The 2016 QSM Challenge: Lessons learned and considerations for a future challenge design

Carlos Milovic[1,2,3], Cristian Tejos[1,2,3], Julio Acosta-Cabronero[4], Pinar Senay Özbay[5], Ferdinand Schwesser[6,7], Jose Pedro Marques[8], Pablo Irarrazaval[1,3,9], Berkin Bilgic[10], Christian Langkammer[11]

[1] Department of Electrical Engineering, Pontificia Universidad Catolica de Chile, Santiago, Chile
[2] Biomedical Imaging Center, Pontificia Universidad Catolica de Chile, Santiago, Chile
[3] Millennium Nucleus for Cardiovascular Magnetic Resonance, Santiago, Chile
[4] Tenoke Ltd., Cambridge, UK
[5] Laboratory of Functional and Molecular Imaging, National Institutes of Health, Bethesda, MD, USA
[6] Buffalo Neuroimaging Analysis Center, Department of Neurology, Jacobs School of Medicine and Biomedical Sciences at the University at Buffalo, The State University of New York, NY, USA
[7] Center for Biomedical Imaging, Clinical and Translational Science Institute, University at Buffalo, The State University of New York, NY, USA
[8] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands.
[9] Institute for Biological and Medical Engineering, Pontificia Universidad Catolica de Chile, Santiago, Chile
[10] Martinos Center for Biomedical Imaging, Harvard Medical School, MA, USA
[11] Department of Neurology, Medical University of Graz, Graz, Austria

Corresponding author: Carlos Milovic, cmilovic@uc.cl

Manuscript type: Full paper

Word count: 5069/5000

Running title: Lessons learned from the 2016 QSM Challenge

Abstract

Purpose*:** The 4th International Workshop on MRI Phase Contrast & QSM (2016, Graz) hosted the first QSM challenge. A single-orientation GRE acquisition was provided, along with COSMOS and the χ_{33} STI component as ground-truths. The submitted solutions differed more than expected depending on the error metric used for optimization and tended to over-regularization. This raised (unanswered) questions about the ground-truths and the metrics utilized. ***Methods: We investigated the influence of background field remnants by applying additional filters. We also estimated the anisotropic contributions from the STI tensor to the apparent susceptibility, to amend the χ_{33} ground-truth and to investigate the impact on the reconstructions. Lastly, we used forward simulations from the COSMOS reconstruction to investigate the impact that has noise on the metrics. ***Results***: Reconstructions compared against the amended STI ground-truth returned lower errors. We show that the background field remnants had a minor impact in the errors. In the absence of inconsistencies, all metrics converged to the same regularization weights, whereas SSIM was more insensitive to such inconsistencies. ***Conclusion***: There was a mismatch between the provided data and the ground-truths due to the presence of unaccounted anisotropic susceptibility contributions and noise. Given the lack of reliable ground-truths when using in vivo acquisitions, simulations are suggested for future QSM challenges.

Keywords: Magnetic susceptibility, quantitative susceptibility mapping, total variation, FANSI.

Introduction

Quantitative Susceptibility Mapping (QSM) is a relatively new MRI technique, where susceptibility changes in tissues are conventionally estimated from Gradient Recalled Echo (GRE) acquisitions[1,2]. The phase of the GRE signal is proportional to changes in the local (macroscopic) magnetic field induced by the underlying distribution of magnetic susceptibility in the tissue[3,4]. The calculation of the magnetic susceptibility distribution from a given magnetization field is an ill-conditioned and also ill-posed inverse problem[5,6]. One of the reasons for the ill-conditioning of the problem is that the dipole convolution kernel, which describes the susceptibility-to-field problem, decays to zero around a double-shaped conical surface at the magic angle around the main field direction. A straightforward inversion implies a division by small numbers, resulting in severe noise-amplification[5,7]. Reasons for the ill-posedness of the problem include the limited field of view coverage, signal dropout due to fast T_2* decay (or lack of signal in bone and air regions), and non-Gaussian noise behavior[8–10]. The majority of the most recently published QSM techniques rely on solving the inverse problem by minimizing an objective function that includes a regularization term. Other QSM techniques rely on modifications to the dipole kernel, or to exploit approaches derived from the Compressed Sensing theory[11]. All these techniques try to reduce streaking artifacts and to prevent noise amplification during the inversion process[12]. Other methodological studies focus on preprocessing steps, such as the removal of background fields (magnetization fields originated from objects outside of the field of view)[13,14], phase unwrapping and multi-echo / multi-coil combination techniques[15,16]. The integration of those techniques and the inversion problem into a single process is referred as the so-called single-step algorithms[17–23].

In addition to those methodological papers, several clinical studies using QSM have been conducted, some of them combining well-established relaxometry measures such as T2*/R2* [24–29]. Generally, it has been observed that the quantified susceptibility values vary depending on the employed methodology[30].

Given such a variety of methods and associated results, a QSM Challenge was proposed in the context of the 4th International Workshop on MRI Phase Contrast & QSM (Graz, Austria, September 2016)[31]. The challenge aimed at defining a faithful reconstruction of magnetic susceptibility. A single head orientation GRE acquisition (from a multi-orientation set) was released to the contestants to test their algorithms. The whole multi-orientation set was used to calculate the susceptibility tensor[32] (STI) and a Calculation Of Susceptibility through Multiple Orientation Sampling (COSMOS)[33] reconstruction. While the χ_{33} term of the susceptibility tensor was used as ground-truth, the COSMOS reconstruction was also provided. Anisotropic elements of the susceptibility tensor (χ_{13} and χ_{23}) were discarded,

assuming that their contributions were negligible in an acquisition with an orientation parallel to the main field. Since most of the algorithms depend on one or more parameters that must be fine-tuned, the ground truth was provided to allow for optimization of the parameters, ensuring that submitted susceptibility maps represent the best results that can be achieved with a certain algorithm. Totally 27 reconstructed QSM images were submitted by 15 teams, and 4 metrics were evaluated (Root Mean Square Error - RMSE, High Frequency Error Norm - HFEN, Structural Similarity Index Metric - SSIM, and error in anatomical ROIs) in separate categories. Henceforth, teams optimized their results to minimize (or maximize in the case of SSIM) a certain quality metric. It turned out that most of numerically well performing approaches produced (visually) over-regularized susceptibility maps either blurry or piece-wise smooth. Reconstructions (and the optimal parameter values) also varied upon the employed quality metric. Furthermore, many approaches performed well according to a particular metric, but poorly according to others. Subsequent discussions and analyses pointed at several possible sources of inconsistency that might have degraded the given phase data and therefore its correspondence to the provided ground truth susceptibility. These effects include incomplete background field removal, noise, anisotropic susceptibility and microstructure contributions, flow and motion artifacts, and co-registration (i.e. misalignment of the multiple orientation acquisitions) errors. In addition, questions were raised about the behavior of the selected quality metrics, mainly whether or not they could provide the same optimal parameters.

In this paper, we investigated the viability of the provided χ_{33} susceptibility map as ground truth for single orientation acquisitions and compared it with the COSMOS reconstruction and an STI derived scalar projection. We also investigated the impact of noise and the influence of background field residuals in the reconstructions.

Methods

MRI data

The 2016 QSM Reconstruction Challenge dataset (available at http://qsm.neuroimaging.at) included rf-spoiled 3D-GRE scans acquired on a 3T system (Siemens Tim Trio) using a 32-channel head-coil. Data were acquired with 1.06-mm isotropic voxels, 15-fold Wave-CAIPI[34] acceleration, 240×196×120 matrix size, echo time (TE)/repetition time (TR)=25/35ms, and flip angle=15°. Twelve head orientations were used for COSMOS and STI calculations. Challenge participants were provided with a single transverse-oriented acquisition (from the multi-orientation dataset), the raw phase and a

background filtered phase, additionally containing the anatomical MPRAGE[35] data[31]. Unwrapping was performed using a Laplacian algorithm[36], while background removal was performed using the Laplacian Boundary Value (LBV)[37] approach. Since this method is not able to completely remove transmit/receiver coil contributions, a polynomial fit was also used to remove these components[31,38]. In the analyses presented here, we worked on the same data set as the one provided by the challenge organizers. We also used the same error metrics (RMSE, HFEN and SSIM) and parameters, as those available in[31] (http://www.neuroimaging.at/pages/qsm.php). Please note that RMSE values are normalized by the L2-norm of the given ground truth, as percentages. The HFEN metric is also normalized, by the L2-norm of the high frequency coefficients of the ground truth. The SSIM is not normalized.

