Quantitative Susceptibility Mapping: Report from the 2016 Reconstruction Challenge

Christian Langkammer, Ferdinand Schweser, Karin Shmueli, Christian Kames, Xu Li, Li Guo, Carlos Milovic, Jinsuh Kim, Hongjiang Wei, Kristian Bredies, Sagar Buch, Yihao Guo, Zhe Liu, Jakob Meineke, Alexander Rauscher, José P. Marques, Berkin Bilgic

(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 Imaging Center, 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 authors:
Christian Langkammer <christian.langkammer@medunigraz.at>
Ferdinand Schweser <schweser@buffalo.edu>
Berkin Bilgic <berkin@nmr.mgh.harvard.edu>
ABSTRACT

Purpose

The aim of the 2016 QSM reconstruction challenge was to test the ability of various QSM algorithms to faithfully recover the underlying susceptibility from noisy phase data.

Methods

Data set: Single orientation gradient-echo images of a healthy volunteer acquired at 3 Tesla 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.

Evaluation: 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 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 very 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.
INTRODUCTION

Quantitative susceptibility mapping (QSM) allows the determination of 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 were 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).

A variety of algorithms have 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, the susceptibility maps resulting from these algorithms appear to have 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 (http://www.qsm2016.com). The primary goal of the challenge was to test the ability of various QSM algorithms to recover faithfully the underlying susceptibility distribution from a healthy volunteer’s phase data. 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.

Details of the challenge were presented 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 (http://qsm.neuroimaging.at) starting from May 12th, 2016, and the deadline for 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.

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 solutions or gels with known magnetic susceptibilities, i.e. regions of piece-wise constant magnetic susceptibility. Such a geometry allows 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 true 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.

Some authors have used numerical models to evaluate their algorithms. However, the phase measured in vivo contains contributions from sources other than isotropic bulk magnetic susceptibility such as chemical exchange effects (30), anisotropic susceptibility (31,32) and microstructure (33–36). The magnitude of these effects is not yet completely understood and they are, therefore, difficult to replicate correctly in a numerical model. 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 anisotropic magnetic susceptibility into account, we employed the susceptibility tensor imaging (STI) approach (37). 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.

The rationale behind providing a reference susceptibility map to the contestants was to eliminate the confounding effects of algorithm parameter choices. With the availability of the reference map, contestants were able to optimize algorithmic parameters and submit the best result they could achieve with their algorithm.

**Selection of the reference**

The candidates for gold standard susceptibility were either COSMOS reconstruction (27), or $\chi_{33}$ 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 may be regarded as a susceptibility map averaged over all 12 orientations of the head with respect to the main magnetic field. 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 $\chi_{33}$ of the STI solution as the reference. Based on STI theory (37), the k-space phase $\Theta(k)$, when the main field lies along $H$ in the subject frame, is given by

$$\Theta(k) = \frac{1}{3} H^T \cdot X \cdot H - H \cdot k \frac{k^T X H}{k^2}$$  \[1\]
where $k$ is a vector of all k-space coordinates and $X$ is the susceptibility tensor in the subject frame and $(\cdot)^T$ denotes matrix transposition. When the acquisition is performed in the transverse plane relative to the subject coordinates, i.e. $H = [0,0,1]^T$, the signal equation becomes

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

If we ignore the off-diagonal terms $\chi_{13}$ and $\chi_{23}$, this leads to the popular relation used for dipole inversion from data acquired at transverse orientation relative to $B_0$:

$$\Theta(k) = \left(\frac{1}{3} - \frac{k_z^2}{k^2}\right) \chi_{33},$$

$$= D \chi_{33},$$  \[3\]

in which case $\chi_{33}$ becomes the reference susceptibility that gives rise to the observed phase signal and $D$ is the dipole kernel in k-space. As illustrated in figure 1, the assumption $\chi_{13} = \chi_{23} \approx 0$ is not strictly valid, which will be examined further in the Discussion.

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 contestans as inputs for susceptibility mapping algorithms 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 as it is routinely acquired in clinical brain imaging studies and certain QSM algorithms use this image as a prior information input.
- 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 $4^{\text{th}}$-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 ensure that variability in submitted susceptibility maps was primarily due to the inversion algorithm used rather than due to differences in background field removal techniques. However, as single-step QSM methods are designed to simultaneously solve background-field removal and inversion problems, those algorithms could use the unprocessed wrapped phase GRE images.
- A brain mask obtained from Brain Extraction Tool (41) was also provided to reduce confounding effects resulting from the use of different masks.
- The reference susceptibility map $\chi_{33}$ 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 $\chi_{33}$ 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 data set). 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 allowed the contestants to focus on optimizing their algorithmic parameters without spending time writing scripts for 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.

