



## Partial volume-aware assessment of multiple sclerosis lesions

Mário João Fartaria<sup>a,b,c,\*</sup>, Alexandra Todea<sup>d</sup>, Tobias Kober<sup>a,b,c</sup>, Kieran O'Brien<sup>e,f</sup>,  
Gunnar Krueger<sup>g</sup>, Reto Meuli<sup>b</sup>, Cristina Granziera<sup>h,i</sup>, Alexis Roche<sup>b,a,c,1</sup>,  
Meritxell Bach Cuadra<sup>b,j,c,1</sup>

<sup>a</sup> Advanced Clinical Imaging Technology (HC CMEA SUI DI PI), Siemens Healthcare AG, Lausanne, Switzerland

<sup>b</sup> Department of Radiology, Lausanne University Hospital (CHUV), and University of Lausanne (UNIL), Lausanne, Switzerland

<sup>c</sup> Signal Processing Laboratory (LTS 5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

<sup>d</sup> Department of Radiology, Pitié-Salpêtrière Hospital, Neuchâtel, Switzerland

<sup>e</sup> Centre for Advanced Imaging, University of Queensland, Queensland, Australia

<sup>f</sup> Siemens Healthcare Pty. Ltd., Brisbane, Queensland, Australia

<sup>g</sup> Siemens Healthcare Ltd, Zürich, Switzerland

<sup>h</sup> Neurologic Clinic and Policlinic, Departments of Medicine, Clinical Research and Biomedical Engineering, University Hospital Basel and University of Basel, Basel, Switzerland

<sup>i</sup> Translational Imaging in Neurology (ThINK) Basel, Department of Medicine and Biomedical Engineering, University Hospital Basel and University of Basel, Basel, Switzerland

<sup>j</sup> Medical Image Analysis Laboratory (MIAL), Centre d'Imagerie BioMédicale (CIBM), Lausanne, Switzerland

### ARTICLE INFO

#### Keywords:

Partial volume  
Multiple sclerosis  
MRI  
Lesion segmentation

### ABSTRACT

White-matter lesion count and volume estimation are key to the diagnosis and monitoring of multiple sclerosis (MS). Automated MS lesion segmentation methods that have been proposed in the past 20 years reach their limits when applied to patients in early disease stages characterized by low lesion load and small lesions. We propose an algorithm to automatically assess MS lesion load (number and volume) while taking into account the mixing of healthy and lesional tissue in the image voxels due to partial volume effects. The proposed method works on 3D MPRAGE and 3D FLAIR images as obtained from current routine MS clinical protocols. The method was evaluated and compared with manual segmentation on a cohort of 39 early-stage MS patients with low disability, and showed higher Dice similarity coefficients (median DSC = 0.55) and higher detection rate (median DR = 61%) than two widely used methods (median DSC = 0.50, median DR < 45%) for automated MS lesion segmentation. We argue that this is due to the higher performance in segmentation of small lesions, which are inherently prone to partial volume effects.

### 1. Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease of the central nervous system, characterized by inflammation, demyelination, axonal loss, and gliosis (Noseworthy et al., 2000). Current MS diagnostic and follow-up criteria are exploiting magnetic resonance (MR) imaging metrics of lesion load, i.e. lesion count and location, as well as activity, i.e. gadolinium enhancement due to blood-brain barrier disruption (Rovira et al., 2015). The presence and spatial pattern of focal lesions in MR images (“dissemination in space”) and the appearance of new lesions (“dissemination in time”) are key components of current diagnosis criteria (Filippi et al., 2016; Polman et al., 2011). The identification of focal pathology and of new lesions as well as their size

changes in follow-up scans are important to perform an early diagnosis, quantifying ongoing disease activity and monitor treatment effects (Filippi et al., 2016). Consequently, one current research focus has been the development of automated MS lesion segmentation (Garcia-Lorenzo et al., 2013; Lladó, Ganiler, et al., 2012; Lladó, Oliver, et al., 2012).

Automated segmentation approaches are either unsupervised or supervised. Unsupervised approaches typically apply clustering algorithms that make use of both image intensity from one or several MR contrasts (T1-weighted, T2-weighted, proton-density-weighted, and/or 2D fluid-attenuated inversion recovery, FLAIR) and spatial information derived from probabilistic atlases of healthy tissues and/or topological constraints (Schmidt et al., 2012; Shiee et al., 2008; Souplet et al., 2008; Tomas-Fernandez and Warfield, 2015; Van Leemput et al., 2001).

\* Corresponding author at: Advanced Clinical Imaging Technology (HC CMEA SUI DI PI), Siemens Healthcare AG, Lausanne, Switzerland.

E-mail address: [mario.fartaria\\_de\\_oliveira@siemens-healthineers.com](mailto:mario.fartaria_de_oliveira@siemens-healthineers.com) (M.J. Fartaria).

<sup>1</sup> The last two authors contributed equally to this work.

Supervised approaches, on the other hand, rely on manually annotated image sets used for training (Anbeek et al., 2004; Brosch et al., 2016; Fartaria, Bonnier, et al., 2016; Morra et al., 2008; Sweeney et al., 2014).

We note that existing automated MS lesion segmentation methods were mostly evaluated on patients in advanced MS stages, who have a high disability, high numbers of lesions, and large lesion volumes (Datta and Narayana, 2013; Sajja et al., 2006; Steenwijk et al., 2013; Sweeney et al., 2013). They, however, showed substantially lower performance when applied to subjects with lower disease burden, i.e. lower lesion load and lesions of smaller size and volume (Anbeek et al., 2004; Cabezas et al., 2014; Fartaria, Bonnier, et al., 2016; Steenwijk et al., 2013; Sweeney et al., 2013). Automated MS lesion segmentation seems to be barely used in clinical practice, where detecting new lesions, particularly of small size, is of key importance for early diagnosis and follow-up of MS patients (Garcia-Lorenzo et al., 2013).

We showed in previous work that small lesions are strongly affected by partial volume (PV) effects, rendering their detection, segmentation and volume estimation challenging (Fartaria, O'Brien, et al., 2017). Taking into account PV could consequently improve the detection of small lesions and the overall lesion volume estimation. This idea was investigated in the mid-90s by (Johnston et al., 1994), who proposed a semi-automated method that takes partial volumes into account using neighborhood and histogram analysis. The same group later added pre- and post-processing steps of image enhancement and mathematical morphology to improve the discrimination between healthy WM and lesions (Johnston et al., 1996). Recently, (Khademi and Moody, 2015) performed image classification using mixed tissue labels as in (Cuadra et al., 2005; Shattuck et al., 2001) and estimated the PV fraction in mixed classes using spatial image gradient analysis. Variants of this approach using hierarchical mixture models are employed in (Galimzianova et al., 2016; Sudre et al., 2015).

Here, we propose a novel method for MS lesion segmentation that relies on a Bayesian PV estimation algorithm inspired by the “mixel” model originally proposed by (Choi et al., 1991), which leads to an ill-posed estimation problem for which (Roche and Forbes, 2014) proposed regularizing priors. We further included spatial constraints to estimate realistic concentration maps of healthy tissues (WM, GM, CSF), and pathological brain tissue. These concentration maps are used to directly compute lesion volumes rather than applying a correction of PV effects in initial hard tissue classification as in previous methods (Johnston et al., 1996; Khademi and Moody, 2015; Wu et al., 2006). Our approach does not rely on edge detection and therefore has the potential to assess PV effects in small lesions without clearly defined boundaries.

