Statistical analyses of motion-corrupted MRI relaxometry data

Running title: Analysis of motion-corrupted MRI data

Nadège Corbin[1,2], Rita Oliveira[3], Quentin Raynaud[3], Giulia Di Domenicantonio[3], Bogdan Draganski[3,4], Ferath Kherif[3], Martina F. Callaghan[2], Antoine Lutti[3]

[1]Centre de Résonance Magnétique des Systèmes Biologiques, UMR5536, CNRS/University Bordeaux, Bordeaux, France

[2]Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, University College London, London, UK

[3]Laboratory for Research in Neuroimaging, Department of Clinical Neurosciences, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland.

[4]Neurology Department, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Code availability The analysis code used in this study is available here [Oliveira et al., 2023]:

https://zenodo.org/record/7612032#.Y-OwlXbMKUl

The presented method is also available in the hMRI-toolbox:

https://hmri-group.github.io/hMRI-toolbox/

Data availability Material related to this paper is available here [Raynaud et al., 2023]:

https://zenodo.org/record/7692074#.ZACk9HaZOUk

This material includes maps of $R2^*_{(2)}$, $R2^*_{(3)}$, R1 and MTsat from 123 participants, and the results of the analyses presented in this manuscript conducted on this data.

Funding statement This work was supported by the Swiss National Science Foundation (grant no 320030_184784 (AL)), the Partridge foundation and the ROGER DE SPOELBERCH Foundation. The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome [203147/Z/16/Z]. BD is supported by the Swiss National Science Foundation (NCCR Synapsy, project grant numbers 32003B_135679, 32003B_159780, 324730_192755 and CRSK-3_190185) and the Fondation Leenaards. The MRI data were acquired on the MRI platform of the Clinical Neuroscience Department, Lausanne University Hospital.

Conflict of interests disclosure The authors declare no potential conflict of interest

Ethics approval statement This study received approval from the local Ethics Committee and all participants gave their written informed consent prior to participation.

Abstract

Consistent noise variance across data points (i.e. homoscedasticity) is required to ensure the validity of statistical analyses of MRI data conducted using linear regression methods. However, head motion leads to degradation of image quality, introducing noise heteroscedasticity into ordinary-least square analyses. The recently introduced QUIQI method restores noise homoscedasticity by means of weighted least square analyses in which the weights, specific for each dataset of an analysis, are computed from an index of motion-induced image quality degradation. QUIQI was first demonstrated in the context of brain maps of the MRI parameter R2*, which were computed from a single set of images with variable echo time. Here, we extend this framework to quantitative maps of the MRI parameters R1, R2*, and MTsat, which are computed from multiple sets of images. QUIQI allows for optimization of the noise model by using metrics quantifying heteroscedasticity and free energy. QUIQI restores homoscedasticity more effectively than insertion of an image quality index in the analysis design and yields higher sensitivity than simply removing the datasets most corrupted by head motion from the analysis. In sum, QUIQI provides an optimal approach to group-wise analyses of a range of quantitative MRI parameter maps that is robust to inherent homoscedasticity.

1. Introduction

The use of linear regression methods for the analysis of brain MRI data is widespread in neuroscience and neurological studies. Linear regression methods model MRI data as a linear combination of explanatory variables that may pertain to disease evolution [Beveridge et al., 2018; Ong et al., 2021; Panda et al., 2019; van der Plas et al., 2021; Scott et al., 2003], treatment [Plaikner et al., 2018], environmental factors [Hu et al., 2022; Ong et al., 2022] or phenotypes [Boots et al., 2020; Honigberg et al., 2020; Papadaki et al., 2019]. To estimate the coefficients of the combination, linear regression methods assume uncorrelated noise across measurements, sampled from normal distributions with equal variances (i.e. *homoscedasticity*). The assumption of homoscedasticity must be satisfied to ensure the validity of statistical inferences arising from the analyses of the relationship between explanatory variables and MRI data, e.g. using Student T-tests or F-tests [Hayes and Cai, 2007].

Head motion during data acquisition degrades the quality of MR images and affects brain feature estimates computed from the data [Castella et al., 2018; Esteban et al., 2017; Mortamet et al., 2009; Reuter et al., 2015; Rosen et al., 2018; Savalia et al., 2017]. In particular, motion degradation impacts the noise level of relaxometry estimates computed from raw MRI data [Castella et al., 2018]. Heterogeneous degrees of motion degradation across cohorts leads to the homoscedasticity assumption in Ordinary Least Square (OLS) analyses being violated [Lutti et al., 2022]. One solution consists of excluding the data most affected by head motion, identified from an image-based index of data quality ('*Motion Degradation Index*', MDI) [Castella et al., 2018; Esteban et al., 2017; Mortamet et al., 2009; Pizarro et al., 2016; Reuter et al., 2015; Rosen et al., 2018; Savalia et al., 2017]. However, determining a suitable threshold value for the index to exclude datasets of poor quality is challenging [Gilmore et al., 2021]. As a result, the resulting decrease in cohort size may lead to a sub-optimal reduction in analysis sensitivity.

The Weighted Least Square (WLS) alternative consists of including the estimate of the variance in each dataset as weights into the linear regression model. WLS analyses constitute an optimal solution because they guarantee residual homoscedasticity and preserve sensitivity. However the variance, i.e. the noise level in each individual dataset, is generally unknown. The recently introduced QUIQI method [Lutti et al., 2022] allows the variance to be estimated from the value of the MDI for each dataset through use of Restricted Maximum Likelihood (REML, [Friston et al., 2002b]). QUIQI was shown to be more effective at ensuring residual homoscedasticity and at optimizing analysis sensitivity than simple exclusion of the datasets with high MDI values.

The QUIQI method was first demonstrated for the analysis of brain maps of the transverse relaxation rate R2*, computed from a single set of raw image volumes acquired on the MRI scanner with variable echo time (TE) [Lutti et al., 2022]. R2* is primarily a correlate of iron content [Fukunaga et al., 2010; Stüber et al., 2014; Yao et al., 2009], but with sensitivity to other tissue metrics, notably myelin and water content [Bagnato et al., 2018; Hametner et al., 2018]. To disentangle the likely drivers of

parameter change or difference, neuroscience studies commonly combine the analysis of multiple, complementary MRI parameters ('*Multi-Parameter Mapping*', [Callaghan et al., 2014; Carey et al., 2018; Schall et al., 2020]). For example, the MRI parameters R1 and MTsat show increased sensitivity to myelin content within brain tissue [Henkelman et al., 2001; Lutti et al., 2014; Sereno et al., 2013]. However, the estimation of R1 and MTsat requires several sets of raw MR images with different contrasts, and the motion level may vary across these images. Also, each MRI parameter is computed from the raw images using specific signal models [Helms et al., 2008a; Helms et al., 2008b], and is differentially sensitive to head motion as a result [Balbastre et al., 2022; Mohammadi et al., 2022]. The applicability of the QUIQI method to analyses of different types of MRI parameter maps remains to be demonstrated.

In this work, we investigate noise heteroscedasticity induced by head motion in OLS analyses of quantitative relaxometry MRI data (qMRI) computed from several sets of images. Using the global and local metrics of homoscedasticity introduced in [Lutti et al., 2022], we assess the ability of the QUIQI method to ensure the validity of statistical analyses of data degraded by motion. Multiple models of the relationship between noise estimates and indices of motion-induced image degradation are considered. We identify the optimal noise model from the global and local metrics of homoscedasticity, combined with the free energy estimates provided by REML. We compare the sensitivity of the WLS statistical analyses conducted with QUIQI with that of standard OLS analyses after exclusion of the datasets most affected by motion degradation. Finally, we compare WLS analyses conducted with QUIQI with an alternative approach based on inserting the MDI into the design matrix of OLS analyses.

