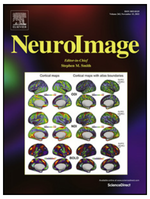
NeuroImage 262 (2022) 119529

Contents lists available at ScienceDirect

NeuroImage

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



Error quantification in multi-parameter mapping facilitates robust estimation and enhanced group level sensitivity

Siawoosh Mohammadi[a,b,*], Tobias Streubel[a,b], Leonie Klock[c], Luke J. Edwards[b], Antoine Lutti[d], Kerrin J. Pine[b], Sandra Weber[c], Patrick Scheibe[b], Gabriel Ziegler[e,f], Jürgen Gallinat[c], Simone Kühn[c,g], Martina F. Callaghan[h], Nikolaus Weiskopf[b,i], Karsten Tabelow[j]

[a] *Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[b] *Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[c] *Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[d] *Laboratory for Research in Neuroimaging, Department of Clinical Neurosciences, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland*
[e] *Institute of Cognitive Neurology and Dementia Research, Otto-von-Guericke-University Magdeburg, Germany*
[f] *Germany Weierstrass Institute for Applied Analysis and Stochastics, German Center for Neurodegenerative Diseases, Magdeburg, Berlin, Germany*
[g] *Max Planck Institute for Human Development, Lise Meitner Group for Environmental Neuroscience, Berlin, Germany*
[h] *Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, UCL, London, UK*
[i] *Felix Bloch Institute for Solid State Physics, Leipzig University, Linnéstraße 5, Leipzig 04103, Germany*
[j] *Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, Berlin 10117, Germany*

ARTICLE INFO

Keywords:
Multi-parameter mapping
Quantitative MRI
Error propagation
Signal-to-noise ratio
Robust estimate

ABSTRACT

Multi-Parameter Mapping (MPM) is a comprehensive quantitative neuroimaging protocol that enables estimation of four physical parameters (longitudinal and effective transverse relaxation rates R_1 and R_2^*, proton density PD, and magnetization transfer saturation MT_{sat}) that are sensitive to microstructural tissue properties such as iron and myelin content. Their capability to reveal microstructural brain differences, however, is tightly bound to controlling random noise and artefacts (e.g. caused by head motion) in the signal. Here, we introduced a method to estimate the local error of PD, R_1, and MT_{sat} maps that captures both noise and artefacts on a routine basis without requiring additional data. To investigate the method's sensitivity to random noise, we calculated the model-based signal-to-noise ratio (mSNR) and showed in measurements and simulations that it correlated linearly with an experimental raw-image-based SNR map. We found that the mSNR varied with MPM protocols, magnetic field strength (3T vs. 7T) and MPM parameters: it halved from PD to R_1 and decreased from PD to MT_{sat} by a factor of 3-4. Exploring the artefact-sensitivity of the error maps, we generated robust MPM parameters using two successive acquisitions of each contrast and the acquisition-specific errors to down-weight erroneous regions. The resulting robust MPM parameters showed reduced variability at the group level as compared to their single-repeat or averaged counterparts. The error and mSNR maps may better inform power-calculations by accounting for local data quality variations across measurements. Code to compute the mSNR maps and robustly combined MPM maps is available in the open-source hMRI toolbox.

1. Introduction

Quantitative magnetic resonance imaging (qMRI) can be more reproducible than conventional MRI (e.g. T1-weighted MRI typically used for morphometric analysis Paus et al., 1999), especially between sites, as it aims to remove the impact of hardware variations on measured maps (Weiskopf et al., 2013; Cercignani and Bouyagoub, 2018). Quantification is typically achieved by acquiring multiple MRI contrasts to disentangle the mixture of different physical MR parameters present in conventional MRI, and correcting for instrumental variation through the acquisition of additional calibration measurements (Weiskopf et al., 2021). Multi-parameter mapping (MPM) provides a comprehensive approach to quantify multiple markers (such as longitudinal relaxation rate R_1, proton density PD, effective transverse relaxation rate R_2^*, and magnetization transfer saturation MT_{sat}) in a time-efficient MRI protocol composed of three multi-echo SPoiled Gradient Recalled echoes (SPGR) with PD-, T1-, and MT-weighting and additional calibration measurements (Helms et al., 2008a; Weiskopf et al., 2013) (Fig. 1).

---

* Corresponding author at: Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.
E-mail address: s.mohammadi@uke.de (S. Mohammadi).

https://doi.org/10.1016/j.neuroimage.2022.119529.
Received 29 December 2021; Received in revised form 20 July 2022; Accepted 1 August 2022

General

hMRI	*In vivo* histology using MRI
MPM	Multi-parameter mapping
SPGR	SPoiled Gradient Recalled echoes
qMRI	Quantitative MRI
VBM	Voxel-based morphometry
WM	White matter
GM	Grey matter

MPM derivatives

SNR	Signal-to-noise ratio
mSNR	Model-based signal-to-noise ratio
$SNR_{\mathrm{raw}}^{\mathrm{PDw}}$	Raw image-based SNR, see Eq. (1)
PD	Proton density
R_1 (T_1)	Longitudinal relaxation rate (time)
R_2^*	Effective transverse relation rate
MT_{sat}	Magnetisation transfer saturation rate
dPD	Error of PD
$dR1$	Error of R_1
dMT	Error of MT_{sat}
R_1^{RO} (PD^{RO}, $MT_{\mathrm{sat}}^{\mathrm{RO}}$)	Robustly combined R_1 (PD and MT_{sat}) values from two-repeat acquisition, see Eq. (7)
R_1^{AM} (PD^{AM}, $MT_{\mathrm{sat}}^{\mathrm{AM}}$)	Arithmetic-mean combined R_1 (PD and MT_{sat}) values from two-repeat acquisition
SEM	Standard-error-of-the-mean
rSEM	Relative SEM

Measurement and simulations

S_{PDw}^0 (S_{T1w}^0, $S_{\mathrm{M;Tw}}^0$)	PD-weighted SPGR signal fitted at zero echo time. In brackets: the same for the T1- and MT-weighted SPGR signal
B_1^+	Transmit field
M_0	Equilibrium magnetization
$\alpha_{\mathrm{PD}}^{\mathrm{nom}}$ ($\alpha_{\mathrm{T1}}^{\mathrm{nom}}$)	Nominal flip angle for PD (T1)-weighted signal
σ	Standard deviation of noise
$N(0,\sigma^2)$	Zero-mean, additive Gaussian noise with variance σ^2
TE	Echo time
TR	Repetition time
TA	Total acquisition time

MPM parameters are sensitive to key biological microstructure features, e.g., myelin density and iron content (Kirilina et al., 2020), as well as volumetric changes. For example, the MPM parameters R_1, PD, and MT_{sat} have demonstrated utility in revealing aging processes (Callaghan et al., 2014), assessing clinical pathology (Freund et al., 2013; David et al., 2019) and illuminating behaviourally-relevant brain microstructure (Whitaker et al., 2016; Ziegler et al., 2019), and are known to be sensitive to macromolecular content and thus correlate with myelin density (West et al., 2018; Mohammadi and Callaghan, 2021). This sensitivity to microstructural tissue properties can be expected to vary between quantitative metrics depending on the underlying MR contrast mechanisms (Edwards et al., 2018; Weiskopf et al., 2021).

The comprehensive acquisition providing the four aforementioned quantitative parameters makes the MPM protocol particularly attractive for large-scale neuroimaging studies (Whitaker et al., 2016; Taubert et al., 2020; Clark et al., 2021) and clinical trials (Leutritz et al., 2020). When planning these kind of studies, a key question is whether we can objectively determine which metric will have the greater statistical sensitivity to the effect of interest under the influence of noise and artefacts (e.g. due to subject movement). In particular, clinical studies or those recruiting special populations depend on reliable power estimates to assess feasibility and efficiency. Power estimates and heuristics from traditional structural MRI techniques such as voxel-based morphometry (VBM) or other anatomical shape analyses cannot be translated to the analysis of quantitative MPMs, since they target largely different mechanisms (cluster of neighbouring voxels via the Jacobi-determinant modulation in VBM vs. single-voxel quantification in MPMs) and result in different characteristics for metrics such as scan-rescan reproducibility (Schnack et al., 2010). The sensitivity of the MPM parameters under the influence of random noise is often determined by the signal-to-noise ratio (SNR) of the underlying weighted volumes. Since we combine multiple weighted volumes with different SNRs to compute MPM parameters, the noise propagation into the MPM parameter estimates plays an important role for power estimation in MPM-based studies.

While sensitivity to microstructural brain differences promises the detection of more subtle anatomical effects than accessible by VBM (Weiskopf et al., 2015), quantitative MRI maps are expected to be more susceptible to image artefacts than the constituent images. Moderate artefacts in the constituent images of quantitative MRI parameters can be amplified after nonlinear combination, e.g. when calculating the quantitative PD, R_1 and MT_{sat} maps from the MPM raw data (Fig. 1). As a consequence, the variability of quantitative MRI parameters across a cohort is usually a composition of the true anatomical variability and variability due to biases caused by instrumental, physiological, and movement related outliers and noise (Weiskopf et al., 2014; Castella et al., 2018; Lutti et al., 2022). The outliers can significantly reduce the effective SNR of the data at the group level and thereby the sensitivity to microstructural changes. Thus, the MPM approach would greatly benefit from the quantification of parameter-specific errors that can routinely identify outliers on a voxel-wise basis.

In this study, we introduce a new method to estimate error maps for each of the three quantitative MPM parameters R_1, PD, and MT_{sat} on a routine basis without the acquisition of any additional data (the error in R2* has been investigated elsewhere Weiskopf et al., 2014). The error maps are sensitive to two different types of variation: random noise, on the one side, and artefactual variation (e.g. due to imaging or subject motion), on the other. As a measure of the influence of random noise, we introduce the so-called model-based signal-to-noise ratio (mSNR) that is defined in analogy to the standard SNR, i.e. as the ratio between the MPM parameter and its error. First, we illustrate how the sensitivity to the two types of variation (random noise and artefacts) manifest themselves in the error and the mSNR maps. In two follow-up analyses, we investigate each of the two types of variation in more detail. The random-noise sensitivity is evaluated by quantifying the relation between the mSNR and the standard raw-image based SNR using both simulations and measurements. The artefact sensitivity is used to robustly combine MPM estimates from two successively acquired sets of MPM raw data (i.e. multi-echo SPGR images with PD-, T1-, and MT-weighting), where the acquisition-specific error maps are used to down-weight erroneous MPM parameters on a voxel-wise basis. We test the hypothesis that the robustly combined MPM estimates have lower image-artefact-related variability at the group level than the MPM estimates obtained from the arithmetic mean across acquisitions, or those obtained from a single MPM acquisition.

2. Background

This section contains terminology and background information about the measurement of the signal-to-noise ratio (SNR) from raw data, the MPM estimation framework, and how the error (and by extension the model-based SNR) can be estimated within this framework. The raw-data SNR (Eq. (1)) is used as a reference to validate the utility of the mSNR and its capacity to estimate the per-acquisition SNR in the MPM parameters.

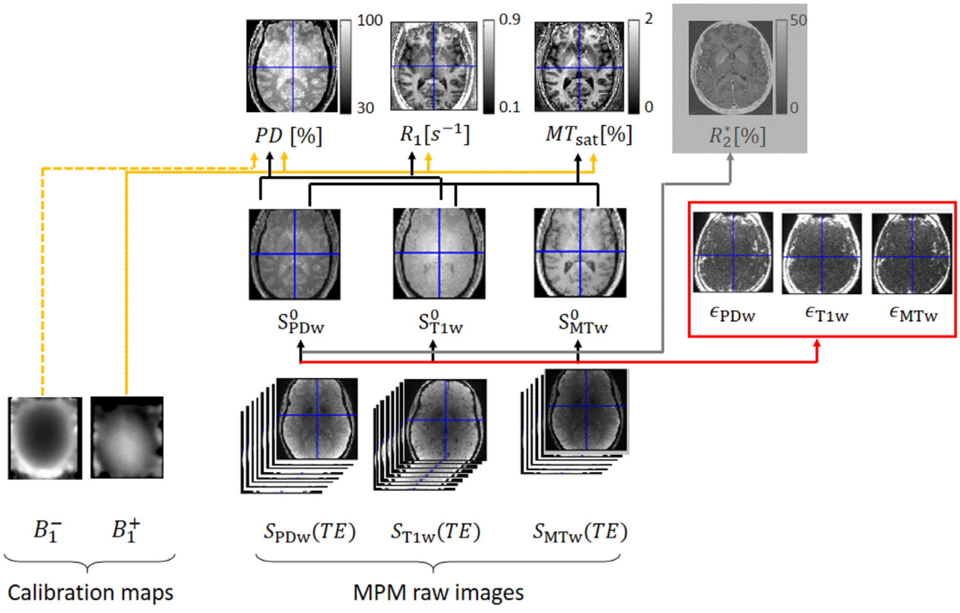


Fig. 1. From raw data to parameter maps. Bottom row: the multi-parameter mapping (MPM) raw data as well as the receive (B_1^-) and transmit (B_1^+) field maps acquired via dedicated calibration measurements. Middle row: the spoiled gradient-recall echo (SPGR) images with different contrasts at echo time (TE) zero ($S_{\text{PDw}}(TE=0) \equiv S^0_{\text{PDw}}$, $S_{\text{T1w}}(TE=0) \equiv S^0_{\text{T1w}}$, $S_{\text{MTw}}(TE=0) \equiv S^0_{\text{MTw}}$, middle row) as well as the three contrast-specific uncertainties (ϵ_{PDw}, ϵ_{T1w}, ϵ_{MTw}, red box, middle row), each of which summarizes the root-mean-square difference between modelled and measured signal per contrast (*Background*, Section 2). Top row: proton density (PD), longitudinal relaxation rate (R_1), magnetization transfer saturation (MT_{sat}), and effective transverse relaxation rate (R_2^*), which is not considered here and thus greyed out. In the MPM framework (Tabelow et al., 2019), the three quantitative parameters PD, R_1, and MT_{sat} are calculated from S^0_{PDw}, S^0_{T1w}, S^0_{MTw} after correction for B_1^- and B_1^+ fields.

