Automatic brain segmentation using fractional signal modeling of a multiple flip angle, spoiled gradient-recalled echo acquisition.

Ahlgren, André; Wirestam, Ronnie; Ståhlberg, Freddy; Knutsson, Linda

Published in:
Magma

DOI:
10.1007/s10334-014-0439-2

2014

Automatic Brain Segmentation using Fractional Signal Modelling of a Multiple Flip Angle Spoiled Gradient-Recalled Echo Acquisition

André Ahlgren (1), Ronnie Wirestam (1), Freddy Ståhlberg (1,2,3), Linda Knutsson (1)

(1) Department of Medical Radiation Physics, Lund University, Lund, Sweden
(2) Department of Diagnostic Radiology, Lund University, Lund, Sweden
(3) Lund University Bioimaging Center, Lund University, Lund, Sweden

Address (applies to all four authors):
Department of Medical Radiation Physics
Lund University
Barngatan 2B
Skåne University Hospital, Lund
SE-221 85 LUND
Sweden

Corresponding author:
André Ahlgren
E-mail address: Andre.Ahlgren@med.lu.se
Telephone: +46 46-173146
Fax: +46 46-178540

Word count: 5500

Number of figures: 10

Number of tables: 1

Number of references: 32

Number of online resources: 1

Keywords: Brain segmentation, Partial volume estimation, T1 mapping, SPGR, VFA, FRASIER

Abstract

Object: The aim of this study was to demonstrate a new automatic brain segmentation method in MRI.

Materials and Methods: The signal of a spoiled gradient-recalled echo (SPGR) sequence acquired with multiple flip angles was used to map T1, and a subsequent fit of a multi-compartment model yielded parametric maps of partial volume estimates of the different compartments. The performance of the proposed method was assessed through simulations as well as in vivo experiments in five healthy volunteers.

Results: Simulations indicated that the proposed method was capable of producing robust segmentation maps with good reliability. Mean bias was below 3% for all tissue types, and the corresponding similarity index (Dice's coefficient) was over 95% (SNR=100). In vivo experiments yielded realistic segmentation maps, with comparable quality to results obtained with an established segmentation method. Relative whole brain cerebrospinal fluid, grey matter, and white matter volumes were (mean±SE) 6.8±0.5%, 47.3±1.1%, and 45.9±1.3% for the proposed method, and 7.5±0.6%, 46.2±1.2%, and 46.3±0.9% for the reference method.

Conclusion: The proposed approach is promising for brain segmentation and partial volume estimation. The straightforward implementation of the method is attractive, and protocols that already rely on SPGR-based T1 mapping may employ this method without additional scans.

Introduction

In many MRI applications, it is of relevance to be able to distinguish between different tissue types or compartments within a given dataset. For example, tissue separation may be critical for complete assessment of the function in a given structure [1]. Separation of different compartments can also be utilized in the actual calculation of parametric maps [2], while, in other applications, it may be important to know the volume of the structures of interest [3].

A tissue segmentation method generally generates parametric maps containing information about the spatial distribution of the compartments of interest. These maps may be binary or continuous, where the latter may be referred to as fractional segmentation or partial volume estimation. Many MRI datasets, including functional and motion-sensitized MRI, are of low spatial resolution, implying that a fractional segmentation is useful since it may be used for correction of partial volume effects [4].

The present work is focused on brain segmentation, specifically segmentation of grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Many existing brain segmentation methods in MRI are automatic and employ modelling of signal intensity distributions in a morphological scan, and/or prior information about the spatial distributions of different compartment classes (e.g., from atlases) [5]. These methods generally employ multiple mathematically advanced algorithms, which make them difficult to implement. The wide variety of existing segmentation methods has been thoroughly reviewed in the literature [1,6].

Simple segmentation methods, based on quantitative mapping and subsequent modelling of the MRI signal, could be a viable alternative to the established methods. Although such segmentation methods do exist, they are still not readily available [7,8]. One way of performing such a data-driven segmentation is to use quantitative T1 mapping, as suggested by Shin et al. [7], employing a multi inversion time (multi-TI) inversion recovery (IR) sequence. This concept, referred to as ‘FRActional Signal mapping from InvErsion Recovery’ (FRASIER), has been successfully adapted to low-resolution perfusion MRI for calibration purposes [2] and partial volume correction [9,10]. The FRASIER method exploits the difference in T1 between different compartments by modelling the dynamic IR signal, in every voxel, as a sum of the three characteristic recovery curves of the three brain compartments [7]. This intuitive concept is most likely applicable to a wide variety of quantitative and volumetric MRI studies, but we hypothesize that it is also possible to

reformulate it for use with other MRI sequences. We demonstrate this by adapting the IR-based segmentation method by Shin et al. [7] to another common method for quantitative T1 mapping, namely the spoiled gradient-recalled echo (SPGR) sequence with variable flip angle (VFA) acquisition [11,12]. The new segmentation method (which incorporates VFA T1 mapping) is dubbed 'SPoiled Gradient-Recalled echo based SEGmentation' (SPGR-SEG). Results from simulations and five in vivo pilot studies are presented to demonstrate the performance and experimental validity of the proposed method.

Materials and Methods

Brain segmentation

The signal equation for a SPGR sequence is given by:

$$S(\theta) = S_0 \sin(\theta) \frac{1 - \mathrm{e}^{-TR/T_{1,t}}}{1 - \cos(\theta)\, \mathrm{e}^{-TR/T_{1,t}}} \qquad [1]$$

where S_0 depends on longitudinal magnetization at thermal equilibrium and receiver coil sensitivity, θ is the flip angle (FA), TR is the repetition time and $T_{1,t}$ is the longitudinal relaxation time. Note that the transverse relaxation term is omitted, assuming a short echo time. Given that data have been collected with two or more different flip angles, the signal can be modelled as a function of the varying flip angles, and Eq. 1 can be solved for M_0 and $T_{1,t}$. For T1 mapping, this is often accomplished using two low flip angles [13].

Assuming that the VFA signal can be modelled as a sum of three compartments, i.e., CSF, GM, and WM, each with a unique partial signal (signal contribution) and a representative T1 value, Eq. 1 is modified as:

$$S(\theta) = \sin(\theta) \sum_i f_{s,i} \frac{1 - \mathrm{e}^{-TR/T_{1,i}}}{1 - \cos(\theta)\, \mathrm{e}^{-TR/T_{1,i}}} \qquad [2]$$

where $f_{s,i}$ and $T_{1,i}$ are the fractional signal and the longitudinal relaxation time of compartment $i = \{CSF, GM, WM\}$, respectively, and $\sum_i f_{s,i} = M_0$. Eq. 2 can be rewritten in matrix form as $\mathbf{S_m} = \mathbf{XF_s} + \mathbf{e}$, where $\mathbf{S_m} = [S(\theta_1), S(\theta_2), \ldots, S(\theta_N)]^T$ is the discretely measured SPGR signal (for different FAs), $\mathbf{e}$ is the measurement error, and

$$X_{ji} = \sin(\theta_j)\frac{1 - \mathrm{e}^{-TR/T_{1,i}}}{1 - \cos(\theta_j)\,\mathrm{e}^{-TR/T_{1,i}}} \quad [3]$$

$$\mathbf{F_s} = \left[f_{s,CSF}, f_{s,GM}, f_{s,WM}\right]^T \quad [4]$$

$\mathbf{F_s}$ is subsequently solved by linear least squares estimation, i.e., $\mathbf{F_s} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{S_m}$, where $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is the pseudo-inverse of $\mathbf{X}$. Since longitudinal magnetization is proportional to proton density, we can correct for the water content of each compartment to yield fractional volumes:

$$f_{v,i} = \frac{f_{s,i}}{\rho_i} \quad [5]$$

where ρ_i is the volumetric water density of compartment i. Finally, the fractional volumes are constrained by $\sum_i f_{v,i} = 1$ to obtain normalized values. It is important to note that, in contrast to the VFA method for T1 mapping, the SPGR-SEG method needs at least three different flip angles to separate three compartments. Representative $T_{1,i}$ values of each compartment need to be estimated prior to the segmentation. This can be accomplished by first mapping $T_{1,t}$ with the VFA method, and then estimating the mean compartment T1 values from a whole-brain histogram.

Flip angle correction

T1 quantification by the VFA approach requires knowledge of the achieved flip angle, but inhomogeneities in the transmit field ($B1^+$) generally results in a spatially varying flip angle. At field strengths above 1.5 T such inhomogeneities are non-negligible, and mapping of the local flip angle is essential for accurate T1 values [14]. The proposed segmentation method exploits the SPGR signal equation in a similar way, and is also dependent on the local flip angle. Hence, a local flip angle correction factor k was measured using the double angle method (DAM) [15,16]: Two gradient echo signals with long TR and different nominal flip angles, so that $\alpha_{2,nom} = 2\alpha_{1,nom}$, where acquired. Since $S(\alpha) \approx S_0\sin(\alpha)$ when $TR \gg T_1$, the signal ratio is $\lambda = S(\alpha_2)/S(\alpha_1) = \sin(\alpha_2)/\sin(\alpha_1)$. Assuming that $\alpha_2 = 2\alpha_1$, the achieved flip angle is given by $\alpha_1 = \arccos(\lambda/2)$, and the flip angle correction factor is subsequently given by $k = \alpha_1/\alpha_{1,nom}$. The correction factor is then applied voxel-wise to the nominal flip angles of the VFA acquisition. Note that this approach assumes similar slice profiles for the two flip angles.

