Hindawi
Radiology Research and Practice
Volume 2021, Article ID 1898461, 18 pages
https://doi.org/10.1155/2021/1898461

Research Article

Comparison of Phase Estimation Methods for Quantitative Susceptibility Mapping Using a Rotating-Tube Phantom

Kathryn E. Keenan,[1] **Ben P. Berman**,[2] **Slávka Rýger**,[1] **Stephen E. Russek**,[1] **Wen-Tung Wang**,[3] **John A. Butman**,[4] **Dzung L. Pham**,[3] **and Joseph Dagher**[2]

[1]*National Institute of Standards and Technology, Physical Measurement Laboratory, 325 Broadway, Boulder, CO 80305, USA*
[2]*The MITRE Corporation, 7515 Colshire Dr, McLean, VA 22102, USA*
[3]*Henry M. Jackson Foundation, 10 Center Drive, Bethesda, MD 20892, USA*
[4]*Clinical Center, National Institutes of Health, 10 Center Drive, Bethesda, MD 20814, USA*

Correspondence should be addressed to Ben P. Berman; bberman@mitre.org

Received 16 July 2021; Accepted 8 October 2021; Published 24 November 2021

Academic Editor: André Luiz Ferreira Costa

Quantitative Susceptibility Mapping (QSM) is an MRI tool with the potential to reveal pathological changes from magnetic susceptibility measurements. Before phase data can be used to recover susceptibility ($\Delta\chi$), the QSM process begins with two steps: data acquisition and phase estimation. We assess the performance of these steps, when applied without user intervention, on several variations of a phantom imaging task. We used a rotating-tube phantom with five tubes ranging from $\Delta\chi = 0.05$ ppm to $\Delta\chi = 0.336$ ppm. MRI data was acquired at nine angles of rotation for four different pulse sequences. The images were processed by 10 phase estimation algorithms including Laplacian, region-growing, branch-cut, temporal unwrapping, and maximum-likelihood methods, resulting in approximately 90 different combinations of data acquisition and phase estimation methods. We analyzed errors between measured and expected phases using the probability mass function and Cumulative Distribution Function. Repeatable acquisition and estimation methods were identified based on the probability of relative phase errors. For single-echo GRE and segmented EPI sequences, a region-growing method was most reliable with Pr (relative error <0.1) = 0.95 and 0.90, respectively. For multiecho sequences, a maximum-likelihood method was most reliable with Pr (relative error <0.1) = 0.97. The most repeatable multiecho methods outperformed the most repeatable single-echo methods. We found a wide range of repeatability and reproducibility for off-the-shelf MRI acquisition and phase estimation approaches, and this variability may prevent the techniques from being widely integrated in clinical workflows. The error was dominated in many cases by spatially discontinuous phase unwrapping errors. Any postprocessing applied on erroneous phase estimates, such as QSM's background field removal and dipole inversion, would suffer from error propagation. Our paradigm identifies methods that yield consistent and accurate phase estimates that would ultimately yield consistent and accurate $\Delta\chi$ estimates.

1. Introduction

Quantitative Susceptibility Mapping (QSM)[1–3] is a method to estimate magnetic susceptibility of tissue from the phase of the magnetic resonance (MR) signal. QSM has potential clinical utility for characterizing neurological diseases [4–6], blood oxygen content [7], iron overload in the heart and liver [8], and quantitative tracking of contrast agent bolus perfusion [9, 10].

Repeatability and reproducibility of QSM have been assessed in phantoms and human subjects using different scanners, magnetic field strengths, and data processing methods. While some studies report high repeatability [11–18], both *in vivo* and in phantoms, recent *in vivo* studies report lower reproducibility across MRI scanners with the same data processing method [19] and across QSM algorithms using the same input data [20]. These conflicting results limit the clinical adoption of QSM.

A typical QSM process requires four steps: data acquisition (Step 1), phase estimation (Step 2), background field removal (Step 3), and magnetic susceptibility reconstruction (Step 4). Recently proposed methods combine Steps 2–4 into fewer steps [21].

In its standardization efforts, the QSM community has actively evaluated competing methods [17, 22, 23], in particular for methods in Steps 3 and 4 of the process [20]. However, the selection of the "best" QSM method is difficult for various reasons: (a) the appropriate definition of a quality metric, for example, accuracy versus repeatability; (b) competing quality metrics that favor different algorithms [20]; (c) the lack of a gold standard *in vivo*; (d) algorithm performance depending on imaging application (*in vivo* versus phantoms); and (e) the large number of methods for each QSM step, which would render any exhaustive validation effort to be combinatorial and quickly untenable.

We present an experimental setup that allows for an exhaustive quantitative analysis of all four QSM steps. This framework uses a rotating-tube phantom design introduced in the work of Erdevig et al. [24], which uses tubes that rotate, within a background solution, relative to the main magnetic field, B_0. The design enables the analysis of MRI data obtained in objects at any orientation, using common QSM techniques. The closed-form theoretical relationship between the magnetic field and magnetic susceptibility in the sample allows for mapping the magnetic field to magnetic susceptibility without having to solve the dipole inversion problem [25].

Our contributions include (a) a framework for evaluation of repeatability and reproducibility of QSM algorithms and (b) rigorous analysis of common methods in Steps 1 and 2 of the QSM process. We focus on the performance of QSM Steps 1 and 2 for the following reasons:

(a) There exist a large number of methods in each of the four QSM steps. We replicate here approximately 90 different combinations of Steps 1 and 2 methods. If we analyze all four steps simultaneously, the resulting data set would grow combinatorially and become difficult to interpret.

(b) Errors introduced early in the QSM process propagate downstream and have been overlooked in validation studies. For example, Olsson et al. used only one acquisition sequence and one phase estimation algorithm [22].

(c) It is well understood that phase unwrapping, a common algorithm used in Step 2, is a nondeterministic polynomial-time-hard problem (in two dimensions and higher) that often relies on user intervention and careful parameter tuning. Therefore, it is important to isolate the impact of such a problem in the QSM processing methods.

(d) Susceptibility weighted imaging [26, 27], electrical properties tomography [28], thermometry [29], flow [30], and elastography [31] use MR phase information, and this work can inform those applications.

We analyze Steps 1 and 2 of the QSM process to understand which are sufficiently robust to be executable without user intervention and independent of scanner, sequence, and other parameter variations.

2. Methods

We used a rotating-tube phantom (Figure 1(a)) to explore the reproducibility of phase estimates obtained after Steps 1 and 2. MRI data was acquired with different pulse sequences (Step 1), at nine angles, and the performance of a wide variety of phase estimation methods was compared to theory. The rotating-tube phantom was designed to take advantage of the analytical model for a long cylinder at an angle, θ, with respect to B_0. The internal z-axis field offset can be derived from Maxwell's equations and is shown to be [25]

$$\delta B_{\text{in}} = \frac{\Delta\chi B_0}{6}\left(3\cos^2\theta - 1\right), \quad (1)$$

where $\Delta\chi$ is the susceptibility difference between the inside and outside of the cylinder and χ^2 terms are ignored.

2.1. Experimental Setup. The rotating-tube phantom consists of five cylindrical, polypropylene tubes (80 mm length and 10 mm outer diameter; Nalgene cryogenic storage vials) alternated orthogonally along the central axis of a larger cylinder (610 mm length and 140 mm outer diameter) containing water (Figure 1). Each tube contains one of the following: 0.5 mM $GdCl_3$, 1.0 mM $GdCl_3$, or 3.2 mM $CuSO_4$ (Figure 1(b)). This phantom is modifiable, and the range of susceptibility values can be focused on particular areas of interest. Here, the range of susceptibility values was selected to span those observed *in vivo* from venous blood (1.0 mM $GdCl_3$) to deep grey matter structures (0.5 mM $GdCl_3$) to the lower limits of MRI detection (3.2 mM $CuSO_4$) [32]. A rod extends from the internal rotation gears through the phantom and outside the MRI scanner, allowing the tubes to be manually rotated. Example MR images are shown for the three primary planes (Figure 1(c)) and in a 3D rendering of the tubes (Figure 1(d)). Temperature of the water was continuously monitored via a fiber optic probe (OpSens Medical, Québec, QC, Canada).

The paramagnetic salt solution $\Delta\chi$ values were estimated from susceptibility theory and corrected using Curie's law with the experimentally measured temperatures (21.0°C–22.0°C). $\Delta\chi$ values were 0.336 ppm and 0.168 ppm for the 1 mM $GdCl_3$ and 0.5 mM $GdCl_3$, respectively, and 0.0804 ppm for the 3.2 mM $CuSO_4$. Additional details on the calculation of $\Delta\chi$ are in the Supplementary Material.

2.2. Data Acquisition. MR data was collected on a 3T Siemens Biograph mMR (MR-PET scanner, Syngo MR E11 software) with a 6-element torso array and 9-element spine array coil for a total of 15 elements. To assess the reproducibility of phase estimation across pulse sequences, we acquired data with four gradient echo (GRE) pulse sequences (details in Table 1):

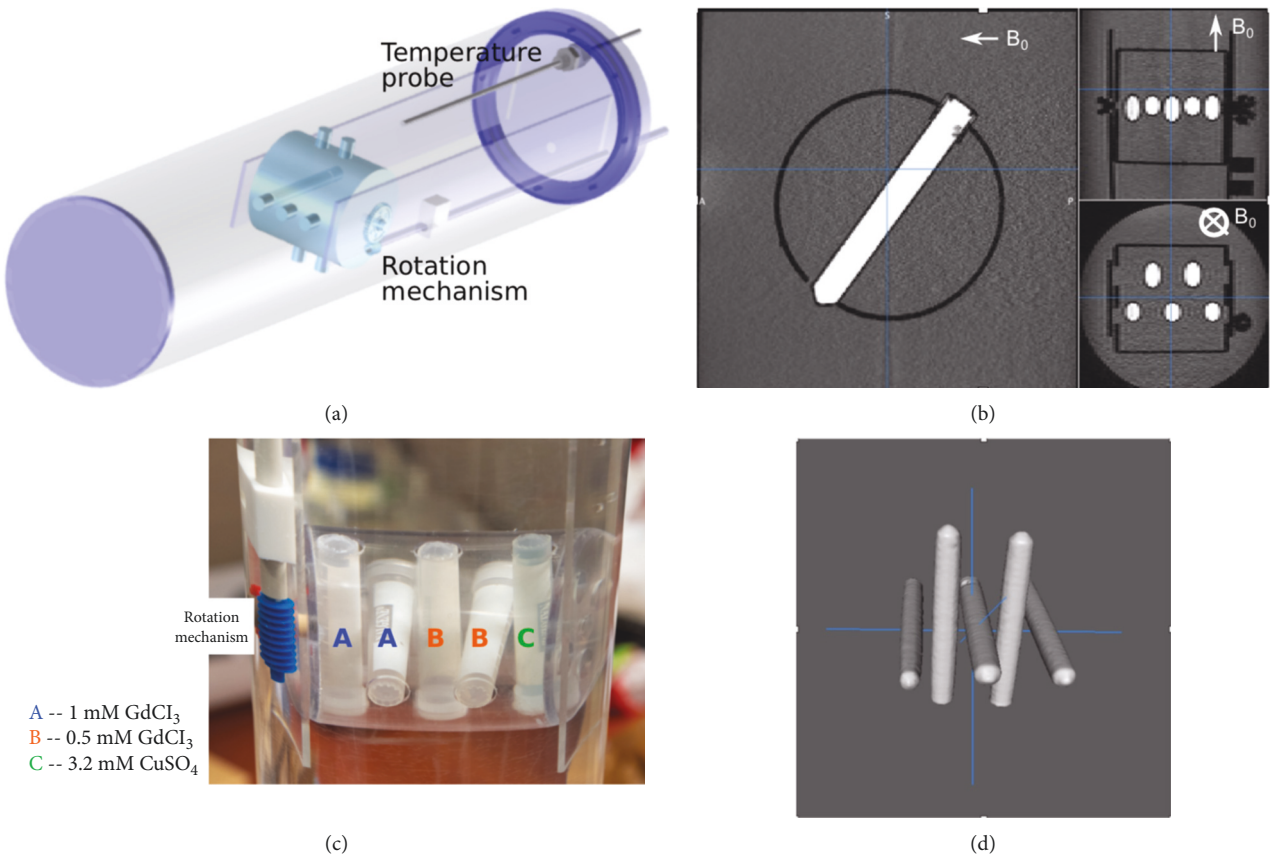


Figure 1: (a) 3D diagram of the rotating-tube phantom design with five smaller cylindrical samples along with temperature probe. (b) Magnitude images of the sagittal (one tube), axial, and coronal (multiple tube) views. (c) A photograph of the 5 tubes. (d) 3D rendering from MRI magnitude images.

