproducibility drops for inferior regions like the cerebellum, where B1 is low, especially for the C1 sites with the 24-channel coils (Fig. 2: B1). For the QSM maps, the standard deviation between sites is also shown to visualize more difficult regions independent of the normalization to the ppb scale. QSM had higher deviations near cavities (e.g. frontal hot spot near paranasal and frontal sinuses) and CSF-tissue boundaries. CEST rNOE had good agreement between sites and no hot spots showed up in the examined slab except for tissue/brain boundaries. The relaxometry with the ME-GRE approach showed lower reproducibility in regions with low B1, e.g. cerebellum and brain stem analogous to the MP2RAGE, but also a gradient between the center and peripheral regions was found especially for the T1 maps. The higher central variability was driven by the differences between the measurements with different RF coils where input power was limited/different.

can identify most reliably at the coarse resolution of the sequence. Our results show that this approach assessed the mean necessary power for the SB volume reliably between subjects and sites. The peaks of the B1 histograms in the WB were adjusted within small error margins of about 5% for all sites that were not limited by the maximum RF power allowable to conform to SAR constraints for all measurements (Fig. 3: Histogram). As the protocols were optimized at the 32-channel RF coil sites, the maximum input power (transmitter reference voltage = 350 V) was determined prior to the study with an extra margin (15%) to meet SAR constraints for all measurements in the protocol without having to change imaging parameters. However, the 24-channel coil type at the C1 sites had lower transmitter efficiency but used more restrictive SAR limits as the 32-channel RF coil type and, therefore, the necessary input power to reach a comparable flip angle in the adjustment volume exceeded the maximum applicable input power to stay within the SAR limit. We used the fixed maximum applicable power at these sites instead, and consequently B1 after adjustment was about 15% lower than at the C2-C4 sites, which also narrowed the distributions of relative B1 and shifted them toward lower B1 values. Of course, this could have been avoided if the protocol had been adjusted at a C1 site or if we had used a larger safety margin as others did in a similar approach (Clarke et al., 2020). Nevertheless, the different power adjustments at the C1 sites gave us additional information on the B1 accuracy for the correction of the relaxometry maps that will be discussed later.

The B0 field was shimmed with the vendor-provided iterative shimming procedure in a rectangular VOI enclosing the cerebrum. The B0

mapping after adjustment clearly shows different field patterns between the C1-C4 and the C5 sites (Fig. 4: B0) due to different applied orders of correction terms for B0 shimming. Looking at the B0 histograms of the WB (Fig. 3: Histogram) or SB volumes and measuring the FWHM of the peaks, however, we could also find that there were differences between the older generation configurations, detectable in both subjects, favoring the C1-C2 sites with the less powerful gradient coil. As the shim coils are integrated into the gradient coil, there might be hardware differences that caused this behavior. As we applied the iterative shimming procedure in the same way at all sites, it might also be possible that the setup at the C3-C4 sites would need more iterations to converge to similar B0 homogeneity.

4.2. MP2RAGE

The MP2RAGE derived image data had the highest reproducibility among all quantitative methods in our study (Fig. 9: Reproducibility) and is widely used at 7 T as an anatomical T1-weighted imaging method for segmentation and quantitative analysis. Compared to our previous traveling heads experiment of the same subjects (Voelker et al., 2016), important parameters that affect the image quality and possibly affect reproducibility were changed. The inversion pulse of the sequence (HS4 (Tannús and Garwood, 1997) instead of TR-FOCI (Hurley et al., 2009, O'Brien et al., 2014)), the power calibration (about 40% higher transmit power) and the image processing pipeline were different compared to the initial study (Voelker et al., 2016), but our results are still highly