## 2. Method

### 2.1. Partial volume estimation

We consider a set of  $n_c$  images of a given subject acquired from different MR image sequences and previously submitted to various pre-processing steps including alignment, bias field correction and skull stripping. Consistent with (Choi et al., 1991; Pham and Prince, 2000; Roche and Forbes, 2014; Van Leemput et al., 2003), we assume that the vector of image intensities  $y_i$  at a voxel  $i$  in the total intra-cranial mask relates to an unknown vector of tissue concentrations  $q_i$ , with  $q_{i,t} \geq 0$  (the concentration of tissue  $t$  at voxel  $i$  is  $\geq 0$ ) and  $\sum_t q_{i,t} = 1$  (the sum of the  $n_t$  tissue concentrations at a voxel  $i$  is equal to one), through the statistical relation:

$$y_i = Mq_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, V) \quad (1)$$

where  $M$  is an  $n_c \times n_t$  matrix representing the mean tissue intensities for each channel ( $n_t$  is the number of distinct tissues and  $M_{ct}$  represents the mean intensity of tissue  $t$  in channel  $c$ ), and  $N$  represents a multi-variable Normal distribution with zero mean and covariance matrix  $V = \text{diag}(\sigma_1^2, \dots, \sigma_{n_c}^2)$ . We assumed that, in each MR sequence, image

intensities are corrupted with independent stationary Gaussian white noise, as a first-order approximation to the non-central chi noise distribution that takes into account the coil combination in MRI (Larsson et al., 2003). In this work, we consider  $n_t = 4$  tissues: CSF, GM, WM and lesions, as well as  $n_c = 2$  contrasts: magnetization-prepared rapid gradient echo (MPRAGE) and 3D FLAIR.

It was recognized 25 years ago (Choi et al., 1991) that recovering the voxel-wise tissue concentrations  $q_i$  from the above multichannel image model leads to an ill-posed inverse problem if  $n_t > n_c + 1$ , as it is the case in our application. Recently, (Roche and Forbes, 2014) proposed a prior concentration model to regularize the problem when formulated via Bayesian maximum a posteriori (MAP) estimation:

$$\pi(q_1, q_2, \dots, q_{n_v}) \propto \exp \left[ \frac{-1}{2} \sum_i q_i^T A q_i - \frac{\beta}{2} \sum_{i,j \in N_i} \|q_i - q_j\|^2 \right] \quad (2)$$

where  $n_v$  is the total number of intra-cranial voxels,  $A$  is a symmetric penalty matrix with zero diagonal and positive off-diagonal elements,  $\beta$  is a positive constant, and  $N_i$  is the neighborhood of voxel  $i$  according to the 6-topology (the 6 adjacent voxels connectedness in 3-dimensions). Both the elements of  $A$  and  $\beta$  are hyperparameters to be tuned in a learning phase. While  $\beta$  controls the amount of spatial smoothness of tissue concentration maps, the purpose of  $A$  is to disentangle intensity fluctuations due to noise from PV effects. Each non-diagonal element acts as a penalty on the mixing of distinct tissues in a voxel, hence limiting spurious concentration variations when a single tissue is present. For instance, the larger  $A_{12}$ , the less likely voxels contain both CSF and GM.

We propose to generalize the prior model of (Roche and Forbes, 2014) by allowing voxel-dependent penalty matrices  $A_i$  including non-zero diagonal elements in order to penalize tissues locally. This avoids confusing GM and lesions, which have similar intensity signatures in both MPRAGE and 3D FLAIR. Specifically, let  $\pi^{GM}$  and  $\pi^{WM}$  be a probabilistic atlas-based prior probability map for the GM and WM, respectively. We set the diagonal elements of  $A_i$  corresponding to CSF, GM, WM and lesions, via:

$$A_i = \begin{pmatrix} 0 & a_1 & a_2 & a_3 \\ a_1 & a_4(1 - \pi_i^{GM}) & a_5 & a_6 \\ a_2 & a_5 & 0 & a_7 \\ a_3 & a_6 & a_7 & a_8(1 - \pi_i^{WM}) \end{pmatrix}$$

where the parameters  $a_1$  to  $a_8$  are pre-tuned with the smoothness parameter  $\beta$ , which are assumed voxel-independent in our particular implementation.

Following (Roche and Forbes, 2014), we estimate the tissue concentrations by MAP, yielding a quadratic programming problem:

$$\min_{q_1, \dots, q_{n_v}} \sum_i \left[ (y_i - M^T q_i)^T V^{-1} (y_i - M^T q_i) + q_i^T A_i q_i + \beta \sum_{j \in N_i} \|q_i - q_j\|^2 \right],$$

where each  $q_i$  is searched in the multidimensional simplex. The solution is evaluated using an iterative scheme that loops over the intra-cranial voxels, and solves for the associated concentration vector  $q_i$  with all other concentration vectors held fixed using an active set algorithm (Nocedal and Wright, 2006). This method proves very robust in practice, and typically converges in  $< 25$  iterations.

### 2.2. Imaging parameters

The noise variance matrix  $V$  is initially assumed to be zero, and is iteratively re-estimated by MAP concurrently with the tissue concentrations (see Section 2.1), yielding the update rule:

$$V = \text{diag} \left[ \frac{1}{n_v} \sum_i (y_i - M^T q_i)(y_i - M^T q_i)^T \right] \quad (3)$$

which is performed after a complete tissue concentration re-estimation loop over the intracranial voxels.

Conversely, the matrix  $M$  of mean tissue intensities is held fixed during the estimation of tissue concentrations, having been determined beforehand by histogram matching (Nyul et al., 2000) with a reference patient. Histogram matching produces a piecewise linear intensity transformation  $f$  from the reference patient image to the input image.  $M$  is estimated as  $M = f(M_{ref})$ , where  $M_{ref}$  is the mean intensity matrix of the reference patient, which in this work was determined from manually drawn regions of interest containing no PV effects for CSF, GM, WM, and lesions.

### 2.3. Hyperparameter tuning

The same reference patient used to set the tissue mean intensities  $M$  (see Section 2.2) was used to tune the hyperparameters  $A$  and  $\beta$  so as to minimize the Hellinger distance  $H$ , between the manual lesion segmentation binary mask  $M$ , and the lesion concentration map output by the PV estimation algorithm  $Q_{Les}$ :

$$H(Q_{Les}, M) = \frac{1}{\sqrt{2}} \cdot \sqrt{\sum_i (\sqrt{q_{Les,i}} - \sqrt{m_i})^2} \quad (4)$$

Two elements of  $A$  were fixed to very large values in order to prescribe mixing of CSF with WM and CSF with lesions ( $a_2 = a_3 = 1 \times 10^{10}$ ). The other parameters were optimized using Powell's method, yielding  $a_1 = 11.25$ ,  $a_4 = 14.33$ ,  $a_5 = 0.47$ ,  $a_6 = 12.21$ ,  $a_7 = 1.33$ ,  $a_8 = 16.93$ , and  $\beta = 0.54$ .

## 3. Experimental set

### 3.1. Data

The study was approved by the Ethics Committee of our institution and all subjects gave written informed consent prior to participation.

Thirty-nine patients (14 males, 25 females, median age 34 years, age range: 20–60 years) with early relapsing-remitting MS, (disease duration < 5 years from diagnosis) and Expanded Disability Status Scale (EDSS) score between 1 and 2 (median EDSS = 1.5), were scanned on a 3 T MRI system (Magnetom Trio a Tim System, Siemens Healthcare GmbH, Erlangen, Germany) using a 32-channel head coil. The MRI protocol included: (i) high-resolution MPRAGE (TR/TE/TI = 2300/900 ms, voxel size =  $1 \times 1 \times 1.2 \text{ mm}^3$ ), and (ii) 3D FLAIR (TR/TE/TI = 5000/394/1800 ms, voxel size =  $1 \times 1 \times 1.2 \text{ mm}^3$ ), (iii) magnetization-prepared 2 rapid acquisition gradient echo (MP2RAGE, TR/TE/TI1/TI2 = 5000/2.89/700/2500 ms, voxel size =  $1 \times 1 \times 1.2 \text{ mm}^3$ ), all acquired in the same session without patient repositioning. MS lesions were identified and marked by one radiologist and one neurologist (6 and 11 years of experience, respectively) by consensus using FLAIR and MP2RAGE images. A trained technician then delineated the lesion volumes in each image, which we considered to be the reference for lesion load and volume (Bonnier et al., 2015; Kober et al., 2012).

