Quantitative Susceptibility Mapping: Report from the 2016 Reconstruction Challenge

Christian Langkammer[1†], Ferdinand Schweser[2†], Karin Shmueli[3†], Christian Kames[4], Xu Li[5], Li Guo[6], Carlos Milovic[7], Jinsuh Kim[8], Hongjiang Wei[9], Kristian Bredies[10], Sagar Buch[11], Yihao Guo[6], Zhe Liu[12], Jakob Meineke[13], Alexander Rauscher[4], José P. Marques[14], Berkin Bilgic[15†]

ABSTRACT

Purpose: The aim of the 2016 quantitative susceptibility mapping (QSM) reconstruction challenge was to test the ability of various QSM algorithms to recover the underlying susceptibility from phase data faithfully.

Methods: Gradient-echo images of a healthy volunteer acquired at 3 Tesla in a single orientation with 1.06 mm isotropic resolution. A reference susceptibility map was provided, which was computed using the susceptibility tensor imaging algorithm on data acquired at 12 head orientations.

Susceptibility maps calculated from the single orientation data were compared against the reference susceptibility map. Deviations were quantified using the following metrics: root mean squared error (RMSE), structure similarity index (SSIM), high-frequency error norm (HFEN), and the error in selected white and grey matter regions.

Results: Twenty-seven submissions were evaluated. Most of the best scoring approaches estimated the spatial frequency content in the ill-conditioned domain of the dipole kernel using compressed sensing strategies. The top ten maps in each category had similar error metrics but substantially different visual appearance.

Conclusion: Because QSM algorithms were optimized to minimize error metrics, the resulting susceptibility maps suffered from over-smoothing and conspicuity loss in fine features such as vessels. As such, the challenge highlighted the need for better numerical image quality criteria.

(1) Department of Neurology, Medical University of Graz, Austria (2) Buffalo Neuroimaging Analysis Center, Department of Neurology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, The State University of New York, Buffalo, NY, USA; Clinical and Translational Science Institute, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, The State University of New York, Buffalo, NY, USA (3) Department of Medical Physics and Biomedical Engineering, University College London, UK (4) UBC MRI Research Centre, Department of Physics and Astronomy, University of British Columbia, Vancouver, Canada (5) F.M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute, Baltimore, MD, USA; Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, MD, USA (6) Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University, Guangzhou, China (7) Department of Electrical Engineering, Pontificia Universidad Catolica de Chile, Santiago, Chile; Biomedical Imaging Center, Pontificia Universidad Catolica de Chile, Santiago, Chile (8) Department of Radiology, University of Illinois at Chicago, IL, USA (9) Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA (10) Institute of Mathematics and Scientific Computing, University of Graz, Austria (11) The MRI Institute for Biomedical Research, Waterloo, Ontario, Canada (12) Department of Biomedical Engineering, Cornell University, Ithaca, New York, USA (13) Philips Research Europe, Hamburg, Germany (14) Donders Centre for Cognitive Neuroimaging, Radboud University, The Netherlands (15) Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Harvard Medical School, MGH, Boston, MA, USA

(†) These authors contributed equally to this work

Corresponding author: Ferdinand Schweser, Email: schweser@buffalo.edu

Published in MRM: http://onlinelibrary.wiley.com/doi/10.1002/mrm.26830/full (DOI: 10.1002/mrm.26830)

INTRODUCTION

Quantitative susceptibility mapping (QSM) aims to determine a basic physical property (i.e. tissue magnetic susceptibility) *in vivo* that is highly sensitive to tissue molecular composition and disease-induced tissue damage (1–5). QSM solves an inverse field-to-source problem, calculating the underlying magnetic susceptibility distribution from gradient-echo (GRE) phase images. Early concepts for QSM have been introduced two decades ago (6–12) and more refined methods have been introduced recently to allow the calculation of susceptibility with reduced reconstruction artefacts from a single orientation in the clinical setting (13,14). The clinical value of QSM is currently being explored and holds great promise for vascular, inflammatory and neurodegenerative diseases of the brain (15–19). As such, the QSM field is rapidly developing, QSM is increasingly being used in clinical studies of neurological disorders, and applications outside the brain are being explored (20–24). The quantitative nature of the technique promises to provide biomarkers that allow the clinical monitoring of disease progression and treatment effects. However, especially considering the quantitative nature of QSM, clinical translation will require a thorough understanding of the reproducibility and accuracy of susceptibility measurements with QSM. Also, for a comparative assessment of QSM-based literature reports, it is important to understand how comparable susceptibility values are if they were reconstructed with different QSM algorithms.

A variety of algorithms has been developed for the numerical solution of the field-to-source inverse problem at the heart of QSM. However, although QSM is supposed to yield a physical tissue property, susceptibility maps calculated with different algorithms from the same dataset can show substantial differences, as illustrated in a recent review by Wang and Liu (1). To systematically compare and quantitatively assess the many available algorithms, we implemented the first QSM reconstruction challenge in the context of the 4th International Workshop on MRI Phase Contrast and Quantitative Susceptibility Mapping, held from September 26th to 28th, 2016 at the Medical University of Graz, Austria (www.qsm2016.com). The primary goal of the challenge was to test the ability of various QSM algorithms to recover the underlying susceptibility distribution from a healthy volunteer's phase data faithfully. The secondary goal was to provide a common reference dataset that would help benchmark not only existing QSM algorithms but also methods that would be developed in the future.

The challenge was announced at the Electro-Magnetic Tissue Properties (EMTP) (formerly SWI) study group meeting at the 2016 annual meeting of the International Society for Magnetic Resonance in Medicine (ISMRM) in Singapore on May 12th, 2016. Data and instructions could be downloaded from the workshop website (qsm.neuroimaging.at) starting from May 12th, 2016, and the deadline for a submission of reconstructed susceptibility maps was September 15th, 2016. The results of the evaluation of submitted maps were presented and discussed at the QSM workshop in Graz on September 27th, 2016. Additionally, the present report includes the input from the discussions in Graz and at the ISMRM EMTP study group meeting in Honolulu on April 26th, 2017.

METHODS

General considerations on input and reference data

In the literature, evaluation of susceptibility mapping algorithms is frequently performed using numerical phantoms (25,26) or acquired phantom data (27–29). Most physical phantoms used have consisted of compartments filled with solutions or gels of different magnetic susceptibilities, i.e. regions of piece-wise constant magnetic susceptibility. Such geometries allow a near-perfect recovery of the underlying susceptibility distribution using regularization of the inverse problem with total (generalized) variation (TV/TGV) or morphological priors because the piece-wise constant constraints and priors exactly match the actual susceptibility distribution. Using a physical phantom would, therefore, put these types of algorithms at a competitive advantage compared to other algorithm types. Moreover, a piece-wise constant susceptibility distribution is not a realistic model of magnetic susceptibility in the brain.

A limitation of numerical models is that contributions from sources other than isotropic bulk magnetic susceptibility such as chemical exchange effects (30), anisotropic susceptibility (31,32) and microstructure (33–36) are difficult to model because

the magnitude of these effects *in vivo* is not yet completely understood. Furthermore, physiological noise, flow, and partial volume effects are difficult to model realistically.

To address the shortcomings of physical phantoms and numerical models, in this challenge we decided to use a human susceptibility map measured *in vivo* as a reference. Attempting to take magnetic susceptibility anisotropy into account, we employed the susceptibility tensor imaging (STI) approach (37) to determine the reference map. STI reconstructs the susceptibility tensor distribution without any regularization or morphological priors. From the susceptibility tensor, it is possible to estimate the expected susceptibility distribution that would be measured with a single-angle susceptibility mapping technique. This effective susceptibility distribution was used as the reference susceptibility map in the challenge as described below.

We decided to provide the reference susceptibility map to the contestants to reduce the potentially negative impact of sub-optimal algorithm-specific parameter choices on the challenge outcome. The availability of the reference allowed the contestants to optimize algorithmic parameters properly and then submit the best scoring result they could achieve with their algorithm.

Selection of the reference

The candidates for gold standard susceptibility were either a COSMOS (27) susceptibility map, or χ_{33} from the STI solution (37). The benefits of these two maps as reference susceptibility distributions include **i)** they are calculated without numerical regularization and, therefore, no spatial smoothing or incorporated prior information, and **ii)** high signal-to-noise ratio (SNR) since both maps are computed from joint processing of images acquired at 12 orientations of the head with respect to B_0.

