Received: 25 July 2023 | Revised: 1 December 2023 | Accepted: 4 December 2023

DOI: 10.1002/mrm.29983

RESEARCH ARTICLE

Magnetic Resonance in Medicine

Quantifying 3D MR fingerprinting (3D-MRF) reproducibility across subjects, sessions, and scanners automatically using MNI atlases

Andrew Dupuis[1] | **Yong Chen**[2,3] | **Michael Hansen**[4] | **Kelvin Chow**[5] | **Jessie E. P. Sun**[2] | **Chaitra Badve**[3] | **Dan Ma**[1] | **Mark A. Griswold**[2] | **Rasim Boyacioglu**[2]

[1]Department of Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio, USA

[2]Department of Radiology, Case Western Reserve University, Cleveland, Ohio, USA

[3]Department of Radiology, University Hospitals, Cleveland, Ohio, USA

[4]Microsoft Research, Redmond, Washington, USA

[5]Siemens Medical Solutions USA, Inc, Chicago, Illinois, USA

Correspondence
Andrew Dupuis, Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, 44106.
Email: andrew.dupuis@case.edu

Funding information
Research reported in this publication was supported by the National Institutes of Health (USA) under award numbers, Grant/Award Numbers: R01EB032709, R01CA269604, R01NS109439, R01CA266702, R01CA282516

Abstract

Purpose: Quantitative MRI techniques such as MR fingerprinting (MRF) promise more objective and comparable measurements of tissue properties at the point-of-care than weighted imaging. However, few direct cross-modal comparisons of MRF's repeatability and reproducibility versus weighted acquisitions have been performed. This work proposes a novel fully automated pipeline for quantitatively comparing cross-modal imaging performance in vivo via atlas-based sampling.

Methods: We acquire whole-brain 3D-MRF, turbo spin echo, and MPRAGE sequences three times each on two scanners across 10 subjects, for a total of 60 multimodal datasets. The proposed automated registration and analysis pipeline uses linear and nonlinear registration to align all qualitative and quantitative DICOM stacks to Montreal Neurological Institute (MNI) 152 space, then samples each dataset's native space through transformation inversion to compare performance within atlas regions across subjects, scanners, and repetitions.

Results: Voxel values within MRF-derived maps were found to be more repeatable ($\sigma_{T1} = 1.90$, $\sigma_{T2} = 3.20$) across sessions than vendor-reconstructed MPRAGE ($\sigma_{T1w} = 6.04$) or turbo spin echo ($\sigma_{T2w} = 5.66$) images. Additionally, MRF was found to be more reproducible across scanners ($\sigma_{T1} = 2.21$, $\sigma_{T2} = 3.89$) than either qualitative modality ($\sigma_{T1w} = 7.84$, $\sigma_{T2w} = 7.76$). Notably, differences between repeatability and reproducibility of in vivo MRF were insignificant, unlike the weighted images.

Conclusion: MRF data from many sessions and scanners can potentially be treated as a single dataset for harmonized analysis or longitudinal comparisons without the additional regularization steps needed for qualitative modalities.

KEYWORDS

Bland Altman, MR fingerprinting, precision, quantitative imaging, relaxation mapping, repeatability

1 | INTRODUCTION

Whereas most clinical MR applications still regularly employ qualitative image generation as the default technique, research studies are consistently demonstrating areas where quantitative MR techniques may exhibit a direct clinical benefit over qualitative alternatives.[1] One quantitative technique is MR fingerprinting (MRF), a US Food and Drug Administration–approved MRI method that performs simultaneous measurements of various tissue properties, including T_1 and T_2 relaxation times.[2] Quantitative analysis of MRF maps has been beneficial in better understanding and tracking of healthy function, disease diagnosis, and disease progression in areas such as the brain,[3–7] heart,[8,9] liver,[10–12] prostate,[13] and kidney.[14,15]

However, before any widespread clinical adoption, quantitative methods such as MRF need to offer significant, provable advantages over established, qualitative MRI scans. In this study, we establish reproducibility and repeatability within MRF techniques. We then demonstrate how reproducibility and repeatability are improved in MRF compared to current clinical imaging techniques, bringing more sensitive diagnostic tools to current imaging suites. Additionally, throughout this study, we determine techniques for creating measurable benchmarks of success and ensuring consistent techniques for future convenience and reliability within the world of MRF imaging, factors that are not consistently considered in qualitative imaging.

Whereas qualitative weighted imaging techniques regulate noise characteristics,[16] spatial accuracy,[17] and signal uniformity,[18] the resultant images are not required to reproduce exact contrasts for different scanners and sites with the same input parameters. These weighted images do not have to be comparable directly to previously acquired images and thus cannot be reliably considered quantifiable MR images.[19] On the contrary, in prior phantom[20] and in vivo[21–26] studies, MRF images were able to quantify T_1 and T_2 values reliably and reproducibly. Some reproducibility studies used 2D acquisitions[22] with partial brain coverage and manually drawn regions of interest (ROIs)[21]; others used 3D MRF, where the coefficient of variation, intraclass correlation, mean gray matter (GM) and white matter (WM) T_1 and T_2 values,[23] cortical thickness, and subcortical region volumes[25] were studied. However, there are few cross-modal comparisons of MRF (or other quantitative MRI techniques) against conventional weighted imaging in terms of repeatability and reproducibility. Compared to the larger body of MRF work, only a few previous studies specifically compared the repeatability of MRF[24] or reproducibility of T_1 mapping[27] against conventional T_1-weighted (T_1w) imaging. None of these prior studies include the entire imaging and analysis chain as a composite source of variability.

The goal of this study was to present a fully integrated acquisition and online reconstruction and analysis framework for 3D MRF that forms an automated, reproducible, traceable pipeline. We then evaluate both the repeatability and reproducibility of 3D MRF compared to clinically standard T_1w and T_2w imaging using this pipeline. To establish repeatability and reproducibility, previously determined definitions of traceability and uncertainty[28] are imbued with specific meanings within specific parameters of MRF. We compared the mean values of different Montreal Neurological Institute (MNI) brain atlas regions to examine the in vivo repeatability and reproducibility of MRF and conventional weighted imaging. We used a fully automated registration pipeline designed to maximize cross-modality spatial coherence and to establish specific benchmarks for repeatability and reproducibility, using automated analysis. We also ensured that all analysis techniques used are available in version-controlled and traceable venues.

2 | METHODS

2.1 | Study design

Ten healthy volunteers (36.8 ± 14.9 years; five men, five women) gave written consent and were scanned in this study according to the applicable institutional review board-approved protocol. Volunteers were imaged over two sessions occurring on different days, with one session each on two 3 T scanners running different software versions (Magnetom Vida, VA20 and VA31, Siemens Healthcare, Erlangen, Germany).

All imaging was performed by using a 20-channel head coil.

This study aims to compare the intrasession, intersession, and interscanner regional reproducibility between MRF and two qualitative acquisition approaches. Each scan session consisted of three sets of acquisitions, with each set consisting of three series of images: 3D-MRF SSFP with a B_1 mapping prescan, product 3D-MPRAGE, and product multislice 2D turbo spin echo (TSE). All were acquired with a FOV of $250 \times 250 \times 150\ mm^3$ and a spatial resolution of $1 \times 1 \times 2.5\ mm^3$.

The first set, referred to as the "original" set on each scanner, served as an initial baseline for each sequence. Immediately following the "original" set, a "repetition" set was acquired, consisting of the same three sequences in the same order. This "repetition" set is intended to serve as the intrasession test–retest because the subject was left in the scanner and the imaging FOV was copied directly from the "original." After completion of all "repetition" images, subjects were asked to leave the scanner, stand up, walk around, and then asked to reenter the scanner. A new localizer was acquired to establish the subject's new

position, while also forcing a recalibration of the scanner's acquisition system and reshimming of the B_0 and B_1 fields. Finally, a "reposition" set was acquired to serve as an intersession test–retest. Subjects were asked to return within 7 days to repeat the entire protocol on the second scanner. The scanner order was randomized between subjects to minimize the potential for scanner-specific effects on the results.

2.2 | Sequences

The MRF data was acquired first in every set for all the subjects and sessions. Before acquiring MRF data, to minimize simulation mismatch errors that result from B_1+ inhomogeneities, the RF transmit field (B_1+) was mapped for later use during pattern matching.

The MRF sequence is an inversion recovery SSFP–based sequence initiated with an adiabatic inversion pulse, followed by the acquisition of a series of 960 time points. This is then repeated for each partition in a stack-of-spirals 3D approach with an acceleration factor of 2 in the slice direction.[6] An extra 2 s pause was added between each partition to allow for longitudinal relaxation. A 2π dephasing moment in the slice direction was used within each TR.[29] A fixed TR of 10.5 ms and a fixed TE of 1.7 ms were used for every time point. The acquisition time for each partition was 10.1 s for MRF with a total acquisition time of 362.4 s for all partitions. A figure showing the flip angle pattern used in this study, as well as a human-readable javascript object notation (JSON) specification of all relevant MRF pulse parameters, is available in the Supporting Information S1. Variable density spiral trajectory with 48 spiral arms rotated by 7.5° between successive time points was designed for a FOV of $250 \times 250 \times 150$ mm and a matrix size of $256 \times 256 \times 60$. The trajectories were measured on scanner one using the approach described by Duyn et al.,[30] and the resulting trajectory was used in the nonuniform fast Fourier transform (NUFFT) gridding operator regardless of scanner.