Simulations

A digital phantom of a normal brain (subject 04) from BrainWeb [17] (http://brainweb.bic.mni.mcgill.ca/brainweb/) was used as ground truth in the simulations. The resolution was subsampled to a 2 mm isotropic voxel size (from 1 mm isotropic), which is similar to the in vivo resolution applied in this study (see 'In vivo experiments'). $T_{1,i}$ values were set to 4.3 s, 1.3 s and 0.8 s for CSF [7], GM and WM [18], respectively, and water content and flip angle correction factor were set to unity. Simulated SPGR signals were generated, based on Eq. 2, using FA=2°/5°/10°/15°/20°/25°/30°, and a TR of 11 ms. Gaussian noise was added with an amplitude corresponding to signal-to-noise ratios (SNRs) ranging from 10 to 200, where SNR was defined as the maximum unperturbed GM SPGR signal value (i.e., the signal value at Ernst angle for GM) divided by the noise standard deviation.

Furthermore, to assess the performance in the presence of more pronounced partial volume effects, the digital phantom was subsampled four times, yielding a 4 mm isotropic resolution, and subsequent segmentation was performed for an SNR of 100. To assess the potential of segmenting additional compartments, e.g., pathologies, a dataset from NITRC TumorSim [19] (http://www.nitrc.org/projects/tumorsim/) was used, were a simulated solid tumor acted as an arbitrary lesion in this work. The digital phantom was based on the same normal brain as the one from BrainWeb. The performance with a fourth compartment was tested with $T_{1,lesion}$ set to 0.5 s, 1.5 s, and 2.5 s, for an SNR of 100.

The sensitivity to errors in flip angle and T1 estimation was investigated by simulations. For a ground truth corresponding to 100% GM, errors in estimated GM volume was calculated for a 10% variation in T1 value (1.17-1.43 s) and for a 20% variation in flip angle (k=0.8-1.2).

In vivo experiments

To evaluate the in vivo performance of SPGR-SEG, five healthy subjects (two females and three males, mean age 31±3.4 y) were scanned with a conventional 3D SPGR/FLASH sequence. The study was approved by the local ethics committee, and all volunteers gave written informed consent. The experiment was performed on a 3T MRI unit (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany) using the following parameters: 52 slices, 3 mm slice thickness, 128×128 matrix, 1.72×1.72 mm^2 in-plane resolution, FA=2°/5°/10°/15°/20°/25°/30°, TR/TE=11/4.2 ms, at a total scan time of 4 min 40 s. For flip angle correction, two 2D SPGR sequences were employed with the following parameters: 52

slices, 3 mm slice thickness, 64×64 matrix, 3.44×3.44 mm^2 in-plane resolution, FA=45°/90°, TR/TE=10000/2 ms, at a total scan time of 12 min 4 s. For comparison purposes, data from a multi-TI IR experiment (IR turbo/fast spin echo) were also acquired using the following parameters: 52 slices, 3 mm slice thickness, 128×128 matrix, 1.72×1.72 mm^2 in-plane resolution, TR/TE=4200/11 ms, TI=50/250/500/750/1000/1500/2000/2500 ms, at a total scan time of 15 min 52 s. All sequences employed parallel imaging (GRAPPA with acceleration factor 2) and $B1^-$ inhomogeneity correction (Prescan Normalize).

Post-processing

All post-processing was executed using in-house developed software (MATLAB 2012a, The MathWorks, Inc., Natick, MA, USA), based on the theory described above. A whole-brain mask, used to remove background voxels in the fitting procedures, was produced using the Robust Brain Extraction software [20] (http://www.nitrc.org/projects/robex/).Flip angle correction maps were calculated using DAM, and the flip angles were subsequently corrected voxel-wise. Motion correction was performed separately on the SPGR and IR data, using Elastix [21] (http://elastix.isi.uu.nl/). For each voxel, Eq. 1 was solved for M_0 and $T_{1,t}$ by nonlinear regression. Longitudinal relaxation properties of GM and WM were estimated from the position of the largest peaks in a whole brain $T_{1,t}$ histogram. Due to the few voxels containing high fractions of CSF, mean T1 in CSF was estimated from a manual ROI placed in the lateral ventricles.

Fractional segmentation was accomplished by constrained linear least squares estimation so that $\mathbf{F_s} \geq 0$, using the proposed model (Eq. 2). Fractional volumes were calculated from the fractional signals according to Eq. 5, assuming water contents of 100%, 89% and 73%, for CSF, GM and WM, respectively [22]. The simulated data were post-processed in the same manner, although assuming error-free $T_{1,i}$ estimation. Fractional segmentation based on the IR data was performed as proposed by Shin et al. [7].

Analysis

The results from the simulations were compared to the true fractional values (ground truth) by calculating accuracy, precision, volume agreement (VA) and volume overlap (VO), for each compartment. Accuracy was defined as the difference between the estimated fractional volume and the ground truth, averaged over the entire brain (with a positive value corresponding to an overestimation). Precision was defined as the root-mean-square deviation

of the corresponding difference (Eq. 7). VA is the agreement in total (whole brain) compartment volume [23], and VO is a similarity measure reflecting the spatial correlation of fractional volumes (cf. Dice's coefficient) [24]. The VO value can be interpreted as the voxel-wise correspondence in compartment volume (overlap) normalized by the average compartment volume in that voxel. Since the estimated fractional volumes are perfectly aligned with the ground truth (i.e., the true spatial distribution of compartments within any given voxel is identical for both maps), the proportion of overlap (nominator in Eq. 9) may be defined as the minimum of the two fractional volume values [24]. VA and VO are both within the interval 0-1, where a high value corresponds to a high agreement and overlap, respectively.

The four performance measures were calculated as follows, for any given compartment i:

$$Accuracy_i = \frac{1}{N}\sum\nolimits_j \left(\tilde{f}_{v,i,j} - f_{v,i,j}\right) \quad [6]$$

$$Precision_i = \sqrt{\frac{1}{N}\sum\nolimits_j \left(\tilde{f}_{v,i,j} - f_{v,i,j}\right)^2} \quad [7]$$

$$VA_i = \left|1 - \frac{\left|\Sigma_j \tilde{f}_{v,i,j} - \Sigma_j f_{v,i,j}\right|}{\Sigma_j \tilde{f}_{v,i,j} + \Sigma_j f_{v,i,j}}\right| \quad [8]$$

$$VO_{i,j} = \frac{\min\left(\tilde{f}_{v,i,j}, f_{v,i,j}\right)}{0.5\left(\tilde{f}_{v,i,j} + f_{v,i,j}\right)} \quad [9]$$

Here, $\tilde{f}_{v,i,j}$ is the estimated fractional volume for compartment i and voxel j, $f_{v,i,j}$ is the corresponding ground truth value, and N is the total number of voxels. To include all types of errors, accuracy and precision were calculated over the entire phantom for all compartments. The drawback of this approach is that each deviation will be included three times (since errors in the three volume maps are not independent), and an alternative approach is to average the measures in the respective compartments only, which results in alternative measures referred to as $Accuracy_i'$ and $Precision_i'$. This was accomplished by producing binary masks, where each voxel was defined as belonging to the compartment with the highest fractional volume (i.e., no voxels were completely omitted), and then calculating the measures in those masks. The VO calculation yields parametric maps, but the values are unreliable in voxels with low fractional volume and, therefore, the mean and standard deviation (SD) of VO were calculated

using the binary masks. In addition, the goodness of fit was assessed by calculating the voxel-wise root mean square error (RMSE), i.e., the square root of the residual sum of squares divided by the number of degrees of freedom. Normalized RMSE (nRMSE) was calculated by dividing RMSE by the voxel-wise maximum SPGR signal.

The in vivo data were analyzed through whole brain compartment volumes, corresponding VA, and whole-brain histograms of fractional volumes for the two methods. Visual inspection and measures of nRMSE was used to further assess the performance in vivo. The nRMSE values were calculated in a way similar to the simulations, i.e., by dividing RMSE by the range of observable values in every voxel, where the range was defined as the maximum SPGR signal at the Ernst angle for SPGR, and two times the longitudinal magnetization at thermal equilibrium (M_0) for IR. In vivo SNR was estimated, similar to the simulations, as the mean maximum obtainable signal in GM divided by the standard deviation in a manually placed background ROI (corrected for Rayleigh distribution).

Results

Simulations

Figure 1 displays an example of parametric results from the simulation study (SNR=100), including ground truth, estimated volumes using SPGR-SEG, and difference maps. It is clear that the largest errors arise due to volume underestimation in GM. Voxels with a mixture of all three compartments are misclassified as GM (see edge of the lateral ventricles). This observation was not present for the original phantom with 1×1×1 mm^3 resolution (data not shown).

Density plots of the entire digital phantom (240082 voxels), comparing ground truth with the estimated volumes, are displayed in Figure 2, for SNR=100. Random deviations from the identity line are most prominent for GM, in accordance with Figure 1 where the GM map appears slightly noisier than the other maps. Comparison of segmentation quality for three different SNRs (10/50/100) is presented in Figure 3. The rightmost column displays the corresponding nRMSE (goodness of fit). Calculated performance measures are given in Figure 4 for six different SNRs (10/25/50/100/150/200). All performance measures, as well as the goodness of fit, improved with increasing SNR. The values in the plots of Figure 4 can be found in table format (Online Resource 1).

Segmentation results for low resolution data (28019 voxels), and for the lesion data (i.e., four compartments), are presented in Table 1 for SNR=100. Spatial segmentation quality for the lesion data set is visualized in Figure 5 for the three different simulated lesion T1 values. Best overall segmentation quality was obtained for $T_{1,lesion}$=0.5 s.

Figure 6 displays the bias in estimated GM volume due to errors in estimated mean T1 (Fig. 6a) and due to errors in flip angle correction (Fig. 6b-c). Figure 6b assumes error-free T1 estimation even though the flip angle is incorrect (which corresponds to, for example, using literature T1 values or estimating T1 with an independent method), whereas Figure 6c assumes that an (incorrect) GM T1 value is estimated from the erroneous data points. The simulation showed that, for a 10% variation in the representative T1 of GM, the average error in the partial volume estimation of pure GM was 8.1±4.7% (mean±SD), and the corresponding maximum error was 17.4%. A 20% variation in effective flip angle yielded an average error of 33.6±22.0% (maximum error 91.1%) with independent T1 estimation, and an average error of 2.0±4.0% (maximum error 26.1%) when a representative GM T1 value was estimated from the uncorrected data.