COSMOS models susceptibility as a scalar, isotropic property, ignoring its orientation dependence. A COSMOS susceptibility map reflects the effective magnetic susceptibility averaged over all 12 orientations of the head. Therefore, we concluded that COSMOS susceptibility maps would not provide an accurate reference for single-angle susceptibility mapping with the head in the normal position, particularly in regions with anisotropic magnetic susceptibility, such as white matter. To mitigate this orientation bias, we chose χ_{33} of the STI solution as the reference. Based on STI theory (37), the Fourier domain phase $\boldsymbol{\Theta(k)}$, when the main field lies along $\boldsymbol{H}$ in the subject frame, is given by

$$\boldsymbol{\Theta(k)} = \frac{\mathbf{1}}{\mathbf{3}}\boldsymbol{H^T \cdot X \cdot H} - \boldsymbol{H \cdot k}\frac{\boldsymbol{k^T \cdot X \cdot H}}{\boldsymbol{k^2}} \qquad \textbf{[1]}$$

where $\boldsymbol{k}$ is a vector of all Fourier domain coordinates and $\boldsymbol{X}$ is the susceptibility tensor in the subject frame and $()^{\boldsymbol{T}}$ denotes matrix transposition. When the acquisition is performed in the transverse plane relative to the subject coordinates, i.e. $\boldsymbol{H} = [0,0,1]^T$, the signal equation becomes

$$\boldsymbol{\Theta(k)} = \left(\frac{1}{3} - \frac{k_z^2}{k^2}\right)\chi_{33} - \frac{k_z}{k^2}\left(k_x\,\chi_{13} + k_y\,\chi_{23}\right) \qquad \textbf{[2]}$$

The relationship commonly used in single-orientation QSM assumes that the terms with the off-diagonal tensor elements, χ_{13} and χ_{23}, are negligible:

$$\boldsymbol{\Theta(k)} = \left(\frac{1}{3} - \frac{k_z^2}{k^2}\right)\chi_{33}, = \mathbf{D}\,\chi_{33}, \qquad \textbf{[3]}$$

where $\mathbf{D}$ is the dipole kernel in the Fourier domain. Equation [3] motivates the use of χ_{33} as the reference susceptibility that gives rise to the observed phase signal.

Data and source code

MRI data were acquired in a healthy female volunteer (age 30) at a 3T system (Tim Trio, Siemens Healthcare GmbH, Erlangen, Germany) with Institutional Review Board approval from Massachusetts General Hospital.

The imaging data provided to the contestants as inputs for the susceptibility mapping included the following datasets:

- 3D gradient-echo magnitude and wrapped phase images acquired with axial slab orientation (and the head in the normal supine position).
- A magnetization-prepared rapid gradient-echo (MPRAGE) image (38) matching the GRE volume because MPRAGE images are routinely acquired in clinical brain imaging studies and certain QSM algorithms use an MPRAGE image as an input, i.e. as prior information.
- A background-field corrected tissue phase image. We used the Laplacian Boundary Value (LBV) method (39) after transmit phase removal by fitting and subtracting a 4^{th}-order 3D-polynomial. LBV was used because it outperformed all other proposed background-field correction methods in a recent comparison study (40). This image was provided in an attempt to eliminate a potential variability in submitted susceptibility maps due to differences in background field removal techniques. However, as single-step QSM methods are designed to solve background-field removal and inversion problems simultaneously, those algorithms could use the unprocessed wrapped phase GRE images.
- A brain mask obtained from the FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) Brain Extraction Tool (41) was also provided to reduce confounding effects resulting from the use of different masks.
- The reference susceptibility map χ_{33} which was calculated using STI (37). The GRE phase images from each head orientation were affine-registered to the axial slab orientation (reference position), masked and the background fields removed as described for the single orientation case above. This local field information was then fed into an iterative LSQR solver (42) to estimate all components of the symmetric susceptibility tensor and provide the tensor element χ_{33} as reference susceptibility map.

3D GRE with Wave-CAIPI acquisition (43) was used to acquire images of the head with 1.06 mm isotropic resolution in 12 different orientations with respect to B_0 (the head orientation table can be found in the downloadable dataset). Further sequence parameters were TE / TR = 25 / 35 ms, BW = 100 Hz/pixel and a 94-s acquisition time for each head orientation with 15-fold acceleration using a Siemens 32 channel head coil. Roemer/SENSE coil combination was employed (44,45), which used sensitivities estimated from reference acquisitions made with both, head and body coil reception. Wave-CAIPI is an accelerated acquisition/reconstruction technique that substantially reduces the scan time, which is especially useful for multi-orientation scans. Despite 15-fold acceleration, the average g-factor penalty due to parallel imaging reconstruction was only 9%. Thus, aliasing artifacts or noise amplification are not expected to impact the resulting susceptibilities (43).

MPRAGE acquisition employed the same resolution and matrix size as 3D-GRE, and sampled 4 echoes using TE1 = 2.05 ms, echo spacing = 1.84 ms, TR = 2510 ms, inversion time (TI) = 1200 ms, BW=651 Hz/pixel and flip angle = 7°. The acquisition took 5 min 39 s using 2-fold GRAPPA acceleration (46). The magnitude images at all 4 echo times were combined by computing the root-sum-of-squares (47), and the combined magnitude image was provided to the participants.

In addition to the imaging data, MATLAB (The MathWorks, Natick, MA) source code was provided for the numerical evaluation of the data set according to the error metrics described in detail below. This code allowed the contestants to focus on optimizing their algorithmic parameters without spending time writing scripts for the calculation of error metrics. The source code also included the widely utilized fast QSM reconstructions, thresholded k-space division (TKD) (28) and a closed-form L2-regularized algorithm (48) to provide contestants with a direct performance comparison to these algorithms.

The images and the Matlab code for the QSM reconstruction challenge will remain available at http://qsm.neuroimaging.at. In addition to the data provided at the time of the challenge and described above, the archive currently also contains the GRE data magnitude and phase data acquired in all 12 orientations. The images provided are shown in Figure 1.

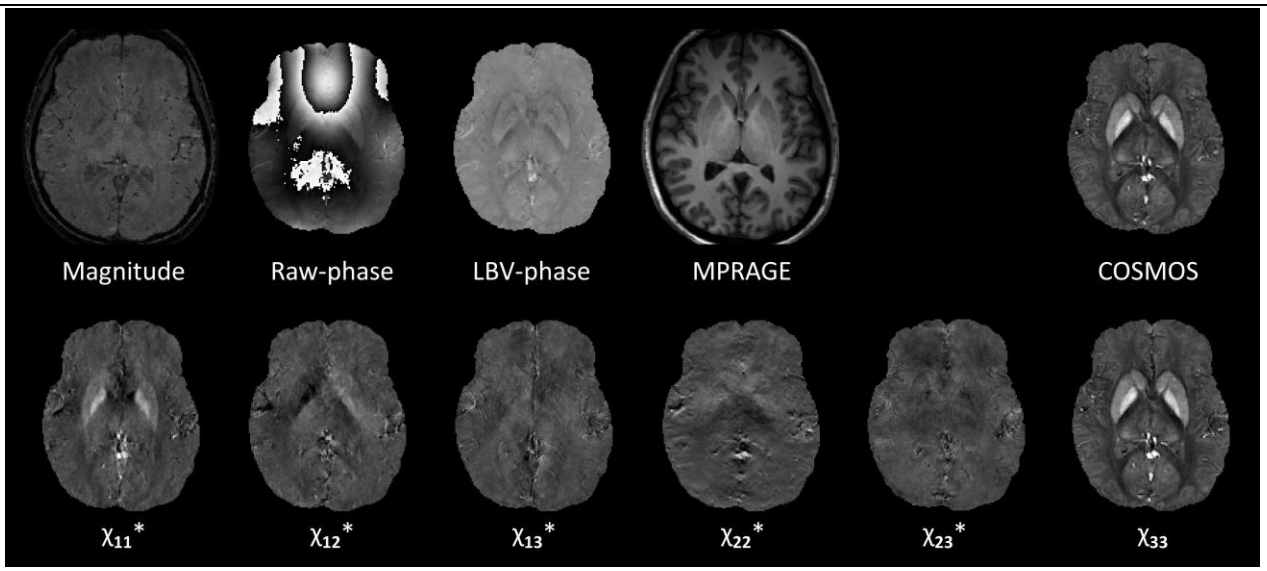


Figure 1: Image data provided to the contestants. The susceptibility maps are scaled from -0.1 to 0.25 ppm, the raw phase is scaled between $\pm\pi$ radians and the LBV-phase image is scaled from -0.05 to 0.05 radians. With the exception of χ_{33}, the reconstructed susceptibility tensor component images (marked here with asterisks) were not provided for the reconstruction challenge, but are now included in the downloadable data set at http://qsm.neuroimaging.at.

Numerical measures of QSM reconstruction quality

We employed quantitative error metrics to evaluate the difference between the reference susceptibility map and the submitted susceptibility maps. As well as the root mean squared error (RMSE), which is commonly used in the literature, we employed three additional error measures, which are often utilized in the fields of computer vision and image reconstruction:

- The high-frequency error norm (HFEN) (49), which aims to measure the fidelity at high spatial frequencies. HFEN applies a Laplacian of a Gaussian (LoG) filter on the reference and input volumes and reports the L2 norm of their difference, normalized by the norm of the LoG-filtered reference.
- The structural similarity index (SSIM) (50), which is a combined measure obtained from three complementary components (luminance similarity, contrast similarity, and structural similarity). SSIM aims to provide a metric that better reflects the "visual" similarity to the reference.
- The absolute value of the mean error in selected anatomical structures (ROI error). To this end, we manually defined ROIs in white matter (genu and splenium of corpus callosum, frontal white matter, occipital white matter, capsula interna) and grey matter nuclei (globus pallidus, putamen, caudate nucleus, red nucleus, substantia nigra, and dentate nucleus) on the reference susceptibility map, χ_{33}.