comparable. Differences occurred, for example, due to the additional B1 correction of the MP2RAGE data with the DREAM B1 maps, which we did not perform in the first study. Therefore, the T1 mapping results are more consistent over the whole brain, and measured T1 values for the subcortical VOIs are slightly lower compared to our previous results similar to Haast et al. (Haast et al., 2018). The inter-site reproducibility of the size of the subcortical volumes could be increased by a factor of 1.5 comparing the larger ($>2\text{ cm}^3$) supratentorial VOIs to our previous analysis. This can be explained by the different application of the FAST segmentation algorithm in MNI space instead of in the original image space and confirms the findings of Feng et al. (2017), who found increased accuracy by a similar method. The different application of the segmentation algorithm did also slightly influence the measured volume of some subcortical VOIs between both studies, e.g. the volume of the left and right putamen, that was segmented 5.4 cm^3 and 5.8 cm^3 for subject 1 and 2 in the old study was measured 5.9 cm^3 and 6.1 cm^3 in this study respectively. The inter-site reproducibility of the corresponding T1 measurements, however, did not increase in this study. Among the already described differences of the study setups and the new C5 site configuration that was not available in the previous study, the additional B1 correction may also introduce additional measurement errors. The measurement errors are additional to the MP2RAGE inaccuracies as both methods are combined. While this combination should increase the accuracy of the measurements, the reproducibility may be reduced and, therefore, did not increase similar to the volume size reproducibility. The measured GM and WM volumes as well as the corresponding T1 relaxation times were lower than in the previous study. A reason for this lies in changes made to the image post processing, e.g. the brain extraction masks that exclude arteries and dura, the additional GDC correction of the data, which requires a different partial volume filtering, and the tissue segmentation based on the B1-corrected images instead of the uncorrected image data. While these changes did alter the extracted volumes, the reproducibility of the measurements stayed at similarly high levels as in the previous work. The reproducibility of the method can be compared to the results of Okubo et al. (2016) at 3 T. They compared the reproducibility of MP2RAGE at 3 T using SPM (Wellcome Trust Centre for Neuroimaging, London, UK) based image processing. They measured a CoV between 2.08% and 2.89% for the T1 values in deep gray matter subcortical VOIs of the putamen, globus pallidus, thalamus and others between healthy subjects at the same site. A comparison to analyses with other image processing at 7 T can be made with the data of Haast et al. (2016), who used similar preprocessing and did a surface-based segmentation and parcellation using Freesurfer (<http://surfer.nmr.mgh.harvard.edu/>), leading to intra-subject CoVs of 1.63% in GM regions of MP2RAGE based T1 maps. In this context, our results prove to be very reproducible, which strengthens our findings that the individual system configuration in the multi-site experiment had an influence on the reproducibility of our 7 T MP2RAGE data. The more homogenous C5 site configuration yielded higher inter-site reproducibility than the more heterogeneous C1-C4 sites. This shows that differences in major imaging components (e.g. the magnet design, the gradient coil or the RF coil) reduce the inter-site reproducibility compared to a multi-site experiment where all imaging components of the systems are identical.

Deviations in image distortion were identified between the different site configurations via co-registration and overlaying the registered images (Fig. 5: Co-registration performance). The image data of the sites with gradient coils of lower strength ("AS95", C1-C2) could be co-registered with rigid transformation (translation + rotation). The sites with gradient coils of higher strength ("SC72", C3-C5) needed the additional affine co-registration with the higher degrees of freedom (rigid + scaling and shearing) to overlap sufficiently to the other sites. The linearity of the "SC72" gradient coil is lower compared to the other gradient coil model. This may result in image distortion which could only be partially corrected by the GDC provided by the vendor. In this experiment we applied the GDC offline with the vendor provided correc-

tion terms and verified this at some sites using both online and offline corrections. The first measurement at the C2 site was chosen as reference for the co-registration as it comprises the higher gradient linearity as well as the 32ch RF coil setup. A crosscheck with the second measurement at the C5 site ERL as co-registration reference had similar results (Inline Supplementary Figure 2) for both subjects. Image distortion will generally be moderated by B0 homogeneity (Wang et al., 2004) and differences may therefore occur due to B0 shimming (e.g. 2nd or 3rd order shimming). The position of the head may also be important for GDC as the distortion and the error in the correction increase with the distance from isocenter, which may amplify differences between individual sites or rescans. For more detailed analysis future studies are necessary with more advanced measures of voxel or surface based morphometry.

Apart from the observed image distortion differences between the configurations the volumetric measurements of the segmentations at the MP2RAGE data did not show systematic differences between individual site configurations. One major reason is that the volume quantification does not reflect differences of that shape of an object and therefore is not an ideal method to detect distortion differences. A better measure would be to extract the actual shape of the VOIs or the overlap after co-registration. Additionally, most of the segmented subcortical volumes that were used for quantification lay near the isocenter of the magnet, rendering image distortion evoked by gradient nonlinearity minimal.