### 3.2. Pre-processing

A patient with relatively high lesion load was chosen as a reference to train the PV estimation algorithm (see Sections 2.2 and 2.3) and was subsequently excluded from the ensuing statistical analysis. ELASTIX (Klein et al., 2010) was used to confirm that the rigid transformation relating FLAIR and MPRAGE was below 1 voxel according to the resulted transformation parameters. All images were further skull-stripped using an in-house method (Schmitter et al., 2014), and corrected for intensity inhomogeneities using the N4 algorithm (Tustison et al., 2010). Fuzzy in-house templates of prior WM and GM probabilities were non-rigidly registered using ELASTIX onto each image volume to produce the prior maps  $\pi^{GM}$  and  $\pi^{WM}$  (see Section 2.1). Such

priors were obtained using the DARTEL tool of SPM8 (Ashburner and Friston, 2009), from a data set of 136 MR scans of healthy subjects acquired at our institution.

### 3.3. Comparison with freely available software

We compared the results obtained using our method on the cohort of early MS patients with two widely used and freely available WM lesion segmentation methods. One method, the Lesion Segmentation Tool (LST), is based on unsupervised outlier rejection and region growing, and is distributed as an SPM toolbox (Schmidt et al., 2012). The other method, LesionTOADS (LTOADS), implements an unsupervised atlas-based fuzzy clustering and is distributed as a plug-in for MIPAV (Shiee et al., 2010). Default parameters were used in both methods, except for the LST where we set the initial threshold ( $\kappa = 0.05$ ) based on the best Dice similarity coefficient (DSC) values obtained in the cohort (Schmidt et al., 2012).

## 4. Validation and results

### 4.1. Lesion detection and segmentation

In this section, we evaluate lesion detection and segmentation by the detection rate (DR), false positive rate (FPR), and Dice similarity coefficient (DSC). These metrics rely on the overlap between the automated segmentation and the reference. Fig. 1 illustrates the lesion maps obtained using the different considered methods.

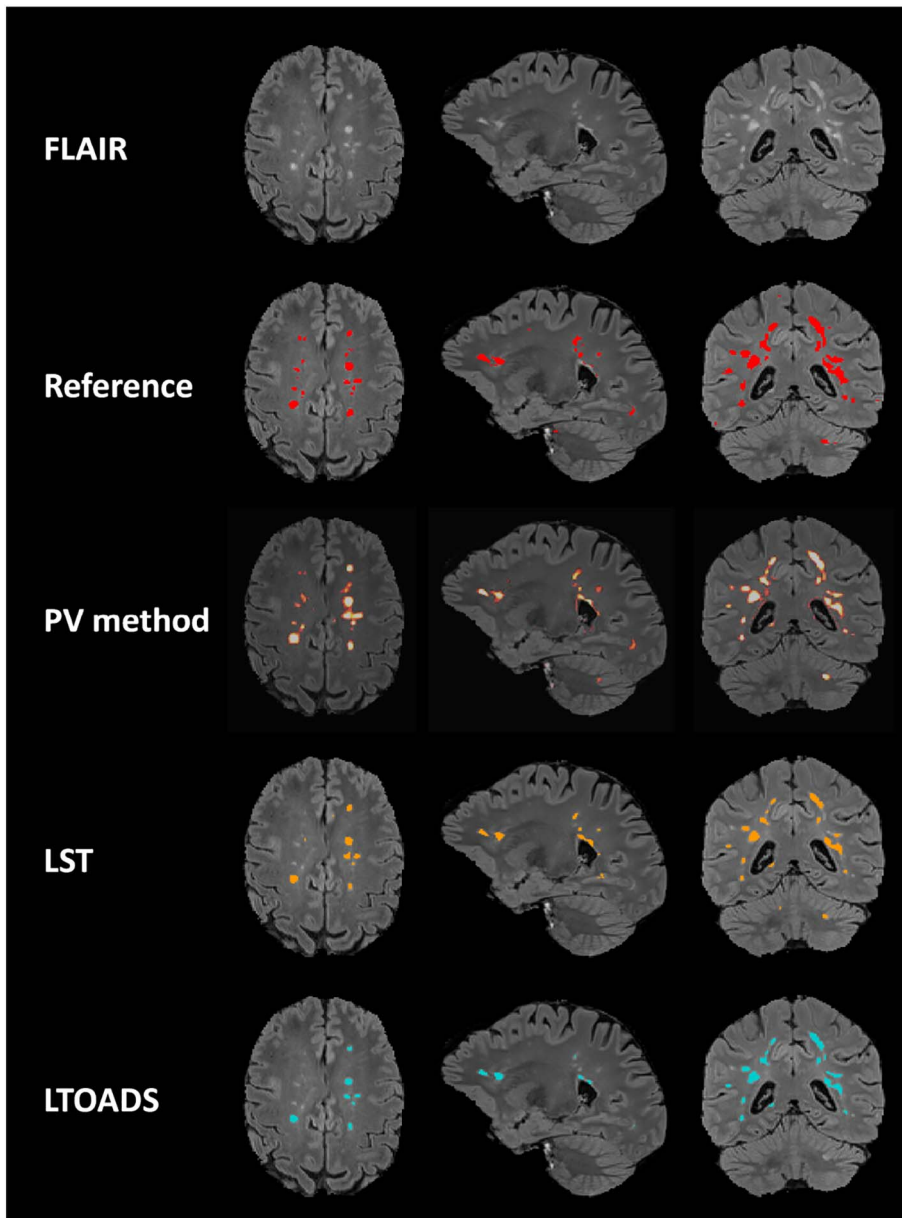
Fig. 2 shows an example of a small and a large lesion, where the PV effect becomes apparent. The small lesion is substantially affected by partial volume, having a maximum voxel concentration of 55% lesional tissue (45% WM concentration for the same voxel) according to the PV-method. In the large lesion, PV is evident on the lesion border, where the voxel concentration of lesional tissue varies between 12% and 98%. The lesion centre has 100% of concentration of lesional tissue, according to the PV-method, indicating “pure” lesion signal.

Since our PV approach provides a concentration map of lesional tissue, a threshold needs to be applied for comparison with the binary masks provided by the other methods. Two lesion concentration map thresholds were used to match the median FPR (across the entire dataset) of the PV-method with LST and LTOADS, respectively. The FPR was obtained by dividing the number of lesions in the automated segmentation that do not overlap with any lesion in the reference with the number of overall lesions in the automated segmentation (Styner et al., 2008). An optimal concentration map threshold was also derived from a receiver operating characteristic (ROC) analysis based on the best trade-off between DR and FPR.

Diagnostic criteria define a minimal diameter of 3 mm for MS lesion (Chris H. Polman et al., 2005). Recently, the meaningful of this diameter size was proved for 3D sequences at 3T (Grahel et al., 2017). If we approximate the lesion shape to a sphere, we end up to a minimum lesion volume of  $\approx 15 \mu\text{L}$ . However, applying this threshold remove most of the lesions that are fully prone to PV effects. In this work we presented the results using two different definitions of minimum lesion volume:  $\approx 3 \mu\text{L}$  and  $\approx 15 \mu\text{L}$ . Our rationale in applying a threshold of  $\approx 3 \mu\text{L}$  was to validate the performance of the methods for small lesions fully prone to PV effects, that might be relevant for early diagnosis and follow-up.