These error metrics were calculated for each submitted map. For RMSE, HFEN and ROI error, smaller values denote better performance, whereas SSIM is normalized between 0 and 1, with 1 being the best possible result. The implementation of these error metrics was provided to the contestants as Matlab source code together with the downloadable image data set.

Our current challenge format did not include a comparison of the reconstruction speed of the different algorithms as the QSM images were processed on the individual computers of the respective research groups, using different development environments (e.g. Matlab, Python, C++, CUDA).

All susceptibility values are reported as ppm in the following.

RESULTS

Brief description of the algorithms used by the contestants

Overall, 27 susceptibility maps from 13 groups were evaluated. The algorithms either used the provided pre-processed (background removed) phase or the raw, wrapped phase. Several algorithms used the GRE magnitude for stabilization of the dipole inversion, and one approach (PHILIPS DTV) also utilized the MPRAGE images.

The algorithms are briefly described in table 1 and images of a central transverse slice of all algorithms are shown in figure 2.

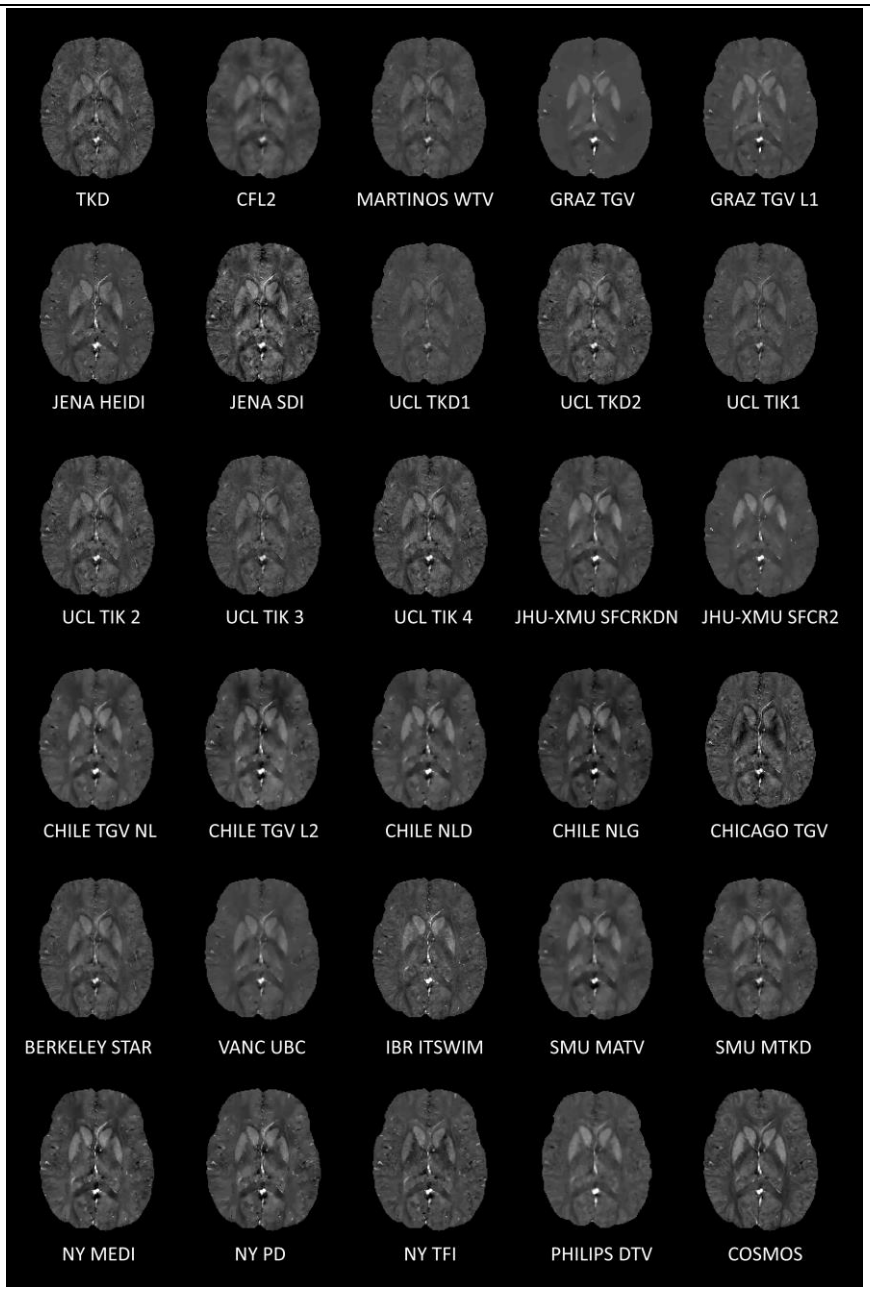


Figure 2: A single transverse slice from all QSM reconstructions submitted for the challenge. QSM images are scaled from -0.1 to 0.25 ppm. See Table 1 for a brief description of all the QSM reconstruction algorithms shown here.

Numerical results: Winners

Table 2 shows the results of the top ranked algorithms in each evaluation category. When all submissions were evaluated, RMSE ranged from 69.0 to 140.9 (median = 83.9), HFEN from 63.5 to 127.3 (median = 75.9), SSIM from 0.94 to 0.63 (median = 0.82) and ROI ERROR from 0.016 to 0.039 ppm (median = 0.020).The winning QSM reconstructions are also depicted in detail in figure 3.

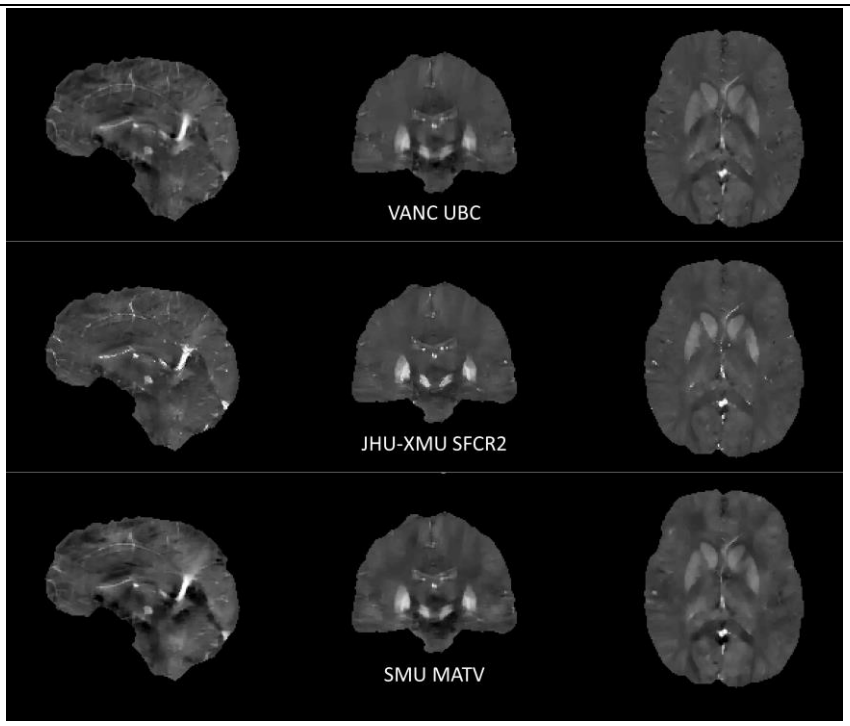


Figure 3: Sagittal, coronal and axial slices of QSM reconstructions of the winners in each category: RMSE (UBC), HFEN and SSIM respectively (SFCR2), and ROI error (MATV). QSM images are scaled from -0.1 to 0.25 ppm.

Winning Approach: RMSE

The winner in the RMSE category was the approach developed by Alexander Rauscher's team at the University of British Columbia, Canada. This algorithm used a weighted variant of a two-step dipole inversion algorithm (51). It adopts an incremental dipole inversion strategy (52–54), dividing the Fourier domain into well-conditioned and ill-conditioned regions. In the first step, the well-conditioned region is reconstructed by solving $\phi = \mathbf{F^{-1}DF}\,\chi_{well}$ using an LSMR solver (55), where ϕ is the local field is spatial space, $\mathbf{F}$ is the forward Fourier transform and $\mathbf{F^{-1}}$ is the inverse Fourier transform. In order to avoid streaking artifacts the implicit regularization properties of Krylov subspace methods (56) are used by terminating the iterative process after only 5 iterations.

To reconstruct the ill-conditioned region, a weighted total variation minimization problem was solved:

$$\chi^* = argmin_\chi \|\chi\|_{\mathrm{WTV}} + \frac{\mu}{2}\|\mathbf{M}\chi - \chi_{well}\|_2^2 \quad \textbf{[4]}$$

where $\mathbf{M} = \mathbf{F^{-1}}(|\mathbf{D}| > \delta)\mathbf{F}$ is a sampling matrix taking the value 1 in the well-conditioned region and 0 in the ill-conditioned region according to a threshold δ applied to $|\mathbf{D}|$, μ is the regularization parameter, $\|\chi\|_{\mathrm{WTV}} = \sum W\,|\nabla\chi|$ is the weighted anisotropic total variation, and $\mathrm{W} = 1/(|\nabla\chi_{well}| + 10^{-6})$ is a weighting matrix derived from the gradient (∇) of the well-conditioned susceptibility map (χ_{well}) reconstructed in step 1. The minimization was solved using the alternating

direction method of multipliers (ADMM) (57). The parameters used were $\mu = 6 \cdot 10^4$ and $\delta = 0.197$. The reconstruction time was 5.7 seconds.