2. Methods

2.1 MRI acquisition

MRI data was acquired in a large cohort of 1432 healthy research participants as part of the 'BrainLaus' study (https://www.colaus-psycolaus.ch/professionals/brainlaus/ [Loued-Khenissi et al., 2022; Trofimova et al., 2021]). The acquisition protocol included multi-echo T1-weighted (T1w), Proton Density-weighted (PDw) and Magnetization Transfer-weighted (MTw) scans, conducted with a custom-made 3D FLASH sequence. B1 mapping data were also acquired to correct for the effect of transmit field inhomogeneity on the qMRI maps [Lutti et al., 2010; Lutti et al., 2012]. Relevant acquisition parameters are available in [Lutti et al., 2022]. The total acquisition time was 27 minutes.

2.2 Map computation and processing

Computation and spatial processing of the qMRI maps were performed offline with the hMRI toolbox [Tabelow et al., 2019], as described in detail below. Image analysis was conducted using the SPM software (www.fil.ion.ucl.ac.uk/spm, Wellcome Centre for Human Neuroimaging) and customized scripts written in Matlab (R2021a, The MathWorks Inc., Natick, MA, USA).

2.2.1. *Computation of the qMRI maps*

Maps of the MRI parameters R1, MTsat and R2* were computed from the raw T1w, PDw and MTw images of each participant. The R2* maps were computed using the ESTATICS approach from the regression of the log signal of the raw images with respect to their echo times [Weiskopf et al., 2014]. To investigate the effect of the number of raw images involved in the map calculation on homoscedasticity, maps of R2* were computed from the T1w and PDw images only ('$R2^*_{(2)}$ maps') and from the T1w, PDw and MTw images ('$R2^*_{(3)}$') (results from R2* maps computed from one set of images only ('$R2^*_{(1)}$') are described in [Lutti et al., 2022]).

Maps of R1 were computed from the T1w and PDw acquisitions using the rational approximation of the Ernst equation [Helms et al., 2008a]. MTsat maps were estimated using the T1w, PDw, and MTw acquisitions using a closed form expression of the MT weighted signal [Helms et al., 2008b]. In total, the performance of QUIQI was assessed on four types of quantitative maps: $R2^*_{(2)}$, $R2^*_{(3)}$, R1 and MTsat maps.

2.2.2. Normalization and segmentation

Complete details of the image processing steps are available in [Lutti et al., 2022]. The MTsat maps were used for spatial normalisation of the data into the Montreal Neurological Institute (MNI) template space. The MTsat maps were segmented into maps of grey and white matter probabilities using Unified Segmentation [Ashburner and Friston, 2005]. The nonlinear diffeomorphic algorithm Dartel [Ashburner, 2007] was used for inter-subject registration of the tissue classes. To preserve the quantitative estimates, the qMRI maps were normalized into the MNI space following the voxel-based quantification procedure proposed in [Draganski et al., 2011].

2.2.3. Motion Degradation Index

The MDI used for QUIQI analysis is computed by the hMRI toolbox [Tabelow et al., 2019]. This MDI was introduced in a validation study against the history of head motion that occurred during data acquisition [Castella et al., 2018]. It is calculated as the standard deviation across white matter of R2* maps computed separately from each set of multi-echo T1w, PDw and MTw images: each set of raw images was assigned a specific value of the MDI.

2.3 Inserting a motion degradation index into image analysis

QUIQI ensures noise homoscedasticity in analyses of MRI data affected by motion degradation by means of weighted least square analyses (WLS). The weights, specific to each dataset of an analysis, are computed from a noise covariance matrix (V) modelled as a linear combination of matrices built from the MDI values of the analysis data ('basis functions').

From the heuristic analysis of the empirical relationship between analysis residuals and the MDI [Lutti et al., 2022], these basis functions were set to contain powers of the MDI values:

$$V_i = \sum_{n=1}^{N_{raw}} \sum_{\alpha=0}^{\alpha_{max}} \lambda_{\alpha,n} MDI_{n,i}^{\alpha} \qquad (1)$$

where V_i is the ith diagonal element of the covariance matrix V, i.e. the noise estimate of the ith dataset. $MDI_{n,i}^{\alpha}$ is the αth power of the MDI of the ith dataset and the nth contrast weighting. The basis functions and the resulting noise covariance matrices V were diagonal because noise is uncorrelated between datasets (i.e. participants).

Estimation of V involved the MDI of each raw image involved in the computation of the qMRI data to be analysed (index n in eq.1), and separate basis functions were computed from each MDI. *Nraw* is the number of raw images involved in this computation: *Nraw* =2 for R2*$_{(2)}$ and R1 maps as these data were estimated from the PDw and T1w raw images. *Nraw*=3 for MTsat and R2*$_{(3)}$ maps as these data were estimated from the PDw, T1w and MTw raw images (see section 2.2.1).

The optimal value of α_{max}, the maximum power of the MDI used to compute the noise covariance matrix, was determined in a model comparison (see section 2.5.1).

V and $\lambda_{\alpha,n}$ were estimated using the Restricted Maximum Likelihood (REML) algorithm as implemented in SPM12 [Friston et al., 2002a]. The subsequent estimation of the weights ($W = \frac{1}{\sqrt{V}}$) for WLS analyses was conducted with the standard analysis tools of SPM12.

2.4 Measures of analysis validity

The results of OLS and WLS analyses were compared using local and global metrics of heteroscedasticity, and a measure of free energy. These metrics were also used to identify the optimal model of the dependence of the noise in qMRI maps on image quality.

2.4.1 Residual heteroscedasticity

We characterized noise heteroscedasticity from the maps of the residuals ϵ, computed for each participant and qMRI maps after estimation of the coefficients of the linear model.

Global heteroscedasticity: We computed a global noise index $var(\epsilon)$, as the spatial variance of the residual maps across a whole tissue type (i.e. grey or white matter) [Lutti et al., 2022]. Consistent with Eq 1., the effect of motion degradation on the global noise index was modelled as a cubic polynomial combination of the MDI of all the raw images involved in the computation of the analysis data. For

R2*$_{(2)}$ and R1 maps, the MDI of the T1w and PDw raw images were considered. For R2*$_{(3)}$ and MTsat maps, the MDI of the T1w, PDw and MTw raw images were considered. Fitting the global noise index with the assumed polynomial combination of the MDI led to the estimation of $var(\epsilon)_{fit}$, the modelled dependence of the global noise level on motion degradation.

The global heteroscedasticity index was taken as the coefficient of determination $R^2 = 1 - \frac{\sum_{i=1}^{N} (var(\epsilon)_i - var(\epsilon)_{fit_i})^2}{\sum_{i=1}^{N} var(\epsilon)^2}$. A high value of R^2 (~1) indicates a strong dependence of the global noise index on motion degradation and high heteroscedasticity. Conversely, a low value of R^2 (~0) indicates a weak dependence of the global noise index on motion degradation and low noise heteroscedasticity.

A graphical rendering of heteroscedasticity is achieved by plotting $var(\epsilon)$ against $var(\epsilon)_{fit}$: a linear relationship between $var(\epsilon)$ and $var(\epsilon)_{fit}$ reflects high heteroscedasticity. The absence of relationship between $var(\epsilon)$ and $var(\epsilon)_{fit}$ reflects low heteroscedasticity.

Local heteroscedasticity: At the local scale, we assessed noise heteroscedasticity in each image voxel using the Engle's Arch test [Engle, 1982], which tests for no linear relationship between consecutive samples of a series of squared residuals. At a given voxel, the residual series was computed from the residual maps at this voxel location. The series samples were organized according to the predicted impact of motion degradation by arranging them in ascending order of the polynomial combination of MDIs ($var(\epsilon)_{fit}$), associated to the tissue class of the given voxel.

The Engle's Arch test was conducted at each voxel of the residual maps, with a maximum lag of 40 points. The local index of heteroscedasticity was the proportion of voxels with significant heteroscedasticity (i.e. rejecting the null hypothesis), calculated after false discovery rate correction using the Benjamini-Hochberg procedure ($p<0.05$; [Glickman et al., 2014]).