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

***Figure 1 appears near here***
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 field, we employed three additional error measures:

- The high frequency error norm (HFEN) (49)
- The structural similarity index (SSIM) (50)
- The absolute mean error in selected anatomical structures (ROI error). We manually outlined 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, dentate nucleus) in the reference susceptibility map $\chi_{33}$.

These error metrics were calculated for each submitted map, where $\chi$ is expressed in ppm. For RMSE, HFEN and ROI error, smaller values denote better performance, whereas SSIM is normalized between 0 and 1, with 1 being the best achievable result.

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 1 approach (PHILIPS DTV) also utilized the MPRAGE images.

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

Numerical results – “Winners”

Table 2 shows the results of the top ranked algorithms in each evaluation category. The winning QSM reconstructions are also depicted in detail in figure 3.
Winning Approach- RMSE

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

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

\[
\chi^* = \arg\min_{\chi} \|\chi\|_{WTV} + \frac{\mu}{2} \|M\chi - \chi_{well}\|_2^2 \quad [4]
\]

where \( M = F^{-1}(D > \delta)F \) is a sampling matrix taking the value 1 in the well-conditioned region and 0 otherwise according to a threshold \( \delta \) applied to \( |D| \), \( \mu \) is the regularization parameter, \( \|\chi\|_{WTV} = \sum W |\nabla \chi| \) is the weighted anisotropic total variation, and \( W = 1/(|\nabla \chi_{well}| + 10^{-6}) \) is a weighting matrix derived from the gradient (\( \nabla \)) of the well-conditioned susceptibility map \( \chi_{well} \) reconstructed in step 1. The minimization was solved using 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 SFCR2 algorithm proposed by the team from Johns Hopkins University, Maryland, USA, led by Xu Li, was the winner in both HFEN and SSIM categories. The SFCR2 result was obtained by using 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 k-space with two regularization constraints:

\[
\hat{\chi} = \arg\min_{\chi} \lambda_1 \|\text{diag}(M)\chi_k(k) - \text{diag}(M)F\chi\|_2^2 + \|P_{mag}\nabla \chi\|_1 + \lambda_2 \|R\chi\|_2^2 \quad [5]
\]
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 to 1 otherwise). The fidelity mask $R$ for the L2 regularization was generated by combining masks obtained via thresholding a preliminary QSM map $\chi_k(k)$ 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). $M$ is a binary mask indicating the well-conditioned region in k-space, i.e. $M = D > \delta$ where $\delta$ is a threshold on the dipole kernel in k-space. 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:

\[
\chi = \arg\min_{\chi} ||W(\phi - F^{-1}DF\chi)||_2^2 + ||P_{\chi} \nabla \chi||_1 + \gamma_2 ||R\chi||_2^2
\]

where the structural prior $P_{\chi}$ 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 and the same fidelity mask $R$ 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 the team from Southern Medical University, Guangzhou, China led by Yanqiu Feng. 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:

\[
\chi = \arg\min_{\chi} ||W(\phi - F^{-1}DF\chi)||_2^2 + \alpha ||P_{mag} \nabla \chi||_1 + \beta ||(1 - P_{mag}) \nabla \chi||_1
\]

where $W$ is a data weighting matrix to compensate the measured field noise (60), $\phi$ is the measured local field, $\nabla$ is the image domain forward difference operator in three dimensions, and $\alpha$ and $\beta$ are the regularization parameters. The regularization parameters used were $\alpha = 0.003, \beta = 0.0009$.

**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, which was designed to match conventional clinical acquisitions as closely as possible with respect to the resolution, readout bandwidth, echo time and coverage. In the following, we summarize the results, discuss the limitations and highlight the lessons learned from the Challenge.
Summary of results

The SFCR2 algorithm won in two categories, SSIM and HFEN, and finished second in the RMSE ranking. The other winners, in terms of RMSE and ROI accuracy, were the algorithms from UBC and SMU. The top three algorithms in the RMSE ranking relied on compressed sensing (CS). As opposed to regularized inversion where the entire k-space is affected by regularization, in CS approaches only the ill-conditioned k-space regions of the susceptibility map were estimated. This was probably the key feature that allowed these top-ranking approaches to perform well, although these solutions 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 leverage incoherent aliasing artifacts arising from pseudo-random undersampling of k-space (61). The dipole artifacts in QSM reconstruction, however, appear more structured due to undersampling only near the magic angle in k-space. The incoherent aliasing requirement of CS can be met in part by using a wavelet transform, which spreads the dipole artifacts across the wavelet subbands (54).

