Published in final edited form as:

J Magn Reson Imaging. 2017 November ; 46(5): 1474–1484. doi:10.1002/jmri.25671.

Methods for the computation of templates from quantitative magnetic susceptibility maps (QSM) – toward improved atlas- and voxel-based analyses (VBA)

Jannis Hanspach, BS¹, Michael G. Dwyer, PhD¹, Niels P. Bergsland, PhD^{1,2}, Xiang Feng, MS³, Jesper Hagemeyer, PhD¹, Nicola Bertolino, MS¹, Paul Polak, MS¹, Jürgen R. Reichenbach, PhD^{3,4}, Robert Zivadinov, MD, PhD^{1,5}, and Ferdinand Schweser, PhD^{1,5}

¹Buffalo 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

²Magnetic Resonance Laboratory, IRCCS Don Gnocchi Foundation, Milan, Italy

³Medical Physics Group, Institute of Diagnostic and Interventional Radiology, Jena University Hospital, Friedrich Schiller University Jena, Jena, TH, Germany

⁴Michael Stifel Center for Data-driven and Simulation Science Jena, Friedrich Schiller University Jena, Jena, TH, Germany

⁵MRI Clinical and Translational Research Center, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, The State University of New York, Buffalo, NY, USA

Abstract

Purpose—To develop and assess a method for the creation of templates for VBA and atlas-based approaches using quantitative magnetic susceptibility mapping (QSM).

Material and Methods—We studied four strategies for the creation of magnetic susceptibility brain templates, derived as successive extensions of the conventional template generation (CONV) based on only T₁-weighted (T_{1w}) images. One method that used only T_{1w} images involved a minor improvement of CONV (U-CONV). One method used only magnetic susceptibility maps as input for template generation (DIRECT), and the other two used a linear combination of susceptibility and T_{1w} images (HYBRID) and an algorithm that directly used both image modalities (MULTI), respectively. The strategies were evaluated in a group of $N=10$ healthy human subjects and semi-quantitatively assessed by three experienced raters. Template quality was compared statistically via worth estimates (WEs) obtained with a log-linear Bradley-Terry model.

Results—The overall quality of the templates was better for strategies including both susceptibility and T_{1w} contrast (MULTI: WE=0.62; HYBRID: WE=0.21), but the best method depended on the anatomical region of interest. While methods using only one modality resulted in

lower WEs, lowest overall WEs were obtained when only T₁w images were used (DIRECT: WE=0.12; U-CONV: WE=0.05).

Conclusion—Template generation strategies that employ only magnetic susceptibility contrast or both magnetic susceptibility and T₁w contrast produce templates with the highest quality. The optimal approach depends on the anatomical structures of interest. The established approach of using only T₁w images (CONV) results in reduced image quality compared to all other approaches studied.

Keywords

voxel-based analysis (VBA); template; magnetic resonance imaging (MRI); quantitative susceptibility mapping (QSM); normalization; T₁-weighted imaging

INTRODUCTION

Quantitative magnetic susceptibility mapping (QSM) is a novel magnetic resonance imaging (MRI) technique that provides quantitative information about volume-average magnetic susceptibility of tissues (1, 2). The method is currently considered as the most sensitive imaging biomarker for studying brain iron homeostasis (3, 4), which is suspected to be disturbed in several neurological disorders, such as Alzheimer's disease, Parkinson's disease, and multiple sclerosis (5, 6).

Hitherto, most imaging-based brain iron studies have relied on *region-of-interest* (ROI) based analyses, in which ROIs are defined on every single brain image individually, often employing T₁-weighted (T₁w) images and fully-automated software algorithms. However, the analysis of brain images using these algorithms is subject to three significant limitations: a) results are generally prone to inaccuracies of the automated ROI definition (7), b) they are limited to signal changes in selected white matter regions and a few *major* deep brain nuclei, such as putamen and globus pallidus, and c) image contrast on T₁w images can be weak, depending on myelin and iron concentrations, which can affect the automated algorithms and introduce a group-dependent study bias. Furthermore, the intricate anatomical substructure of basal ganglia nuclei, which is well discernible on susceptibility maps (8), cannot be assessed with current automated ROI-based techniques.

Nonlinearly deforming, or “warping”, brain images into a standardized anatomical space, also referred to as *normalizing*, allows overcoming several limitations of ROI-based methods. For example, it allows a *voxel-based analysis* (VBA) free of *a priori* structural assumptions. Alternatively, a *brain atlas* may be constructed by defining ROIs on the template, which can be transformed to the original images for a custom, operator-independent ROI-analysis. As a target for the deformation operations of the normalization serves a standardized brain image with high contrast of all relevant brain structures. This standardized brain is also called *brain template*.

The normalization of brains is traditionally based on T₁w images and a T₁w brain template because T₁w imaging is part of most established neuroimaging protocols and provides excellent contrast between *cortical* gray and white matter. However, if the primary target of

research is *deep* gray matter (DGM), such as in the more recent brain iron research, the utility of T₁w images is limited because DGM nuclei have a weak contrast relative to surrounding white matter on T₁w images (Figure 1). In particular, similar to a T₁w-based ROI analysis, the poor contrast of DGM on T₁w images may potentially result in improper matching of these structures with detrimental consequences for the VBA or atlas-based analysis of susceptibility maps. This limitation may be particularly relevant for the study of small subnuclear regions with strong susceptibility contrast but weak contrast on T₁w images, such as the internal structures of the globus pallidus (9). In general, normalization procedures should exploit the intricate anatomical contrast of DGM available on susceptibility maps. However, dedicated methods for the creation of magnetic susceptibility templates have not been introduced or systematically compared. In the present work, we aimed to present and assess different methods for the creation of magnetic susceptibility templates.