Experiment design

1. Background field remnants.

Background field removal is one part of the QSM pipeline where results may vary substantially depending on the algorithm used. The QSM challenge data was processed using LBV and polynomial fit steps. To evaluate the quality of the provided local phase, and how background field remnants could impact in the error metrics, we included two additional methods to reduce the remaining background fields. The first method is the recently proposed Weak-harmonics QSM (WH-QSM) reconstruction[19], where an estimate of the background field is jointly reconstructed with the susceptibility map, by imposing a weak-harmonics regularization to decouple terms. The second method is the popular Projection onto Dipole Fields[39] (PDF) algorithm, followed by a TV-FANSI reconstruction[10] (300 iterations, and $\mu = 100*\alpha$, μ is an internal Lagrangian weight and α is the regularization weight). Both methods were chosen to avoid further ROI reduction and to minimize over-filtration. We optimized the metric scores by an exhaustive search, for different regularization weights (α), with multiplicative steps of $10^{0.2}$. We compared these results with reconstructions performed using TV-FANSI and no additional background removal step. Error metrics were evaluated using χ_{33} and COSMOS as ground truth.

2. Ensuring the validity of the QSM physical model

2.1 Forward phase estimates with anisotropic components.

One way to compare the validity of the provided ground-truths is to estimate the differences between

the calculated magnetization field generated by such susceptibility maps with the single-orientation acquired phase. To account for anisotropic components, we calculated the whole susceptibility tensor from the in vivo dataset using all 12 orientations, provided by the challenge organizers, and then applied the full forward STI model to synthesize the phase of a single acquisition in the same orientation as the main field[32]:

$$\frac{F\Phi_{STI}}{\gamma H_0 TE} = \frac{1}{3} F\chi_{33} - \frac{k_z^2}{k^2} F\chi_{33} - \frac{k_z k_x}{k^2} F\chi_{13} - \frac{k_z k_y}{k^2} F\chi_{23}, \qquad \text{Eq. 1}$$

This simulated phase was compared with the acquired single orientation phase, as well as the simulated fields from the χ_{33} and COSMOS ground-truths (obtained by convolution with the dipole kernel). RMSE calculation was repeated for each case.

2.2 Solving the inverse-problem with anisotropic components.

The susceptibility distribution calculated by QSM methods is an apparent susceptibility that incorporates both isotropic and anisotropic components. The latter is the result of micro-structure effects and the tensor nature of the susceptibilities of certain tissues. Whereas some studies[40-42] have described how micro-structure affects phase contrast and its impact on both QSM and STI, we are interested in providing a scalar apparent susceptibility ground-truth from STI. For this effect, a projection of the STI data onto a scalar value is required. To assess the impact of the anisotropic components on the QSM estimations, and how it compares with the χ_{33} component, a new ground-truth must be derived. A direct division of the synthetic STI-derived phase (Eq. 1) by the dipole kernel yielded:

$$\chi_{STI} = \chi_{33} + F^{-1}\left(\frac{-k_z}{k^2 - 3k_z^2}\right)\left(k_x F\chi_{13} + k_y F\chi_{23}\right), \qquad \text{Eq. 2}$$

This may be interpreted as:

$$\chi_{STI} = \chi_{33} + \chi_{corr}, \qquad \text{Eq. 3}$$

where χ_{corr} is a correction term that incorporates projections of the anisotropic components. However, χ_{corr} cannot be analytically derived since it involves division by zero-valued coefficients within the "magic cone". Approximations to this correction term were computed using TKD[5] (threshold = 0.17)

and Tikhonov[6] (λ = 6E-2) regularization methods, to obtain two new proposed ground-truths: $\chi_{STI\text{-}TKD}$ and $\chi_{STI\text{-}L2}$. We chose the L2 and TKD algorithms for these correction terms to avoid giving TV-FANSI an unfair advantage if variational penalties were also applied as regularization in this step. Algorithms´ parameters were chosen for the apparent well-behavior when working with the single acquisition data and were directly applied for this correction factor estimation. QSM reconstructions were performed using TV-FANSI[10] (maximum 50 iterations), with an exhaustive search of the optimal regularization weight using multiplicative $10^{0.1}$ steps from 10^{-6} to 10^{-2}. Metric scores were evaluated using $\chi_{STI\text{-}TKD}$, $\chi_{STI\text{-}L2}$, χ_{33} and COSMOS as ground truth, respectively.

Alternatively, from Eq. 1, we may instead use the anisotropic χ_{13} and χ_{23} terms to forward simulate a phase correction term for the acquired data. This amended phase corresponds to Φ_{33}, the scalar magnetization of a susceptibility distribution χ_{33} (i.e. convolving by the dipole kernel). Then, it becomes possible to compare QSM reconstructions using χ_{33} as ground truth by using either the acquired phase or the amended phase as input data for the algorithm (Eq. 4).

$$F\Phi_{33} = \gamma H_0 TE\left(\frac{1}{3} - \frac{k_z^2}{k^2}\right) F\chi_{33} = \Phi_{acq} + \gamma H_0 TE\left(\frac{k_z k_x}{k^2} F\chi_{13} + \frac{k_z k_y}{k^2} F\chi_{23}\right) \qquad \text{Eq.4}$$

To see if background field remnants had still an effect in the reconstructions when anisotropic components are taken into consideration (the cumulative effect), we repeated the background removal analysis described earlier (section 1) for the best scoring (lowest RMSE/HFEN and larger SSIM) phase/ground-truth set.

3. Impact of SNR to metrics and ground-truth solutions.

Data provided for the challenge corresponded to a single-orientation acquired using Wave-CAIPI with a high acceleration factor, thus enabling relatively short acquisition times (all 12 head orientations were acquired within one hour). In turn, this high acceleration factor produced low SNR data. Statistical analysis of the real and imaginary components in the signal-void background gave a peak SNR between 40 to 60. In order to assess the influence of noise in the reconstruction quality (measured by the metrics used in the challenge) and the correlation between the predicted optimal parameters by each metric, we designed an analytic simulation experiment. We used COSMOS-based forward simulations, where the phase data were simulated by convolving with the dipole kernel. In these experiments, susceptibility values outside the brain were set to zero (the mean susceptibility value). We used the

calculated phase and the provided magnitude data (normalized between 0 and 1) to create real and imaginary images. Gaussian noise was added to each channel in the complex image domain, with SNR settings from 2 to 128 in a dyadic sequence. Noise-corrupted phase images were used directly as input for TV-FANSI reconstructions. Because only local phases were simulated, objects outside the region of interest were not included and consequently, unwrapping or background field removal methods were not required. Individual optimal reconstruction parameters were found for each error metric and SNR level.

The influence of background remnants and anisotropic components was also studied, on two phases forward simulated at SNR = 32 and SNR = 64, in concordance to our estimated SNR range for the in vivo acquisition. We included the remnants found by the PDF method, and the phase derived from the anisotropic tensor elements, χ_{13} and χ_{23} (Eq. 1) in this analysis. Metric scores were evaluated using TV-FANSI reconstructions and compared with the in vivo metric scores (obtained from the experiments described in previous sections).

In all the presented experiments, QSM reconstructions were evaluated with RMSE, HFEN and SSIM (with K = [0.01, 0.03] and L = 255 parameters[43], as used by the organizers of the challenge), with respect to each ground-truth. To provide further insights about quality metrics, we also include the analysis of metrics not used in the challenge, such as the correlation coefficient (CC), mutual information[44] (MI), root mean squared error of the gradient of susceptibilities (GXE), mean absolute deviation (MAD) and the mean absolute deviation of the gradient of susceptibilities (MADGX). GXE, MAD and MADGX errors are normalized by the L2-norm and L1-norm of the gradients of the ground truth, and the L1-norm of the ground truth, as correspond for each metric. Source code to evaluate all the described metrics is available at the FANSI-Toolbox site: http://gitlab.com/cmilovic/FANSI-Toolbox

Results

1. Background field remnants.

Estimations of the background field remnants using PDF and WH-QSM are presented in Figure 1. Visually, both methods yield similar remnants near the boundaries of the ROI, especially at lower and upper regions. Smooth remnants seem to be also present in deep brain areas. In both cases, remnant background phase information seems to be smaller in magnitude than the local phase data, revealing a

minor contribution.

Susceptibility calculations were performed using the provided local phase, and the local phase further filtered by PDF using TV-FANSI. The supplied phase was also used as input for WH-QSM, which jointly estimated the background and performed the susceptibility calculation. Error scores for these results are presented in Table 1.

***** Table 1 near here *****

Both background filtering methods show small improvements (lower than 5,5%) for all three metrics (except for HFEN, using PDF+TV-FANSI and COSMOS as ground truth which scored a 1,4% higher error). Optimal regularization weights (Table 2) were similar between methods (with WH-QSM obtaining slightly smaller optimal regularization weights for most metrics), and results were still visually over-regularized (Supplementary Figure S1). RMSE seems to be more sensible to background remnant errors than HFEN and SSIM.

***** Table 2 near here *****

2. Ensuring the validity of the QSM physical model

2.1 Forward phase estimates with anisotropic components.

The simulated field (i.e. the phase data) estimated from the STI equation (Eq. 1) returned lower RMSE values (relative to the phase from the single acquisition) than forward simulations using χ_{33} (-20.8%) and COSMOS (-9.4%), as seen in Figure 2. Phase differences are also reported in Supplementary Figure S2 with higher contrast. Notably, the STI-based phase error map (Figure 2D) contains no medium or large-scale anatomical features. Differences are more prominent in the venous system, which seem to be independent on the orientation of the vessels.