Winning Approach: HFEN and SSIM

The JHU-XMU SFCR2 algorithm developed by Xu Li's team at Johns Hopkins University, Maryland, USA, was the winner in both HFEN and SSIM categories and used a two-step structural feature based collaborative reconstruction (SFCR) algorithm (58). In the first step, an interim susceptibility map $\hat{\chi}$ was reconstructed by using a compressed sensing (CS) model in the Fourier domain with two regularization constraints:

$$\hat{\chi} = \text{argmin}_{\chi}\, \lambda_1 \|diag(\mathbf{M})\chi_k(\mathbf{k}) - diag(\mathbf{M})\mathbf{F}\chi\|_2^2 + \left\|P_{mag}\nabla\chi\right\|_1 + \lambda_2\|\mathbf{R}\chi\|_2^2 \quad \textbf{[5]}$$

where the structural prior P_{mag} was derived by thresholding the gradient amplitude of the magnitude image with 30% voxels considered as edges for L1 regularization (in P_{mag} edges were set to 0 and regions with no edges to 1). The fidelity mask $\mathbf{R}$ for the L2 regularization was generated by combining masks obtained via thresholding a preliminary QSM map $\chi_k(\mathbf{k})$ calculated with TKD and its gradient (similar to Fig. 4 in (58), with thresholds of 0.04 ppm for QSM and 0.1 for its gradient norm square). $\mathbf{M}$ is a binary mask indicating the well-conditioned region in the Fourier domain, i.e. $\mathbf{M} = |\mathbf{D}| > \delta$ where δ is a threshold on the dipole kernel in the Fourier domain. Parameters chosen for this step were $\delta = 0.19$, $\lambda_1 = 50$ and $\lambda_2 = 2$ and processing was terminated after three iterations. The final susceptibility map was then fitted in the spatial domain using weighted minimization:

$$\chi = \text{argmin}_{\chi}\, \gamma_1\|W(\phi - \mathbf{F}^{-1}\mathbf{D}\mathbf{F}\chi)\|_2^2 + \left\|P_{\hat{\chi}}\nabla\chi\right\|_1 + \gamma_2\|\mathbf{R}\chi\|_2^2 \quad \textbf{[6]}$$

where the structural prior $P_{\hat{\chi}}$ was extracted from the interim susceptibility map $\hat{\chi}$ (the solution of Eq. 5) with similar 30% edge voxels, $W = 1/|\phi|^{1/3}$ is a weighting matrix calculated from the local field ϕ and the same fidelity mask $\mathbf{R}$ as in the first step was used. Regularization parameters chosen for this step were $\gamma_1 = 50$ and $\gamma_2 = 1$, and iterative processes were terminated after 2 iterations.

Winning Approach: ROI accuracy

The winner in this category was the morphology-adaptive total variation (MATV) algorithm developed by Yanqiu Feng's team from Southern Medical University, Guangzhou, China. This algorithm first classifies the imaging target into smooth and non-smooth regions by thresholding the magnitude gradient map (59). In the dipole inversion, the regularization weights are adapted according to local morphological information: voxels in smooth regions are assigned larger TV regularization weights than in non-smooth regions. The QSM reconstruction via the MATV algorithm can be formulated as follows:

$$\chi = \text{argmin}_{\chi}\|\mathrm{W}(\phi - \mathbf{F}^{-1}\mathbf{D}\mathbf{F}\chi)\|_2^2 + \ \alpha\left\|P_{mag}\nabla\chi\right\|_1 + \ \beta\left\|(1 - P_{mag})\nabla\chi\right\|_1 \quad \textbf{[7]}$$

where W is a data weighting matrix to compensate the measured field noise (60) and α and β are the regularization parameters. The regularization parameters used were $\alpha = 0.003, \beta = 0.0009$.

Given only a marginal difference to the above described approach, we would like to acknowledge also the Primal-Dual and Forward Gradient Implementation (PD) algorithm developed by Yi Wang's team from Cornell University, New York, USA (61).

DISCUSSION

The QSM 2016 reconstruction challenge established a framework for the numerical comparison of QSM algorithms. We limited the challenge to a single data set that matched conventional clinical acquisitions closely with respect to the resolution, readout bandwidth, echo time and coverage. In the following, we summarize the results, discuss the limitations of the design of the challenge, and highlight the lessons learned.

Summary of results

The JHU-XMU SFCR2 algorithm won in two categories, SSIM and HFEN, and finished second in the RMSE ranking. The other winners, regarding RMSE and ROI accuracy, were the VANC UBC and SMU MATV algorithms. The top three algorithms in the RMSE ranking relied on reconstruction approaches known from compressed sensing (CS) MRI. As opposed to regularized inversion, where the entire Fourier domain is affected by regularization, in CS-approaches only the ill-conditioned Fourier sub-domain of the susceptibility map is estimated by minimizing a sparsity enforcing metric. This limitation to only a sub-space of the Fourier domain was probably the key for allowing these top-ranking approaches to produce the best reconstruction accuracies. However, also these winning maps were not ideal from a visual or radiological point of view, suffering from over-smoothing, and conspicuity loss in fine structures (Figures 2 and 3). CS techniques employed in accelerated MR data acquisition exploit incoherent aliasing artifacts arising from pseudo-random under-sampling of the k-space (62). The dipole artifacts in QSM reconstruction, however, appear more structured due to under-sampling only near the magic angle in the Fourier domain. Although the incoherent aliasing prerequisite for CS was fully not met, we think that these strategies performed well due to two main reasons:

(i) Because the missing content in the ill-conditioned region is a relatively small portion (e.g. 20-30%) of the Fourier domain, its estimation is easier, and potential blurring artifacts mainly impact this conical region, whereas the majority of the Fourier spectrum of the susceptibility cannot be altered to minimize the employed sparsity or smoothness metric (54).

ii) CS methods involving wavelet penalties do enjoy partial incoherence because the undersampling artifacts are distributed across the wavelet scales. The incoherence in both Total Variation and wavelet domains can be further improved by randomly under-sampling the ill-conditioned region (54).

Despite these points, CS methods are not necessarily immune to over-smoothing if they allow a reduction of the data consistency with a large regularization parameter. In this case, the data consistency becomes less important than the prior information, i.e. well-conditioned frequency content is no longer kept intact in favor of matching the CS constraint.

TKD and CFL2 solutions were provided as benchmark algorithms. The performance metrics RMSE / HFEN / SSIM for these algorithms were: 86.5 / 82.0 / 0.77 for TKD and 81.2 / 75.5 / 0.81 for CFL2. The winning algorithms had metrics: 69.0 / 63.5 / 0.94, corresponding to an improvement of 18% in RMSE, 19% in HFEN and 16% in SSIM over CFL2. The improvement in ROI accuracy was smaller, CFL2 ranked seventh in this category. We conclude that if the average susceptibilities inside specific grey and white matter ROIs are desired, a method as simple as CFL2 may provide sufficient accuracy. The submitted algorithms, however, provided a marked improvement in artifact mitigation and retention of high frequency features relative to the CFL2 benchmark.

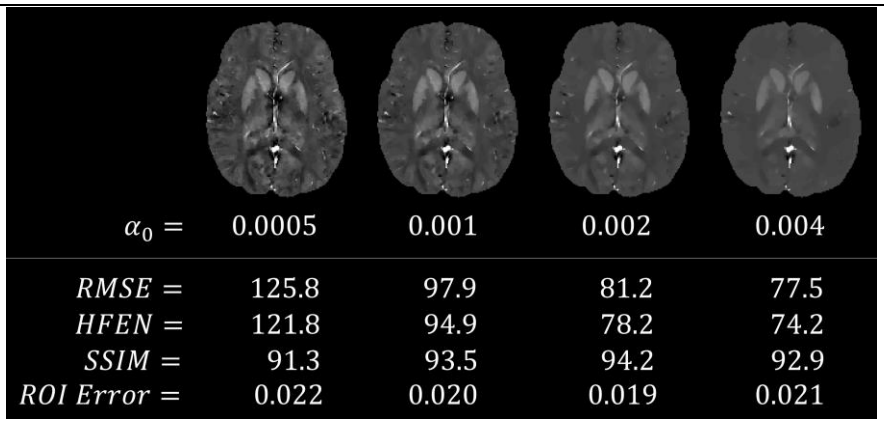
In the last few years, several research groups have proposed single-step QSM algorithms, which estimate the underlying susceptibility directly from the raw phase without separate interim phase processing. Despite the fact that a very specific phase filtering pipeline (LBV + polynomial fitting) was applied to create the reference susceptibility maps, the single-step algorithms were capable of providing competitive results despite the processing pipeline bias for multi-step approaches in this challenge.