MATERIALS AND METHODS

Template Creation Strategies

We studied four advanced methods for the creation of QSM brain templates, derived as successive extensions of the conventional template generation based on T₁w images (CONV). In CONV, T₁w images were used to create a brain template that represents image features from all individual input T₁w images. We obtained a *susceptibility* template with this approach by, first, registering susceptibility maps to T₁w images, then using the deformation transforms of the T₁w-images resulting from the T₁w-template creation process to normalize the susceptibility maps, and finally averaging all deformed susceptibility maps. The four advanced methods are described in the following (see also Table 1):

- i. Conventional approach with up-sampling (U-CONV): T₁w images may have a different in-plane resolution than susceptibility maps (here, reconstructed $1 \times 1 \text{ mm}^2$ vs. $0.5 \times 0.5 \text{ mm}^2$), depending on the pulse sequence parameters employed. To assess the impact of a reduced spatial in-plane resolution of the T₁w template compared to the susceptibility maps, U-CONV up-sampled the T₁w images to the voxel size of the susceptibility maps before applying method CONV.
- ii. Direct approach (DIRECT): Magnetic susceptibility maps provide high contrast in several anatomical regions where contrast can be low or even absent on T₁w images (Figure 1), depending on pulse sequence parameters. The T₁w-imaging based approaches (CONV and U-CONV) do not exploit this improved contrast on susceptibility maps for the calculation of the deformation transforms. DIRECT aimed to improve the template quality by applying the template generation algorithm directly to susceptibility maps instead of T₁w images.
- iii. T₁w/QSM hybrid approach (HYBRID): The HYBRID approach combined T₁w images with susceptibility maps into a hybrid image contrast to exploit the image information of both modalities (10). This method aimed to improve the deformation transform calculation compared to the direct approach (DIRECT). We obtained the final susceptibility templates by applying the calculated

deformation transforms to the original susceptibility maps, followed by an averaging of the deformed maps.

- iv. Multivariate approach (MULTI): Different from the HYBRID approach, multivariate template generation algorithms directly utilize for the template generation multiple *individual* images of the same subjects, preferably with image contrast that has a complementary information content (11). These algorithms minimize an objective function that depends on both image contrasts rather than only on one image contrast. Using individual T_1w images and susceptibility maps for the template generation is expected to result in a better exploitation of the complementary anatomical contrast compared to the HYBRID approach because each image contrast is fully accounted for by a separate term in the optimization function.

Subjects, Data Acquisition, and Reconstruction

We evaluated all four advanced strategies and the traditional strategy in a group of $N=10$ healthy human subjects (5 female, 5 male). The subjects were randomly picked from a database of research MRI exams at our institution. Selection criteria included that both T_1w images and susceptibility maps had been acquired in the same imaging session and subjects had had a normal neurological examination and no history of neurologic disorders or chronic psychiatric disorders. The subjects were chosen from a sub-cohort with a narrow age range of 50 ± 1 years (range 49 to 51 years) to minimize variations of brain morphology and image contrast due to age-related atrophy and age-related iron-load in the DGM (12). We assumed that a highly homogeneous group would yield a template with optimal delineation of susceptibility variations characteristic for the mean age of the group, allowing studying algorithm-related quality differences in a relatively controlled setting. We chose a cohort size of $N=10$ because it had been suggested that template generation stabilizes at around ten subjects (13), which means that subjects from the same demographic cohort deliver a similar template and the addition of more subjects does not improve the template generation.

MRI was performed on a 3 Tesla MRI system (Signa Excite HD 12.0; GE Healthcare, Milwaukee, Wisconsin) with an eight-channel head-and-neck coil. For QSM we used a clinical un-accelerated 3D single-echo spoiled GRE sequence with flow compensation in read and slice directions, $256\times 192\times 128\text{mm}^3$ field-of-view (FOV), $512\times 192\times 64$ matrix, echo time (TE)/repetition time (TR) 22ms/40ms, bandwidth (BW) 13.89kHz, 12° flip angle, and an acquisition time of 8:46 min:sec. Parallel imaging acceleration of the GRE sequence and phase image reconstruction were not supported by the scanner platform. Hence, we saved the raw k-space data for offline reconstruction of susceptibility maps. T_1w images were acquired using a 3D inversion recovery magnetization-prepared fast spoiled gradient echo (IR-FSPGR) sequence with TE/inversion time (TI)/TR=2.8ms/900ms/5.9ms, flip angle= 10° , $256\times 192\times 184$ matrix, and an FOV of $256\times 192\times 184\text{mm}^3$, resulting in 1mm isotropic resolution at an acquisition time of 9:18 min:sec. All scans were prescribed parallel to the sub-callosal line in an axial-oblique orientation, and no averaging was performed. Written informed consent was obtained from all the participants, and the study protocols were approved by the local institutional review board.