2.4.2 Free energy

REML estimates the hyperparameters λ that are used to compute the matrix V (Eq.1) by maximizing the evidence lower bound objective (ELBO) function, a measure of negative variational free energy. Intuitively, the ELBO favours the accuracy and penalizes the complexity of the model. We used the ELBO estimates provided by the REML implementation in SPM12 [Friston et al., 2002a] to identify the optimal model of the relationship between noise in the qMRI maps and the degradation of the raw images induced by head motion (see section 2.5.1).

2.5 Image analyses

Linear regression analysis was conducted on qMRI maps of R2*$_{(2)}$, R2*$_{(3)}$, R1 and MTsat. This allowed the assessment of noise heteroscedasticity in WLS and OLS analyses, separately for different MRI parameters and different numbers of raw image types involved in their estimation. The quantitative MRI data were modelled as the linear combination of 5 regressors that represented age, square and cubic values of age, gender and brain volumes.

2.5.1 Noise model comparison

The coefficients of the general linear model were estimated from the full dataset (N=1432). OLS analyses were conducted with α_{max}= 0 in Eq.1, i.e. assuming uniform noise level across all datasets as would be done in a standard analysis. WLS analyses were conducted with 5 models of the noise covariance matrix: α_{max}=2 to 5 with $\lambda \in \mathbb{R}$, and α_{max}=4 with $\lambda \in \mathbb{R}^+$. The latter model was considered to assess the effect of enforcing positivity of the λ hyperparameters on noise heteroscedasticity and free energy, and for consistency with [Lutti et al., 2022].

OLS and WLS analyses were compared by considering differences in heteroscedasticity and free energy. Across the WLS analyses, the optimal noise model was identified as that which led to the lowest values of heteroscedasticity and the largest increase in ELBO.

2.5.2 Age sensitivity

We compared the sensitivity of WLS analyses against that of standard OLS analyses using statistical F-tests of age-related differences in $R2^*_{(2)}$, $R2^*_{(3)}$, R_1, and MTsat. The analyses were performed on a subsample of the full dataset, obtained by randomly selecting up to 10 images per age bin of 5 years. The size of the resulting dataset was 123, close to the typical cohort size of similar studies [Callaghan et al., 2014].

The WLS analyses were conducted with the optimal noise model (see section 2.5.1). OLS analyses were conducted after exclusion of the datasets most degraded by head motion, i.e. with the highest MDI values averaged across the raw images involved in the calculation of the qMRI maps. Exclusion of 3, 7, 13, 20 and 30% of the most degraded datasets was considered [Castella et al., 2018; Esteban et al., 2017; Mortamet et al., 2009; Pizarro et al., 2016; Reuter et al., 2015; Rosen et al., 2018; Savalia et al., 2017]. The optimal fraction of excluded datasets was identified as the smallest value leading to local and global heteroscedasticity estimates comparable to those obtained in WLS analyses with the optimal noise model.

2.5.3 Specificity

To assess the specificity of the OLS and WLS analyses, we recorded the rate of false positives in two types of analyses frequently conducted in neuroscience studies. I) In a subset of the full dataset with up to 10 images per age bin of 5 years (N = 123), the participants' age was randomly assigned from a uniform distribution ranging from the minimum to the maximum age of the data subsets and statistical F-tests of age-related differences in $R2^*_{(2)}$, $R2^*_{(3)}$, R_1 and MTsat maps. II) In a subset of the full dataset within a narrow age range (56–58 y.o.; N = 129), we conducted unbalanced comparisons of a group of 10 qMRI maps with a group of 119 maps, using two-sample T-tests. Given that no age-related effect would be expected in these analyses, any significant effects were deemed to be false positives.

We conducted analyses I and II 1,000 times for different subsets of data and monitored the rate of significant results, i.e. false positives, across repetitions at the cluster level ($p < 0.05$, FWE-corrected), with a cluster forming threshold of $p < 0.001$ uncorrected.

2.5.4 Inserting the Motion Degradation Index in the analysis design

In a subset of data in a narrow age range (56–58 y.o.; N = 129), we considered an alternative method to QUIQI to correct motion degradation effects, which consists of inserting the MDI as a confounding factor in the analysis design. Here, powers of 1 to 4 of the MDI values were included as regressors in the design matrix. We compared noise heteroscedasticity between OLS analyses and WLS analyses with the optimal noise model ($\alpha_{max} = 4, \lambda \in \mathbb{R}$, identified from 2.5.1).

3. Results

3.1 Noise model comparison

With standard OLS analyses, global heteroscedasticity ranges from 0.44 to 0.7 across the different types of analysed qMRI data, and the fraction of voxels with significant local heteroscedasticity ranges from 0.42 to 0.85 (Fig. 1). Global and local heteroscedasticity are largely comparable across the analyses of $R2^*_{(2)}$ and $R2^*_{(3)}$ maps, suggesting little effect of the number of raw images involved in the computation of the qMRI maps on heteroscedasticity. The generally lower level of heteroscedasticity in analyses of R1 and MTsat maps suggests a bigger effect of the type of analysed qMRI data.

WLS analyses reduce global and local heteroscedasticity for all types of qMRI data, models of covariance matrix and tissue types (Fig.1). Local heteroscedasticity is below 0.05 with $\alpha_{max} =$ 2,4 or $5, \lambda \in \mathbb{R}$). With these models, the global measure R^2 is below 0.16. The two alternative models ($\alpha_{max} = 3, \lambda \in \mathbb{R}$ and $\alpha_{max} = 4, \lambda \in \mathbb{R}^+$) lead to higher global and local heteroscedasticity, particularly for MTsat data.

Across the noise models with ($\lambda \in \mathbb{R}$), the increase in ELBO compared to OLS analyses reach a maximum for $\alpha_{max} = 3$, closely followed by $\alpha_{max} = 4$. Noise models with $\alpha_{max} = 2$ and $\alpha_{max} = 5$ lead to a decrease in ELBO in analyses of MTsat maps. The gains in ELBO compared to OLS analyses are higher in analyses of $R2^*_{(2)}$ data followed by $R2^*_{(3)}$, R1, and MTsat data, suggesting an effect of both the type of analysis data and of the number of raw image volumes involved in their computation.

Enforcing positivity for the REML hyperparameter (i.e. $\lambda \in \mathbb{R}^+$) further increased the ELBO. However, noise heteroscedasticity is comparably high for this model of the noise covariance.

Noise homoscedasticity is essential to ensure the validity of statistical inference and was the key requirement in our selection of the optimal model of the noise covariance. Among the models that ensure noise homoscedasticity ($\alpha_{max} = 2,4$ or $5, \lambda \in \mathbb{R}$), we deemed $\alpha_{max} = 4, \lambda \in \mathbb{R}$ to be optimal as it leads to the highest increase in ELBO compared to OLS analyses, for all types of analysis data and both tissue types. With this model, analysis residuals depend only weakly on the value of the MDI of the analyses data (i.e. global heteroscedasticity $R^2 \leq 0.16$ see Fig.2A). The ELBO systematically increases compared to OLS analyses, by an amount that varies according to the type of qMRI data and number of raw image volumes involved in the map calculation (see Fig.2B).

3.2 Age sensitivity

Differences in age sensitivity between OLS and WLS analyses conducted on the full dataset show substantial effects, both positive and negative (Supporting Figure S1). This is consistent with the effect of noise heteroscedasticity in OLS analyses, which might lead to under- or over-estimation of the noise level - both of which invalidate associated inferences. To ensure a valid comparison of age sensitivity with WLS analyses, we investigated the exclusion from the OLS analyses of the datasets most affected by head motion, which we identified from their high MDI values. Overall, global and local noise heteroscedasticity decrease with increasing fraction of excluded datasets (Fig.3). After exclusion of the 30% of the datasets with the highest MDIs, global heteroscedasticity lies in the same range as WLS analyses overall (R^2~0.1-0.2), except for $R2^*_{(2)}$ and $R2^*_{(3)}$ data in white matter (R^2=0.34). However, local heteroscedasticity remains generally higher than for WLS analyses (<0.05), and reaches 0.12 in R2* data and 0.25 in the white matter of MTsat data (solid bars in Fig.3). Higher exclusion fractions, which would have reduced local heteroscedasticity further, were deemed too prohibitive to be investigated.