We evaluated lesion detection (knowing that lesion number is the main MRI biomarker for diagnosis in MS) through the DR, defined by dividing the number of lesions in the automated segmentation that overlap with a lesion in the reference, with the number of overall lesions in the reference (Styner et al., 2008). The lesion segmentation agreement with the reference was evaluated using DSC:

$$\text{DSC} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (6)$$



**Fig. 1.** Axial, sagittal, and coronal views from three different patients showing the detection and segmentation results obtained from different methods. From top to bottom. 3D FLAIR slice, manual segmentation (reference), proposed partial volume (PV) method, and two freely available software: Lesion Segmentation Tool (LST), and LesionTOADS (LTOADS).

where *TP*: number of true positives; *FP*: number of false positives; *FN*: number of false negatives; (“true” relates to marked voxels in the reference). The Wilcoxon signed-rank tests were used to compare the DSC results between methods.

To match the FPR of LST (median FPR = 51%) and LTOADS (median FPR = 42%), thresholds of 32% and 38% were applied to the lesion concentration maps, respectively. We found that the optimal threshold for the generation of the binary lesion maps, according to the ROC analysis was 40%. PV-method presented the better detection performance for any lesion size category for all the applied thresholds to the concentration maps (Fig. 3). For a threshold of 32%, PV-method achieved better performance than LST in all lesion size categories (3–14  $\mu\text{L}$ : 30.1%, 13.8%; 15–20  $\mu\text{L}$ : 60.5%, 22.6%; 21–50  $\mu\text{L}$ : 76.5%, 50.8%; 51–100  $\mu\text{L}$ : 94.7%, 86.8%. DR values for PV-method and LST, respectively.), except for big lesions (lesion volume > 100  $\mu\text{L}$ ), where both methods detected 98.5% of the lesions. Overall, PV-method achieved a DR = 60.7%, against the 43.4% obtained from LST. For a threshold of 38%, PV-method achieved higher performance when compared to LTOADS in all lesion size categories (3–14  $\mu\text{L}$ : 22.8%, 12.9%; 15–20  $\mu\text{L}$ : 56.3%, 17.9%; 21–50  $\mu\text{L}$ : 72.7%, 45.3%; 51–100  $\mu\text{L}$ :

93%, 73.2%; > 100  $\mu\text{L}$ : 98.5%, 84.6%. DR values for PV-method and LTOADS, respectively.). Overall, PV-method achieved a DR = 56.2%, against the 37.8% obtained from LTOADS. Similar behaviour was found, when the optimal threshold of 40% was applied to the concentration maps, where PV-method reached a overall DR of 54.5%. Although the DR is lower compared to the values obtained with other thresholds, the PV-method still showed the best performance when compared to LST and LTOADS. For a threshold of 32% and considering a minimum lesion volume of  $\approx 3.2 \mu\text{L}$ , PV-method achieved a median DSC of 0.55 significantly superior ( $P < .001$ ) to LST (median DSC of 0.50, Fig. 4A). The same behaviour was observed when a threshold of 38% was applied to the PV map, the method presented the same median DSC of 0.55, which was significantly ( $P < .05$ ) higher than LTOADS (median DSC of 0.51, Fig. 4A). When the optimal threshold of 40% was applied, PV-method achieved a median DSC of 0.53 significantly superior ( $P < .001$ ) to LST but non-significantly different from LTOADS ( $P = .06$ , Fig. 4A). Lastly, similar behaviour was found when the minimum lesion volume was set to  $\approx 15 \mu\text{L}$ . However, median DSC values appeared to be higher in all the methods: 0.53 for PV-method with a threshold of 40%; 0.59 for PV-method with a threshold of 32%;

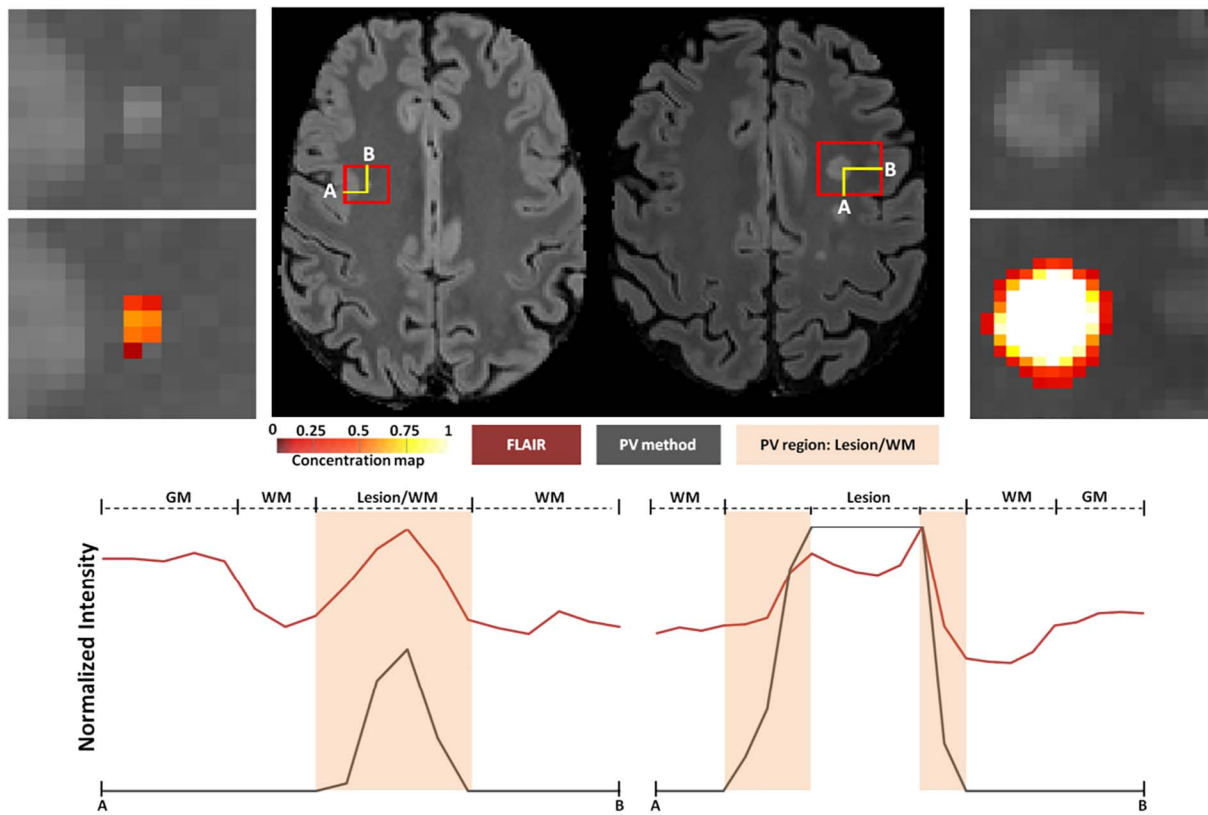


Fig. 2. Partial volume effect in a small (left side), and large (right side) lesion. The small lesion is fully affected by partial volume, having a maximum concentration of 55% of lesional tissue in the path A to B, according to the proposed partial volume (PV) method. The big lesion is prone to partial volume effects in the border, detected as low lesion concentration by the proposed PV-method. In this case, effects of partial volume in the transitions between healthy and lesional tissue are clearly seen. Intensity profiles of the path A to B, for FLAIR intensities, and PV lesions maps where shown on the bottom. Smooth orange regions in the profile plots represent regions of partial volume (mix of lesional tissue and WM tissue).