Next, a product 3D-MPRAGE sequence was acquired to represent a T_1-weighted contrast and to be the structural baseline image within the intrasession, intersession, and interscanner registration pipelines. The product 3D-MPRAGE sequence was used with the following acquisition parameters: TR = 2100 ms, TE = 2.59 ms, TI = 900 ms, and TA = 4 m: 53 s.

Finally, a product multislice 2D-TSE acquisition was acquired to represent a T_2w contrast with the following acquisition parameters: TR = 10 620 ms, TE = 93 ms, turbo factor = 17, TA = 3 m: 2 s. Product MPRAGE and TSE sequences had the same FOV, matrix size, and position as the 3DMRF scan. Product MPRAGE and TSE scan parameters were fixed across all subjects and selected with the assistance of a research technologist to match standard clinical contrasts. Vendor-default prescan normalization was left enabled for all weighted imaging sequences to match the settings used in clinical practice at our institution, and to avoid including biases in our data that are well accounted for by standardized techniques with existing regulatory approval.

2.3 | Reconstruction

Reconstruction of the product sequences, as well as the B_1+ mapping sequence, was performed by the standard pipeline provided by the scanner/vendor. Header-complete digital imaging and communication in medicine (DICOM) standard images were exported and saved within directories specific to each subject, scanner, and acquisition set.

3D-MRF data was reconstructed online via a custom Gadgetron Kubernetes cluster using the framework for image reconstruction environment (FIRE) interface prototype.[31] All data were also converted to ISMRM raw data (ISMRM-RD)[32] format for future analysis or retrospective reconstruction. The SNR-constrained realtime compression provided by the FIRE prototype was disabled used to avoid potential differences in compression implementation across software platforms—instead, uncompressed raw data was sent to a remote reconstruction server via a secure shell (SSH) connection. Reconstruction was performed within a graphics processing unit enabled Docker container containing the precalculated MRF dictionary, singular value decomposition (SVD) compression matrix,[33] and spiral density compensation function. In order to conserve memory on the reconstruction server, time-domain SVD compression and 2D nonuniform fast Fourier transform (NUFFT) were performed on each partition on a rolling basis as raw data was received. Once all partition data has been uploaded at the end of the scan, a through-partition 1D fast Fourier transform was applied to generate the 3D SVD-space images. Finally, the resulting SVD images and the associated B_1+ maps were then pattern-matched based on maximum inner product to obtain T_1 and T_2 maps.[21] The resulting voxel wise quantitative maps were then returned to the scanner via the SSH tunnel and stored as header complete DICOM images by the FIRE prototype interface. A figure detailing the online reconstruction pipeline used is available in the Supporting Information Figure S2.

2.4 | Post-processing

The post-processing pipeline inputs were the DICOM image series organized in a semantic hierarchical file

scheme. All post-processing performed on this file hierarchy was performed by an automated pipeline based on a version-controlled Docker container to ensure identical settings, and procedures were applied to all data. Before the post-processing pipeline could be run, DICOM images were converted to nuroimaging informatics technology initiative (NIfTI) files using the open-source dcm2niix package (v1.0.20220720),[34] with consistent naming and versioning enforced by a subject-specific file map stored within each subject's DICOM repository. The resulting unified NIfTI directory structure was then processed by a Python-based image registration and statistical analysis pipeline utilizing the NiBabel[35] NIfTI management package and NiPype's[36] NIfTI pipelining system and FMRIB software library (FSL, 6.0)[37] interface extension.

Acquired images went through post-processing to establish a shared registration space among all image series for a specific subject, across all sets and scanners. The first step of the post-processing pipeline is the generation of synthetic MPRAGE and TSE contrast images for all MRF image sets as a basis for intermodality linear registrations. Whereas there are existing approaches to generating synthetic contrasts from MRF time series images using convolutional networks, this work requires pixel-precise coregistration that can be corrupted by generative approaches, such as U-NETs. Instead, synthetic contrasts were generated using a simple TensorFlow (2.11.0)[38] voxelwise regression network trained on MRF T_1/T_2 parameter maps as inputs, versus registered MPRAGE and TSE contrasts as outputs, to provide MRF-space synthetic contrasts with adequate similarity to each qualitative approach such that performant registration is possible within FSL. For network training, an initial registration was performed between MPRAGE, TSE, and T_1/T_2 map pairs; quantitative maps were manually masked to avoid the influence of free-space spiral artifacts on the network. All data from three prior volunteers were linearized into independent, two-input-one-output voxel datasets for each conventional contrast. From this linearized voxel table, data were split into training (50%) and test (50%) sets, and a dense two-layer regression model with a data normalizer was trained using the Adam optimizer with a mean absolute error loss function. The resulting network functions similarly to a color map, with unique combinations of T_1 and T_2 values corresponding to unique grayscale values based on the qualitative images used for training. This simple nonconvolutional model was then saved and used for voxelwise prediction of synthetic image values from each volunteer's T_1/T_2 map pairs.

Following synthetic image generation, all imaging sets from each subject were linearly registered together using FMRIB's linear image registration tool.[39,40] Within each image set, the product TSE images were first registered to the MRF-derived synthetic TSE images, resulting in a transformation matrix TSE to MRF space. Then, the MRF-derived synthetic MPRAGE images were registered to the product MPRAGE images. This process yielded two sets of linear transformation matrices, bringing all images within each set to the shared space of the product MPRAGE contrast.

Within each scanner, the "original" MPRAGE image series was used as a basis to which the "repetition" and "reposition" MPRAGE image series were linearly registered. As a result, linear transformations were known for all images from a single scanner to a single image space as defined by the "original" MPRAGE image series. This process was repeated for each scanner, and each scanner's respective "original" MPRAGE image series was linearly registered to the other, yielding a full chain of invertible transformation matrices for all image series to a single shared space. Finally, a single nonlinear warp field was calculated via FMRIB's Nonlinear Image Registration Tool to establish a transformation from the "first" scanner's "original" MPRAGE series to MNI-152-2 mm template space.

By inverting and combining all linear and nonlinear transformations throughout the registration chain, MNI-152 space atlases, ROIs, or other labels can be projected into each source image series. The registration flow is further detailed in Figure 1.

All raw data, DICOMs, NIfTI-converted images, linear and nonlinear transformation matrices, warped label maps, and atlas-label voxel dictionaries were uploaded to Azure Blob storage (Version 2022-11-02, Microsoft, Redmond, WA, USA) for validation and retrospective analysis. Access via a Python API is available under a data-sharing agreement to encourage further investigation and statistical analysis of the data.

2.5 | Statistical analysis

Inverse transformations from the registration were used to generate warped atlas label maps in the original image spaces for each subject/scanner/set/series combination. Eighteen well-defined, homogenous regions from the Harvard–Oxford subcortical atlas were selected for regional comparison. For each tissue compartment in every image series, the mean, median, and SD of the voxel values were calculated by using the warped series-specific label maps as a voxel mask and generating arrays of all constituent voxel values from each compartment. No post-scan normalization was applied outside of the vendor's online reconstruction—all regional statistics were generated from DICOM values in exported datasets.

The repeatability of the intrasession, intersession, and interscanner cases was assessed via comparison of the