The age sensitivity of the WLS analyses was compared with that of OLS analyses after exclusion of the 30% of datasets with the highest MDIs. Statistical F-maps of age-related changes in qMRI data, obtained from WLS analyses, exhibit predominant features that are consistent with previous findings from the literature (Fig.4, [Callaghan et al., 2014]). These include an increase in R2* with age in sub-cortical areas attributed to a local increase in iron concentration, and a decrease in MTsat and R1 in frontal white matter attributed to fibre demyelination. With WLS analyses, statistical scores are larger and the significance threshold is lower due to the higher number of samples (see inset in Fig. 4). The number of voxels above significance increases by a factor 2.8 to 4.3 with WLS analyses, except for the R1 and MTsat parameters in white matter, where the increase is of 20% and 5% respectively (Fig. 5). With WLS analyses, the spatial distribution of significant voxels shows enhanced symmetry between the left and right hemispheres (Fig. 5). Regions of significant voxels also show improved spatial continuity.

3.3 Specificity

OLS analyses and WLS analyses with the optimal noise model ($\alpha_{max} = 4, \lambda \in \mathbb{R}$) show equivalent rates of false positives in both F-tests and imbalanced two-sample T-tests (Table 1). On average, OLS and WLS analyses show significant results at the cluster level in 5.4% and 6.5% of the tests, in agreement with the expected rate of 5%. The rates of false positives are slightly lower for the tests with shuffled age regressor but remain comparable between OLS and WLS analyses (3.1% and 2.5% respectively).

3.4 Inserting the Motion Degradation Index in the analysis design

The level of noise heteroscedasticity in WLS analyses conducted with QUIQI was compared with that of OLS analyses after insertion of the MDI into the design matrix. Residual heteroscedasticity remains present in OLS analyses ($R^2_{WLS} < R^2_{OLS}$), for all data and tissue types (Fig. 6).

4. Discussion

Analyses of MRI data using linear regression rely on the assumption of noise homoscedasticity to ensure the validity of statistical inference. Degradation of MRI data quality due to head motion invalidates this assumption. The resulting mis-estimation of the noise variance leads to erroneous statistical inference and increased risks of false positives or negatives. The QUIQI method restores the validity of statistical analyses by conducting weighted least-square estimations, with weights that are computed from an index of data degradation due to head motion. The benefit of QUIQI has recently been demonstrated from a large dataset of R2* maps computed from a single set of raw images [Lutti et al., 2022]. Here, we extended this method to brain maps of the MRI parameters R2*, R1 and MTsat, computed from multiple sets of raw images. The impact of restoring statistical validity is substantial and consistent with under- or over-estimation of the noise variance in OLS analyses (Supporting Figure S1). WLS analyses show higher sensitivity to brain differences across a dataset than OLS analyses that excluded the most corrupted data (Fig. 4-5). WLS analyses are also more effective at ensuring homoscedasticity than insertion of the MDIs as regressors in the design matrix of the analysis.

QUIQI requires modelling of the noise covariance matrix from a Motion Degradation Index (MDI) of the analysis data. The relationship between analysis residuals and the MDI in an OLS analysis constitutes a good baseline to infer a suitable form for this model [Lutti et al., 2022]. For the current application, we chose to base this model on polynomial combinations of the MDI of each of the raw images involved in the computation of the R2*, R1 and MTsat maps. We compared candidate models, with different polynomial degrees and constraints on the hyperparameter λ, from measures of local and global noise heteroscedasticity, and from ELBO - a measure of free energy. WLS analyses strongly reduced heteroscedasticity for all the types of qMRI data analysed here. Also, the same subset of models consistently led to optimal restoration of homoscedasticity. From this subset, we identified the optimal noise model that led to the highest increase in ELBO compared to OLS analyses. The use of higher powers of the MDI increased the complexity of the noise covariance model, leading to a decrease in the ELBO.

The sensitivity of OLS analyses and WLS analyses conducted with QUIQI were compared in the context of age-related effects [MacDonald and Pike, 2021]. For our comparison, the 30% of datasets with the highest MDI were removed from the OLS analyses to achieve a similar level of homoscedasticity than the WLS analyses. The higher sensitivity of the WLS analyses conducted with QUIQI arises from its ability to exploit the full dataset to increase statistical power. With WLS analyses, the spatial distribution of significant voxels shows improved biological plausibility in the form of greater spatial continuity and enhanced symmetry between the left and right hemispheres. This increase in sensitivity preserves specificity, i.e. the rate of false positives in the statistical analyses remained under control.

QUIQI is available for use within the hMRI toolbox (https://hmri-group.github.io/hMRI-toolbox/). This implementation relies on the REML algorithm of SPM12 that is commonly used to correct for temporal correlations and group-level analyses of fMRI data [Friston et al., 2002a]. Following specification of the design matrix ('*factorial design*'), the *QUIQI_Build* module builds a dictionary of basis functions from MDI values provided by the user. By default, *QUIQI_Build* computes the basis functions as powers of the MDI that are specified by the user. Alternatively, the noise model can be readily adapted to different analysis datasets and MDIs (see hmri_quiqi_build.m). By default, no constraint is imposed on the hyperparameters (see [Lutti et al., 2022] for details on adding a positivity constraint). The *QUIQI_Check* module can be used after image analysis to estimate global heteroscedasticity in the data. Here, the degree of the polynomial used to estimate $\mathrm{var}(\varepsilon)_{\mathrm{fit}}$ is independent of the powers of the MDI used to model the noise covariance matrix in *QUIQI Build*.

5. Conclusion

Degradation of image quality due to head motion invalidates the homoscedasticity assumption in the statistical analysis of MRI data. Here, we extended the QUIQI method to the analysis of brain maps of the MRI parameters R2*, R1 and MTsat, computed from multiple sets of raw images. QUIQI restores homoscedasticity and the validity of statistical inference, and allows for optimization of the noise model using specially-dedicated metrics of heteroscedasticity and free energy. QUIQI is more effective at ensuring homoscedasticity than regressing out the image quality indices, and yields higher sensitivity

than removal of the datasets most corrupted by head motion from the analysis. QUIQI is available in the hMRI toolbox for the study of brain differences from MRI data.

References

Ashburner J, Friston KJ (2005): Unified segmentation. NeuroImage 26:839–851.

Ashburner J (2007): A fast diffeomorphic image registration algorithm. NeuroImage 38:95–113.

Bagnato F, Hametner S, Boyd E, Endmayr V, Shi Y, Ikonomidou V, Chen G, Pawate S, Lassmann H, Smith S, Welch EB (2018): Untangling the R2* contrast in multiple sclerosis: A combined MRI-histology study at 7.0 Tesla. PLOS ONE 13:e0193839.

Balbastre Y, Aghaeifar A, Corbin N, Brudfors M, Ashburner J, Callaghan MF (2022): Correcting inter-scan motion artifacts in quantitative R 1 mapping at 7T. Magn Reson Med. https://hal.science/hal-03646797.

Beveridge JE, Machan JT, Walsh EG, Kiapour AM, Karamchedu NP, Chin KE, Proffen BL, Sieker JT, Murray MM, Fleming BC (2018): Magnetic resonance measurements of tissue quantity and quality using T 2 * relaxometry predict temporal changes in the biomechanical properties of the healing ACL. J Orthop Res 36:1701–1709.

Boots EA, Castellanos KJ, Zhan L, Barnes LL, Tussing-Humphreys L, Deoni SCL, Lamar M (2020): Inflammation, Cognition, and White Matter in Older Adults: An Examination by Race. Front Aging Neurosci 12.