Table 1: All data acquisition parameters.

Sequence	Scan ID	TR (ms)	Echo times (ms)	Bandwidth (Hz)	Alpha (deg)	Resolution (mm^3)	Acquisition time (min:s)
Single-echo GRE (SEGE)	1	25	16	80	15	$1.0 \times 1.0 \times 2.0$	1 : 59
	2	35	25.7	80	15	$0.5 \times 0.5 \times 2.0$	4 : 35
	3	45	30	80	15	$1.0 \times 1.0 \times 2.0$	3 : 35
Multiecho GRE (MEGE)	1	25	2.5, 6.2, 9.9, 13.6, 17.3, 21.0	500	15	$1.0 \times 1.0 \times 2.0$	1 : 59
	2	35	3, 7.8, 12.6, 17.4, 22.2, 27.0	530	15	$0.5 \times 0.5 \times 2.0$	4 : 35
	3	45	2.5, 6.2, 9.9, 13.6, 17.3, 21.0, 24.7, 28.4, 32.1, 35.7, 39.5	500	15	$1.0 \times 1.0 \times 2.0$	3 : 35
sEPI	1	72	27, ETL = 15	430	20	$1.0 \times 1.0 \times 2.0$	0 : 31
	2	72	27, ETL = 15	430	20	$0.5 \times 0.5 \times 2.0$	0 : 52
MAGPI	1	25	4.1, 8.9, 12.6, 16.3, 20.0	190	15	$1.0 \times 1.0 \times 2.0$	1 : 59
	2	35	7.6, 13.0, 17.3, 21.5, 25.7, 30.0	250	15	$0.5 \times 0.5 \times 2.0$	4 : 35
	3	45	6.6, 13.0, 18.5, 23.8, 29.0, 34.2, 39.4	200	15	$1.0 \times 1.0 \times 2.0$	3 : 35

(i) Single-echo GRE (SEGE): This is a commonly chosen protocol with QSM and other susceptibility-based techniques [33], wherein a single-echo time TE is measured as close to T_2^* of tissue of interest (here, the target is 60 ms for 1 mM $GdCl_3$). This maximizes the phase SNR at this T_2^*. To maximize the magnitude SNR at the chosen TE and TR, we set the readout bandwidth at its lowest possible value.

(ii) Segmented Echo Planar Imaging (sEPI): A recently proposed sEPI sequence was shown to possess similar quality phase images as SEGE [34], while acquiring full brain coverage much faster than SEGE. As with SEGE, phase images were generated at a single TE at the center of the echo train.

(iii) Multiecho GRE (MEGE): This protocol acquires multiple TEs in a single TR. The challenge with this technique is the choice of the echo spacing ΔTE and readout bandwidth BW. A short ΔTE reduces the likelihood of aliasing in the phase domain but introduces noise. A long ΔTE yields phase images with better SNR but suffers from potentially unrecoverable phase-aliasing errors. For example, to unwrap frequency offsets of ±150 Hz, ΔTE must be less than 3.33 ms. A common approach [33, 35] is to acquire data with a short ΔTE and, in order to recover SNR efficiency similar to SEGE sequences [36], acquire as many echoes as possible in a TR. However, due to hardware limitations, the readout bandwidth, BW, places a lower limit on ΔTE. We aimed to select the shortest ΔTE possible at the highest BW attainable with the MR system. This choice minimizes the likelihood of phase wrap errors, which may not be recoverable by all phase unwrapping algorithms. We elaborate on this choice further in the Discussion section.

(iv) MAGPI: This is an MEGE sequence that uses preoptimized echo times and bandwidths selection that, when paired with a corresponding phase estimation algorithm, yields maximum-likelihood optimal phase estimates in the presence of wrapping, noise, and phase-offset errors [37, 38].

All sequences were 3D excitations of a 128 mm × 128 mm × 128 mm slab (64 slices) with sagittal slab-selection and phase encode along the B_0 direction. We used anisotropic voxels to boost SNR, a common practice for QSM and Susceptibility-Weighted Imaging [26, 27, 39]. MAGPI and the sEPI sequences used alternating gradient polarity, while the other sequences did not. Autoshimming was completed prior to the first data acquisition, and then the same shim parameters were used over time, over all sequences and methods.

We assessed the effect of in-plane resolution and TR on phase estimate reproducibility with each protocol (Table 1). We also examined the reproducibility of phase estimates across nine different angles by advancing the apparatus approximately 18 degrees per turn.

2.3. Phase Estimation. Images generated in Steps 2–3 of QSM are commonly referred to as frequency, phase, or field maps, depending on the units of the data. We interchangeably use these names in this work depending on context, and, in our analysis, we convert all phase images to frequency via a simple scalar multiplication. Ten phase algorithms were selected to estimate the frequency offset image [37, 40–47]. Table 2 lists the methods used for each pulse sequence; multiple codes were downloaded from freely available resources (e.g., MEDI) and integrated with the pipeline.

All phase estimation methods were applied with default parameters in 3D over the entire acquisition volume and were implemented and run blinded to the theoretical solution. Apart from the MAGPI algorithm, which operates on raw k-space channel-uncombined data, all methods (including MAGPI-unopt) operated on unprocessed phase data obtained using the vendor-provided adaptive-coil-combine method. While this is "unprocessed phase data," different vendors may apply different filters, phase corrections, or other adjustments that could influence the phase quality. Here, adaptive-coil-combine describes the algorithm used by this vendor to combine multichannel coil data [48]. Each unwrapping method used the SNR in the magnitude image to guide unwrapping orientation in the phase domain: for example, the Laplacian-based methods used this SNR to mask the entire image, while others (region-growing, GBC) masked the phase values in regions with poor SNR. SNR was measured as approximately 30 dB in water, similar to *in vivo* values ranging from 25 to 30 dB.

For MEGE sequences, the multiecho data is processed using the five following categories of algorithms:

(1) Spatial phase unwrapping at each echo, followed by temporal combination of the resulting images using a weighted averaging method, that is, phase SNR-optimal [49].

(2) Spatial phase unwrapping at each echo with weighted averaging (as in 1), but a 1D phase unwrapping step is used just before weighted averaging. This is meant to correct any remaining aliasing that spatial unwrapping failed to correct.

(3) Direct temporal phase estimation (Slope, Division) applied in complex domain. These methods correctly unwrap the phase over time, provided the inherent frequency is less than the Nyquist frequency associated with the echo spacing.

(4) Temporal phase combination (as in 3), followed by 3D spatial phase unwrapping to correct errors encountered with temporal phase estimation.

(5) Maximum-likelihood-based combination of multiecho and multichannel data (MAGPI) [37]. This method solves the phase estimation problem on a voxel-by-voxel basis, without resorting to spatial averaging techniques.

All phase images are eventually converted to frequency offset (Hz) by dividing by 2π x phase evolution time.

Table 2: All phase estimation methods. The first six methods were combined with other common phase processing techniques to process the multiecho data, as described in the Methods section.

Sequence	Method	Summary
SEGE, MEGE, sEPI	Unprocessed	Default channel combined output phase image produced by the scanner
	Phun	Region-based algorithm [46]
	GBC	Goldstein's branch cut method [42]
	MEDI.RG	Region growing algorithm from MEDI Toolbox [45, 47]
	MEDI.LP	Laplacian algorithm from MEDI Toolbox [44, 45]
	Laplace	Laplacian algorithm [40]
MEGE	Slope	Computes the frequency offset from the slope of the complex data across echoes [41]
	Division	Computes the frequency offset from the complex division of the data at successive echoes [43]
	MAGPI-unopt	Applies the MAGPI postprocessing algorithm using a random (unoptimized) subset of echoes from the MEGE sequence. Since MAGPI requires uneven echo spacing, only four unequally spaced echoes were selected [37]
MAGPI	MAGPI	Maximum-likelihood estimate of phase from optimal echo spacing [37]

2.4. Adjusting for Field due to Rotation of Apparatus. To accurately estimate $\Delta\chi$ from the phase images, we need to remove the global field effects resulting from the tubes rotating in the magnetic field, as well as field effects from the apparatus itself. We call this process "frequency referencing" (FR).

The scalar magnitude of the complete field inside a voxel can be written as follows:

$$\delta B(r,\theta) = \delta B_{\Delta\chi}(r,\theta) + \delta B_0(r,\theta), \tag{2}$$

where r is the spatial coordinate of the voxel and θ is the rotation angle of the tube relative to B, $\delta B_{\Delta\chi}(r,\theta)$ is the field caused by magnetic susceptibility variations (such as the one due to a homogeneous cylindrical object immersed in a homogeneous sphere), and $\delta B_0(r,\theta)$ is an unknown field offset component. We separate $\delta B_0(r,\theta)$ into a component that varies only spatially and a component that varies only with the rotation of the apparatus:

$$\delta B_0(r,\theta) = \delta B_0^{bkg}(r) + \delta B_0^{rot}(\theta). \tag{3}$$

At a given angle of rotation, sources of spatially varying global offsets $\delta B_0^{bkg}(r)$ are field inhomogeneity (imperfections of magnetic field/coils), bulk magnetic susceptibility of the apparatus [50, 51], and coil phase offset [52]. As the apparatus is rotated, in the absence of "shimming" at the console, the center frequency will be shifted due to the bulk magnetic susceptibility of the entire apparatus [53, 54]. Our goal is to extract $\Delta\chi$ by fitting $\delta B(r,\theta)$ to the angle of rotation θ, inside the tube. $\delta B_0^{bkg}(r)$ is a nuisance parameter that can be easily accounted for during the fitting process by allowing for a constant shift to the cosine.

First, we compute an estimate of $\delta B_0^{rot}(\theta)$ using the average frequency in a static region outside the "Tube + - Sphere" system (Figure2(a)). The average (indicated by < >) is taken over pixels in a region r_{out} distant from local susceptibility effects:

$$\langle \delta B(r,\theta)\rangle_{r\in r_{\text{out}}} = \langle \delta B_0^{bkg}(r)\rangle_{r\in r_{\text{out}}} + \delta B_0^{rot}(\theta) = C_{r_{\text{out}}} + \delta B_0^{rot}(\theta), \tag{4}$$

where we define $C_{r_{\text{out}}}$ to be a variable that is only a function of the referencing region. Then, the referencing step consists of

$$\delta B_{\text{ref}}(r,\theta) \equiv \delta B(r,\theta) - \langle \delta B(r,\theta)\rangle_{r\in r_{\text{out}}} = \delta B_{\Delta\chi}(r,\theta) + \left(\delta B_0^{bkg}(r) - C_{r_{\text{out}}}\right), \tag{5}$$

thus removing $\delta B_0^{rot}(\theta)$. Our goal is to use the referenced field, $\delta B_{\text{ref}}(r,\theta)$, at every pixel in the tube center, $r = r_{\text{in}}$, to fit the field variation to the angle of rotation θ. The only component that varies with θ is the first term on right side of equation (5). The estimation step is a simple fit with respect to θ, along with an arbitrary shift for the constant: $c = \delta B_0^{bkg}(r) - C_{r_{\text{out}}}$. For the case of a cylinder (equation (1)), the estimate can be obtained by solving

$$\left(\widehat{\Delta\chi}(r_{in}), c^*\right) = \arg\min_{\Delta\chi,c} \left\| \widehat{\delta B}_{\text{ref}}(r_{\text{in}},\theta) - \left(\frac{\Delta\chi}{6}\left(3\cos^2\theta - 1\right)B_0 + c\right)\right\|_2. \tag{6}$$

Fitting c effectively amounts to shifting the midline of the data to match the model (across all angles). We used a bisquare-weighting method to fit this midline. We include a few examples of $\widehat{\Delta\chi}$ estimation using equation (6) in Table S.1.