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 in favor of multi-step approaches in this challenge.

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. However, it was possible for a submission to be optimized for low RMSE by extensive parameter search (e.g. Figure 4). Although this algorithm 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 as measures of high-frequency error and structural
similarity. We thus attempted to create a multi-dimensional performance vector, but we think that finding better metrics is an interesting and open problem.

A major reason why we provided the reference susceptibility map was to ensure that algorithms did not just perform better because the contributing group had spent more time carefully optimizing their processing parameters. With this, we aimed to allow a fair comparison by letting each algorithm perform as well as it could, given a particular pair of input and reference images. However, tuning the regularization parameters to achieve better error metrics led to overly smooth reconstructions in the winning approaches, but notably not in the simpler inversion techniques with scores in the top ten i.e. UCL TIK and UCL TKD, which preserved structural details. This limitation of the current evaluation metrics could be mitigated by incorporating experts’ visual rating of the submitted susceptibility maps. A potential way to amend the RMSE metric could be by comparing the susceptibility map gradient against that in the reference map via \[ \nabla \text{RMSE} = 100 \cdot \frac{\| \nabla (\chi_{33} - \chi_{\text{recon}}) \|_2^2}{\| \nabla \chi_{33} \|_2^2}. \] \( \nabla \text{RMSE} \) would be a more direct measure of the fidelity of high frequency components, and could complement the existing metrics and the visual rating.

**Selection of the reference susceptibility map**

We selected \( \chi_{33} \) instead of the COSMOS solution as standard reference to eliminate the potential orientation bias in the latter susceptibility map. However, the measured local field \( \phi \) suffers from another source of bias, namely the contribution from \( \chi_{13} \) and \( \chi_{23} \) in the transverse plane, as demonstrated in Figure 1. These contributions are clearly non-negligible, as these tensor components have about 70% signal amplitude relative to \( \chi_{33} \).

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 \( \chi_{\text{msa}} = \lambda_1 - (\lambda_2 + \lambda_3)/2 \) where \( \lambda_i \) are the susceptibility tensor eigenvalues. This mask could be refined using white matter segmentation.

In this context, 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). As white matter represents a large brain volume fraction, both 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.

Since the downloadable dataset includes \( \chi_{33} \) and \( \chi_{\text{COSMOS}} \) as well as all components of the susceptibility tensor, future algorithms could report performance metrics relative to any of these. We believe identifying a bias-free gold-standard susceptibility map is an open and important problem in our field.

**Lessons learned for the next QSM reconstruction challenge**

We are fully determined to push forward, improve and extend this research endeavor based on lessons learned from this initial challenge. The feedback from members of the QSM community who attended the
Graz Workshop was very encouraging. 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. These are summarized below:

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. Reference and submitted susceptibility maps could be compared on a per-voxel basis by assessing one-dimensional profiles or correlation coefficients.

III. A better reference map could be found by incorporating the contribution of $\chi_{13}$ and $\chi_{23}$ to the field map in the transverse orientation. A potential way of implementing this could be by rearranging the STI relation in the transverse plane as follows:

$$\Theta(k) = D \chi_{33} - \frac{k_z}{k^2} \left( k_x \chi_{13} + k_y \chi_{23} \right)$$

$$= D \chi_{33} - \frac{k_z^2}{k^2} \chi_{13} + \frac{k_y}{k_z} \chi_{23}$$

[8]

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

$$\Theta(k) = D \chi_{33} - \frac{k_z^2}{k^2} \tilde{\chi} + \frac{1}{3} \tilde{\chi} = D \chi_{33} + \frac{1}{3} \tilde{\chi}$$

[9]

$$\Theta(k) + \frac{1}{3} \tilde{\chi} = D (\chi_{33} + \tilde{\chi})$$

[10]

This suggests that a new ground truth susceptibility could be found by $\chi_{\text{new}} = \chi_{33} + \tilde{\chi}$ and that the input local field data could be amended by $\Theta_{\text{new}} = \Theta + \frac{1}{3} \tilde{\chi}$.

IV. The phase data from all 12 individual orientations could be provided. This data set has already been updated and can be obtained from the webpage.

V. Include the computational efficiency as additional information or category, which would require common access to a single evaluation computer where the processing time of all algorithms is determined.

VI. The challenge could be extended by adding phase data calculated using the forward model (11) from a realistic numerical brain phantom derived from STI or COSMOS susceptibility maps. The data could be made more realistic by adding noise and partial volume artifacts.