4.3. QSM

To reduce the artifacts of the QSM reconstruction, it was very important to exclude inferior regions near cavities that cancelled the signal in the later echoes of the ME-GRE sequence. Therefore, the MP2RAGE derived brain masks were used and combined with HD-BET derived masks for each individual echo. This method reduced the artifacts while also allowing us to normalize the susceptibility of each subject to the whole brain. Since normalization of the data is not straightforward in QSM (Straub et al., 2017, Feng et al., 2018) and may introduce additional reproducibility error, we did not normalize the data further to a reference volume such as the frontal part of the lateral ventricles. Thus, the measured susceptibility values were different between the two subjects. In fact, the susceptibility between the subjects scaled with a factor (1.25) instead of an offset that might be corrected with inter-subject normalization. Consequently, physiological differences between the two subjects might be the reason for the measured susceptibility differences. We also segmented different VOIs in the ventricles and in WM that can be used as a normalization reference and measured the reproducibility to evaluate their use as reference VOIs. These VOIs proved to be very reproducible for inter-site comparisons of the same subjects, and CoV were similar to the VOIs that we used to analyze the QSM maps. For the QSM reproducibility analysis, we used the VOIs segmented with the MP2RAGE, but also manually segmented subcortical GM VOIs like the substantia nigra, red nucleus and dentate nuclei to be comparable to literature values. Both segmentation methods provided similarly high reproducibility for inter-site data. Compared to other groups and methods (Santin et al., 2017, Deh et al., 2015, Feng et al., 2018, Lin et al., 2015, Spincemaille et al., 2019), our QSM analysis method at 7 T performed very well. For example, Santin et al. (2017) compared intra-site reproducibility measures of different reconstruction techniques at 3 T and measured with LBV-MEDI (Laplacian boundary value (Zhou et al., 2014), morphology-enabled dipole inversion (Liu et al., 2011)) intra-subject averaged CoVs between 2% and 4% for red nucleus, substantia nigra and globus pallidus. This is comparable to our results at 7 T, where we also found some exceptions with higher deviations that might be explained by the higher sensitivity at 7 T and consequently higher propensity for reconstruction artifacts. A recent study from the UK7T group (Rua et al., 2020) did an extensive multi-site analysis for QSM data at 7 T. They compared the reproducibility for multi-echo and single-echo QSM data with different co-registration, segmentation methods and referencing from 10 subjects at 5 sites equipped with MR systems from two

different vendors. From this data we extracted average CoV values, compared it to our reproducibility results and found very high agreement with their most reproducible method (single echo, manual ROI, SyN-registration). For example, substantia nigra (6.1%), red nucleus (8.8%), caudate nuclei (7.6%), putamen (5.2%) and globus pallidus (5.5%) compare very well to our average reproducibility values of both subjects (5.8%, 5.6%, 7.8%, 6.7%, 6.7%).

4.4. CEST

The CEST reconstruction performed reliably for the rNOE contrast. The rNOE maps had a very high reproducibility between subjects and sites, but we found a bias between the C5 configuration sites and the older C1-C4 installations. One reason for the observed differences was the thermal instability of the systems' RFPAs caused by the very long CEST pre-saturation pulse chain (140×15 -ms Gaussian shaped pulses were applied with a delay of 10 ms per RF block). The RFPAs were not capable of perfectly handling the long ON time (unblank) duty cycle, and our calibration measurements showed that the flip angles deviate proportionally to the ON time duty cycle (Supplementary data). The flip angle deviations between the individual sites were measured between 3% and 11% (mean = 6%). The individual age or type of the installed RFA might influence these variations. Hence, the calibration of our data with the measured deviation factors reduced the observed differences between the different scanner configurations by 8.1%, but there is still a bias visible in the rNOE maps, particularly as a small gradient with higher rNOE contrast in the dorsal areas of the C5 sites. For Amide and MT the bias between the configurations remained even higher after the corrections (~7%) with either smaller (Amide) or higher (MT) CEST signal measured at the C5 configuration compared to the other configurations. Reasons for these systematic differences need to be examined in further experiments. There might be dynamic inaccuracies of the RFPAs due to the thermal stress of the CEST experiment that cause the observed differences, but it could also be that the different sequence implementation of the readout for the C5 software version leads to T2* effects that may cause deviations between the system configurations. Reproducibility is currently a topic at the CEST community and besides our data there is few data available at 7T. Geades et al. (2017) did report high intra-site and inter-subject reproducibility at 7T, which our data does not comprise. Nevertheless, the inter-site reproducibility found in this experiment is very high and comparable to the results of Deshmane et al. (2019) where CoVs of 7.8% and 6.8% were reported in GM and WM between three subjects at 3 T.