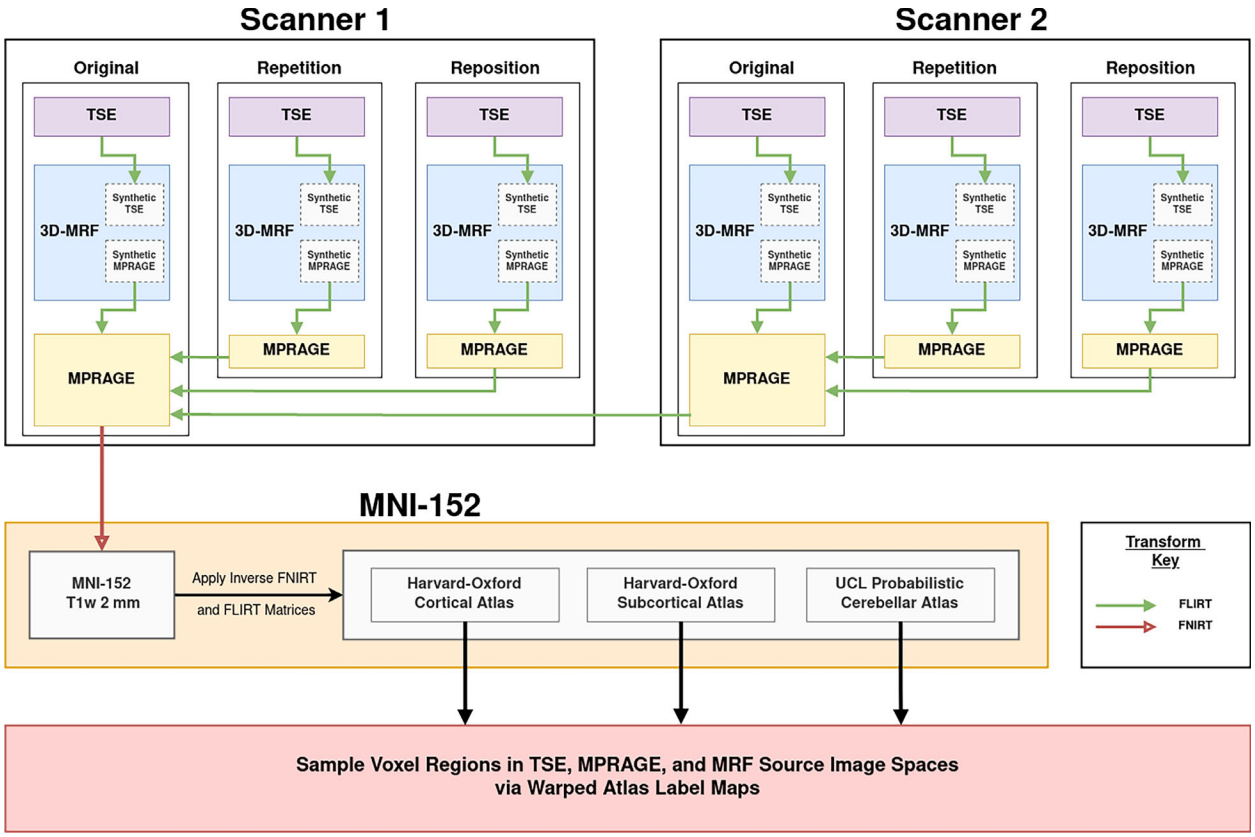

FIGURE 1 Seven-parameter registration via FLIRT was used within subjects. For each imaging set, TSE images were registered to MRF maps via synthetic TSE. MRF was then registered to MPRAGE via synthetic MPRAGE. MPRAGE images from each imaging set and scanner were then registered to the "original" MPRAGE series, which was then nonlinearly warped via FNIRT to MNI-152 space. The linear and nonlinear transformations saved for each scanner/set/series combination were then inverted and used to generate set-specific atlas label maps. These maps were used to sample and save voxel buffers for all atlas regions. FLIRT, first linear registration; FNIRT, first nonlinear registration; MNI, Montreal Neurological Institute; MRF, MR fingerprinting; TSE, turbo spin echo.

mean values of registered regions across paired samples from the same subject. The unweighted mean and SD of the differences in mean value across all subject and regions is reported for each modality and case. Additionally, because regions have vastly different volumes, a weighted mean and weighted SD[41] for the performance of each modality across subjects and regions was performed. The resulting bias and agreement metrics were then compared to establish the relative stability of each imaging approach.

The resulting aggregate data from each region and image type were then used to generate Bland–Altman plots. The plots compared the stability of regional mean values for all qualitative and quantitative image series, across all subjects in the intrasession, intersession, and interscanner cases.

3 | RESULTS

Sample MRF maps, resulting synthetic images, and the associated MPRAGE (T_1w) and TSE (T_2w) images are shown in Figure 2 for a representative subject. Other subjects' sample maps are available in the Supporting Information Figures S3-S12. Maps visualizing automated registration pipeline performance for the same representative subject across sessions and scanners are illustrated in Figure 3. Similar maps for all the subjects are available in the Supporting Information Figures S13-S22. Bulk registration errors were not seen across any of the image series, and no manual intervention or registration correction was performed outside of the automated pipeline.

Atlas registration performance maps were generated for each subject to demonstrate the performance of the FNIRT nonlinear registration between MNI-152-2 mm space and the first scanner's original MPRAGE image series. The results for the same representative subject and the other subjects are in Figure 4 and the Supporting Information S23-S32, respectively. The aggregated mean T_1 and T_2 values and their respective SDs derived from 3D MRF for the examined MNI-152 atlas regions across all sessions on both scanners for all subjects are presented in Table 1. Bland–Altman plots were prepared comparing the

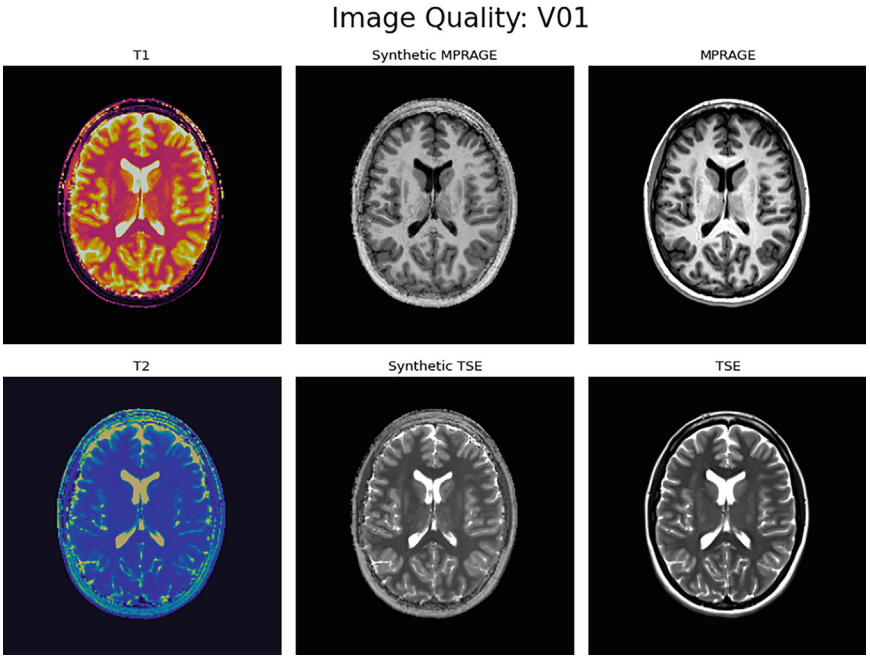

FIGURE 2 Synthetic MPRAGE (T_1w) and TSE (T_2w) images generated from MRF maps were used as the input to the registration pipeline. Both synthetic qualitative contrasts showed similar intraregional contrast compared to the MPRAGE and TSE imaging acquired. T_{1W}, T_1-weighted; T_{2W}, T_2-weighted.

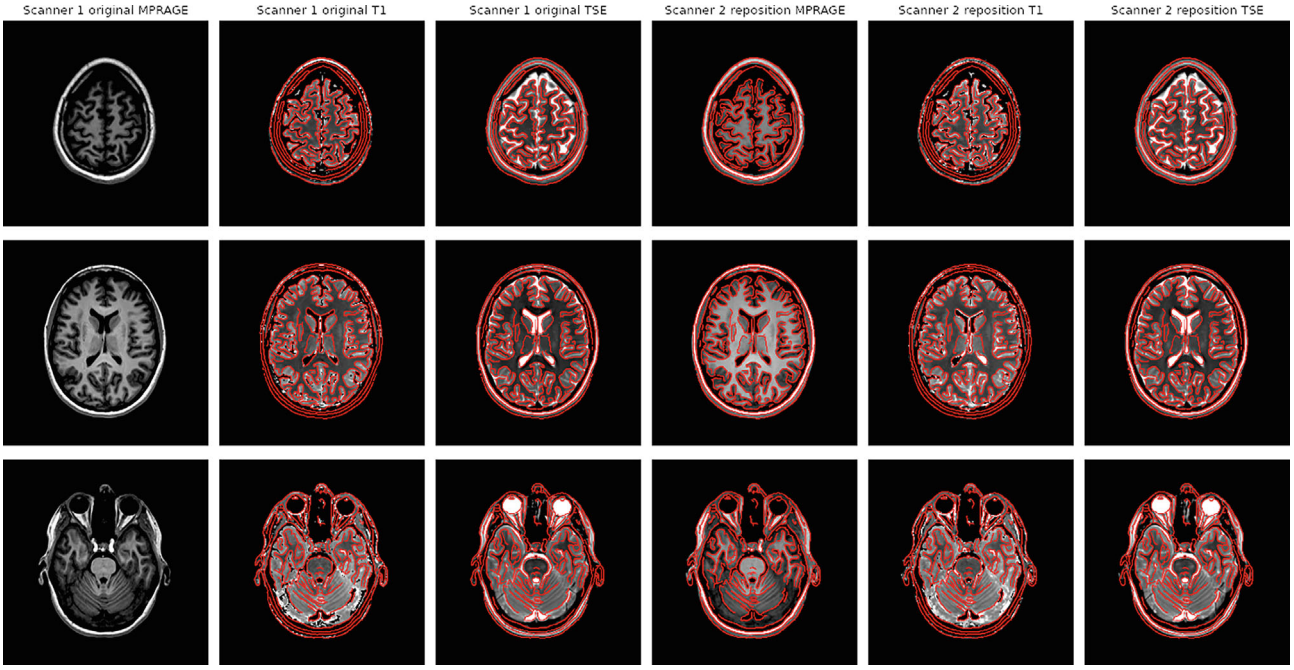

FIGURE 3 The results of canny edge detection ($\sigma = 1$) applied to the first scanner's original MPRAGE series projected over the first scanner's original T_1 map (represented here in grayscale) and TSE series and applied to the second scanners reposition MPRAGE, T_1 map, and TSE series.

repeatability of 3D-MRF, TSE, and MPRAGE acquisitions using regional means compared between immediate repetitions (Figure 5) and subject repositions on the same scanner (Figure 6).