2.2 Solving the inverse-problem with anisotropic components.

Estimated $\chi_{corr\text{-}L2}$ and $\chi_{corr\text{-}STI}$ correction images are shown in Figure 3. These corrections are incorporated into the χ_{33} ground-truth (χ_{STI} maps), for optimization of the reconstructions with the single orientation acquisition. The χ_{STI} maps returned better quality metric scores than χ_{33}, and at least similar to COSMOS (Table 3). This indicates that the proposed χ_{STI} solutions constitute more consistent ground truths for single orientation QSM. Even so, the amended Φ_{33} phase with anisotropic components turned χ_{33} into the best ground truth in this comparison. Notably, optimal regularization parameters for QSM reconstructions (for a given target metric) were smaller using the STI-based ground-truths and the amended Φ_{33} phase, thus achieving less over-regularized results (Figure 4 and Supplementary Information Figure S4). In addition, regularization weights were more similar across different metrics than parameter optimizations performed with the provided ground-truths. This greater convergence of the metrics indicates that incompatibilities between the input data and the proposed ground-truth are being minimized.

***** Table 3 near here *****

Repeating the background removal analysis on this amended phase yielded small improvements (less than 2%) for susceptibility maps using WH-QSM and no improvement for PDF (see Supplementary Information Figure S5 and Table S1 for details). This is a minor improvement, compared to the ~15% gains shown in Table 3. We found a 4.8% RMSE improvement using the remnants calculated from Φ_{33} to correct the forward simulation of the STI phase. In contrast, the background remnant calculated from the acquired phase returned a 3.6% RMSE improvement when comparing to the forward simulations using χ_{33} and COSMOS. Both remnant maps are presented in Supplementary Figure S3.

3. Impact of SNR to metrics and ground-truth solutions.

TV-FANSI reconstructions and optimal metric scores using forward simulated phases with COSMOS as ground truth for SNR=2 to SNR=128 are displayed in Figure 5A-C. Simulations at higher SNR levels (10^3 and 10^6, not shown in Figure 5) reveal a minimal error level at approximately 17.6% RMSE, 15.7% HFEN and 0.99 SSIM. Scores are almost perfect for SNR=10^6 if no magnitude information is incorporated in the simulation of the complex normal noise (0.02% RMSE, 0.01 HFEN and 1.0 SSIM), revealing that this plateau is caused by the truncation of the magnetization fields due to existence of

regions without MRI signal. Simulations done at SNR = 20 yield errors comparable to those found in the best in-vivo scores, which is not the case of the supplied data (with an SNR at least in the 40-60 range, as shown in Supplementary Information Figure S6). In a realistic SNR range, RMSE and HFEN scores reveal that the noise level of the single acquisition cannot explain the scores obtained by reconstructions using χ_{33} as ground-truth. Both our χ_{STI} proposed ground-truth and working with the Φ_{33} amended phase reach a score level closer to the baseline.

In terms of the regularization weight (Figure 5D), optimal values for the HFEN metric tended to be the lowest in simulations, whereas SSIM yielded the lower values working with the in vivo data. Maximum differences in the optimal weights between metrics (at a given SNR) grows exponentially with noise, at a similar rate, with a high correlation between the metrics. Additional metrics (Supplementary Information Figure S7) seem also to be correlated. MI seemed to be the least reliable.

When background remnants and anisotropic components are taken into consideration in our simulations (Figure 6 and Supplementary Information Figure S8) the mismatch is closer in line with the behavior of the challenge data. Anisotropic components were the largest source of mismatch, with background remnants having a smaller contribution when both terms are considered together.

Exhaustive metric scores for reconstructions of the provided in vivo phase, using both χ_{STI} maps, χ_{33} and COSMOS as ground truth, are presented in Supplementary Information Figure S9. This figure includes all the analyzed metrics.

Discussion

The analysis of the submitted results to the 2016 QSM Reconstruction Challenge[31] raised a few issues about the challenge design (input data and ground truth) and the metrics. Errors (measured by the metrics) using COSMOS as ground-truth were significantly smaller to those using χ_{33}, in opposition to the assumption that χ_{33} was a better representation of the apparent susceptibility in the z-axis (by discarding any anisotropic contribution). Furthermore, most QSM submitted results were over-regularized. And finally, the regularization weights depended significantly on the metric chosen to be optimized for. Because of this, some algorithms high-scoring in one category performed worse in the others. To explain these observations, we conducted a series of experiments to determine the role of background remnants, anisotropic components and noise in the metric scores, and visual appeal of the

results.

Background filtered reconstructions show small improvements for error metrics, independent of the ground truth. This indicates that although background remnants were present in the provided local phase, their impact on the reconstructions is secondary to other effects. Optimal regularization weights were consistent between methods (with WH-QSM obtaining slightly smaller optimal weights).

Additionally, arguing against the relevance of background remnants is that all the multi-orientation data were filtered using the same pipeline (LBV + polynomial fit). Although this may introduce errors in the STI estimation, background ghosting or boundary errors should be mostly contained in the isotropic χ_{33} component, with a biased estimation similar to the one in the single-orientation reconstruction.

With background remnants discarded as main contributors to degradation of the metric scores, we analyzed the validity of assuming χ_{33} as the ground truth for a single orientation acquisition. The assumption that anisotropic magnetizations and microstructure effects were negligible was invalidated by comparing the forward simulated phases using χ_{33}, COSMOS, and the complete STI equation. RMSE of the phase calculated from χ_{33} was the highest, whereas the phase from the STI equation was the lowest (by 9.4% with respect to COSMOS and 20.8% to χ_{33}). The remnant phase map using the full STI model (Fig. 2D) has no large-scale features, and major discrepancies that are found in the vasculature may be associated with the blurring effect created by spatial interpolation in the co-registration step (in the process of calculating the STI tensor), or flow effects in this non-compensated acquisition. An additional source of uncertainty is the error generated at the STI estimation by the use of a limited set of orientations. As shown by Li et al[45], RMSE could rise up to 30% for acquisitions similar to those in the 2016 Challenge dataset (in terms of SNR and maximum angulation). Nevertheless, this error should account only for a fraction of the estimated anisotropy, given the structures visible in the remnant map, and could be related more to microstructure effects. Unfortunately, this source of error cannot be estimated due to the lack of a real ground truth in this in vivo setting. The remnant map also discards large co-registration errors as a potential source of mismatch, due to the lack of contiguous sharp edge features with a different sign that are created by the misalignment of structures. Some of the medium-scale features, especially near the para-nasal cavities and cerebellum, have similarities with the background remnants shown in Figure Supplementary S2. This explains why the metric scores overall improve when these fields are removed. The COSMOS forward simulated phase partially incorporated anisotropic components, and thus, reconstruction metrics using COSMOS as reference achieved overall better scores than using χ_{33}.

QSM reconstructions that incorporated the anisotropic tensor components (either by comparing to the

χ_{STI} ground truths or using the amended Φ_{33} phase) resulted in significantly lower errors than when using χ_{33} or COSMOS as ground truth. In practice, these tensor elements incorporate contributions from susceptibility anisotropic effects as micro-structure effects and chemical shielding. Since calculating the χ_{corr} term involves a dipole inversion, we used two algorithms that do not impose piece-wise constant solutions to avoid an unfair advantage to TV-FANSI in the metric scores. This is partially reflected in the reconstructions using Φ_{33} as input, where the anisotropic components were incorporated in the acquired phase and were jointly reconstructed, leading to the lowest error scores. Further background field subtraction improved the scores marginally. Optimal regularization weights associated with these reconstructions also were smaller (however, yielding more visually appealing results) and more consistent between metrics.

The analysis of the forward simulated reconstructions at different SNR scenarios gave more insights to explain the metric scores and the differences in their optimal values. Given that the actual SNR of the single orientation data is in the 40-60 range, noise does not sufficiently explain the measured (worse) metric scores in in vivo acquisitions. Optimal regularization weights were also smaller for our simulations in this noise range, yielding less over-regularization. While SSIM tended to return the lower regularization weights with in vivo data, in the simulations HFEN consistently yielded lower regularization weights, while SSIM tended to return similar values to those derived from using the RMSE. The discrepancy between the optimal regularization weights given by different metrics was also smaller for simulations. These tests revealed that SSIM is less sensitive to local discrepancies than HFEN and RMSE. HFEN promotes sharper (less regularization) results than SSIM and RMSE in the absence of local mismatches. L2 norm-based error metrics (HFEN and RMSE) tend to minimize the total energy of the mismatches, distributing it among neighboring pixels, thus promoting over-regularized results. Consequently, this is an indication that other sources for the mismatch between the acquisitions and the ground-truth are present, beyond noise.