0.57 for PV-method with a threshold of 38%; 0.52 for LST; and 0.52 for LTOADS.

#### 4.2. Total lesion volume

The total lesion volume (TLV) per patient was estimated using each method, and compared with the reference using the two definitions of minimum lesion volume ( $\approx 3 \mu\text{L}$  and  $\approx 15 \mu\text{L}$ , Fig. 5). The Wilcoxon signed-rank test was used to compare the results between methods. When the minimum lesion volume was set to  $\approx 3 \mu\text{L}$ , the PV-method showed the best agreement in terms of TLV compared to the other methods ( $P < .001$ ). Similar behaviour was found when the minimum lesion volume was set to  $15 \mu\text{L}$ .

Correlations between the reference and each method, for both definitions of minimum lesion volume, were also studied using the

Spearman's rank correlation coefficient ( $\rho_s$ ) and the coefficient of determination (R-squared,  $R^2$ ). As shown in Fig. 6, the PV-method and LST methods were found to be more highly correlated with the manually determined TLVs than the LTOADS method. For a minimal lesion volume of  $\approx 3 \mu\text{L}$ , the overall Spearman's rank correlation coefficient was  $\rho_s = 0.94$  for the proposed method,  $\rho_s = 0.90$  for LST, and  $\rho_s = 0.84$  for LTOADS. Both LST and LTOADS presented a higher underestimation of the TLVs for patients with high/moderate lesion loads, as reflected by the coefficient of determination with respect to the identity line:  $R^2 = 0.33$  for LST and  $R^2 = 0.27$  for LTOADS, against  $R^2 = 0.76$  for the proposed PV-method. Similar results were observed when minimal lesion volume was set to  $\approx 15 \mu\text{L}$  (see Fig. 6).

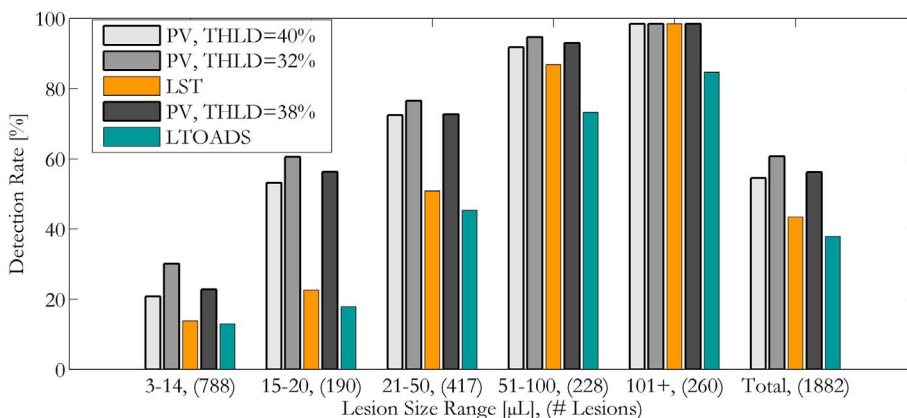
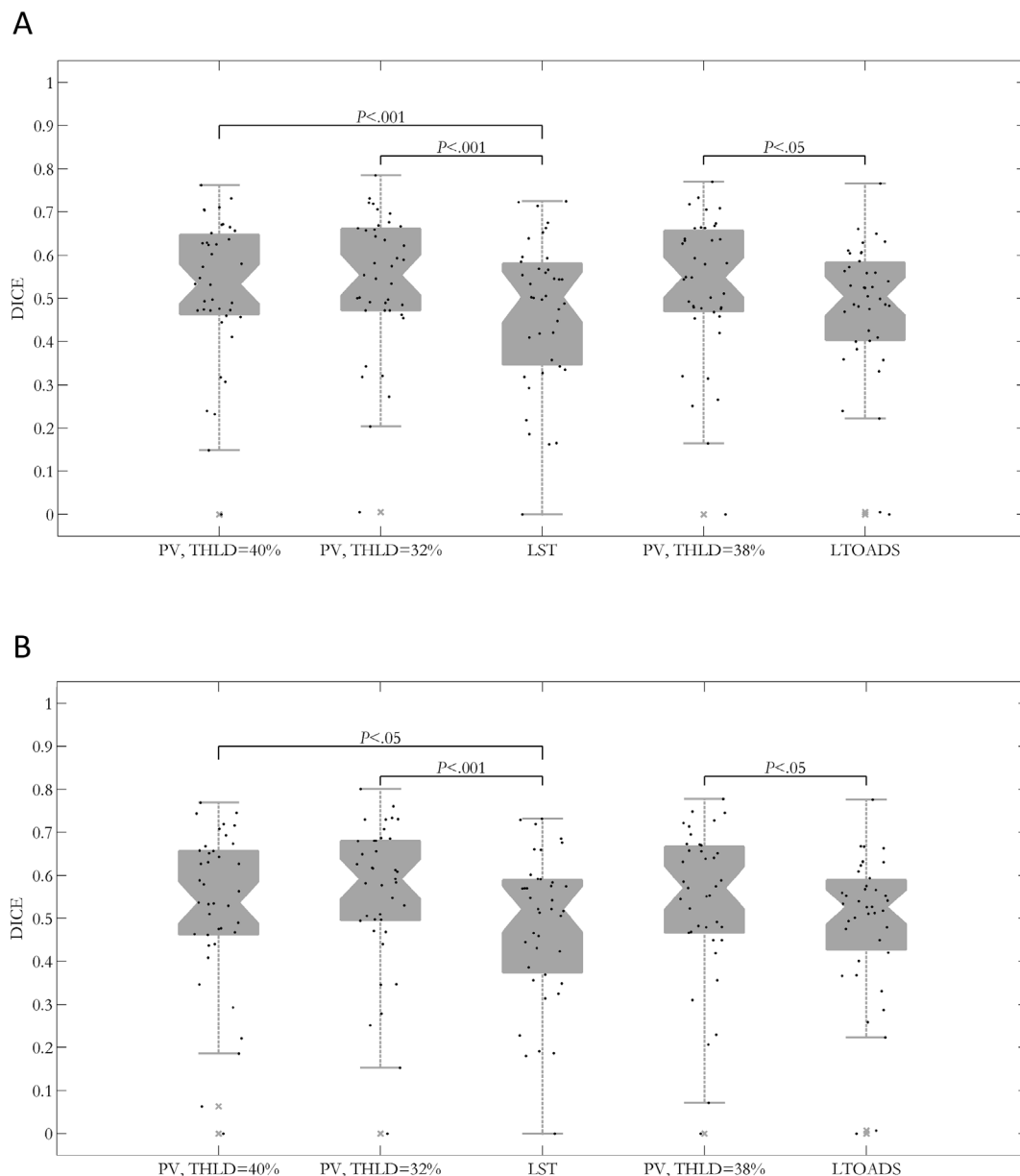


Fig. 3. Detection rate for different ranges of lesion size obtained using the proposed partial volume (PV) method (using three different thresholds, THLD), Lesion Segmentation Tool (LST), and LesionTOADS (LTOADS). Numbers in parenthesis represent the number of lesions for the respective lesion size range according to the reference.