Reproducibility was tested by comparing all combinations of subjects, scanners, and sets (Figure 7). Table 2 summarizes the mean percent differences and SDs of 3D-MRF, TSE, and MPRAGE across the intrasession, intersession, and interscanner cases. In all cases except TSE, smaller regions have higher variability and skew the aggregate reproducibility unproportionally. When the size of each region is included in the estimation of the weighted SD[41] for the overall repeatability and reproducibility (Table 3), the performance of T_1, T_2, and MPRAGE better reflect the visual and histogrammatic similarity of the regions.

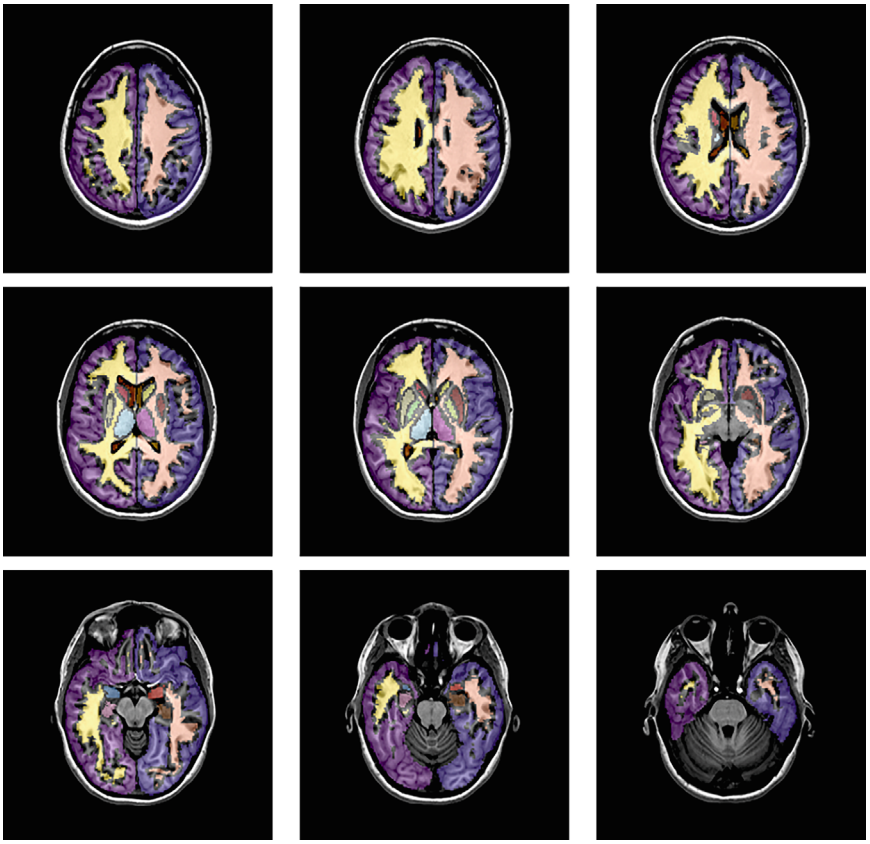

FIGURE 4 The atlas label maps from MNI space were inversely warped and eroded to remove mislabeling artifacts introduced by nonlinear interpolation of integer label values. The resulting atlas region overlays were colorized according to the legend used in the subsequent Bland Altman plots.

TABLE 1 Fingerprinting-derived T_1 and T_2 values of MNI-152 regions.

Tissue compartment	T_{1mean} (ms)	σ_{T1} (ms)	T_{2mean} (ms)	σ_{T2} (ms)
Cerebral white matter	914.1	40.5 (4.4%)	45.9	2.3 (5.0%)
Cerebral cortex	1889.4	129.5 (6.9%)	124.6	15.8 (12.6%)
Lateral ventricles	4341.4	447.9 (10.3%)	467.0	26.7 (5.7%)
Thalamus	1175.5	73.7 (6.3%)	49.6	3.7 (7.5%)
Caudate	1338.4	66.1 (4.9%)	51.5	6.0 (11.6%)
Putamen	1234.0	53.6 (4.3%)	45.0	4.6 (10.1%)
Pallidum	938.7	53.2 (5.7%)	30.3	3.3 (11.0%)
Hippocampus	1621.8	93.1 (5.7%)	76.0	9.9 (13.0%)
Amygdala	1463.4	55.2 (3.8%)	62.5	3.6 (5.7%)

Note: Measured mean T_1 and T_2 values and associated SDs in the selected tissue compartments, derived from 3D MRF maps across all sessions on both scanners for all subjects.
Abbreviations: MNI, Montreal Neurological Institute; MRF, MR fingerprinting.

4 | DISCUSSION

This study aimed to provide an online MRF reconstruction and analysis framework while also investigating the reproducibility of MRF and conventional weighted imaging. From the start, we developed a traceable and online 3DMRF reconstruction that outputs DICOMs directly to the scanner. We then evaluated our fully automatic online 3D MRF reconstruction, as well as our cross-modality registration and analysis pipeline, to determine whether in vivo 3D MRF repeatability and reproducibility meets or exceeds that of vendor product MPRAGE and TSE.

The brains of healthy volunteers were imaged using MPRAGE, TSE, and MRF protocols across multiple sets following varying perturbations to the subject, repeated on different scanners on multiple days. The consistent volume-weighted biases and SDs in Table 3 indicate that the T_1 and T_2 values generated by in vivo 3D-MRF

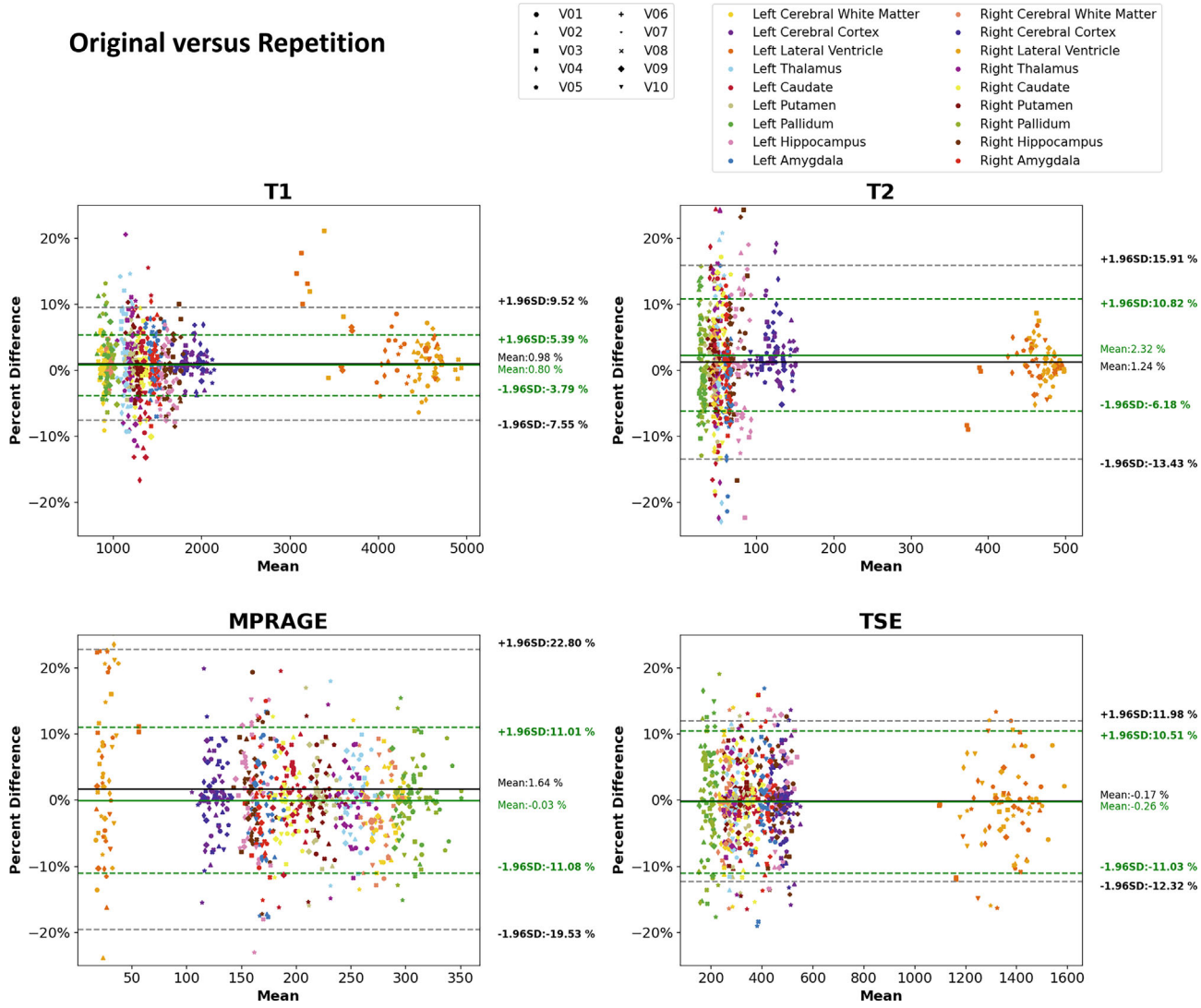

FIGURE 5 The repeatability between the "original" and "repetition" sets on each scanner was compared for all subjects. This represents a same-session test–retest on the same scanner hardware because the subject remained in the bore, and the scanner has not performed adjustments between acquisitions. The gray and green dashed lines show the confidence intervals for the mean and weighted (by region volume) mean cases.