2.1. Conventional SNR estimates of raw images

As a reference method to estimate SNR, we use the difference-image based SNR estimated from the raw multi-echo SPGR images with PD-weighting of a two-repeat measurement (Price et al., 1990; Reeder et al., 2005) or in short raw-image based SNR. For a twice repeated MPM measurement, the raw-image based SNR is estimated for the SPGR image with PD-weighting ($SNR^{\text{PDw}}_{\text{raw}}$) at the shortest echo time. The estimate of the mean signal is obtained from a small ROI by taking the mean of the signals from repeat 1 and 2, i.e. $\frac{1}{2}\langle S^{(1)}_{\text{PDw}} + S^{(2)}_{\text{PDw}}\rangle$, and this value is divided by the standard deviation of the difference of the signals from repeat 1 and 2 across the ROI, i.e. $\frac{1}{\sqrt{2}}\text{std}(S^{(1)}_{\text{PDw}} - S^{(2)}_{\text{PDw}})$, giving:

$$SNR^{\text{PDw}}_{\text{raw}} = \frac{1}{\sqrt{2}} \frac{\langle S^{(1)}_{\text{PDw}} + S^{(2)}_{\text{PDw}}\rangle}{\text{std}\left(S^{(1)}_{\text{PDw}} - S^{(2)}_{\text{PDw}}\right)} . \tag{1}$$

This method has been demonstrated to be a robust estimate of the SNR across different acceleration factors (Reeder et al., 2005; Dietrich et al., 2007). Note, that an equivalent metric to that in Eq. (1) can also be calculated for the other two contrasts (MTw and T1w) and other echo times. Without loss of generality we focused on $SNR^{\text{PDw}}_{\text{raw}}$ in this study.

2.2. Error in MPM parameters

The noise level in the multi-echo SPGR images with PD-, T1-, and MT-weighting is estimated from the contrast-specific uncertainties (ϵ_{PDw}, ϵ_{T1w}, ϵ_{MTw} in Fig. 1) derived from the root-mean-square (rms) difference of the predicted and measured signal decay across echo times. The errors in the MPM parameters differ from the associated contrast-specific uncertainties of the PD-, T1-, and MT-weighted SPGR signal as illustrated in Fig. 2. This section summarizes how these metrics are related.

Error propagation: To estimate the error of each quantitative map, we calculated the first order propagation of error in R_1, PD, and MT_{sat} under the assumption of uncorrelated errors between the SPGR and calibration measurements. For example, the error of R_1 was derived to be:

$$dR_1\left(S_{\text{T1w}}, S_{\text{PDw}}, B_1^+\right) = \sqrt{\left(\frac{dR_1}{dS_{\text{T1w}}} dS_{\text{T1w}}\right)^2 + \left(\frac{dR_1}{dS_{\text{PDw}}} dS_{\text{PDw}}\right)^2 + \left(\frac{dR_1}{dB_1^+} dB_1^+\right)^2} \tag{2}$$

with $R_1 \approx \frac{S^0_{\text{PDw}}\ \alpha_{\text{PD}}/TR_{\text{PD}} - S^0_{\text{T1w}}\ \alpha_{\text{T1}}/TR_{\text{T1}}}{2(S^0_{\text{T1w}}/\alpha_{\text{T1}} - S^0_{\text{PDw}}/\alpha_{\text{PD}})}$.

Local variations in the transmit field B_1^+ are incorporated into the flip angles via $\alpha_{\text{PD}} = B_1^+ \alpha^{\text{nom}}_{\text{PD}}$ ($\alpha_{\text{T1}} = B_1^+ \alpha^{\text{nom}}_{\text{T1}}$) with $\alpha^{\text{nom}}_{\text{PD}}$ ($\alpha^{\text{nom}}_{\text{T1}}$) being the nominal flip angles (Helms et al., 2008a; Lee et al., 2017; Tabelow et al., 2019), S^0_{PDw} (S^0_{T1w}) being the signal approximated at zero echo time $S_{\text{PDw}}(TE=0)$ ($S_{\text{T1w}}(TE=0)$) using the linearized SPGR signal fit (see next paragraph), and dS_{R1} and dS_{PD} being the contrast-specific uncertainties.

Linearized SPGR signal: To estimate the signal variation for each contrast, we first solved the joint model of the SPGR signals with PD-, T1-, and MT-weighting using the linearized exponential signal decay as introduced in the ESTATICS model (Weiskopf et al., 2014):

$$\vec{y} = \underline{X}\,\vec{\beta} + \vec{\mathrm{e}} \tag{3}$$

with $\vec{y} = \begin{pmatrix} \vec{y}_{\text{PDw}} \\ \vec{y}_{\text{T1w}} \\ \vec{y}_{\text{MTw}} \end{pmatrix}$ and $\vec{y}_m = \ln\begin{pmatrix} S_m(TE_1) \\ \vdots \\ S_m(TE_{N_m}) \end{pmatrix}$, $\underline{X}$ being an $N \times 4$ design matrix with rows $X_{k,i} = [\delta_{\text{PDw},i},\ \delta_{\text{T1w},i},\ \delta_{\text{MTw},i},\ -TE_k]$, N_m is the number of echoes for each contrast $m \in \{\text{PDw},\ \text{T1w}, \text{MTw}\}$, $\delta_{j,i}\begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ is the Kronecker delta, $\vec{\beta} = \begin{pmatrix} \ln(S^0_{\text{PDw}}) \\ \ln(S^0_{\text{T1w}}) \\ \ln(S^0_{\text{MTw}}) \\ R_2^* \end{pmatrix}$ are regression coefficients, and the model-fit residual vector $\vec{\mathrm{e}}$ is composed of elements e_k that are equal to the difference between predicted and measured logarithmic signal for the k-th MPM volume $k \in \{1, \ldots, N_{\text{PDw}} + N_{\text{T1w}} + N_{\text{MTw}}\}$. While it has been previously shown that the elements of $\vec{\mathrm{e}}$ can be used to down-weight outliers for robust estimation of R_2^* (Weiskopf et al., 2014), here, we use the rms difference of the predicted and measured signal as an estimate for the variation in the signal per contrast:

$$\epsilon_{\text{PDw}} \equiv \sqrt{\frac{1}{N_{\text{PDw}}} \sum_{j=1}^{N_{\text{PDw}}} \left(S_{\text{PDw}}(TE_j) - \exp\left(X_{j,:} \cdot \vec{\beta}\right)\right)^2}, \tag{4a}$$

$$\epsilon_{\text{T1w}} \equiv \sqrt{\frac{1}{N_{\text{T1w}}} \sum_{j=1}^{N_{\text{T1w}}} \left(S_{\text{T1w}}(TE_j) - \exp\left(X_{N_{\text{PDw}}+j,:} \cdot \vec{\beta}\right)\right)^2}, \tag{4b}$$

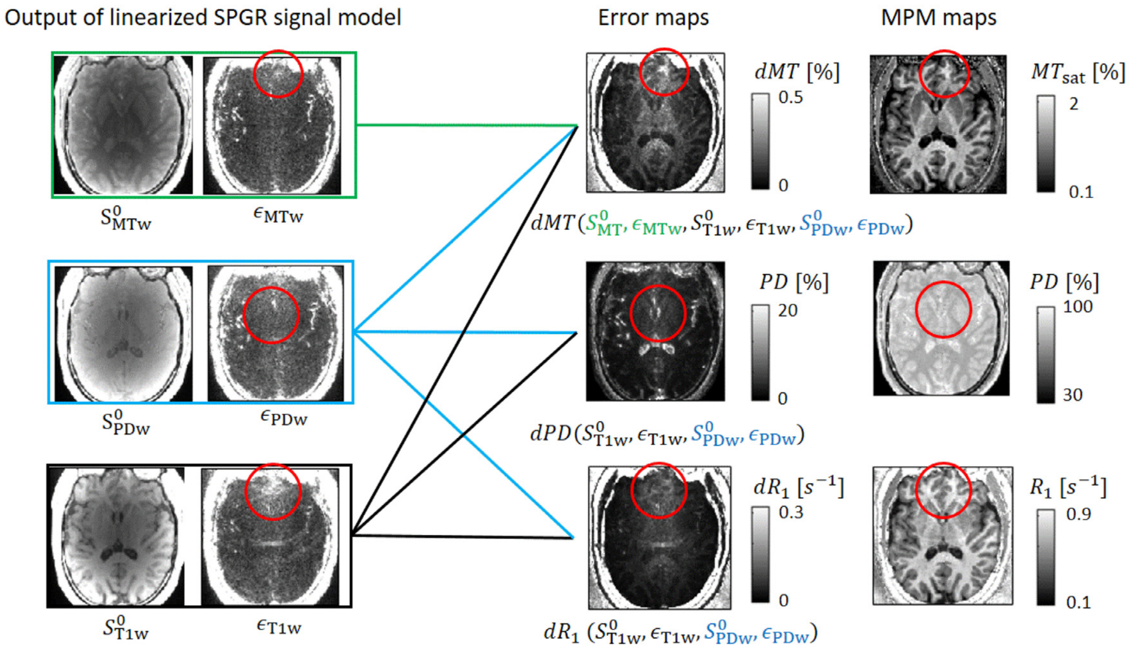


Fig. 2. Error maps, their dependencies on MPM data and sensitivity to artefacts. Left column: the offset (S^0_{MTw}, S^0_{PDw}, and S^0_{T1w}) and contrast-specific uncertainties ($\in_{\mathrm{MTw}}$, $\in_{\mathrm{PDw}}$, and $\in_{\mathrm{T1w}}$) of the SPGR signal as described in Fig. 1. Middle column: the error maps and their dependencies (coloured lines) on offset and contrast-specific uncertainties. Right column: the MPM maps. The MT-error map (top-middle) depends on all offsets and contrast-specific uncertainties. In contrast, the R_1- (bottom row) and PD- (middle row) error maps (dR_1 and dPD), depend only on S^0_{T1w} and S^0_{PDw} as well as on the corresponding contrast-specific uncertainties $\in_{\mathrm{T1w}}$ and $\in_{\mathrm{PDw}}$. The red circles highlight regions with higher error or higher contrast-specific uncertainties and the corresponding region in the MPM parameters. Higher values in contrast-specific uncertainties are often smeared out (e.g. in $\in_{\mathrm{T1w}}$) and localized increases are not necessarily accompanied by a biased MPM parameter (e.g. red circle in bottom row), whereas higher values in the error maps circle top row are co-localized with biased MPM parameters (e.g. circle in top row).

$$\in_{\mathrm{MTw}} \equiv \sqrt{\frac{1}{N_{\mathrm{MTw}}}\sum_{j=1}^{N_{\mathrm{MTw}}}\left(S_{\mathrm{MTw}}(TE_j) - \exp\left(X_{N_{\mathrm{PDw}}+N_{\mathrm{T1w}}+j,:}\cdot\vec{\beta}\right)\right)^2}. \qquad (4c)$$

In the following, the measures in Eq. (4) are denoted as contrast-specific uncertainties (Fig. 1).

Error maps: Using the contrast-specific uncertainties (Eq. (4)) to characterise the noise within a given MPM acquisition, the error maps can now be estimated using the concept of error propagation. For the example of the error of R_1 (Eq. (2)), the following is obtained:

$$dR_1 \approx \sqrt{\left(\frac{dR_1}{dS_{\mathrm{T1w}}}\in_{\mathrm{T1w}}\right)^2 + \left(\frac{dR_1}{dS_{\mathrm{PDw}}}\in_{\mathrm{PDw}}\right)^2}, \qquad (5)$$

where, the error in the transmit field is assumed to be negligible $dB_1^+ \approx 0$. The same approach can be used to generate error maps for the PD and MT_{sat} estimates. Details of the derivation and the formulae of the errors for each parameter (dR_1, dPD, dMT) can be found in Background Supplementary Materials B1–B3 and their implementation in the hMRI toolbox can be found here: www.hMRI.info.

Sensitivity of error maps to imaging artefacts: The relation between the error maps, the intercept of the linearized SPGR-signals which enters the error maps via the derivative of the MPM contrasts (e.g. for dR_1 see the formulae for $\frac{dR_1}{dS_{\mathrm{T1w}}}$ and $\frac{dR_1}{dS_{\mathrm{PDw}}}$ in the Background Supplementary Material B1–B3, Eqs. (B3) and (B4)) and the contrast-specific uncertainties is illustrated in Fig. 2. We hypothesize that the error maps are sensitive to the factors that contribute to image degradation captured by the contrast-specific uncertainties (Eq. (4)), such as head motion during the acquisition of each image volume (Weiskopf et al., 2014; Castella et al., 2018), physiological artefacts, or sequence-specific artefacts (e.g. noise-enhancement due to parallel-imaging). Our hypothesis is motivated by the observation that artefacts in the contrast-specific uncertainty maps are often smeared out across the whole brain (e.g. in $\in_{\mathrm{T1w}}$, Fig. 2) and localized increases are not necessarily accompanied by a biased MPM parameter. High values in the error maps, however, are localized and coincide with noticeable variation in the associated MPM parameter maps (red circles in top row, Fig. 2).