Among the submissions, one approach employed parallel computing on GPU hardware for rapid dipole inversion (CHICAGO TGV). Despite solving the same underlying mathematical problem as other TGV based methods, this has yielded different performance metrics due to (i) using different regularization parameters and a different number of iterations, (ii) different implementations of the mathematical libraries and (iii) double precision (CPU) versus single precision (GPU) computation also has a substantial impact on iterative methods as any numerical differences accumulate.

However, the main discussion points of this reconstruction challenge were the identification of performance metrics that would be representative of susceptibility image quality and the selection of reference susceptibility maps.

How representative are RMSE, HFEN, and SSIM of susceptibility map quality?

All three measures are global error metrics aiming to summarize the mismatch against a reference image in a single number. We intentionally allowed the contestants to optimize for low RMSE by extensive parameter search. Although the algorithm applied for the post-challenge experiments shown in Figure 4 yielded highly over-regularized smooth QSM images, the resulting RMSE was only approximately 10% higher than that of the winning algorithm. RMSE is a simple global error metric and is usually not a reliable indicator of visual quality or over-smoothing by itself. Recognizing this, we added HFEN and SSIM to create a multi-dimensional performance vector that would allow us to probe differences between image features. However, the limited visual quality of the submitted susceptibility maps leads us to the conclusion that it will be important to find better metrics for the evaluation of susceptibility map quality.



$\alpha_0 =$	0.0005	0.001	0.002	0.004
$RMSE =$	125.8	97.9	81.2	77.5
$HFEN =$	121.8	94.9	78.2	74.2
$SSIM =$	91.3	93.5	94.2	92.9
$ROI\ Error =$	0.022	0.020	0.019	0.021

Figure 4: QSM algorithms were optimized to minimize error metrics in this challenge. This figure shows results of the GRAZ TGV algorithm with varying regularization parameter α_0. While the QSM image with $\alpha_0 = 0.004$ (right) suffered from over-smoothing and conspicuity loss in fine features such as vessels and the cortex, the RMSE was better than for the normally utilized $\alpha_0 = 0.0005$ (left). QSM images are scaled from -0.1 to 0.25 ppm.

The reason why we provided the reference susceptibility map was to ensure that each algorithm produced the best scoring susceptibility map for the given metrics. However, a major outcome of this challenge turned out to be that the chosen numerical metrics, which are intensively applied in computer vision research, were problematic because they favored over-smoothing of the reconstructed susceptibility maps. Over-regularization was consistently observed as a strategy to improve all error metrics, leading to an unexpected visual appearance of the susceptibility maps that differs from the typical appearance of the maps known from the literature.

Further insights on this issue could be gained by comparing the three fidelity metrics (RMSE, HFEN, and SSIM) and quantitative accuracy inside regions of interest (ROI error). The simple CFL2 method ranked 3rd in ROI accuracy, while not making it into the top 10 in any of the fidelity metrics. Despite performing well when average values were considered inside ROIs, its image quality suffered from streaking and blurring artifacts, which were better captured by RMSE, HFEN, and SSIM metrics. However, it is more difficult to gain insights from a comparison between the three fidelity metrics. RMSE, HFEN, and SSIM aimed to capture overall error, high-frequency deviation, and "visual" fidelity, respectively.

In future evaluations, some limitations of the metrics could be mitigated by incorporating experts' visual rating of the submitted susceptibility maps. A potential way to amend the RMSE metric could be to compare the gradients of the susceptibility map against those in the reference map via $\nabla\mathrm{RMSE} = 100 \cdot \|\nabla(\chi_{33} - \chi_{recon})\|_2^2 \,/\, \|\nabla\chi_{33}\|_2^2$. The metric $\nabla\mathrm{RMSE}$ may provide a more direct measure of the fidelity of high frequency components, and could complement the existing metrics and the visual rating.

However, while optimization in respect to certain quality measures will require further systematic investigations, the degree of regularization should be also chosen according to the subsequent usage of the QSM images depending whether this is anatomical ROI evaluation, voxel-based analysis or visual inspection by radiologists.

Selection of the reference susceptibility map

We selected χ_{33} instead of the COSMOS solution as standard reference to eliminate the potential orientation bias in the latter susceptibility map. However, this required the assumption that phase contributions from the off-diagonal tensor terms χ_{13} and χ_{23} in the transverse plane are negligible. As demonstrated in Figure 1, these contributions are non-negligible because the tensor elements can have about 70% amplitude relative to χ_{33}.

One potential way to combine the strengths of both reference map candidates in future challenges would be to mask out the anisotropic regions in the COSMOS map. Such an anisotropy mask could be obtained by thresholding the STI anisotropy defined as $\chi_{msa} = \lambda_1 - (\lambda_2 + \lambda_3)/2$ where λ_i are the susceptibility tensor eigenvalues. This mask could be refined using a white matter segmentation.

Also related to limitations of the employed reference, there is clear evidence that the microstructural compartmentalization of magnetic susceptibility in white matter and its water distribution has a significant impact on the observed phase images (33–36). These effects are not accounted for by either COSMOS, STI or any of the single orientation reconstruction methods, yielding an error in susceptibility values in fiber bundles (33) that is difficult to estimate. As white matter represents a relatively large brain volume fraction, both white matter (WM) measurements and whole brain metrics will be affected by these microstructural effects, and a particular regularization might inadvertently improve the metrics without resulting in a more accurate or precise reconstruction.

In summary, the existence of phase contrast related to off-diagonal tensor elements poses the question of what is the perfect susceptibility map reconstructed from a single-angle phase image? The presence of non-susceptibility contrast mechanisms including chemical exchange-induced frequency shifts (30), which are currently not accounted for by multi-orientation QSM algorithms poses the question of how we can measure the gold standard *in-vivo* susceptibility map. Both seem to be open questions in our field.

Lessons learned from the first QSM reconstruction challenge

We are fully determined to push forward, improve and extend this research endeavor based on lessons learned from this initial challenge. In particular, the feedback from members of the QSM community who attended the Graz Workshop and the EMTP study group meeting at the ISMRM 2017 was encouraging to proceed with an evaluation of the various algorithmic approaches to better understand the potential and limitations of QSM. The main suggestions and recommendations addressed the limitations of the performance metrics for evaluation of the submitted susceptibility maps and the choice of the reference map. We list the conclusions from the various discussions in the following:

I. Instead of relying entirely on error metrics, it would be informative for experienced radiologists and QSM experts to perform a visual assessment of submitted susceptibility maps.
II. The challenge could be divided into two parts, where the first part would assess the quantitative accuracy with respect to a known ground truth. To this end, the challenge could comprise phase data obtained using the forward model (11) on a realistic numerical brain phantom derived from STI or COSMOS susceptibility maps. The data could be made more realistic by adding noise, partial volume artifacts and contributions from anisotropic susceptibility sources. The second part would involve *in-vivo* patient data and aim to assess the robustness of the methods in a clinical scenario, in the potential presence of motion, blooming and signal dropout artifacts. For parameter tuning, a COSMOS reference from a healthy control could be provided.
III. Reference and submitted susceptibility maps could be compared on a per-voxel basis by assessing one-dimensional profiles or correlation coefficients.

IV. A better *in-vivo* reference map could be created by incorporating the contribution of χ_{13} and χ_{23} into the field map provided to the contestants (transverse orientation). A potential way of implementing this could be by rearranging the STI relation in the transverse plane as follows:

$$\begin{aligned}\boldsymbol{\Theta}(\boldsymbol{k}) &= \mathbf{D}\,\chi_{33} - \frac{k_z}{k^2}\left(k_x\,\chi_{13} + k_y\,\chi_{23}\right) \\ &= \mathbf{D}\chi_{33} - \frac{k_z^2}{k^2}\left(\frac{k_x}{k_z}\,\chi_{13} + \frac{k_y}{k_z}\,\chi_{23}\right)\end{aligned} \quad \textbf{[8]}$$

now defining $\tilde{\chi} \triangleq \frac{k_x}{k_z}\,\chi_{13} + \frac{k_y}{k_z}\,\chi_{23}$,

$$\boldsymbol{\Theta}(\boldsymbol{k}) = \mathbf{D}\chi_{33} - \frac{k_z^2}{k^2}\tilde{\chi} + \frac{1}{3}\tilde{\chi} - \frac{1}{3}\tilde{\chi} \quad \textbf{[9]}$$

$$\boldsymbol{\Theta}(\boldsymbol{k}) + \frac{1}{3}\tilde{\chi} = \mathbf{D}(\chi_{33} + \tilde{\chi}) \quad \textbf{[10]}$$

Equation 10 suggests that a new ground truth susceptibility could be created by $\chi_{new} \triangleq \chi_{33} + \tilde{\chi}$ and that the input local field data could be amended by $\boldsymbol{\Theta}_{new} \triangleq \boldsymbol{\Theta} + \frac{1}{3}\tilde{\chi}$. A numerical challenge in computing $\tilde{\chi}$ would be the division by k_z for the plane of frequencies where $k_z = 0$. To address this, we can interpret division by k_z as integration along z in image-space, and multiplication by k_x (or k_y) as differentiation along x (or y) axes (1). In a discrete implementation, integration would correspond to summation over z indices and differentiation would be the difference between neighboring voxels along x (or y) axes.