were repeatable whether a scan is repeated immediately (T_1: 0.80% ± 2.34%, T_2: 2.32% ± 4.34%) or with a reposition of the subject on the same scanner (T_1: 0.30% ± 1.90%, T_2: 2.07% ± 3.20%), and reproducible on a different scanner on a different day (T_1: −1.02% ± 2.21%, T_2: −3.24% ± 3.89%). Most importantly, regardless of scanner and session, the intrasubject variations of both T_1 and T_2 were found to be lower than the variations within T_1 and T_2 regions across the sampled population shown in Table 1.

Considering the observed negligible differences between intrascanner and interscanner variations, the apparent reproducibility of in vivo 3D-MRF offers multiple opportunities: data from many sessions, scanners, and sites can potentially be treated as a single dataset for harmonized analysis. Similarly, structural or statistical intrasubject comparisons are valid across scanners or sessions for the proposed 3D-MRF pipeline without any additional data regularization steps. The same is not true of the baseline product imaging methods that we tested here.

Because the direct output of MRF is an actual quantifiable measurement, additional criteria and opportunities in terms of repeatability and reproducibility need to be satisfied, which might not apply to conventional weighted imaging at the scanner output level.

The reproducibility of 3D MRF was investigated in various common clinical situations in a traceable framework via bounds of uncertainty that we set and explored through cortical and subcortical regional mean values. Defining and monitoring traceability and uncertainty with structured boundaries will be useful for careful

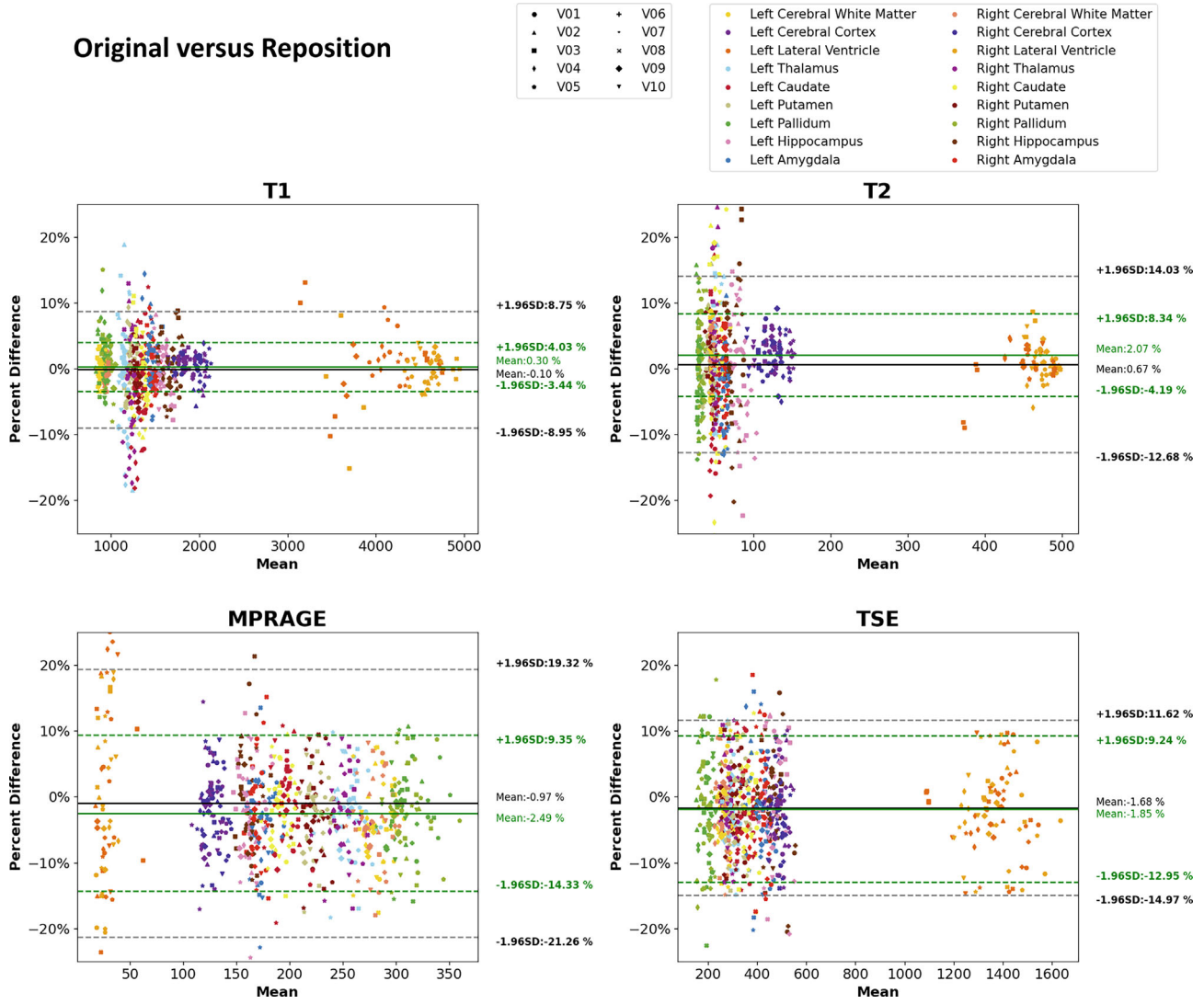

FIGURE 6 The repeatability between the "original" and "reposition" sets on each scanner were compared for all subjects. Between acquisitions, the subject was asked to leave the bore and walk around before being repositioned inside, triggering a reshimming of the system and frequency adjustments. This represents a cross-session test–retest on the same scanner hardware. The gray and green dashed lines show the confidence intervals for the mean and weighted (by region size) mean cases.

integration into clinical settings and to further ensure physician confidence.[28]

4.1 | Traceability

The data analysis for this study recognized the importance of traceable research documenting all the stages of a study from data acquisition to presentation of the results. This structured record becomes functionally important as a foundation for studies involving multiple scanners, sites, and even across vendors. In our framework, we defined and ensured traceability at the data acquisition, reconstruction, and post-processing steps. After the fully integrated acquisition and automated online reconstruction, the DICOM images are fed back to the MR console via the FIRE interface prototype. This allows MR radiologists and technologists to interact in real time with MRF quantitative maps in their preferred environments (PACS or MR console) in the vendor coordinate system, with full DICOM capability.

The next step in the traceability chain was the analysis pipeline for automated post-processing of the DICOM images. The cloud-based and version-controlled registration and regional analysis pipeline can support future applications and more complex analyzes that use MRF in longitudinal or multi-center large scale studies. This scale of traceability for every step ensures that comparisons across different MRF variants, sites, scanners, and even vendors are possible and valid. Due to the pipeline's