By conducting forward simulations at SNR=32 and SNR=64, which is the expected noise level range of the in vivo acquisition, we explored the influence of background remnants and anisotropic components to the metric scores. This experiment revealed that anisotropic tensor components had a larger influence on the errors than background remnants. Both effects combined increased the errors to a level similar to those found in the challenge, confirming that these two were the major sources of mismatch. Optimal regularization weights also were larger when these effects were included, showing a similar behavior to the in vivo reconstructions for the RMSE and HFEN metrics. The SSIM metric,

on the other hand, revealed a larger invariance to these effects. In addition, other error sources might be responsible for minor mismatches of the in vivo reconstruction and ground truth. For example, blurring of the multiple-orientation data due to interpolation might also play a role in this discrepancy. Please note that we are ignoring noise and other effects in the STI or COSMOS estimations as sources of errors. By comparing our reconstruction to noisy or flawed ground-truths (due to the limited available range of angulation values) we are under-estimating the influence of noise and anisotropic contributions. Unfortunately, these errors are hard to estimate without an appropiate ground-truth.

Other global metrics may also be considered, as proposed in Supplementary Information Figure S7-S9, which presented similar trends to RMSE, HFEN, and SSIM. Such metrics may provide additional information about the source of the errors. For image analysis purposes, other alternatives to consider in the future may follow multiscale or vectorized metrics[46], or the inspection of the SSIM map, which reveal more information regarding the nature of the errors. Unfortunately, such metrics or methods are not directly applicable to rank results for a challenge contest as compared to scalar scores.

Finally, while other potential sources of errors such as vascular flux, motion, etc. might be also considered, it remains an open question on how to properly numerically compare single orientation with multiple orientation reconstructions. Methods to project the susceptibility tensor onto the apparent scalar susceptibility should be explored and validated in future investigations. These should considerate projections onto different head orientations to validate the methods. This may be performed by simply rotating all the images to set a new reference "central" acquisition, thus creating a new, rotated, susceptibility tensor to account for the anisotropic contributions to the new χ_{33} component. Experimental validation of these methods should also consider the errors incurred in the susceptibility tensor estimation, and estimate how noise impacts the tensor-to-scalar projection. Due to the limitations of in vivo STI acquisitions, this may be complemented with ex vivo acquisitions.

Conclusions

The 2016 QSM Reconstruction Challenge used a highly accelerated Wave-CAIPI GRE sequence to provide single and multiple orientation reconstructions. Despite the low SNR of the provided single orientation phase, noise alone cannot explain the poor metrics scores found in the reconstructions using χ_{33} as ground truth, and why using COSMOS as ground truth significantly improved these scores. As demonstrated in the experiments in this paper, a non-negligible mismatch was found between the

provided local phase data and using χ_{33} component of the STI tensor as ground truth. Anisotropic magnetization contributions and microstructural effects incorporated in the χ_{13} and χ_{23} elements proved to be the major source for this mismatch. A more adequate ground truth may be obtained by projecting these anisotropic contributions into the apparent scalar susceptibility. However, because of the necessity to perform a dipole inversion, this estimation is not trivial and requires further investigation. COSMOS reconstructions can capture some of these effects and may be better suited as ground-truth targets, although differences might be relevant in some structures (white matter and the vascular system). A secondary contribution to the mismatch is the presence of minor background field remnants.

Considerations and Outlook for future QSM challenges

The majority of biomedical imaging challenges (such as from MICCAI) conducted in the past years aimed segmentation and classification[47]. However, owing to the underlying physical problem without a ground truth, comparing the performance of QSM reconstruction algorithms is more complex than what can be reflected by a single numeric error measure, and the question on how to measure the quality of the reconstructions still remains open. Therefore, for future challenges it could be advisable to consider the overall performance instead of relying on a single metric. Multiple metric categories is also an option. To avoid controversy, categories should be announced beforehand, and the scoring source code made available with the release of the ground truth dataset[47].

In the case of working with in vivo data, thorough background removal steps should be carried out to provide clean local phases. This may be performed by including an additional background field removal step, such as PDF or V-SHARP. Major possible sources of errors could be also masked out by using smaller regions of interest. Given all the challenges involved in using multiple orientation reconstructions as ground-truth, and the unfeasibility of generating a single orientation ground truth, in vivo data remains a challenging ground for QSM reconstruction challenges.

Instead, analytic simulations of the acquired signal should be encouraged. Whole head susceptibility distributions may be derived from in vivo data to provide ground-truth with a realistic brain geometry, from which scalar (isotropic) magnetization fields may be calculated. Piece-wise analytic phantoms should be avoided to prevent unfair advantages to variational approaches (e.g. TV or TGV regularizations). In addition, analytic simulations may be even further extended in the future to incorporate the background field removal step and unwrapping and could provide a complete GRE signal simulation platform on which all different steps of the QSM reconstruction pipeline can be tested

and evaluated[48].

Another consideration for a future challenge design is to include an evaluation of a priori optimization of the results. Parameter fine-tuning without knowing the ground-truth is a non-trivial task. Having the ground-truth hidden from the participants should allow the community to evaluate how good the analytic predictors of optimal parameters are, and how reliable or reproducible the results may be.

In conclusion, the presence of unaccounted for anisotropic susceptibility contributions, noise, and the lack of reliable ground-truths using in vivo acquisitions, strongly suggests synthetic data and forward simulations for future QSM challenge designs.

Acknowledgments

We thank FONDECYT 1191710, CONICYT Programa PIA-Anillo ACT192064 and Millenium Science Initiative of the Ministry of Economy, Development and Tourism, grant Nucleus for Cardiovascular Magnetic Resonance for their funding support. CL acknowledges funding from the Austrian Science Fund (FWF grants KLI523 and P30134).

References

1. Mittal S, Wu Z, Neelavalli J, Haacke EM. Susceptibility-weighted imaging: Technical aspects and clinical applications, part 2. *Am J Neuroradiol*. 2009;30:232-252. doi:10.3174/ajnr.A1461

2. Haacke EM, Liu S, Buch S, Zheng W, Wu D, Ye Y. Quantitative susceptibility mapping: Current status and future directions. *Magn Reson Imaging*. 2015;33:1-25. doi:10.1016/j.mri.2014.09.004

3. Salomir R, De Senneville BD, Moonen CTW. A fast calculation method for magnetic field inhomogeneity due to an arbitrary distribution of bulk susceptibility. *Concepts Magn Reson*. 2003;19B:26-34. doi:10.1002/cmr.b.10083

4. Marques JPP, Bowtell R. Application of a fourier-based method for rapid calculation of field inhomogeneity due to spatial variation of magnetic susceptibility. *Concepts Magn Reson Part B Magn Reson Eng*. 2005;25:65-78. doi:10.1002/cmr.b.20034

5. Shmueli K, de Zwart J a, van Gelderen P, Li T-Q, Dodd SJ, Duyn JH. Magnetic susceptibility mapping of brain tissue in vivo using MRI phase data. *Magn Reson Med*. 2009;62:1510-1522. doi:10.1002/mrm.22135

6. de Rochefort L, Liu T, Kressler B, Liu J, Spincemaille P, Lebon V, Wu J, Wang Y. Quantitative susceptibility map reconstruction from MR phase data using bayesian regularization: validation and application to brain imaging. *Magn Reson Med*. 2010;63:194-206. doi:10.1002/mrm.22187

7. Wharton S, Schäfer A, Bowtell R. Susceptibility mapping in the human brain using threshold-based k-space division. *Magn Reson Med*. 2010;63:1292-1304. doi:10.1002/mrm.22334

8. Liu T, Wisnieff C, Lou M, Chen W, Spincemaille P, Wang Y. Nonlinear formulation of the magnetic field to source relationship for robust quantitative susceptibility mapping. *Magn Reson Med*. 2013;69:467-476. doi:10.1002/mrm.24272

9. Wang S, Liu T, Chen W, Spincemaille P. Noise effects in various quantitative susceptibility mapping methods. *IEEE Trans Biomed Eng.* 2013;60:3441-3448.

10. Milovic C, Bilgic B, Zhao B, Acosta-Cabronero J, Tejos C. Fast nonlinear susceptibility inversion with variational regularization. *Magn Reson Med*. 2018;80:814-821. doi:10.1002/mrm.27073

11. Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn Reson Med 2007; 58:1182-1195.