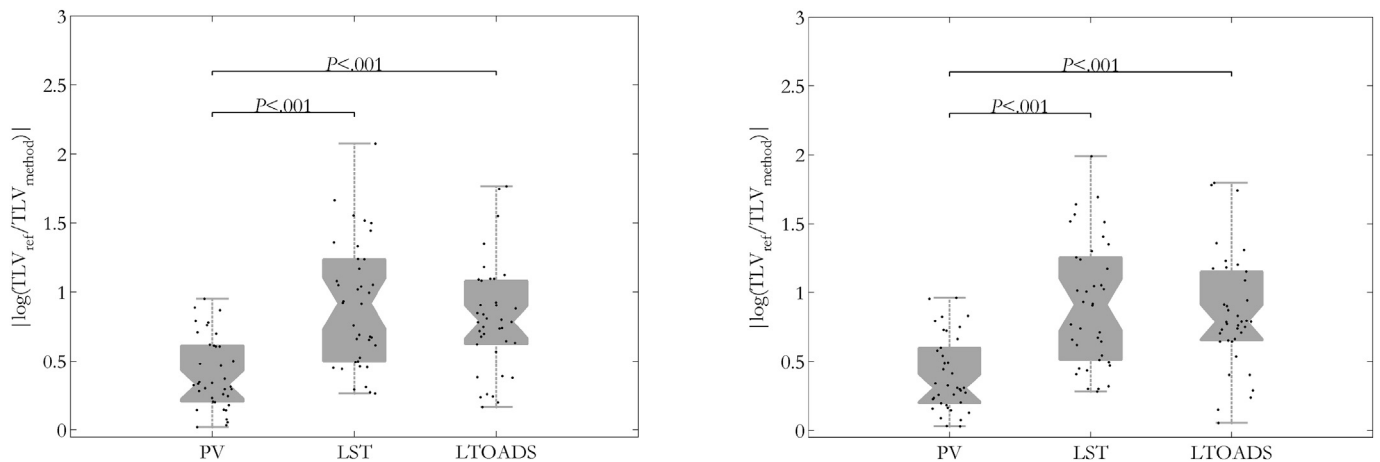
**Fig. 4.** Boxplots of Dice similarity coefficient between manual segmentation and automated lesion segmentation using the proposed partial volume (PV) method (using three different thresholds, THLD), Lesion Segmentation Tool (LST), and LesionTOADS (LTOADS). Results were computed using two different definitions of minimum lesion volume:  $\approx 3 \mu\text{L}$  (panel A) and  $\approx 15 \mu\text{L}$  (panel B). The crosses in the plot represent outliers in the cohort.

## 5. Discussion

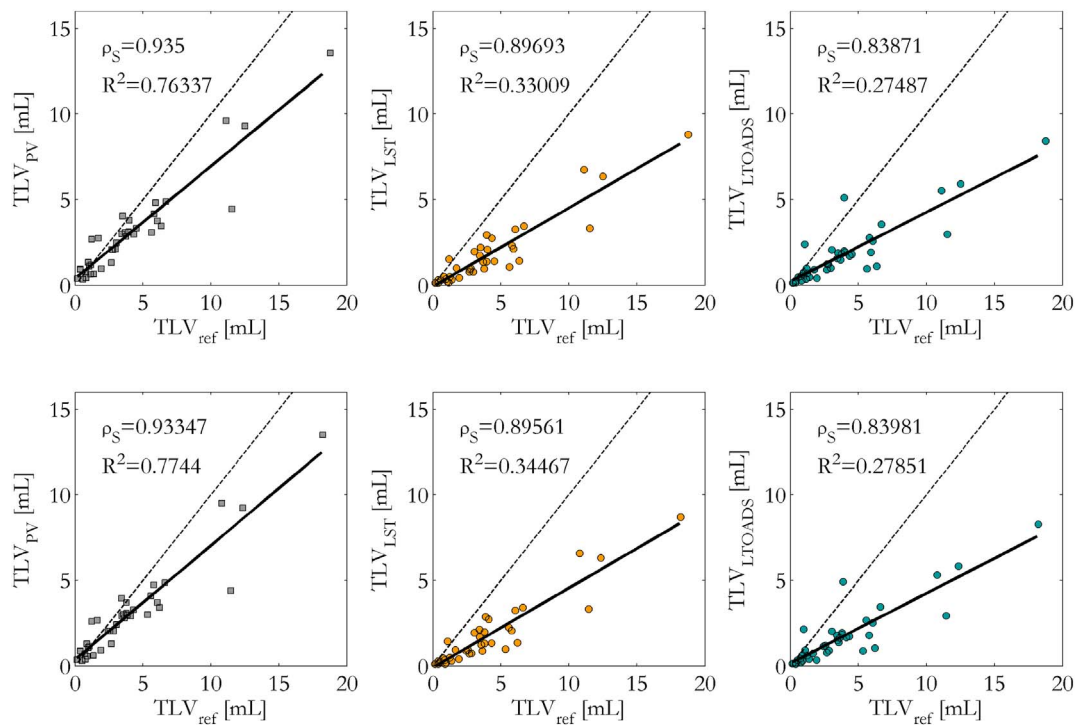
We developed and validated a method to automatically detect and segment MS lesions in patients at an early disease stage and with low disability scores. Our method is based on a Bayesian PV estimation using the “mixel” model (Choi et al., 1991; Roche and Forbes, 2014), extended to lesion detection by including spatial constraints from atlas-based probability maps of GM and WM. Our approach was developed and optimized for images from the routine MS clinical protocol in our institution (3D MPRAGE and 3D FLAIR). Other clinical images, such as T2-weighted or proton density, may be also used in our multi-channel approach, although the hyper parameters require to be tuned accordingly. Most MS lesion segmentation methods reported in the literature show low performance when applied to patient data exhibiting low lesion load and small lesions (Anbeek et al., 2004; Cabezas et al., 2014; Galimzianova et al., 2016; Sajja et al., 2006; Schmidt et al., 2012; Steenwijk et al., 2013). However, accurate detection especially in the

early stages of MS can be crucial for initial diagnosis and subsequent treatment monitoring (Filippi et al., 2016). Here, our goal was to improve the detection of very small lesions that are partially or totally affected by PV effects. Moreover, we aimed to better estimate lesion volume through better delineation of lesion borders, which are prone to PV effects.

Our results showed the feasibility of the proposed method to detect and segment WM MS lesions, including lesions of small size, as well as the importance of taking into account PV for lesion detection, lesion segmentation, and lesion volume estimation. We considered two different definitions of minimal lesion size: 1) the one proposed by the MS diagnostic criteria (Polman et al., 2005) of 3 mm diameter and 2) a smaller size of  $\approx 3 \mu\text{L}$  in volume in order to include small lesions fully prone to PV effects. PV-method showed the best results, when compared to LST and LTOADS, for both definitions of minimum lesion size. However, the improvements were more evident for a minimum lesion volume of  $\approx 3 \mu\text{L}$ , where detection of small lesions by the proposed



**Fig. 5.** Boxplots of total lesion volume (TLV) difference between reference and automated lesion segmentation using the proposed partial volume (PV) method, and two freely available software: Lesion Segmentation Tool (LST), and LesionTOADS (LTOADS). Results were computed using two different definitions of minimum lesion volume:  $\approx 3 \mu\text{L}$  (left panel) and  $\approx 15 \mu\text{L}$  (right panel). The crosses in the plot represent outliers in the cohort.



**Fig. 6.** Total lesion volume (TLV) correlation between the reference and automated lesion segmentation (from left to right) using the proposed partial volume (PV) method, and two freely available software: Lesion Segmentation Tool (LST), and LesionTOADS (LTOADS). Results were computed using two different definitions of minimum lesion volume:  $\approx 3 \mu\text{L}$  (first row) and  $\approx 15 \mu\text{L}$  (second row). The respective Spearman's rank correlation coefficient ( $\rho_s$ ), and coefficient of determination ( $R^2$ ) are given. The dash line represents the identity line ( $TLV_{reference} = TLV_{method}$ ), and the solid line represents the best linear fit to the points.

method significantly improved the total lesion volume estimation when compared to LST, and LTOADS.