V. Include the computational efficiency as additional information or a separate category, which would require access to a single evaluation computer for all contestants on which the processing time of all algorithms could be accurately determined and compared.

VI. The data consistency could be used as an additional metric. To this end, one could use the submitted susceptibility maps in a forward field simulation (11) and compare the resulting phase against the measured phase.

VII. Susceptibility maps should be compensated for the known systematic underestimation before quality metrics are calculated. This approach would avoid underestimation resulting in poor metrics despite the reconstruction being of otherwise high quality.

VIII. The susceptibility could be evaluated exclusively in deep grey matter structures where QSM is more likely to be correct given the absence of highly anisotropic fiber bundles.

IX. The mutual information, cross correlation and $\nabla\mathrm{RMSE}$ between the reference and submitted maps could be included as additional quality metrics.

X. Multi-echo phase data could be provided to allow field maps to be derived by fitting the phase over echo times (63–65).

Most of these suggestions are feasible but may require additional data processing and acquisition. We have already updated the downloadable data set to include the magnitude and phase data from all 12 directions. This dataset could facilitate extensions such as an STI challenge, or future research towards computation of a better reference map. Since this dataset includes χ_{33} and χ_{COSMOS} as well as all components of the susceptibility tensor, future publications on new algorithms may report performance metrics relative to any of these.

Another interesting avenue to explore could be issuing sub-challenges with clinical data from populations with different diseases. Such a challenge would be an excellent opportunity to test the robustness of the algorithms in the clinical setting, and performance evaluation would benefit from the experience of neuro-radiologists. However, the lack of a true gold standard reference renders the quantitative assessment of susceptibility maps beyond the description of apparent artifacts difficult.

CONCLUSION

The substantive differences between the various submitted susceptibility maps highlight a critical limitation of current regularized QSM techniques: the appearance of the resulting susceptibility maps depends strongly on the algorithm used and the associated parameter choices. Hence, a direct comparison of results from studies employing different QSM algorithms and parameters is challenging. Consequently, in the EMTP study group meeting at the ISMRM 2017 it was consensually decided that another challenge will be designed based on the lessons learned from the present challenge.

ACKNOWLEDGEMENTS

This QSM 2016 reconstruction challenge was embedded in the framework of the ISMRM-endorsed 4th International QSM workshop in 2016. We are grateful for the support of the QSM workshop steering committee and the opportunity to present the challenge concept at the EMTP study group meeting at the ISMRM 2016 in Singapore. We would like to thank the workshop's local organizing committee led by Stefan Ropele and his team for hosting this event and providing generous prizes for the winning QSM algorithms. The contestants would like to acknowledge participating team members: Julio Acosta-Cabronero (DZNE, Magdeburg, Germany), Lijun Bao (Xiamen University, China), Emma Biondetti (University College London, UK), Andreas Deistung (Jena University Hospital, Germany), Max Diefenbach (Technical University of Munich, Germany), Emma Dixon (University College London, UK), Yanqiu Feng (Southern Medical University, China), E. Mark Haacke (Wayne State University, Detroit, MI, USA), Anita Karsa (University College London, UK), Ulrich Katscher (Philips Research Europe, Hamburg, Germany), Chunlei Liu (Berkeley, CA, USA), Saifeng Liu (Institute for Biomedical Research, Waterloo, Canada), Stefan Ropele (Medical University of Graz, Austria), Cristian Tejos (Pontificia Universidad Catolica, Chile), David Vicente (University of Graz, Austria), Yuyao Zhang (Berkeley, CA, USA), Bo Zhao (Martinos Center, Boston, USA).

Disclosure

The organizers (CL, FS and BB) participated in this QSM challenge even though they were responsible for the concept, data acquisition, pre-processing, and evaluation of submitted QSM images. However, unlike in image reconstruction challenges that only provide a subset of the data for reconstruction and then test against a gold standard that is unknown to the challenge participants, the same reference images were available to all participants in the present challenge. Consequently, there was no specific advantage for the organizers.

Funding information

This work was supported by the Austrian Science Fund (FWF grant numbers: KLI523 and P30134). Research reported in this publication was partially funded by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR001412. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

1. Wang Y, Liu T. Quantitative susceptibility mapping (QSM): Decoding MRI data for a tissue magnetic biomarker. *Magn. Reson. Med.* 2015;73:82–101. doi: 10.1002/mrm.25358.
2. Liu C, Li W, Tong K a., Yeom KW, Kuzminski S. Susceptibility-weighted imaging and quantitative susceptibility mapping in the brain. *J. Magn. Reson. Imaging* 2015;42:23–41. doi: 10.1002/jmri.24768.
3. Haacke EM, Liu S, Buch S, Zheng W, Wu D, Ye Y. Quantitative Susceptibility Mapping: Current Status and Future Directions. *Magn. Reson. Imaging* 2014;33:1–25. doi: 10.1016/j.mri.2014.09.004.
4. Deistung A, Schweser F, Reichenbach JR. Overview of quantitative susceptibility mapping. *NMR Biomed.* 2016. doi: 10.1002/nbm.3569.
5. Schweser F, Deistung A, Reichenbach JR. Foundations of MRI phase imaging and processing for Quantitative Susceptibility Mapping (QSM). *Z. Med. Phys.* 2016;26:6–34. doi: 10.1016/j.zemedi.2015.10.002.
6. Holt RW, Diaz PJ, Duerk JL, Bellon EM. MR susceptometry: an external-phantom method for measuring bulk susceptibility from field-echo phase reconstruction maps. *J. Magn. Reson. Imaging* 1994;4:809–818. doi: 10.1002/jmri.1880040612.
7. Sepulveda NG, Thomas IM, Wikswo JP. Magnetic susceptibility tomography for three-dimensional imaging of diamagnetic and paramagnetic objects. *IEEE Trans. Magn.* 1994;30:5062–5069. doi: 10.1109/20.334296.
8. Hwang SN, Wehrli FW. The Calculation of the Susceptibility-Induced Magnetic Field from 3D NMR Images with Applications to Trabecular Bone. *J. Magn. Reson. Ser. B* 1995;109:126–145. doi: 10.1006/jmrb.1995.0002.
9. Haacke EMM, Cheng NYY, House MJJ, Liu Q, Neelavalli J, Ogg RJJ, Khan A, Ayaz M, Kirsch W, Obenaus A. Imaging iron stores in the brain using magnetic resonance imaging. *Magn. Reson. Imaging* 2005;23:1–25.
10. Yeo DTB, Fessler JA, Kim B. Motion robust magnetic susceptibility and field inhomogeneity estimation using regularized image restoration techniques for fMRI. *Med. Image Comput. Comput. Assist. Interv.* 2008;11:991–8.
11. Marques JP, 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;25B:65–78. doi: 10.1002/cmr.b.20034.
12. Li L. Magnetic susceptibility quantification for arbitrarily shaped objects in inhomogeneous fields. *Magn. Reson. Med.* 2001;46:907–16.
13. 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.
14. Shmueli K, de Zwart JA, 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;62:1510–22. doi: 10.1002/mrm.22135.
15. Liu C, Wei H, Gong N, Cronin M, Dibb R, Decker K. Quantitative Susceptibility Mapping: Contrast Mechanisms and Clinical Applications. *Tomography* 2015;1:3–17. doi: 10.18383/j.tom.2015.00136.
16. 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.
17. Stüber C, Pitt D, Wang Y. Iron in Multiple Sclerosis and Its Noninvasive Imaging with Quantitative Susceptibility Mapping. *Int. J. Mol. Sci.* 2016;17:100. doi: 10.3390/ijms17010100.
18. Langkammer C, Pirpamer L, Seiler S, Deistung A, Schweser F, Franthal S, Homayoon N, Katschnig-Winter P, Koegl-Wallner M, Pendl T, Stoegerer EM, Wenzel K, Fazekas F, Ropele S, Reichenbach JR, Schmidt R, Schwingenschuh P. Quantitative Susceptibility Mapping in Parkinson's Disease. *PLoS One* 2016;11:e0162460. doi: 10.1371/journal.pone.0162460.
19. Eskreis-Winkler S, Zhang Y, Zhang J, Liu Z, Dimov A, Gupta A, Wang Y. The clinical utility of QSM: disease diagnosis, medical management, and surgical planning. *NMR Biomed.* 2016. doi: 10.1002/nbm.3668.
20. Sharma SD, Hernando D, Horng DE, Reeder SB.

Quantitative susceptibility mapping in the abdomen as an imaging biomarker of hepatic iron overload. *Magn. Reson. Med.* 2015;74:673–83. doi: 10.1002/mrm.25448.

21. Dibb R, Qi Y, Liu C. Magnetic susceptibility anisotropy of myocardium imaged by cardiovascular magnetic resonance reflects the anisotropy of myocardial filament α-helix polypeptide bonds. *J. Cardiovasc. Magn. Reson.* 2015;17:60. doi: 10.1186/s12968-015-0159-4.

22. Wei H, Dibb R, Decker K, Wang N, Zhang Y, Zong X, Lin W, Nissman DB, Liu C. Investigating magnetic susceptibility of human knee joint at 7 tesla. *Magn. Reson. Med.* 2017. doi: 10.1002/mrm.26596.