2.3. Model-based SNR (mSNR)

Following the definition of SNR, we introduce here the model-based SNR maps (in short mSNR maps) for each MPM parameter map. This is defined as the ratio between a quantitative MPM map and its corresponding error map. For example, the mSNR map for the R_1 parameter was calculated as follows:

$$mSNR^{\mathrm{R1}} = \frac{R_1}{dR_1}. \qquad (6)$$

To avoid divergence for very small error values, a threshold was used to set the mSNR map to zero when $\mathrm{dR1} < 10^{-4}$ (the same was done for $mSNR^{\mathrm{MT}}$ and $mSNR^{\mathrm{PD}}$, with $\mathrm{dMT} < 10^{-4}$ and $\mathrm{dPD} < 10^{-2}$, respectively). These thresholds were chosen heuristically. Note that the MPM-specific mSNRs are estimated per MPM acquisition Eqs. (5) and (6).

2.4. Robust combination of MPM parameters

When multiple MPM datasets are available, the resulting parameters can be combined by a simple arithmetic mean. Here we propose an alternative robust combination, of potentially erroneous MPM parameters, by exploiting their empirical error maps. The idea is formulated for the case of two distinct imaging repeats but can be generalized to multiple repeats.

To generate robust MPM parameters (denoted by superscript "RO") from a two-repeat protocol a function of the error maps is used to weight each repeat according to their voxel-wise error. For the example of R_1, the robust-combination was defined as follows on a voxel-by-voxel basis:

$$R_1^{\mathrm{RO}} = w_{R1}\ R_1^{(1)} + \left(1 - w_{R1}\right)\ R_1^{(2)}, \tag{7}$$

where, $w_{R1} = \frac{1}{mw_{R1}} f(\frac{dR_1^{(1)}}{dR_1^{(2)}})$; $f(x) = \frac{1}{\exp(100\ (x-1)/k)+1}$ is the Fermi function; $R_1^{(1,2)}$; $dR_1^{(1,2)}$ are the longitudinal relaxation rates and their respective errors from repeats (1) and (2), all defined on a per voxel basis; and $mw_{R1} = \max_{\vec{r}}(f(\vec{r}))$ is defined as the maximum weight across voxels $\vec{r}$. The parameter k tunes the sensitivity of the weights with respect to the ratio of errors: small k leads to high sensitivity. The parameter ($k = 10\%$) was heuristically optimized for one subject that showed strong motion artefacts in the scan-rescan measurements (see Supplementary Materials S1, "Robust combination and two-repeat multi-parameter mapping protocol") and applied to the rest of the subjects. The robust combine option can to be enabled via the local default file of the hMRI toolbox (hmri_def.wcomb), where also the k parameter can be adjusted for different protocols (e.g., hmri_def.wcombparams.kt = 10).

3. Materials and methods

3.1. Subjects and MRI

Subjects: 20 healthy volunteers participated in this study. 18 were measured with protocol 1, one with protocol 2 at 3T, and one with protocol 3 at 7T. We excluded two participants measured with protocol 1 from the analysis: one due to excessive, unsalvageable levels of movement and one due to image reconstruction problems. We included the remaining 16 participants in the reported group analysis (age: 20–54 years; $mean(SD) = 32.63(8.55)$ years; 7 female, 9 male). Exclusion criteria were any psychiatric disorders, assessed via the Mini-International Neuropsychiatric Interview (Ackenheil et al., 1999), neurological diseases, head trauma or metallic implants. Participants provided written informed consent and were compensated for their participation. The local ethics committees at University Medical Center Hamburg-Eppendorf and Medical Faculty of the University of Leipzig approved the study (PV5141; LPEK_006_Kühn; Reg.-No. 273-14-25082014; WF-74/16).

Post mortem validation: Whole brains were obtained via the Department of Legal Medicine, Medical Centre Hamburg-Eppendorf.

MRI protocol: Scans were performed on three MRI systems (Siemens Healthineers, Erlangen, Germany): 3T PRISMA (protocol 1), 3T PRISMA-fit (protocol 2), and Magnetom 7T (protocol 3). For protocols 1 and 2, the body coil was used for transmission (Tx) and the 64-channel receiver head-coil for reception (Rx). For protocol 3, an integrated 1-channel Tx/32-channel Rx head coil (Nova Medical, Wilmington, MA, USA) was used. Whole brain MR images were acquired using the MPM (Weiskopf et al., 2013) protocol, including three differently weighted (MT-, PD- and T_1-weightings) multi-echo SPGR contrasts (protocol 3 only acquired two weightings, PD- and T_1-weighting). The effect of spatial inhomogeneities of the radio-frequency transmit field B_1^+ on the quantitative MRI maps (Lutti et al., 2014) was corrected using specially-dedicated data acquired using the 3D echo-planar imaging (EPI) spin-echo and stimulated echo method (Lutti et al., 2010; 2012), available in the hMRI toolbox (Tabelow et al., 2019). Image distortions in the EPI data were corrected using B0-field mapping data (2D double-echo FLASH sequence). For protocols 1 and 2, the automatic transmit adjust procedure was used, whereas for protocol 3, the transmit voltage was calibrated using an initial low-resolution transmit field map to be optimal over the occipital lobe.

The sequence parameters for protocols 1-3 are summarized in Table 1.

The acquisition of all multi-echo SPGR contrasts was repeated for each individual. This was done within a single imaging session at 3T (i.e. protocols 1 and 2 contained two "runs" of each contrast) and in two separate imaging sessions at 7T (i.e. protocol 3 contained only one run).

The total scan time of both runs was about 28 min (= 2 × 11 + 6) for protocol 1, about 33 min (= 2 × 13.5 + 6) for protocol 2, and about 84 min (= 2x(36 + 6)) min for protocol 3.

For protocol 3, participant motion was monitored and corrected prospectively by an optical tracking system (Kineticor, Honolulu, HI, USA) (Callaghan et al., 2015). Each volunteer was scanned while wearing a mouth guard with a passive Moiré pattern marker used for tracking (manufactured by the Department of Cardiology, Endodontology and Periodontology, University Medical Center Leipzig; comparable to Papoutsi et al., 2020). The dataset acquired with protocol 3 was taken from the study by (McColgan et al., 2021).

In all protocols, parallel imaging was performed using generalised autocalibrating partial parallel acquisition (GRAPPA) (Griswold et al., 2002) with acceleration factors of 3 (protocol 1) or 4 (protocols 2 and 3). We used a range of protocols which have been designed for neuroscience studies (protocol 2, e.g.: Ellerbrock and Mohammadi, 2018; protocol 3: McColgan et al., 2021; protocol 1 is used in a running study), to assess the utility of the proposed error and mSNR maps. These spanned spatial resolution and static magnetic field strength (protocols 1 and 2: 1 mm isotropic at 3T vs. protocol 3: 0.5 mm isotropic at 7T) as well as different acceleration strategies to reduce scan time (2 × 2 acceleration in protocols 2 and 3 vs. 3 × 1 acceleration plus partial Fourier sampling in protocol 1) while keeping field strength and spatial resolution fixed.

3.2. Map creation and spatial processing

Map creation and spatial processing was performed using modules in SPM12 version v7771 (Friston et al., 2006) and a branch of the hMRI toolbox (Tabelow et al., 2019) available here: https://github.com/hMRI-group/hMRI-toolbox/releases/tag/errormaps. For B_1^+ correction of the MT_{sat} maps, we used a heuristic correction factor as detailed in (Helms, 2015; Helms et al., 2021). Note, that the error and mSNR maps will be part of the official hMRI toolbox (www.hMRI.info) upon publication. To be generated, they need to be enabled in the local default folder (hmri_def.errormaps = true). An instruction of where to find the error maps and how to generate the robust combination is given in Supplementary Materials S2 and will become available on the hMRI wiki page (https://github.com/hMRI-group/hMRI-toolbox/wiki).

Rigid-body registration: To ensure that the data from both repeats were in the same space, the second dataset was registered to the first using a rigid-body transformation (spm_coreg). To this end, first the MPM maps were aligned to the MNI-space template ("avg152T1") with the auto-align module in the hMRI toolbox (Tabelow et al., 2019) using the MT_{sat} map as "source image" (i.e. the image that was used for estimating the transformation parameters). Then, the MT_{sat} maps were thresholded (MT_{sat}>0 and MT_{sat}<5 p.u.) to remove unreasonable or extreme values and segmented (spm_segment) into grey and white matter tissue probability maps (TPMs). These grey and white matter TPMs were combined and used as source and target images for inter-repeat registration, with the TPMs from the second repeat being the source and the TPMs of the first repeat being the target image. The estimated transformation parameters were applied to all maps from the second repeat. The reason for using the grey and white matter TPMs for registration instead of the original maps was to reduce potential confounding effects of motion artefacts in the individual MPM images on the accuracy of the registration.

Non-linear spatial registration and spatial smoothing in common group space: First the MPM maps of each subject were transformed into MNI space using the geodesic shooting nonlinear registration tools (spm_shoot, Ashburner and Friston, 2011). To estimate the nonlinear transformation that maps each individual brain into common space, high-quality grey and white matter tissue probability maps were generated per subject by segmenting the arithmetic mean of the MT_{sat} maps across both repeats. Then, the hMRI toolbox was used for tissue-specific smoothing (Tabelow et al., 2019).

Table 1
MRI parameters.

Acquisition parameters	Protocol 1	Protocol 2	Protocol 3
Field strength	3T	3T	7T
PDw & T1w min. TE : ΔTE : max. TE | # of echoes	2.3:2.3:18.4 [ms] | 8	2.3:2.3:18.4 [ms] | 8	2.8: 2.8:16 [ms] | 6
MTw min. TE : ΔTE : max. TE | # of echoes	2.3: 2.3:13.8 [ms] | 6	2.3: 2.3:13.8 [ms] | 6	None
Repetition time (TR)	25.0 [ms]	25.0 [ms]	25.0 [ms]
Acquisition resolution	$1 \times 1 \times 1$ [mm^3]	$1 \times 1 \times 1$ [mm^3]	$0.5 \times 0.5 \times 0.5$ [mm^3]
Acceleration factor in phase and partition directions	3×1	2×2	2×2
Partial Fourier in phase | partition directions	6/8 | 6/8	OFF | OFF	OFF | OFF
Flip angle PDw, MTw	6°	6°	5°
Flip angle T1w	21°	21°	24°
MT pulse angle	220°	220°	None
MT pulse length	4 [ms]	4 [ms]	None
MT pulse off res. freq.	2 [kHz]	2 [kHz]	None
MT pulse shape	Gaussian	Gaussian	None
ApS	3:46 [min]	4:42 [min]	18 [min]

Different sequence parameters for protocols 1–3. Note that protocol 3 does not include an MTsat map. Abbreviations: FoV, field of view, MT(w), magnetization transfer (weighted), PD(w), proton density (weighted), T1w, T1 weighted, T1, longitudinal relaxation time, TE, echo time, ApS, acquisition time per set of multi-echo SPGR images, SPGR, spoiled gradient-recall echo.

3.3. Analysis

Three analyses were performed, which we describe in detail below. In the first analysis, we assessed two different features of the error and mSNR maps: their sensitivity to the image SNR and to image artefacts. Second, the relation between mSNR and the image noise of SPGR raw images was quantitatively investigated. Third, the artefact-associated variability of MPM parameters at the group level was investigated for different combinations of a two repeat MPM protocol.

Analysis I: Illustration of error and mSNR maps: On the single-subject level, we investigated how the mSNR varied for different MPM protocols and how imaging artefacts in the MPM parameter maps appeared in the error and the mSNR maps. For the same subject, we used two repeats with varying artefact levels. Local artefacts were visually identified as biases in the respective MPM parameter maps. The correspondence between error and respective MPM parameter maps in erroneous regions was depicted. Further, it was investigated whether a robust combination of the MPM parameter values per repeat, i.e. a weighted-average towards the less erroneous repeat, was less biased than the arithmetic mean average of both repeats.

Analysis II: Quantifying the relation between raw-image-based and model based SNR: The propagation of SNR into mSNR was characterized by simulations and *in vivo* measurements. Hereby, we used the raw-image-based SNR of the PD-weighted image ($SNR_{\mathrm{raw}}^{\mathrm{PDw}}$, see Eq. (1)) as a proxy for the image SNR, both in the simulations and measurements. To assess the dependence between SNR^{PDw} and $mSNR^{m}$ for varying SNRs, a linear model was fit to the data:

$$mSNR_j^m = A_j^m \; SNR_j^{\mathrm{PDw}} + B_j^m, \tag{8}$$

with A_j^m and B_j^m being the fitted parameters, $m \in \{\mathrm{PD}, \mathrm{R1}, \mathrm{MT}\}$ and j being the index that specifies whether simulated or measured data was used ($j \in \{\mathrm{sim}, \mathrm{meas}\}$).