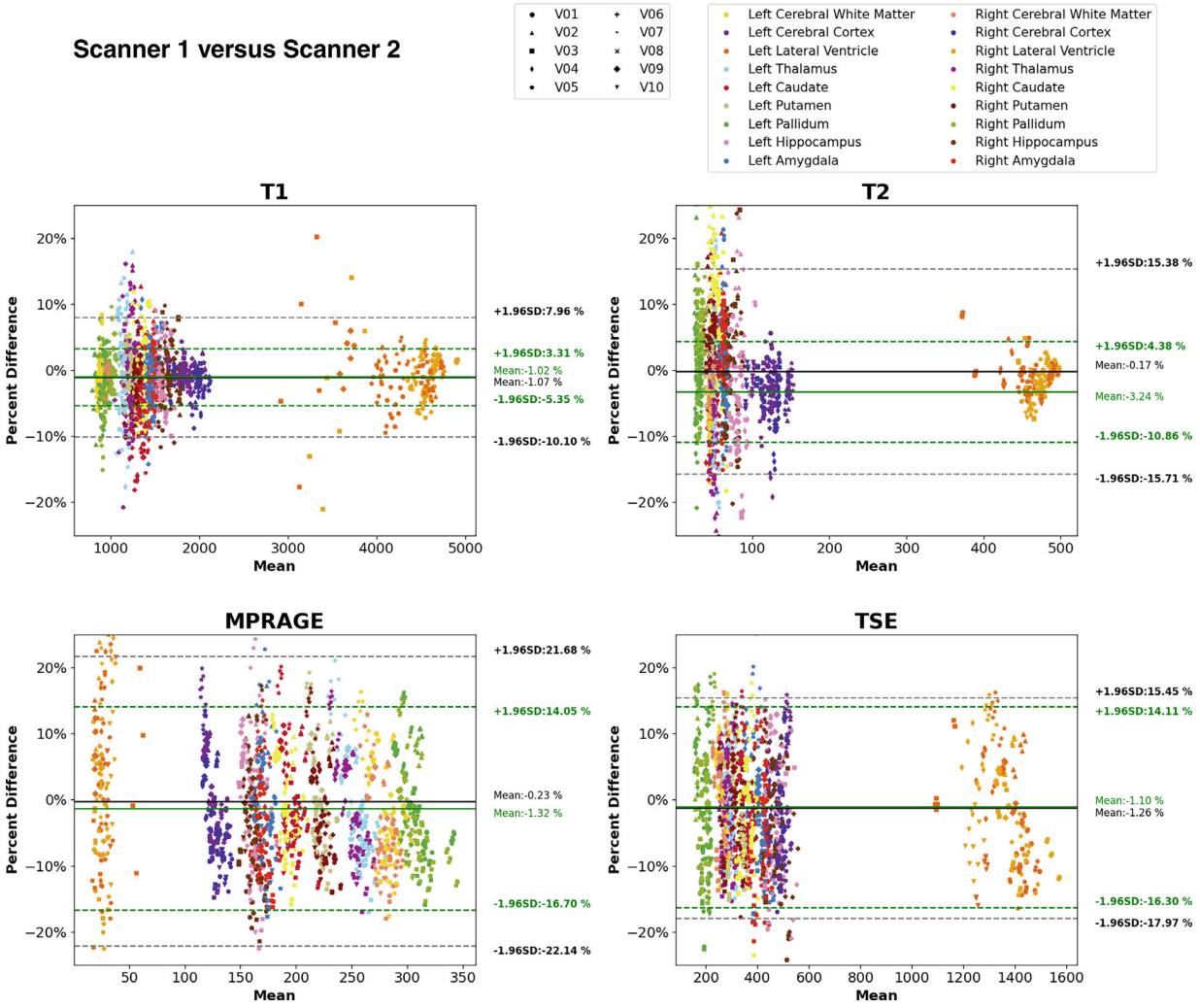

FIGURE 7 The reproducibility across all combinations of subjects, scanners, and sets was compared. The data included is the full matrix of nine combinations of original, repetition, and reposition sets across both scanners. The gray and green dashed lines show the confidence intervals for the mean and weighted (by region size) mean cases.

flexible infrastructure, other tools or software packages can also be integrated at any level. In prior studies, the use of offline reconstructions and manual analysis pipelines often impeded the use of cross-dataset registration and statistical tools. This resultant loss of significant portions of the metadata present in clinical sequences such as position, scale, and subject identifiers made it difficult for cross-modal comparisons. Holding the history of the pipeline accountable could further increase the confidence in MRF and help usher in its adoption in clinical settings.

4.2 | Uncertainty

After the traceability chain was established and documented for in vivo quantitative mapping with MRF, the next step evaluated the associated uncertainty of the quantitative maps.

Quantitative tissue properties and imaging biomarkers are only meaningful when measurement uncertainties are provided. The goal of the uncertainty evaluation was not to define a measure of error for the quantitative maps but rather to provide guidance for the decision-making process based on the maps. Like UK Biobank,[42] large scale studies that also include reproducible MRF quantitative maps can be enabled to define population-based normative tissue property values with known uncertainties. Eventually, decisions can be made about the individual patients directly at the MR console with confidence, such as manually or automatically flagging significant findings with respect to a population reference (obtained from large scale studies) or a past measurement of the same subject (longitudinal).

The MRF reconstruction pipeline inputted the B_1 maps acquired and reconstructed with the standard vendor sequence and corrected for the B_1 inhomogeneity by

TABLE 2 Repeatability and reproducibility performance by modality, unweighted.

Mean ± SD (bias ± agreement)	Intrascanner		Interscanner
	Same-session	Cross-session	Cross-session
T_1 (%)	0.98 ± **4.36**	−0.10 ± **4.51**	−1.07 ± **4.61**
T_2 (%)	1.24 ± **7.48**	0.67 ± **6.81**	−0.17 ± **7.93**
MPRAGE (%)	1.64 ± **10.80**	−0.97 ± **10.35**	−0.23 ± **11.18**
TSE (%)	−0.17 ± **6.20**	−1.68 ± **6.78**	−1.26 ± **8.52**

Note: Summary of mean value intrascanner repeatability and interscanner reproducibility for selected MNI-152 regions between conventional MPRAGE and TSE imaging versus 3D-MRF. Standard deviation, the primary criteria for each method's repeatability, is shown in bold.
Abbreviation: TSE, turbo-spin-echo.

TABLE 3 Repeatability and reproducibility performance by modality, weighted by region sizes.

Weighted mean ± weighted SD (bias ± agreement)	Intrascanner		Interscanner
	Same-session	Cross-session	Cross-session
T_1 (%)	0.80 ± **2.34**	0.30 ± 1.**90**	−1.02 ± **2.21**
T_2 (%)	2.32 ± **4.34**	2.07 ± **3.20**	−3.24 ± **3.89**
MPRAGE (%)	−0.03 ± **5.63**	−2.49 ± **6.04**	−1.32 ± **7.84**
TSE (%)	−0.26 ± **5.49**	−1.85 ± **5.66**	−1.10 ± **7.76**

Note: Summary of weighted mean value intrascanner repeatability and interscanner reproducibility for selected MNI-152 regions between conventional MPRAGE and TSE imaging versus 3D-MRF. Regional differences are weighted by voxel population size on a per-set, per-subject basis to represent the whole-brain aggregate performance of each acquisition approach. Standard deviation, the primary criteria for each method's repeatability, is shown in bold.

matching each voxel's fingerprint to the portion of the dictionary with the voxel's relative B_1. Among the previous MRF repeatability and reproducibility studies, only Kőzdőrfer et al. corrected for B_1 inhomogeneities.[21] B_1 correction eliminates the bias from MRF T_2 maps improving accuracy[43]; thus, it is expected to lower the variability between sessions. Different scanners and software versions can have varying limits and adjustments of the RF power and can affect the T_2 contrast for weighted imaging and MRF time series unless accounted for. MRF data shows reduced variability compared to conventional images and MRF's ability to consider variable and inhomogeneous B_1 could be a factor.

Previously, some MRF repeatability and reproducibility studies used 2D MRF[21,22] rather than the volumetric 3D acquisitions commonly used in neuroradiological clinical practice. Besides using 2D acquisitions, manually drawn ROIs often formed the basis for intrasession, intersession, and interscanner comparisons of the resulting maps.[21,26] These ROIs potentially introduced errors due to human intervention in the processing pipeline, while also reducing the scalability of the study and reducing the potential for intermodality comparisons of MRF results against clinical standard contrasts provided by vendor product implementations. For these studies, 2D in vivo brain repeatability was shown to be 2%–3% for T_1 and 5%–8% for T_2[22]; 2%–3% for T_1 and 3%–8% for T_2.[21] Reproducibility was slightly lower for both studies: 3%–8% for T_1 and 8%–14% for T_2[22]; 3.4% for T_1 and 8% for T_2.[21]

Two other studies that investigated the repeatability and reproducibility of a 3D in vivo brain MRF data with different analyzes reported similar results.[23,25] Buonincontri et al. based the analysis on average GM, WM, and CSF relaxation times and reported <2% T_1 and <5% repeatability, and 6% GM T_1 and 10% GM T_2 reproducibility.[22] With automated segmentation of the same 3D MRF data into cortical and subcortical regions, Fujita et al. reported repeatability (cortical: T_1 4% and T_2 6%, subcortical: T_1 1.3% and T_2 5%) and reproducibility of T_1 and T_2 (cortical: T_1 2.2% and T_2 6.7%, subcortical: T_1 3.2% and T_2 5.8%), as well as cortical thickness and subcortical volumes.[25]

4.3 | Qualitative imaging

MPRAGE and TSE are common acquisition schemes for diagnostic MRI of the brain and are frequently used in clinical practice. Both techniques are qualitative acquisitions that are adjusted to maximize contrast between specific tissues and normalized by proprietary vendor reconstructions; therefore, they fare not expected to have reproducible intensity for certain tissues.

As a result of changes in receiver tuning, coil loading, and image autoscaling, we expected a linear bias (mean shift in reproducibility) would still govern any inter- and intrasession regional variations.

