

# Acknowledgements



# Publications

The presented thesis is a compendium of the following research articles:

- **Sergi Valverde**, Eloy Roura, Arnau Oliver, and Xavier Lladó. Automated brain tissue segmentation of MR images in the presence of white matter lesions. *Submitted to Medical Image analysis*. 2016.
- **Sergi Valverde**, Arnau Oliver, Eloy Roura, Deborah Pareto, Joan Carles Vilanova, Lluís Ramió-Torrentà, Jaume Sastre-Garriga, Xavier Montalban, Àlex Rovira and Xavier Lladó. Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling SPM8 toolboxes. *NeuroImage Clinical*. 2016. Elsevier. [JCR N IF 2.526, Q2(6/14)].
- **Sergi Valverde**, Arnau Oliver, Yago Díez, Mariano Cabezas, Joan Carles Vilanova, Lluís Ramió-Torrentà, Àlex Rovira, and Xavier Lladó. Evaluating the effects of white matter multiple sclerosis lesions on the volume estimation of six brain tissue segmentation methods. *American Journal of Neuroradiology*. Vol. 36(6), pp. 1109-1115. 2015. American Society of Neuroradiology. [JCR RNMMI IF 3.589, Q1(19/125)].
- **Sergi Valverde**, Arnau Oliver, Mariano Cabezas, Eloy Roura and Xavier Lladó. Comparison of ten brain tissue segmentation methods using revisited IBSR annotations. *Journal of Magnetic Resonance Imaging*. Vol. 41, Issue 1, pp. 93-101. January 2015. DOI: 10.1002/jmri.24517. Wiley. [JCR RNMMI IF: 3.210 Q1(23/125)].
- **Sergi Valverde**, Arnau Oliver, and Xavier Lladó. A white matter lesion-filling approach to improve brain tissue volume measurements. *NeuroImage: Clinical*. Vol. 6, pp 86-92. 2014. Elsevier. [JCR N IF 2.526, Q2(6/14)].

The rest of publications derived from this PhD thesis are the following:

## Journals

- Eloy Roura, Nicolae Sarbu, Arnau Oliver, **Sergi Valverde**, Sandra Gonzàlez, Ricard Cervera, Nuria Bargalló and Xavier Lladó. Automated detection of Lupus white matter lesions in MRI images. *Submitted to Frontiers in Human Neuroscience*. 2016. *Frontiers*. [JCR PS IF 3.626, Q1(13/76)].
- Eloy Roura, Arnau Oliver, Mariano Cabezas, **Sergi Valverde**, Deborah Pareto, Joan Carles Vilanova, Lluís Ramió-Torrentà, Àlex Rovira and Xavier Lladó. A toolbox for multiple sclerosis lesion segmentation. *Neuroradiology on October 2015, Vol. 57, Issue 10, pp. 1031-1043*. DOI: 10.1007/s00234-015-1552-2. Springer. [JCR RNMMI IF 2.485, Q2(41/125)].
- Mariano Cabezas, Arnau Oliver, **Sergi Valverde**, Brigitte Beltrán, Joan Carles Vilanova, Lluís Ramió-Torrentà, Àlex Rovira and Xavier Lladó. BOOST: a supervised approach for multiple sclerosis lesion segmentation. *Journal of Neuroscience Methods*. Vol. 237, pp 108-117, 2014. Elsevier. [JCR N IF 2.025, Q3(174/252)].
- Yago Díez, Arnau Oliver, Mariano Cabezas, **Sergi Valverde**, Robert Martí, Joan Carles Vilanova, Lluís Ramió-Torrentà, Àlex Rovira and Xavier Lladó. Intensity based methods for brain MRI longitudinal registration. A study on multiple sclerosis patients. *Neuroinformatics*. Vol 12(3), pp 365-379, 2014. Springer. [JCR CSTM IF 2.825, Q1(13/102)]

## Conferences

- Eloy Roura, Arnau Oliver, Mariano Cabezas, **Sergi Valverde**, Deborah Pareto, Joan Carles Vilanova, Lluís Ramió-Torrentà, Àlex Rovira and Xavier Lladó. An SPM12 extension for multiple sclerosis lesion segmentation. *SPIE Medical Imaging 2016*. February 2016, San Diego, USA.
- **Sergi Valverde**, Arnau Oliver, Eloy Roura, Deborah Pareto, Joan Carles Vilanova, Lluís Ramió-Torrentà, Jaume Sastre-Garriga, Xavier Montalban, Àlex Rovira and Xavier Lladó. Evaluation of two automated lesion segmentation and filling pipelines for brain tissue segmentation of multiple sclerosis patients. *ECTRIMS 2015. Multiple Sclerosis*. October 2015, Barcelona, Spain. [JCR CN IF:4.472 Q1(25/191)].
- Eloy Roura, Arnau Oliver, Mariano Cabezas, **Sergi Valverde**, Deborah Pareto, Joan Carles Vilanova, Lluís Ramió-Torrentà, Àlex Rovira and Xavier Lladó.

---

A toolbox for segmenting multiple sclerosis lesions using T1w and FLAIR images. *ECTRIMS 2015. Multiple Sclerosis.* October 2015, Barcelona, Spain. [JCR CN IF:4.472 Q1(25/191)].

- **Sergi Valverde**, Arnau Oliver, Deborah Pareto, Joan Carles Vilanova, Àlex Rovira , Lluís Ramió-Torrentà and Xavier Lladó. SLF: a MS white matter lesion filling toolbox for the SPM software. *ECTRIMS 2014. Multiple Sclerosis.* September 2014, Boston, USA. [JCR CN IF:4.822 Q1(22/192)].
- Ester Quintana, Brigitte Beltrán, **Sergi Valverde**, René Robles-Cedeno, Hec-tor Perkal, Xavier Lladó, José Manuel Fernández-Real and Lluís Ramió-Torrentà. Expression of miRNAs in multiple sclerosis cerebrospinal fluid and their re-lation to MR activity. *ECTRIMS 2014. Multiple Sclerosis.* September 2014, Boston, USA. [JCR CN IF:4.822 Q1(22/192)].
- Ester Quintana, Brigitte Beltrán, **Sergi Valverde**, René Robles-Cedeno, Hec-tor Perkal, Xavier Lladó, José Manuel Fernández-Real and Lluís Ramió-Torrentà. Relación entre la expresión de mirnas en LCR y variables de RM. *Neurología*, vol 29, pp 66-67. 2014. [JCR CN IF:1.322 Q3(142/191)].
- **Sergi Valverde**, Arnau Oliver, Mariano Cabezas, Yago Díez, Jordi Freixenet, Xavier Lladó, Joan Carles Vilanova, Àlex Rovira and Lluís Ramió-Torrentà. A quantitative study of the effects of White Matter MS lesions on tissue segmen-tation methods. *ECTRIMS 2013. Multiple Sclerosis.* October 2013, Copen-hagen, Denmark. [JCR CN IF:4.472 Q1(25/191)].



# Acronyms

- ANN** Artificial Neural Network  
**BET** Brain Extraction Tool  
**BSE** Brain Surface Extractor  
**CSF** Cerebrospinal Fluid  
**EDSS** Extended Disability Status Scale  
**FANTASM** Fuzzy and Noise Tolerant Adaptive Segmentation Method  
**FAST** FMRIB's Automated Segmentation Tool  
**FCM** Fuzzy-C Means  
**FLAIR** Fluid Attenuated Inversion Recovery  
**FSL** FMRIB Software Library  
**GAMIXTURE** Image segmentation toolbox based on genetic algorithm and mixture model optimization  
**GM** Gray Matter  
**CNS** Central Nervous System  
**MARGA** Multispectral Adaptive Region Growing Algorithm  
**MRI** Magnetic Resonance Image  
**MS** Multiple Sclerosis  
**KNN** K-Nearest Neighbor  
**PD** Proton Density  
**PVC** Partial Volume Classifier  
**SVPASEG** Image segmentation toolbox based on local Markov random fields  
**SPM** Statistical Parametric Mapping  
**T1-w** T1-weighted  
**T2-w** T2-weighted  
**WM** White Matter



# Contents

<b>Acknowledgements</b>	i
<b>Publications</b>	iii
<b>Acronyms</b>	vii
<b>1 Introduction</b>	1
<b>Introduction</b>	1
1.1 Research context . . . . .	1
1.1.1 Multiple Sclerosis . . . . .	1
1.1.2 Magnetic resonance imaging in MS . . . . .	2
1.1.3 Image analysis in MS . . . . .	3
1.2 Research background . . . . .	6
1.3 Objectives . . . . .	7
1.3.1 Document structure . . . . .	9
<b>2 Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations.</b>	13
<b>3 Evaluating the Effects of White Matter Multiple Sclerosis Lesions on the Volume Estimation of 6 Brain Tissue Segmentation Methods.</b>	23
<b>4 A white matter lesion-filling approach to improve brain tissue volume measurements.</b>	31

<b>5 Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling.</b>	<b>39</b>
<b>6 Automated tissue segmentation of MR brain images in the presence of white matter lesions.</b>	<b>49</b>
<b>7 Main results and discussion</b>	<b>69</b>
<b>Main results and discussion</b>	<b>69</b>
7.1 Effect of WM lesions on tissue segmentation . . . . .	69
7.2 Effect of lesion filling in tissue segmentation . . . . .	71
7.3 Effect of automating lesion segmentation and filling on tissue segmentation . . . . .	72
7.4 Fully automated tissue segmentation of images containing WM lesions	73
7.5 Related collaborations . . . . .	75
<b>8 Conclusions</b>	<b>77</b>
8.1 Future work . . . . .	79

# Chapter 1

## Introduction

In this first chapter, we introduce the reader to the research context and background of the thesis. Afterwards, we situate this work within the research line of our group, and finally we describe the proposed objectives.