12. Kee Y, Liu Z, Zhou L, Dimov A, Cho J, de Rochefort L, Seo JK, Wang Y. Quantitative Susceptibility Mapping (QSM) Algorithms: Mathematical Rationale and Computational Implementations. *IEEE Trans Biomed Eng*. 2017;64:2531-2545. doi:10.1109/TBME.2017.2749298

13. Fortier V, Levesque IR. Phase processing for quantitative susceptibility mapping of regions with

large susceptibility and lack of signal. *Magn Reson Med*. 2018;79:3103-3113. doi:10.1002/mrm.26989

14. Schweser F, Robinson SD, de Rochefort L, Li W, Bredies K. An illustrated comparison of processing methods for phase MRI and QSM: removal of background field contributions from sources outside the region of interest. *NMR Biomed*. 2017;30:e3604. doi:10.1002/nbm.3604

15. Robinson SD, Bredies K, Khabipova D, Dymerska B, Marques JP, Schweser F. An illustrated comparison of processing methods for MR phase imaging and QSM: combining array coil signals and phase unwrapping. *NMR Biomed*. 2017;30:e3601. doi:10.1002/nbm.3601

16. Eckstein K, Dymerska B, Bachrata B, Bogner W, Poljanc K, Trattnig S, Robinson SD. Computationally Efficient Combination of Multi-channel Phase Data From Multi-echo Acquisitions (ASPIRE). *Magn Reson Med*. 2018;79:2996-3006. doi:10.1002/mrm.26963

17. Chatnuntawech I, McDaniel P, Cauley SF, Gagoski BA, Langkammer C, Martin A, Grant PE, Wald LL, Setsompop K, Adalsteinsson E, Bilgic B. Single-step quantitative susceptibility mapping with variational penalties. *NMR Biomed*. 2017;30:e3570. doi:10.1002/nbm.3570

18. Langkammer C, Bredies K, Poser BA., Barth M, Reishofer G, Fan AP, Bilgic B, Fazekas F, Mainero C, Ropele S. Fast quantitative susceptibility mapping using 3D EPI and total generalized variation. *Neuroimage*. 2015;111:622-630. doi:10.1016/j.neuroimage.2015.02.041

19. Milovic C, Bilgic B, Zhao B, Langkammer C, Tejos C, Acosta-Cabronero J. Weak-harmonic regularization for quantitative susceptibility mapping. *Magn Reson Med*. 2019;81:1399-1411. doi:10.1002/mrm.27483

20. Liu Z, Kee Y, Zhou D, Wang Y, Spincemaille P. Preconditioned total field inversion (TFI) method for quantitative susceptibility mapping. *Magn Reson Med*. 2017;78:303-315. doi:10.1002/mrm.26331

21. Liu T, Zhou D, Spincemaille P, Yi W. Differential approach to quantitative susceptibility mapping without background field removal. In: *Proceedings of the 22nd Annual Meeting of ISMRM, Milan, Italy*. ; 2014:597.

22. Wei H, Cao S, Zhang Y, Guan X, Yan F, Yeom KW, Liu C. Learning-based single-step quantitative susceptibility mapping reconstruction without brain extraction. Neuroimage. 2019;202:116064. doi: 10.1016/j.neuroimage.2019.116064

23. Sun H, Ma Y, MacDonald ME, Pike GB. Whole head quantitative susceptibility mapping using a least-norm direct dipole inversion method. Neuroimage. 2018;179:166-175. doi: 10.1016/j.neuroimage.2018.06.036.

24. Langkammer C, Krebs N, Goessler W, Scheurer E, Ebner F, Yen K, Fazekas F, Ropele S. Quantitative MR imaging of brain iron: a postmortem validation study. *Radiology*. 2010;257:455-462. doi:10.1148/radiol.10100495

25. Li W, Langkammer C, Chou Y-H, Petrovic K, Schmidt R, Song AW, Madden DJ, Ropele S, Liu

C. Association between increased magnetic susceptibility of deep gray matter nuclei and decreased motor function in healthy adults. *Neuroimage*. 2015;105:45-52. doi:10.1016/j.neuroimage.2014.10.009

26. Acosta-Cabronero J, Cardenas-Blanco A, Betts MJ, Butryn M, Valdes-Herrera JP, Galazky I, Nestor PJ. The whole-brain pattern of magnetic susceptibility perturbations in Parkinson's disease. *Brain*. 2017;140:118-131. doi:10.1093/brain/aww278

27. Acosta-Cabronero J, Williams GB, Cardenas-Blanco A, Arnold RJ, Lupson V, Nestor PJ. In vivo quantitative susceptibility mapping (QSM) in Alzheimer's disease. *PLoS One*. 2013;8:e81093. doi:10.1371/journal.pone.0081093

28. Reichenbach JR, Schweser F, Serres B, Deistung A. Quantitative Susceptibility Mapping: Concepts and Applications. *Clin Neuroradiol*. 2015;25:225-230. doi:10.1007/s00062-015-0432-9

29. Deistung A, Schweser F, Reichenbach JR. Overview of quantitative susceptibility mapping. *NMR Biomed*. 2017;30:e3569. doi:10.1002/nbm.3569

30. Deng W, Boada F, Poser BA, Schirda C, Stenger VA. Iterative projection onto convex sets for quantitative susceptibility mapping. Magn Reson Med. 2015 Feb;73:697-703. doi: 10.1002/mrm.25155

31. Langkammer C, Schweser F, Shmueli K, Kames C, Li X, Guo L, Milovic C, Kim J, Wei H, Bredies K, Buch S, Guo Y, Liu Z, Meineke J, Rauscher A, Marques JP, Bilgic B. Quantitative susceptibility mapping: Report from the 2016 reconstruction challenge. *Magn Reson Med*. 2018;79:1661-1673. doi:10.1002/mrm.26830

32. Liu C. Susceptibility tensor imaging. Magn Reson Med. 2010;63:1471-1477. doi:10.1002/mrm.22482

33. Liu T, Spincemaille P, De Rochefort L, Kressler B, Wang Y. Calculation of susceptibility through multiple orientation sampling (COSMOS): A method for conditioning the inverse problem from measured magnetic field map to susceptibility source image in MRI. *Magn Reson Med*. 2009;61:196-204. doi:10.1002/mrm.21828

34. Bilgic B, Gagoski BA, Cauley SF, Fan AP, Polimeni JR, Grant PE, Wald LL, Setsompop K. Wave-CAIPI for highly accelerated 3D imaging. *Magn Reson Med*. 2015;73:2152-2162. doi:10.1002/mrm.25347

35. Brant-Zawadzki M, Gillan GD, Nitz WR. MPRAGE: a three-dimensional, T1-weighted, gradient-echo sequence--initial experience in the brain. *Radiology*. 1992;182:769-775. doi:10.1148/radiology.182.3.1535892

36. Schofield M a, Zhu Y. Fast phase unwrapping algorithm for interferometric applications. *Opt Lett*. 2003;28:1194-1196. http://www.ncbi.nlm.nih.gov/pubmed/12885018.

37. Zhou D, Liu T, Spincemaille P, Wang Y. Background field removal by solving the Laplacian

boundary value problem. *NMR Biomed*. 2014;27:312-319. doi:10.1002/nbm.3064

38. Schweser F, Atterbury M, Deistung A, Lehr BW, Sommer K. Harmonic phase subtraction methods are prone to B1 background components. In: *Proc Intl Soc Mag Reson Med 19*. Montreal; 2011.

39. Liu T, Khalidov I, de Rochefort L, Spincemaille P, Liu J, Tsiouris AJ, Wang Y. A novel background field removal method for MRI using projection onto dipole fields (PDF). NMR Biomed. 2011;24:1129-1136. doi:10.1002/nbm.1670

40. Wharton S, Bowtell R. Effects of white matter microstructure on phase and susceptibility maps. *Magn Reson Med*. 2015;73:1258–1269. doi:10.1002/mrm.25189

41. Cronin MJ, Bowtell R. Quantifying MRI frequency shifts due to structures with anisotropic magnetic susceptibility using pyrolytic graphite sheet. Sci. Rep. 2018;8:6259

42. Yablonskiy D, Sukstanskii A. Lorentzian effects in magnetic susceptibility mapping of anisotropic biological tissues. Journal of Magnetic Resonance. 2018;292:129-136.

43. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Trans Image Process. 2004;13:600-612. doi:10.1109/TIP.2003.819861

44. Moddemeijer, R. On Estimation of Entropy and Mutual Information of Continuous Distributions. Signal Processing. 1989;16:233-246

45. Li X, Vikram DS, Lim I, Jones CK, Farrell J, van Zijl P. Mapping magnetic susceptibility anisotropies of white matter in vivo in the human brain at 7T. NeuroImage. 2012,62:314-330.

46. Zhou Wang, Bovik AC. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. IEEE Signal Process Mag. 2009;26:98-117. doi:10.1109/MSP.2008.930649

47. Maier-Hein L, Eisenmann M, Reinke A, Onogur S, Stankovic M, Scholz P, Arbel T, Bogunovic H, Bradley AP, Carass A, Feldmann C, Frangi AF, Full PM, van Ginneken B, Hanbury A, Honauer K, Kozubek M, Landman BA, März K, Maier O, Maier-Hein K, Menze BH, Müller H, Neher PF, Niessen W, Rajpoot N, Sharp GC, Sirinukunwattana K, Speidel S, Stock C, Stoyanov D, Taha AA, van der Sommen F, Wang CW, Weber MA, Zheng G, Jannin P, Kopp-Schneider A. Why rankings of biomedical image analysis competitions should be interpreted with care. Nat Commun. 2018;9:5217. doi: 10.1038/s41467-018-07619-7

48. Marques JP, Bilgic B, Meineke J, Milovic C, Chan K-S, van der Zwaag W, Hedouin R, Langkammer C, and Schweser F. Towards QSM Challenge 2.0: Creation and Evaluation of a Realistic Magnetic Susceptibility Phantom. Proc. 27th Annual Meeting of the ISMRM, Montreal, Canada, 2019.