4.5. Relaxometry

The relaxometry with the ME-GRE approach provided very reproducible inter-site results. The PD maps were nearly as reproducible as the T1 maps measured with the MP2RAGE method, with the limitation that the multi-parametric mapping could not be measured for all configurations (C5 was excluded due to problems with the sequence implementation). The T1 values measured with ME-GRE are slightly less reproducible but also differ systematically from the values of the MP2RAGE measurement due to signal contributions from the faster and slower relaxing components being weighted differently (Jutras et al., 2017; Rioux et al., 2016), and different magnetization transfer effects (Teixeira et al., 2019). Main characteristics of our results are similar to the multi-site data of Leutritz et al. (2020) or Weiskopf et al. at 3 T. Compared to Weiskopf et al. (2013) with a similar setup at 3 T, where they measured inter-site CoVs for GM, WM and caudate nuclei between 4.6% and 6% for R1, 2.7% and 3.6% for PD and 11.4% and 20.3% for R2* at three different sites with the same type of MRI scanner, our results showed higher reproducibility for most VOIs at 7 T. Even though we found that the different input power used at the 24-channel RF coil site configuration (C1) systematically reduced the inter-site reproducibility between all sites for our measurements, especially for the R1 maps, a

gradient in the R1 maps correlating to the B1 map profile was found for the C2-C4 configurations and a flatter profile was measured at the C1 sites. As the correct R1 calculation is proportional to the square of the measured B1 (Helms et al., 2008), the observed differences point toward the accuracy of transmit field correction with the DREAM B1 mapping method. DREAM is a relatively robust technique with an optimum flip angle of about 55° and a relatively high dynamic range between 10° and 70° with good accuracy (Nehrke and Börnert, 2012). However, it is known that for higher flip angles ($> 80^\circ$), DREAM tends to underestimate the true flip angle (Nehrke and Börnert, 2012). This might be the key to the apparent R1 differences we found here. The portion of the B1 values higher than $1.2 \times$ nominal flip measured in the cranial brain (Fig. 3: Histogram) was much higher for the C2-C4 sites than at the C1 configuration. Therefore, reducing B1 or measuring B1 with lower nominal flip angle would have been beneficial for the accuracy of the maps and should reduce the overestimation of central T1 similar to the C1 configuration data. A better accuracy and higher dynamic range of the DREAM B1 mapping could be accomplished by using multiple nominal flip angles (Olsson et al., 2020) or alternative B1 mapping methods may be used (Pohmann and Scheffler, 2013). The PD and as expected the T2* maps were less influenced by this issue, and the inter-site reproducibility was similar between central VOIs and peripheral areas. PD map estimation can be biased by T2* effects. In our study, we applied extrapolation to TE = 0 ms to account for this, although it assumes that transversal relaxation is monoexponential. In single pool relaxometry it is known that inadvertent magnetization transfer effects lead to sources of variation in R1 and PD maps, introducing a dependence on the RF power level (Teixeira et al., 2019). Although we attempted to control transmitter calibration, input RF power during the ME-GRE scans was not perfectly consistent between sites and may further explain inter-site variation.

5. Conclusion

Our data show that standardized state-of-the-art quantitative imaging methods lead to a very high reproducibility for inter-site comparisons in the same subjects between different UHF sites and MR system architectures, even across 15 years of operation. However, some variability was observed, especially when different hardware platforms/software versions or different RF coils were used, or the imaging protocols had to be adapted to account for these differences. Therefore, special care has to be taken in multicenter studies if the detectable effect is on the order of the variation we found in this study. Nevertheless, the multicenter reproducibility at 7 T for supratentorial quantitative brain imaging was found to be similar to literature data for 3 T, which is a promising finding for UHF MRI. The data we acquired may be used for further algorithm and image processing testing on reproducibility and can be accessed online (DOI: [dx.doi.org/10.5281/zenodo.4117947](https://doi.org/10.5281/zenodo.4117947)).

Data availability

The image data published in this study is available the public repository zenodo (DOI: [dx.doi.org/10.5281/zenodo.4117947](https://doi.org/10.5281/zenodo.4117947)) and can be accessed by any researcher.

Compliance with ethical standards

Funding

This work was supported by a grant from the German Research Foundation (DFG): project title German Ultrahigh Field Imaging, project number 324669110, grant numbers LA 1325/7-1, QU 154/5-1, SP 632/9-1.

Funding by the Deutsche Forschungsgemeinschaft (DFG) for the Heisenberg professorship of F. B. Laun is gratefully acknowledged (LA

2804/12-1). Funding by the German Ministry of Education and Research for M.R. Stefanescu and M. Hock is gratefully acknowledged (Grant/Award No. 01EO1504).

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Most of the institutes or sites involved in this study have or had and institutional research agreement with Siemens Healthcare. UHF adapted imaging sequences were provided by Siemens Healthcare GmbH.

NW was a speaker at an event organized by Siemens Healthcare and was reimbursed for the travel expenses.