Callaghan MF, Freund P, Draganski B, Anderson E, Cappelletti M, Chowdhury R, Diedrichsen J, FitzGerald THB, 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.

Carey D, Caprini F, Allen M, Lutti A, Weiskopf N, Rees G, Callaghan MF, Dick F (2018): Quantitative MRI provides markers of intra-, inter-regional, and age-related differences in young adult cortical microstructure. Neuroimage 182:429–440.

Castella R, Arn L, Dupuis E, Callaghan MF, 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.

Draganski B, Ashburner J, Hutton C, Kherif F, Frackowiak RSJ, Helms G, Weiskopf N (2011): Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). NeuroImage 55:1423–1434.

Engle R (1982): Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United-Kingdom Inflation. Econometrica 50:987–1007.

Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ (2017): MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PloS One 12:e0184661.

Friston KJ, Glaser DE, Henson RNA, Kiebel S, Phillips C, Ashburner J (2002a): Classical and Bayesian inference in neuroimaging: applications. NeuroImage 16:484–512.

Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J (2002b): Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16:465–483.

Fukunaga M, Li T-Q, van Gelderen P, de Zwart JA, Shmueli K, Yao B, Lee J, Maric D, Aronova MA, Zhang G, Leapman RD, Schenck JF, Merkle H, Duyn JH (2010): Layer-specific variation of iron content in cerebral cortex as a source of MRI contrast. Proc Natl Acad Sci 107:3834–3839.

Gilmore AD, Buser NJ, Hanson JL (2021): Variations in structural MRI quality significantly impact commonly used measures of brain anatomy. Brain Inform 8:7.

Glickman ME, Rao SR, Schultz MR (2014): False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J Clin Epidemiol 67:850–857.

Hametner S, Endmayr V, Deistung A, Palmrich P, Prihoda M, Haimburger E, Menard C, Feng X, Haider T, Leisser M, Köck U, Kaider A, Höftberger R, Robinson S, Reichenbach JR, Lassmann H, Traxler H, Trattnig S, Grabner G (2018): The influence of brain iron and myelin on magnetic susceptibility and effective transverse relaxation - A biochemical and histological validation study. NeuroImage 179:117–133.

Hayes AF, Cai L (2007): Using heteroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation. Behav Res Methods 39:709–722.

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.

Henkelman RM, Stanisz GJ, Graham SJ (2001): Magnetization transfer in MRI: a review. NMR Biomed 14:57–64.

Honigberg MC, Pirruccello JP, Aragam K, Sarma AA, Scott NS, Wood MJ, Natarajan P (2020): Menopausal age and left ventricular remodeling by cardiac magnetic resonance imaging among 14,550 women. Am Heart J 229:138–143.

Hu B, Cha J, Fullerton JM, Hesam-Shariati S, Nakamura K, Nurnberger JI, Anand A (2022): Genetic and environment effects on structural neuroimaging endophenotype for bipolar disorder: a novel molecular approach. Transl Psychiatry 12.

Loued-Khenissi L, Trofimova O, Vollenweider P, Marques-Vidal P, Preisig M, Lutti A, Kliegel M, Sandi C, Kherif F, Stringhini S, Draganski B (2022): Signatures of life course socioeconomic conditions in brain anatomy. Hum Brain Mapp 43:2582–2606.

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, Dick F, Sereno MI, Weiskopf N (2014): Using high-resolution quantitative mapping of R1 as an index of cortical myelination. NeuroImage 93:176–188.

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. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3299646/.

MacDonald ME, Pike GB (2021): MRI of healthy brain aging: A review. NMR Biomed 34:e4564.

Mohammadi S, Streubel T, Klock L, Edwards LJ, Lutti A, Pine KJ, Weber S, Scheibe P, Ziegler G, Gallinat J, Kühn S, Callaghan MF, Weiskopf N, Tabelow K (2022): Error quantification in multi-parameter mapping facilitates robust estimation and enhanced group level sensitivity. NeuroImage 262. https://doi.org/10.1016/j.neuroimage.2022.119529.

Mortamet B, Bernstein MA, Jack CR, Gunter JL, Ward C, Britson PJ, Meuli R, Thiran J-P, Krueger G, Alzheimer's Disease Neuroimaging Initiative (2009): Automatic quality assessment in structural brain magnetic resonance imaging. Magn Reson Med 62:365–372.

Oliveira R, Raynaud Q, Corbin N, Di Domenicantonio G, Callaghan MF, Lutti A (2023): QUIQI II Analysis Script - Statistical analyses of motion-corrupted MRI relaxometry data. Zenodo. https://zenodo.org/record/7612032.

Ong T, Bharatha A, Alsufayan R, Das S, Lin AW (2021): MRI predictors for brain invasion in meningiomas. Neuroradiol J 34:3–7.

Ong YY, Pang WW, Huang JY, Aris IM, Sadananthan SA, Tint M-T, Yuan WL, Chen L-W, Chan YH, Karnani N, Velan SS, Fortier MV, Choo J, Ling LH, Shek L, Tan KH, Gluckman PD, Yap F, Chong Y-S, Godfrey KM, Chong MF-F, Chan S-Y, Eriksson JG, Wlodek ME, Lee YS, Michael N (2022): Breastfeeding may benefit cardiometabolic health of children exposed to increased gestational glycemia in utero. Eur J Nutr 61:2383–2395.

Panda A, Obmann VC, Lo W-C, Margevicius S, Jiang Y, Schluchter M, Patel IJ, Nakamoto D, Badve C, Griswold MA, Jaeger I, Ponsky LE, Gulani V (2019): MR fingerprinting and ADC mapping for characterization of lesions in the transition zone of the prostate gland. Radiology 292:685–694.

Papadaki E, Kavroulakis E, Kalaitzakis G, Karageorgou D, Makrakis D, Maris TG, Simos PG (2019): Age-related deep white matter changes in myelin and water content: A T2 relaxometry study. J Magn Reson Imaging 50:1393–1404.

Pizarro RA, Cheng X, Barnett A, Lemaitre H, Verchinski BA, Goldman AL, Xiao E, Luo Q, Berman KF, Callicott JH, Weinberger DR, Mattay VS (2016): Automated Quality Assessment of Structural Magnetic Resonance Brain Images Based on a Supervised Machine Learning Algorithm. Front Neuroinformatics 10. https://www.frontiersin.org/articles/10.3389/fninf.2016.00052.

Plaikner M, Kremser C, Zoller H, Jaschke W, Henninger B (2018): Monitoring iron overload: Relationship between R2∗ relaxometry of the liver and serum ferritin under different therapies. J Clin Imaging Sci 8.

van der Plas E, Gutmann L, Thedens D, Shields RK, Langbehn K, Guo Z, Sonka M, Nopoulos P (2021): Quantitative muscle MRI as a sensitive marker of early muscle pathology in myotonic dystrophy type 1. Muscle Nerve 63:553–562.

Raynaud Q, Oliveira R, Draganski B, Kherif F, Corbin N, Di Domenicantonio G, Callaghan MF, Lutti A (2023): QUIQI II Dataset - Statistical analyses of motion-corrupted MRI relaxometry data. Zenodo. https://zenodo.org/record/7692074.

Reuter M, Tisdall MD, Qureshi A, Buckner RL, van der Kouwe AJW, Fischl B (2015): Head motion during MRI acquisition reduces gray matter volume and thickness estimates. NeuroImage 107:107–115.

Rosen AFG, Roalf DR, Ruparel K, Blake J, Seelaus K, Villa LP, Ciric R, Cook PA, Davatzikos C, Elliott MA, Garcia de La Garza A, Gennatas ED, Quarmley M, Schmitt JE, Shinohara RT, Tisdall MD, Craddock RC, Gur RE, Gur RC, Satterthwaite TD (2018): Quantitative assessment of structural image quality. NeuroImage 169:407–418.

Savalia NK, Agres PF, Chan MY, Feczko EJ, Kennedy KM, Wig GS (2017): Motion-related artifacts in structural brain images revealed with independent estimates of in-scanner head motion. Hum Brain Mapp 38:472–492.