In vivo experiments

Figure 7 displays examples of segmentation maps produced by SPGR-SEG in all five subjects. Figure 8 depicts a single slice comparison of the in vivo segmentation results using the proposed method (SPGR-SEG) and the reference method (FRASIER). The rightmost panels display zoomed parts of the respective maps, and the bottom row displays the corresponding nRMSE (goodness of fit). The spatial distributions of fractional volumes agreed well with known brain structures for both methods, although differences between the methods are noticeable, for example, in deep grey matter (DGM) structures. The mean whole brain nRMSE for all five subjects was (mean ± standard error of mean; SE) 3.5±0.06% for SPGR-SEG and 3.5±0.04% for FRASIER. Although the mean nRMSE was equal for both methods, the spatial distribution of nRMSE values differed between the methods (Fig. 8).

Mean in vivo SNR was (mean±SE) 582±32 for SPGR-SEG and 376±35 for FRASIER. The mean number of voxels in the extracted brain was (mean±SE) 179664±8818 (c.f. the digital phantom which consisted of 240082 voxels). Mean relative whole brain CSF/GM/WM volumes in all subjects were (mean±SE) 6.8±0.5% / 47.3±1.1% / 45.9±1.3% for SPGR-SEG and 7.5±0.6% / 46.2±1.2% / 46.3±0.9% for FRASIER. None of the whole brain compartment

volumes were significantly different between the two methods (two-sided t-test, α=0.05). The corresponding VA between the methods was (mean±SE) 0.95±0.01%, 0.98±0.002% and 0.98±0.004 for CSF, GM and WM, respectively.

Figure 9 shows a comparison of the segmentation results in five slices of one representative subject. The maps are similar overall, but differences in, for example, DGM and ventricles (middle slice) are noticeable. The distributions of fractional volumes in all five subjects and for both methods are shown in Figure 10 in the form of histogram plots. There are significant differences between the corresponding histograms of the two methods.

Discussion

Simulations

For a simulated SNR of 100, the proposed method produced segmentation results of visually good quality, although a general underestimation in GM voxels, spread throughout the simulated brain, was observed (Fig. 1-2). The general underestimation of GM volumes was confirmed by the performance plots (Fig. 4). In particular, the density plot in Figure 2b shows that SPGR-SEG had a tendency to underestimate the GM volume in voxels with a high fraction of GM. Furthermore, Figure 2a suggested that the CSF volume was generally slightly overestimated. For the discretization used in the density plots (2% intervals), the mode of the estimated fractional CSF volume was 24%, 52% and 82%, for a ground truth of 20%, 50% and 80%, respectively. Apart from this, the density plots suggest an excellent correspondence between the ground truth and the estimated volumes with Pearson correlation coefficients around 0.99. Figure 1 also demonstrates that SPGR-SEG misclassified voxels with a mixture of all three compartments (edge of lateral ventricles). Although this error is obvious and expected, border zones with three adjacent components are rare, and this type of misclassification can be minimized by employing high resolution acquisitions.

The simulation study suggested that SPGR-SEG performs well at SNR levels above 100 (Fig. 3-4). The measure of accuracy was close to zero for all compartments for SNR≥100, whereas the segmentation quality for SNR<50 suggested that SPGR-SEG is likely to be inapplicable at lower SNR levels. The precision measure was lowest (i.e., best precision) for CSF and WM, and stabilized at approximately 3% as SNR increased. The difference between whole brain measures (Accuracy/Precision) and compartmental measures (Accuracy'/Precision') were

most apparent for GM, with a worse accuracy and precision for compartmental values. This is consistent with Figure 1 (GM difference map), where the majority of the errors can be seen in GM. The accuracy and precision of CSF and WM, on the other hand, were, for the same reason, better for the compartmental measures. For an SNR of 100, the mean nRMSE was 1.09±0.57 (mean±SD), i.e., the residuals had a mean magnitude comparable to 1% of the maximum signal, which is reasonable since the noise magnitude was approximately 1% of the maximum signal in GM (according to the SNR definition).

The corresponding volume agreement and volume overlap values were also satisfactory in the simulation study. For SNR=100, all VA values were over 0.96 which suggests an excellent agreement. The total true volume and estimated volume for each compartment was 17%/19% (CSF), 49%/48% (GM), and 34%/34% (WM). A VO value above 0.7 has been regarded excellent in the literature [25]. Normally, VO reflects the volume of voxels classified as belonging to the same class, across two measurements, normalized by the average volume for that tissue category [24]. Here, fractional values are compared between a ground truth and a simulated measurement, which makes the interpretation less straightforward. On the other hand, a VO of around 0.97 (SNR=100) was obtained which suggests very good similarity, regardless of the exact interpretation. Error-free flip angle correction and $T_{1,i}$ estimation was assumed in the simulations, and the performance is expected to decrease when uncertainties are present in these steps.

The segmentation accuracy of the low-resolution data set was similar to the standard data set (Table 1), with a minor overall decrease in performance. This is encouraging regarding the possibility to apply SPGR-SEG in low-resolution quantitative MRI applications.

The lesion simulation results (Table 1 and Fig. 5) demonstrated that the multi-component modelling becomes significantly more difficult with additional compartments. In particular, Table 1 shows that the performance was reduced in most cases. This is expected since more compartments generally yield an increase in the amount of possible solutions. Only considering the number of possible binary combinations, they increase from 7 to 15 when extending the model to include a fourth compartment. The lower performance for the lesion data sets was accompanied by a small reduction in nRMSE, which is related to the increase in free parameters (i.e, overfitting). Figure 5 suggests that SPGR-SEG selects solutions with a high volume in one compartment, rather than highly mixed solutions (in consistence with Fig. 1). This is especially clear in the lesion segmentation with $T_{1,lesion}$=2.5 ms (bottom right map

in Fig. 5), where voxels with a mixture of CSF and GM are classified as lesion. As mentioned above, a higher resolution will reduce this type of misclassifications. For example, the performance of the lesion segmentation was improved when using the original 1×1×1 mm^3 phantom resolution (data not shown). A similar increase in performance could also be achieved by repeating the sampling or acquiring additional flip angles. Hence, segmentation of additional components using SPGR-SEG (for example, a pathological lesion) seems challenging, and good data quality and/or an extended model, taking into account additional tissue information, may be needed to enable satisfactory results. Furthermore, it is of particular importance that the additional component has a representative T1 value which is different from those of the existing compartments.

The sensitivity of SPGR-SEG to heterogeneities in T1 corresponded to an average error of 8.1% for a 10% variation in GM T1 (Fig. 6a). Whereas the simulated T1 variation may be realistic in some areas, DGM structures often have significantly lower T1 values due to myelin content and iron depositions [7]. Segmentation performance in such structures will depend on the difference in T1 between whole-brain GM and DGM, as well as how the segmentation method handles mixed signals. Shin et al. pointed out that, although DGM is mostly composed of GM, it also contains myelinated axons which, in the strict sense, can be regarded as fractions of WM [7]. However, as with the FRASIER method, SPGR-SEG does not account for iron content and, therefore, GM volumes in DGM are likely to be underestimated.

Further simulations showed that the method may be very sensitive to uncorrected variations in FA (33.6% average error for a 20% variation; Fig. 6b). However, this result assumed error-free T1 estimation, which is unlikely when the FAs are wrong. Still, if this were the case, errors in GM volume would be very large without FA correction. More realistically, T1 is estimated from the same data as used for segmentation, and in such a case the average error was reduced to 2% (Fig. 6c), which emphasizes the high coupling between flip angle and T1 in the model. This situation highlights another unrealistic extreme, which is primarily applicable to a global error in achieved FA. The $B1^+$ field is generally inhomogeneous across the brain, and uncorrected errors in local FA will, firstly, broaden the peaks in the whole brain T1 histogram hampering the estimation of $T_{1,i}$ values, and, secondly, introduce spatially varying errors in partial volumes estimates. A 20% variation in achieved FA is realistic at 3 T [26] and, therefore, we can expect errors of the order of 10-30% if FA correction is not included. The error is close to zero for α/α_{nom}>1 when T1 is estimated from the data (Fig.

6c), which is the result of two effects. Firstly, for $1<\alpha/\alpha_{nom}<1.2$ the estimated T1 in GM is still very different from the fixed T1 value for CSF (4.3 s). Secondly, the nonnegative constraint on the least squares problem avoids overestimation in GM volume (which would correspond to a negative error in Fig. 6c). This is true until the flip angle error becomes so large that the estimated T1 in GM falls above the T1 value for CSF, which happens around α/α_{nom} =1.77 (which is unlikely at 3 T [26]).

In vivo experiments

The proposed segmentation method yielded realistic and robust segmentation results in all investigated subjects (Fig. 7). In general, the SPGR-SEG output agreed well with the segmentation maps produced by the reference method (Fig. 8-9), which was also suggested by the similarity in whole-brain compartment volumes and high volume agreement. Still, distinct differences between the methods was observed across all subjects, and one of the most noticeable differences was that DGM structures (e.g. thalamus, putamen and caudate nucleus) seemed to be more clearly delineated and had higher volumes in the GM map originating from SPGR-SEG (Fig. 8-9). This difference, although hard to explain with certainty, may be a consequence of differences in SNR and image quality in the center of the brain. SPGR-SEG and FRASIER estimate partial volumes and, therefore, DGM structures are determined to be a mixture of GM and WM [7] (see simulation discussion). In many applications, binary delineation of DGM may be of interest and magnetization transfer mapping has been proposed as a feasible way to yield higher GM fractions in DGM structures [27].