The result, however, was a muddling of the underlying interregional contrasts, evidenced by the low bias and high variability within an individual session (MPRAGE: −0.03% ± 5.63%, TSE: −0.26% ± 5.49%), across sessions (MPRAGE: −2.49% ± 6.04%, TSE: −1.85% ± 5.66%), and across scanners (MPRAGE: −1.32% ± 7.84%, TSE: −1.10% ± 7.76%). Because MPRAGE images form the basis for the registration approach, the comparatively lower reproducibility of MPRAGE is not due to poor within-subject or MNI registration because any registration errors that may contribute to variability in MPRAGE image sets would have propagated to 3D-MRF and TSE images.

Whereas in some cases the bias of the differences was lower for qualitative modalities than the MRF-derived

values, the SD of the differences was always higher for qualitative modalities than MRF. Low-to-moderate bias, especially linear bias across a measurement parameter, can be easily corrected via calibration during standard scanner quality assurance and maintenance procedures. In fact, both qualitative methods assessed in this study already benefit from post hoc calibration due to the vendor applying both surface-coil intensity normalization and noise prewhitening in the product image reconstruction pipeline. The evaluated MRF reconstruction does not implement any form of scanner- or sequence-specific normalization.

The contrast variability of the weighted images must be interpreted by radiologists with appropriate window and level adjustments to accommodate reading images acquired in different sessions.

Additionally, most qualitative or quantitative post-processing methods using raw DICOM images could be affected by the contrast variability because the algorithms behind the techniques rely on certain contrast between tissues.[44] Eck et al.[45] showed that many radiomics features, extracted from TSE, are not robust when image contrast, resolution, and acceleration factors are changed. Scanner software upgrades might cause additional problems for longitudinal studies due to B_1 variations and signal saturation.[19] Reproducibility is critical for longitudinal comparisons of disease states and for the training of direct or convolutional inference networks based on value-normalized large-scale datasets, which are becoming more common. Because the MRF quantitative maps are found to be more reproducible than MPRAGE or TSE in common clinical scenarios, future cross-scanner or cross-site large scale studies would be justified in using MRF, instead of or additional to conventional imaging. Automated image analysis tools, such as FSL and FreeSurfer, can also be expanded to operate on quantitative MRF maps.

4.4 | Study limitations and future work

This study investigated the reproducibility of MRF and conventional imaging on a pixel basis rather than regional volume or other structural metrics. T_1w contrast is the sole input for most open-source brain analysis tools; yet, an analysis engine could be optimized to extract regions from each contrast/relaxation map separately. An additional consideration is that synthetic weighted contrast generation from quantitative maps, required to run most analysis tools, is not straightforward, and optimization of these methods is out of the scope of this paper.

A linear regression lookup table approach was used in lieu of direct Bloch simulation approaches because accurate proton density maps were not immediately available from the online reconstruction process used in this study. Direct substitution of M0 as proton density yielded inconsistent contrasts versus ground truth MPRAGE, resulting in the development of the proposed 2-to-1 lookup table approach, which substantially improved robustness of the automated registration and skull stripping processes needed in this work and was therefore used in the described limited capacity. The synthetic images generated with the regression network provided an MPRAGE- and TSE-like contrast only for the registration purposes. As illustrated in Figures 2 and 3, coupled with the same anatomy and similar position between modalities, the synthetic images had adequate tissue contrast to ensure an accurate registration within a set. The reliability of the resulting maps suggests that synthetic contrast generation based on MRF maps may allow for system- and session-agnostic T_1/T_2 weighted contrasts with reproducibility magnitudes exceeding the current clinical standard. Future work will focus on creating better synthetic weighted images, which could be used to compare reproducibility at the structural volume or biomarker level for MRF and conventional weighted imaging with established community tools.

5 | CONCLUSIONS

To improve traceability with minimal manual interventions, we presented a fully automated data acquisition, reconstruction, and analysis pipeline for 3D-MRF. Reproducibility of quantitative MRF maps and qualitative MPRAGE and TSE images were evaluated over sessions and scanners by comparing mean values from MNI brain atlas regions. The proposed MRF acquisition, reconstruction, and analysis pipeline was found to be more repeatable and reproducible than qualitative methods, which should open the door to wider clinical adoption and widespread use.

CONFLICT OF INTEREST STATEMENT

Authors Yong Chen, Dan Ma, and Mark A. Griswold have licensed patents to Siemens related to MR fingerprinting and receive royalties. Author Michael Hansen is an employee of Microsoft Research. Author Kelvin Chow is an employee of Siemens Medical Solutions USA.

DATA AVAILABILITY STATEMENT

The code that supports the findings of this study are available at https://doi.org/10.5281/zenodo.8184908.[46] All

image data, in NIfTI format, are available at https://doi.org/10.5281/zenodo.8183344.[47]

ORCID

Andrew Dupuis https://orcid.org/0000-0002-3714-4812
Yong Chen https://orcid.org/0000-0001-6183-2693
Michael Hansen https://orcid.org/0000-0002-8087-8731
Kelvin Chow https://orcid.org/0000-0003-0698-1746
Jessie E. P. Sun https://orcid.org/0000-0003-4659-9883
Chaitra Badve https://orcid.org/0000-0001-8547-8019
Dan Ma https://orcid.org/0000-0003-1664-9579
Mark A. Griswold https://orcid.org/0000-0002-3011-6747
Rasim Boyacioglu https://orcid.org/0000-0002-3099-2640

REFERENCES

1. Seiler A, Nöth U, Hok P, et al. Multiparametric quantitative MRI in neurological diseases. *Front Neurol*. 2021;12:640239.
2. Ma D, Gulani V, Seiberlich N, et al. Magnetic resonance fingerprinting. *Nature*. 2013;495:187-192.
3. Springer E, Cardoso PL, Strasser B, et al. MR fingerprinting-a radio genomic marker for diffuse gliomas. *Cancers (Basel)*. 2022;14:723.
4. Hung S-C, Chen Y, Yap P-T, Lin W. Magnetic resonance fingerprinting of the pediatric brain. *Magn Reson Imaging Clin N Am*. 2021;29:605-616.
5. Mostardeiro TR, Panda A, Witte RJ, et al. Whole-brain 3D MR fingerprinting brain imaging: clinical validation and feasibility to patients with meningioma. *MAGMA*. 2021;34:697-706.
6. Ma D, Jiang Y, Chen Y, et al. Fast 3D magnetic resonance fingerprinting for whole-brain coverage. *Magn Reson Med*. 2018;79:2190-2197.
7. Chen Y, Chen M-H, Baluyot KR, et al. MR fingerprinting enables quantitative measures of brain tissue relaxation times and myelin water fraction in the first five years of life. *Neuroimage*. 2019;186:782-793.
8. Eck BL, Yim M, Hamilton JI, et al. Cardiac magnetic resonance fingerprinting: potential clinical applications. *Curr Cardiol Rep*. 2023;25:119-131.
9. Hamilton JI, Pahwa S, Adedigba J, et al. Simultaneous mapping of T_1 and T_2 using cardiac magnetic resonance fingerprinting in a cohort of healthy subjects at 1.5T. *J Magn Reson Imaging*. 2020;52:1044-1052.
10. Fujita S, Sano K, Cruz G, et al. MR fingerprinting for liver tissue characterization: a histopathologic correlation study. *Radiology*. 2023;306:150-159.
11. Velasco C, Cruz G, Jaubert O, Lavin B, Botnar RM, Prieto C. Simultaneous comprehensive liver T_1, T_2, T1rho, and fat fraction characterization with MR fingerprinting. *Magn Reson Med*. 2022;87:1980-1991.
12. Huang SS, Boyacioglu R, Bolding R, MacAskill CJ, Chen Y, Griswold MA. Free-breathing abdominal magnetic resonance fingerprinting using a pilot tone navigator. *J Magn Reson Imaging*. 2021;54:1138-1151.
13. Lo LW, Panda A, Jiang Y, Ahad J, Gulani V, Seiberlich N. MR fingerprinting of the prostate. *MAGMA*. 2022;35:557-571.
14. MacAskill CJ, Markley M, Farr S, et al. Rapid B_1-insensitive MR fingerprinting for quantitative kidney imaging. *Radiology*. 2021;300:380-387.
15. Hermann I, Chacon-Caldera J, Brumer I, et al. Magnetic resonance fingerprinting for simultaneous renal T_1 and T1rho mapping in a single breath-hold. *Magn Reson Med*. 2020;83:1940-1948.
16. NEMA. Determination of Signal-to-Noise Ratio (SNR) in Diagnostic Magnetic Resonance Imaging. *NEMA Standards Publication MS 1-2008*. National Electrical Manufactures Association; 2020:R2014-R2020.
17. NEMA. Determination of Two-Dimensional Geometric Distortion in Diagnostic Magnetic Resonance Images. *NEMA Standards Publication MS 2-2008*. National Electrical Manufactures Association; 2020:R2014-R2020.
18. NEMA. Determination of Image Uniformity in Diagnostic Magnetic Resonance Images. *NEMA Standards Publication MS 3-2008*. National Electrical Manufactures Association; 2020:R2014-R2020.
19. Keenan KE, Gimbutas Z, Dienstfrey A, Stupic KF. Assessing effects of scanner upgrades for clinical studies. *J Magn Reson Imaging*. 2019;50:1948-1954.
20. Jiang Y, Ma D, Keenan KE, Stupic KF, Gulani V, Griswold MA. Repeatability of magnetic resonance fingerprinting T_1 and T_2 estimates assessed using the ISMRM/NIST MRI system phantom. *Magn Reson Med*. 2017;78:1452-1457.
21. Körzdörfer G, Kirsch R, Liu K, et al. Reproducibility and repeatability of MR fingerprinting relaxometry in the human brain. *Radiology*. 2019;292:429-437.
22. Buonincontri G, Biagi L, Retico A, et al. Multi-site repeatability and reproducibility of MR fingerprinting of the healthy brain at 1.5 and 3.0T. *Neuroimage*. 2019;195:362-372.
23. Buonincontri G, Kurzawski JW, Kaggie JD, et al. Three dimensional MRF obtains highly repeatable and reproducible multi-parametric estimations in the healthy human brain at 1.5T and 3T. *Neuroimage*. 2021;226:117573.
24. Fujita S, Buonincontri G, Cencini M, et al. Repeatability and reproducibility of human brain morphometry using three-dimensional magnetic resonance fingerprinting. *Hum Brain Mapp*. 2021;42:275-285.
25. Fujita S, Cencini M, Buonincontri G, et al. Simultaneous relaxometry and morphometry of human brain structures with 3D magnetic resonance fingerprinting: a multicenter, multiplatform, multifield-strength study. *Cerebral Cortex*. 2022;33:729-739.
26. Lo W-C, Bittencourt LK, Panda A, et al. Multicenter repeatability and reproducibility of MR fingerprinting in phantoms and in prostatic tissue. *Magn Reson Med*. 2022;88:1818-1827.
27. Voelker MN, Kraff O, Goerke S, et al. The traveling heads 2.0: multicenter reproducibility of quantitative imaging methods at 7 Tesla. *Neuroimage*. 2021;232:117910.
28. Cashmore MT, McCann AJ, Wastling SJ, McGrath C, Thornton J, Hall MG. Clinical quantitative MRI and the need for metrology. *Br J Radiol*. 2021;94:20201215.
29. Jiang Y, Ma D, Seiberlich N, Gulani V, Griswold MA. MR fingerprinting using fast imaging with steady state precession (FISP) with spiral readout. *Magn Reson Med*. 2014;74:1621-1631.
30. Duyn JH, Yang Y, Frank JA, van der Veen JW. Simple correction method for k-space trajectory deviations in MRI. *J Magn Reson*. 1998;132:150-153.