VII. The phase calculated from participants’ QSM results using the forward model (11) could be compared against the original measured phase.

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 and cross correlation between 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 (62–64).
Most of these suggestions are easy to implement but some would 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 could facilitate extensions such as an STI challenge, or future research towards computation of a better reference map.

Another interesting avenue to explore could be issuing sub-challenges with clinical data from populations with different diseases. This would be a great 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. We are determined to improve this challenge and hopeful that this will lead to further insight into the design of dipole inversion algorithms.

The substantive differences between the 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. We conclude that direct comparison of results from studies employing different QSM algorithms and parameters is difficult, even when providing a reference susceptibility map.

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 Dr 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, MA, 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 provide only 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.

REFERENCES


28. Wharton S, Schäfer A, Bowtell R. Susceptibility mapping in the human brain using threshold-based...


# TABLES

Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Input Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>TKD (provided)</td>
<td>Truncated K-space Division (28) with zeroes at the ill-conditioned regions (cone) in k-space, threshold = 0.19.</td>
<td>LBV</td>
</tr>
<tr>
<td>CFL2 (provided)</td>
<td>Closed-form L2-regularized inversion (48).</td>
<td>LBV</td>
</tr>
<tr>
<td>MARTINOS WTV</td>
<td>Compressed Sensing compensated QSM (54), accelerated reconstruction using ADMM optimization.</td>
<td>LBV</td>
</tr>
<tr>
<td>GRAZ TGV</td>
<td>Total generalized variation (TGV) based method incorporating phase unwrapping, background field removal and dipole inversion in a single iteration (65).</td>
<td>RAW</td>
</tr>
<tr>
<td>GRAZ TGV L1</td>
<td>Total generalized variation (TGV) based method (65) with additional L1 magnitude stabilization.</td>
<td>RAW</td>
</tr>
<tr>
<td>JENA HEIDI</td>
<td>Hybrid algorithm 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.</td>
<td>LBV</td>
</tr>
<tr>
<td>JENA SDI</td>
<td>TKD algorithm with extreme thresholding of the dipole kernel and underestimation compensation based on the deconvolution point-spread function (66).</td>
<td>LBV</td>
</tr>
<tr>
<td>UCL TKD 1</td>
<td>TKD as in (14,66) i.e. without zeroes inside the k-space cone. A threshold of $\delta = \frac{2}{3}$ was used with no correction for $\chi$ underestimation.</td>
<td>LBV</td>
</tr>
<tr>
<td>UCL TIK</td>
<td>Closed-form Tikhonov inversion as alluded to in (67) and mentioned in (1) as a Tikhonov-regularized minimal norm solution. 1 had $\alpha = 0.0588$ and no correction for $\chi$ underestimation. 2 had $\alpha = 0.0588$ and correction for $\chi$ underestimation with a factor of 1.65. 4 had $\alpha = 0.025$ and correction for $\chi$ underestimation with a factor of 1.30.</td>
<td>LBV</td>
</tr>
<tr>
<td>Institution</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>JHU-XMU SFCR KDN</td>
<td>Based on the 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.</td>
<td></td>
</tr>
<tr>
<td>JHU-XMU SFCR2</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>CHILE TGV L2</td>
<td>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 (68).</td>
<td></td>
</tr>
<tr>
<td>CHILE TGV NL</td>
<td>Non-linear 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.</td>
<td></td>
</tr>
<tr>
<td>CHILE NLD</td>
<td>Discretization of the dipole kernel based on (69). It uses finite differences and the DFT to achieve an analytical solution in the Fourier domain.</td>
<td></td>
</tr>
<tr>
<td>CHILE NLG</td>
<td>Dipole kernel defined in space by the Green’s function, integrating it for each voxel (70).</td>
<td></td>
</tr>
<tr>
<td>CHICAGO TGV</td>
<td>Algorithm based on the TGV QSM method (65), implemented on GPU-hardware (CUDA 7.5, NVIDIA GeForce GTX 980Ti).</td>
<td></td>
</tr>
<tr>
<td>BERKELEY STAR</td>
<td>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 regularization approach (71).</td>
<td></td>
</tr>
<tr>
<td>VANC UBC</td>
<td>LSMR solver (55) followed by weighted compressed sensing minimization – see winning approach in category RMSE.</td>
<td></td>
</tr>
<tr>
<td>IBR ITSWIM</td>
<td>Variable regularization threshold for inverse process / k-space improvement and dual binary mask approach for the iterative algorithm.</td>
<td></td>
</tr>
<tr>
<td>SMU MATV</td>
<td>Morphology-Adaptive Total Variation (MATV) separates target susceptibility into smooth and non-smooth regions, where the latter are</td>
<td></td>
</tr>
</tbody>
</table>
assigned smaller TV weights than smooth regions during dipole inversion (59) – see winning approach in the ROI accuracy category.