We apply this frequency referencing method after each phase estimation algorithm. To investigate repeatability with respect to the location of frequency reference estimate, we apply this process in 13 different regions selected across static areas of the phantom (Figure 2(a)).

2.5. Error Analysis. We use the theoretically determined values of $\Delta\chi$ to predict the field values at each angle that would have been measured with ideal methods in Steps 1 and 2. We then compute the error (Hz) between measured frequency $\widehat{\delta B}_{\text{ref}}(r_{\text{in}},\theta)$ and expected frequency offset:

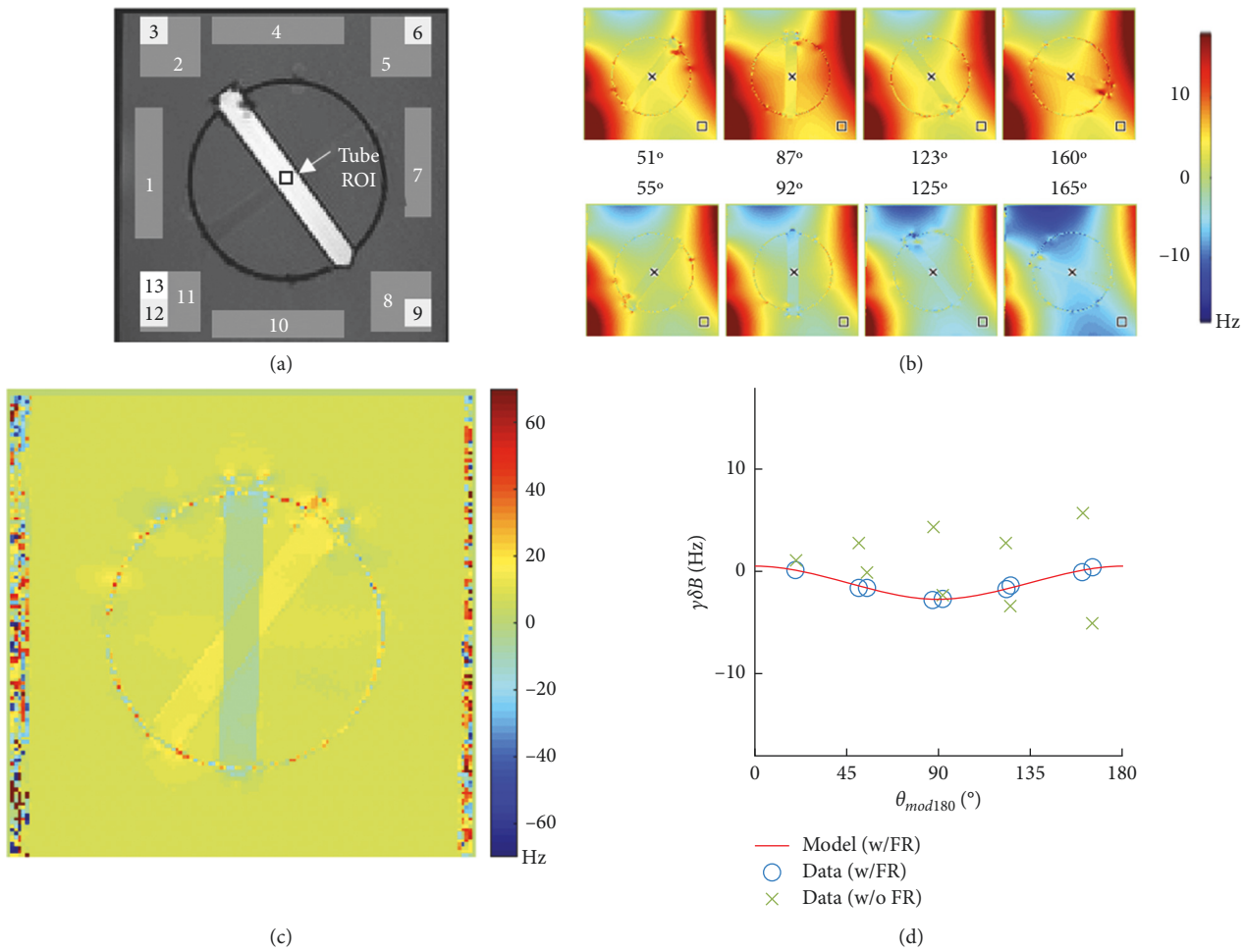


Figure 2: (a) An example magnitude GRE image showing both the tube ROI (small black box in center) used for the analysis and the placement of the 13 background ROIs. (b) Example frequency images in tube 3 (0.5 mM GdCl3) obtained with the single GRE sequence (1 mm, TR 45 ms) and reconstructed with MEDI.RG phase estimation method. (c) Also using SEGE + MEDI.RG, the field difference between two angles ($\delta\mathbf{B}_{\mathbf{total}}(\mathbf{r}, \theta_2) - \delta\mathbf{B}_{\mathbf{total}}(\mathbf{r}, \theta_1)$) showing that this difference is attributable to the spatially invariant component (in homogeneous areas) and a spatially variant component (in areas close to material boundaries). The spatially invariant component of the field difference is removed with frequency referencing (FR). (d) A plot of the frequency against the angle of rotation (modulo 180°) is shown for each of the data without frequency referencing (in green (x), data after frequency referencing using the frequency reference ROI #9 (blue circles), and the model's prediction of the frequency (solid red line).

$$\varepsilon(r_{\text{in}}, \theta) = \widehat{\delta B}_{\text{ref}}(r_{\text{in}}, \theta) - \left(\frac{\Delta\chi_{\text{true}}}{6}\left(3\cos^2\theta - 1\right)B_0 + c^*\right), \tag{7}$$

where c^* is obtained by solving equation (6) with $\Delta\chi$ set to $\Delta\chi_{\text{true}}$.

To account for any dependence on tube content, we compute the absolute relative error:

$$\epsilon_{\mathbf{r}}(\mathbf{r}_{\mathbf{in}}, \boldsymbol{\theta}) = \frac{\left|\epsilon(\mathbf{r}_{\mathbf{in}}, \boldsymbol{\theta})\right|}{\mathbf{B}_0\Delta\boldsymbol{\chi}}, \tag{8}$$

where the absolute value is used instead of the signed error due to the irrelevance of sign in this context.

Error statistics were computed for each voxel in the tube ROI (mean tube ROI sizes were 17 pixels at 1 mm resolution and 57 pixels at 0.5 mm resolution), for each slice (2 slices per tube), each tube (5 total), each angle (9 total rotations), using each applicable phase estimation method (from a possible 10), each background phase removal ROI (13 total), and each sequence with its respective resolution and TR variations (11 total). Cumulatively, 2.36 million frequency values were analyzed in this experiment.

The large number of data points allows us to extract statistics about ε and ε_r including their probability mass function, $\Pr(\varepsilon)$ and $\Pr(\varepsilon_r)$. The probability of ε can exhibit a multimodal distribution and therefore is not a Gaussian. While we can report the absolute bias and standard deviations from such a distribution, it would not be descriptive of $\Pr(\varepsilon)$. A more practical measure is the likelihood of observing absolute relative errors less than or equal to a threshold, τ. This is obtained by integrating $\Pr(\varepsilon_r)$ between 0 and τ a measure known as the Cumulative Distribution Function (CDF),

$$F_{\varepsilon_r}(\tau) = \Pr(\varepsilon_r \leq \tau). \quad (9)$$

The CDF can be used to capture phase errors that are dominated by outliers, as well as phase errors that result from generally poor/unreliable model fitting. The ideal CDF is a step function, and any presence of outliers/large errors yields a CDF with slow convergence to 1. The frequency of large errors is seen from the magnitude of the deviation of the CDF from 1.0 at any given threshold.

3. Results

Figure 2(a) shows a magnitude GRE image of tube 3 (0.5 mM $GdCl_3$) in one orientation. Figure 2(b) shows frequency offset images corresponding to different rotations. Figure 2(c) shows a typical field difference between two angles ($\delta B(r, \theta_2) - \delta B(r, \theta_1)$). According to equation (3), this difference is equal to $\delta B_{\Delta\chi}(r, \theta_2) - \delta B_{\Delta\chi}(r, \theta_1) + \delta B_0^{rot}(\theta_2) - \delta B_0^{rot}(\theta_1)$, predicting spatial variations only in locations close to areas with susceptibility changes and a constant field in homogeneous locations. This is precisely what we observe in Figure 2(c). Figure 2(d) shows a plot of the frequency values for a voxel inside the tube prior to frequency referencing (symbol "x"). The resulting plot does not follow the expected sinusoid (solid line). After applying the frequency referencing step, we observe the expected sinusoidal shape (symbol "o").

Figure 3 shows the resulting histogram of frequency errors and resulting CDF. From Figures 3(a) and 3(b), we note the probability of ε for this example exhibits a multimodal distribution. We can see from the CDF in Figure 3(c) that the probability of observing errors less than 10% ($F_{\varepsilon_r}(0.1)$) is about 66%.

Table 3 summarizes the error statistics for all combination of sequences and algorithms. The first column lists the sequence type, and the second and third columns indicate the name and category of each postprocessing algorithm, respectively. We report the mean and standard deviations of both ε and ε_r. We also report $F_{\epsilon_r}(0.1)$ pooled over all background ROIs, as well as the range (minimum, $F_{\varepsilon_{rW}}(0.1)$, and maximum, $F_{\varepsilon_{rB}}(0.1)$) of $F_{\varepsilon_r}(0.1)$ encountered in those ROIs. These quantities are computed from data that includes all tubes and angles. A representative subset of these results is selected for more detailed analysis and illustration in Figures 4–6.

Figure 4 shows a subset of frequency-offset images for different sequence + algorithm pairs, at 3 of the 9 angles of rotation. This figure illustrates typical challenges with phase estimation methods. For example, in Figure 4(a), we observe phase unwrapping errors in SEGE + GBC, with abrupt jumps across contiguous regions. The corresponding frequency versus angle plot (last column in Figure 4(a)) shows that incorrect frequency referencing in these areas (square in figure) yields occasional mismatch between measurement and predictions at certain angles. SEGE + Laplace demonstrates a smoothly varying frequency map across the FOV; however, the resulting data deviates from the expected theoretical values at almost every angle (Figure 4(b)). MEGE + Slope, a direct temporal phase estimation method (MEGE category 3), exhibits phase wrapping errors when the underlying frequency value is larger than the bandwidth allowable by ΔTE(Figure 4(c)). Placing a frequency referencing ROI in these areas yields incorrect values at the respective angles. Note that the example frequency reference ROI (#9 in Figure 2) is meant to highlight the phase errors or artifacts observed. Figures 4(d) and 4(e) show that the results from MAGPI-unopt and MAGPI are consistent with those predicted from theory.