The spatial difference in nRMSE between SPGR-SEG and FRASIER is related to differences in the obtained MRI signal. Whereas an IR sequence has a maximum MRI signal primarily related to the magnetization at thermal equilibrium, the maximum MRI signal of a SPGR sequence is strongly related to T1 relaxation. The SNR was also higher for the SPGR acquisition, compared to the IR acquisition.

Differences between the methods may also related to the difference in sequence design and, subsequently, signal modelling. For example, a high precision in T1 mapping is vital for both methods, and the difference in T1 estimation methodology is likely to propagate to differences in estimated fractional volumes. In particular, VFA is more sensitive to variations in excitation flip angle, whereas IR T1 mapping rely on robust inversion efficiency. However, neither segmentation method depends on absolute T1 values, but rather on each compartment

showing a representative estimable T1 value. Nevertheless, flip angle correction is important in SPGR-SEG since spatial variations of the achieved flip angle introduce errors in the model, and also obstruct the $T_{1,i}$ estimation (see simulation discussion). In this work, calculated T1 values, based on SPGR acquisition, were overestimated compared with literature values (data not shown). The deviation from literature values could be due to a number of reasons. For example, it has been suggested that absolute T1 mapping by the VFA method requires absolute FA calibration [26], correction for RF pulse shape [28], and correction for incomplete RF spoiling [29].

The fractional volume histograms (Fig. 10) show that there were differences in the distribution of compartment partitions between the two segmentation methods. The small difference for $f_{v,CSF}$>0.5 (Fig. 10a) was also seen in the CSF maps (e.g., lateral ventricles in Fig. 9). Even larger differences were seen for $f_{v,CSF}$<0.5, but the source of this difference was difficult to analyze. High $f_{v,GM}$ values were systematically underestimated in the simulations, which possibly could explain the general lower amount of voxels with a high $f_{v,GM}$ in SPGR-SEG, compared to FRASIER (Fig. 10b). FRASIER also showed a larger amount of low $f_{v,GM}$<0.5, which could be associated with the overall lower $f_{v,GM}$ values in DGM for FRASIER, compared to SPGR-SEG. The larger amount of voxels with high fractions of WM in FRASIER (Fig. 10c) was also observable in the zoomed panels of Figure 8. The differences imply that, currently, partial volume estimates will depend on the choice of method and, therefore, these differences need to be further investigated.

The total acquisition time was 15 min 52 s for FRASIER and 16 min 44 s for SPGR-SEG. Both methods can, in a significantly shorter time, produce segmentation results of comparable quality as was presented here, although it will require special sequences that were not available in the present study. For example, whole-brain FRASIER data acquisition has previously been performed in 4 min 32 s by applying Look-Locker read-out [7], and SPGR-SEG could be accelerated by using ultra-short TR (e.g., TurboFLASH) combined with faster $B1^+$ mapping methods (e.g., actual flip-angle imaging [30] or Bloch-Siegert shift [31]).

Comparison of simulations and in vivo experiments

The in vivo SNR values seemed high in both methods, which is explained by the somewhat unconventional definition (used in order to be consistent with the simulations). Still, in vivo segmentation maps (Fig. 7-9) were almost completely free from noise using both methods,

roughly to be compared with a simulated SNR≥100 in Figure 3. This means that the SPGR-SEG acquisition did not suffer from significant disturbance from thermal noise in the segmentation performance, and that higher resolution acquisitions should be feasible.

The in vivo mean nRMSE was 3.5% across both methods, which compares to a simulated SNR of approximately 30 (Figure 4d), i.e., a significantly lower value compared to the estimated SNR in vivo. This illustrates the difficulty in comparing simulations with real data, and that the simulations should only be regarded as an indication of the performance of SPGR-SEG. For example, the in vivo measurements will suffer from perturbations not included in the simulations, e.g., imaging artifacts, errors in flip angle correction and compartmental variations in T1. The in vivo VA between FRASIER and SPGR-SEG was ≥0.95 for all compartments. Comparing with the VA values in the simulations (Online Resource 1 and Fig. 4), this corresponds to an SNR of around 100. Note, however, that in the simulations, a ground truth is compared with estimated values, whereas for the in vivo data two different experimental methods are compared.

Prospects

The proposed method is simple, robust and does not depend on the use of spatial priors, complex modelling or computationally intensive optimization. The development of SPGR-SEG was largely based on the work by Shin et al. [7], and since our algorithm was easily derived we believe that similar segmentation methods, plausibly exploiting additional sequences and quantitative parameters, will emerge. It should also be noted that, although this work demonstrates the application to brain segmentation, the methodology may be directly applicable to any structure or object. For the possibility of separating additional compartments in the brain, e.g., fat, vessels, tumors or other pathologies, one of the most important factors is that the representative T1 value differs from the other compartments. Successful segmentation of pathologic areas from healthy tissue is of great importance in clinical research, and future work of interest includes investigating the possibility of achieving this with SPGR-SEG.

The proposed approach is particularly suitable for applications in which SPGR T1 mapping is already included, e.g., Dynamic Contrast Enhanced MRI (DCE-MRI) [32], where implementation may be possible without additional scans. A faster acquisition would make the method even more attractive since it may be easily added to existing protocols.

Note that transversal relaxation effects are disregarded in the present work. Although T1 quantification with VFA is insensitive to such effects, due to the linear scaling of the signal, the proposed segmentation method will not be insensitive since different compartments have different T2*. Although the errors are expected to be minor, inclusion of T2* relaxation effects would reduce those errors and possibly also improve segmentation performance of extra compartments.

In this work, seven flip angles were used to allow for a reasonable acquisition time, whereas a rigid optimization of protocol parameters (e.g., TR and number and choice of FAs) is a highly complex issue, which is likely to yield better segmentation performance. In the further development of SPGR-SEG, areas of interest includes assessing the reproducibility, optimizing imaging parameters, and designing an acquisition protocol with a shorter, clinically more acceptable, scan time.

Conclusion

We have proposed a new method, called SPGR-SEG, for automatic brain segmentation in MRI, based on quantitative T1 mapping with SPGR sequences. It produced robust fractional volume maps of comparable quality to a reference method. The proposed method only requires access to a standard SPGR sequence and is therefore immediately feasible at most MRI sites. Initial simulation results are encouraging with regard to reliability, and results from in vivo experiments confirm the feasibility of acquiring brain segmentation maps under realistic experimental conditions. Simulations suggested that segmentation of additional compartments, e.g., pathologies, is challenging with the current protocol. We believe that similar segmentation methods, based on fractional MR signal modelling, will emerge and show significant usefulness in clinical research.

Acknowledgements

This project was supported by the Swedish Research Council (grant nos. 13514, 2005-6910, 2007-3974 and 2007-6079), the Crafoord foundation, the Lund University Hospital Donation Funds and the Swedish Cancer Society, grant no. 2009/1076.

References

1. Balafar MA, Ramli AR, Saripan MI, Mashohor S (2010) Review of brain MRI image segmentation methods. Artif Intell Rev 33:261-274

2. Ahlgren A, Petersen E, Ståhlberg F, Wirestam R, Knutsson L (2011) Using fractional segmentation for estimation of the equilibrium magnetization of arterial blood in model-free arterial spin labeling. In: Proceedings of the 28th Annual Meeting ESMRMB, Leipzig, p 554

3. Giorgio A, De Stefano N (2013) Clinical use of brain volumetry. J Magn Reson Imaging 37:1-14

4. Asllani I, Borogovac A, Brown TR (2008) Regression algorithm correcting for partial volume effects in arterial spin labeling MRI. Magn Reson Med 60:1362-1371

5. Ashburner J, Friston KJ (2005) Unified segmentation. NeuroImage 26:839-851

6. Clarke LP, Velthuizen RP, Camacho MA, Heine JJ, Vaidyanathan M, Hall LO, Thatcher RW, Silbiger ML (1995) MRI segmentation: Methods and applications. Magn Reson Imaging 13:343-368

7. Shin W, Geng X, Gu H, Zhan W, Zou Q, Yang Y (2010) Automated brain tissue segmentation based on fractional signal mapping from inversion recovery Look-Locker acquisition. NeuroImage 52:1347-1354

8. West J, Warntjes JB, Lundberg P (2012) Novel whole brain segmentation and volume estimation using quantitative MRI. Eur Radiol 22:998-1007

9. Petr J, Schramm G, Hofheinz F, Langner J, van den Hoff J (2013) Partial volume correction in arterial spin labeling using a Look-Locker sequence. Magn Reson Med 70:1535-1543

10. Ahlgren A, Petersen E, Ståhlberg F, Wirestam R, Knutsson L (2012) Partial volume correction in model-free arterial spin labeling. In: Proceedings of the 29th Annual Meeting ESMRMB, Lisbon, p 110

11. Christensen KA, Grant DM, Schulman EM, Walling C (1974) Optimal determination of relaxation times of fourier transform nuclear magnetic resonance. Determination of spin-lattice relaxation times in chemically polarized species. J Phys Chem 78:1971-1977

12. Fram EK, Herfkens RJ, Johnson GA, Glover GH, Karis JP, Shimakawa A, Perkins TG, Pelc NJ (1987) Rapid calculation of T1 using variable flip angle gradient refocused imaging. Magn Reson Imaging 5:201-208

13. Deoni SC, Rutt BK, Peters TM (2003) Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. Magn Reson Med 49:515-526

14. Cheng H, Wright G (2006) Rapid high-resolution T1 mapping by variable flip angles: accurate and precise measurements in the presence of radiofrequency field inhomogeneity. Magn Reson Med 55:566-574

15. Insko EK, L B (1993) Mapping of the radiofrequency field. J Magn Reson A 103:82– 85

16. Stollberger R, Wach P (1996) Imaging of the active B1 field in vivo. Magn Reson Med 35:246-251

17. Aubert-Broche B, Griffin M, Pike GB, Evans AC, Collins DL (2006) Twenty new digital brain phantoms for creation of validation image data bases. IEEE Trans Med Imag 25:1410-1416