### 1.1 Research context

#### 1.1.1 Multiple Sclerosis

The human nervous system can be divided into the central nervous system (CNS) consisting on the brain and the spinal chord, and the peripheral nervous system, which connects the CNS with the sense organs [7]. CNS is mainly constituted by two tissue components: gray matter (GM), which consists of neuronal cell bodies; and white matter tissue (WM), which is mainly composed of myelinated axon tracts [63]. In the case of the brain, it is mostly composed by GM and WM, both evolved by the Cerebro-spinal fluid (CSF), which provides basic mechanical and immunological protection to the brain inside the skull [63].

Multiple sclerosis (MS) is the most common chronic immune-mediated disabling neurological disease of the CNS [64], in which the insulating covers of the nerve cells in the spinal chord and brain are damaged [18]. Nowadays, MS is the most frequent non-traumatic neurological disease that causes more disability in young adults. It follows a similar behavior also seen in other putative autoimmune diseases, and affects twice as many women as it does in men [19]. It has a low incidence in childhood, but it increases rapidly in adulthood reaching a peak between 25 and 35 years, and then slowly declines, becoming rare at 50 and older [8]. So far, the world estimate for the disease is between 1.3 to 2.5 million cases, being relatively common

in Europe, the United States, Canada, New Zealand, and parts of Australia, but rare in Asia, and in the tropics and subtropics of all continents [8].

MS is characterized by areas of inflammation, demyelination, axonal loss, and gliosis scattered throughout the CNS, often causing motor, sensorial, vision, coordination, deambulation, and cognitive impairment [17]. Demyelination is the process of progressive damage to the protective covering (myelin sheath) around the axon of the neurons. Demyelinated axons conduct impulses at reduced or spontaneous velocity causing impairment in sensation, movement and cognition [18]. The different clinical courses of the disease are generally grouped in four subtype forms [43]. The *Relapsing/Remitting* (RRMS) form of the disease is characterized by exacerbation times where symptoms are present. These periods are followed by periods of remission, where the patient recovers partial or totally from the disease symptoms. The *Secondary Progressive* (SPMS) form is characterized by a gradual intensification of symptoms between affection relapses. The *Progressive remitting* (PRMS) form is typified by an increase in the relapsing times with significant recovery but with worsening symptoms in new relapsing intervals. Lastly, the *Primary Progressive* (PPMS) form is characterized by a severe decrease of remitting times with special localization in the brain. In general, 50% of RRMS patients after 10 years develop the SPMS form of the disease. After 25 years, the 90% of RRMS patients would develop the SPMS form.

### 1.1.2 Magnetic resonance imaging in MS

Magnetic Resonance Imaging (MRI) is a noninvasive medical imaging technique that is used in radiology to generate image representations of different internal anatomical organs and physiological processes of the body. Over the last 40 years, MRI have evolved as a clinical modality [33], and in particular as an essential tool for the diagnosis and evaluation of central nervous system disorders such as MS [24]. On MRI, MS plaques are well-delimited regions with hypo-intense signal intensity with respect to GM on T1-weighted (T1-w), while hyper-intense with respect to GM on T2-weighted (T2), Proton Density (PD) and Fluid Attenuated Inversion Recovery (FLAIR) modalities (see Figure 1.1).

In this aspect, new criteria for MS diagnosis and monitoring has been revised in the last years [48], due to the MRI sensitivity to reveal focal white matter (WM) lesion plaques and disease activity in time and space [26]. Additionally, various studies have also analyzed the correlation between MRI brain tissue atrophy measures and MS disability status, showing that tissue loss is an important component of the disease progression [14, 25, 27, 56]. Tissue loss seems to increase through the course of MS with a similar rate between 0.3% and 0.5% per year, and independently of the MS subtype [21, 56]. In general, GM atrophy is more associated with disability

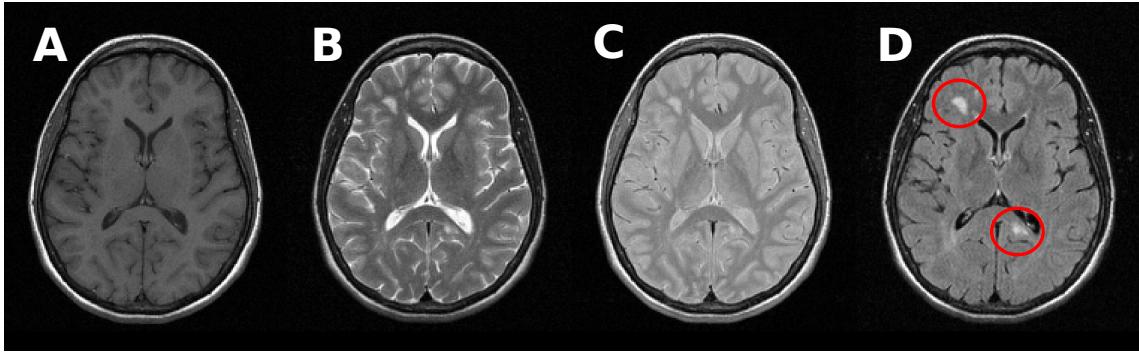


Figure 1.1: MRI image modalities. A) T1-weighted (T1-w) image sequence. B) T2-weighted (T2-w) image sequence. C) Proton Density (PD) image sequence. D) Fluid Attenuated Inversion Recovery (FLAIR) sequence. MS plaques are shown inside red circles on the FLAIR modality. MS plaques are hypo-intense with respect to GM and WM in T2-w, PD and FLAIR sequences, while hypo-intense with respect to WM on the T1-w modality.

changes than WM atrophy [28], and not only in the RRMS and SPMS MS subtypes [27, 56], but also in CIS patients where several studies have shown a significantly greater ventricular cavities and an associated GM loss on MRI scans of CIS patients that will develop MS compared to those who not [13, 25].

### 1.1.3 Image analysis in MS

Manual analysis of brain images is unfeasible in practice, given the large number of three-dimensional MRI slices for each patient and the possible intra/inter observer variability between experts [8]. This has led to the development since the early nineties of a wide number of lesion and tissue segmentation methods, with the aim to reduce the execution time and the inherent variability of manual annotations [16, 32, 39].

#### Pre-processing of MRI images

Acquired brain MRI volumes incorporate non brain tissue parts of the head such as eyes, fat, spinal cord or brain skull. Brain tissue extraction from non-brain tissue is commonly referred in the literature as skull-stripping (see Figure 1.2 B and C). Skull-stripping has a direct effect on the performance of automated methods, as differences in skull stripping would lead into unexpected results in the tissue classification if skull or eyes are included as brain tissue [1, 49]. Among the different proposed

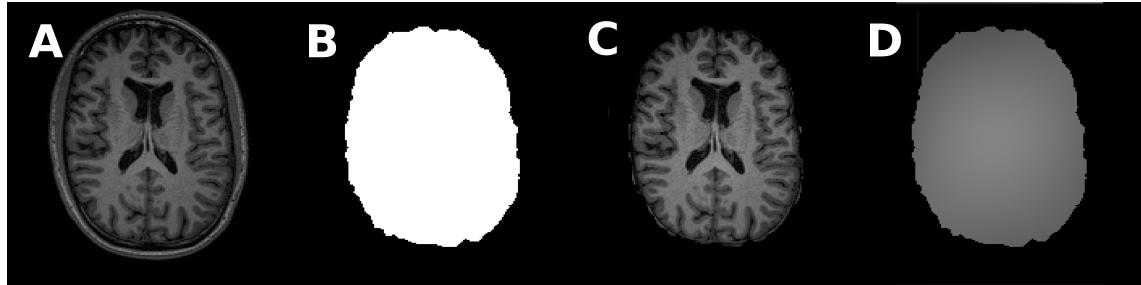


Figure 1.2: MRI pre-processing steps. A) T1-weighted (T1-w) image sequence. B) Computed brain mask and C) skull stripped T1-w sequence using the BET approach [62]. D) Estimated T1-w bias-field using the N3 method proposed by [61].

methods for skull-stripping [1, 40, 54], the Brain Extraction Tool (BET) [62], and the Brain Surface Extractor (BSE) [59] are the most commonly used methods by the neuro-imaging community.