The CDF of ε_r collects the errors, such as those observed in Figure 4, over a variety of acquisition and processing parameters. Figure 5 shows $F_{\varepsilon_r}(\tau)$ for all algorithms when ε_r is pooled over all voxels, background ROIs, slices, tubes, and sequence variations. This represents an overall summary of algorithm behavior, irrespective of which parameter was used in acquisition and postprocessing. We see that MAGPI attains a nearly ideal CDF, with 0.91 probability of relative errors, ε_r, less than 0.1 ($F_{\varepsilon_r}(0.1) = 0.91$) and rapidly converges to 1 (Figure 5). MAGPI and MAGPI-unopt achieve similar CDFs, with MAGPI performing slightly better, as expected. MEDI.RG and GBC phase unwrapping methods, both based on region growing, have similar CDFs, with $F_{\varepsilon_r}(0.1) = 0.69$ and $F_{\varepsilon_r}(0.1) = 0.70$, respectively. The unprocessed phase images have the most artifacts and, thus, the lowest CDF across all ε_r. Figure 5 focuses on the CDF for ε_r in [0, 1.0] to highlight the different convergence pattern (distribution/frequency of errors) in that domain. The CDF extends beyond $\varepsilon_r = 1.0$ for any occurrence of relative errors greater than 100%.

Next, we explore the behavior of $F_{\varepsilon_r}(\tau)$ as a function of data acquisition strategy. In Figures 6(a)–6(c), we group results by three sequence types: SEGE, sEPI, and MEGE. Since MAGPI is a multiecho sequence, we include MAGPI in the MEGE category. For each CDF curve $F_{\epsilon_r}(\tau)$, ϵ_r is pooled across all pixels, background ROIs, slices, tubes, and variations of TR and resolution within that sequence type. We also explore variability of $F_{\epsilon_r}(\tau)$ with the frequency referencing method. For each sequence, we show CDFs in the ROI with the maximum $F_{\epsilon_r}(\tau = 0.1)$ ($F_{\epsilon_{rB}}$, Figures 6(d)–6(f)) and the ROI with the minimum $F_{\epsilon_r}(\tau = 0.1)$ ($F_{\epsilon_{rW}}$, Figures 6(g)–6(i)) for a given data set. The separation between $F_{\epsilon_{rB}}$ and $F_{\epsilon_{rW}}$ demonstrates the robustness (or lack thereof) of a method to frequency reference ROI selection.

Table 4 shows the dependence of the CDF on scan variability and tube contents. Because the performance of some methods is dominated by frequency reference ROI (seen in Table 3 and Figure 6), Table 4 shows results for the largest $\mathbf{F}_{\epsilon_r}(10\%)$ for a given sequence + method pair observed over all frequency reference ROIs, for every method, sequence, scan variation, tube, and slice.

4. Discussion

A reliable pulse sequence protocol, with repeatable and reproducible phase estimation, is a necessary step to develop robust QSM methods for clinical use. Previous work examined the reproducibility of certain QSM methods using phantoms [22, 25, 55], simulation [22, 25, 56], and human subjects [19, 56]. We used a rotating-tube phantom to

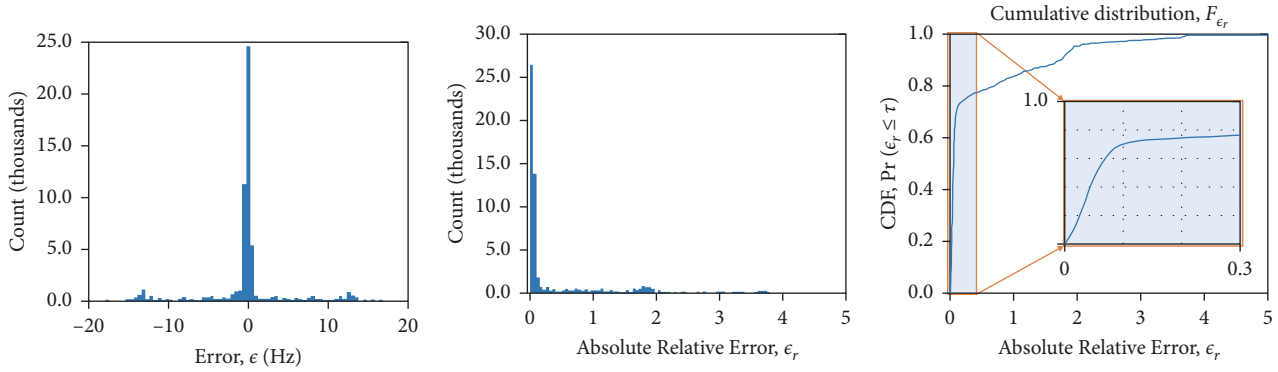
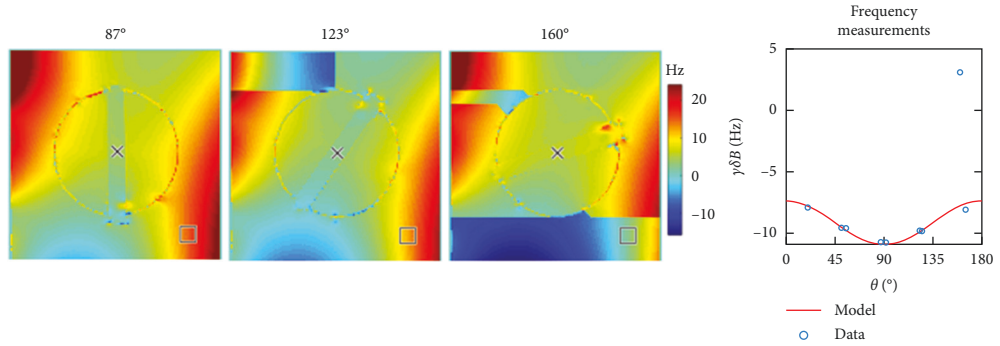


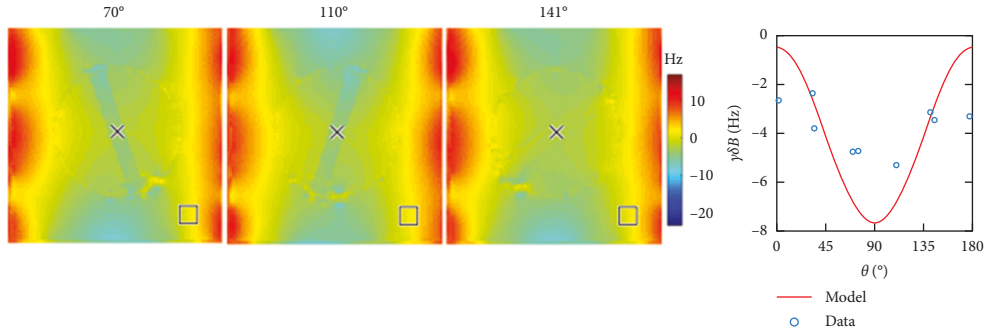
Figure 3: (a) The histogram of the error ε (in Hz) when the measurement is obtained with the SEGE + GBC phase estimation method. The error data is pooled over all ROI voxels, all backgrounds, rotations, and tubes. (b) The histogram of the absolute relative error $\boldsymbol{\varepsilon_r}$ seen with the same measurement. (c) The corresponding Cumulative Distribution Function $\mathbf{F}_{\epsilon_r}$ of the error in (b). The inset figure shows a portion of this CDF within [0, 0.3] to better illustrate the distributions within reasonable error range. In this case, the CDF shows that the probability of obtaining $\boldsymbol{\varepsilon_r}$ less than 0.1 is around 0.66 with the SEGE + GBC pair. The maximum observed ϵ was 27.0 Hz, and the maximum observed $\boldsymbol{\varepsilon_r}$ was 6.85, illustrating the occasional large errors that may result with SEGE + GBC.

Table 3: Error statistics for each of the estimation methods and acquisition protocols. The mean and standard deviation of error measurements ε (Hz), as well as the absolute relative errors ε_r, are provided in columns 4–7. Columns 8-9 show $F_{\epsilon_r}(0.1)$ of the overall data, as well as the range (min and max) of $F_{\epsilon_r}(0.1)$ across reference backgrounds.

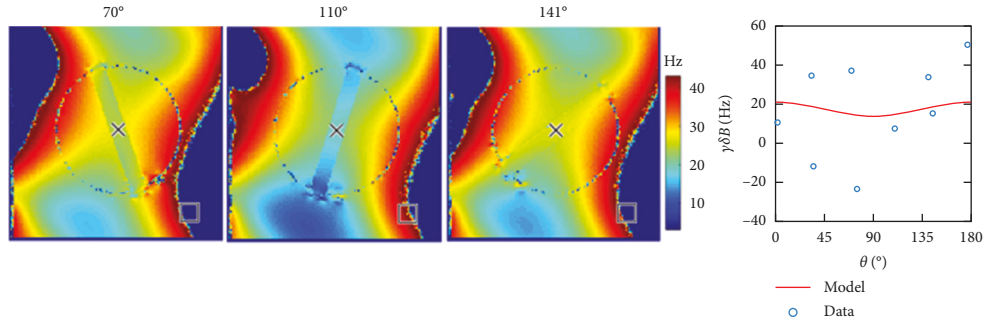
Protocol	Postprocessing	Method	μ_ϵ(Hz)	σ_ϵ (Hz)	μ_{ϵ_r}	σ_{ϵ_r}	$\mathbf{F}_{\epsilon_r}(0.1)$	$[\mathbf{F}_{\epsilon_{rW}}(0.1), \mathbf{F}_{\epsilon_{rB}}(0.1)]$
SEGE	Spatial 3D unwrap	Unprocessed	0.29	3.85	0.52	0.92	0.48	[0.24, 0.76]
		Phun	0.21	5.12	0.35	1.38	0.67	[0.49, 0.89]
		GBC	−0.22	4.83	0.53	1.08	0.66	[0.47, 0.82]
		MEDI.RG	−0.19	1.60	0.08	0.22	0.90	[0.84, 0.96]
		MEDI.LP	−0.01	1.07	0.14	0.12	0.50	[0.38, 0.62]
		Laplace	0.00	1.01	0.14	0.15	0.54	[0.40, 0.66]
sEPI	Spatial 3D unwrap	Unprocessed	−0.34	4.19	0.58	0.76	0.43	[0.32, 0.52]
		Phun	−0.10	5.23	0.73	1.30	0.45	[0.31, 0.56]
		GBC	−0.32	3.80	0.42	0.97	0.66	[0.31, 0.91]
		MEDI.RG	1.03	4.14	0.32	1.15	0.85	[0.79, 0.91]
		MEDI.LP	0.00	1.14	0.12	0.12	0.57	[0.51, 0.65]
		Laplace	0.00	0.95	0.12	0.11	0.58	[0.47, 0.68]
Multiecho	1. Spatial 3D unwrap ⟶ temporal average	(i) Unprocessed	0.01	2.60	0.48	0.41	0.16	[0.08, 0.23]
		(ii) Phun	0.39	4.75	0.30	1.23	0.57	[0.48, 0.69]
		(iii) GBC	0.31	2.21	0.15	0.36	0.78	[0.59, 0.91]
		(iv) MEDI.RG	−0.37	3.47	0.66	0.95	0.35	[0.09, 0.63]
		(v) MEDI.LP	−0.01	0.89	0.11	0.09	0.58	[0.49, 0.67]
		(vi) Laplace	0.00	0.82	0.12	0.12	0.58	[0.44, 0.68]
	2. Spatial 3D unwrap ⟶ temporal 1D unwrap ⟶ temporal average	(i) Phun	0.19	7.66	0.94	2.23	0.55	[0.12, 0.96]
		(ii) GBC	0.20	7.78	0.96	2.25	0.55	[0.12, 0.96]
		(iii) MEDI.RG	0.13	7.95	0.97	2.29	0.55	[0.12, 0.96]
		(iv) MEDI.LP	−0.01	0.89	0.11	0.09	0.58	[0.49, 0.67]
		(v) Laplace	0.00	0.82	0.12	0.13	0.58	[0.45, 0.68]
	3. Temporal estimation	(i) Slope	0.26	9.28	1.15	2.79	0.55	[0.13, 0.96]
		(ii) Division	0.26	9.27	1.15	2.78	0.55	[0.13, 0.95]
	4. Temporal estimation ⟶ spatial 3D unwrap	(i) Phun	0.04	11.04	1.17	3.02	0.53	[0.06, 0.95]
		(ii) GBC	0.17	9.18	1.10	2.69	0.54	[0.09, 0.95]
		(iii) MEDI.RG	0.47	62.07	10.11	16.54	0.21	[0.00, 0.76]
		(iv) MEDI.LP	0.16	3.87	0.54	1.16	0.43	[0.20, 0.70]
		(v) Laplace	0.22	6.89	0.89	2.05	0.36	[0.10, 0.66]
	5. MAGPI	(i) MAGPI unopt	0.00	0.30	0.05	0.05	0.90	[0.78, 0.97]
		(ii) MAGPI	0.00	0.27	0.05	0.04	0.91	[0.84, 0.98]