Simulated $SNR_{\mathrm{sim}}^{\mathrm{PDw}}$ and $mSNR_{\mathrm{sim}}^{m}$: To simulate the noisy SPGR signal, we added complex-valued Gaussian noise to the rational approximation of the Ernst equation (Ernst and Anderson, 1966; Helms et al., 2008a) with heuristic approximation of the MT-pulse effect (Helms et al., 2008b). Then, the absolute value of the noisy signal was calculated as follows:

$$S\left(TR, TE, R_1, R_2^*, M_0, \alpha, MT_{\mathrm{sat}}, \sigma\right) = \mathrm{M}_0 \left\| \alpha \frac{TR\ R_1}{\frac{\alpha^2}{2} + MT_{\mathrm{sat}} + TR\ R_1} \exp\left(-TE\ R_2^*\right) + \frac{1}{\sqrt{2}}(x + iy) \right\| \text{ with } x, y \sim N\left(0, \sigma^2\right) \tag{9}$$

Table 2
Simulation parameters.

	M_0^{GT}[p.u.]	$R_1^{\mathrm{GT}}[s^{-1}]$	$R_2^{*GT}[s^{-1}]$	$MT_{\mathrm{sat}}^{\mathrm{GT}}$[p.u.]
White matter	69.8	0.94	22	1.59
Grey matter	77.6	0.70	15	1.04

The ground truth parameters (with superscript GT) that were used to simulate the signal in Eq. (9) for grey and white matter. The other parameters (repetition time: TR, echo time: TE) were as in protocol 1. Finally, the zero-mean, additive Gaussian noise $N(0, \sigma^2)$ with varying standard deviation was added ($\sigma \in \{0.002, \ldots, 0.1\}$), each time with 5000 noise realisations. This resulted in $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$ values between 62 and 2.

The multi-echo SPGR signals were simulated using Eq. (9) and the parameters in Table 2. To achieve PD- and T1-weighting we use the respective flip angles for protocol 1 (see Table 1) and $MT_{\mathrm{sat}} = 0$ in Eq. (9). The MT-weighted multi-echo SPGR signal was simulated using the same parameters as for the PD-weighted signal only with the difference that the MT_{sat} value in Table 2 was used. Then, $mSNR^{\mathrm{PD}}$, $mSNR^{\mathrm{R1}}$, and $mSNR^{\mathrm{MT}}$ were calculated from the simulated SPGR signals using the proposed approach (see Background and Supplementary Materials B1–B3). Additionally, $SNR_{\mathrm{sim}}^{\mathrm{PDw}}$ was calculated from the simulated signal using Eq. (1). Note that the simulation was performed separately for white matter and grey matter ground truth parameters (Table 2) to investigate the influence of tissue type on the relation between $SNR_{\mathrm{sim}}^{\mathrm{PDw}}$ and $mSNR^{m}$. Also note, that the same simulation was used to investigate the relation between $mSNR^{m}$ and the experimental SNR of the MPM parameters (see Supplementary Materials S3).

Measured $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$ and $mSNR^{m}$: To partition the MRI brain volumes into regions of interest (ROIs) with varying SNR, we used the Oxford-Harvard atlas for grey and white matter (Frazier et al., 2005; Desikan et al., 2006; Makris et al., 2006; Goldstein et al., 2007). Hereby, it was assumed that regions closer to the skull have higher SNR because they are closer to the head coil than regions within the centre of the brain. Note that this analysis assumes that the proposed SNR measures are independent of the tissue type (i.e. whether it is grey or white matter) but solely depend on their distance to the head coil.

The SNR measures were estimated for each subject in individual space, to prevent interpolation artefacts associated with the spatially non-linear registration into common space affecting the estimation of $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$. To this end, ROIs were projected into individual space using the inverse of the spatial transformations estimated in the “spatial processing section”. Then, the individual tissue probability maps for white

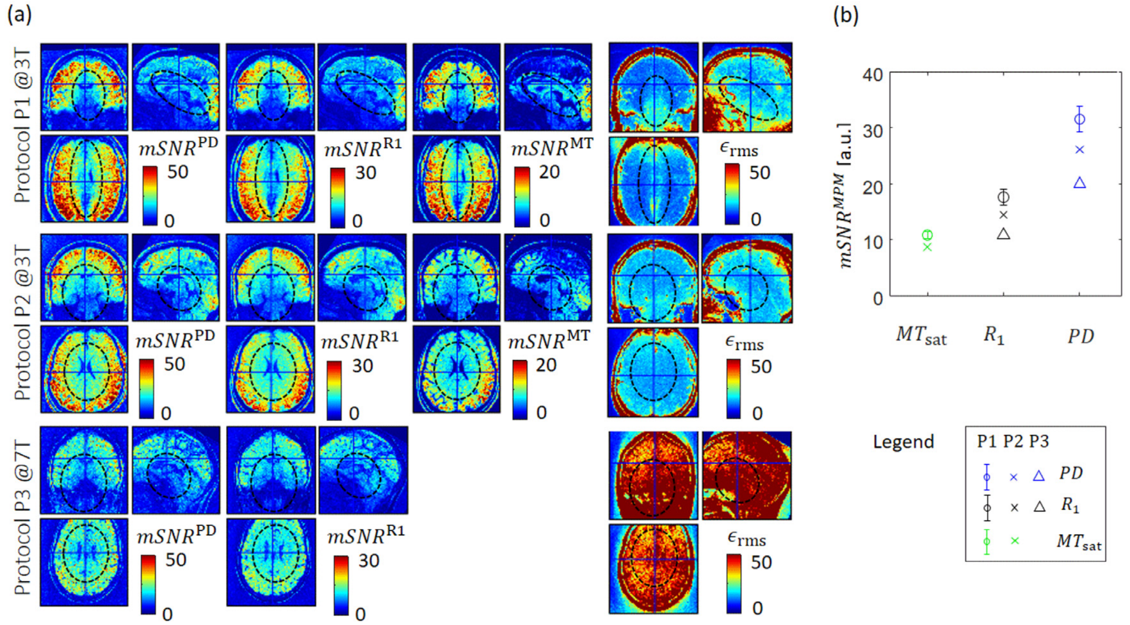


Fig. 3. Variation of mSNR maps across the brain for different MPM protocols. (a): mSNR maps for three MPM protocols (P1-P3) are depicted (1st row: P1, 2nd row: P2, and 3rd row: P3): $mSNR^{\mathrm{PD}}$ (1st column), $mSNR^{\mathrm{R1}}$ (2nd column), $mSNR^{\mathrm{MT}}$ (3rd column). Fourth column: the averaged across contrast-specific uncertainties is depicted ($\epsilon_{\mathrm{rms}} = \sqrt{\frac{1}{3}(\epsilon_{\mathrm{PDw}}^2 + \epsilon_{\mathrm{T1w}}^2 + \epsilon_{\mathrm{MTw}}^2)}$, 4th column). (b) The mSNR parameters averaged across the brain for protocols P1 (circle), P2 (cross), and P3 (triangle) for the three MPM parameters (PD: blue, R_1: black, MT_{sat}: green). Note that P1 was calculated across a group of healthy subjects (standard deviation across group in black) whereas P2 and P3 were calculated only for a single subject (no standard deviation). $mSNR^m$ and ϵ_{rms} in arbitrary units. The values in the mSNR maps decrease towards the centre of the brain accompanied by an increase in the ϵ_{rms} values. While the mSNR map measured with protocol 1 showed a strong left-right gradient (middle row, left: ellipse with large eccentricity), the mSNR maps measured with protocols 2 (and two out of three mSNR maps measured with protocol 3) showed a circular shaped area of decreased values (middle row: middle and right). Note that, since no MT measurement was available for protocol 3, ϵ_{rms} was the mean of the remaining two residuals.

and grey matter were multiplied with each individual ROI and thresholded at 90%. Only ROIs containing more than 100 voxels were used for the analysis. Then, $\mathrm{SNR}_{\mathrm{raw}}^{\mathrm{PDw}}$ and $\mathrm{mSNR}^{\mathrm{m}}$ were estimated within each ROI for each subject. This experiment was performed for the data acquired with protocol 1.

Additionally, the subject-averaged $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$ per ROI was projected into a template image in group space for visualization purposes only. The spatial dependence of mSNR maps was qualitatively compared for two different protocols at 3T (protocol 1 and 2) and one at 7T (protocol 3). Then, the average mSNR value across the brain was quantified per MPM parameter and protocol (averaged across subjects for protocol 1).

Analysis III: Artefactual variation at the group level: The variability of MPM parameters across a group of healthy subjects is expected to be a combination of the true anatomical variability in the cohort and artefactual variability caused by instrumental, physiological, and subject-movement related noise. While the latter can be reduced by averaging, the former cannot. To assess the variability across the group, we calculated the standard-error-of-the-mean (SEM) for four sets of MPM parameters in MNI space after tissue-specific hMRI smoothing (see section “spatial processing”). The four sets of MPM parameters were generated either from repeats (1) and (2) separately, or from the arithmetic mean (AM) across repeats or their robust combination (RO in Eq. (7)).

To assess the effect of instrumental, physiological, and subject-movement on the variability, the SEM of the arithmetic-mean combined MPM parameters (SEM_{AM}^m) were used as reference (with $m \in \{\mathrm{MT}, \mathrm{PD}, R1\}$) and compared to the SEM of the other sets of MPM parameters via their relative difference: $rSEM_i^m = (\frac{SEM_i^m}{SEM_{\mathrm{AM}}^m} - 1) \times 100$ with $i \in \{1, 2, \mathrm{RO}\}$. Hereby, a negative (positive) $rSEM_i^m$ would indicate that the variability in the i dataset is smaller (larger) than in the reference AM dataset. Since it is expected that instrumental, physiological, and subject-movement artefacts increase the variability of the estimated parameters, negative $rSEM_i^m$ values in each dataset were interpreted as a reduction of artefactual variability and thus an increase of sensitivity towards group differences whereas positive $rSEM_i^m$ values were interpreted as a decline of sensitivity. Note that the scan time of the two-repeat measurement (28 min = 2 × 11 min + 6 min) was only 65% longer than the scan time of the one-repeat measurements (17 min = 1 × 11 min + 6 min) because the calibration measurements (~ 6 min) were not repeated.

4. Results

4.1. Analysis I: Illustration of error and mSNR maps

This analysis illustrates how variations due to image SNR and artefacts manifest themselves in the mSNR and error maps for different MPM protocols and at different field strengths. The different protocols showed a decrease of mSNR towards the centre of the brain. The variation of mSNR as a function of coil configuration and head positioning is illustrated in Figs. S4.1 and S4.2 (see Supplementary Materials S4). Although protocols 1 and 2 were measured with similar parameters (e.g. spatial resolution, TE, TR) and instruments (e.g. at 3T and with a 64ch head-coil), the decrease towards the centre of the brain was differently shaped within the respective mSNR maps: the mSNR maps measured with protocol 1 had an elliptical pattern of lower mSNR centrally, the long axis of which ran anterior-posterior whereas the central region of lower mSNR was circular for protocol 2, but decreased less rapidly than protocol 1 (dashed lines in Fig. 3a). The decline of mSNR was accompanied by an increase in the error map values (see Supplementary Materials, S5). Interestingly, the shape of mSNR decline in protocol 1 was not only visible in the error maps (see Supplementary Materials, S5) but also in the average contrast-specific uncertainty maps for protocol 1 (1st row, 4th column, Fig. 3a), indicating that the noise or artefact-level is higher in the

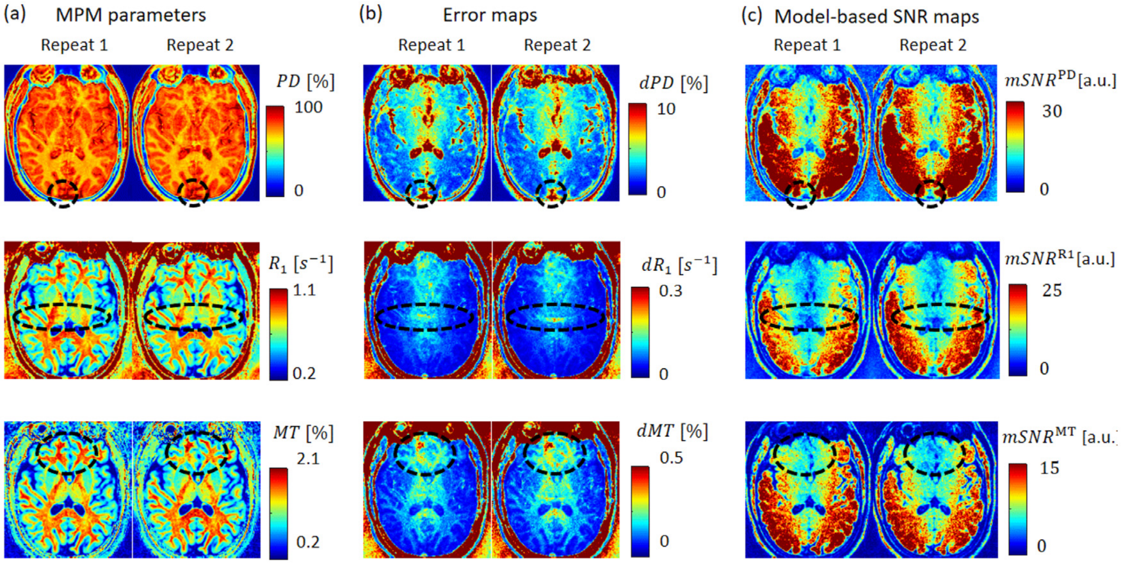


Fig. 4. Artefacts in the MPM, error, and mSNR maps. Depicted are the quantitative MPM parameters (left column), the error maps (middle column), and corresponding mSNR maps (right column) for a subject, measured in two repeats with varying artefacts using protocol 1. The mSNR maps have the same contrast, even though the contrast of the associated MPM map varies (top: PD, top: R_1, bottom: MT_{sat}). The artefacts in the parameter maps (highlighted by dashed circles) manifest themselves as increased error-map values and reduced mSNR-map values. Three artefacts were identified with potentially different origin: physiological noise that remained almost the same between repeats at the superior sagittal sinus (top row), aliasing artefacts that varied between repeats (middle row), and voluntary subject motion artefacts that strongly varied between repeats (bottom row). Note that intensity ranges for the three mSNR maps differ.

areas of decreased mSNR and varies with protocol. Although $mSNR^{\mathrm{PD}}$ and $mSNR^{\mathrm{R1}}$ maps measured with protocol 3 showed smaller gradients towards the centre of the brain as compared to their counterparts measured with protocols 1 and 2, they revealed the same trend (higher mSNR values towards the cortex and lower towards the centre of the brain). On average, across the brain, we found that the mSNR values decreased from protocol 1 to 3 (Fig. 3b): the mSNR of protocol 1 was about 1.2 times larger than the mSNR of protocol 2, whereas the ratio between the mSNR values of protocols 1 and 3 was about 1.6 for PD and R_1. The ratio of $mSNR^{\mathrm{PD}}$ to $mSNR^{\mathrm{R1}}$ was similar across protocols: 1.79 for protocol 1, 1.80 for protocol 2, and 1.82 for protocol 3.