Complex-valued GRE images for QSM were reconstructed offline from the raw k-space data on a $512 \times 512 \times 64$ spatial matrix using sum-of-squares for magnitude images and scalar phase matching for phase images (14). The k-space data were zero-padded in the phase-encoding direction before the processing to achieve an isotropic in-plane resolution. Distortions due to imaging gradient non-linearity were compensated for as described by Polak et al. (15). Phase images were unwrapped with a best-path algorithm (16), background-field corrected with V-SHARP (3, 17) (maximum radius 5mm; truncated singular value decomposition (TSVD) threshold 0.05), and converted to magnetic susceptibility maps using *homogeneity enabled incremental dipole inversion* (HEIDI) (18). HEIDI employs an incremental inversion process that reconstructs Fourier components of the susceptibility distribution close to the Null-space of the forward problem based on prior knowledge about the distribution extracted from magnitude and phase images. All other Fourier components are reconstructed without explicit regularization.

Image Pre-processing

The image pre-processing steps are schematically illustrated in Figure 2 and described in the following:

T₁-weighted images—To reduce B₁-related intensity inhomogeneities, we applied a nonparametric non-uniform B-spline-based intensity normalization (19) to all T₁w images (N4-ITK). Variations of the image intensity ranges of the T₁w images between subjects were homogenized to a nominal range of 0 to 4095 (using 32-bit single-precision floating-point-format) by applying a decile-based piecewise linear intensity normalization algorithm (20). To focus the template generation on brain structures instead of skull, neck, or muscles, we skull-stripped the pre-processed T₁w images before applying the template building algorithm (FSL BET (21); default parameters). For method U-CONV, we resampled the T₁w images to the voxel size of the susceptibility maps before template construction and image pre-processing using linear interpolation (mri_convert; FreeSurfer 5.3.0, Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA) (blue arrows in Figure 2).

Susceptibility maps and GRE magnitude images—Atlas generation algorithms have traditionally been optimized for T₁w images, where image intensities are positive and nominal intensity values range in the hundreds to thousands. However, the dynamic range of voxel values on susceptibility maps is very different from this, with both negative and positive image intensities in the order of parts-per-million (ppm). When unaltered susceptibility maps were processed with method DIRECT they showed severe deformations, were translated to the bottom or outside of the FOV, were massively rotated, or the size of the brain was significantly altered. With the MULTI approach, the T₁w template generation was successful, but the QSM template looked like an original single subject map normalized to the T₁w template.

To overcome malfunctions associated with the unconventional image intensity values, we rescaled the susceptibility maps before applying the DIRECT and MULTI approaches such that the voxel values spanned a range typical for T₁w imaging. A similar strategy has recently been employed by Visser et al. (22). For methods DIRECT and MULTI we linearly

rescaled each voxel value χ of the susceptibility maps according to the following equation: $\rho = |\alpha \chi + \beta|$ where ρ is the rescaled voxel value and the coefficients α and β were chosen such that the typical intensity range of the susceptibility maps (-0.1 to $+0.2$ ppm) was mapped to an intensity range that was similar to that of the intensity-normalized T_1w images. Taking the absolute value prevented the presence of any voxels with negative values. For our intensity-normalized T_1w images, a histogram analysis showed that 99% of the voxels of our T_1w images had intensities between 250 and 4000, motivating an empirical definition of the coefficients as $\alpha=12500 \text{ ppm}^{-1}$ and $\beta=1500$. Note that the exact values of these parameters depend on the normalization of the T_1w images and sequence parameters, and are not generally valid.

We reduced B_1 -related signal inhomogeneities in GRE magnitude images by applying the same bias field correction as for the T_1w images. We did not bias-correct the susceptibility maps because they represent a physical property and did not suffer from slowly varying inhomogeneities.

Intra-subject transformations—For strategies MULTI and HYBRID, T_1w images had to be registered to the susceptibility maps. To this end, we first determined the linear transformations reflecting the different imaging slab prescriptions of the two sequences (tkregister; FreeSurfer 5.3.0). Subsequently, to compensate for inter-scan head motion, we performed a rigid-body registration of the transformed T_1w images to the GRE magnitude images with the affine transformation as an initializing transform (Advanced Normalization Tools [ANTs]; version 2.0.0; <http://stnava.github.io/ANTs>). We applied the computed transforms to all T_1w images resulting in a set of T_1w images co-registered to the respective susceptibility maps. Conversion of transformation matrices between the coordinate systems of the different software packages was achieved using c3d (ITK-SNAP; www.itksnap.org).

HYBRID images—We combined image contrast of original (unscaled) susceptibility maps and co-registered T_1w images for the HYBRID approach according to the following formula: $\text{HYBRID} = |T_1w + a \cdot \text{QSM}|$. To systematically study the effect of different values of a and to determine the optimal value, we created eight sets of hybrid images with parameter a ranging from -15000 to -1000 ppm^{-1} in steps of 2000 ppm^{-1} . We chose negative values for a because they converted the hyper-intense DGM contrast on susceptibility maps into hypo-intense contrast on HYBRID images, similar to typical T_1w image contrast. The range of studied values of a was justified as follows: The highest susceptibility values in the basal ganglia are typically around $+0.2$ ppm, whereas the image intensities in corresponding regions on our (intensity-normalized) T_1w images was around 3000. Hence, for $a=-15000 \text{ ppm}^{-1}$ the hybrid image contrast in these regions equaled zero. Lower values of a would have resulted in reduced intra-nuclear contrast due to saturation effects. Figure 3 shows hybrid images created with different parameters a , demonstrating the varying degree of DGM contrast. Contrast in iron-laden DGM regions was weak for $a=0$ and gradually increased with decreasing a .