DSC is one of the most commonly-used metrics to measure the segmentation quality (Lladó, Oliver, et al., 2012); however, this metric is not appropriate to evaluate the segmentation of small objects (Schmidt et al., 2012). Despite this, DSC is widely used (Brosch et al., 2016; Galimzianova et al., 2016; Jain et al., 2015) even to this type of data characterized with low lesion load. Interestingly, the median of DSC for the PV-method was significantly higher than for the other two studied methods.

Our study shows that modelling the PV effect improves volumetric measurements as well as the detection of small lesions, which is of particular importance for MS patients in an early disease phase. The clinical relevance of lesions with diameter smaller than 3 mm is not yet

established. However, their counting can be important not only for early-diagnosis but also for follow-up assessment (new lesion count) and to monitor the treatment response.

A limitation of our method is the detection of “black holes” which are hypointense in FLAIR images. Since the method relies on image intensity, hypointense voxels in FLAIR images from “black holes” will most likely be classified as CSF. In this particular data set of early MS stages, there were few “black holes”, however future developments based on CSF prior information should be done and tested in data sets of patients at later MS stages and with higher disability. Furthermore, the method should be tested on images from different scanners and protocols. This will enable us to test the robustness of the histogram matching approach (see Section 2.2) used to homogenise the intensity scale across images from different acquisitions. Lastly, it should be

considered that the tuning of the hyper-parameters was performed using only one patient as a reference. This reference was found to be the most representative case in terms of lesion size and distribution. However, future validation should be performed in order to evaluate the influence of the reference patient on hyper-parameter tuning and consequently on lesion segmentation performance.

Extensions of the presented method should focus on exploring mathematical morphology operations to better disentangle GM from lesions, and consequently improve the lesion detection. This could particularly improve the detection in regions where the method presents a relatively modest performance due to the lesion location and very similar intensities between lesional and normal-appearing GM tissue, as e.g. seen in juxtacortical lesions. Future work will also aim to extend the method to cortical lesion detection, by incorporating input from advanced sequences like double-inversion recovery (DIR), and magnetization-prepared 2 rapid acquisition with gradient echo (MP2RAGE), which has been recently shown to be more sensitive to cortical lesions (Fartaria, Bonnier, et al., 2016).