Fig. 4 shows the relation between each MPM parameter, the associated error and mSNR maps. The mSNR maps had the same contrast for all MPM parameters, while the contrast in the error maps reflects the MPM parameter contrast (PD: high in grey matter and low in white matter, MT_{sat} and R_1: low in grey matter and high in white matter).

As expected, artefacts in the MPM maps (highlighted by dashed circles) manifested differently in the error and mSNR maps: while the error in regions affected by artefacts was increased, the mSNR value was decreased because the latter depends reciprocally on the error. Three artefactual regions with potentially different artefact-causes were identified. The artefact in the top row may be caused by physiological noise due to flow artefacts because it was located in the superior sagittal sinus and did not vary between repeats. In the middle row, an aliasing artefact was identified which varied between repeats and thus might be enhanced by involuntary subject motion. The artefact highlighted in the bottom row was most likely a result of involuntary subject motion because it varied between the two repeats.

Fig. 5 illustrates regionally localized artefacts in the MPM parameters that were captured by the error maps, became less pronounced in the arithmetic mean, and could be partly removed in the robust combination (Fig. 5 for R_1, the corresponding figures for PD and MT_{sat} can be found in Supplementary Materials S6). In the Supplementary Materials (S6, Fig. S6.3), we demonstrated that these artefacts were related to involuntary subject-motion, appeared as an increased error, and that the errors could be used in a two-repeat combination to reduce the bias in the MPM parameters.

4.2. Analysis II: quantifying relation between raw-image and model based SNR

Here, we quantify a linear relation between image SNR and mSNR at the group level and in simulations. Fig. 6 depicts the linear relation between $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$ and the $mSNR^{\mathrm{m}}$ of the three MPM parameters ($m \in \{\mathrm{PD}, \mathrm{R1}, \mathrm{MT}\}$) in simulation (Fig. 6a–c) and in measurements (Fig. 6d–f) with slopes and intercepts reported in Table 3. For the measurement, variation in $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$ and $mSNR^{\mathrm{m}}$ was achieved by measuring the respective metrics within 111 regions of interest (ROIs) averaged across 16 healthy subjects using protocol 1 (Fig. 7).

The range of each SNR measure was as follows: 18-54 for $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$, 11-54 for $mSNR^{\mathrm{PD}}$, 6-30 for $mSNR^{\mathrm{R1}}$, and 4-18 for $mSNR^{\mathrm{MT}}$. We observed a linear relation between $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$ and the mSNRs with highly significant p-values for the slopes but non-significant p-values for the intercept in the simulated data (significance level: $p < 0.05$ with the null hypothesis in each case being that the parameter is zero). The slope of the mSNR parameter curve was steepest for $mSNR^{\mathrm{PD}}$, followed by $mSNR^{\mathrm{R1}}$, and was smallest for $mSNR^{\mathrm{MT}}$. The simulations revealed that the slope of the mSNR parameter curve was systematically smaller for WM than for GM (2% for $mSNR^{\mathrm{R1}}$, 6% for $mSNR^{\mathrm{PD}}$, and 14% for $mSNR^{\mathrm{MT}}$). We found a similar trend for the ratio of the slopes between measurements and simulations when taking the fitted $mSNR^{\mathrm{PD}}$-parameter as reference: the slope of $mSNR^{\mathrm{PD}}$ was 1.8 to 2 times higher than the slope of $mSNR^{\mathrm{R1}}$, whereas it was 3.2 to 4.5 times higher for $mSNR^{\mathrm{MT}}$, indicating that MTsat requires a much bigger gain in $SNR_{\mathrm{raw}}^{\mathrm{PDw}}$ for them to translate into gains in $mSNR^{\mathrm{MT}}$ as compared to PD. Moreover, we found that the fitted intercept differed substantially between measurements and simulations (Table 3).

4.3. Analysis III: reducing artefactual variation at the group level

In this analysis, we show how artefactual variations at the group level vary for different combinations of a two-repeat MPM acquisition. Variability at the group level was assessed by the standard-error-of-the-mean (SEM). For WM, the SEM of the arithmetic mean (AM) of the MPM parameters showed opposite variability between SEM^{PD} and SEM^{MT}, e.g.: SEM^{PD} was high in the cortical spinal tracts whereas SEM^{MT}

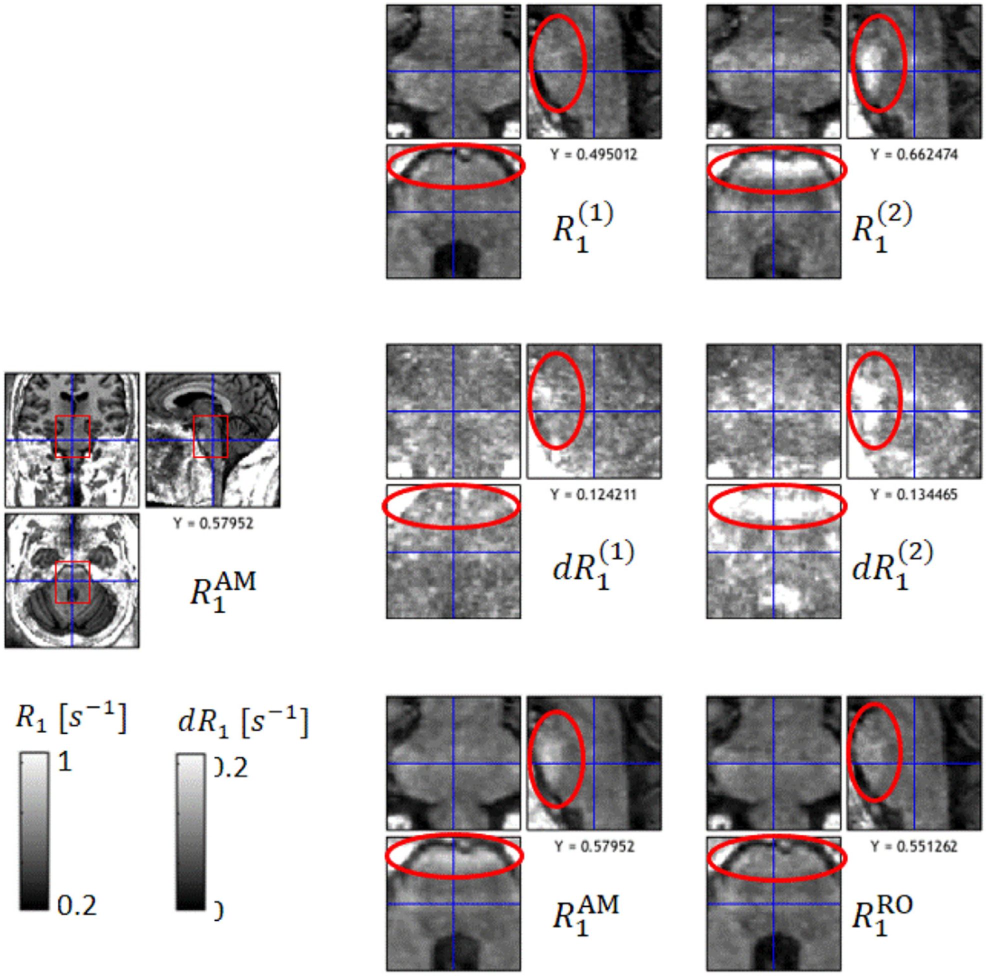


Fig. 5. Reduced artefacts in robustly-combined longitudinal relaxation rate (R_1) parameter. Depicted are: two successive repeats of the R_1 map using protocol 1 with superscript (1) and (2) (top row), the associated error maps for each repeat (middle row), and their arithmetic mean and robustly combined average with superscript *AM* and *RO* (bottom row). An area is magnified (red box, left column), where the error maps were sensitive to artefacts (hyper intensities) and the robust combined R_1 contained fewer artificially increased values than the arithmetic mean (circle) and single-repeat R_1 maps.

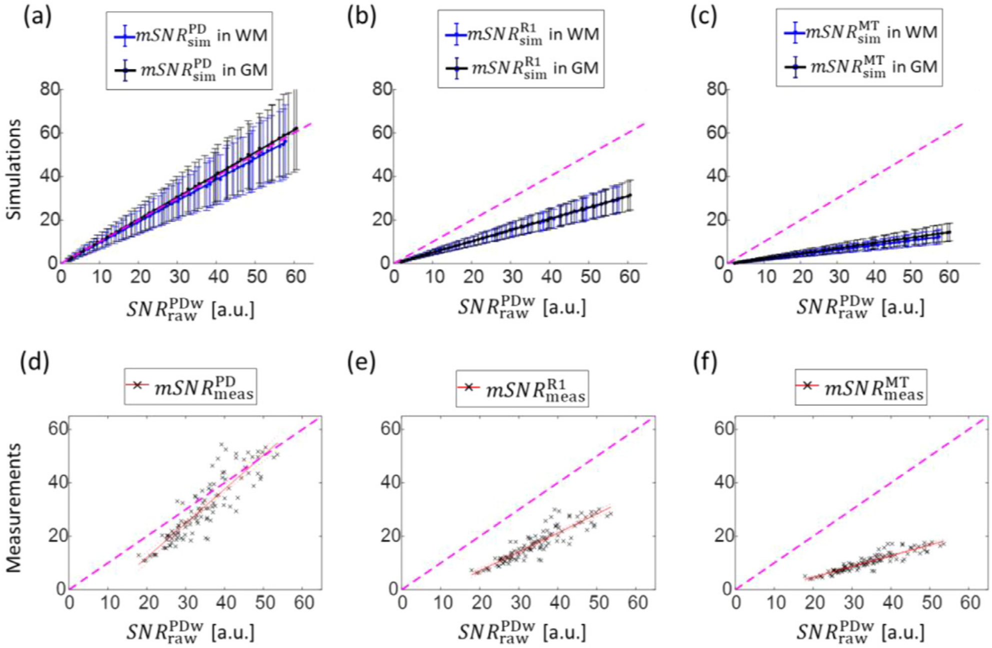


Fig. 6. Relating mSNR to the image SNR of the PD-weighted acquisition. Depicted is mSNR as a function of SNR_{raw}^{PDw} (Eq. (1)) using simulations (a–c) and measurements across a group of healthy subjects (d–f). (a–c): The simulated mSNR is depicted as a function of SNR_{raw}^{PDw} with mean (circle) and standard deviation (errorbar) across 5000 noise realisations for the PD (a), R_1 (b), and MT_{sat} (c) parameters (Eq. (9)). (c–e): For the ROIs depicted in Fig. 7 the mSNR parameters (crosses in a: $mSNR^{MT}$, b: $mSNR^{R1}$, c: $mSNR^{PD}$) are plotted against the corresponding SNR_{raw}^{PDw} value. A heuristic linear relation is fitted between the mSNR and SNR_{raw}^{PDw} data (Eq. (8)), both for simulated and measured data. The dashed line in magenta is the unity line.

Table 3

Simulated and measured relation between model-based and raw image-based SNR.