Schall M, Iordanishvili E, Mauler J, Oros-Peusquens A-M, Shah NJ (2020): Increasing body mass index in an elderly cohort: Effects on the quantitative MR parameters of the brain. J Magn Reson Imaging 51:514–523.

Scott RC, King MD, Gadian DG, Neville BGR, Connelly A (2003): Hippocampal abnormalities after prolonged febrile convulsion: A longitudinal MRI study. Brain 126:2551–2557.

Sereno MI, Lutti A, Weiskopf N, Dick F (2013): Mapping the Human Cortical Surface by Combining Quantitative T1 with Retinotopy. Cereb Cortex N Y NY 23:2261–2268.

Stüber C, Morawski M, Schäfer A, Labadie C, Wähnert M, Leuze C, Streicher M, Barapatre N, Reimann K, Geyer S, Spemann D, Turner R (2014): Myelin and iron concentration in the human brain: a quantitative study of MRI contrast. NeuroImage 93 Pt 1:95–106.

Tabelow K, Balteau E, Ashburner J, Callaghan MF, 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.

Trofimova O, Loued-Khenissi L, DiDomenicantonio G, Lutti A, Kliegel M, Stringhini S, Marques-Vidal P, Vollenweider P, Waeber G, Preisig M, Kherif F, Draganski B (2021): Brain tissue properties link cardio-vascular risk factors, mood and cognitive performance in the CoLaus|PsyCoLaus epidemiological cohort. Neurobiol Aging 102:50–63.

Weiskopf N, Callaghan MF, 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. https://www.frontiersin.org/articles/10.3389/fnins.2014.00278/full.

Yao B, Li T-Q, Gelderen P van, Shmueli K, de Zwart JA, Duyn JH (2009): Susceptibility contrast in high field MRI of human brain as a function of tissue iron content. NeuroImage 44:1259–1266.

$R2^*_{(2)}$

		GM	WM
Two-sample T-test (cluster level)	OLS	0.054	0.040
	WLS	0.067	0.055
F-test (cluster level)	OLS	0.031	0.027
	WLS	0.023	0.022

$R2^*_{(3)}$

		GM	WM
Two-sample T-test (cluster level)	OLS	0.055	0.043
	WLS	0.065	0.066
F-test (cluster level)	OLS	0.032	0.028
	WLS	0.027	0.022

R1

		GM	WM
Two-sample T-test (cluster level)	OLS	0.096	0.036
	WLS	0.088	0.047
F-test (cluster level)	OLS	0.055	0.019
	WLS	0.038	0.018

MT

		GM	WM
Two-sample T-test (cluster level)	OLS	0.069	0.039
	WLS	0.063	0.066
F-test (cluster level)	OLS	0.036	0.018
	WLS	0.030	0.021

Table 1 : Specificity is preserved with WLS analyses. Specificity of WLS and OLS analyses of R2*(2), R2*(3), R1 and MTsat data, in grey and white matter. The rate of false positive clusters ($p < 0.05$) were obtained with two tests: 1/ Group comparison of two sub-groups within a narrow age range with a two-sampled T-test 2/ Age-associated differences after random assignment of the participants' age with an F-Test. Any positive result is a false positive. The cluster forming threshold was $p < 0.001$ uncorrected.

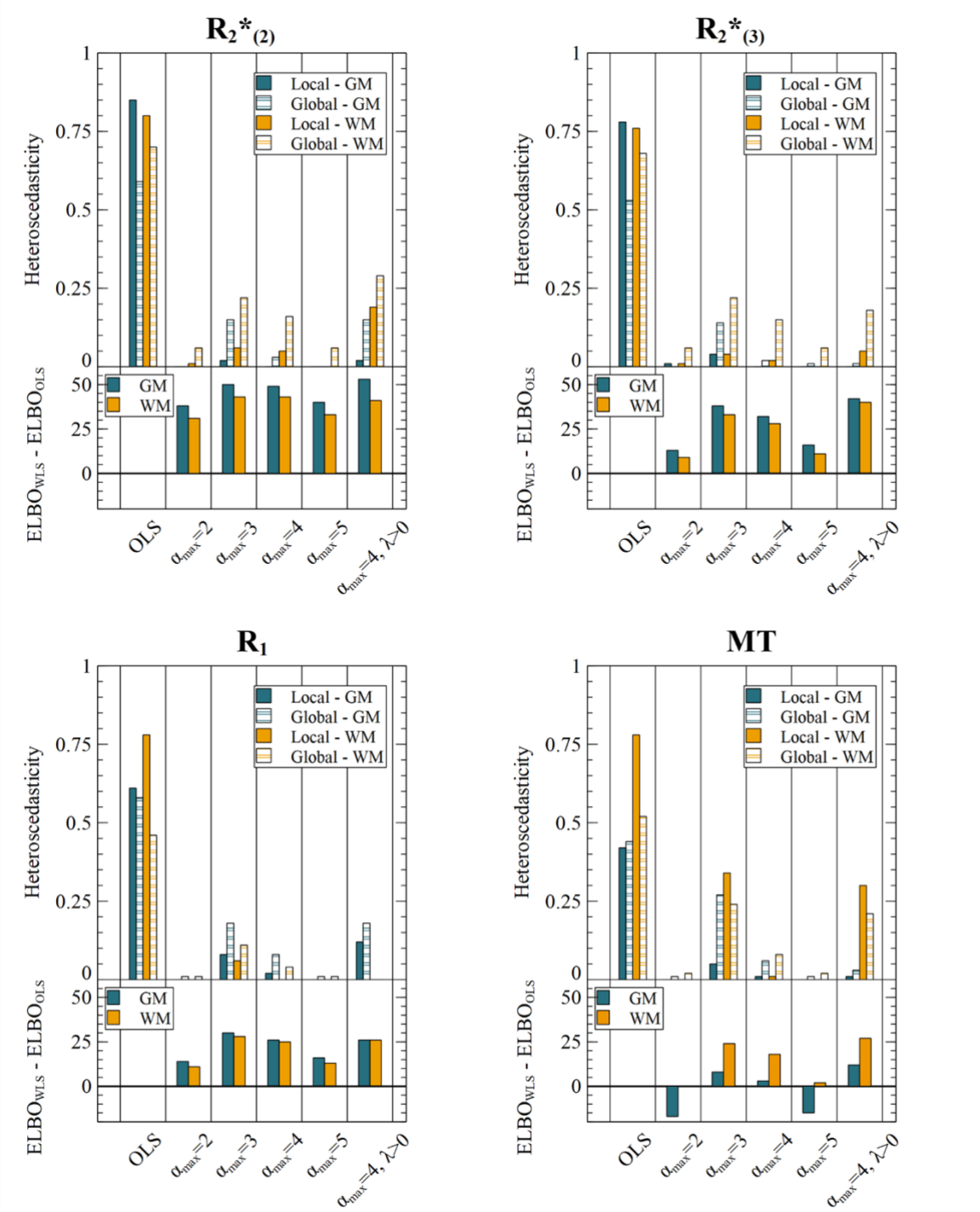

Figure 1: WLS analyses reduce heteroscedasticity and increase ELBO. Each graph shows heteroscedasticity and ELBO levels for different qMRI maps R2*$_{(2)}$ (top left), R2*$_{(3)}$ (top right), R1 (bottom left) and MTsat (bottom right)). The top part of each graph shows local and global heteroscedasticity levels for OLS analyses and WLS analyses with different noise models (i.e. α_{max}), in white matter (WM) and grey matter (GM). The bottom part shows differences in ELBO between WLS and OLS analyses. WLS analyses strongly reduce noise heteroscedasticity, for all noise models and qMRI maps. WLS analyses also lead to an increase in ELBO, except for two noise models in the GM of MTsat maps.