Furthermore, inherent characteristics of the MRI acquisition process such as differences in the magnetic field, bandwidth filtering of the data or eddy currents driven by field gradients usually derive in image artifacts that may also have a negative impact on the performance of methods [60]. In these cases, intensity correction of MRI images is either performed before lesion/tissue segmentation, or as an integrated part of the tissue segmentation pipeline (see Figure 1.2 D). Among the former available strategies proposed [3, 37], the N3 [61], and N4 [68] methods are currently the de-facto standard tools used for intensity correction.

### Automated lesion segmentation

MRI based diagnostic criteria for MS has led to an increasing need to analyze quantitatively focal MS lesions in individual and temporal studies [9, 48]. Different sequences such as T2-w, PD and FLAIR are often used in lesion classification, as MS lesions appear brighter than GM and WM on them. However, WM lesions often present a similar signal intensity profile to CSF on T2-w. In contrast, FLAIR sequences suppress fluids from the image, restraining the CSF tissue effects on the acquired image, although some severe T2-w hyper-intense lesions appear similar to CSF in FLAIR [36].

A wide number of automated lesion segmentation techniques have been proposed during the last years [30, 42]. In these methods, classification is based either in supervised or unsupervised learning. Supervised learning methods employ a training set of correctly-identified observations that are used as prior information to learn the lesion characteristics. Newer proposed strategies integrate spatial decision forest

[31], statistical methods [65], patch-based models [35] or adaptive dictionary learning strategies [22]. In contrast, unsupervised learning methods do not use any prior information in the classification task, which involves grouping data into categories based on some measure of inherent similarity or distance characteristic of the input images. Among these, most recent methods include probabilistic models which separate WM lesions from normal-appearing tissue by considering lesions as an outlier class [36, 38, 67], or techniques that make use of the signal intensity of lesions on FLAIR to apply several thresholding methods with post-processing steps to automatically segment lesions [53, 57].

### Automated brain tissue segmentation in MS

The existent correlation between brain tissue atrophy measures and MS disability status [25, 27], has increased the necessity to develop robust automated brain tissue segmentation methods capable to perform accurate brain tissue volume measurements [34]. However, automated segmentation of brain tissue is still a challenging problem due to the complexity of the images, existence of lesions, differences in tissue intensities, noise, intensity inhomogeneities and the absence of models of the anatomy that fully capture the possible deformations in each structure [8, 39].

A wide number of brain tissue segmentation methods have been proposed so far. General purpose intensity based methods usually perform tissue segmentation on T1-w sequences, as this modality clearly separates gray matter from white matter. These include probabilistic strategies based on Bayesian inference [4, 46, 55, 59], Markov Random Fields models [6, 66, 72], or unsupervised clustering methods such as Zhang2001, *Fuzzy logic* [11, 47]. In contrast, supervised learning approaches also combines T1-w sequences with other modalities such as T2-w and PD using *K-Nearest-Neighbor* classifiers [20, 70], *Support Vector Machines* [2, 69], *Random Forests* [71, 45], or trained *Gaussian mixture models* [51].

However, different studies <sup>1</sup> have shown that tissue abnormalities found in MS image patients such as WM lesions reduce the accuracy of tissue segmentation methods [5, 15]. Effectively, WM lesions on T1-w are hypo-intense with respect to normal-appearing WM, and therefore, lesion voxels that are classified as GM are distorting the overall GM volume. However, lesion voxels may also have an effect in the observed differences in normal-appearing tissue. WM lesions which are actually classified as WM decrease the mean overall signal intensity of the WM, causing that GM voxels with signal intensities similar to WM lesions may be also mis-classified as WM. In contrast, if WM lesions are classified as GM, normal-appearing WM voxels with signal intensities similar to lesions may be miss-classified as GM.

---

<sup>1</sup>Em puc referenciar jo AJNR2015?

### Lesion filling

In MS, hypo-intense WM lesions have to be pre-processed before tissue segmentation in order to reduce the effects of WM lesions on tissue segmentation. Historically, WM lesions have been masked-out of the T1-w before segmentation, and their volume have been added to WM afterwards [14]. Although this method effectively reduces the error in tissue volume, it has been shown that this approach is not optimal [5, 15].

In this aspect, several strategies have proposed to in-paint lesions on the T1-w with signal intensities of the normal-appearing WM before tissue segmentation [5, 15, 44, 58], a process which is usually known in the literature as lesion filling. However, most of the available lesion filling methods require manual delineations of lesions, which may be tedious, challenging and time-consuming task depending of the characteristics of the image [42]. When available, lesion filling has demonstrated not only a significant reduction in the associated errors of WM lesions in tissue volume measurements [50], but also in image registration [12, 23, 58] and cortical thickness measurements [44].

## 1.2 Research background

This thesis is located within the framework of different research projects associated to the Computer Vision and Robotics research group (VICOROB), a research institute of the University of Girona<sup>2</sup>. VICOROB has been working on several medical image analysis projects since 1996, mainly in segmentation and registration of mammography images. Lately in 2010, the research group started a fruitful collaboration with several medical MS research teams with the aim to develop new automated techniques capable to segment MS lesions and to perform atrophy measurements that can be transferred to experts for clinical use. In particular, our research in the MS field has been carried out within the following research projects:

1. [2009 – 2012] PI09/91918 “SALEM: Segmentación Automática de Lesiones de Esclerosis Múltiple en imágenes de resonancia magnética” awarded by the Instituto Carlos III.
2. [2010 – 2012] VALTEC09-1-0025 “Salem: toolkit para la segmentación automática de lesiones de Esclerosis Múltiple en resonancia magnética” awarded in 2009 by the Generalitat de Catalunya within the ”Projectes de valorització VALTEC”.

---

<sup>2</sup>[www.vicorob.udg.edu](http://www.vicorob.udg.edu)

3. [2015 – 2017] TIN2014-55710-R: “Herramientas de neuroimagen para mejorar el diagnóstico y el seguimiento clínico de los pacientes con Esclerosis Múltiple” awarded in 2014 by the spanish call Retos de investigación 2014.
4. [2015 – 2019] BiomarkEM.cat: To develop, validate and transfer to clinical practice robust tools and totally automated for measuring new lesions and changes on the brain volume within MS patients. Awarded in 2015 by the Fundació Marató de TV3.

Since then, the research group has published original contributions in different fields such as image pre-processing [54], MS lesion segmentation [9, 10, 42, 53], temporal analysis [29, 41], image registration [23, 52], and tissue segmentation [8]. All this works have been published in partnership with different medical MS teams from:

- from the Hospital Vall d’Hebron: Dr. Rovira, who is the director of the “Unitat de Ressonància Magnètica-Centre Vall d’Hebron” (URMVH) and has participated in several research projects funded by public and private institutions in the last few years, as well as Dr. Pareto and technicians Huerga and Corral. This group is part of the MAGNIMS network, a European network of centres that share an interest in the MS study through MRI.
- from the Clínica Girona / Hospital Santa Caterina: Dr. Vilanova and Dr. Barceló are the codirectors of the “Unitat de Ressonància Magnètica” at the Clínica Girona and are members of several national and international radiology societies.
- from the Hospital Josep Trueta: Dr. Ramió-Torrentà, who is the current coordinator of the ”Unitat de Neuroimmunología i Esclerosis Múltiple”, as well as Drs. Robles and Beltrán, who work for the radiology unit.

## 1.3 Objectives

As part of the TIN2014-55710-R and BiomarkEM.cat research project frameworks, the main goal of this thesis is:

**to develop a novel fully automated brain tissue segmentation method capable of computing accurate tissue volume measurements of MS patient images.**

This objective refers to the brain tissue segmentation of MS patient images at a specific time but we do not consider the differences in tissue volume at different stages.

Different stages have to be covered first in order to fulfill the main proposed goal. The aim of each of the proposed stages is to permit us to gain a better knowledge of the different parts that compose a fully automated tissue segmentation method for MS images containing lesions. Hence, we detail each of the proposed stages to cover:

- **to analyze the state of the art of tissue segmentation methods.** The first stage aims to review different proposed tissue segmentation techniques in order to understand their advantages and drawbacks. Methods are evaluated on public databases that incorporate manual tissue annotations, which permits to perform a quantitative analysis of the accuracy of the methods.
- **to study the effect of WM lesions on tissue segmentation of MS patient images.** Although it is known that the inclusion of WM lesions on tissue segmentation distort the brain volume measurements, this effect has not been studied and compared across different tissue segmentation methods. In this aspect, the second stage is focused on the analysis of the effects of WM lesions on the tissue distributions of a set of tissue segmentation approaches using multi-center 1.5T MS data from different scanners. Our hypothesis here is that a better knowledge of the correlation between lesion attributes, such as signal intensity and lesion size, and the observed differences in tissue volume of the analyzed algorithms may be beneficial to design a tissue segmentation method for MS.
- **to reduce the effect of WM lesions on tissue segmentation of MS patient images.** As said in section 1.1.3, WM lesions have to be pre-processed before tissue segmentation in order to reduce the effects of WM lesions on tissue segmentation. In this aspect, the third stage is two-fold: first, to compare the accuracy of different proposed lesion filling techniques in the literature, analyzing their accuracy on both 1.5T and 3T data. Once we have analyzed the benefits and drawbacks of each proposed method, we aim to propose a new lesion filling pipeline in order to overcome the possible limitations of existent methods. The proposed method will be available for the research community.
- **to analyze the effect of automated WM lesion segmentation and filling on the tissue segmentation.** Although lesion filling techniques have already been successfully applied to reduce the effect of WM lesions on tissue segmentation, usually WM lesions are annotated manually before segmentation. In contrast, the effect of both automated lesion segmentation and filling

on tissue segmentation is still unclear. The fourth stage of the thesis aims to understand the effects of the inherent errors in automated lesion segmentation on the posterior lesion filling and tissue segmentation. Thus, we aim to compare the accuracy of different state-of-the-art fully automated pipelines that incorporate lesion segmentation, lesion filling and tissue segmentation on MS data, in order to analyze the extend of the effect of remaining lesions on the differences in tissue segmentation, which may be beneficial to update our acquired knowledge in previous stages.