31. Chow K, Kellman P, Xue H. Prototyping image reconstruction and analysis with FIRE. *Proceedings of SCMR Annual Scientific Sessions*, Virtual Meeting. SCMR; 2021.
32. Inati SJ, Naegele JD, Zwart NR, et al. ISMRM raw data format: a proposed standard for MRI raw datasets. *Magn Reson Med*. 2017;77:411-421.
33. McGivney DF, Pierre E, Ma D, et al. SVD compression for magnetic resonance fingerprinting in the time domain. *IEEE Trans Med Imaging*. 2014;33:2311-2322.
34. Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J Neurosci Methods*. 2016;264:47-56.
35. Brett M, Markiewicz C, Hanke M, et al. NiBabel: Access a cacophony of neuro-imaging file formats. 2023. Accessed November 9, 2022. Available at: https://nipy.org/nibabel/
36. Gorgolewski K, Burns CD, Madison C, et al. Nipype: a flexible, lightweight, and extensible neuroimaging data processing framework in python. *Front Neuroinform*. 2011;5:13.
37. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *Neuroimage*. 2012;62:782-790.
38. Developers TensorFlow. TensorFlow (v2.11.0). Zenodo; 2022. doi:10.5281/zenodo.10126399
39. Jenkinson M, Smith S. A global optimization method for robust affine registration of brain images. *Med Image Anal*. 2001;5:143-156.
40. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 2002;17:825-841.
41. Heckert NA, Filliben JJ. NIST handbook 148: Dataplot reference manual, volume 2: let subcommands and library functions. *National Institute of Standards and Technology Handbook Series*. NIST; 2003. Accessed November 19, 2022. https://www.itl.nist.gov/div898/software/dataplot/refman2/homepage.htm and https://www.itl.nist.gov/div898/software/dataplot/refman2/ch2/weightsd.pdf
42. Alfaro-Almagro F, Jenkinson M, Bangerter NK, et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage*. 2018;166:400-424.
43. Ma D, Coppo S, Chen Y, et al. Slice profile and B_1 corrections in 2D magnetic resonance fingerprinting. *Magn Reson Med*. 2017;78:1781-1789.
44. Leung KK, Clarkson MJ, Bartlett JW, et al. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *Neuroimage*. 2010;50:516-523.
45. Eck B, Chirra PV, Muchhala A, et al. Prospective evaluation of repeatability and robustness of radiomic descriptors in healthy brain tissue regions in vivo across systematic variations in T_2-weighted magnetic resonance imaging acquisition parameters. *J Magn Reson Imaging*. 2021;54:1009-1021.
46. Dupuis A, Chen Y, Griswold MA, Boyacioglu R. *Python code for quantifying 3D magnetic resonance fingerprinting (3D-MRF) reproducibility across subjects, sessions, and scanners automatically using MNI atlases*. Zenodo; 2023. doi:10.5281/zenodo.8184908
47. Dupuis A, Chen Y, Hansen M, et al. *Intrasession, intersession, and interscanner qualitative and quantitative MRI datasets of healthy brains at 3.0T*. Zenodo; 2023. Accessed July 25, 2023. https://zenodo.org/records/8183344

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Figure S1. Sequence flip angles for each of the 960 timepoints per partition used for the magnetic resonance fingerprinting (MRF) acquisition. TR and TE were held constant at 10.5 and 2.2 ms respectively. The flip angle was smoothly varied in a pseudo sinusoidal pattern between 0 and 57 degrees. Forty-eight spiral interleaves were acquired in a wrapping sequential pattern that repeats 20 times.

Figure S2. System architecture for the online Kubernetes-based 3D-MRF Reconstruction. Data is sent from the scanner to Azure via an SSH tunnel between the scanner's host and an SSH jump pod within the Kubernetes cluster. One or multiple GPU-enabled nodes then share the load of storing temporary dependencies and reconstructing datasets that arrive on the cluster. Logs are stored to a persistant Prometheus appliance for debugging and monitoring purposes.

Figure S3. Image quality for V01. Synthetic MPRAGE and TSE images generated from MRF maps were used as the input to the registration pipeline. Both synthetic qualitative contrasts showed similar intraregional contrast compared to the MPRAGE and TSE imaging acquired.

Figure S4. Image quality for V02.

Figure S5. Image quality for V03.

Figure S6. Image quality for V04.

Figure S7. Image quality for V05.

Figure S8. Image quality for V06.

Figure S9. Image quality for V07.

Figure S10. Image quality for V08.

Figure S11. Image quality for V09.

Figure S12. Image quality for V10.

Figure S13. Linear registration performance for V01. The results of canny edge detection applied to the first scanner's original MPRAGE series projected over the first scanner's original T_1 and TSE series and the second scanner's reposition MPRAGE, T_1, and TSE series.

Figure S14. Linear registration performance for V02.

Figure S15. Linear registration performance for V03.

Figure S16. Linear registration performance for V04.

Figure S17. Linear registration performance for V05.

Figure S18. Linear registration performance for V06.

Figure S19. Linear registration performance for V07.

Figure S20. Linear registration performance for V08.

Figure S21. Linear registration performance for V09.

Figure S22. Linear registration performance for V10.

Figure S23. Atlas registration quality for V01. Atlas label maps from Montreal Neurological Institute (MNI) space were inversely warped and eroded to remove mislabeling

artifacts introduced by nonlinear interpolation of integer label values. The resulting atlas region overlays were colorized according to the legend used in the subsequent Bland Altman plots.

Figure S24. Atlas registration quality for V02.
Figure S25. Atlas registration quality for V03.
Figure S26. Atlas registration quality for V04.
Figure S27. Atlas registration quality for V05.
Figure S28. Atlas registration quality for V06.
Figure S29. Atlas registration quality for V07.
Figure S30. Atlas registration quality for V08.
Figure S31. Atlas registration quality for V09.
Figure S32. Atlas registration quality for V10.

Data S1. Sequence parameters for the SSFP MRF acquisition used in this study. Repetition Time (TR), Echo Time (TE), Flip Angle (FA), RF Phase (PH), and readout spiral (ID) are specified for each of the 960 acquisition timepoints in units of microseconds (TR, TE) and degrees (FA, PH). ID refers to individual interleaves of the factor 48 undersampled spiral trajectory, with each ID referring to a spiral arm rotated by ID * (360/48) degrees from the base spiral arm shape (ID = 0).

How to cite this article: Dupuis A, Chen Y, Hansen M, et al. Quantifying 3D MR fingerprinting (3D-MRF) reproducibility across subjects, sessions, and scanners automatically using MNI atlases. *Magn Reson Med*. 2024;1-15. doi: 10.1002/mrm.29983