Template Generation Strategies

For methods CONV, U-CONV, DIRECT, and HYBRID we used the diffeomorphic greedy symmetric normalization (Greedy-SyN) transformation model to build the templates (buildtemplateparallel, ANTs; deactivated bias field correction). In each iteration step, this method registers all brain images to an intermediate template obtained by averaging the registered brain images of the previous iteration step. For the MULTI approach, we used SyN for multivariate neuroanatomy (SyNMN) (23) (antsMultivariateTemplateConstruction, ANTs) with rescaled susceptibility maps (analog DIRECT) and co-registered T₁w images as input and otherwise default parameters. This algorithm uses a symmetric diffeomorphic registration algorithm, which minimizes an objective function with a similarity criterion that depends on both image contrasts rather than only on one image contrast. The effect of the modality-weighting factor in the similarity metric of SyNMN was studied by executing the algorithm with the following different weights (QSM×T₁w): 1×0, 1×0.5, 1×1, 0.5×1, 0×1. Modality weights of 1×0 and 0×1 mean that only the susceptibility map and only the T₁w images are used for the template generation, respectively. In particular, the use of 1×0 is similar to approach DIRECT, whereas 0×1 is similar to approach U-CONV.

In all cases, templates were created by an iterative process. First, a simple mean of uncoregistered images was taken, and all images were then rigid-body-aligned to the mean to produce an improved rough template (initial template). This was iterated four times with affine registration and using cross-correlation similarity metric, and only one iteration on a four times under-sampled grid. Next, the resulting affine template was used as the initial target for a set of non-linear iterations. Interim results of the algorithms (at different iteration steps) were visually checked for transformation failures, and the particular strategy was excluded from subsequent analyses if such errors occurred.

For all strategies, we obtained the susceptibility template by, first, applying the non-linear template-generation transformations to the native susceptibility maps and, then, averaging the image intensities of all transformed susceptibility maps. For strategies CONV and U-CONV, we, first, affine-registered the GRE magnitude images to the respective T₁w images using ANTs and then used the resulting registration matrices as initialization transforms for the non-linear deformation of susceptibility maps to the T₁w template space. To improve comparability of the CONV template with the other templates, we linearly resampled the template to the voxel size of the susceptibility maps (mri_convert; FreeSurfer).

All computations were carried out on a Linux workstation (Ubuntu 12.04) with 48 cores (dual Intel Xenon E5–2697v2 CPUs at 2.7 GHz) and 396 GB RAM. Eight cores were dedicated to the present study and the computation time of the complete processing pipeline was roughly 10 hours for each method.

Analysis

It is difficult to quantitatively evaluate and compare template creation strategies using objective image comparison metrics, because a ground-truth does not exist. To the best of our knowledge, there is no reliable method to quantify template quality. Hence, we decided to perform a visual rating and derive a rank order of quality. Three raters (FS/PP/NB) with

several years of experience in both general clinical neuroimaging and QSM (11, 9, and five years, respectively) were asked to rate the visual similarity of image contrast on the templates to that on the ten original single-subject (input) susceptibility maps. To this end, we presented raters the ten original input susceptibility maps immediately before the assessment of the templates. Raters were blinded to the experimental hypothesis, the template creation technique, and parameters used. Each template was compared against all other templates in pairs and raters had the ability to toggle between the two templates of each pair. One of the two presented templates had to be identified as better than the other.

To account for the different contrast of anatomical regions on susceptibility maps and T₁w images, we carried out four separate ratings that focused on various anatomical regions: basal ganglia, thalamus, motor cortex, and venous vasculature. For each region, the rating was based on the same two-dimensional slices optimally showing the region of interest. Figure 4 shows the four regions and representative slices of the rating process. As a first step, we assessed the quality of HYBRID and MULTI templates created with different hybrid factors and modality weights, respectively. The parameters yielding the best templates for the particular strategy and anatomical region were then compared to templates obtained with the other approaches in separate ratings.

Statistics

To convert from pairwise ratings to a specific interval ranking scale when evaluating the raters' data, we used a log-linear Bradley-Terry model, a generalized linear model. Briefly, individual worth estimates (WEs) were derived by using the *prefmod* R-package (24) for R (25). The WEs π_j and π_k define the probability p to prefer an object j over an object k :

$$p(j > k) | \pi_j, \pi_k = \frac{\pi_j}{\pi_j + \pi_k}.$$
 Plotting the WEs provides an elegant way to visualize a ranking of the competing items (26).

Inter-rater reliability was estimated by calculating *Cohen's Kappa coefficient* for all pairs of raters and pair-wise ratings. Coefficients below 0.4 were considered as poor consistency, coefficients between 0.4 and 0.75 as fair to good, and over 0.75 as excellent (27).