- **to propose a new fully automated tissue segmentation method for MS patient images.** Finally, we aim to benefit from the acquired knowledge on previous stages to propose a **fully automated tissue segmentation method** able to deal with images of MS patients with different level of brain atrophy and lesion load. In this last stage, we aim to validate the accuracy of the proposed method by comparing it with the state of the art in tissue segmentation in MS. The proposed method will be publicly available for the research community.

### 1.3.1 Document structure

A graphical description of the structure of the thesis is shown in Figure 1.3. Connections between chapters in Figure 1.3 depict the conceptual link between chapters :(. The rest of the document is organized as follows:

- **Chapter 2. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations.** In chapter 2, we present a comprehensive comparison of the accuracy of 10 brain tissue segmentation methods on two public MRI databases.
- **Chapter 3. Evaluating the Effects of White Matter Multiple Sclerosis Lesions on the Volume Estimation of 6 Brain Tissue Segmentation Methods.** After reviewing different tissue segmentation techniques in public data, we perform a detailed analysis of the effects of WM lesions on the brain tissue volume measurements of six of these tissue segmentation methods using MS data from different hospital centers.
- **Chapter 4. A white matter lesion-filling approach to improve brain tissue volume measurements.** In this chapter, we propose a new technique to fill WM lesions on 1.5T and 3T data, validating its accuracy with other proposed methods in the literature.

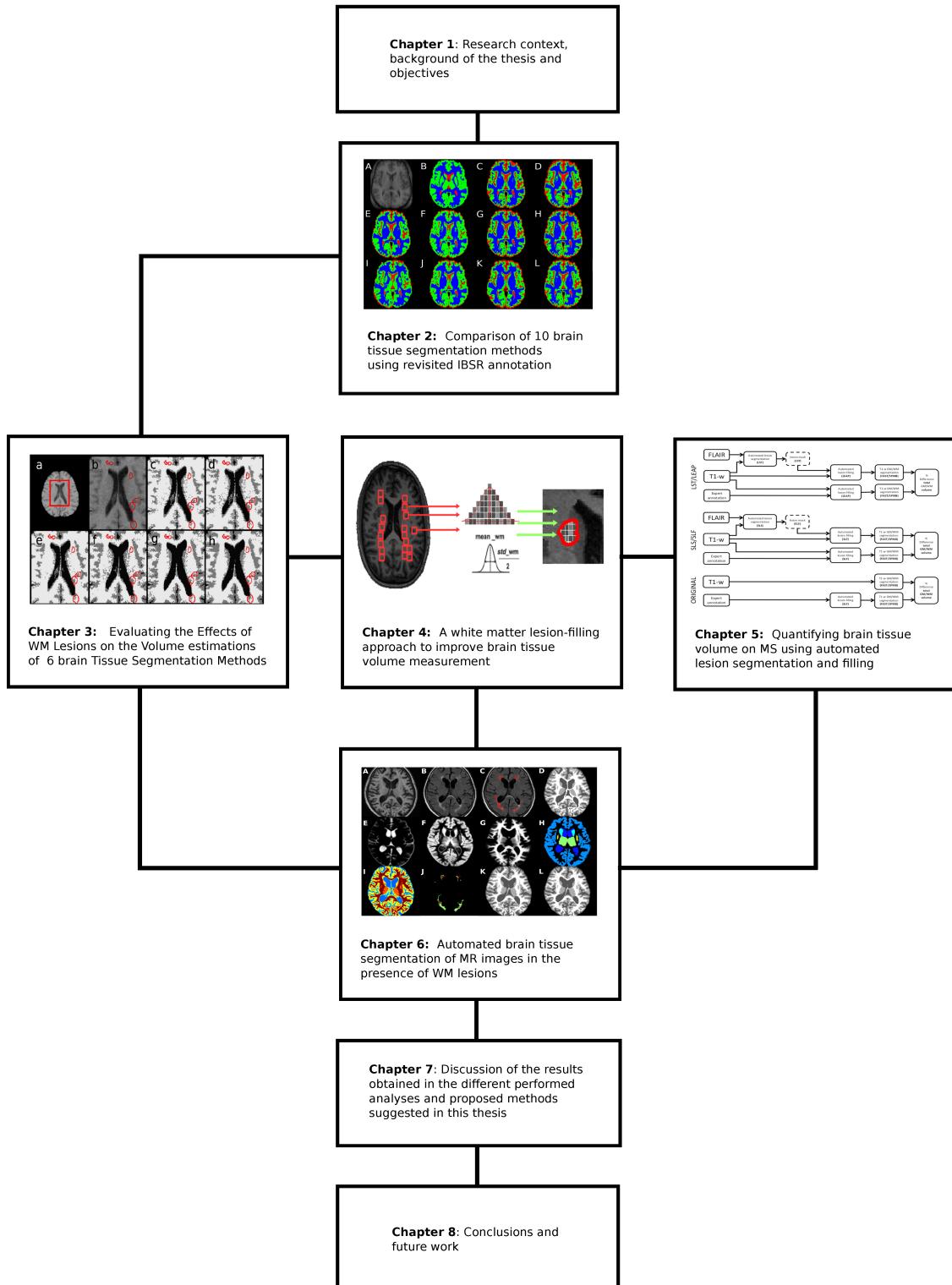


Figure 1.3: Organization of the document. Preliminary chapter 1 describes the research context and the main objectives of this thesis. Chapters 2 to 6 introduce the main contributions of this work. Chapter 7 proposes a general discussion of the results obtained in Chapters 2 to 6. Finally, the main conclusions and the proposed future work are defined in Chapter 8. Connections between chapters depict a conceptual link between chapters.

- **Chapter 5. Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling.** In this chapter we present a detailed evaluation of the performance of different automated pipelines that incorporate lesion segmentation, lesion filling and tissue segmentation on MS data.
- **Chapter 6. Automated brain tissue segmentation of MR images in the presence of white matter lesions.** We propose a novel fully automated tissue segmentation pipeline designed to deal with MS patient images containing lesions. We validate the accuracy of the proposed method with other state-of-the-art techniques.
- **Chapter 7. Results and discussion.** This chapter provides an comprehensive discussion of the results obtained in the different performed analyses and proposed methods suggested in this thesis.
- **Chapter 8. Conclusions and future work.** Finally, the main conclusions based on the contributions of this thesis are defined. Based on this conclusions, we summarize different future works to improve or extend the suggested approach.



## Chapter 2

# Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations.

In this chapter, we perform a quantitative evaluation of the accuracy of 10 automated brain tissue segmentation methods. Methods are compared based using the public databases IBSR20 and IBSR18<sup>1</sup>. The performance of the methods is evaluated by ranking their accuracy based on the significant differences with respect the rest of methods. This proposed evaluation has been published in the following paper:

Paper published in Journal of Magnetic Resonance in Medicine (JMRI)

Volume: 41, Issue: 1, Pages: 93-101, Published: January 2014

DOI: 10.1002/jmri.24517

Quality Index: 3.21 (Quartile 1)

---

<sup>1</sup>no existeixen ja :(

## Original Research

## Comparison of 10 Brain Tissue Segmentation Methods Using Revisited IBSR Annotations

Sergi Valverde, MS,\* Arnaud Oliver, PhD, Mariano Cabezas, PhD,  
Eloy Roura, MS, and Xavier Lladó, PhD

**Purpose:** Ground-truth annotations from the well-known *Internet Brain Segmentation Repository* (IBSR) datasets consider Sulcal cerebrospinal fluid (SCSF) voxels as gray matter. This can lead to bias when evaluating the performance of tissue segmentation methods. In this work we compare the accuracy of 10 brain tissue segmentation methods analyzing the effects of SCSF ground-truth voxels on accuracy estimations.