Credit authorship contribution statement

Maximilian N. Voelker: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Oliver Kraff:** Conceptualization, Methodology, Validation, Investigation, Writing - review & editing, Visualization, Supervision. **Steffen Goerke:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing - review & editing. **Frederik B. Laun:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing - review & editing. **Jannis Hanspach:** Methodology, Validation, Formal analysis, Data curation, Writing - review & editing. **Kerrin J. Pine:** Conceptualization, Methodology, Investigation, Validation, Formal analysis, Data curation, Writing - review & editing. **Philipp Ehse:** Conceptualization, Methodology, Investigation, Validation, Writing - review & editing. **Moritz Zaiss:** Conceptualization, Methodology, Validation, Writing - review & editing. **Andrzej Liebert:** Investigation, Validation, Data curation. **Sina Straub:** Conceptualization, Methodology, Investigation, Validation, Data curation. **Korbinian Eckstein:** Methodology, Software. **Simon Robinson:** Conceptualization, Methodology, Investigation. **Armin N. Nagel:** Conceptualization, Methodology, Investigation, Writing - review & editing. **Maria R. Stefanescu:** Investigation, Validation, Data curation. **Astrid Wollrab:** Investigation, Validation, Data curation. **Sabrina Klitz:** Investigation, Validation, Data curation. **Jörg Felder:** Investigation, Validation, Data curation. **Michael Hock:** Investigation, Validation, Data curation. **Dario Bosch:** Formal analysis, Methodology, Visualization. **Nikolaus Weiskopf:** Conceptualization, Writing - review & editing. **Oliver Speck:** Project administration, Funding acquisition, Conceptualization, Writing - review & editing. **Mark E. Ladd:** Project administration, Funding acquisition, Conceptualization, Writing - review & editing. **Harald H. Quick:** Project administration, Funding acquisition, Conceptualization, Writing - review & editing, Supervision.

Ethical standards

The study was conducted according to the principles expressed in the Declaration of Helsinki. The protocol was approved by the local university institutional review board.

Informed consent

Informed consent was obtained from all participants included in the study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.neuroimage.2021.117910](https://doi.org/10.1016/j.neuroimage.2021.117910).

References

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., Feb. 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54 (3), 2033–2044.

Bazin, P.-L., Weiss, M., Dinse, J., Schäfer, A., Trampel, R., Turner, R., Jun. 2014. A computational framework for ultra-high resolution cortical segmentation at 7T. *Neuroimage* 93, 201–209.

Breitling, J., et al., Nov. 2019. Adaptive denoising for chemical exchange saturation transfer MR imaging. *NMR Biomed.* 32 (11).

Breuer, F.A., Blaimer, M., Heidemann, R.M., Mueller, M.F., Griswold, M.A., Jakob, P.M., Mar. 2005. Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging. *Magn. Reson. Med.* 53 (3), 684–691.

Chen, Y., Liu, S., Wang, Y., Kang, Y., Haacke, E.M., 2018. STrategically Acquired Gradient Echo (STAGE) imaging, part I: creating enhanced T1 contrast and standardized susceptibility weighted imaging and quantitative susceptibility mapping. *Magn. Reson. Imaging* 46, 130–139.

Clarke, W.T., et al., Feb. 2020. Multi-site harmonization of 7 tesla MRI neuroimaging protocols. *Neuroimage* 206, 116335.

Deh, K., et al., 2015. Reproducibility of quantitative susceptibility mapping in the brain at two field strengths from two vendors. *J. Magn. Reson. Imaging* 42 (6), 1592–1600.

Deistung, A., Schäfer, A., Schweser, F., Biedermann, U., Turner, R., Reichenbach, J.R., Jan. 2013. Toward in vivo histology: a comparison of quantitative susceptibility mapping (QSM) with magnitude-, phase-, and R2-imaging at ultra-high magnetic field strength. *Neuroimage* 65, 299–314.

Deshmane, A., et al., 2019. 3D gradient echo snapshot CEST MRI with low power saturation for human studies at 3T. *Magn. Reson. Med.* 81 (4), 2412–2423.

Eckstein, K., et al., Jun. 2018. Computationally efficient combination of multi-channel phase data from multi-echo acquisitions (ASPIRE). *Magn. Reson. Med.* 79 (6), 2996–3006.

Ehse, P., Brenner, D., Stirnberg, R., Pracht, E.D., Stöcker, T., Apr. 2019. Whole-brain B1-mapping using three-dimensional DREAM. *Magn. Reson. Med.* mrm.27773.

Ernst, R.R., Anderson, W.A., Jan. 1966. Application of fourier transform spectroscopy to magnetic resonance. *Rev. Sci. Instrum.* 37 (1), 93–102.