## References

- Anbeek, P., Vincken, K.L., van Osch, M.J., Bisschops, R.H., van der Grond, J., 2004. Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage* 21 (3), 1037–1044. <http://dx.doi.org/10.1016/j.neuroimage.2003.10.012>.
- Ashburner, John, Friston, Karl J., 2009. Computing average shaped tissue probability templates. *NeuroImage* 45 (2), 333–341.
- Bonnier, Guillaume, Roche, Alexis, Romascano, David, Simioni, Samanta, Meskaldji, Djalel Eddine, Rotzinger, David, et al., Du Pasquier, Renaud, 2015. Multicontrast MRI quantification of focal inflammation and degeneration in multiple sclerosis. *Biomed. Res. Int.* 2015.
- Brosch, Tom, Tang, Lisa Y.W., Yoo, Youngjin, Li, David K.B., Traboulsee, Anthony, Tam, Roger, 2016. Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Trans. Med. Imaging* 35 (5), 1229–1239.
- Cabezas, M., Oliver, A., Roura, E., Freixenet, J., Vilanova, J.C., Ramio-Torrenta, L., et al., Lladó, X., 2014. Automatic multiple sclerosis lesion detection in brain MRI by FLAIR thresholding. *Comput. Methods Prog. Biomed.* 115 (3), 147–161. <http://dx.doi.org/10.1016/j.cmpb.2014.04.006>.
- Choi, Hwan Soo, Haynor, David R., Kim, Yongmin, 1991. Partial volume tissue classification of multichannel magnetic resonance images—a mixel model. *IEEE Trans. Med. Imaging* 10 (3), 395–407.
- Cuadra, Meritxell Bach, Cammoun, Leila, Butz, Torsten, Cuisenaire, Olivier, Thiran, J.-P., 2005. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Trans. Med. Imaging* 24 (12), 1548–1565.
- Datta, S., Narayana, P.A., 2013. A comprehensive approach to the segmentation of multichannel three-dimensional MR brain images in multiple sclerosis. *Neuroimage Clin.* 2, 184–196. <http://dx.doi.org/10.1016/j.nicl.2012.12.007>.
- Fartaria, Mário João, Bonnier, Guillaume, Roche, Alexis, Kober, Tobias, Meuli, Reto, Rotzinger, David, et al., Granziera, Cristina, 2016. Automated detection of white matter and cortical lesions in early stages of multiple sclerosis. *J. Magn. Reson. Imaging* 43 (6), 1445–1454. <http://dx.doi.org/10.1002/jmri.25095>.
- Fartaria, Mário João, O'Brien, Kieran, Sorega, Alexandra, Bonnier, Guillaume, Roche, Alexis, Falkovskiy, Pavel, ... Granziera, Cristina, 2017. An ultra-high field study of cerebellar pathology in early relapsing-remitting multiple sclerosis using MP2RAGE. *Investig. Radiol.* 52 (5), 265–273. <http://dx.doi.org/10.1097/RLI.0000000000000338>.
- Filippi, Massimo, Rocca, Maria A., Ciccirelli, Olga, De Stefano, Nicola, Evangelou, Nikos, Kappos, Ludwig, et al., Frederiksen, Jette L., 2016. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. *Lancet Neurol.* 15 (3), 292–303.
- Galimzianova, Alfiya, Pernuš, Franjo, Likar, Boštjan, Špiclin, Žiga, 2016. Stratified mixture modeling for segmentation of white-matter lesions in brain MR images. *NeuroImage* 124, 1031–1043.
- García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D.L., Collins, D.L., 2013. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Med. Image Anal.* 17 (1), 1–18. <http://dx.doi.org/10.1016/j.media.2012.09.004>.
- Grahl, S., Pongratz née Biberacher, V., Schmidt, P., Engl, C., Bussas, M., Radetz, A., et al., Muhlau, M., 2017. Defining a Minimal Meaningful Lesion Size in Multiple Sclerosis. European Committee for Treatment and Research in Multiple Sclerosis -ECTRIMS.
- Jain, Saurabh, Sima, Diana M., Ribbens, Annemie, Cambron, Melissa, Maertens, Anke, Van Hecke, Wim, et al., Daams, Marita, 2015. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *Neuroimage Clin.* 8, 367–375.
- Johnston, Brian, Atkins, M. Stella, Booth, Kellogg S., 1994. Partial Volume Segmentation in 3D of Lesions and Tissues in Magnetic Resonance Images. Paper Presented at the Medical Imaging 1994.
- Johnston, B., Atkins, M.S., Mackiewicz, B., Anderson, M., 1996. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. *IEEE Trans. Med. Imaging* 15 (2), 154–169. <http://dx.doi.org/10.1109/42.491417>.
- Khademi, April, Moody, Alan R., 2015. Multiscale Partial Volume Estimation for Segmentation of White Matter Lesions Using Flair MRI. Paper Presented at the 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI).
- Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2010. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* 29 (1), 196–205. <http://dx.doi.org/10.1109/TMI.2009.2035616>.
- Kober, T., Granziera, C., Ribes, D., Browaeys, P., Schlupe, M., Meuli, R., et al., Krueger, G., 2012. MP2RAGE multiple sclerosis magnetic resonance imaging at 3 T. *Investig. Radiol.* 47 (6), 346–352. <http://dx.doi.org/10.1097/RLI.0b013e31824600e9>.
- Larsson, Erik G., Erdogmus, Deniz, Yan, Rui, Principe, Jose C., Fitzsimmons, Jeffrey R., 2003. SNR-optimality of sum-of-squares reconstruction for phased-array magnetic resonance imaging. *J. Magn. Reson.* 163 (1), 121–123.
- Lladó, Xavier, Ganiler, Onur, Oliver, Arnau, Cabezas, Mariano, Freixenet, Jordi, Valls, Laia, et al., Rovira, Àlex, 2012. Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology* 54 (8), 787–807.
- Lladó, Xavier, Oliver, Arnau, Cabezas, Mariano, Freixenet, Jordi, Vilanova, Joan C., Quiles, Ana, et al., Rovira, Àlex, 2012. Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches. *Inf. Sci.* 186 (1), 164–185. <http://dx.doi.org/10.1016/j.ins.2011.10.011>.
- Morra, J., Tu, Zhuowen, Toga, A., Thompson, P., 2008. Automatic segmentation of MS lesions using a contextual model for the MICCAI grand challenge. In: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, pp. 1–7.
- Nocedal, Jorge, Wright, Stephen J., 2006. *Numerical Optimization*, 2nd ed. Springer.
- Noseworthy, John H., Lucchinetti, Claudia, Rodriguez, Moses, Weinschenker, Brian G., 2000. Multiple Sclerosis. *N. Engl. J. Med.* 343 (13), 938–952. <http://dx.doi.org/10.1056/NEJM200009283431307>.
- Nyul, L.G., Udupa, J.K., Zhang, X., 2000. New variants of a method of MRI scale standardization. *IEEE Trans. Med. Imaging* 19 (2), 143–150. <http://dx.doi.org/10.1109/42.836373>.
- Pham, Dzung L., Prince, Jerry L., 2000. Unsupervised Partial Volume Estimation in Single-channel Image Data. Paper Presented at the Mathematical Methods in Biomedical Image Analysis, 2000. Proceedings. IEEE Workshop on.
- Polman, Chris H., Reingold, Stephen C., Edan, Gilles, Filippi, Massimo, Hartung, Hans-Peter, Kappos, Ludwig, et al., Wolinsky, Jerry S., 2005. Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald Criteria”. *Ann. Neurol.* 58 (6), 840–846. <http://dx.doi.org/10.1002/ana.20703>.
- Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M., et al., Wolinsky, J.S., 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann. Neurol.* 69 (2), 292–302. <http://dx.doi.org/10.1002/ana.22366>.
- Roche, Alexis, Forbes, Florence, 2014. *Partial Volume Estimation in Brain MRI Revisited Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014*. Springer, pp. 771–778.
- Rovira, Alex, Wattjes, Mike P., Tintore, Mar, Tur, Carmen, Yousry, Tarek A., Sormani, Maria P., ... on behalf of the Magnims study group, 2015. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—clinical implementation in the diagnostic process. *Nat. Rev. Neurol.* <http://dx.doi.org/10.1038/nrneuro.2015.106>. (advance online publication).
- Sajja, B.R., Datta, S., He, R.J., Mehta, M., Gupta, R.K., Wolinsky, J.S., Narayana, P.A., 2006. Unified approach for multiple sclerosis lesion segmentation on brain MRI. *Ann. Biomed. Eng.* 34 (1), 142–151. <http://dx.doi.org/10.1007/s10439-005-9009-0>.
- Schmidt, P., Gaser, C., Arsic, M., Buck, D., Forschler, A., Berthele, A., ... Muhlau, M., 2012. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage* 59 (4), 3774–3783. <http://dx.doi.org/10.1016/j.neuroimage.2011.11.032>.
- Schmitter, Daniel, Roche, Alexis, Maréchal, Bénédicte, Ribes, Delphine, Abdulkadir, Ahmed, Bach-Cuadra, Meritxell, ... Maeder, Philippe, 2014. An evaluation of volume-based morphology for prediction of mild cognitive impairment and Alzheimer's disease. *Neuroimage Clin.* 7, 7–17.
- Shattuck, David W., Sandor-Leahy, Stephanie R., Schaper, Kirt A., Rottenberg, David A., Leahy, Richard M., 2001. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* 13 (5), 856–876.
- Shiee, Navid, Bazin, P., Pham, Dzung L., 2008. Multiple sclerosis lesion segmentation using statistical and topological atlases. In: *Grand Challenge Work.: Mult. Scler. Lesion Segm. Challenge*, pp. 1–10.
- Shiee, N., Bazin, P.L., Ozturk, A., Reich, D.S., Calabresi, P.A., Pham, D.L., 2010. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage* 49 (2), 1524–1535. <http://dx.doi.org/10.1016/j.neuroimage.2009.09.005>.
- Souplet, Jean-Christophe, Lebrun, Christine, Ayache, Nicholas, Malandain, Grégoire, 2008. An automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions. Paper Presented at the The MIDAS Journal-MS Lesion Segmentation (MICCAI 2008 Workshop).
- Steenwijk, M.D., Pouwels, P.J., Daams, M., van Dalen, J.W., Caan, M.W., Richard, E., ... Vrenken, H., 2013. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). *Neuroimage Clin* 3, 462–469. <http://dx.doi.org/10.1016/j.nicl.2013.10.003>.
- Styner, Martin, Lee, Joohwi, Chin, Brian, Chin, M., Commowick, Olivier, Tran, H., ... Warfield, S., 2008. 3D segmentation in the clinic: a grand challenge II: MS lesion segmentation. *MIDAS J.* 2008, 1–6.
- Sudre, Carole H., Cardoso, M. Jorge, Bouvy, Willem H., Biessels, Geert Jan, Barnes, Josephine, Ourselin, Sebastien, 2015. Bayesian model selection for pathological neuroimaging data applied to white matter lesion segmentation. *IEEE Trans. Med. Imaging* 34 (10), 2079–2102.
- Sweeney, E.M., Shinohara, R.T., Shiee, N., Mateen, F.J., Chudgar, A.A., Cuzzocreo, J.L., ...



- Crainiceanu, C.M., 2013. OASIS is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *Neuroimage Clin* 2, 402–413. <http://dx.doi.org/10.1016/j.nicl.2013.03.002>.
- Sweeney, E.M., Vogelstein, J.T., Cuzzocreo, J.L., Calabresi, P.A., Reich, D.S., Crainiceanu, C.M., Shinohara, R.T., 2014. A comparison of supervised machine learning algorithms and feature vectors for MS lesion segmentation using multimodal structural MRI. *PLoS One* 9 (4), e95753. <http://dx.doi.org/10.1371/journal.pone.0095753>.
- Tomas-Fernandez, Xavier, Warfield, Simon Keith, 2015. *A Model of Population and Subject (MOPS) Intensities With Application to Multiple Sclerosis Lesion Segmentation*.
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 29 (6), 1310–1320. <http://dx.doi.org/10.1109/TMI.2010.2046908>.
- Van Leemput, Koen, Maes, Frederik, Vandermeulen, Dirk, Colchester, Alan, Suetens, Paul, 2001. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans. Med. Imaging* 20 (8), 677–688.
- Van Leemput, Koen, Maes, Frederik, Vandermeulen, Dirk, Suetens, Paul, 2003. A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans. Med. Imaging* 22 (1), 105–119.
- Wu, Y., Warfield, S.K., Tan, I.L., Wells 3rd, W.M., Meier, D.S., van Schijndel, R.A., et al., Guttman, C.R., 2006. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. *NeuroImage* 32 (3), 1205–1215. <http://dx.doi.org/10.1016/j.neuroimage.2006.04.211>.