RESULTS

Parameter Optimization for HYBRID and MULTI

The preselection of combination coefficients for the HYBRID approach resulted in the following four coefficients that were then compared to one another: -9000, -11000, -13000, -15000, in ppm⁻¹ respectively. Optimal WEs depended on the region, with -11000 ppm⁻¹ being the best coefficient in basal ganglia, -13000 ppm⁻¹ in the thalamus and motor cortex, and -15000 ppm⁻¹ and -11000 ppm⁻¹ (same WEs) in the veins. MULTI yielded the best templates when weights were similar (either 1 or 0.5) for both modalities (best in basal ganglia: 1×0.5; veins: 1×0.5; motor cortex: 1×1). In the thalamus, the best template was obtained for weight 1×0, which means that only the susceptibility maps were used for the template generation, similar to strategy DIRECT. The weight 0×1, similar to approach U-CONV, yielded the lowest WE in all regions. The results of the log-linear Bradley-Terry model with different combination coefficients (HYBRID) and modality weights (MULTI)

are visualized in Figure 5. Figure 6 shows exemplary slices of MULTI templates with different modality weights in the regions of basal ganglia (a–e) and motor cortex (f–j). Differences between the templates, such as contrast, sharpness or delineation of anatomical structures, were subtle but clearly discernible upon closer side-by-side inspection.

Comparison of Template Generation Strategies

Total and individual WEs of the four regions' WEs of the different approaches are illustrated in Figure 7. DIRECT received the highest WE for the basal ganglia and thalamus, followed by MULTI. In veins and motor cortex HYBRID and MULTI received the highest WEs. CONV and U-CONV had consistently low WEs in all regions. Figure 8 shows all templates generated by the different strategies and in the various brain regions.

Statistical Analysis Of The Template Comparisons

Table 2 lists *Cohen's kappa coefficients* for the inter-rater agreement of the different ratings. Consistency was poor or fair (kappa below 0.4 or slightly higher than 0.4) for all preselection ratings, reflecting a low impact of the parameter choices on the resulting templates. Excellent (kappa > 0.75) or good (just below 0.75) consistency was observed for the final rating, indicating substantial differences between the individual templates that were reproducibly identified by different raters.

DISCUSSION

To the best of our knowledge, the present study is the first published attempt to create and evaluate susceptibility-based brain templates. While QSM is a relatively recent technique, practical issues with negative intensities may explain this observation when the algorithms that are currently available for template generation are directly applied to susceptibility maps without prior rescaling. We found that rescaling of susceptibility values before template generation is crucial for the current ANTs algorithms. Most image processing and analysis algorithms have originally been developed for standard (magnitude) images and have difficulties handling susceptibility maps that have image intensities equally distributed around zero.

While the CONV method consistently resulted in the lowest quality WEs in our study, all approaches led to relatively comparable templates with acceptable quality. Generally, it could be observed that regions with high contrast on susceptibility maps and low contrast on T₁w images (basal ganglia, thalamus) benefited from the incorporation of the susceptibility contrast into the template creation, whereas regions with low contrast on susceptibility maps and high contrast on T₁w images (motor cortex) benefited from T₁w image contrast. Upon detailed inspection, we found that the multivariate approach (MULTI) yielded a consistently high quality in all four anatomical regions analyzed with high ranks of the total log-linear Bradley-Terry model WEs. Although slightly better template quality was obtained in specific anatomical areas with the HYBRID approach (veins) and with the DIRECT approach (basal ganglia and thalamus), this leads to the conclusion that MULTI may be considered as a “one-fits-all” approach, if T₁w images are available.

A particular advantage of the DIRECT method over the other methods is that it does not rely on T₁w images. This reduces the complexity of the analysis workflow and enables a template-based analysis in settings in which T₁w images are not available, such as in the routine clinical setting. If T₁w images are available, further improvement of MULTI and HYBRID may be achieved by including additional MRI contrasts and employing advanced classification methods, such as support vector machines, instead of a simple weighted averaging, respectively. This will be subject to future research. In particular, additional contrasts could be derived from GRE imaging, such as R₂* maps (not available in this study), susceptibility weighted images (SWI), the GRE magnitude, or phase images. However, it should be considered that these images often show considerably weaker image contrast than susceptibility maps and, in particular, compared to susceptibility maps, SWI and phase images suffer from orientation dependence, which may limit their benefit for normalization. The inclusion of non-GRE image information, e.g. diffusion data, also might improve the template generation for MULTI and HYBRID.

The inferiority of the CONV and U-CONV approach compared to the susceptibility-based approaches in the present study illustrated the limitations of using images with reduced susceptibility contrast as anatomical prior for the analysis of susceptibility maps. The reduced quality stems from only using T₁w information, since regions with high susceptibility contrast are not taken into account in the template generation, and thus the averaging process reduces the susceptibility contrast.

Several groups have recently applied such approaches for the analysis of susceptibility maps, which will benefit from the incorporation of high-quality susceptibility templates. Rudko et al. (28), Acosta-Caronero et al. (29), Darki et al. (30), Hinoda et al. (31), Li et al. (32), and Lauzon et al. (33) employed a method similar to CONV by registering susceptibility maps to T₁-weighted (T₁w) images and using standard brain templates created from T₁w images for the normalization step. Carpenter et al. (34) used a study-specific template created from pediatric T₁w images. Goubran et al. (35) created group-specific templates from GRE magnitude images. Cobzas et al. (36) and Lim et al. (37) employed (segmented) single-subject images as registration targets. Lim et al. proposed to use the GRE magnitude of a single selected subject as the template and propagated atlas-labels from the subject's susceptibility map. Cobzas et al. (36) employed a multi-atlas alignment strategy which used images of four individuals as templates representing particular subgroups of the study cohort (different ages in that case). The use of susceptibility templates generated from a large group of subjects rather than single-subject maps would reduce the theoretical bias in these studies due to image features specific to the selected subjects.