18. Wansapura JP, Holland SK, Dunn RS, Ball WS, Jr. (1999) NMR relaxation times in the human brain at 3.0 tesla. J Magn Reson Imaging 9:531-538

19. Prastawa M, Bullitt E, Gerig G (2009) Simulation of brain tumors in MR images for evaluation of segmentation efficacy. Med Image Anal 13:297-311

20. Iglesias JE, Liu CY, Thompson PM, Tu Z (2011) Robust brain extraction across datasets and comparison with publicly available methods. IEEE Trans Med Imaging 30:1617-1634

21. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP (2010) elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 29:196-205

22. Donahue MJ, Lu H, Jones CK, Edden RA, Pekar JJ, van Zijl PC (2006) Theoretical and experimental investigation of the VASO contrast mechanism. Magn Reson Med 56:1261-1273

23. Liu T, Young G, Huang L, Chen NK, Wong ST (2006) 76-space analysis of grey matter diffusivity: methods and applications. NeuroImage 31:51-65

24. Cardenas VA, Ezekiel F, Di Sclafani V, Gomberg B, Fein G (2001) Reliability of tissue volumes and their spatial distribution for segmented magnetic resonance images. Psychiat Res-Neuroim 106:193-205

25. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC (1994) Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Trans Med Imag 13:716-724

26. Wang J, Qiu M, Kim H, Constable RT (2006) T1 measurements incorporating flip angle calibration and correction in vivo. J Magn Reson 182:283-292

27. Helms G, Draganski B, Frackowiak R, Ashburner J, Weiskopf N (2009) Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps. NeuroImage 47:194-198

28. Parker GJ, Barker GJ, Tofts PS (2001) Accurate multislice gradient echo T(1) measurement in the presence of non-ideal RF pulse shape and RF field nonuniformity. Magn Reson Med 45:838-845

29. Preibisch C, Deichmann R (2009) Influence of RF spoiling on the stability and accuracy of T1 mapping based on spoiled FLASH with varying flip angles. Magn Reson Med 61:125-135

30. Yarnykh VL (2007) Actual flip-angle imaging in the pulsed steady state: a method for rapid three-dimensional mapping of the transmitted radiofrequency field. Magn Reson Med 57:192-200

31. Sacolick LI, Wiesinger F, Hancu I, Vogel MW (2010) B1 mapping by Bloch-Siegert shift. Magn Reson Med 63:1315-1322

32. Wirestam R (2012) Using contrast agents to obtain maps of regional perfusion and capillary wall permeability. Imaging Med 4:423-442

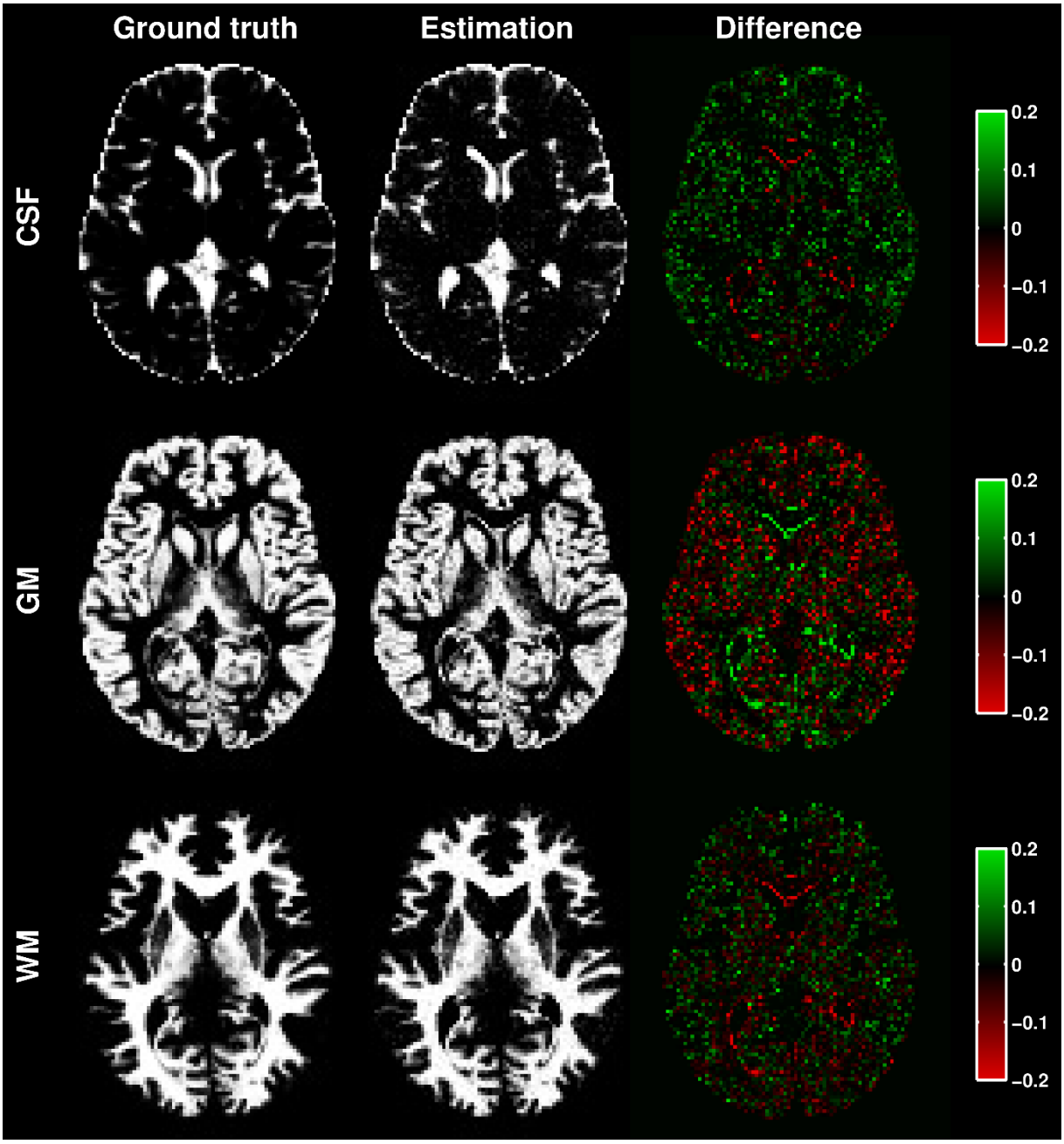

Fig. 1 Simulation results for SNR=100 depicting one slice of the digital phantom (ground truth), estimated volumes in the same slice, and the voxel-wise difference (accuracy). In the difference maps, positive values (green) correspond to an overestimation, and negative values (red) correspond to an underestimation

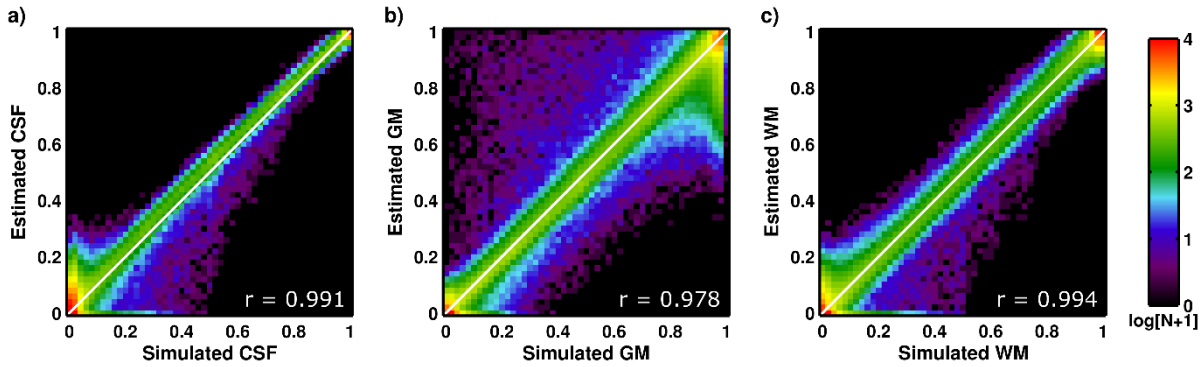

Fig. 2 Density plots of the estimated volumes versus the simulated ground truth, for SNR=100, in a) CSF, b) GM, and c) WM. White solid lines correspond to the identity line, and Pearson correlation coefficients are noted in white text. Note that the bin count *N* is logarithmized to facilitate the dynamic range of the color scale