23. Xie L, Sparks MA, Li W, Qi Y, Liu C, Coffman TM, Johnson GA. Quantitative susceptibility mapping of kidney inflammation and fibrosis in type 1 angiotensin receptor-deficient mice. *NMR Biomed.* 2013;26:1853–63. doi: 10.1002/nbm.3039.

24. Straub S, Laun FB, Emmerich J, Jobke B, Hauswald H, Katayama S, Herfarth K, Schlemmer H-P, Ladd ME, Ziener CH, Bonekamp D, Röthke MC. Potential of quantitative susceptibility mapping for detection of prostatic calcifications. *J. Magn. Reson. Imaging* 2017;45:889–898. doi: 10.1002/jmri.25385.

25. 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.

26. 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. doi: 10.1002/nbm.3570.

27. 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.

28. Wharton S, Schäfer A, Bowtell R. Susceptibility mapping in the human brain using threshold-based k-space division. *Magn. Reson. Med.* 2010;63:1292–304. doi: 10.1002/mrm.22334.

29. Erdevig HE, Russek SE, Carnicka S, Stupic KF, Keenan KE. Accuracy of magnetic resonance based susceptibility measurements. *AIP Adv.* 2017;7:56718. doi: 10.1063/1.4975700.

30. Shmueli K, Dodd SJ, Li T-QQ, Duyn JH. The contribution of chemical exchange to MRI frequency shifts in brain tissue. *Magn. Reson. Med.* 2011;65:35–43. doi: 10.1002/mrm.22604.

31. Lee J, Shmueli K, Fukunaga M, van Gelderen P, Merkle H, Silva AC, Duyn JH. Sensitivity of MRI resonance frequency to the orientation of brain tissue microstructure. *Proc. Natl. Acad. Sci. U. S. A.* 2010;107:5130–5. doi: 10.1073/pnas.0910222107.

32. Li W, Wu B, Avram A V, Liu C. Magnetic susceptibility anisotropy of human brain in vivo and its molecular underpinnings. *Neuroimage* 2012;59:2088–97. doi: 10.1016/j.neuroimage.2011.10.038.

33. Wharton S, Bowtell R. Effects of white matter microstructure on phase and susceptibility maps. *Magn. Reson. Med.* 2015;73:1258–69. doi: 10.1002/mrm.25189.

34. Yablonskiy DA, Sukstanskii AL. Effects of biological tissue structural anisotropy and anisotropy of magnetic susceptibility on the gradient echo MRI signal phase: theoretical background. *NMR Biomed.* 2016. doi: 10.1002/nbm.3655.

35. Luo J, He X, Yablonskiy D a. Magnetic susceptibility induced white matter MR signal frequency shifts--experimental comparison between Lorentzian sphere and generalized Lorentzian approaches. *Magn. Reson. Med.* 2014;71:1251–63. doi: 10.1002/mrm.24762.

36. He X, Yablonskiy DA. Biophysical mechanisms of phase contrast in gradient echo MRI. *Proc. Natl. Acad. Sci. U. S. A.* 2009;106:13558–63. doi: 10.1073/pnas.0904899106.

37. Liu C. Susceptibility tensor imaging. *Magn. Reson. Med.* 2010;63:1471–7. doi: 10.1002/mrm.22482.

38. Mugler JP, Brookeman JR. Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magn. Reson. Med.* 1990;15:152–7.

39. 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.

40. 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.* 2016. doi: 10.1002/nbm.3604.

41. Smith SM. Fast robust automated brain extraction. *Hum. Brain Mapp.* 2002;17:143–155. doi: 10.1002/hbm.10062.

42. Paige CC, Saunders MA. LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Trans. Math. Softw.* 1982;8:43–71. doi: 10.1145/355984.355989.

43. Bilgic B, Xie L, Dibb R, Langkammer C, Mutluay A, Ye H, Polimeni JR, Augustinack J, Liu C, Wald LL, Setsompop K. Rapid multi-orientation quantitative susceptibility mapping. *Neuroimage* 2016;125:1131–41. doi: 10.1016/j.neuroimage.2015.08.015.

44. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* 1999;42:952–962.

45. Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM. The NMR phased array. *Magn. Reson. Med.* 1990;16:192–225.

46. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.* 2002;47:1202–1210. doi: 10.1002/mrm.10171.

47. Fujimoto K, Polimeni JR, van der Kouwe AJW, Reuter M, Kober T, Benner T, Fischl B, Wald LL. Quantitative comparison of cortical surface reconstructions from MP2RAGE and multi-echo MPRAGE data at 3 and 7T. *Neuroimage* 2013. doi: 10.1016/j.neuroimage.2013.12.012.

48. Bilgic B, Chatnuntawech I, Fan AP, Setsompop K, Cauley SF, Wald LL, Adalsteinsson E. Fast image reconstruction with L2-regularization. *J. Magn. Reson. Imaging* 2013;0:1–11. doi: 10.1002/jmri.24365.

49. Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imaging* 2011;30:1028–41. doi: 10.1109/TMI.2010.2090538.

50. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 2004;13:600–12.

51. Kames C, Wiggermann V, Rauscher A. Rapid two-step QSM without a priori information. In: Proceedings of the 24th Annual Meeting ISMRM. ; 2016. p. 256.

52. Li W, Wu B, Liu C. Quantitative susceptibility mapping of human brain reflects spatial variation in tissue composition. *Neuroimage* 2011;55:1645–56. doi: 10.1016/j.neuroimage.2010.11.088.

53. Schweser F, Sommer K, Deistung A, Reichenbach JR. Quantitative susceptibility mapping for investigating subtle susceptibility variations in the human brain. *Neuroimage* 2012;62:2083–100. doi: 10.1016/j.neuroimage.2012.05.067.

54. Wu B, Li W, Guidon A, Liu C. Whole brain susceptibility mapping using compressed sensing. *Magn. Reson. Med.* 2012;67:137–47. doi: 10.1002/mrm.23000.

55. Fong DC-L, Saunders M. LSMR: An Iterative Algorithm for Sparse Least-Squares Problems. *SIAM J. Sci. Comput.* 2011;33:2950–2971. doi: 10.1137/10079687X.

56. Wang S, Liu T, Chen W, Spincemaille P, Wisnieff C, Tsiouris AJ, Zhu W, Pan C, Zhao L, Wang Y. Noise Effects in Various Quantitative Susceptibility Mapping Methods. *IEEE Trans. Biomed. Eng.* 2013;60:3441–3448. doi: 10.1109/TBME.2013.2266795.

57. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Found. Trends® Mach. Learn.* 2010;3:1–122. doi: 10.1561/2200000016.

58. Bao L, Li X, Cai C, Chen Z, van Zijl P. Quantitative Susceptibility Mapping using Structural Feature based Collaborative Reconstruction (SFCR) in the Human Brain. *IEEE Trans. Med. Imaging* 2016;62:1–1. doi: 10.1109/TMI.2016.2544958.

59. Guo L, Guo Y, Mei Y, Guan J, Chen W, Feng Y. Quantitative Susceptibility Map Reconstruction from MR Phase Data Using Morphology-Adaptive Total Variation. In: Proc. Intl. Soc. Mag. Reson. Med. ; 2017. p. 5452.

60. Liu J, Liu T, de Rochefort L, Ledoux J, Khalidov I, Chen W, Tsiouris AJ, Wisnieff C, Spincemaille P, Prince MR, Wang Y. Morphology enabled dipole inversion for quantitative susceptibility mapping using structural consistency between the magnitude image and the susceptibility map. *Neuroimage* 2012;59:2560–8. doi:

10.1016/j.neuroimage.2011.08.082.

61. Kee Y, Deh K, Dimov A, Spincemaille P, Wang Y. Primal-dual and forward gradient implementation for quantitative susceptibility mapping. *Magn. Reson. Med.* 2017. doi: 10.1002/mrm.26627.

62. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* 2007;58:1182–95. doi: 10.1002/mrm.21391.

63. Gilbert G, Savard G, Bard C, Beaudoin G. Quantitative comparison between a multiecho sequence and a single-echo sequence for susceptibility-weighted phase imaging. *Magn. Reson. Imaging* 2012;30:722–30. doi: 10.1016/j.mri.2012.02.008.

64. Wu B, Li W, Avram AV, Gho S-MM, Liu C. Fast and tissue-optimized mapping of magnetic susceptibility and T2* with multi-echo and multi-shot spirals. *Neuroimage* 2012;59:297–305. doi: S1053-8119(11)00780-4 [pii]10.1016/j.neuroimage.2011.07.019.

65. Biondetti, Emma, Karsa A, Thomas DL, Shmueli K. Evaluating The Accuracy of Susceptibility Maps Calculated from Single-Echo versus Multi- Echo Gradient-Echo Acquisitions. In: Proc. Intl. Soc. Mag. Reson. Med. ; 2017. p. 2258.

66. Langkammer C, Bredies K, Poser B a., 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.

67. Schweser F, Deistung A, Sommer K, Reichenbach JR. Toward online reconstruction of quantitative susceptibility maps: superfast dipole inversion. *Magn. Reson. Med.* 2013;69:1582–94. doi: 10.1002/mrm.24405.