In this study, we relied on a subject group with particularly narrow age distribution. The use of a homogeneous subject group in this study was motivated by a reduced effect of inter-subject contrast variations and an improved ability to assess the effect of different template generation strategies. It can be expected that wider age-range results in diminishing clarity of anatomical structures on the template because the susceptibility maps used for the normalization are more different. To test this hypothesis, we applied all algorithms also to a set of images from a group of subjects with a wider age range ($N=10$ healthy human subjects; 5 female, 5 male; 19 to 76 years). We obtained a comparable template quality and

could qualitatively confirm the generalizability of the key results in the present work (data not shown). The subject group consisted of only healthy subjects. Another interesting arena for future research is the study of pathology conspicuity when templates are created from patient data. For example, it may be possible to study changes in iron distribution patterns between study groups by comparing templates created individually from the groups. It is expected that our methods are easily applicable to patient data. However, this is subject of future work and needs to be validated.

We employed different publicly available software toolboxes for pre-processing and transforming images (FSL, ITK, FreeSurfer, and ANTs). This mix of tools may appear to complicate the re-implementation of the presented steps. However, despite the template generation algorithms, all methods (such as bias-field correction) may be replaced by similar algorithms as long as the algorithms correctly perform the required operation. There is no reason to assume that the use of alternative methods have a significant effect on the generated template.

Our study has several limitations. A visual rating, even by experienced raters, is not objective. In particular, the criteria for a higher template quality are difficult to define. Our raters relied on their experience with susceptibility contrast. We tried to mitigate the subjectivity of this approach by using multiple blinded raters and allowing a side-by-side comparison of all techniques with one another. Another limitation is that the registration algorithms based on diffeomorphic transformations could yield to inaccurate representations of the anatomic regions. Such inaccurate representations are difficult to detect. However, the most frequent use of templates is as a registration target. For this use, it is most important that the template facilitates a proper normalization of the brain images, whereas anatomically incorrect representation of anatomical regions is only a minor issue. Furthermore, we only carried out four ratings on various anatomical regions: basal ganglia, thalamus, motor cortex and venous vasculature and did not look at more regions. Other anatomical regions, such as the substantia nigra or red nuclei, have also been studied with QSM and could have been rated and analyzed as well in order to improve the validity of the analysis. Nevertheless, we considered the four chosen regions as representative for typical contrast of brain regions on T₁w images and susceptibility maps.

In conclusion, the highest quality templates of magnetic susceptibility in this study were obtained with approaches that either relied on only susceptibility contrast (DIRECT) or exploited both susceptibility and T₁w contrast (HYBRID and MULTI). The optimal approach depended on the anatomical structures of interest. The conventional approach using a T₁w template (CONV, U-CONV) resulted in significantly reduced image quality compared to all other approaches studied. Our study provides a foundation for future research and clinical application of QSM in large-cohort studies to explore changes in iron homeostasis in normal aging and diseases of the human brain.

Acknowledgments

Grant Support

Research reported in this publication was funded by a RISE (Research Internships in Science and Engineering) worldwide stipend of the German Academic Exchange Service (DAAD) awarded to Jannis Hanspach and 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. Support is also acknowledged from the bilateral PPP – USA 2015/2016 program (Project 57131995 – Quantitative MRI for Neurological Studies) of the German Academic Exchange Service (DAAD) awarded to Jürgen R. Reichenbach.

Abbreviation Key

QSM	quantitative susceptibility mapping
ROI	region-of-interest
VBA	voxel-based analysis
T₁w	T ₁ -weighted
DGM	deep gray matter
CONV	conventional approach
U-CONV	conventional approach with up-sampling
DIRECT	direct approach with rescaling
ppm	parts-per-million
HYBRID	T ₁ w/QSM hybrid approach
MULTI	Multivariate approach with rescaling
WE	Worth Estimate
GRE	gradient recalled echo
FOV	field-of-view
TE	echo time
TR	repetition time
IR-FSPGR	3D inversion recovery magnetization-prepared fast spoiled gradient echo
TI	inversion time
BW	bandwidth
TSVD	truncated singular value decomposition
HEIDI	homogeneity enabled incremental dipole inversion
ANTs	Advanced Normalization Tools