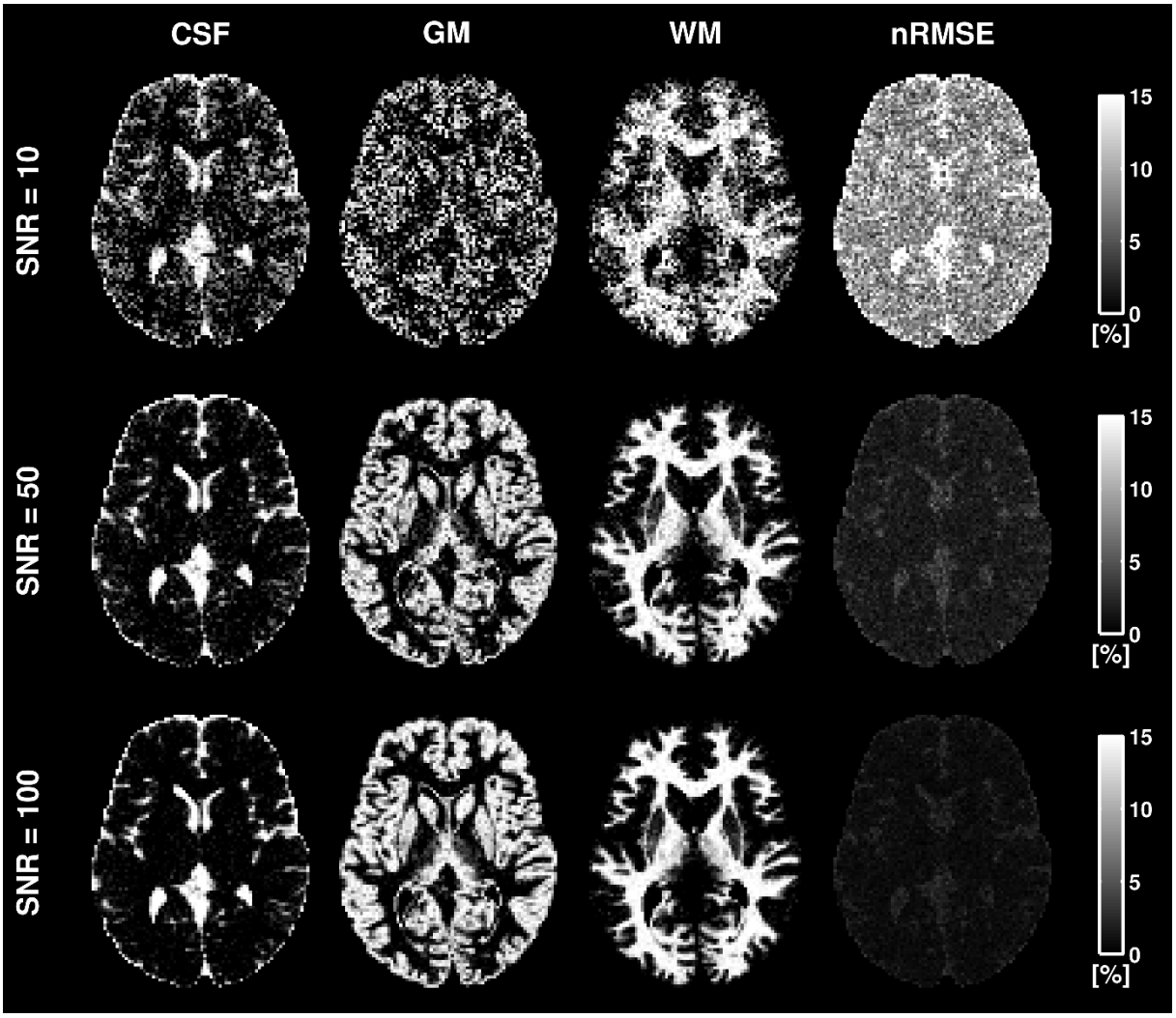

Fig. 3 Illustration of simulated segmentation quality for three different SNR levels (10, 50 and 100). Last column displays the voxel-wise goodness of fit, defined as the normalized root-mean-squared-error (nRMSE), i.e., RMSE in percentage of the maximum SPGR signal

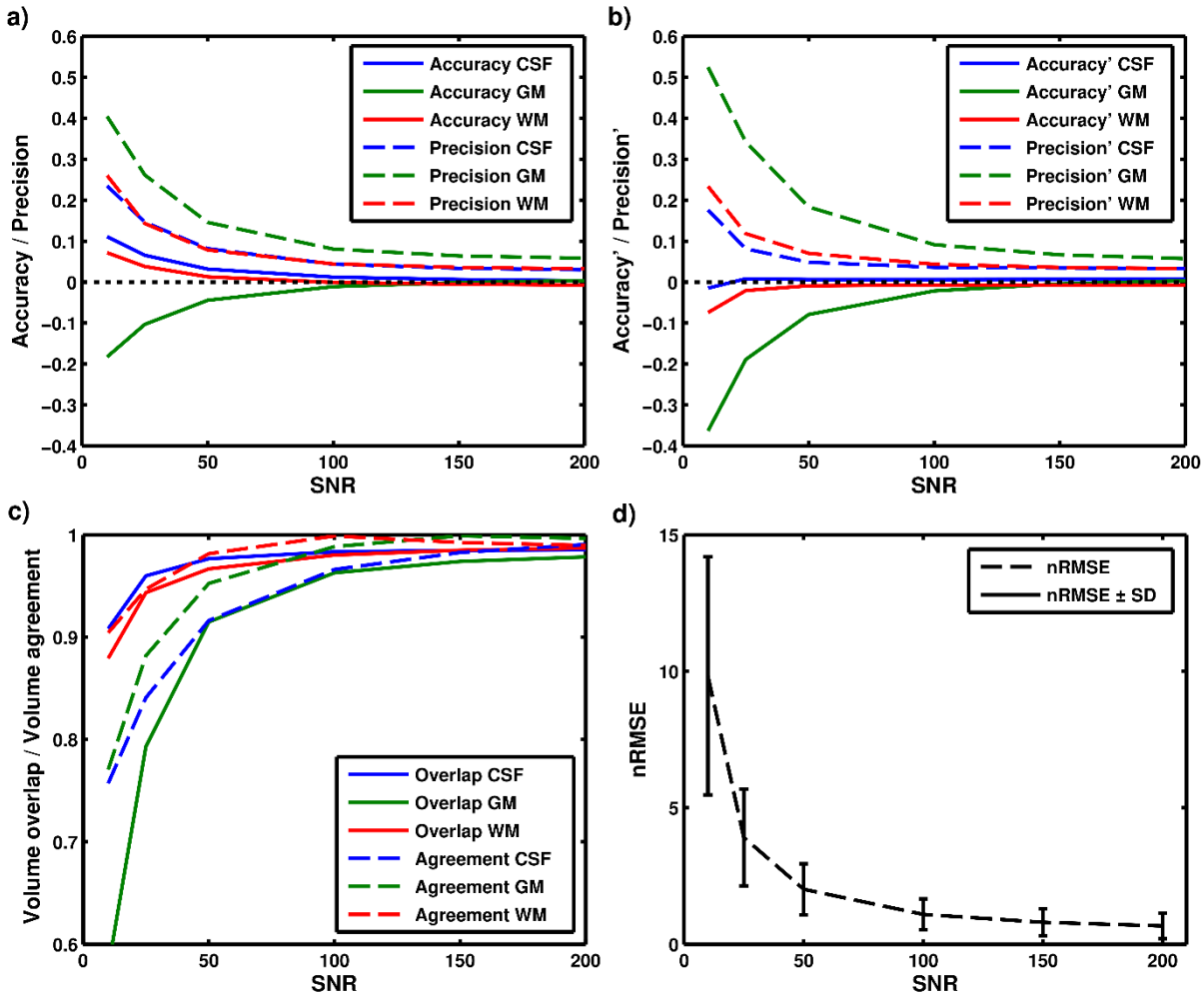

Fig. 4 Performance measures and goodness-of-fit (nRMSE) from the simulation study. In a-c, blue, green and red color corresponds to CSF, GM and WM, respectively. a) Accuracy (solid lines) and precision (dashed lines) averaged over the entire digital phantom. b) Accuracy (solid lines) and precision (dashed lines) averaged over the respective compartment. c) Volume overlap (solid lines) and volume agreement (dashed lines). d) nRMSE averaged over the entire digital phantom, with error bars corresponding to one standard deviation

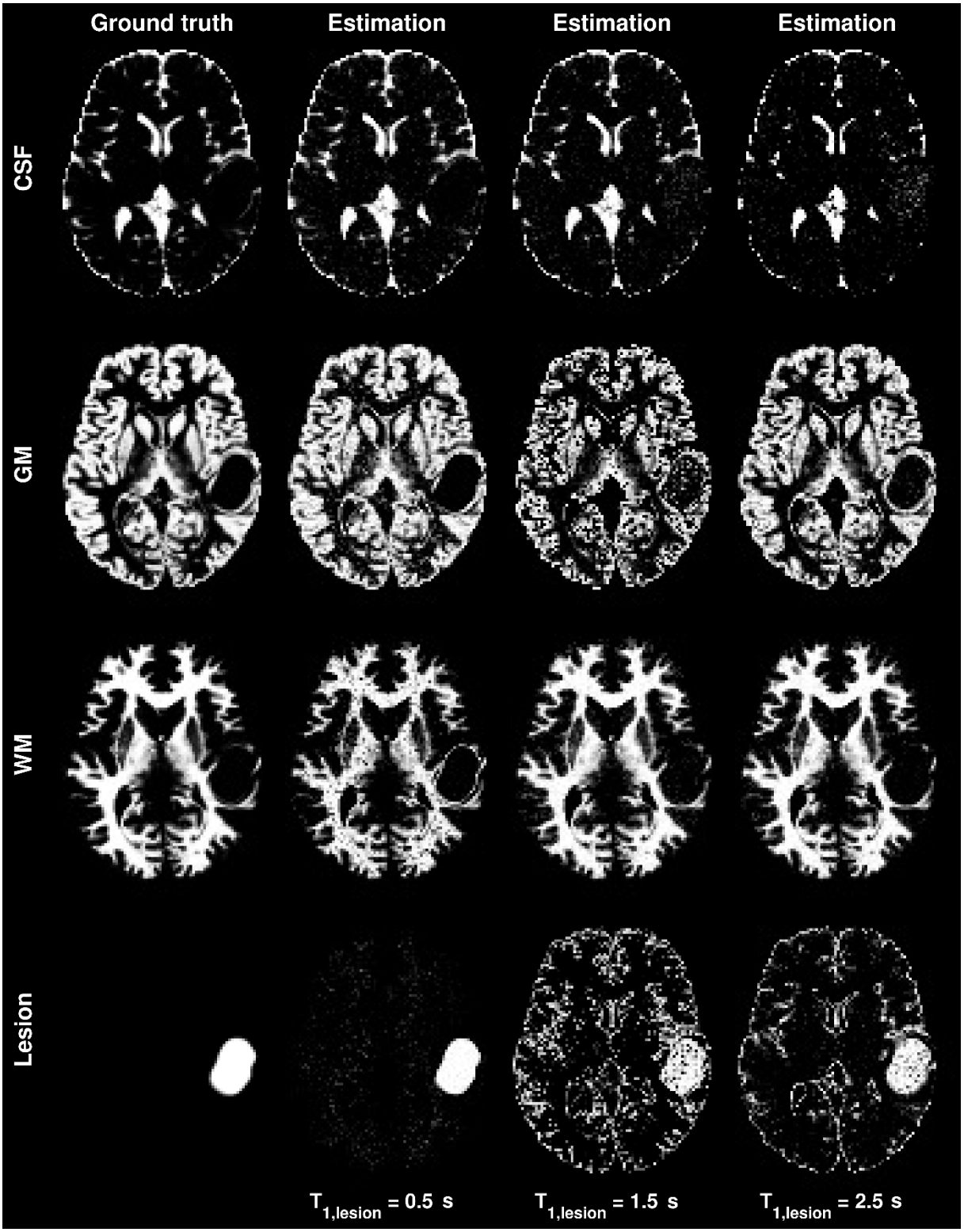

Fig. 5 Simulated segmentation quality when a fourth component (e.g., lesion/pathology) is present, for SNR=100. The left-most column displays the ground truth and the other columns display segmentation results for three different T1 values (0.5/1.5/2.5 s) of the simulated lesion