“FDA clears first 7T magnetic resonance imaging device,” *U.S. Food and Drug Administration*, 2017. [Online]. Available: <https://www.fda.gov/news-events/press-announcements/fda-clears-first-7t-magnetic-resonance-imaging-device>.

Feng, X., et al., Jun. 2017. An improved FSL-FIRST pipeline for subcortical gray matter segmentation to study abnormal brain anatomy using quantitative susceptibility mapping (QSM). *Magn. Reson. Imaging* 39, 110–122.

Feng, X., Deistung, A., Reichenbach, J.R., Feb. 2018. Quantitative susceptibility mapping (QSM) and R2* in the human brain at 3 T. *Z. Med. Phys.* 28 (1), 36–48.

Feng, X., Deistung, A., Reichenbach, J.R., Feb. 2018. Quantitative susceptibility mapping (QSM) and R2* in the human brain at 3 T. *Z. Med. Phys.* 28 (1), 36–48.

Geades, N., Hunt, B.A.E., Shah, S.M., Peters, A., Mouglin, O.E., Gowland, P.A., Aug. 2017. Quantitative analysis of the z-spectrum using a numerically simulated look-up table: application to the healthy human brain at 7T. *Magn. Reson. Med.* 78 (2), 645–655.

Glasser, M.F., et al., Oct. 2013. The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124.

Griswold, M.A., et al., Jun. 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.* 47 (6), 1202–1210.

Haacke, E.M., Liu, S., Buch, S., Zheng, W., Wu, D., Ye, Y., 2015. Quantitative susceptibility mapping: current status and future directions. *Magn. Reson. Imaging* 33 (1), 1–25.

Haast, R.A.M., Ivanov, D., Formisano, E., Uludağ, K., 2016. Reproducibility and reliability of quantitative and weighted T1 and T2* mapping for myelin-based cortical parcellation at 7 tesla. *Front. Neuroanat.* 10 (NOV), 1–17.

Haast, R.A.M., Ivanov, D., Uludağ, K., 2018. The impact of B1+ correction on MP2RAGE cortical T1 and apparent cortical thickness at 7T. *Hum. Brain Mapp* 39 (6), 2412–2425.

Harkins, K.D., et al., Mar. 2016. The microstructural correlates of T1 in white matter. *Magn. Reson. Med.* 75 (3), 1341–1345.

Helms, G., Dathe, H., Dechent, P., Mar. 2008. Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magn. Reson. Med.* 59 (3), 667–672.

Helms, G., Dathe, H., Kallenberg, K., Dechent, P., Dec. 2008. High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. *Magn. Reson. Med.* 60 (6), 1396–1407.

Henkelman, R.M., Stanisz, G.J., Graham, S.J., Apr. 2001. Magnetization transfer in MRI: a review. *NMR Biomed.* 14 (2), 57–64.

Hurley, A.C., et al., 2009. Tailored RF pulse for magnetization inversion at ultrahigh field. *Magn. Reson. Med.* p. NA-NA.

Isensee, F., et al., Dec. 2019. Automated brain extraction of multisequence MRI using artificial neural networks. *Hum. Brain Mapp.* 40 (17), 4952–4964.

Jones, C.K., et al., Aug. 2013. Nuclear Overhauser enhancement (NOE) imaging in the human brain at 7T. *Neuroimage* 77, 114–124.

Jovicich, J., et al., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 30 (2), 436–443.

Jutras, J.D., Wachowicz, K., Gilbert, G., De Zanche, N., 2017. SNR efficiency of combined bipolar gradient echoes: comparison of three-dimensional FLASH, MP2RAGE, and multiparameter mapping with VFA-FLASH and MP2RAGE. *Magn. Reson. Med.* 77 (6), 2186–2202.

Katrin, K., 2020. 7 tricks for 7T CEST - improving reproducibility of multi-pool evaluation. CEST Conference talk session 5.

Kim, J., Wu, Y., Guo, Y., Zheng, H., Sun, P.Z., May 2015. A review of optimization and quantification techniques for chemical exchange saturation transfer MRI toward sensitive in vivo imaging. *Contrast Media Mol. Imaging* 10 (3), 163–178.

Kirilina, E., et al., Oct. 2020. Superficial white matter imaging: Contrast mechanisms and whole-brain in vivo mapping. *Sci. Adv.* 6 (41).

Ladd, M.E., et al., Dec. 2018. Pros and cons of ultra-high-field MRI/MRS for human application. *Prog. Nucl. Magn. Reson. Spectrosc.* 109, 1–50.