LITERATURE REFERENCES

1. Haacke EM, Liu S, Buch S, Zheng W, Wu D, Ye Y. Quantitative susceptibility mapping: current status and future directions. *Magn Reson Imaging*. 2015; 33:1–25. [PubMed: 25267705]
2. Liu C, Wei H, Gong N-J, Cronin M, Dibb R, Decker K. Quantitative Susceptibility Mapping: Contrast Mechanisms and Clinical Applications. *Tomogr J Imaging Res*. 2015; 1:3–17.
3. Schweser F, Deistung A, Lehr BW, Reichenbach JR. Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: An approach to in vivo brain iron metabolism? *NeuroImage*. 2011; 54:2789–2807. [PubMed: 21040794]
4. 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.
5. Berg D, Youdim MBH. Role of iron in neurodegenerative disorders. *Top Magn Reson Imaging TMRI*. 2006; 17:5–17. [PubMed: 17179893]
6. Ropele S, Kiladonk ID, Wattjes MP, et al. Determinants of iron accumulation in deep grey matter of multiple sclerosis patients. *Mult Scler J*. 2014; 20:1692–1698.
7. Derakhshan M, Caramanos Z, Giacomini PS, et al. Evaluation of automated techniques for the quantification of grey matter atrophy in patients with multiple sclerosis. *NeuroImage*. 2010; 52:1261–1267. [PubMed: 20483380]
8. Deistung A, Schäfer A, Schweser F, Biedermann U, Turner R, Reichenbach JR. Toward in vivo histology: A comparison of quantitative susceptibility mapping (QSM) with magnitude-, phase-, and R2*-imaging at ultra-high magnetic field strength. *NeuroImage*. 2013; 65:299–314. [PubMed: 23036448]
9. Ide S, Kakeda S, Ueda I, et al. Internal structures of the globus pallidus in patients with Parkinson's disease: evaluation with quantitative susceptibility mapping (QSM). *Eur Radiol*. 2015; 25:710–718. [PubMed: 25361824]
10. Schweser F, Feng X, Battle R, et al. Improved deep gray matter segmentation using anatomical information from quantitative susceptibility maps. *Milan*. 2014; 22:1787. 22.
11. Avants B, Duda JT, Kim J, et al. Multivariate Analysis of Structural and Diffusion Imaging in Traumatic Brain Injury. *Acad Radiol*. 2008; 15:1360–1375. [PubMed: 18995188]
12. Hallgren B, Sourander P. The effect of age on the non-haemin iron in the human brain. *J Neurochem*. 1958; 3:41–51. [PubMed: 13611557]
13. Avants BB, Tustison N, Song G. Advanced normalization tools (ANTS). *Insight J*. 2009; 2:1–35.
14. Hammond KE, Lupo JM, Xu D, et al. Development of a robust method for generating 7.0 T multichannel phase images of the brain with application to normal volunteers and patients with neurological diseases. *NeuroImage*. 2008; 39:1682–1692. [PubMed: 18096412]
15. Polak P, Zivadinov R, Schweser F. Gradient Unwarping for Phase Imaging Reconstruction. In *Proc Intl Soc Mag Reson Med*. 2015:23.
16. Abdul-Rahman HS, Gdeisat MA, Burton DR, Lalor MJ, Lilley F, Moore CJ. Fast and robust three-dimensional best path phase unwrapping algorithm. *Appl Opt*. 2007; 46:6623–6635. [PubMed: 17846656]
17. Wu B, Li W, Guidon A, Liu C. Whole brain susceptibility mapping using compressed sensing. *Magn Reson Med*. 2012; 67:137–147. [PubMed: 21671269]
18. Schweser F, Sommer K, Deistung A, Reichenbach JR. Quantitative susceptibility mapping for investigating subtle susceptibility variations in the human brain. *NeuroImage*. 2012; 62:2083–2100. [PubMed: 22659482]
19. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imaging*. 2010; 29:1310–1320. [PubMed: 20378467]
20. Nyul LG, Udupa JK, Xuan Zhang. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging*. 2000; 19:143–150. [PubMed: 10784285]
21. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp*. 2002; 17:143–155. [PubMed: 12391568]

22. Visser E, Keuken MC, Douaud G, et al. Automatic segmentation of the striatum and globus pallidus using MIST: Multimodal Image Segmentation Tool. *NeuroImage*. 2016; 125:479–497. [PubMed: 26477650]
23. Avants BB, Duda JT, Zhang H, Gee JC. Multivariate normalization with symmetric diffeomorphisms for multivariate studies. *Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv*. 2007; 10(Pt 1):359–366.
24. Hatzinger R, Dittrich R. Prefmod: An R package for modeling preferences based on paired comparisons, rankings, or ratings. *J Stat Softw*. 2012; 48:1–31.
25. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013.
26. Fienberg SE, Larntz K. Log linear representation for paired and multiple comparisons models. *Biometrika*. 1976; 63:245–254.
27. Fleiss, JL., Levin, B., Paik, MC. *Statistical Methods for Rates and Proportions*. 2nd. New York: John Wiley & Sons; 1981.
28. Rudko DA, Solovey I, Gati JS, Kremenutzky M, Menon RS. Multiple Sclerosis: Improved Identification of Disease-relevant Changes in Gray and White Matter by Using Susceptibility-based MR Imaging. *Radiology*. 2014; 272:851–864. [PubMed: 24828000]
29. Acosta-Cabronero J, Betts MJ, Cardenas-Blanco A, Yang S, Nestor PJ. In Vivo MRI Mapping of Brain Iron Deposition across the Adult Lifespan. *J Neurosci*. 2016; 36:364–374. [PubMed: 26758829]
30. Darki F, Nemmi F, Möller A, Sitnikov R, Klingberg T. Quantitative susceptibility mapping of striatum in children and adults, and its association with working memory performance. *NeuroImage*. 2016; 136:208–214. [PubMed: 27132546]
31. Hinoda T, Fushimi Y, Okada T, et al. Quantitative Susceptibility Mapping at 3 T and 1.5 T: Evaluation of Consistency and Reproducibility. *Invest Radiol*. 2015; 50:522–530. [PubMed: 25900085]
32. Li W, Langkammer C, Chou Y-H, et al. Association between increased magnetic susceptibility of deep gray matter nuclei and decreased motor function in healthy adults. *NeuroImage*. 2015; 105:45–52. [PubMed: 25315786]
33. Lauzon ML, McCreary CR, McLean DA, Salluzzi M, Frayne R. Quantitative susceptibility mapping at 3 T: comparison of acquisition methodologies. *NMR Biomed*. 2016
34. Carpenter KLH, Li W, Wei H, et al. Magnetic susceptibility of brain iron is associated with childhood spatial IQ. *NeuroImage*. 2016; 132:167–174. [PubMed: 26899787]
35. Goubran M, Rudko DA, Santyr B, et al. In vivo normative atlas of the hippocampal subfields using multi-echo susceptibility imaging at 7 Tesla: Susceptibility Atlas of the Hippocampus at 7 T. *Hum Brain Mapp*. 2014; 35:3588–3601. [PubMed: 24339427]
36. Cobzas D, Sun H, Walsh AJ, Lebel RM, Blevins G, Wilman AH. Subcortical gray matter segmentation and voxel-based analysis using transverse relaxation and quantitative susceptibility mapping with application to multiple sclerosis: Voxel-based QSM and R_2^* in MS. *J Magn Reson Imaging*. 2015; 42:1601–1610. [PubMed: 25980643]
37. Lim IAL, Faria AV, Li X, et al. Human brain atlas for automated region of interest selection in quantitative susceptibility mapping: Application to determine iron content in deep gray matter structures. *NeuroImage*. 2013; 82:449–469. [PubMed: 23769915]