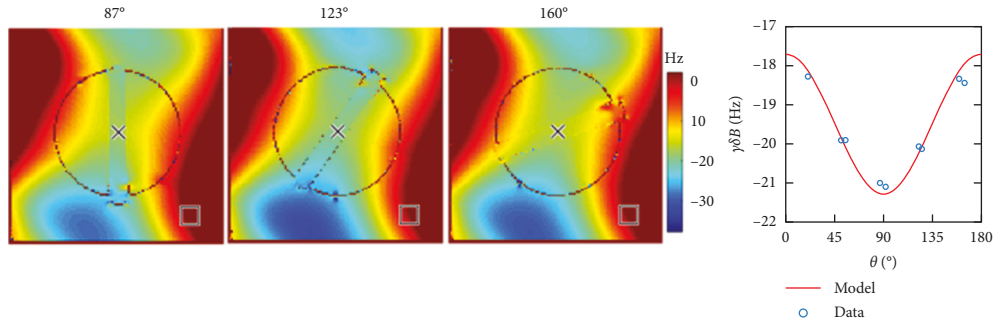
(a)



(b)

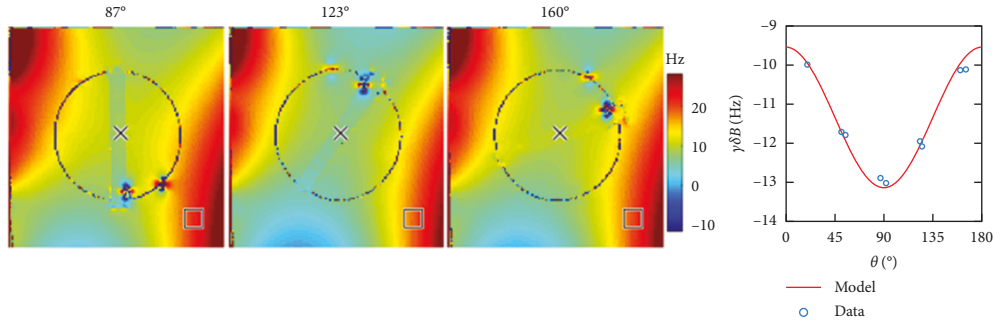


(c)



(d)

Figure 4: Continued.



(e)

Figure 4: Each row in this figure shows five different examples of sequence (at 1 mm resolution, TR = 45 ms) and phase estimation method pairs for three different angles of rotation in different tubes. These examples were selected to illustrate the spatial nature of different phase estimations errors and artifacts. The corresponding plots in the last column show the resulting frequency measurements as a function of angle of rotation (modulo 180°), for a voxel inside the tube, after frequency referencing (blue circle). The predicted frequency offset as obtained from equation (1) is also shown in solid red line. The background ROI (#9) used is shown with a square overlaid on the frequency maps. (a) SEGE + GBC shows a phase wrapping error in tube 3 in the frequency reference ROI for the 160-degree angle. (b) SEGE + Laplace shows smoothly varying frequency maps; however, the values deviate from the expected result at every angle. (c) MEGE + Slope, without 3D phase unwrapping, shows the presence of phase wrapping in areas with large frequency values. The frequency reference ROI is within a phase wrapped region across all angles. (d, e) MEGE + MAGPI-unopt and MAGPI's phase estimation show frequency maps consistent with values predicted from the model.

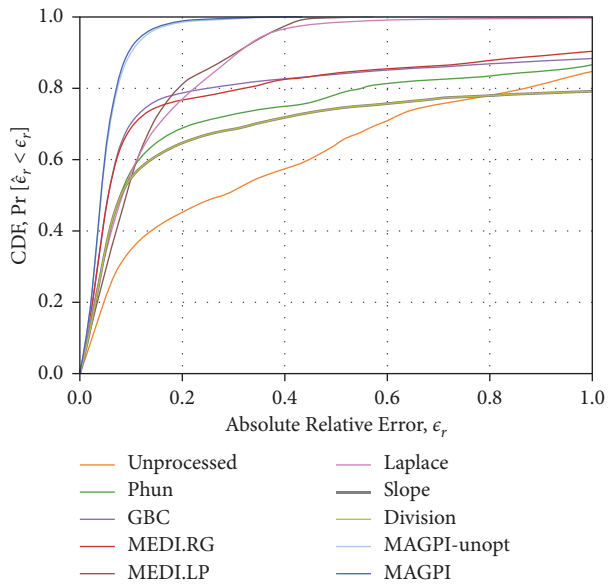


Figure 5: CDF $\mathbf{F}_{\epsilon_r}$ of the absolute relative error obtained for each of the phase estimation methods studied in this work, when the error is pooled over all sequences, slices, tubes, angles, and frequency referencing ROIs. Note that some methods do not converge to probability of 1, due to the presence of errors greater than 100%.

quantitatively evaluate methods used for data acquisition and phase estimation. This is not the first study using long tubes nor is it the first to position tubes relative to B_0; however, compared to previous work [22, 24, 25, 55, 57–59], more acquisition and phase estimation methods were considered. Specifically, we used ~90 different combinations of pulse sequences and phase estimation methods to analyze millions of measurements from different ROIs, tube contents, rotations, and sequence parameters. This vast amount of data ultimately allowed us to estimate the probability distribution of phase error with every QSM method, along with other important statistics. Additional phase estimation methods could be retrospectively used on the data set, which we aim to make publicly available.

Our results showed varying degrees of accuracy and precision over all tested methods. For example, while the majority of methods resulted in μ_ϵ less than 1 Hz (note the particularly small μ_ϵ with MEDI.LP, Laplace, MEGE + MAGPI-unopt, and MAGPI), the only methods with μ_{ϵ_r} <10% are SEGE + MEDI.RG (7.6%), MEGE + MAGPI-unopt (5.3%), and MAGPI (5.0%). We observed a similar trend with precision, whereby methods

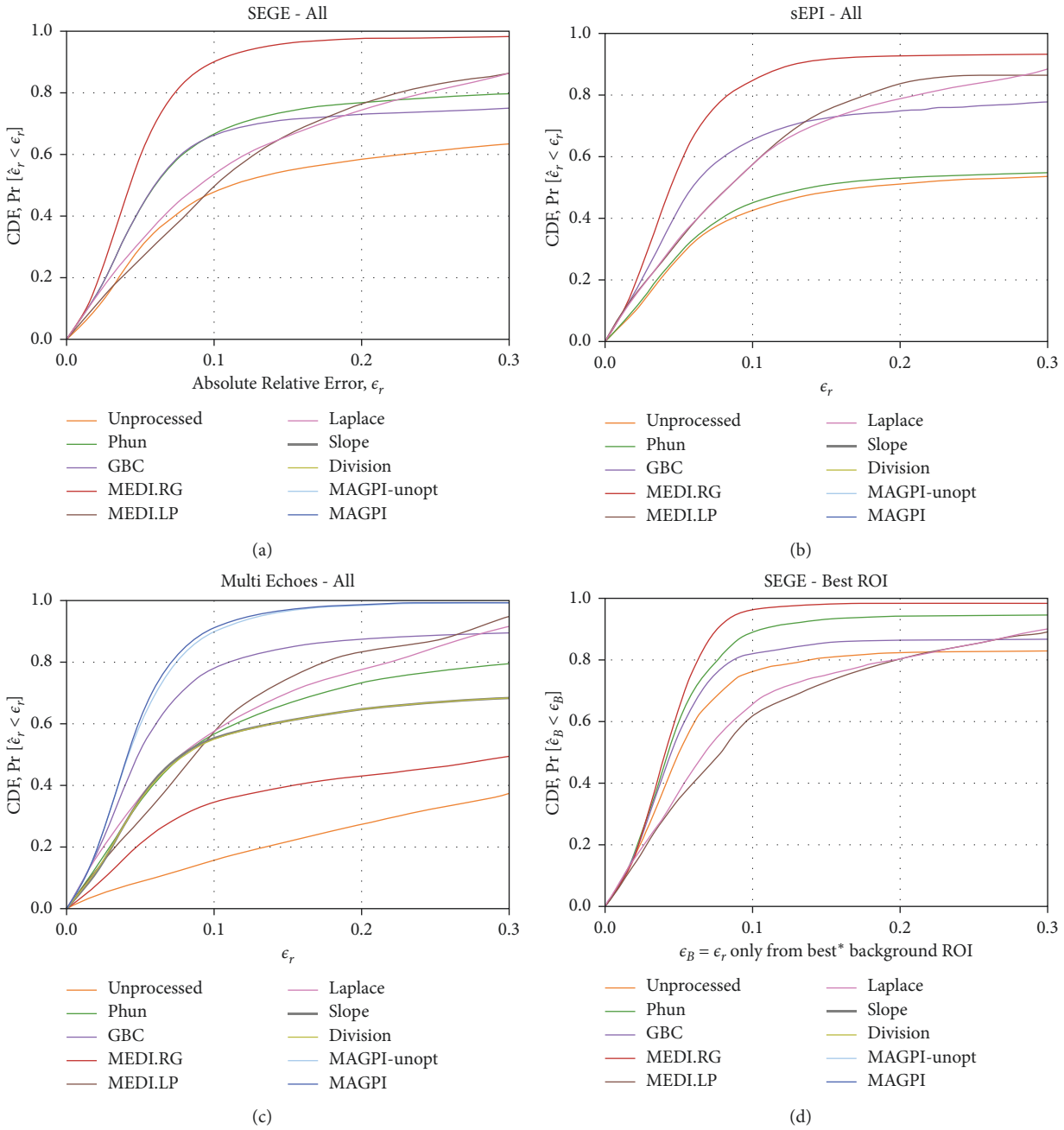


Figure 6: Continued.