- Langkammer, C., et al., Sep. 2012. Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study. *Neuroimage* 62 (3), 1593–1599.
- Leutritz, T., et al., 2020. Multiparameter mapping of relaxation (R_1 , R_2^*), proton density and magnetization transfer saturation at 3 T: a multicenter dual-vendor reproducibility and repeatability study. *Hum. Brain Mapp.* April, 1–16.
- Li, W., Avram, A.V., Wu, B., Xiao, X., Liu, C., 2014. Integrated Laplacian-based phase unwrapping and background phase removal for quantitative susceptibility mapping. *NMR Biomed.* 27 (2), 219–227.
- Lin, P.Y., Chao, T.C., Wu, M.L., 2015. Quantitative susceptibility mapping of human brain at 3T: a multisite reproducibility study. *Am. J. Neuroradiol.* 36 (3), 467–474.
- Liu, T., et al., Sep. 2011. Morphology enabled dipole inversion (MEDI) from a single-angle acquisition: comparison with COSMOS in human brain imaging. *Magn. Reson. Med.* 66 (3), 777–783.
- Louapre, C., et al., Nov. 2015. Beyond focal cortical lesions in MS. *Neurology* 85 (19), 1702–1709.
- MacKay, A., Laule, C., Vavasour, I., Bjarnason, T., Kolind, S., Mädler, B., May 2006. Insights into brain microstructure from the T2 distribution. *Magn. Reson. Imaging* 24 (4), 515–525.
- Marques, J.P., Gruetter, R., Jul. 2013. New developments and applications of the MP2RAGE sequence - focusing the contrast and high spatial resolution R1 mapping. *PLoS One* 8 (7), e69294.
- Marques, J.P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., Gruetter, R., Jan. 2010. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *Neuroimage* 49 (2), 1271–1281.
- Michelson, A.A., 1927. *Studies in Optics*. University Press.
- Nehrke, K., Börnert, P., Nov. 2012. DREAM-a novel approach for robust, ultrafast, multi-slice B1 mapping. *Magn. Reson. Med.* 68 (5), 1517–1526.
- Nehrke, K., Sprinkart, A.M., Börnert, P., Apr. 2015. An in vivo comparison of the DREAM sequence with current RF shim technology. *Magn. Reson. Mater. Physics, Biol. Med.* 28 (2), 185–194.
- O'Brien, K.R., et al., Oct. 2014. Dielectric pads and low- B1+ adiabatic pulses: complementary techniques to optimize structural T1 w whole-brain MP2RAGE scans at 7 tesla. *J. Magn. Reson. Imaging* 40 (4), 804–812.
- Okubo, G., et al., 2016. MP2RAGE for deep gray matter measurement of the brain: a comparative study with MPRAGE. *J. Magn. Reson. Imaging* 43 (1), 55–62.
- Olsson, H., Andersen, M., Helms, G., Oct. 2020. Reducing bias in DREAM flip angle mapping in human brain at 7T by multiple preparation flip angles. *Magn. Reson. Imaging* 72, 71–77.
- Patenaude, B., Smith, S.M., Kennedy, D.N., Jenkinson, M., Jun. 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 56 (3), 907–922.
- Pohmann, R., Scheffler, K., Mar. 2013. A theoretical and experimental comparison of different techniques for B1 mapping at very high fields. *NMR Biomed.* 26 (3), 265–275.
- Pohmann, R., Speck, O., Scheffler, K., Feb. 2016. Signal-to-noise ratio and MR tissue parameters in human brain imaging at 3, 7, and 9.4 tesla using current receive coil arrays. *Magn. Reson. Med.* 75 (2), 801–809.
- Preibisch, C., Deichmann, R., Jan. 2009. 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.
- Rioux, J.A., Levesque, I.R., Rutt, B.K., 2016. Biexponential longitudinal relaxation in white matter: characterization and impact on T1 mapping with IR-FSE and MP2RAGE. *Magn. Reson. Med.* 75 (6), 2265–2277.
- Rua, C., et al., 2020. Multi-centre, multi-vendor reproducibility of 7T QSM and R_2^* in the human brain: results from the UK7T study. *Neuroimage* 223, 117358 April.
- Santin, M.D., et al., 2017. Reproducibility of R_2^* and quantitative susceptibility mapping (QSM) reconstruction methods in the basal ganglia of healthy subjects. *NMR Biomed.* 30 (4).
- Schofield, M.A., Zhu, Y., Jul. 2003. Fast phase unwrapping algorithm for interferometric applications. *Opt. Lett.* 28 (14), 1194–1196.