	Slope parameter A^{MPM} (p-value)	Offset parameter B^{MPM} (p-value)	Ratio slopes ($\frac{A^{\mathrm{PD}}}{A^{m}}$)
$mSNR^{\mathrm{PD}}_{\mathrm{meas}}$	1.28 (2.19×10^{-8})	-13.45 (2.19×10^{-8})	1
$mSNR^{\mathrm{R1}}_{\mathrm{meas}}$	0.71 (1.88×10^{-41})	-7.07 (1.06×10^{-8})	1.80
$mSNR^{\mathrm{MT}}_{\mathrm{meas}}$	0.40 (3.52×10^{-44})	-3.43 (1.28×10^{-7})	3.20
$mSNR^{\mathrm{PD}}_{\mathrm{sim}}$ in GM	1.03 ($2.05 \times {}^{-73}$)	-0.28 (0.13)	1
$mSNR^{\mathrm{PD}}_{\mathrm{sim}}$ in WM	0.97 (1.31×10^{-76})	-0.16 (0.26)	1
$mSNR^{\mathrm{R1}}_{\mathrm{sim}}$ in GM	0.52 (2.20×10^{-73})	-0.15 (0.12)	1.97
$mSNR^{\mathrm{R1}}_{\mathrm{sim}}$ in WM	0.52 (1.08×10^{-76})	-0.10 (0.17)	1.88
$mSNR^{\mathrm{MT}}_{\mathrm{sim}}$ in GM	0.24 (4.33×10^{-72})	-0.14 (0.004)	4.26
$mSNR^{\mathrm{MT}}_{\mathrm{sim}}$ in WM	0.22 (1.15×10^{-74})	-0.11 (0.002)	4.53

The coefficients of the heuristic linear models (Eq. (8)) that relate the $SNR^{\mathrm{PDw}}_{\mathrm{raw}}$ to the simulated ($mSNR^{m}_{\mathrm{sim}}$) and measured ($mSNR^{m}_{\mathrm{meas}}$) mSNR, summarizing the slopes and intercepts of the curves in Fig. 6, with A^{m} being the slope and B^{m} being the intercept, and the ratio between A^{PD} and A^{m} ($m \in \{PD, \mathrm{R1}, \ \mathrm{MT}\}$). The subscript "sim" refers to the simulated data in Fig. 6a–c and "meas" to the measured data in Fig. 6e and f (for details see methods section "Analysis II"). Note that for the measured mSNR the average across repeats was used.

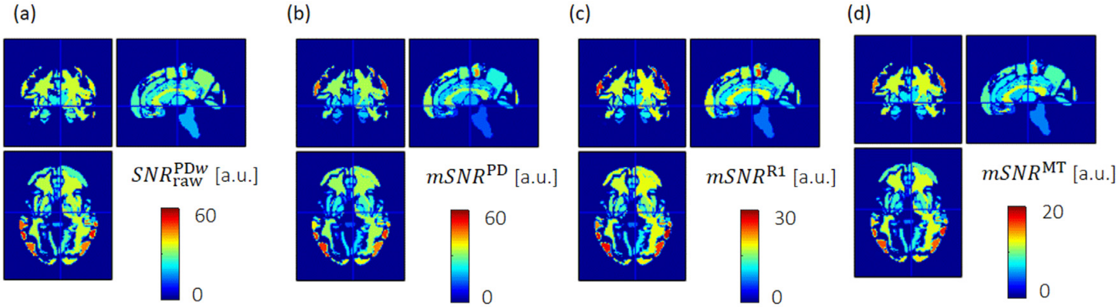


Fig. 7. mSNR and SNR^{PDw}_{raw} maps averaged within regions-of-interest (ROIs) across the brain. Depicted are the following maps for protocol 1: (a) $SNR^{\mathrm{PDw}}_{\mathrm{raw}}$, (b) $mSNR^{\mathrm{PD}}$, (c) $mSNR^{\mathrm{R1}}$, and (d) $mSNR^{\mathrm{MT}}$. While the $SNR^{\mathrm{PDw}}_{\mathrm{raw}}$ was calculated within each ROI using Eq. (1) per subject, the mSNRs were spatially averaged within each ROI per subject. Then, each of the four different SNR metrics was additionally averaged across the group of healthy subjects on a voxel-by-voxel level. The ROIs were selected using the Oxford-Harvard atlas (Frazier et al., 2005; Desikan et al., 2006; Makris et al., 2006; Goldstein et al., 2007). The 111 ROIs out of 117 were used in which the number of voxels was larger than 100.

Table 4

The average relative standard error of the mean (rSEM).

	Repeat 1	Repeat 2	Robust (RO)
$rSEM^{\mathrm{MT}}_{i}$ in WM	7.1 ± 15.0 %	11.1 ± 14.7 %	-0.8 ± 4.6 %
$rSEM^{\mathrm{PD}}_{i}$ in WM	14.4 ± 19.4 %	9.1 ± 19.5 %	-2.9 ± 9.1 %
$rSEM^{\mathrm{R1}}_{i}$ in WM	14.4 ± 19.3 %	9.1 ± 17.5 %	-2.9 ± 8.0 %
$rSEM^{\mathrm{MT}}_{i}$ in GM	10.5 ± 18.6 %	12.6 ± 17.1 %	-0.7 ± 6.7 %
$rSEM^{\mathrm{PD}}_{i}$ in GM	9.0 ± 15.8 %	6.1 ± 14.6 %	-1.0 ± 8.3 %
$rSEM^{\mathrm{R1}}_{i}$ in GM	18.7 ± 21.0 %	5.3 ± 16.7 %	-4.7 ± 10.6 %

The rSEM is estimated with respect to the SEM of the arithmetic-mean (AM) combined MPM parameters (PD, R_1, and MT_{sat}): $rSEM^{m}_{i} = (\frac{SEM^{m}_{i}}{SEM^{m}_{\mathrm{AM}}} - 1) \times 100$ with $m \in \{\mathrm{PD}, \mathrm{R1}, \mathrm{MT}\}$ and $i \in \{1, \ 2, \ \mathrm{RO}\}$, where (1) and (2) are the respective repeats of the acquisition and RO is the robustly combined parameters. Rows 1–3 show the rSEM in white matter (WM) and rows 4–6 show the rSEM in grey matter (GM).

was low (arrow in Fig. 8a). For GM, the SEM showed higher values toward the outer edge (Supplementary Materials S7, Fig. S7.1), potentially caused by residual inaccuracies in spatial registration.

All three MPM parameters showed, on average across the brain, a higher variability if the SEM was estimated on the basis of a single repeat as compared to the AM combined MPM parameters (Table 4): across contrasts the relative SEM for repeat 1 and 2 was between 5.3% ($rSEM^{\mathrm{R1}}_{(1)}$ in GM) and 18.7% ($rSEM^{\mathrm{R1}}_{(2)}$ in GM). In some specific regions, however, one of the two repeats showed smaller SEM (blue areas in Figs. 8b and S7.1b, first and second column). Across the brain, the robust combination (RO) based MPM parameters showed a lower variability as compared to the AM combined MPM parameters: across contrasts the relative SEM was between -0.7 ($rSEM^{\mathrm{MT}}_{(\mathrm{RO})}$) and -4.7% ($rSEM^{\mathrm{R1}}_{(\mathrm{RO})}$). Again, in some localized regions the corresponding $rSEM$ maps were also positive, indicating a higher variability after robust combination in specific regions (Figs. 8b and S7.1b, third column).

5. Discussion

For three quantitative MPM parameters (R_1, PD, MT_{sat}) we introduced a method to estimate the associated error and model-based signal-to-noise ratio (mSNR) maps without the need to acquire additional data. First, we illustrated that the error and mSNR maps capture the random noise variations associated with instrumental features (e.g. head coil configuration), different protocols (e.g. acceleration factors), as well as noise sources related to artefacts (e.g. subject motion). Second, we used measurements across a group of healthy subjects together with simulations to show that mSNRs also reflect SNR. We found that they were linearly related to raw-image-based SNR and that their slopes varied between MPM parameters: the slope was highest for PD, lower for R_1 and lowest for MT_{sat}. Third, we exploited the artefact-sensitivity of the error maps to generate robust MPM parameters from a two-repeat MPM protocol. We showed that artefactual group variability was reduced in the two-repeat MPM acquisition as compared to the single-repeat MPMs. Importantly, the variability was lowest when using the robust MPM parameters as compared to the arithmetic-mean combination of MPM parameters.

5.1. Error and mSNR

To efficiently capture the errors in the MPM parameters (R_1, PD, and $M\mathrm{T}_{\mathrm{sat}}$) for each individual MPM experiment, we proposed using error

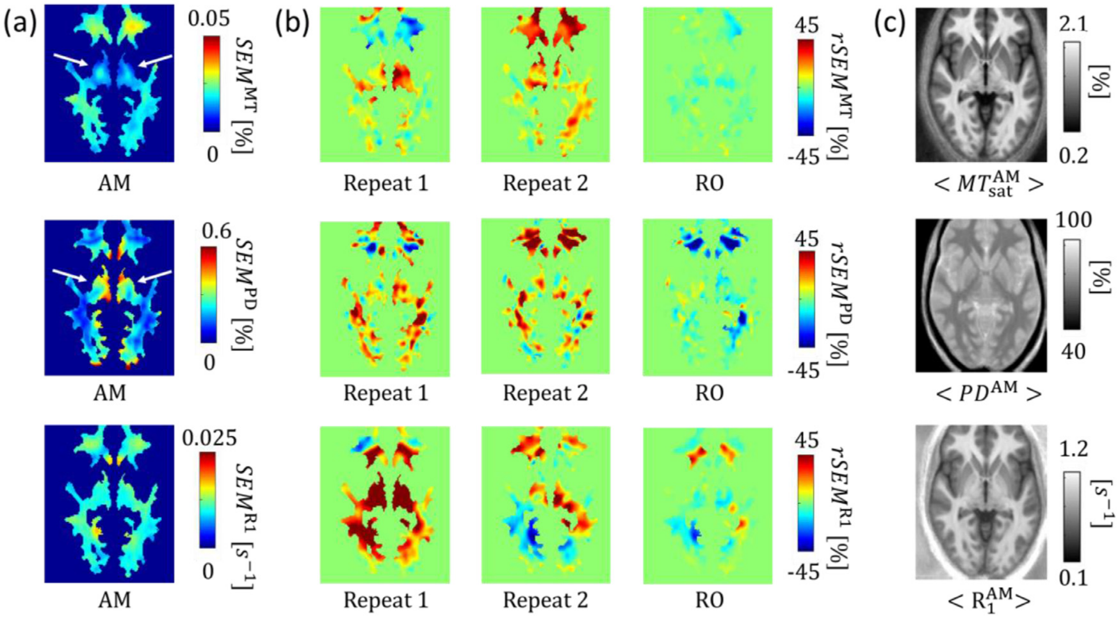


Fig. 8. Group variability across subjects for white matter. The group variability was assessed by the standard-error-of-the-mean (SEM^m) for the quantitative MPM maps MT_{sat} (top row), PD (middle row) and R_1 (bottom row) and is illustrated for white matter ($m \in \{MT, PD, R1\}$). Depicted are (a): SEM^m maps generated from the arithmetic mean (AM) of the two-repeat datasets; (b, from left to right): the relative change of SEM^m (denoted as $rSEM$) for the 1st and 2nd repeat dataset, and their robust combination (RO) using SEM_{AM} as reference; (c): the group-averaged MPM maps. Regions showing reduced SEM are blue and regions showing increased SEM are red. Note that the SEM^m values are not directly comparable, since the scaling of the associated parameters is very different. Note also that the data in (b) might locally exceed the map limits.

propagation of uncorrelated uncertainties and to approximate the noise variance via the contrast-specific uncertainties of the transverse decay for the PD-, T1-, and MT-weighted SGPR signals (Eq. (4)). The error maps capture noise variation due to random noise and noise associated with artefacts such as subject motion.

The random noise sensitivity is best revealed when using the mSNR which is calculated, in analogy to the SNR, as the ratio between the parameter and its error. While the mSNR decreased towards the centre of the brain (Fig. 3a), which is in accordance with the expected decrease in SNR due to the receive field of the head coil (64 ch coil in protocol 1 and 2 and 32 ch coil in protocol 3), it also varied with coil configurations (Fig. S4.1: 32 ch vs. 64 ch) and head position (Fig. S4.2: rotation of a post mortem brain by ~180°). We found that the change of the mSNR from the brain periphery to the centre was different for protocols 1 and 2 although they were using equivalent MR systems (3T and 64 ch head-coil). The most apparent difference in those protocols was the acceleration and the Partial-Fourier (PF) factors: a 3×1 acceleration together with a PF of 6/8 was used in protocol 1 whereas in protocol 2 a 2×2 acceleration and no PF was used. The direction of acceleration in protocol 1 coincided with the direction in which the steeper decline of mSNR values was observed in the respective mSNR maps (left-right direction) as compared to the mSNR acquired with protocol 2. Moreover, the local decrease in mSNR was accompanied by an increase of the contrast-specific uncertainties and errors, meaning that the lower mSNR is driven by a higher noise or artefact level in those regions. One potential reason for the protocol-specific noise-pattern could be an interaction between the g-factor-induced SNR loss (Robson et al., 2008) and Partial Fourier imaging effects. Additionally, the changes in spatial resolution and field strength not only changed the decline of mSNR towards the centre of the brain (protocol 2 showed a steeper decline than protocol 3) but also the averaged mSNR value across the brain.