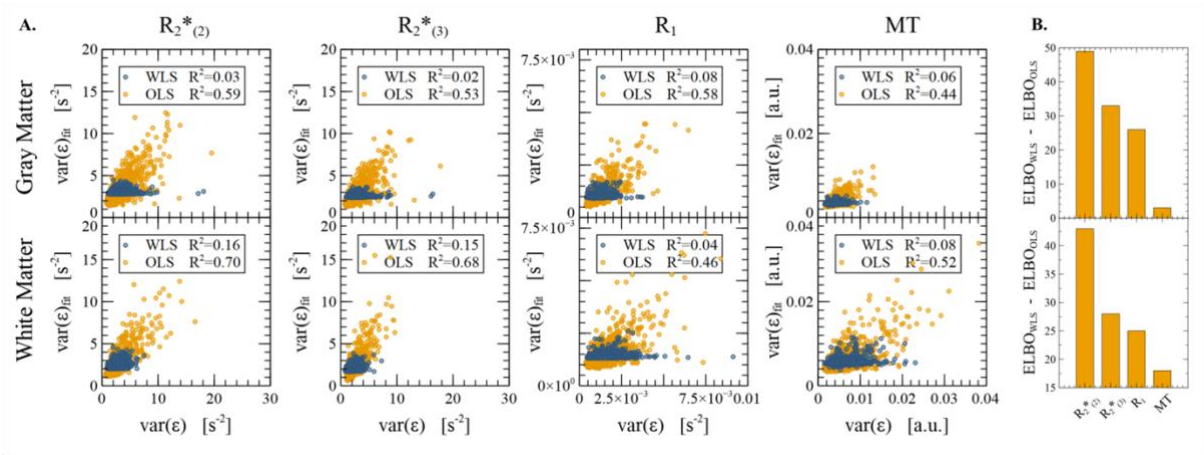

Figure 2: Noise homoscedasticity and ELBO increase with the optimal noise model. (a) With the optimal noise model ($\alpha_{max} = 4, \lambda \in \mathbb{R}$), global heteroscedasticity (R^2) does not exceed 0.16, for all types of qMRI maps and in both grey and white matter. (b) WLS analyses lead to an increase in ELBO, for all types of qMRI maps and in both grey and white matter. This increase varies according to the type of qMRI data and the number of raw image volumes involved in the map calculation.

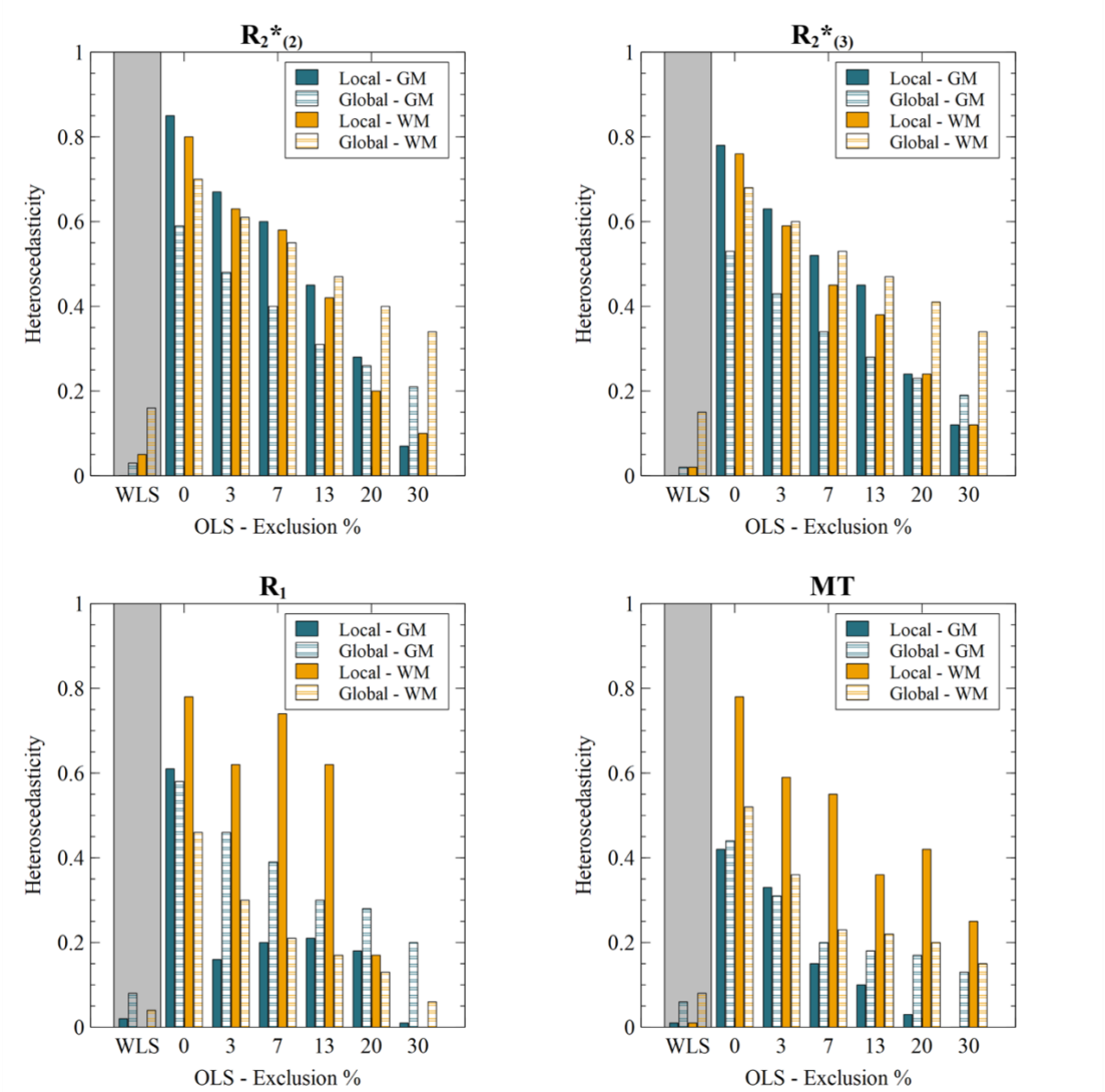

Figure 3: Exclusion of the most degraded datasets improves noise heteroscedasticity in OLS analyses. After exclusion of 30% of the datasets, global heteroscedasticity lies in the same range as WLS analyses overall (R^2~0.1-0.2), except for $R2^*_{(2)}$ and $R2^*_{(3)}$ maps in white matter (R^2=0.34). However local heteroscedasticity remains generally higher than for WLS analyses (<0.05), and reaches 0.12 in R2* data and 0.25 in the white matter of MTsat data.