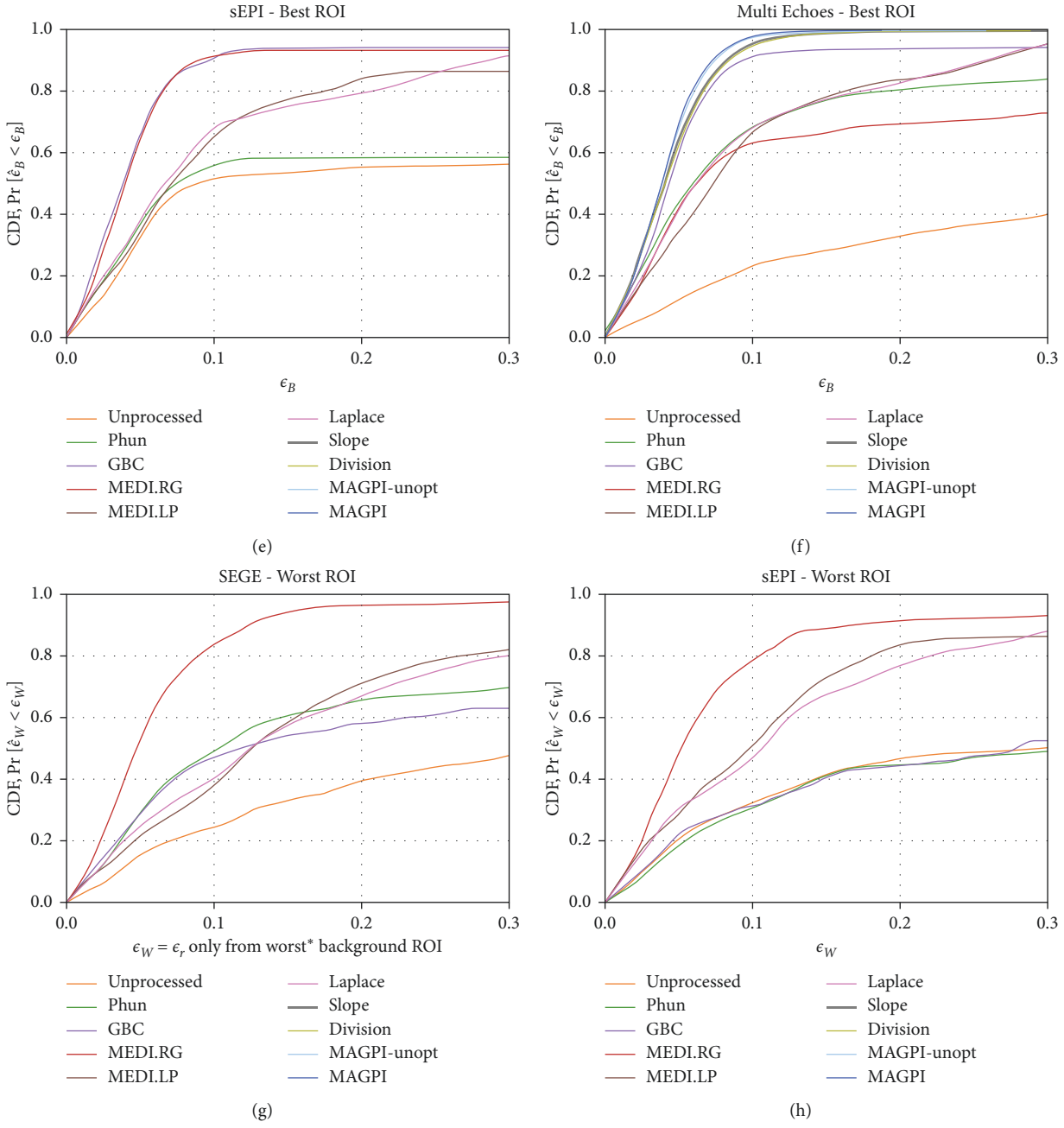
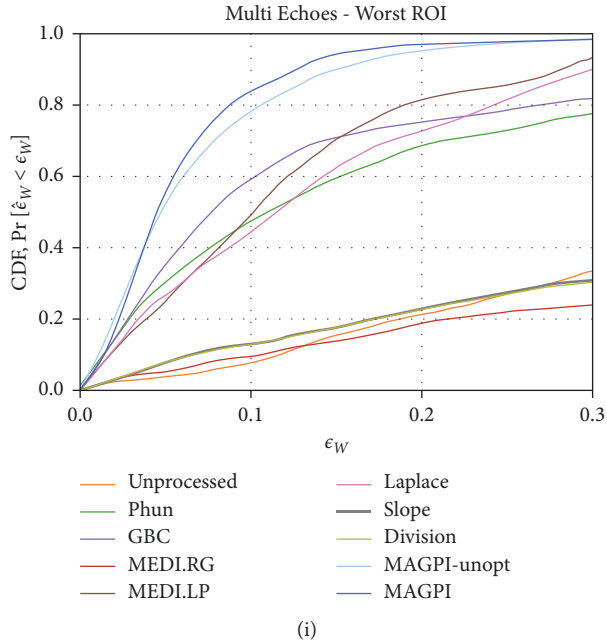


Figure 6: Continued.



(i)

Figure 6: The CDFs for each acquisition and phase estimation method pair. Each column corresponds to a different type of acquisition: single-echo acquisitions (SEGE) are shown in the first column, accelerated acquisitions with sEPI in the second, and multiecho acquisitions (MEGE and MAGPI) in the third. The first row (a–c) shows the CDF when all the error data is pooled, over all voxels, slices, tubes, and background ROIs. Given the large variability in performance for different background ROIs, we show in the second (d–f) and third (g–i) rows the CDFs obtained in the ROIs with the maximum and minimum $\mathbf{F}_{\epsilon_r}(0.1)$, respectively. A highly reproducible method will have a similar curve shape across all plots.

Table 4: Probability of relative errors <0.1 in the background ROI that maximizes $\mathbf{F}_{\epsilon_r}(0.1)$ across sequences, scan variability (ID is as listed in Table 1), and phantom tube components. The tube contents are 1.0 mM $GdCl_3$ in tubes 1 and 2, 0.5 mM $GdCl_3$ in tubes 3 and 4, and 3.2 mM $CuSO_4$ in tube 5. The last column is the standard deviation of all $F_{\epsilon_{rB}}(0.1)$ for the given method.

Method	Sequence	Scan ID 1	Scan ID 2	Scan ID 3	Tube 1	Tube 2	Tube 3	Tube 4	Tube 5	Std $\mathbf{F}_{\epsilon_{rB}}(0.1)$
Unprocessed	SEGE	0.79	0.86	0.47	0.59	0.88	0.83	0.89	0.68	0.32
	MEGE	0.25	0.21	0.27	0.25	0.23	0.25	0.26	0.18	
	sEPI	0.51	0.55	—	1.00	0.00	0.86	0.06	0.77	
Phun	SEGE	0.91	0.96	0.68	0.98	0.88	0.83	0.87	0.88	0.24
	MEGE	0.67	0.73	0.64	0.89	0.51	0.54	0.66	0.80	
	sEPI	0.56	0.58	—	0.92	0.00	0.87	0.37	0.79	
GBC	SEGE	0.96	0.73	0.96	0.75	0.79	0.97	0.93	0.73	0.08
	MEGE	0.95	0.88	0.96	0.89	0.93	0.97	1.00	0.80	
	sEPI	0.89	0.97	—	0.95	0.89	0.90	1.00	0.82	
MEDI.RG	SEGE	0.97	0.96	0.97	1.00	0.94	1.00	1.00	0.89	0.20
	MEGE	0.24	0.76	0.65	0.63	0.61	0.64	0.67	0.63	
	sEPI	0.91	0.92	—	1.00	0.84	0.94	0.97	0.84	
MEDI.LP	SEGE	0.63	0.63	0.59	0.36	0.58	0.53	0.96	0.70	0.18
	MEGE	0.68	0.67	0.64	0.36	0.75	0.53	0.95	0.76	
	sEPI	0.65	0.66	—	0.36	0.63	0.53	1.00	0.78	
Laplace	SEGE	0.57	0.68	0.68	0.36	0.97	0.52	0.85	0.56	0.19
	MEGE	0.62	0.72	0.64	0.36	1.00	0.69	0.90	0.46	
	sEPI	0.66	0.74	—	0.40	0.93	0.58	0.89	0.60	
Slope	MEGE	0.96	0.95	0.97	0.99	1.00	0.95	0.99	0.82	0.06
Division	MEGE	0.95	0.94	0.96	0.99	1.00	0.95	0.99	0.78	0.07
MAGPI-unopt	MEGE	0.97	0.97	0.97	1.00	1.00	0.98	1.00	0.87	0.04
MAGPI	MAGPI	0.97	0.98	0.97	1.00	1.00	0.98	1.00	0.91	0.03

with the lowest relative standard deviation σ_{ϵ_r} were MEGE + MEDI.LP (8.9%), MEGE + MAGPI-unopt (4.8%), and MAGPI (4.1%). The detailed behavior of the error is captured by the CDF (or PDF) of the data (Figure 5). A summary of the CDF is in the second-to-last column of Table 3 where we show the probability of observing relative errors <10%, $F_{\epsilon_r}(10\%)$, which captures the frequency by which *relatively* acceptable errors occur. An advantage of the Laplace-based methods is that they had smooth phase maps with qualitatively no apparent phase jumps. However, analysis showed that Laplace phase images result in quantitatively larger errors (a low $F_{\epsilon_r}(10\%)$) than other methods, suggesting incorrect phase unwrapping results, similar to the work of Chen et al. [60]. Other methods with low $F_{\epsilon_r}(10\%)$ were the unprocessed phase data, MEGE + Slope/Div, and MEGE + MEDI.RG, which had large phase unwrapping errors in a significant proportion of the data. $F_{\epsilon_r}(10\%)$ is an arbitrary point at which we highlight the behavior of F_{ϵ_r} and does not represent the entirety of the distribution of error (or CDF). For example, $F_{\epsilon_r}(10\%)$ of SEGE + MEDI.RG was comparable to MEGE + MAGPI-unopt and MAGPI, despite the comparatively poorer (larger) μ_{ϵ_r} and σ_{ϵ_r} of SEGE + MEDI.RG. This is due to relative errors falling mostly within the chosen 10% threshold for these methods.

We explored the dependence of errors on frequency reference ROI location (Figure 6). Since phase estimation errors (particularly large errors) are undesirable anywhere in the FOV, any spatial variation of the CDF highlights the potential dependence of the method on user intervention and/or its automated processing. We show the range of $F_{\epsilon_r}(10\%)$ observed across the 13 different frequency reference ROIs in the last column of Table 3. The results suggest that the most repeatable methods across background ROIs are SEGE + MEDI.RG, sEPI + MEDI.RG, MEGE + MAGPI-unopt, and MAGPI.

The MEGE data was processed using four broad classes of postprocessing algorithms. We note the following about these algorithms:

(i) The results obtained with methods in Categories 1 and 2 were fundamentally similar. That is, additional 1D-temporal processing does not alter the performance of 3D spatial unwrapping methods (with the exception of Laplacian-based methods, which we discuss below). This redundancy is due to the inherent Nyquist limitation associated with the echo spacing. As a result, we focus on the distinctive results of Category 1: spatial phase unwrapping + weighted averaging of echoes (Table 3).

(ii) Some postprocessing methods used in MEGE Categories 1 and 2 performed more poorly with MEGE than with SEGE (e.g., MEDI.RG). We believe this is due to the relatively larger bandwidth used with MEGE acquisitions, resulting in noisier images at each echo. Higher BW acquisitions were needed with MEGE to accommodate temporal methods (MEGE Categories 3 and 4), which require short echo spacing. It is possible that MEGE + MEDI.RG would perform better with lower BW (wider echo spacing). Due to time/complexity constraints, we were unable to explore every possible MEGE variation that favors specific algorithms. This is a limitation of this study.

(iii) MEGE Category 3 methods (Division/Slope) were straightforward to apply but resulted in a wide range of errors. This is due to large errors observed in frequency reference ROIs where the underlying frequency-offset value is larger than what is allowable by the smallest echo spacing. While such errors are avoidable with shorter echo spacing, this is not always possible (as was the case here) due to hardware constraints on readout bandwidth, resolution, FOV, and so forth. While Slope and Division are straightforward to apply, they result in a suboptimal combination of echoes, with noisy phase estimates.

(iv) MEGE Category 4 methods have a similar performance to that of Category 3 methods. That is, spatial phase unwrapping did not seem to markedly improve the performance of temporal phase unwrapping. This is potentially due to the hard-to-unwrap noisy boundary lines observed with Category 3 methods, as shown in MEGE + Slope example in Figure 4, which are still present with Category 4 methods.

Finally, we explored the dependence of the CDF on scan variability and tube contents (Table 4), and we note the following:

(1) Not all sequence + method pairs were invariant to scan and/or tube content. sEPI + Phun and sEPI + Unprocessed had greatest variability, followed by all of the Laplace-based methods (LP, MEDI.LP), indicating that some methods produced inconsistent phase even in their best-case scenario. Laplace-based techniques generated smoothly varying, though numerically inaccurate phase maps, irrespective of sequence, and sEPI + Phun suffered from sequence-dependent errors.