Noise variations due to artefacts such as subject motion, were well visualized by the error maps but also present in the mSNR. We demonstrated that this second source of variance typically appeared as a local increase in error (and decrease in mSNR) and was in parts accompanied by a bias in the MPM parameters (Figs. 5, S6.1, S6.2, and S6.3). The error depends on the contrast-specific uncertainty, which in turn is estimated per MPM contrast based on the assumption that the measured signal can be explained by a mono-exponential signal decay with TE (see Eq. (4a)–(4c)). Violation of this assumption, e.g., due to a sinc-modulated signal decay caused by susceptibility-related intravoxel dephasing (Hernando et al., 2012) or other deviation from a mono-exponential signal decay due to subject motion (Magerkurth et al., 2011), can increase the uncertainty and lead to a higher error. The latter type of error will most probably vary between repeats whereas the first will not (see examples in Fig. 4). The proposed robust combination of a two-repeat MPM acquisition can only reduce the bias in the MPM parameters for the type of error that varies between repeats (Fig. 5, S6.1, S6.2, and S6.3). It is particularly efficient if, e.g., subject motion is only present in one of the two repeats (Fig. S6.3).

5.2. Raw-image vs. model-based SNR

We found that the mSNR is linearly related to the image-based SNR (Fig. 6) and to the experimental SNR of the MPM parameters (Fig. S3.1). These linear relations confirm that the mSNR is a genuine measure of SNR. Interestingly, the relative slopes between mSNR and image-based SNR showed a similar trend between simulations and measurements: the slope of $mSNR^{\mathrm{PD}}$ was 1.8 to 2 times higher than the slope of $mSNR^{\mathrm{R1}}$, whereas it was 3.2 to 4.5 times higher than for $mSNR^{\mathrm{MT}}$. The relation between $mSNR^{\mathrm{PD}}$ and $mSNR^{\mathrm{R1}}$ will depend on the chosen flip angles and TRs as has previously been shown for SNR in R1 and PD (Helms et al., 2011). A similar argument can be used to understand that $mSNR^{\mathrm{MT}}$ will also be flip angle dependent.

The following application exemplifies how this information can be relevant for large-scale neuroimaging studies, fundamental neuroscience or clinical research studies. If we assume that two groups of subjects possess different myelin densities in the brain (e.g. due to aging Callaghan et al., 2014) and that this difference has the same effect size in PD (via $MTV = 1 - PD/100$, Mezer et al., 2013) and MT_{sat} we would need 3.2–4.5 times higher image SNR to observe the same effect under the influence of noise when using MT_{sat} as a biomarker for myelin instead of PD. Only considering noise, this provides a good argument to skip the MT-weighted contrast and thus to further reduce scan time in time-critical studies.

In contrast to the slope of $mSNR^m$, its offset strongly differed between simulations (almost zero) and measurements (-13 for $mSNR^{\text{PD}}$ to -3 for $mSNR^{\text{MT}}$). One reason for this deviation could be that the mSNR in the experiment was compared between different brain regions with different relaxation rates, proton densities, and MTsat values, each of which might possess a slightly different linear dependency between $mSNR^m$ and $SNR^{\text{PDw}}_{\text{raw}}$ ($m \in \{\text{PD}, \text{R1}, \text{MT}\}$). Our simulations revealed that the slope was consistently smaller in WM than in GM (2% for R_1, 6% for PD and 14% for MT_{sat}). Another reason for the deviation could be the fact that the measured mSNR map is sensitive to both random noise and spatially varying artefacts, e.g., due to parallel imaging, whereas simulations included only random noise.

5.3. Reducing artefactual variation at the group level

The variability at the group level is composed of the true anatomical variability and the artefactual variability induced by different noise sources (e.g. thermal noise, physiological noise, or subject motion). In contrast to the former, the latter type of variability can be reduced by repeated measurements. In accordance with this knowledge, we found that an arithmetic-mean combination of the two repeats reduced the group variability of the MPM parameters (variability reduction: 10.9% in WM and 10.4% in GM) but even more so if the robust combination was used (additional reduction of 2.2% in WM and 2.1% in GM). The additional improvement using robust combination confirms that the reliability of MPM parameters can be further improved when the error map information is exploited to down-weight erroneous MPM values on a per-repeat basis. Overall, we found that the artefactual group variability was more efficiently reduced for the robustly combined R_1 and PD maps than for the robustly combined MT_{sat} maps (Table 4, last column). This observation could be explained by the fact that the mSNR is lowest in MT_{sat} and thus the efficient noise-reduction of the arithmetic-mean combination is almost as relevant as the suppression of outliers in the robust combination. Here, a larger *k*-parameter could help to further improve the rSEM of MT_{sat} (see Supplementary Materials S1 for dependence of *k*-parameter on outlier suppression and noise reduction).

Note, that we focussed here on the direct effects of the two-repeat protocol on the MPM parameters by using the same transformation and tissue segments for all datasets. However, we believe that the higher artefact level in the single-repeat MPM parameter maps will also degrade the segmentation and by consequence the spatial registration, which, in turn, will further degrade the sensitivity to any true group differences.

5.4. Considerations

The two-repeat protocols are longer in scan time than their one-repeat counterparts (e.g. for protocol 1 it is: 17 min vs. 28 min). Since scan time is often the limiting factor, it is important to consider scan time when comparing variability. To do so, here we translate the reduction in variability into an effective increase in sample size, assessed via the standard-error-of-the-mean (SEM). Under the assumption of Gaussian distributed independent data the SEM directly dictates statistical sensitivity (t-score $\propto 1/\text{SEM}$) and scales with one-over the square-root of the sample size N. With these relations in mind, a 13% reduction of SEM^{RO} relative to the SEM of a standard MPM acquisition (SEM^{stand}) would translate to an effective increase of the sample size of 28%. This number is directly proportional to the gain in effective scan time, i.e. cumulative scan-time across all subjects. Since a two-repeat acquisition is required for the robust combination, it comes at the price of an extended MPM protocol that was about 65% longer than the corresponding one-repeat acquisition of protocol 1 (total scan time about: 17 min). In total, the effective scan time of the proposed protocol is about 37% longer than the one-repeat protocol, even if accounting for the improved variability due to robust combination. Thus, the proposed protocol and robust combination might be more useful for specific studies, where a small group of subjects, e.g., patients with a rare disease, are investigated and high-quality data of each subject is of higher priority than scan time, but not for studies where a large number of subjects can be afforded (and poor datasets excluded).

In some regions, we found that the proposed robustly combined MPM parameters showed a higher SEM than the arithmetic mean combination, indicating that the error maps do not always correctly capture erroneous regions. One reason might be that the propagation of uncorrelated errors that was used to generate the error maps relied on the assumption that imaging artefacts were adequately captured by the contrast-specific uncertainties estimated from the linearized SPGR fit. This, however, might not always be valid. For example, the contribution from uncertainties in the B1+ estimate is neglected (i.e. we assume $\frac{dR_1}{dB_1^+} dB_1^+ \approx 0$), but may play an important role as a source of variance in the MPM parameters (Lee et al., 2017), especially for applications at 7T. Another reason for the locally higher rSEM of the robust combination could be the mono-exponential signal model assumption underlying the estimation of the contrast-specific uncertainties in the error maps. Violation of this assumption, e.g. because the signal is better described by a multi-exponential decay associated with multiple compartments, flip-angle dependent R_2^* Milotta et al. (2021), or by a sinc-modulated signal decay due to susceptibility-related intravoxel dephasing, can lead to a bias in the variance estimation and thus in the error. In those voxels, a repeat-dependent motion artefact could manifest in such a way as to reduce the error, e.g., due to interference between a sinc-modulated signal decay and motion artefacts, leading to incorrect up-weighting of the motion-corrupted repeat in the robust combination, which would therefore bias the MPM parameters. Finally, in regions without repeat-dependent motion artefacts, robustly combined MPM parameters will be more noisy, though less biased, than MPM parameters combined with the arithmetic mean approach (see Fig. S1.1c). For those regions, higher rSEM is expected when using the robustly combined MPM parameters than the arithmetic mean combined MPM parameters.

5.5. Generalizability and potential applications

The proposed idea to estimate error maps can in principle be generalised to any quantitative MRI computation method that accommodates multi-echo data. Possibilities include multi-echo MP2RAGE (Metere et al., 2017; Sun et al., 2020) or multi-echo DWI (Eichner et al., 2020).

The knowledge that the mSNR is a genuine measure of SNR can be useful for global and local power analyses. On a local level the mSNR maps might inform studies interested in specific regions regarding which protocol settings, e.g. parallel imaging acceleration or RF head coil, would be optimal. For example, in studies focussing on the hippocampus a less steep decay of mSNR towards the centre of the brain would be preferred, whereas maximal peripheral sensitivity may be preferable for a study interested in the neocortex.

Analysis of the standard deviation of the error maps across the group can reveal additional information about spatially localised artefacts that can potentially alter sensitivity at the group level. For protocol 1, we found a higher standard deviation across the group in the cerebellum for all three MPM parameters and in the posterior region of the brain only for the R_1 parameter (Fig. S8.1 in Supplementary Materials S8).

The error and/or mSNR maps can be directly used for statistical comparisons as confidence measures reflecting variation in SNR and erroneous MPM values due to artefact. This allows use of the error maps to improve the robustness of statistical analyses at the group level without needing a two-repeat MPM acquisition. Further research is necessary to find the best statistical neuroimaging framework (e.g. a linear mixed model) that allows integration of confidence maps at the individual subject level (here the error or mSNR maps) with variability measures at the group level (typically standard-error-of-the mean across subjects).

Alternatively, error and mSNR maps can be used as additional information in group statistics to assess the reliability of observed differences. For example, if statistical significance in voxel-based statistics between two groups is driven by a few outliers, these might be accompanied by particularly high error values. Thus, the error maps can be used to remove or down-weight erroneous regions in group statistics. This application has been shown to improve statistical significance to detect group differences (David et al., 2017).

6. Conclusion

We have introduced a new method to estimate parameter-wise error and model-based SNR (mSNR) maps for three MPM parameters (proton density, PD, longitudinal relaxation rate, R_1, and magnetization transfer saturation, MT_{sat}) on a routine basis without requiring additional data. These new measures can be used to estimate the noise sensitivity of MPM parameters and, if two or more MPM measurements are available, improve their robustness to artefacts such as involuntary subject movement on a per-subject level. The sensitivity to noise might be useful for power-calculations and to compare the suitability of the different MPM parameters as biomarkers in neuroscience or clinical research studies. The improved robustness of MPM parameters might be particularly important in clinical studies where patients with a rare disease are investigated and high data quality is more crucial than high throughput of data. All three advances, the error maps, the mSNR maps, and the robustly combined MPM maps, are available in the open-source hMRI toolbox.

Declaration of Competing Interest

The Max Planck Institute for Human Cognitive and Brain Sciences has an institutional research agreement with Siemens Healthcare. NW holds a patent on acquisition of MRI data during spoiler gradients (US 10,401,453 B2). NW was a speaker at an event organized by Siemens Healthcare and was reimbursed for the travel expenses.

Credit authorship contribution statement

Siawoosh Mohammadi: Conceptualization, Formal analysis, Funding acquisition, Data curation, Investigation, Methodology, Software, Visualization, Writing – original draft. **Tobias Streubel:** Data curation, Writing – review & editing. **Leonie Klock:** Data curation, Writing – review & editing. **Luke J. Edwards:** Software, Writing – review & editing. **Antoine Lutti:** Writing – review & editing. **Kerrin J. Pine:** Data curation, Writing – review & editing. **Sandra Weber:** Data curation, Writing – review & editing. **Patrick Scheibe:** Writing – review & editing. **Gabriel Ziegler:** Writing – review & editing. **Jürgen Gallinat:** Resources, Writing – review & editing. **Simone Kühn:** Funding acquisition, Resources, Writing – review & editing. **Martina F. Callaghan:** Funding acquisition, Conceptualization, Writing – review & editing, Supervision. **Nikolaus Weiskopf:** Funding acquisition, Conceptualization, Writing – review & editing, Supervision. **Karsten Tabelow:** Conceptualization, Methodology, Software, Writing – review & editing, Supervision.

Data availability

The authors do not have permission to share data.

Code availability

The code to generate error and SNR maps, as well as robust MTsat, R1, and PD maps as they were used in this paper is available within the hMRI toolbox here: https://github.com/hMRI-group/hMRI-toolbox/releases/tag/errormaps.

Acknowledgements

We would like to thank Saskia Helbling and Peter McColgan for sharing the 7T dataset originally published in (McColgan et al., 2021). We would like to thank Francisco J. Fritz for acquiring the post-mortem data. This project was funded by the ERA-NET NEURON (hMRIofSCI), the Federal Ministry of Education and Research (BMBF; 01EW1711A and B), and the German Research Foundation (DFG Priority Program 2041 “Computational Connectomics”, MO 2397/5-1, DFG Emmy Noether Stipend: MO 2397/4-1), and the Forschungszentrums Medizintechnik Hamburg (fmthh; grant 01fmthh2017). The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 616905. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the grant agreement No 681094, and is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0137. SK has been funded by the Max Planck Society, two grants from the German Research Foundation (TRR 169/C8, SFB 936/C7) and the European Union (ERC-2016-StG-Self-Control-677804, Baltic Interreg Programme: Baltic Game Industry). MFC is supported by the MRC and Spinal Research Charity through the ERA-NET Neuron joint call (MR/R000050/1). The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome [203147/Z/16/Z]. AL is supported by the Swiss National Science Foundation (grant no 320030_184784) and the Fondation ROGER DE SPOELBERCH.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119529.