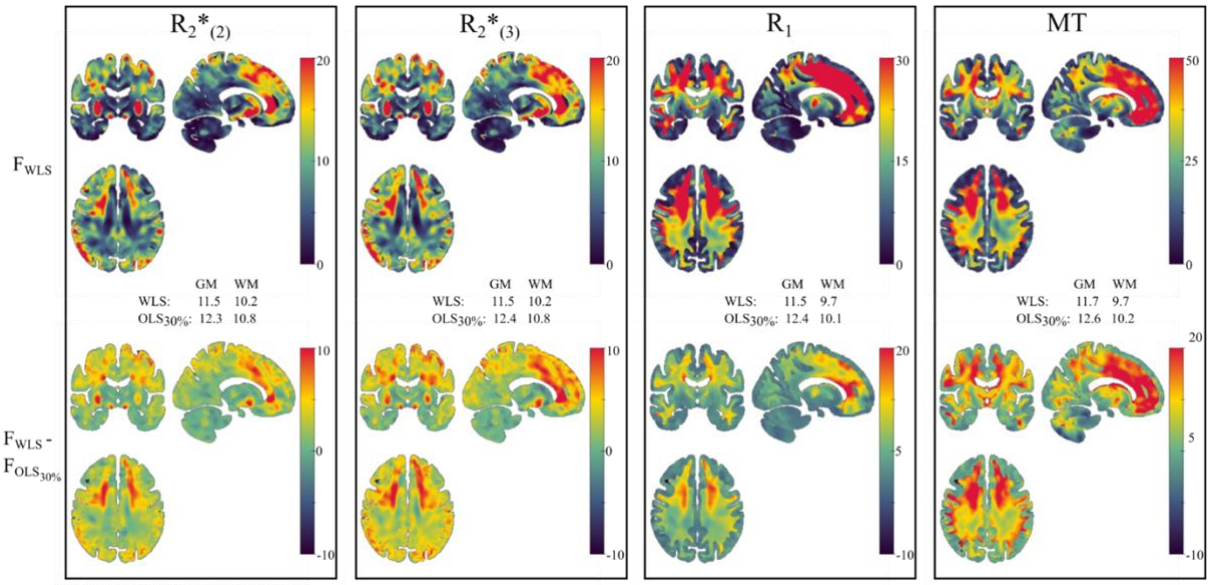

Figure 4: WLS analyses show increased sensitivity to brain differences. Statistical F-maps of age-related changes (F_{WLS}), obtained in WLS analyses of 123 samples of the full dataset, show predominant features that include an increase in R2* with age in sub-cortical areas and a decrease in MTsat and R1 in frontal white matter (top). WLS analyses exhibit higher Fscores than OLS analyses after exclusion of 30% of the most corrupted data, in all types of qMRI data and throughout grey and white matter (bottom). For each type of qMRI maps, the inset indicates the threshold for statistical significance ($p<0.05$, FWE corrected).

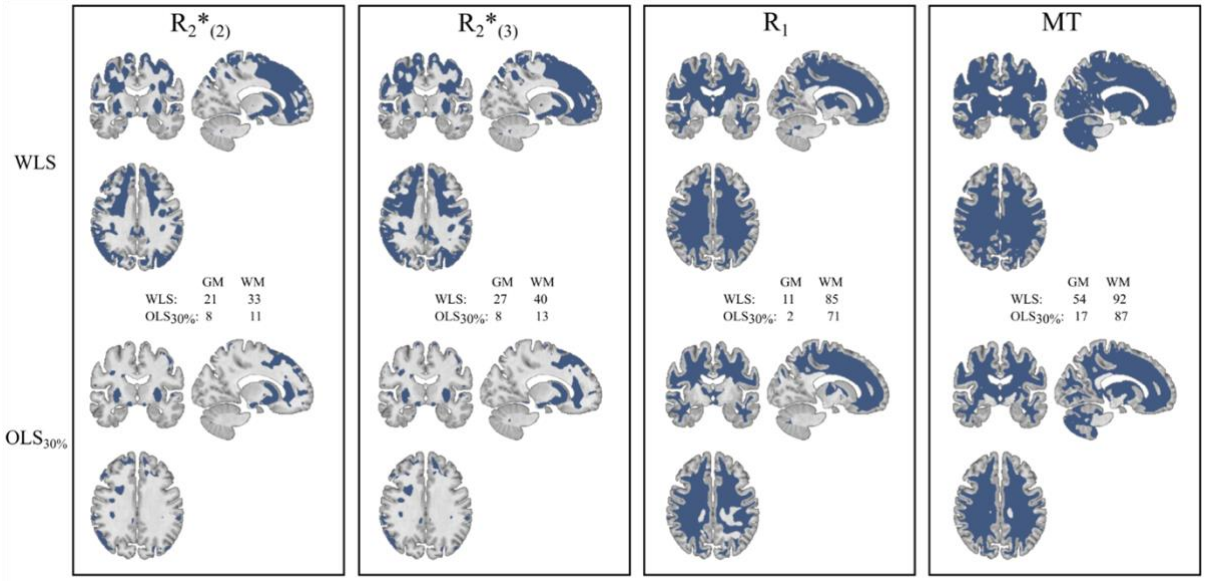

Figure 5: WLS analyses increase the spread of significant voxels in statistical maps. Maps of voxels with significant age-related changes (blue) ($p<0.05$, FWE corrected) show increased spatial extent with WLS analyses (top) than OLS analyses with exclusion of 30% of the most corrupted data (bottom), for all types of qMRI data. The insets indicate the fraction of significant voxels in grey and white matter.

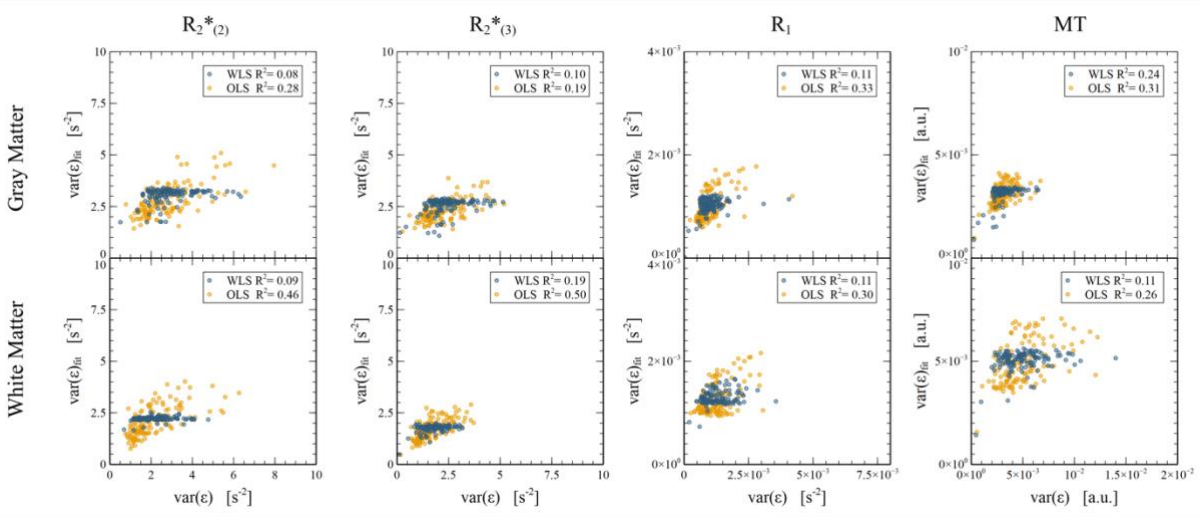

Figure 6: Including the motion degradation index as regressors in the design matrix does not restore homoscedasticity. Despite inserting the MDI in the design matrix, OLS analyses exhibit a high level of global noise heteroscedasticity (high R^2), for all types of qMRI maps and in both grey and white matter. Homoscedasticity is restored with WLS analyses.

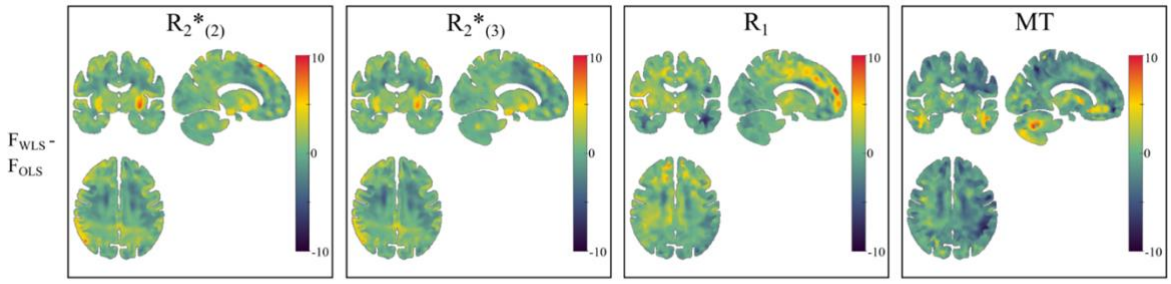

Figure S1: Noise heteroscedasticity leads to under- or over-estimated statistical results. Differences in statistical F-maps between OLS and WLS analyses conducted on the full dataset are substantial and may be positive or negative, consistent with the effect of noise heteroscedasticity in OLS analyses.