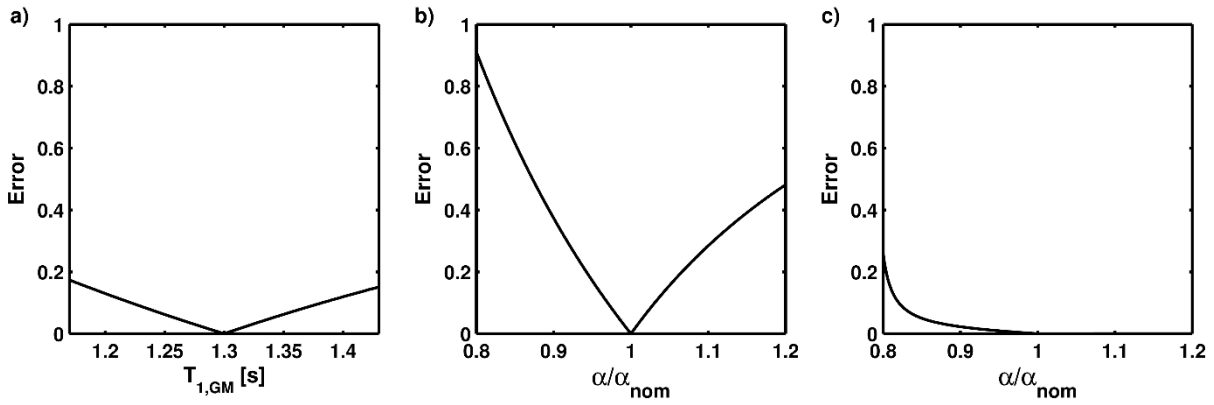

Fig. 6 Bias in estimated GM volume (in a voxel with pure GM) due to variations in T1 value and flip angle. a) Error in estimated GM volume as a function of ±10% variation in T1 (mean $T_{1,GM}$=1.3 s). b) Error in estimated GM volume as a function of ±20% variation in achieved flip angle, assuming error-free T1 estimation. c) Error in estimated GM volume as a function of ±20% variation in achieved flip angle, assuming that T1 is estimated from the erroneous data points

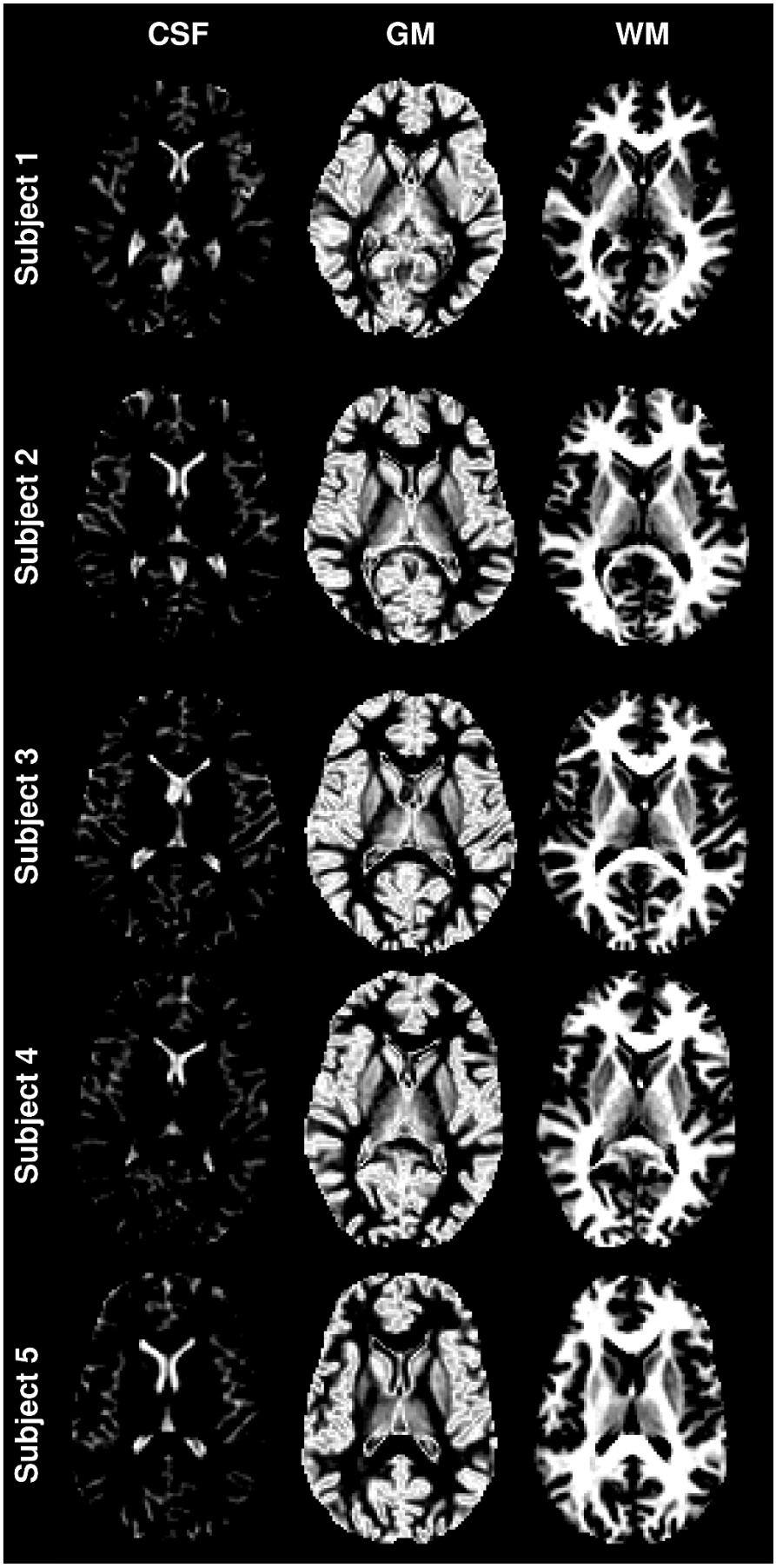

Fig. 7 In vivo segmentation results in on slice from all five subjects using SPGR-SEG