- Schuenke, P., Windschuh, J., Roeloffs, V., Ladd, M.E., Bachert, P., Zaiss, M., Feb. 2017. Simultaneous mapping of water shift and B1 (WASABI)-Application to field-Inhomogeneity correction of CESTMRI data. *Magn. Reson. Med.* 77 (2), 571–580.
- Smith, S.M., Nov. 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17 (3), 143–155.
- Spincemaille, P., et al., 2019. Clinical integration of automated processing for brain quantitative susceptibility mapping: multi-site reproducibility and single-site robustness. *J. Neuroimaging* 29 (6), 689–698.
- Spincemaille, P., et al., Jan. 2020. Quantitative susceptibility mapping: MRI at 7T versus 3T. *J. Neuroimaging* 30 (1), 65–75.
- Straub, S., et al., 2017. Suitable reference tissues for quantitative susceptibility mapping of the brain. *Magn. Reson. Med.* 78 (1), 204–214.
- Straub, S., et al., Nov. 2020. Toward quantitative neuroimaging biomarkers for Friedrich's ataxia at 7 Tesla: susceptibility mapping, diffusion imaging, R_2 and R_1 relaxometry. *J. Neurosci. Res.* 98 (11), 2219–2231.
- Tabelow, K., et al., December 2018. hMRI – A toolbox for quantitative MRI in neuroscience and clinical research. *Neuroimage* 194, 191–210 2019.
- Tannús, A., Garwood, M., Dec. 1997. Adiabatic pulses. *NMR Biomed.* 10 (8), 423–434.
- Teixeira, R.P.A.G., Malik, S.J., Hajnal, J.V., Feb. 2019. Fast quantitative MRI using controlled saturation magnetization transfer. *Magn. Reson. Med.* 81 (2), 907–920.
- Tustison, N.J., et al., Jun. 2010. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 29 (6), 1310–1320.
- Uğurbil, K., Mar. 2018. Imaging at ultrahigh magnetic fields: history, challenges, and solutions. *Neuroimage* 168, 7–32.
- Vaughan, J.T., et al., Jul. 2001. 7T vs. 4T: RF power, homogeneity, and signal-to-noise comparison in head images. *Magn. Reson. Med.* 46 (1), 24–30.
- Voelker, M.N., et al., 2016. The traveling heads: multicenter brain imaging at 7 Tesla. *Magn. Reson. Mater. Phys., Biol. Med.* 29 (3).
- Wang, D., Strugnell, W., Cowin, G., Doddrell, D.M., Slaughter, R., Nov. 2004. Geometric distortion in clinical MRI systems. *Magn. Reson. Imaging* 22 (9), 1211–1221.
- Ward, K., Aletras, A., Balaban, R., Mar. 2000. A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). *J. Magn. Reson.* 143 (1), 79–87.
- Wei, H., et al., Oct. 2015. Streaking artifact reduction for quantitative susceptibility mapping of sources with large dynamic range. *NMR Biomed.* 28 (10), 1294–1303.
- Weiskopf, N., et al., 2013. Quantitative multi-parameter mapping of R_1 , PD^* , MT, and R_2^* at 3T: a multi-center validation. *Front. Neurosci.* 7, 95.
- Weiskopf, N., Mohammadi, S., Lutti, A., Callaghan, M.F., Aug. 2015. Advances in MRI-based computational neuroanatomy. *Curr. Opin. Neurol.* 28 (4), 313–322.
- Windschuh, J., et al., May 2015. Correction of B1 inhomogeneities for relaxation-compensated CEST imaging at 7 T. *NMR Biomed.* 28 (5), 529–537.
- Wu, B., Li, W., Guidon, A., Liu, C., 2012. Whole brain susceptibility mapping using compressed sensing. *Magn. Reson. Med.* 67 (1), 137–147.
- Yushkevich, P.A., et al., Jul. 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31 (3), 1116–1128.
- Zaiss, M., et al., Mar. 2014. Inverse Z-spectrum analysis for spillover-, MT-, and T1 -corrected steady-state pulsed CEST-MRI-application to pH-weighted MRI of acute stroke. *NMR Biomed.* 27 (3), 240–252.
- Zaiss, M., et al., May 2015. Relaxation-compensated CEST-MRI of the human brain at 7T: unbiased insight into NOE and amide signal changes in human glioblastoma. *Neuroimage* 112, 180–188.
- Zaiss, M., Ehse, P., Scheffler, K., 2018. Snapshot-CEST: optimizing spiral-centric-ordered gradient echo acquisition for fast and robust 3D CEST MRI at 9.4 T. *NMR Biomed.* 31 (4), e3879.
- 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.
- Zhou, D., Liu, T., Spincemaille, P., Wang, Y., Mar. 2014. Background field removal by solving the Laplacian boundary value problem. *NMR Biomed.* 27 (3), 312–319.