**Materials and Methods:** The set of methods is composed by FAST, SPM5, SPM8, GAMIXTURE, ANN, FCM, KNN, SVPASEG, FANTASM, and PVC. Methods are evaluated using original IBSR ground-truth and ranked by means of their performance on pairwise comparisons using permutation tests. Afterward, the evaluation is repeated using IBSR ground-truth without considering SCSF.

**Results:** The Dice coefficient of all methods is affected by changes in SCSF annotations, especially on SPM5, SPM8 and FAST. When not considering SCSF voxels, SVPASEG ( $0.90 \pm 0.01$ ) and SPM8 ( $0.91 \pm 0.01$ ) are the methods from our study that appear more suitable for gray matter tissue segmentation, while FAST ( $0.89 \pm 0.02$ ) is the best tool for segmenting white matter tissue.

**Conclusion:** The performance and the accuracy of methods on IBSR images vary notably when not considering SCSF voxels. The fact that three of the most common methods (FAST, SPM5, and SPM8) report an important change in their accuracy suggest to consider these differences in labeling for new comparative studies.

**Key Words:** brain MRI; tissue segmentation; permutation tests; IBSR

**J. Magn. Reson. Imaging** 2014;00:000–000.

© 2014 Wiley Periodicals, Inc.

MOST AUTOMATIC BRAIN tissue segmentation methods are evaluated on common data which permit to

reproduce and compare the results between different studies. In the past 15 years, some public datasets have been proposed based on simulated data such as the Brainweb dataset (1) or real MRI data acquired from different sources such as the Internet Brain Segmentation Repository (IBSR)<sup>1</sup>. IBSR contains two sets of T1-w images acquired from healthy subjects and composed by 20 (IBSR20) and 18 (IBSR18) images, respectively. Both sets provide images and ground-truth segmentations labeled by experts which permit to quantify the accuracy of methods with respect to the images of the dataset (2–5).

However, on both datasets all original ground-truth annotations do not contain sulcal parts of cerebrospinal fluid (CSF) tissue and include it inside the gray matter (GM) segmentation (see Fig. 1). Although this fact is known by authors, most of the studies which make use of the IBSR datasets tend to publish their findings assuming the deviation on CSF tissue.

On all these studies, CSF is not considered (2,6–8), or a weak performance of CSF is obtained compared with GM and white matter (WM) (3,5,9). In contrast, several studies also using IBSR datasets have relabeled sulcal CSF (SCSF) voxels as GM tissue before evaluating the accuracy of methods (10–12) or combined all CSF with GM as a single tissue (13) to minimize the differences between segmentation masks and ground-truth labels.

Changes in CSF tissue labels can not only affect the accuracy of CSF tissue but also the GM estimation. For instance, with original IBSR data, the accuracy of a method segmenting GM tissue can be benefited by misclassified SCSF voxels, as long as they are labeled as GM on ground-truth images. Misclassified SCSF voxels can bias the real accuracy of methods segmenting GM tissue because these voxels are labeled as GM and can compensate the performance on genuine GM tissue. More importantly, on comparative studies using IBSR datasets this fact can induce to inaccurate measurements, and real differences between methods can be hidden behind the bias introduced by SCSF voxels.

In this study, we evaluate the effects of the IBSR annotations on the accuracy of ten commonly used

Department of Computer Architecture and Technology, University of Girona, Girona, (Spain).

Contract grant sponsor: Generalitat de Catalunya; Contract grant number: FI-DGR2013.

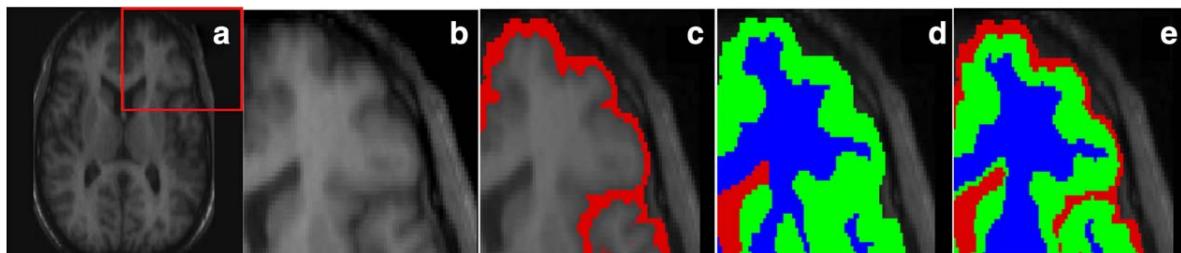
Eloy Roura also holds a research grant from Universitat de Girona (ref UdG BR-GR13).

\*Address reprint requests to: S.V., Campus Montilivi, University of Girona, 17071 Girona (Spain). E-mail: svalverde@eia.udg.edu

Received July 23, 2013; Accepted October 22, 2013.

DOI 10.1002/jmri.24517

View this article online at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).



**Figure 1.** Graphical description of IBSR ground-truth annotations. A: Image IBSR01 from IBSR18 dataset. B: Zoomed view of part of the image. C: Highlighted sulcal CSF voxels. D: IBSR ground-truth for the current image. E: Example of tissue segmentation. SVPASEG correctly classifies the majority of sulcal CSF voxels as CSF. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

automatic brain tissue segmentation methods with the aim to establish a reference mark that can be used in future comparisons between these and new proposed methods.

## MATERIALS AND METHODS

### MRI Image Data

The MRI image data is composed by 20 and 18 T1-w scans of normal subjects from the *Internet Brain Segmentation Repository* (IBSR), available from the Center for Morphometric Analysis at Massachusetts General Hospital. The first set of scans is commonly known in the literature as IBSR20 while the second is known as IBSR18. Although the Center for Morphometric Analysis at Massachusetts General Hospital provide both datasets, the characteristics for each set of images is different.

### IBSR20 Dataset

The IBSR20 image set is composed by 20 T1-w scans with 3.1 mm slice thickness ( $256 \times 63 \times 256$ ). The authors also provide labeled volumes with main tissue annotations for evaluation (GM, WM, and CSF), based on trained experts using a semi-automated intensity contour mapping algorithm (14), and signal intensity histograms. These images are sorted by level of difficulty. The most challenging scans contain important acquisition artifacts and irregularities.

### IBSR18 Dataset

The set is composed by 18 T1-w scans with 1.5 mm slice thickness ( $256 \times 128 \times 256$ ). IBSR18 scans present higher resolution and image quality than IBSR20, with no apparent acquisition artifacts that can bias the accuracy of some scans. Scans are provided after processing them with the Autoseg bias field correction routines from the Center for Morphometric Analysis. The dataset is also supplied with manually labeled volumes into 84 structures obtained using the NVM program<sup>2</sup> and three-class labeled volumes by assigning each of the 84 structures into GM, WM, and CSF tissues.

### Preprocessing

First, we remove from the images all nonbrain tissue parts such as eyes, fat, spinal cord or brain skull using a binary mask created from the provided ground-truth. Afterward, all images from IBSR20 dataset are corrected from possible intensity nonuniformities and acquisition artifacts using the N3 package<sup>3</sup> (15). To guarantee a common preprocessing framework between segmentation techniques, we disabled all possible preprocessing options on methods for all experiments.

### Segmentation Methods

In this study, we evaluate 10 brain tissue segmentation algorithms on T1-w MRI data with the aim to include a wide set of different segmentation techniques and available tools. We include in our study simpler techniques such as ANN, FCM, KNN, and public available toolboxes such as FAST, SPM5, SPM8, PVC, GAMIXTURE, SVPASEG, and FANTASM. Characteristics and requirements for each method are summarized in Table 1. After testing different parameter configurations for each method, we run all methods with default options because they provide the best overall results for both datasets.

FAST (16) and SVPASEG (8) methods are both based on Markov Random Fields models. In the case of FAST, Markov Random Field parameters are estimated using the iterative Expectation Maximization algorithm with K-means initialization. SVPASEG is based on an Iterative Conditional Modes algorithm and initialization parameters are estimated by a real-coded Genetic Algorithm. PVC (17) is based on a Maximum-a-Posteriori classifier and also optimized by the Iterative Conditional Modes algorithm. Tissue distributions are estimated by combining the tissue measurement model with a spatial prior model of the brain.