References

Ackenheil, M., Stotz-Ingenlath, G., Dietz-Bauer, R., Vossen, A., 1999. MINI Mini International Neuropsychiatric Interview, German Version 5.0.0. DSM IV. Psychiatrische Universitätsklinik München, Germany.

Ashburner, J., Friston, K.J., 2011. Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. Neuroimage 55, 954–967.

Callaghan, M.F., Freund, P., Draganski, B., Anderson, E., Cappelletti, M., Chowdhury, R., Diedrichsen, J., FitzGerald, T.H.B., Smittenaar, P., Helms, G., Lutti, A., Weiskopf, N., 2014. Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. Neurobiol. Aging 35, 1862–1872.

Callaghan, M.F., Josephs, O., Herbst, M., Zaitsev, M., Todd, N., Weiskopf, N., 2015. An evaluation of prospective motion correction (PMC) for high resolution quantitative MRI. Front. Neurosci. 9, 97.

Castella, R., Arn, L., Dupuis, E., Callaghan, M.F., Draganski, B., Lutti, A., 2018. Controlling motion artefact levels in MR images by suspending data acquisition during periods of head motion. Magn. Reson. Med. 80, 2415–2426.

Cercignani, M., Bouyagoub, S., 2018. Brain microstructure by multi-modal MRI: is the whole greater than the sum of its parts? NeuroImage 182, 117–127.

Clark, I.A., Callaghan, M.F., Weiskopf, N., Maguire, E.A., 2021. The relationship between hippocampal-dependent task performance and hippocampal grey matter myelination and iron content. Brain Neurosci. Adv. 5 23982128211011923.

David, G., Freund, P., Mohammadi, S., 2017. The efficiency of retrospective artifact correction methods in improving the statistical power of between-group differences in spinal cord DTI. NeuroImage 158, 296–307.

David, G., Seif, M., Huber, E., Hupp, M., Rosner, J., Dietz, V., Weiskopf, N., Mohammadi, S., Freund, P., 2019. *In vivo* evidence of remote neural degeneration in the lumbar enlargement after cervical injury. Neurology 92, e1367–e1377.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980.

Dietrich, O., Raya, J.G., Reeder, S.B., Reiser, M.F., Schoenberg, S.O., 2007. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. J. Magn. Reson. Imaging 26, 375–385.

Edwards, L.J., Kirilina, E., Mohammadi, S., Weiskopf, N., 2018. Microstructural imaging of human neocortex *in vivo*. NeuroImage 182, 184–206.

Eichner, C., Paquette, M., Mildner, T., Schlumm, T., Pléh, K., Samuni, L., Crockford, C., Wittig, R.M., Jäger, C., Möller, H.E., Friederici, A.D., Anwander, A., 2020. Increased sensitivity and signal-to-noise ratio in diffusion-weighted MRI using multi-echo acquisitions. Neuroimage 221, 117172.

Ellerbrock, I., Mohammadi, S., 2018. Four *in vivo* g-ratio-weighted imaging methods: comparability and repeatability at the group level. Hum. Brain Mapp. 39, 24–41.

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

Frazier, J.A., Chiu, S., Breeze, J.L., Makris, N., Lange, N., Kennedy, D.N., Herbert, M.R., Bent, E.K., Koneru, V.K., Dieterich, M.E., Hodge, S.M., Rauch, S.L., Grant, P.E., Cohen, B.M., Seidman, L.J., Caviness, V.S., Biederman, J., 2005. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. Am. J. Psychiatry 162, 1256–1265.

Freund, P., Weiskopf, N., Ashburner, J., Wolf, K., Sutter, R., Altmann, D.R., Friston, K., Thompson, A., Curt, A., 2013. MRI investigation of the sensorimotor cortex and the corticospinal tract after acute spinal cord injury: a prospective longitudinal study. Lancet Neurol. 12, 873–881.

Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D., 2006. Statistical Parametric Mapping: The Analysis of Functional Brain Images, 1st ed. Academic Press, London.

Goldstein, J.M., Seidman, L.J., Makris, N., Ahern, T., O'Brien, L.M., Caviness, V.S., Kennedy, D.N., Faraone, S.V., Tsuang, M.T., 2007. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. Biol. Psychiatry 61, 935–945.

Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn. Reson. Med. 47, 1202–1210.

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

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

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, 669–677.

Helms G (2015) Correction for residual effects of B1+ inhomogeniety on MT saturation in FLASH-based multi-parameter mapping of the brain. In: Proc Intl Soc Magn Reson Med., Canada, Toronto 2015;23:3360.

Helms G, Weiskopf N, Lutti A (2021) Correction of FLASH-based MT saturation in human brain for residual bias of B1-inhomogeneity at 3T. arXiv:210414878 [physics] Available at: http://arxiv.org/abs/2104.14878 [Accessed May 25, 2021].

Hernando, D., Vigen, K.K., Shimakawa, A., Reeder, S.B., 2012. R^*2 mapping in the presence of macroscopic B0 field variations. Magn. Reson. Med. 68, 830–840.

Kirilina, E., Helbling, S., Morawski, M., Pine, K., Reimann, K., Jankuhn, S., Dinse, J., Deistung, A., Reichenbach, J.R., Trampel, R., Geyer, S., Müller, L., Jakubowski, N., Arendt, T., Bazin, P.L., Weiskopf, N., 2020. Superficial white matter imaging: contrast mechanisms and whole-brain *in vivo* mapping. Sci. Adv. 6, eaaz9281.

Lee, Y., Callaghan, M.F., Nagy, Z., 2017. Analysis of the precision of variable flip angle T1 mapping with emphasis on the noise propagated from RF transmit field maps. Front. Neurosci. 11 (106). doi:10.3389/fnins.2017.00106.

Leutritz, T., Seif, M., Helms, G., Samson, R.S., Curt, A., Freund, P., Weiskopf, N., 2020. Multiparameter mapping of relaxation (R1, $R2^*$), proton density and magnetization transfer saturation at 3 T: a multicenter dual-vendor reproducibility and repeatability study. Hum. Brain Mapp. 41, 4232–4247.

Lutti, A, Corbin, N, Ashburner, J, Ziegler, G, Draganski, B, Phillips, C, Kherif, F, Callaghan, MF, Di, Domenicantonio G, 2022. Restoring statistical validity in group analyses of motion-corrupted MRI data. Hum Brain Mapp 43, 1973–1983.

Lutti, A., Hutton, C., Finsterbusch, J., Helms, G., Weiskopf, N., 2010. Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T. Magn. Reson. Med. 64, 229–238.

Lutti, A., Stadler, J., Josephs, O., Windischberger, C., Speck, O., Bernarding, J., Hutton, C., Weiskopf, N., 2012. Robust and fast whole brain mapping of the RF transmit field B1 at 7T. PLoS ONE 7, e32379.

Magerkurth, J., Volz, S., Wagner, M., Jurcoane, A., Anti, S., Seiler, A., Hattingen, E., Deichmann, R., 2011. Quantitative T^*2-mapping based on multi-slice multiple gradient echo flash imaging: retrospective correction for subject motion effects. Magn. Reson. Med. 66, 989–997.

Makris, N., Goldstein, J.M., Kennedy, D., Hodge, S.M., Caviness, V.S., Faraone, S.V., Tsuang, M.T., Seidman, L.J., 2006. Decreased volume of left and total anterior insular lobule in schizophrenia. Schizophr. Res. 83, 155–171.

McColgan, P., Helbling, S., Vaculčiaková, L., Pine, K., Wagstyl, K., Attar, F.M., Edwards, L., Papoutsi, M., Wei, Y., Van den Heuvel, M.P., Tabrizi, S.J., Rees, G., Weiskopf, N., 2021. Relating quantitative 7T MRI across cortical depths to cytoarchitectonics, gene expression and connectomics. Hum. Brain Mapp. 42, 4996–5009.

Metere, R., Kober, T., Möller, H.E., Schäfer, A., 2017. Simultaneous quantitative MRI Mapping of T1, $T2^*$ and magnetic susceptibility with multi-echo MP2RAGE. PLOS One 12, e0169265.

Mezer, A., Yeatman, J.D., Stikov, N., Kay, K.N., Cho, N.J., Dougherty, R.F., Perry, M.L., Parvizi, J., Hua, L.H., Butts-Pauly, K., Wandell, B.A., 2013. Quantifying the local tissue volume and composition in individual brains with magnetic resonance imaging. Nat. Med. 19, 1667–1672.

Milotta, G., Nadege, C., Lambert, C., Lutti, A., Mohammadi, S., Callaghan (n.d.), M., 2021. Characterisation of the flip angle dependence of $R2^*$ in multi-parameter mapping. Proc. Intl. Soc. Magn. Reson. Med. 29, 3060.

Mohammadi, S., Callaghan, M.F., 2021. Towards *in vivo* g-ratio mapping using MRI: unifying myelin and diffusion imaging. J. Neurosci. Methods 348, 108990.

Papoutsi, M., Magerkurth, J., Josephs, O., Pépés, S.E., Ibitoye, T., Reilmann, R., Hunt, N., Payne, E., Weiskopf, N., Langbehn, D., Rees, G., Tabrizi, S.J., 2020. Activity or connectivity? A randomized controlled feasibility study evaluating neurofeedback training in Huntington's disease. Brain Commun. 2 (1), fcaa049. doi:10.1093/braincomms/fcaa049.

Paus, T., Zijdenbos, A., Worsley, K., Collins, D.L., Blumenthal, J., Giedd, J.N., Rapoport, J.L., Evans, A.C., 1999. Structural maturation of neural pathways in children and adolescents: *in vivo* study. Science 283, 1908–1911.

Price, R.R., Axel, L., Morgan, T., Newman, R., Perman, W., Schneiders, N., Selikson, M., Wood, M., Thomas, S.R., 1990. Quality assurance methods and phantoms for magnetic resonance imaging: report of AAPM nuclear magnetic resonance Task Group No. 1. Med. Phys. 17, 287–295.

Reeder, S.B., Wintersperger, B.J., Dietrich, O., Lanz, T., Greiser, A., Reiser, M.F., Glazer, G.M., Schoenberg, S.O., 2005. Practical approaches to the evaluation of signal-to-noise ratio performance with parallel imaging: application with cardiac imaging and a 32-channel cardiac coil. Magn. Reson. Med. 54, 748–754.

Robson, P.M., Grant, A.K., Madhuranthakam, A.J., Lattanzi, R., Sodickson, D.K., McKenzie, C.A., 2008. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. Magn. Reson. Med. 60, 895–907.

Schnack, H.G., van Haren, N.E.M., Brouwer, R.M., van Baal, G.C.M., Picchioni, M., Weisbrod, M., Sauer, H., Cannon, T.D., Huttunen, M., Lepage, C., Collins, D.L., Evans, A., Murray, R.M., Kahn, R.S., Hulshoff Pol, H.E., 2010. Mapping reliability in multicenter MRI: voxel-based morphometry and cortical thickness. Hum. Brain Mapp. 31, 1967–1982.

Sun, H., Cleary, J.O., Glarin, R., Kolbe, S.C., Ordidge, R.J., Moffat, B.A., Pike, G.B., 2020. Extracting more for less: multi-echo MP2RAGE for simultaneous T1 -weighted imaging, T1 mapping, R 2 ∗ mapping, SWI, and QSM from a single acquisition. Magn. Reson. Med. 83, 1178–1191.

Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M.F., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Lutti, A., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Weiskopf, N., Ziegler, G., Mohammadi, S., 2019. hMRI – a toolbox for quantitative MRI in neuroscience and clinical research. NeuroImage 194, 191–210.

Taubert, M., Roggenhofer, E., Melie-Garcia, L., Muller, S., Lehmann, N., Preisig, M., Vollenweider, P., Marques-Vidal, P., Lutti, A., Kherif, F., Draganski, B., 2020. Converging patterns of aging-associated brain volume loss and tissue microstructure differences. Neurobiol. Aging 88, 108–118.

Weiskopf, N., Callaghan, M.F., Josephs, O., Lutti, A., Mohammadi, S., 2014. Estimating the apparent transverse relaxation time ($R2^*$) from images with different contrasts (ESTATICS) reduces motion artifacts. Front. Neurosci. 8, 278.

Weiskopf, N., Edwards, L.J., Helms, G., Mohammadi, S., Kirilina, E., 2021. Quantitative magnetic resonance imaging of brain anatomy and *in vivo* histology. Nat. Rev. Phys. 3, 570–588.

Weiskopf, N., Mohammadi, S., Lutti, A., Callaghan, M.F., 2015. Advances in MRI-based computational neuroanatomy: from morphometry to in-vivo histology. Curr. Opini. Neurol. 28, 313–322.

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. 7, 95.

West, K.L., Kelm, N.D., Carson, R.P., Gochberg, D.F., Ess, K.C., Does, M.D., 2018. Myelin volume fraction imaging with MRI. Neuroimage 182, 511–521.

Whitaker, K.J., et al., 2016. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. PNAS 113, 9105–9110.

Ziegler, G., Hauser, T.U., Moutoussis, M., Bullmore, E.T., Goodyer, I.M., Fonagy, P., Jones, P.B., Consortium, N.S.P.N., Lindenberger, U., Dolan, R.J., 2019. Compulsivity and impulsivity traits linked to attenuated developmental frontostriatal myelination trajectories. Nat. Neurosci. 22, 992–999.