(2) Considering the best-case ROI scenario, we did not observe consistent performance differences between pairing methods with either SEGE or sEPI. The sEPI sequence provides a significant acceleration over SEGE via its segmented GRE approach and performed better with some methods (GBC) and worse with others (Phun).

(3) Table 4 compares Sequence + Method pair performance as a function of variability of scan parameters. For example, MEGE + Slope performed consistently well ($F_{\epsilon_{rB}}(10\%) > 0.95$) irrespective of the scan type used, while SEGE + Laplace was not always able to estimate the correct phase, irrespective of scan. SEGE + GBC, though, only struggled with Scan 2, which had a shorter TR than Scan 1.

(4) The same analysis can be applied to tube contents. MAGPI achieved $F_{\epsilon_{rB}}(10\%)$ close to 1.0 for all tubes and $F_{\epsilon_{rB}}(10\%) = 0.91$ in Tube 5. sEPI + Phun had inconsistent performance across tubes of similar

content (Tubes 1, 2, 3, and 4). Among the methods that maximized $F_{\epsilon_r}(0.1)$, Tube 5 exhibited lower $F_{\epsilon_{rB}}(10\%)$ compared to other tubes, which could be due to challenges defining $\Delta\chi_{\text{th}}$ and the complexity of $CuSO_4$ compared to $GdCl_3$.

Overall, the consistently good performance across tube content, scan, and rotation angle in the absence of phase estimation errors validates the $\Delta\chi$ model (equation (1)) and demonstrates that the rotating-tube phantom itself did not introduce unexpected detrimental effects to the phase measurements. In previous work, QSM performance improved with higher isotropic spatial resolution and higher coverage [56, 61]. Here, slice thickness was 2.0 mm for all scans, which may explain why performance did not drastically change with resolution. Additionally, Zhou et al. [61] and Karsa et al. [56] examined the entire QSM process, including inversion, which was not addressed here. Olsson et al. [22] used a phantom with tubes of Gd in comparable concentrations, though only one tube was used when varying angle with respect to B_0 (five angles). That study used one method to estimate QSM [62–64] over multiple spatial resolutions, volumes, and inversion parameters. Similar to the work of Karsa et al. [56], Olsson et al.'s results improved with increasing resolution and volume coverage, and, similar to our results, Olsson et al. observed errors in phase estimation using [47], compared to the theoretical result.

A limitation of this work is that the geometric structure of the phantom was not identical to that encountered *in vivo*. *In vivo* imaging may present different sources of phase errors not included here (e.g., eddy currents, susceptibility-induced signal drops), and the phantom could present some challenges that are not encountered or are less prominent in the brain. This is a common problem with nearly all phantom studies, and it is counterbalanced by the advantage of having a known truth, which is not possible *in vivo*. Errors observed in phantoms are frequently observed *in vivo*, even when the geometry of the phantom is a gross simplification of human anatomy. For example, in T_1 estimation, while phantoms can be used to refine a method, errors are amplified when methods are applied *in vivo* [65]. The many parameters, sequences, and processing steps considered here are useful to evaluate the robustness of phase estimation techniques, and this work can be viewed as a complement to other efforts seeking to evaluate the accuracy of QSM techniques, such as the use of simulated data [66, 67].

We introduced a wide range of variability to test repeatability and reproducibility of many data acquisition scenarios. While the performance of each method could be improved with additional "intervention" and potentially adapting the acquisition parameters to the intended post-processing methods to be used later, our intent was to assess the ability of existing techniques in diverse imaging scenarios encountered in reality. The degree to which Steps 3 and 4 of QSM are sensitive to the errors introduced in Steps 1 and 2 requires further investigation. This phantom validation study allowed us to set a quantitative limit on the performance of various Steps 1 and 2 methods. Ongoing work focuses on evaluating the performance of a subset of these methods, paired with Steps 3 and 4 methods.

5. Conclusion

In this work, we used a rotating-tube phantom to explore sources of error in QSM data acquisition and phase estimation. To assess the robustness and repeatability of methods, we did not manually intervene. The two most impactful parameters on reproducibility of measurements were (a) acquisition protocol (e.g., single echo or multiple echoes) and (b) phase errors. The most repeatable and reproducible approaches were MAGPI and MAGPI-unopt, both methods based on the maximum-likelihood approach in phase estimation. For the remaining methods, performance varied greatly, even when systematically applied to the same underlying data from the same sequence or with the same method across different sequences. This assessment of which methods are repeatable and reproducible without manual intervention is an important step towards using QSM pipelines in clinical settings without experienced users.

Abbreviations

BW: Bandwidth
CDF: Cumulative Distribution Function
FOV: Field of view
FR: Frequency referencing
GBC: Goldstein's branch cut
GRE: Gradient echo
LP: Laplace
MAGPI: Maximum AmbiGuity distance for Phase Imaging
MEDI: Morphology enabled dipole inversion
MEGE: Multiecho gradient echo
QSM: Quantitative Susceptibility Mapping
RG: Region growing
ROI: Region of interest
SEGE: Single-echo gradient echo
sEPI: Segmented Echo Planar Imaging.

Data Availability

The authors are performing subsequent analysis of the data for an extension of this paper. After completion, the data and code will be made publicly available.

Disclosure

A preprint has previously been published [68]. Certain commercial instruments and software are identified to specify the experimental study adequately. This does not imply endorsement by NIST or that the instruments and software are the best available for the purpose.

Conflicts of Interest

The authors K. E.K, B. P.B, S. C., S. E.R., W.-T.W., J. A.B, and D. L.P. have no conflicts of interest. J. D. is the inventor of the MAGPI method and is the founder of MAGPI LLC.

Acknowledgments

The authors W.-T.W., J. A.B., and D. L.P. would like to acknowledge funding from the Department of Defense in the Center for Neuroscience and Regenerative Medicine and the intramural research program of the National Institutes of Health. The authors K. E.K, S. C., and S. E.R. acknowledge funding from their employer, the National Institute of Standards and Technology. The authors B. P.B. and J. D. acknowledge funding from their employer, the MITRE Corporation.

Supplementary Materials

A supplementary material file is included which provides additional details on the frequency referencing methods (Section S.1) and the determination of reference values (Section S.2). (*Supplementary Materials*)

References

[1] Y. Wang and T. Liu, "Quantitative susceptibility mapping (QSM): decoding MRI data for a tissue magnetic biomarker," *Magnetic Resonance in Medicine*, vol. 73, no. 1, pp. 82–101, 2015.

[2] E. M. Haacke, N. Y. C. Cheng, M. J. House et al., "Imaging iron stores in the brain using magnetic resonance imaging," *Magnetic Resonance Imaging*, vol. 23, no. 1, pp. 1–25, 2005.

[3] I. R. Young, S. Khenia, D. G. T. Thomas et al., "Clinical magnetic susceptibility mapping of the brain," *Journal of Computer Assisted Tomography*, vol. 11, no. 1, pp. 2–6, 1987.

[4] J. H. Jensen, K. Szulc, C. Hu et al., "Magnetic field correlation as a measure of iron-generated magnetic field in homogeneities in the brain," *Magnetic Resonance in Medicine*, vol. 61, no. 2, pp. 481–485, 2009.

[5] K. Shmueli, J. A. de Zwart, P. van Gelderen, T.-Q. Li, S. J. Dodd, and J. H. Duyn, "Magnetic susceptibility mapping of brain tissue in vivo using MRI phase data," *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1510–1522, 2009.

[6] H. Tan, T. Liu, Y. Wu et al., "Evaluation of iron content in human cerebral cavernous malformation using quantitative susceptibility mapping," *Investigative Radiology*, vol. 49, no. 7, pp. 498–504, 2014.

[7] E. M. Haacke, K. Prabhakaran, I. R. Elangovan, Z. Wu, and J. Neelavalli, "Oxygen saturation: quantification, susceptibility weighted imaging in MRI," in *MRI: Basic Concepts and Clinical Applications*, M. Haacke and J. Reichenbach, Eds., John Wiley & Sons, Inc, Hoboken, NJ, USA, pp. 517–528, 2011.

[8] S. D. Sharma, D. Hernando, D. E. Horng, and S. B. Reeder, "Quantitative susceptibility mapping in the abdomen as an imaging biomarker of hepatic iron overload," *Magnetic Resonance in Medicine*, vol. 74, no. 3, pp. 673–683, 2015.

[9] D. Bonekamp, P. B. Barker, R. Leigh, P. C. M. van Zijl, and X. Li, "Susceptibility-based analysis of dynamic gadolinium bolus perfusion MRI," *Magnetic Resonance in Medicine*, vol. 73, no. 2, pp. 544–554, 2015.

[10] E. Lind, L. Knutsson, R. Kämpe, F. Ståhlberg, and R. Wirestam, "Assessment of MRI contrast agent concentration by quantitative susceptibility mapping (QSM): application to estimation of cerebral blood volume during steady state," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 30, no. 6, pp. 555–566, 2017.

[11] K. Deh, T. D. Nguyen, S. Eskreis-Winkler et al., "Reproducibility of quantitative susceptibility mapping in the brain at two field strengths from two vendors," *Journal of Magnetic Resonance Imaging*, vol. 42, no. 6, pp. 1592–1600, 2015.

[12] K. Deh, K. Kawaji, M. Bulk et al., "Multicenter reproducibility of quantitative susceptibility mapping in a gadolinium phantom using MEDI+0 automatic zero referencing," *Magnetic Resonance in Medicine*, vol. 81, no. 2, pp. 1229–1236, 2019.

[13] X. Feng, A. Deistung, and J. R. Reichenbach, "Quantitative susceptibility mapping (QSM) and R 2 ∗ in the human brain at 3 T," *Zeitschrift für Medizinische Physik*, vol. 28, no. 1, pp. 36–48, 2018.

[14] T. Hinoda, Y. Fushimi, T. Okada et al., "Quantitative susceptibility mapping at 3 T and 1.5 T," *Investigative Radiology*, vol. 50, no. 8, pp. 522–530, 2015.

[15] P.-Y. Lin, T.-C. Chao, and M.-L. Wu, "Quantitative susceptibility mapping of human brain at 3T: a multisite reproducibility study," *American Journal of Neuroradiology*, vol. 36, no. 3, pp. 467–474, 2015.

[16] M. D. Santin, M. Didier, R. Valabrègue et al., "Reproducibility of R2∗ and quantitative susceptibility mapping (QSM) reconstruction methods in the basal ganglia of healthy subjects," *NMR in Biomedicine*, vol. 30, no. 4, Article ID e3491, 2017.

[17] S. D. Robinson, K. Bredies, D. Khabipova, B. Dymerska, J. P. Marques, and F. Schweser, "An illustrated comparison of processing methods for MR phase imaging and QSM: combining array coil signals and phase unwrapping," *NMR in Biomedicine*, vol. 30, no. 4, Article ID e3601, 2017.

[18] R. Wang, G. Xie, M. Zhai et al., "Stability of R2∗ and quantitative susceptibility mapping of the brain tissue in a large scale multi-center study," *Scientific Reports*, vol. 28, no. 1, Article ID 45261, 2017.

[19] M. Lancione, G. Donatelli, P. Cecchi, M. Cosottini, M. Tosetti, and M. Costagli, "Echo-time dependency of quantitative susceptibility mapping reproducibility at different magnetic field strengths," *NeuroImage*, vol. 197, pp. 557–564, 2019.

[20] C. Langkammer, F. Schweser, K. Shmueli et al., "Quantitative susceptibility mapping: report from the 2016 reconstruction challenge," *Magnetic Resonance in Medicine*, vol. 79, no. 3, pp. 1661–1673, 2018.