Figure 1.

Illustration of typical contrast differences between T_1w images (a–d) and susceptibility maps (e–h) at the anatomical levels that were used for the assessment of the template quality.

Columns show, from left to right, axial slices of the inferior part of the basal ganglia, the medial part of the basal ganglia and thalamus (weak contrast of thalamus, pulvinar, and globus pallidus), veins (invisible on T_1w images; weak contrast of posterior horn of putamen), and motor cortex (inhomogeneous contrast between white and gray matter on susceptibility maps). Note that the contrast on T_1w images depends on the chosen sequence parameters. The images shown here were acquired with the sequence parameters listed in the Methods section.

Figure 2. Schematic illustration of the pre-processing pipeline. Images used as input for the different template generation approaches are highlighted by red boxes. Blue arrows describe the process for U-CONV.

Figure 3.

HYBRID images of an exemplary subject created with different combination coefficients a in ppm^{-1} . The image contrast ranges from 0 to 3900 (black to white).

Figure 4.

The four brain regions examined in this study indicated as red outlines on an original (unregistered) susceptibility map of one representative subject (a–d). The corresponding images used in the template rating process (e–l). The parameters used to generate the atlases in each column (e–l; from left to right) were: 1. DIRECT/HYBRID (-11000 ppm^{-1}), 2. HYBRID (-13000 ppm^{-1})/MULTI (1×0), 3. CONV/DIRECT, 4. MULTI(1×1)/UCONV.

Figure 5. Log-linear Bradley-Terry ranking of templates generated with different HYBRID combination coefficients, a in ppm^{-1} (a); Ranking of templates obtained with the MULTI approach using different modality weights (b).

Figure 6.
MULTI templates created with various modality weights (QSM×T_{1w}). Shown are slices at the level of the basal ganglia (a–e) and the motor cortex (f–j). The image contrast ranges from -0.1 to 0.2 ppm (black to white). Green arrows point to regions with a high (preferable) contrast, whereas red arrows point to areas with low or unclear contrast in the same area.

Figure 7.

Log-linear Bradley-Terry ranking of all template generation strategies in all four brain regions (a) and in the respective brain region (b). For HYBRID the following values for a (in ppm^{-1}) were used: -11000 , -13000 , -11000 , -13000 (from left to right). A higher ranking indicates better template quality.

Figure 8.

Slices of the templates generated by the different strategies in the evaluated anatomical regions. Green arrows point to regions with a high (preferable) contrast, whereas red arrows point to regions with low or unclear contrast in the same region. The image contrast ranges from -0.1 to 0.2 ppm (black to white). For HYBRID the following parameters for a in ppm^{-1} were used: -11000 , -13000 , -11000 , -13000 and for MULTI the following modality

weights were used: 1×0.5 , 1×0 , 1×0.5 , 1×1 (from basal ganglia to motor cortex, respectively).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Brief description of the presented methods to generate a susceptibility template.

Method	Description
CONV	Conventional approach based on T ₁ w images. A susceptibility template was obtained by normalizing the susceptibility maps to a T ₁ w template followed by averaging of all normalized susceptibility maps.
U-CONV	Equivalent to CONV except for up-sampling the T ₁ w images to the same resolution as the magnetic susceptibility maps before generating the T ₁ w template.
DIRECT	Susceptibility maps were directly used as input for the susceptibility template generation.
HYBRID	T ₁ w images and QSM were linearly combined to a hybrid image which was used for the template generation. Individual susceptibility maps were normalized to this template followed by averaging the normalized susceptibility maps.
MULTI	A multivariate template generation algorithm utilized both T ₁ w and susceptibility maps to generate a susceptibility template.

Table 2

Pairwise Cohen's Kappa coefficient for all ratings (parameter preselection and final assessment) and all three raters.

	Rater1/Rater2	Rater1/Rater3	Rater2/Rater3
MULTI preselection	0.05	0.40	0.45
HYBRID preselection	0.34	0.08	0.25
Final	0.84	0.73	0.89