SPM5 and SPM8 are two of the available versions of the SPM toolbox. SPM5 *segment* and SPM8 *new\_segment* methods are both based on the same algorithm, which comprises the parameter estimations of a Gaussian Mixture Model, atlas registration and bias-field correction at the same time iteratively (18). SPM8 is the current version of the toolbox and introduces

<sup>2</sup><http://neuromorphometrics.org:8080/>

<sup>3</sup><http://www.bic.mni.mcgill.ca/ServicesSoftware/MINC>

Table 1  
Brain Tissue Segmentation Methods Evaluated In This Study \*

Method reference Name	Algorithm characteristics					Implementation	Version	Source
	Algorithm	Image Type	IC	SS				
FAST [16]	HRMF+EM	All	Yes	No	FSL v.5.0 (Sept 2012)	<a href="http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FAST">http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FAST</a>		
SPM5 [18]	GMM	All	Yes	Yes	SPM8 v.4667 (Feb 2012)	<a href="http://www.fil.ion.ucl.ac.uk/spm/">www.fil.ion.ucl.ac.uk/spm/</a>		
SPM8 [19]	GMM	All	Yes	Yes	SPM8 v.4667 (Feb 2012)	<a href="http://www.fil.ion.ucl.ac.uk/spm/">www.fil.ion.ucl.ac.uk/spm/</a>		
GAMIXTURE [20]	GA-GMM	T1	No	No	GAMIXTURE v1.1 (Feb 2007)	<a href="http://www.loni.ucla.edu/Software/GAMixture">http://www.loni.ucla.edu/Software/GAMixture</a>		
ANN [23]	SOM	T1	No	No	MATLAB 7.12	–		
FCM [21]	FCM	T1	No	No	MATLAB 7.12	–		
KNN [24]	KNN	T1,PD	No	No	MATLAB 7.12	–		
SPVASEG [8]	GA-MRF	T1,T2	No	No	v.2.1 (Oct 2010)	<a href="http://www.cs.tut.fi/jupeto/svpaseg/">http://www.cs.tut.fi/jupeto/svpaseg/</a>		
FANTASM [22]	RFCM	T1	Yes	Yes	MIPAV v.R3c (Mar 2012)	<a href="http://mipav.cit.nih.gov/">http://mipav.cit.nih.gov/</a>		
PVC [17]	MAP	T1	Yes	Yes	Brainsuite 11.a (May 2011)	<a href="http://neuroimage.usc.edu/neuro/BrainSuite">http://neuroimage.usc.edu/neuro/BrainSuite</a>		

\*The acronyms for the algorithms stand for: Expectation Maximization (EM), Fuzzy C-Means (FCM), Gaussian Mixture Model (GMM), Genetic Algorithm (GA), Hidden Random Markov Fields (HRMF), Maximum-a-Posteriori (MAP), K-nearest Neighbor (KNN), Markov Random Fields (MRF), Robust Fuzzy C-Means (RFCM), Self Organized Map (SOM). The acronyms for the column definition stand for Intensity Correction (IC) and Skull-stripping (SS).

several differences with respect to SPM5 such as a different treatment of mixing proportions, an improved registration model, an extended set of probabilistic atlases, and a more robust initial registration (19). Similar to SPM approaches, tissue distributions on GAMIXTURE (20) are also obtained by a Gaussian Mixture Model. However, this program estimates both the initial and the successive parameters of the Gaussian Mixture Model by a real-coded Genetic Algorithm.

FCM (21) and FANTASM (22) are based on Fuzzy Clustering techniques. FCM refers to the classical fuzzy-c-means clustering algorithm, which does not take into account spatial information. In contrast, FANTASM extends the FCM approach by modifying the objective function with a penalty term based on the membership of neighbors to other classes that makes the method more robust to noise and acquisition artifacts.

ANN (23) implements a Self Organizing Map or Kohonen network, which clusters image data based on an iterative process of comparison of related changes within voxels organizing unknown data into groups of similar patterns, according to a similarity criterion (e.g., Euclidean distance). Finally, KNN (24) implements a self-trained K-Nearest Neighbor algorithm based on automatic registration of prior probability atlases into the input image, which are used to label training voxels with the class with maximum probability obtained from the tissue atlases.

## Data Analysis

First, we evaluate the accuracy of methods segmenting GM, WM, and CSF tissues on both IBSR20 and IBSR18 datasets by computing the overlap between the segmentation masks and the ground-truth annotations using the Dice similarity coefficient (25):

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}$$

where  $TP$  is the number of true positives between the segmentation result and the ground-truth,  $FP$  is the

number of false positives and  $FN$  is the number of false negatives. Hence, the DSC coefficient ranges from 0 to 1, and a higher Dice coefficient represents more accurate segmentation. Compared with other measures such as the Sensitivity, Specificity, False Negative, and False Positive Rates (13,26,27), which only state on positive or negative outcomes and are only based on size, the Dice coefficient is a compromise between the positive and negative outcomes, allowing a better understanding of the overall similarity and dissemination of the location of the differences between the segmentation masks and the ground-truth (28).

Second, we rank methods on both IBSR20 and IBSR18 datasets using significant pairwise method permutation tests of the obtained Dice values on GM and WM tissues (29,30). Permutation tests select small sets of independent values obtained by the different segmentation methods, choose all possible method pairs, and for each pair permute an arbitrary number of times a random number of values. Permutation tests permit to compute the exact  $P$ -value, and are not limited by any statistical distribution or minimum number of subjects. The  $P$ -value for each method pair is computed by counting the number of times that the mean difference of the permuted pair is higher than the mean difference without permuting. Finally, the significance of the results between the method pair is stored as the percentage of times where the  $P$ -value  $\leq 0.05$ .

For our experiments, we have adapted the implementation provided by Klein et al. (30)<sup>4</sup>. The test returns the mean  $\mu$  and standard deviation  $\sigma$  of the fraction of times when each method provides a higher Dice value than the rest of the methods with significant  $P$ -value  $\leq 0.05$ . Consequently, methods with higher means have passed a higher number of pairwise comparisons with other methods using randomly chosen subsets of values. Methods are presented in ranks determined by the mean and standard

<sup>4</sup>Available for download in [http://www.mindboggle.info/papers/evaluation\\_NeuroImage2009/](http://www.mindboggle.info/papers/evaluation_NeuroImage2009/)

**Table 2**  
Mean Dice Coefficient for Each Method and Tissue Computed From IBSR20 Scans

a. IBSR20 with original ground-truth			
Method	GM	WM	CSF
FAST	0.68±0.06	0.79±0.10	0.13±0.04
SPM5	0.76±0.06	0.80±0.04	0.17±0.07
SPM8	0.78±0.06	0.81±0.08	0.21±0.07
GAMIXTURE	0.77±0.09	0.74±0.16	<b>0.25±0.12</b>
ANN	0.69±0.09	0.77±0.14	0.15±0.06
FCM	0.69±0.09	0.77±0.14	0.14±0.05
KNN	0.64±0.09	0.80±0.06	0.13±0.04
SVPASEG	<b>0.82±0.04</b>	<b>0.81±0.07</b>	0.21±0.06
FANTASM	0.70±0.10	0.77±0.14	0.15±0.06
PVC	0.66±0.11	0.63±0.23	0.13±0.05

b. IBSR20 with ground-truth not considering SCSF			
Method	GM	WM	CSF
FAST	0.82±0.06	0.78±0.12	0.76±0.12
SPM5	0.86±0.03	<b>0.82±0.02</b>	0.82±0.08
SPM8	0.86±0.06	0.81±0.07	<b>0.83±0.08</b>
GAMIXTURE	0.83±0.06	0.74±0.16	0.75±0.10
ANN	0.81±0.08	0.77±0.14	0.76±0.12
FCM	0.81±0.08	0.77±0.14	0.76±0.12
KNN	0.78±0.08	0.80±0.06	0.75±0.12
SVPASEG	<b>0.88±0.04</b>	0.81±0.07	0.76±0.09
FANTASM	0.81±0.08	0.77±0.14	0.77±0.12
PVC	0.79±0.09	0.62±0.24	0.77±0.12

Table (a) shows the results on IBSR20 scans evaluated with original ground-truths. Table (b) shows the results on the same scans when not considering SCSF on the evaluation. Reported values are mean ± standard deviation. The highest mean Dice value for each tissue is shown in bold text.

deviation of the method with highest mean ( $\mu_o$ ,  $\sigma_o$ ). Ranks are decided in terms of the distance of the mean of each method to the best mean observed. Specifically, Rank 1 methods are those in  $(\mu_o - \sigma_o, \mu_o)$ , Rank 2 methods fall in  $(\mu_o - 2\sigma_o, \mu_o - \sigma_o)$  and Rank 3 methods are those in the interval  $(\mu_o - 3\sigma_o, \mu_o - 2\sigma_o)$ . In all our experiments we have repeated the permutation tests within  $N = 1000$ .