68. Kressler B, de Rochefort L, Liu T, Spincemaille P, Jiang Q, Wang Y. Nonlinear regularization for per voxel estimation of magnetic susceptibility distributions from MRI field maps. *IEEE Trans. Med. Imaging* 2010;29:273–81. doi: 10.1109/TMI.2009.2023787.

69. Milovic C, Bilgic B, Zhao B, Acosta-Cabronero J, Tejos C. A Fast Algorithm for Nonlinear QSM Reconstruction. In: Proc. Intl. Soc. Mag. Reson. Med. ; 2017. p. 3669.

70. Milovic C, Pinto JM, Acosta-Cabronero J, Dusek P, Madai VI, Huelnhagen T, Niendorf T, Wuerfel J, Tejos C. Improved magnetic dipole kernel for reconstruction methods in quantitative susceptibility mapping. In: Proc. Intl. Soc. Mag. Reson. Med. ; 2016. p. 2843.

71. Jenkinson M, Wilson JL, Jezzard P. Perturbation method for magnetic field calculations of nonconductive objects. *Magn. Reson. Med.* 2004;52:471–477. doi: 10.1002/mrm.20194.

72. Wei H, Dibb R, Zhou Y, Sun Y, Xu J, Wang N, Liu C. Streaking artifact reduction for quantitative susceptibility mapping of sources with large dynamic range. *NMR Biomed.* 2015;28:1294–303. doi: 10.1002/nbm.3383.

73. Tang J, Liu S, Neelavalli J, Cheng YCN, Buch S, Haacke EM. Improving susceptibility mapping using a threshold-based K-space/image domain iterative reconstruction approach. *Magn. Reson. Med.* 2013;69:1396–407. doi: 10.1002/mrm.24384.

74. Guo Y, Guo L, Mei Y, Feng Y. Enhancing Quantitative Susceptibility Mapping by Using Gradient L2 Regularization with Morphological Priors. In: Proc. Intl. Soc. Mag. Reson. Med. ; 2017. p. 1967.

75. Liu Z, Kee Y, Zhou D, Wang Y, Spincemaille P. Preconditioned total field inversion (TFI) method for quantitative susceptibility mapping. *Magn. Reson. Med.* 2016. doi: 10.1002/mrm.26331.

76. Meineke J, Wenzel F, Wilkinson ID, Katscher U. Quantitative Susceptibility Mapping in Alzheimer's Disease using Joint background-field removal and segmentation-Enhanced Dipole Inversion. In: Proc. Intl. Soc. Mag. Reson. Med. ; 2016. p. 4051.

TABLES

Table 1

Name	Description	Input Phase RAW/LBV*
TKD (provided)	Threshold-based K-space Division (28) with zeroes at the ill-conditioned regions (cone) in k-space, threshold = 0.19.	LBV
CFL2 (provided)	Closed-form L2-regularized inversion (48).	LBV
MARTINOS WTV	Compressed Sensing compensated QSM (54), accelerated reconstruction using ADMM optimization.	LBV
GRAZ TGV	Total generalized variation (TGV) based method incorporating phase unwrapping, background field removal and dipole inversion in a single iteration (66).	RAW
GRAZ TGV L1	Total generalized variation (TGV) based method (66) with additional L1 magnitude stabilization.	RAW
JENA HEIDI	Hybrid algorithm based on Homogeneity Enabled Incremental Dipole Inversion (HEIDI) that solves three sub-domains of k-space using different approaches, depending on the conditioning. The well-conditioned k-space was solved using unregularized LSQR, the critical part of the k-space was recovered by solving a weighted Total Variation Problem with priors derived from phase images, the transition area was derived from the LSQR solution using denoising (53). Parameters defining the three sub-domains were chosen to obtain optimal error measures relative to the gold standard.	LBV
JENA SDI	TKD algorithm with extreme thresholding of the dipole kernel and underestimation compensation based on the deconvolution point-spread function as in superfast dipole inversion (SDI) (67).	LBV
UCL TKD 1	TKD as in (14,67) i.e. without zeroes inside the k-space cone. A threshold of $\delta = \frac{2}{3}$ was used with no correction for χ underestimation.	LBV
UCL TIK	Closed-form Tikhonov inversion as alluded to in (68) and mentioned in (1) as a Tikhonov-regularized minimal norm solution. 1 had $\alpha = 0.0588$ and no correction for χ underestimation. 2 had $\alpha = 0.0588$ and correction for χ underestimation with a factor of 1.65. 4 had $\alpha = 0.025$ and correction for χ underestimation with a factor of 1.30.	LBV
JHU-XMU SFCRKDN	Based on the structural feature-based collaborative reconstruction (SFCR) QSM paper in (58), simplified the L2 regularization terms in M-step and S-step, added de-noising operation, k-space based L1 solver and HEIDI like k-space combination.	LBV
JHU-XMU SFCR2	Based on the SFCR QSM paper in (58), L1 and L2 regularized two-step reconstruction with regularization a priori extracted from magnitude and interim susceptibility maps – see winning approach in categories HFEN and SSIM.	LBV
CHILE TGV L2	Magnitude weighted TGV. Uses an L2 data fidelity term, spatially weighted by the square of the magnitude. First order approximation of the non-linear term (69).	LBV

CHILE TGV NL	Non-linear (NL) TGV result. It uses a non-linear data fidelity term, similar to Liu's nonlinear MEDI but with a fast solver with alternating direction method of multipliers (ADMM) and a mixture of a global and local solvers to deal with the nonlinear equation.	LBV
CHILE NLD	Discretization of the dipole kernel based on (70). It uses finite differences and the DFT to achieve an analytical solution in the Fourier domain.	LBV
CHILE NLG	Dipole kernel defined in space by the Green's function, integrating it for each voxel (71).	LBV
CHICAGO TGV	Algorithm based on the TGV QSM method (66), implemented on GPU-hardware (CUDA 7.5, NVIDIA GeForce GTX 980TI).	RAW
BERKELEY STAR	Streaking artifacts Reduction (STAR) via isolating and calculating strong susceptibility sources automatically, then large and small susceptibility values were reconstructed using a two-level TV approach (72).	LBV
VANC UBC	LSMR solver (55) followed by weighted compressed sensing minimization – see winning approach in category RMSE.	LBV
IBR ITSWIM	Variable regularization threshold for inverse process / k-space improvement with a binary mask including the deep grey matter nuclei and veins used in the iterative algorithm (73).	LBV
SMU MATV	Morphology-Adaptive Total Variation (MATV) separates target susceptibility into smooth and non-smooth regions, where the latter are assigned smaller TV weights than smooth regions during dipole inversion (59) – see winning approach in the ROI accuracy category.	LBV
SMU MTKD	TKD with morphological priors (MTKD). The target susceptibility map is separated into smooth and non-smooth regions by exploiting morphological information. The gradient of the target susceptibility map is forced to be zero in the smooth regions, and to be the gradient of TKD-reconstructed susceptibility map in the non-smooth regions (74).	LBV
NY MEDI	Morphology Enabled Dipole Inversion (MEDI) method using a Bayesian regularization approach that adds spatial priors from the magnitude image (13,25).	LBV
NY PD	Solving the objective of MEDI using the Primal-Dual (PD) formulation of the total variation and a forward difference method for discretization (61).	LBV
NY TFI	The Total Field Inversion (TFI) method simultaneously estimates the background and local fields, preventing error propagation from background field removal to QSM (75).	RAW
PHILIPS DTV	Single-step QSM starting from wrapped raw phase using Directional Total-Variation (DTV) with MPRAGE as a prior for estimating edges (76).	RAW

Table 1 Caption: Brief description of all QSM algorithms participating in the reconstruction challenge * RAW = raw phase (for single step algorithms), LBV = LBV preprocessed phase. The algorithms are named to reflect the team's institution or location followed by an abbreviation related to the technique(s) exploited by each algorithm.

Table 2

RMSE (%)		HFEN (%)		SSIM		ROI ERROR (ppm)	
69.0	VANC UBC	63.5	JHU-XMU SFCR2		JHU-XMU SFCR2		SMU MATV
70.3	JHU-XMU SFCR2	68.8	GRAZ TGV L1	0.94	JHU-XMU SFCRKDN	0.016	NY PD
73.6	MARTINOS WTV	68.9	VANC UBC		NY MEDI		CHILE TGV NL
74.2	PHILIPS DTV	70.9	PHILIPS DTV	0.93	GRAZ TGV	0.017	CHILE NLD
74.6	GRAZ TGV L1	71.8	SMU MATV	0.87	GRAZ TGV L1		SMU MTKD
75.2	UCL TIK 1	73.1	UCL TIK 1		CHILE TGV L2		CFL2
76.6	UCL TKD 1	73.6	MARTINOS WTV	0.84	NY TFI		UCL TIK 2
	GRAZ TGV	74.1	IBR ITSWIM		JENA HEIDI	0.018	UCL TIK 4
77.5	BERKELEY STAR		JHU-XMU SFCRKDN	0.83	CHILE NLD		JHU-XMU SFCRKDN
79.1	SMU MATV	74.2	GRAZ TGV		CHILE TGV NL		CHILE TGV L2

Table 2 caption: Top 10 algorithms with the best scores in each category evaluated for the QSM reconstruction challenge.