Figures

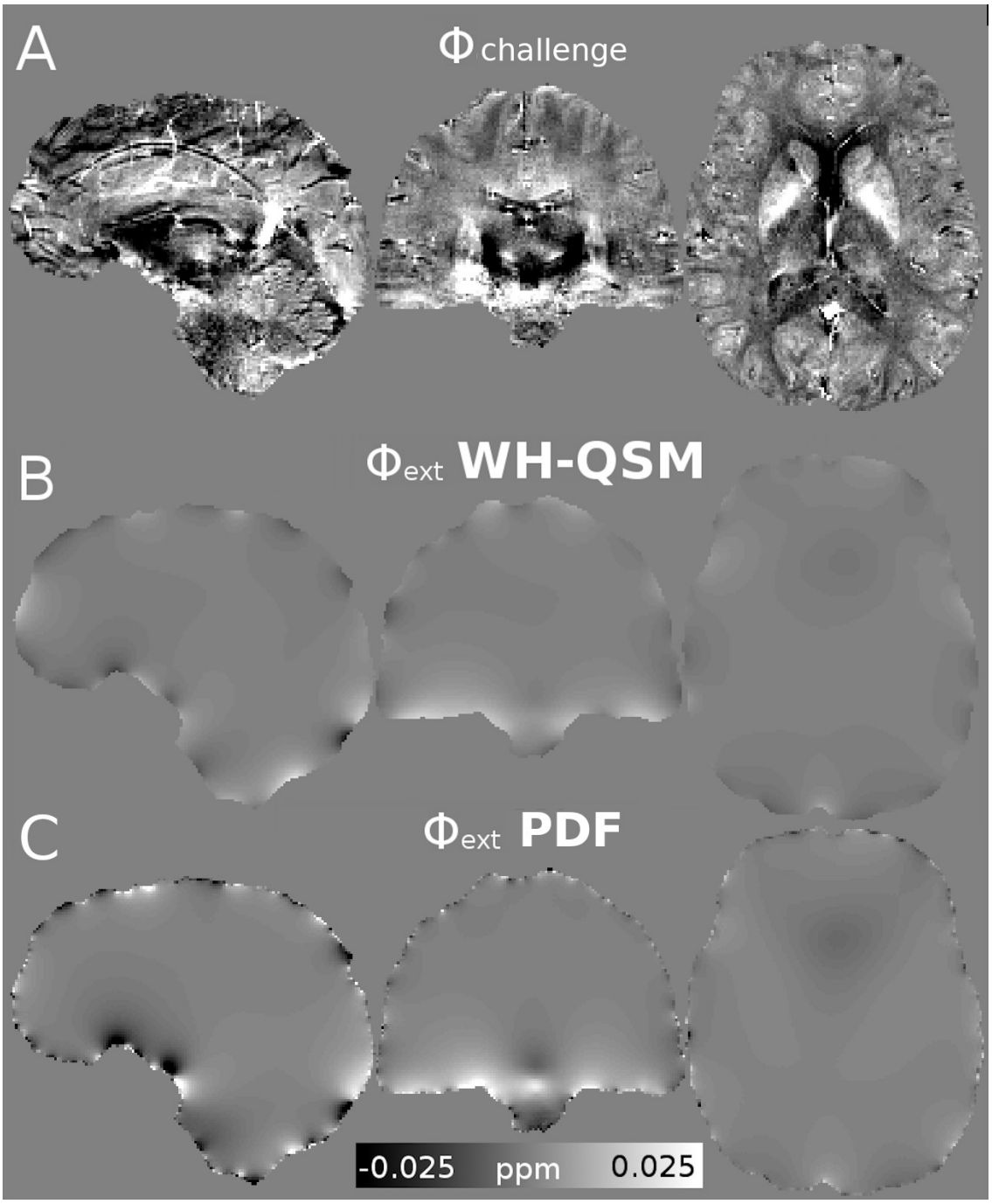

Figure 1. Single orientation acquisition (A) and background remnant models using (B) weak-harmonics (WH-QSM, with β=5, μ_h=0.1, the weak harmonic regularization weights) and (C) the Projection onto Dipole Fields (PDF) methods.

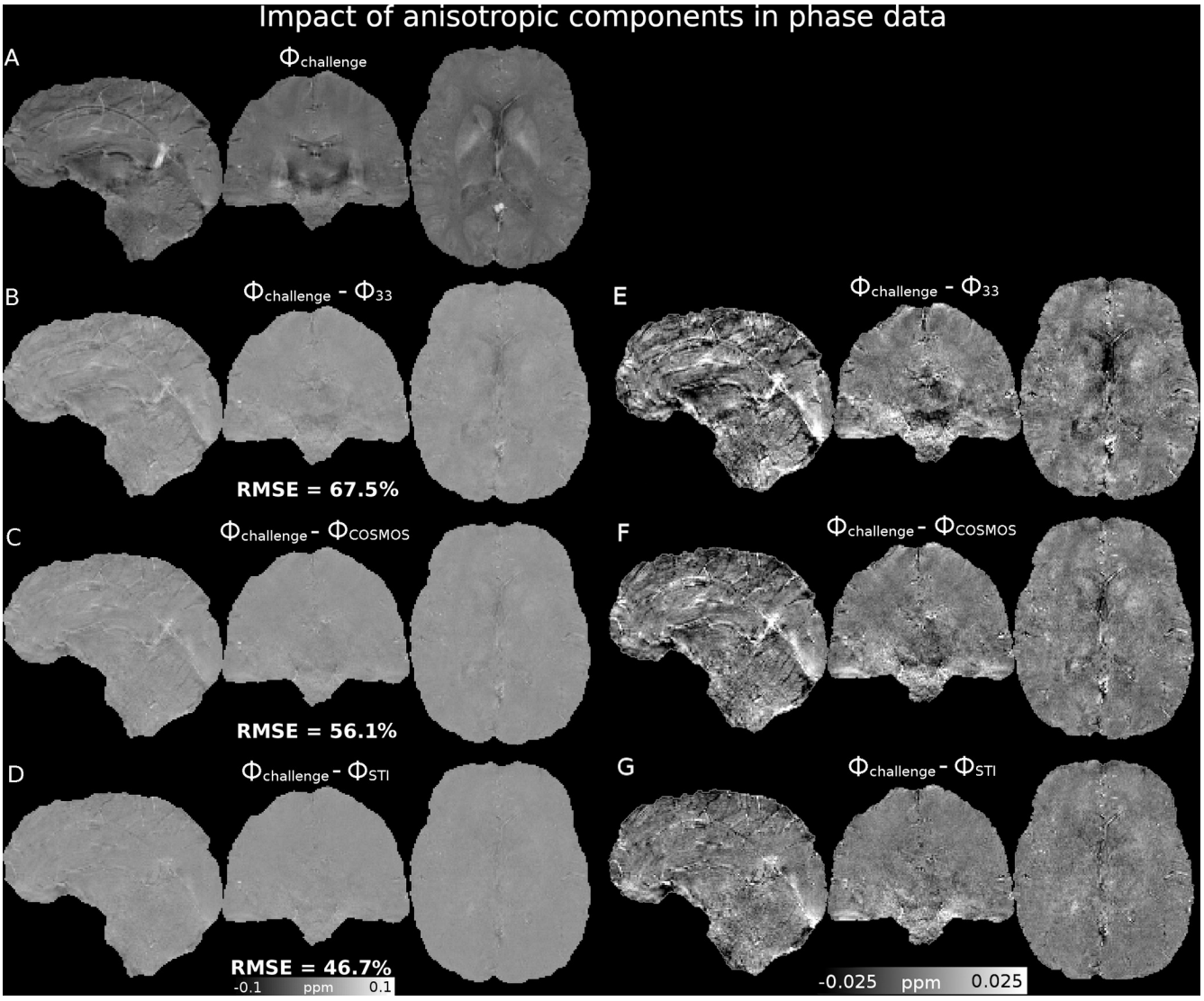

Figure 2. Differences between forward simulated phases and the provided single orientation acquisition (A) for: (B and E) the χ_{33} component, (C and F) COSMOS, and (D and G) the STI equation, including the χ_{13} and χ_{23} anisotropic components. The right column shows the same difference maps as the left column, but with a higher contrast to reveal subtle differences.