## RESULTS

### IBSR20 Dataset

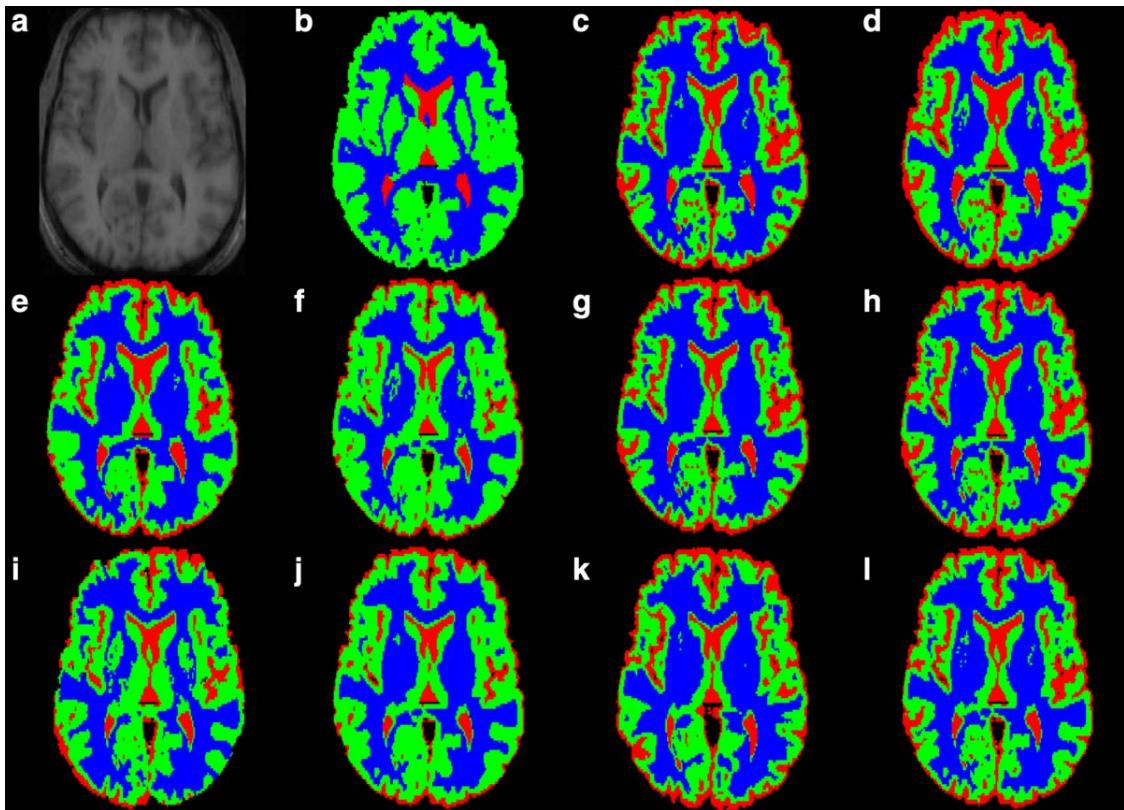
Table 2a shows the Dice overlap coefficients obtained for each method and tissue. SVPASEG provides the highest Dice values on GM, followed by SPM8, GAMIXTURE and SPM5. SVPASEG and both SPM approaches take advantage of the prior atlas information to improve their performance on the most difficult images, which also explains their low standard deviation in comparison with the rest of the methods. GAMIXTURE compensates a lower performance on difficult scans with a similar performance to SVPASEG on easier images. In contrast, FAST, PVC, and KNN provide the lowest Dice values, mostly because their poor performance on CSF. On WM, again SVPASEG

but also SPM5 and SPM8 provide the highest overlap coefficients while PVC is the method of the study with the lowest Dice coefficient.

All methods provide very low values for CSF due to divergences on the classification of SCSF tissue with respect to IBSR20 ground-truth. Methods tend to classify SCSF voxels mostly as CSF but those are labeled as GM in the ground-truth. This fact has also a direct effect on GM tissue because voxels segmented as SCSF are considered as false positives by the ground-truth, decreasing the mean Dice GM values on all methods. In contrast, the number of false positives caused by differences between the segmentation masks and the ground-truth appears to be lower on WM tissue.

To avoid possible influences of SCSF tissue on the overall accuracy of the methods, we recomputed again the Dice overlap coefficient on the same 20 images of the IBSR20 dataset but without taking into account SCSF voxels. Table 2b shows the new values for all methods and tissues. All methods improve their performance after not considering SCSF, especially those with lower Dice values on original images. On WM, SPM5, SVPASEG, SPM8, and KNN are the methods with highest Dice values. Contrary to GM, all methods except PVC provide very similar values on easy scans, and differences on the mean WM Dice values are mostly due to the performance of each method on the most challenging scans. GAMIXTURE and PVC provide the lowest Dice values on WM. The voxels affected by image artifacts appear brighter than normal voxels. The inclusion of these voxels into the tissue distributions of methods that incorporate the voxel intensity into the segmentation process such as FCM, FANTASM, ANN, or GAMIXTURE can increase the threshold between GM and WM tissue intensities, and voxels that would have been classified as WM will be misclassified as GM. The results without considering SCSF show that these differences between methods cannot be attributed to CSF because 8 of 10 methods maintain their performance after the modifications introduced on the ground-truth.

Table 3 shows the ranking of methods returned by the permutation tests for the IBSR20 dataset using the original ground-truth and without considering SCSF voxels. With original ground-truth annotations, SVPASEG outperforms significantly the rest of the methods in 90% of the times on GM. Although SVPASEG performs similarly to SPM5 and SPM8 on difficult scans, the fact that SVPASEG provides the highest Dice values on 15 out of 20 scans makes the method invariant to the different permutations. Permuting the Dice values of several scans with other methods will maintain significant real differences between both methods if one of them has most of the highest values. On WM, SVPASEG, SPM8, SPM5, KNN, and FAST are ranked as Rank 1. The results of the permutation test without considering SCSF reveal that the methods more penalized by original images ground-truth are SPM5, SPM8 and FAST. In the case of SPM5 and SPM8, after not considering SCSF the permutation test classifies both methods in Rank 1 along with SVPASEG while in the case of FAST is



**Figure 2.** Tissue segmentation masks for IBSR\_05 image. Original image (A), ground-truth (B), and segmentation results from FAST (C), SPM5 (D), SPM8 (E), GAMIXTURE (F), ANN (G), FCM (H), KNN (I), SVPASEG (J), FANTASM (K), and PVC (L). In segmentation mask images, CSF tissue is labeled in red, GM in green, and WM in blue.

promoted from Rank 3 to Rank 2. On the contrary, GAMIXTURE and PVC appear to be benefited by the original ground-truth because both methods are moved to Rank 3 after not considering SCSF.

#### IBSR18 Dataset

Table 4a shows the mean Dice values obtained for all methods and tissues with original ground-truth annotations (see also Fig. 2). In general, better results are

Table 3

Permutation Tests for Obtained Dice Overlap Coefficients on the IBSR20 Dataset With Original Ground-Truth and Not Considering SCSF Voxels \*

IBSR20 (original ground-truth)						IBSR20 (not considering SCSF)							
		GM		WM				GM		WM			
	Method	$\mu \pm \sigma$	Method	$\mu \pm \sigma$		Method	$\mu \pm \sigma$	Method	$\mu \pm \sigma$		Method	$\mu \pm \sigma$	
Rank 1	SVPASEG	0.90±0.32	SVPASEG	0.70±0.48	SVPASEG	0.80±0.42	SVPASEG	0.64±0.48	Rank 2	Rank 3	Rank 1	Rank 2	Rank 3
	SPM8	0.60±0.52	SPM5	0.50±0.53	SPM8	0.70±0.48	SPM8	0.60±0.52					
	SPM5	0.50±0.53	KNN	0.40±0.70	SPM8	0.60±0.70	SPM5	0.50±0.53					
	KNN	0.40±0.70	FAST	0.30±0.82	FANTASM	0.30±0.95	ANN	-0.30±0.82					
	FAST	0.30±0.82			FAST	0.10±0.99	FANTASM	-0.30±0.82					
Rank 2	SPM5	0.50±0.71			GAMIXTURE	0.30±0.95	ANN	-0.30±0.82	Rank 3	Rank 2	Rank 1	Rank 2	Rank 3
	SPM8	0.50±0.71			FAST	0.10±0.99	FANTASM	-0.30±0.82					
	GAMIXTURE	0.50±0.71					FCM	-0.30±0.82					
Rank 3	FANTASM	0.03±0.95	ANN	-0.30±0.82	FANTASM	-0.19±0.92	GAMIXTURE	-0.70±0.67	Rank 2	Rank 3	Rank 1	Rank 2	Rank 3
	ANN	-0.03±0.94	FANTASM	-0.30±0.82	ANN	-0.50±0.71	PVC	-0.90±0.32					
	FAST	-0.50±0.70	FCM	-0.30±0.82	FCM	-0.50±0.71							
	FCM	-0.50±0.71	GAMIXTURE	-0.70±0.67	PVC	-0.51±0.52							
	PVC	-0.60±0.52	PVC	-0.91±0.32	KNN	-0.80±0.42							
	KNN	-0.80±0.42											

\*Reported values are mean and standard deviation ( $\mu$ ,  $\sigma$ ) of the fraction of times when each method produced significant  $P$  values. Positive values indicate that on average, the method overperformed the other methods in pair-wise significant tests. Negative values indicate the contrary. Rank 1:  $(\mu_o - \sigma_o, \mu_o + \sigma_o]$ , Rank 2:  $(\mu_o - 2\sigma_o, \mu_o - \sigma_o]$ , Rank 3  $(\mu_o - 3\sigma_o, \mu_o - 2\sigma_o]$ .