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMU MTKD</td>
<td>Truncated k-space division (TKD) with morphological priors (MTKD) LBV (ISMRM 2017, abstract #6095). 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.</td>
</tr>
<tr>
<td>NY MEDI</td>
<td>Morphology Enabled Dipole Inversion (MEDI) method using a Bayesian regularization approach that adds spatial priors from the magnitude image (13,25).</td>
</tr>
<tr>
<td>NY PD</td>
<td>Solving the objective of MEDI using the Primal-Dual (PD) formulation of the total variation and a forward difference method for discretization [[[Accepted in MRM, reference 10.1002/mrm.26627]].</td>
</tr>
<tr>
<td>NY TFI</td>
<td>The Total Field Inversion (TFI) method simultaneously estimates the background and local fields, preventing error propagation from background field removal to QSM (72).</td>
</tr>
<tr>
<td>PHILIPS DTV</td>
<td>Single-step QSM starting from wrapped raw phase using Directional Total-Variation with MPRAGE as a prior for estimating edges (ISMRM 2016, abstract #4051).</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>RMSE (%)</th>
<th>HFEN (%)</th>
<th>SSIM</th>
<th>ROI ERROR (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>69.0</td>
<td>VANC UBC</td>
<td>JHU-XMU SFCR2</td>
<td>SMU MATV</td>
</tr>
<tr>
<td>70.3</td>
<td>JHU-XMU SFCR2</td>
<td>68.8</td>
<td>GRAZ TGV L1</td>
</tr>
<tr>
<td>73.6</td>
<td>MARTINOS WTV</td>
<td>68.9</td>
<td>VANC UBC</td>
</tr>
<tr>
<td>74.2</td>
<td>PHILIPS DTV</td>
<td>70.9</td>
<td>PHILIPS DTV</td>
</tr>
<tr>
<td>74.6</td>
<td>GRAZ TGV L1</td>
<td>71.8</td>
<td>SMU MATV</td>
</tr>
<tr>
<td>75.2</td>
<td>UCL TIK 1</td>
<td>73.1</td>
<td>UCL TIK 1</td>
</tr>
<tr>
<td>76.6</td>
<td>UCL TKD 1</td>
<td>73.6</td>
<td>MARTINOS WTV</td>
</tr>
<tr>
<td></td>
<td>GRAZ TGV</td>
<td>74.1</td>
<td>IBR ITSWIM</td>
</tr>
<tr>
<td>77.5</td>
<td>BERKELEY STAR</td>
<td>JHU-XMU SFCRKDN</td>
<td>0.83</td>
</tr>
<tr>
<td>79.1</td>
<td>SMU MATV</td>
<td>74.2</td>
<td>GRAZ TGV</td>
</tr>
</tbody>
</table>

Table 2 caption: Top 10 algorithms with the best scores in each category evaluated for the QSM reconstruction challenge.
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 ±π radians and the LBV-phase image is scaled from -0.05 to 0.05 radians. With the exception of $\chi_{33}$, the reconstructed susceptibility tensor component images were not provided for the reconstruction challenge, but are now included in the downloadable data set (marked here with asterisks).
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.
Figure 3: Sagittal, coronal and axial slices of QSM reconstructions of the winners in each category: RMSE (UBC), HFEN and SSIM respectively (SFRC2), and ROI error (MATV). QSM images are scaled from -0.1 to 0.25 ppm.
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 $\alpha_0$. 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.

<table>
<thead>
<tr>
<th>$\alpha_0$</th>
<th>0.0005</th>
<th>0.001</th>
<th>0.002</th>
<th>0.004</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>125.8</td>
<td>97.9</td>
<td>81.2</td>
<td>77.5</td>
</tr>
<tr>
<td>HFEN</td>
<td>121.8</td>
<td>94.9</td>
<td>78.2</td>
<td>74.2</td>
</tr>
<tr>
<td>SSIM</td>
<td>91.3</td>
<td>93.5</td>
<td>94.2</td>
<td>92.9</td>
</tr>
<tr>
<td>absErr</td>
<td>0.022</td>
<td>0.020</td>
<td>0.019</td>
<td>0.021</td>
</tr>
</tbody>
</table>