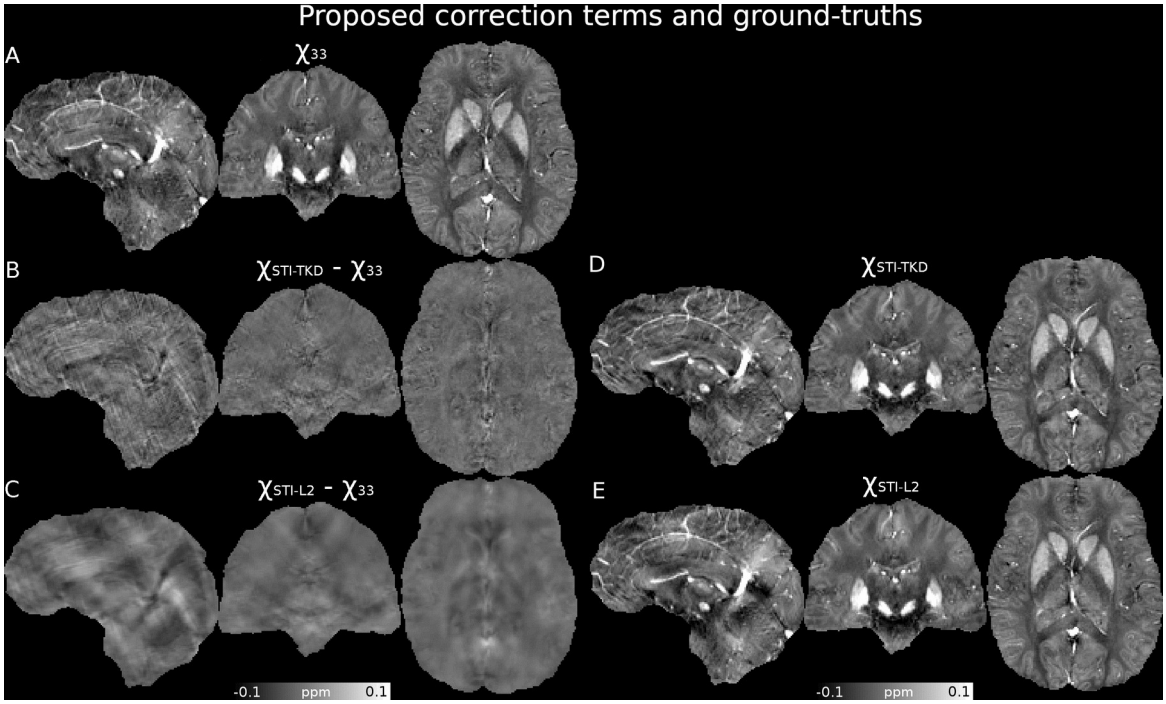

Figure 3. The χ_{33} (isotropic) component and correction factors, χ_{corr} (B and C, derived from χ_{13} and χ_{23} anisotropic components) used to estimate $\chi_{STI-TKD}$ and χ_{STI-L2} respectively (D and E).

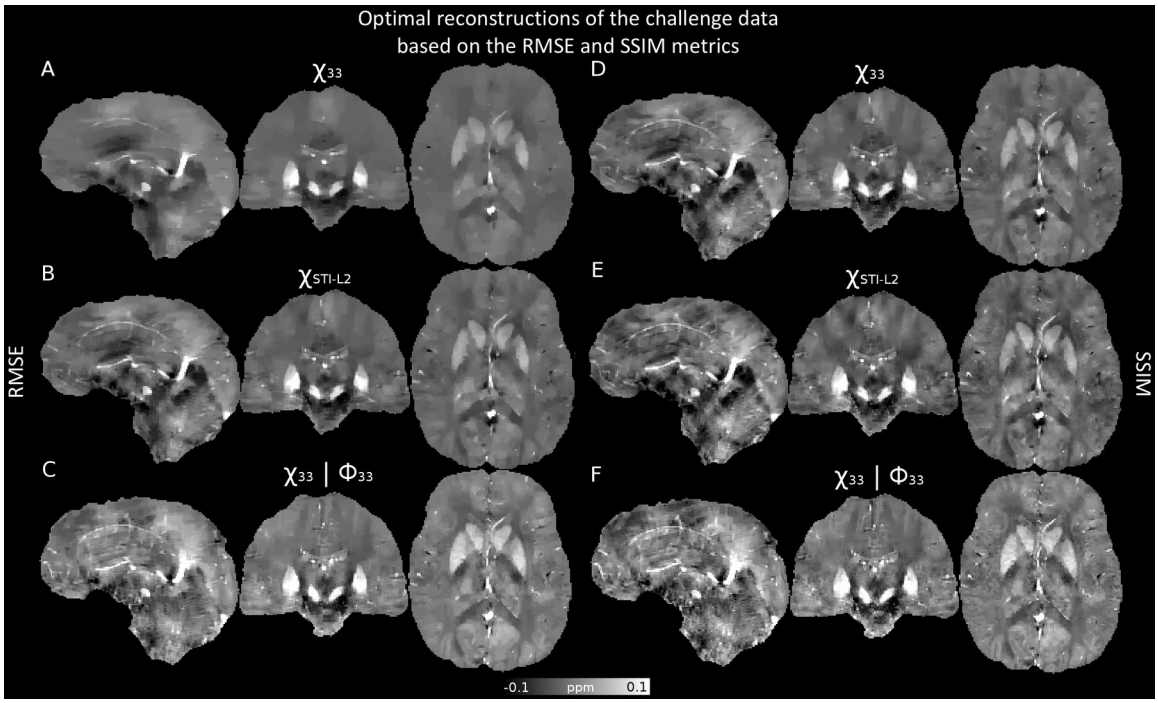

Figure 4. Selected optimal reconstructions of the challenge data, based on the RMSE (left panel) and SSIM (right panel) metrics, using χ_{33} (A,D) and χ_{STI-L2} (B, E) as ground truth. (C ,F) show optimal reconstructions of the amended phase (to remove anisotropic components) and using χ_{33} as ground truth ($\chi_{33}|\Phi_{33}$).

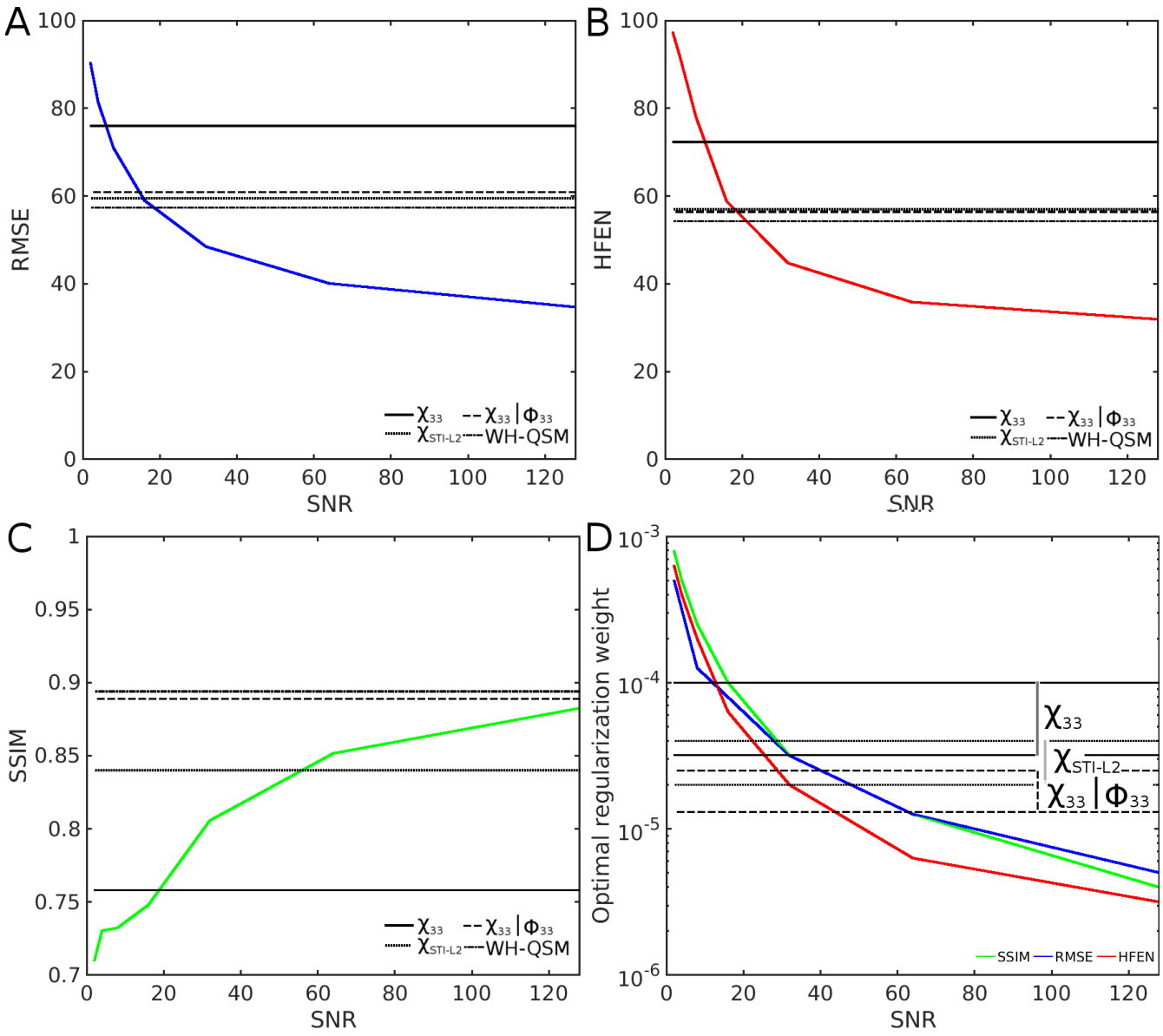

Figure 5. Metric scores of the reconstruction using COSMOS-based forward simulations. Optimal (A) RMSE, (B) HFEN and (C) SSIM scores are presented as function of SNR. Metric scores for in vivo reconstructions using χ_{33} and $\chi_{STI\text{-}L2}$ as ground truth, along with reconstructions of the amended phase with TV-FANSI ($\chi_{33}|\Phi_{33}$) and WH-QSM, are also included. (D) presents the optimal regularization weight for each metric and SNR setting. Ranges for the optimal reconstructions using χ_{33}, $\chi_{STI\text{-}L2}$ and $\chi_{33}|\Phi_{33}$ also included.