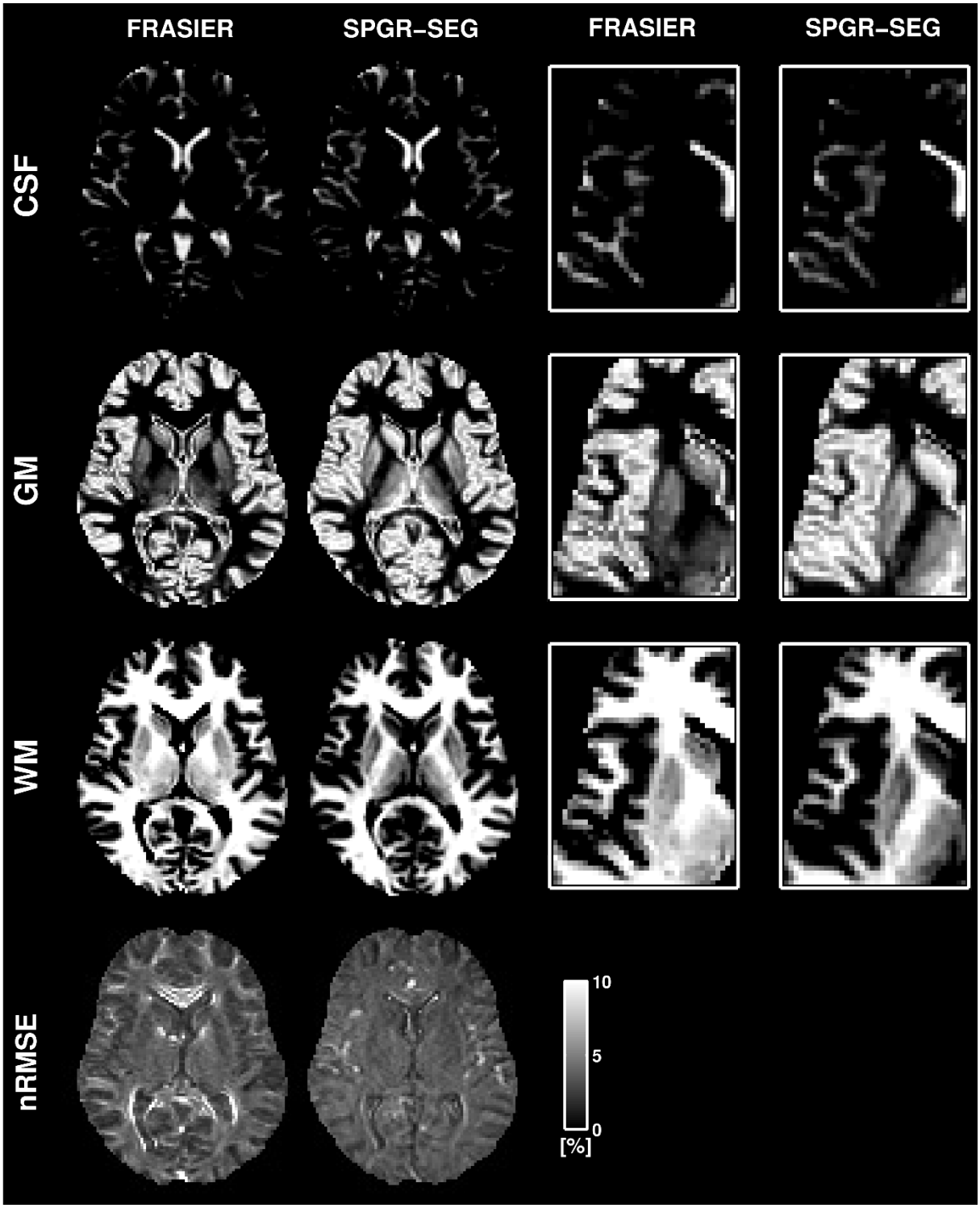

Fig 8 In vivo segmentation results from one slice in one subject using the reference method (FRASIER) and the proposed method (SPGR-SEG). The six panels to the right display zoomed parts of the segmentation maps and the bottom row displays nRMSE of the respective model fit

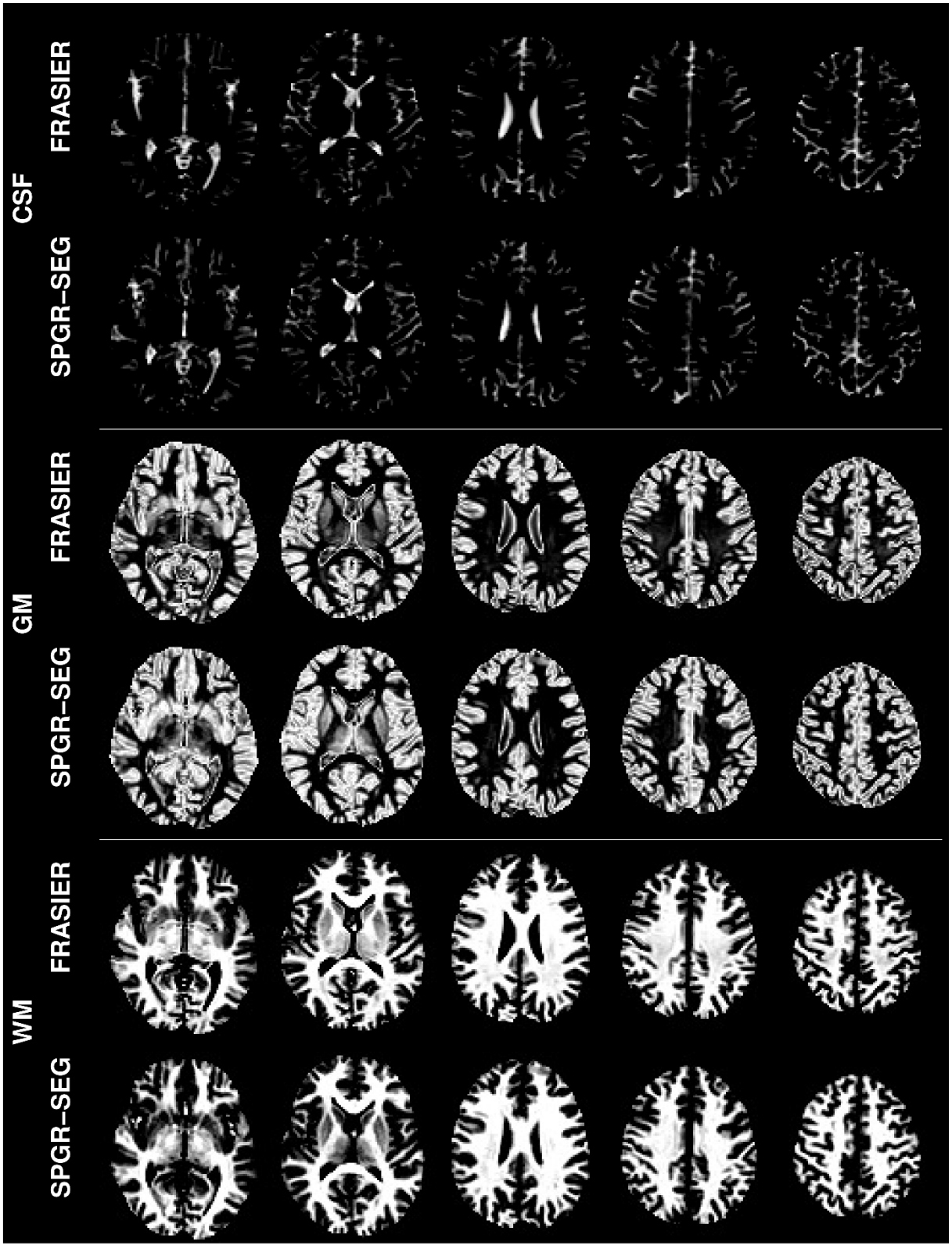

Fig 9 Comparison of in vivo segmentation results using the reference method (FRASIER) and the proposed method (SPGR-SEG) in five different slices of one subject

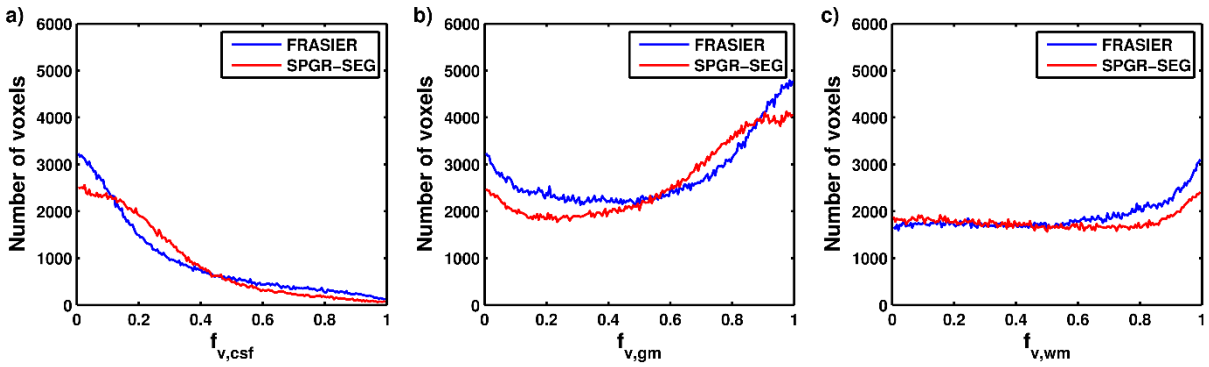

Fig 10 Whole brain histogram plots of in vivo segmentation results, comparing estimated volumes of the reference method (FRASIER; blue line) and the proposed method (SPGR-SEG; red line) in a) CSF, b) GM, and c) WM

Data set	Accuracy CSF / GM / WM (/lesion)	Precision CSF / GM / WM (/lesion)	Accuracy' CSF / GM / WM (/lesion)	Precision' CSF / GM / WM (/lesion)	Volume overlap CSF / GM / WM (/lesion)	Volume agreement CSF / GM / WM (/lesion)	nRMSE
Standard	0.01 / -0.01 / 0.00	0.04 / 0.08 / 0.04	0.01 / -0.02 / -0.01	0.04 / 0.09 / 0.04	0.98 (0.04) / 0.96 (0.05) / 0.98 (0.04)	0.97 / 0.99 / 1.00	1.09 (0.57)
Low res	0.00 / 0.02 / -0.02	0.05 / 0.11 / 0.06	-0.01 / 0.02 / -0.01	0.06 / 0.10 / 0.06	0.96 (0.08) / 0.95 (0.06) / 0.97 (0.08)	0.99 / 0.98 / 0.97	1.07 (0.54)
$T_{1,lesion}$ **0.5 s**	0.01 / 0.01 / -0.03 / 0.01	0.04 / 0.09 / 0.10 / 0.04	0.01 / -0.01 / -0.07 / -0.03	0.04 / 0.09 / 0.15 / 0.08	0.98 (0.05) / 0.96 (0.05) / 0.94 (0.11) / 0.97 (0.10)	0.97 / 0.99 / 0.96 / 0.80	1.07 (0.58)
$T_{1,lesion}$ **1.5 s**	-0.02 / -0.16 / 0.01 / 0.16	0.06 / 0.32 / 0.06 / 0.33	-0.03 / -0.26 / -0.01 / -0.19	0.07 / 0.43 / 0.05 / 0.32	0.96 (0.09) / 0.68 (0.40) / 0.98 (0.05) / 0.82 (0.28)	0.95 / 0.80 / 0.99 / 0.21	1.02 (0.50)
$T_{1,lesion}$ **2.5 s**	-0.06 / -0.05 / 0.00 / 0.11	0.17 / 0.13 / 0.05 / 0.25	-0.25 / -0.06 / -0.01 / -0.13	0.35 / 0.14 / 0.05 / 0.22	0.66 (0.40) / 0.93 (0.09) / 0.98 (0.05) / 0.90 (0.16)	0.77 / 0.95 / 1.00 / 0.28	0.99 (0.44)

Table 1 Results from the simulation of low-resolution data ('low res') and additional compartment ($T_{1,lesion}$ denotes the simulated lesion T1 relaxation time), for SNR=100. One standard deviation is in parenthesis, where applicable