[21] Z. Liu, Y. Kee, D. Zhou, Y. Wang, and P. Spincemaille, "Preconditioned total field inversion (TFI) method for quantitative susceptibility mapping," *Magnetic Resonance in Medicine*, vol. 78, no. 1, pp. 303–315, 2017.

[22] E. Olsson, R. Wirestam, and E. Lind, "MRI-based quantification of magnetic susceptibility in gel phantoms: assessment of measurement and calculation accuracy," *Radiology Research and Practice*, vol. 2018, Article ID 6709525, 13 pages, 2018.

[23] V. Fortier and I. R. Levesque, "Phase processing for quantitative susceptibility mapping of regions with large susceptibility and lack of signal," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3103–3113, 2018.

[24] H. E. Erdevig, S. E. Russek, S. Carnicka, K. F. Stupic, and K. E. Keenan, "Accuracy of magnetic resonance based susceptibility measurements," *AIP Advances*, vol. 7, no. 5, Article ID 056718, 2017.

[25] Y.-C. N Cheng, C.-Y. Hsieh, J. Neelavalli, and E. M. Haacke, "Quantifying effective magnetic moments of narrow cylindrical objects in MRI," *Physics in Medicine and Biology*, vol. 54, no. 22, pp. 7025–7044, 2009.

[26] C. Liu, W. Li, K. A. Tong, K. W. Yeom, and S. Kuzminski, "Susceptibility-weighted imaging and quantitative

susceptibility mapping in the brain," *Journal of Magnetic Resonance Imaging*, vol. 42, no. 1, pp. 23–41, 2015.

[27] S. Liu, S. Buch, Y. Chen et al., "Susceptibility-weighted imaging: current status and future directions," *NMR in Biomedicine*, vol. 30, no. 4, Article ID e3552, 2017.

[28] U. Katscher, T. Voigt, and C. Findeklee, "Electrical conductivity imaging using magnetic resonance tomography," in *Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3162–3164, Munich, Germany, September 2009.

[29] V. Rieke and K. Butts Pauly, "MR thermometry," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 2, pp. 376–390, 2008.

[30] S. Ley, D. Mereles, M. Puderbach et al., "Value of MR phase-contrast flow measurements for functional assessment of pulmonary arterial hypertension," *European Radiology*, vol. 17, no. 7, pp. 1892–1897, 2007.

[31] R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic resonance elastography by direct visualization of propagating acoustic strain waves," *Science*, vol. 269, no. 5232, pp. 1854–1857, 1995.

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

[33] E. M. Haacke, S. Liu, S. Buch, W. Zheng, D. Wu, and Y. Ye, "Quantitative susceptibility mapping: current status and future directions," *Magnetic Resonance Imaging*, vol. 33, no. 1, pp. 1–25, 2015.

[34] P. Sati, D. Thomasson, N. Li et al., "Rapid, high-resolution, whole-brain, susceptibility-based MRI of multiple sclerosis," *Multiple Sclerosis Journal*, vol. 20, no. 11, pp. 1464–1470, 2014.

[35] W. Feng, J. Neelavalli, and E. M. Haacke, "Catalytic multiecho phase unwrapping scheme (CAMPUS) in multiecho gradient echo imaging: removing phase wraps on a voxel-by-voxel basis," *Magnetic Resonance in Medicine*, vol. 70, no. 1, pp. 117–126, 2013.

[36] G. Helms and P. Dechent, "Increased SNR and reduced distortions by averaging multiple gradient echo signals in 3D FLASH imaging of the human brain at 3T," *Journal of Magnetic Resonance Imaging*, vol. 29, no. 1, pp. 198–204, 2009.

[37] J. Dagher and K. Nael, "MAGPI: a framework for maximum likelihood MR phase imaging using multiple receive coils," *Magnetic Resonance in Medicine*, vol. 75, no. 3, pp. 1218–1231, 2016.

[38] J. Dagher and K. Nael, "MR phase imaging with bipolar acquisition," *NMR in Biomedicine*, vol. 30, no. 4, Article ID e3523, 2017.

[39] A. M. Halefoglu and D. M. Yousem, "Susceptibility weighted imaging: clinical applications and future directions," *World Journal of Radiology*, vol. 10, no. 4, pp. 30–45, 2018.

[40] C. J. G. Bakker, H. de Leeuw, and P. R. Seevinck, "Selective depiction of susceptibility transitions using Laplace-filtered phase maps," *Magnetic Resonance Imaging*, vol. 30, no. 5, pp. 601–609, 2012.

[41] N.-k. Chen and A. M. Wyrwicz, "Correction for EPI distortions using multi-echo gradient-echo imaging," *Magnetic Resonance in Medicine*, vol. 41, no. 6, pp. 1206–1213, 1999.

[42] R. M. Goldstein, H. A. Zebker, and C. L. Werner, "Satellite radar interferometry: two-dimensional phase unwrapping," *Radio Science*, vol. 23, no. 4, pp. 713–720, 1988.

[43] K. Lu, T. T. Liu, and M. Bydder, "Optimal phase difference reconstruction: comparison of two methods," *Magnetic Resonance Imaging*, vol. 26, no. 1, pp. 142–145, 2008.

[44] M. A. Schofield and Y. Zhu, "Fast phase unwrapping algorithm for interferometric applications," *Optics Letters*, vol. 28, no. 14, pp. 1194–1196, 2003.

[45] Y. Wang, "Cornell MEDI Toolbox," http://pre.weill.cornell.edu/mri/pages/qsm.html.

[46] S. Witoszynskyj, A. Rauscher, J. R. Reichenbach, and M. Barth, "Phase unwrapping of MR images using ΦUN-a fast and robust region growing algorithm," *Medical Image Analysis*, vol. 13, no. 2, pp. 257–268, 2009.

[47] R. Cusack and N. Papadakis, "New robust 3-D phase unwrapping algorithms: application to magnetic field mapping and undistorting echoplanar images," *NeuroImage*, vol. 16, no. 3 Pt 1, pp. 754–764, 2002.

[48] D. L. Parker, A. Payne, N. Todd, and J. R. Hadley, "Phase reconstruction from multiple coil data using a virtual reference coil," *Magnetic Resonance in Medicine*, vol. 72, no. 2, pp. 563–569, 2014.

[49] B. Wu, W. Li, A. V. Avram, S.-M. Gho, and C. Liu, "Fast and tissue-optimized mapping of magnetic susceptibility and T2 with multi-echo and multi-shot spirals," *NeuroImage*, vol. 59, no. 1, pp. 297–305, 2012.

[50] J. L. R. Andersson, M. S. Graham, I. Drobnjak, H. Zhang, and J. Campbell, "Susceptibility-induced distortion that varies due to motion: correction in diffusion MR without acquiring additional data," *NeuroImage*, vol. 171, pp. 277–295, 2018.

[51] J. A. Lundman, M. Bylund, A. Garpebring, C. Thellenberg Karlsson, and T. Nyholm, "Patient-induced susceptibility effects simulation in magnetic resonance imaging," *Physics and Imaging in Radiation Oncology*, vol. 1, pp. 41–45, 2017.

[52] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller, "The NMR phased array," *Magnetic Resonance in Medicine*, vol. 16, no. 2, pp. 192–225, 1990.

[53] B. Gruber, M. Froeling, T. Leiner, and D. W. J. Klomp, "RF coils: a practical guide for nonphysicists," *Journal of Magnetic Resonance Imaging*, vol. 48, no. 3, pp. 590–604, 2018.

[54] K. C. Chu, Y. Xu, J. A. Balschi, and C. S. Springer Jr, "Bulk magnetic susceptibility shifts in NMR studies of compartmentalized samples: use of paramagnetic reagents," *Magnetic Resonance in Medicine*, vol. 13, no. 2, pp. 239–262, 1990.

[55] C.-Y. Hsieh, Y.-C. N. Cheng, J. Neelavalli, E. M. Haacke, and R. J. Stafford, "An improved method for susceptibility and radius quantification of cylindrical objects from MRI," *Magnetic Resonance Imaging*, vol. 33, no. 4, pp. 420–436, 2015.

[56] A. Karsa, S. Punwani, and K. Shmueli, "The effect of low resolution and coverage on the accuracy of susceptibility mapping," *Magnetic Resonance in Medicine*, vol. 81, no. 3, pp. 1833–1848, 2019.

[57] M. J. Cronin, N. Wang, K. S. Decker, H. Wei, W.-Z. Zhu, and C. Liu, "Exploring the origins of echo-time-dependent quantitative susceptibility mapping (QSM) measurements in healthy tissue and cerebral microbleeds," *NeuroImage*, vol. 149, pp. 98–113, 2017.

[58] Y. Kanazawa, Y. Matsumoto, M. Harada, H. Hayashi, T. Matsuda, and H. Otsuka, "Appropriate echo time selection for quantitative susceptibility mapping," *Radiological Physics and Technology*, vol. 12, no. 2, pp. 185–193, 2019.

[59] W. Zheng, H. Nichol, S. Liu, Y.-C. N. Cheng, and E. M. Haacke, "Measuring iron in the brain using quantitative susceptibility mapping and X-ray fluorescence imaging," *NeuroImage*, vol. 78, pp. 68–74, 2013.

[60] M.-C. Hsieh, C.-Y. Tsai, M.-C. Liao, J.-L. Yang, C.-H. Su, and J.-H. Chen, "Quantitative susceptibility mapping-based microscopy of magnetic resonance venography (QSM-mMRV) for in vivo morphologically and functionally assessing cerebromicrovasculature in rat stroke model," *PLoS One*, vol. 11, no. 3, Article ID e0149602, 2016.

[61] D. Zhou, J. Cho, J. Zhang, P. Spincemaille, and Y. Wang, "Susceptibility underestimation in a high-susceptibility phantom: dependence on imaging resolution, magnitude contrast, and other parameters," *Magnetic Resonance in Medicine*, vol. 78, no. 3, pp. 1080–1086, 2017.

[62] L. de Rochefort, T. Liu, B. Kressler et al., "Quantitative susceptibility map reconstruction from MR phase data using bayesian regularization: validation and application to brain imaging," *Magnetic Resonance in Medicine*, vol. 63, no. 1, pp. 194–206, 2010.

[63] J. Liu, T. Liu, L. de Rochefort et al., "Morphology enabled dipole inversion for quantitative susceptibility mapping using structural consistency between the magnitude image and the susceptibility map," *NeuroImage*, vol. 59, no. 3, pp. 2560–2568, 2012.

[64] T. Liu, C. Wisnieff, M. Lou, W. Chen, P. Spincemaille, and Y. Wang, "Nonlinear formulation of the magnetic field to source relationship for robust quantitative susceptibility mapping," *Magnetic Resonance in Medicine*, vol. 69, no. 2, pp. 467–476, 2013.

[65] N. Stikov, M. Boudreau, I. R. Levesque, C. L. Tardif, J. K. Barral, and G. B. Pike, "On the accuracy of T1mapping: searching for common ground," *Magnetic Resonance in Medicine*, vol. 73, no. 2, pp. 514–522, 2015.

[66] Q. S. M. C. O. Committee, B. Bilgic, C. Langkammer et al., "QSM reconstruction challenge 2.0: design and report of results," *Magnetic Resonance in Medicine*, vol. 86, no. 3, pp. 1241–1255, 2021.

[67] J. P. Marques, J. Meineke, C. Milovic et al., "QSM reconstruction challenge 2.0: a realistic in silico head phantom for MRI data simulation and evaluation of susceptibility mapping procedures," *Magnetic Resonance in Medicine*, vol. 86, no. 1, pp. 526–542, 2021.

[68] K. E. Keenan, B. P. Berman, and S. C. Carnicka, "Validation of data acquisition and phase estimation for quantitative susceptibility mapping with a rotating-tube phantom," *bioRxiv*, 2021, preprint.