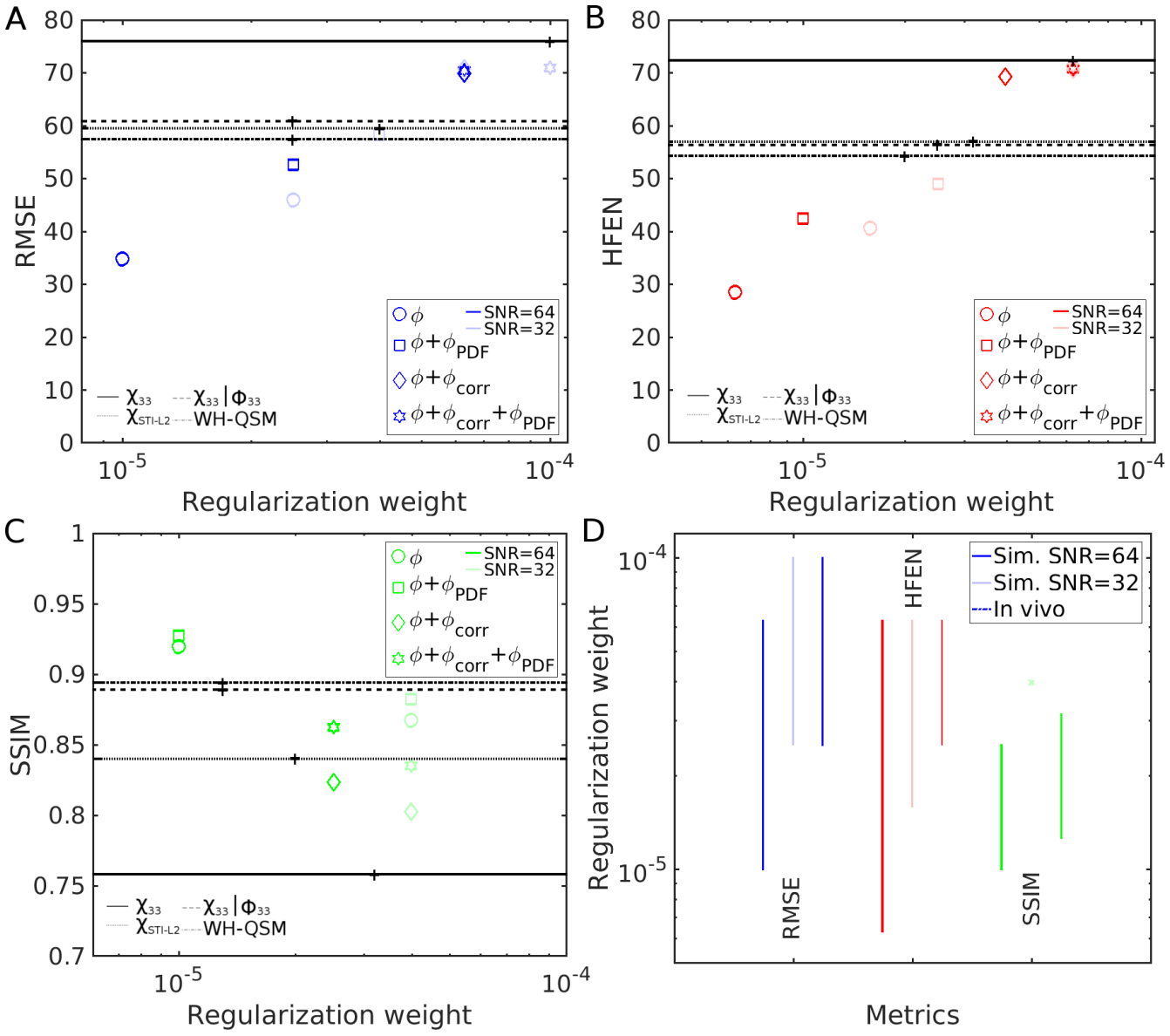

Figure 6. (A) RMSE, (B) HFEN and (C) SSIM optimal scores for reconstructions at SNR=64 and SNR=32 and their regularization weight. Simulations included the phase forward calculated using COSMOS Φ with the addition of the background remnants obtained using PDF (Φ_{PDF}) and the anisotropic phase contributions (Φ_{corr}) calculated from the χ_{13} and χ_{23} components of the STI tensor. (D) Shows the range span of optimal regularization weight values by the different simulations and in vivo reconstructions, for a given metric.

Tables

Table 1: Quality metric scores of QSM reconstructions with additional background filtering

	RMSE	COSMOS HFEN
TV-FANSI	64.49	59.85
WH-QSM	60.05	58.05

Table 2: Optimal Regularization weight found for each QSM reconstruction

	RMSE	COSMOS HFEN
TV-FANSI	6.3E-05	4.0E-05
WH-QSM	4.0E-05	4.0E-05

Table 3: Quality metric scores and optimal QSM regularization parameters of ground-truth solutions obtained with TV-FANSI.

	RMSE	α
χ_{33}	76.0	1.0E-04
COSMOS	62.6	5.0E-05
$\chi_{\text{STI-TKD}}$	65.2	5.0E-05

Supporting Information captions

Supporting Information Figure S1. Selected optimal reconstructions of the challenge data, based on RMSE (left column) and SSIM (right column), using χ_{33} as ground truth (A). Panels show reconstructions of the provided local phase using (B, E) TV-FANSI, (C, F) WH-QSM (with joint background removal process), and (D, G) TV-FANSI of the PDF filtered local phase.

Supporting Information Figure S2. Difference between the forward simulated phases using χ_{33} and COSMOS. Note the windowing.

Supporting Information Figure S3. (A) and (B) reveal the background extracted from the acquired phase and the amended phase, respectively, using WH-QSM (β=500, μ_h=10, the weak harmonic regularization weights).

Supporting Information Figure S4. Difference maps between optimal results using RMSE and SSIM, for the same ground-truth.

Supporting Information Figure S5. Selected optimal reconstructions of the amended phase (without anisotropic contributions) based on the RMSE and SSIM metrics, using χ_{33} as ground truth. Panels show reconstructions using (A, D) TV-FANSI, (B, E) WH-QSM (with joint background removal process), and (C, F) TV-FANSI of the background filtered local phase, using PDF.

Supporting Information Figure S6. Effect of noise level on forward simulations. (A) Single orientation phase data and forward simulations of Φ_{STI} at SNR levels: (B) 20, (C) 32, (D) 40, (E) 60, (F) 100 and (G) noiseless.

Supporting Information Figure S7. Metric scores of the reconstruction using COSMOS-based forward simulations. Optimal (A) CC, (B) MI, (C) GXE, (D) MAD and (E) MADGX scores are presented as function of SNR. Metric scores for in vivo reconstructions using χ_{33} and $\chi_{STI\text{-}L2}$ as ground truth, along with reconstructions of the amended phase with TV-FANSI ($\chi_{33}|\Phi_{33}$) and WH-QSM, are also included. (F) presents the optimal regularization weight for all metrics and SNR setting. Ranges for the optimal reconstructions using χ_{33}, $\chi_{STI\text{-}L2}$ and $\chi_{33}|\Phi_{33}$ also included.

Supporting Information Figure S8. (A) RMSE, (B) HFEN, (C) SSIM, (D) CC, (E) MI, (F) GXE, (G) MAD and (H) MADGX scores for reconstructions at SNR=64 and SNR=32 as function of the reconstruction weight. Simulations included the phase forward simulatedcalculated using COSMOS Φ with the addition of the background remnants obtained using PDF (Φ_{PDF}) and the anisotropic phase

contributions (Φ_{corr}) calculated from the χ_{13} and χ_{23} components of the STI tensor. (I) Shows the range span of optimal regularization weight values by the different simulations and in vivo reconstructions, for all given metrics.

Supporting Information Figure S9. Comparison of all proposed quality metrics as function of the regularization weigh for in vivo reconstructions. (A) RMSE, (B) HFEN, (C) SSIM, (D)CC, (E) MI, (F) GXE, (G) MAD and (H) MADGX. Reconstructions times are also provided (I).

Supporting Information Table S1. Quality metric scores and optimal regularization weights of QSM reconstructions of the amended phase Φ_{33} with additional background filtering.