Table 4  
Mean Dice Coefficient for Each Method and Tissue Computed From IBSR18 Scans

a. IBSR18 with original ground-truth			
Method	GM	WM	CSF
FAST	0.74±0.04	<b>0.89±0.02</b>	0.12±0.05
SPM5	0.68±0.07	0.86±0.02	0.10±0.05
SPM8	0.81±0.02	0.88±0.01	<b>0.17±0.08</b>
GAMIXTURE	0.78±0.08	0.87±0.02	0.15±0.09
ANN	0.70±0.07	0.87±0.03	0.11±0.06
FCM	0.70±0.06	0.88±0.03	0.11±0.06
KNN	0.79±0.03	0.86±0.03	0.16±0.07
SVPASEG	<b>0.81±0.03</b>	0.86±0.02	0.16±0.07
FANTASM	0.71±0.06	0.88±0.03	0.11±0.06
PVC	0.70±0.08	0.83±0.07	0.13±0.06

b. IBSR18 with ground-truth not considering SCSF			
Method	GM	WM	CSF
FAST	0.88±0.01	<b>0.89±0.02</b>	0.47±0.18
SPM5	0.89±0.02	0.87±0.02	<b>0.79±0.08</b>
SPM8	<b>0.91±0.01</b>	0.88±0.01	0.77±0.08
GAMIXTURE	0.89±0.03	0.87±0.02	0.52±0.15
ANN	0.87±0.03	0.88±0.03	0.52±0.15
FCM	0.88±0.02	0.88±0.03	0.52±0.15
KNN	0.87±0.03	0.86±0.03	0.46±0.16
SVPASEG	0.90±0.01	0.87±0.02	0.57±0.13
FANTASM	0.88±0.02	0.88±0.03	0.53±0.15
PVC	0.83±0.08	0.84±0.07	0.52±0.15

Table (a) shows the results on IBSR18 scans evaluated with original ground-truths. Table (b) shows the results on the same scans when not considering SCSF on the evaluation. Reported values are mean ± standard deviation. The highest mean Dice value for each tissue is shown in bold text.

obtained compared with IBSR20 images, due to the higher spatial resolution and the better image quality of IBSR18 images. SVPASEG and SPM8 provide the highest Dice values on GM, followed by KNN and GAMIXTURE. Again, it appears that SVPASEG and SPM8 take advantage of the prior atlas information to outperform the other methods. Surprisingly, SPM5 provides the lowest Dice value on original IBSR18 images. Unexpected low values for SPM5 on GM are caused by two factors: first, the provided high standard deviation suggests that the mean Dice value of the method has been affected by a low performance on punctual images. Second, although both SPM5 and SPM8 use a probabilistic atlas to guide the segmentation, the atlas is different among the two versions. The low Dice values yield by SPM5 appear to be caused by the own SPM5 atlas. The probability of SCSF voxels to pertain to CSF determines the amount of SCSF voxels that will be classified as CSF. Compared with the SPM8, this probability appears higher in the SPM5 atlas and most of the SCSF voxels are classified as CSF by the method.

On experiments without considering SCSF (see Table 4b), SPM5 provides values similar to the best methods on GM. FAST is also affected by this aspect, and Dice values on images without considering SCSF are notably higher than those evaluated with original

ground-truth. On WM, FAST is the method that provides the highest accuracy, followed by SPM8, FANTASM and FCM. In contrast, we observe that the performance of SVPASEG on WM with this second dataset is lower than previous results with IBSR20 images.

The ranking of methods returned by the permutation tests for the IBSR18 dataset with original and modified ground-truths is shown in Table 5. On images with original ground-truth, SVPASEG is the best ranked method on GM tissue, followed by SPM8, GAMIXTURE and KNN. The rest of methods are assigned to Rank 3. On WM, FAST is the only method that is ranked in the first group, while FANTASM is ranked in the second group and the rest of methods in Rank 3. As we have seen on IBSR20 images, the permutation tests on both tissues are again influenced by not considering SCSF. Thus, on GM FAST is now classified in Rank 1, while FANTASM and FCM are assigned to Rank 2. On WM, SPM8 is now assigned to Rank 2 in detriment of FANTASM, which is assigned to Rank 3.

## DISCUSSION

In the literature, there are numerous studies comparing the accuracy of their proposed methods with some of the methods of our study (31–33). In fact, we have reviewed other studies which also used IBSR20 and IBSR18 datasets to compare their results with our findings. These previous studies have used original IBSR ground-truths without extracting SCSF voxels.

On IBSR20 data, our results are similar to those reported in other studies for FAST (6,9), SPM5 (6), SPM8 (9), and SVPASEG (31). Our results for FCM and FAST are slightly higher than those reported also by Shahvaran et al (31) while lower for PVC than those published by Shattuck et al (17). The fact that we are observing differences in the accuracy for the same method between studies (FAST [6,9,31]) can be caused by changes in the preprocessing pipeline, initialization parameters or changes in the skull-stripping masks. On IBSR18 data, our findings are also similar to those published in other studies for FAST (32), SPM8 (32), FANTASM (33), GAMIXTURE (32), and SVPASEG (8).

On images evaluated with original ground-truths, the accuracy of methods is in general lower on IBSR20 than IBSR18 images, especially on WM. This fact is explained by the acquisition artifacts found in several images from IBSR20 that have a direct effect on WM tissue distributions, reducing the mean value of the methods, and increasing their variability. Furthermore, the lack of SCSF labeling on ground-truth masks appears to have a weak impact on WM tissue because the improvement on the performance of methods after not considering SCSF is inappreciable. In contrast, we have found that differences in SCSF have a direct impact on GM tissue accuracy. All methods tend to segment SCSF tissue as CSF which decreases the Dice values for both GM and CSF tissues.

Table 5

Permutation Tests for Obtained Dice Overlap Coefficients on the IBSR18 Dataset With Original Ground-Truth and Not Considering SCSF Voxels\*

IBSR18 (Original ground-truth)								IBSR18 (not considering SCSF)								
GM				WM				GM				WM				
	Method	$\mu \pm \sigma$			Method	$\mu \pm \sigma$			Method	$\mu \pm \sigma$				Method	$\mu \pm \sigma$	
Rank 1	SVPASEG	0.80±0.42	FAST	0.90±0.32	SPM8	0.70±0.48	FAST	0.90±0.32	SPM8	0.70±0.48	FAST	0.90±0.32	SPM8	0.70±0.48	FAST	0.90±0.32
	SPM8	0.70±0.48			SVPASEG	0.60±0.52										
	GAMIXTURE	0.50±0.71			FAST	0.40±0.70										
	KNN	0.40±0.84			GAMIXTURE	0.32±0.47										
Rank 2				FANTASM	0.32±0.65	FANTASM	0.10±0.74	SPM8	0.38±0.68							
					FCM	-0.22±0.79										
Rank 3	FAST	-0.10±0.88	SPM8	0.25±0.63	KNN	-0.29±0.49	FCM	0.25±0.63	SPM5	-0.32±0.64	FANTASM	0.24±0.63	SPM5	-0.32±0.64	FANTASM	0.24±0.63
	FANTASM	-0.11±0.86	FCM	0.19±0.57	SPM5	-0.50±0.71	ANN	0.24±0.63		-0.50±0.71	ANN	0.24±0.63		-0.50±0.71	ANN	0.24±0.63
	PVC	-0.40±0.52	ANN	0.11±0.56	PVC	-0.79±0.41	SPM5	-0.20±0.42		-0.79±0.41	SPM5	-0.20±0.42		-0.79±0.41	SPM5	-0.20±0.42
	SPM5	-0.59±0.51	GAMIXTURE	-0.13±0.56	KNN	-0.29±0.49	GAMIXTURE	-0.38±0.68		-0.29±0.49	KNN	-0.33±0.40		-0.29±0.49	KNN	-0.33±0.40
	FCM	-0.60±0.52	SVPASEG	-0.27±0.40	SPM5	-0.49±0.52	SVPASEG	-0.50±0.53		-0.49±0.52	SVPASEG	-0.50±0.53		-0.49±0.52	SVPASEG	-0.50±0.53
	ANN	-0.60±0.52	KNN	-0.28±0.40	PVC	-0.60±0.52	PVC	-0.60±0.52		-0.60±0.52	PVC	-0.60±0.52		-0.60±0.52	PVC	-0.60±0.52

\*Reported values are mean and standard deviation ( $\mu, \sigma$ ) of the fraction of times when each method produced significant  $P$  values. Positive values indicate that on average, the method over-performed the other methods in pair-wise significant tests. Negative values indicate the contrary. Rank 1:  $(\mu_o - \sigma_o, \mu_o]$ , Rank 2:  $(\mu_o - 2\sigma_o, \mu_o - \sigma_o]$ , Rank 3  $(\mu_o - 3\sigma_o, \mu_o - 2\sigma_o]$ . (as table 4).

The majority of the reviewed studies use the Dice coefficient to evaluate the accuracy of methods segmenting GM, WM, and CSF tissues. However, other measures such as the Sensitivity and Specificity, False Positive Fraction, and False Negative Fraction (13,26,27) can be used. To compare our results with these coefficients, we re-computed the accuracy of all ten methods using the Sensitivity and Specificity measures. These coefficients are inversely related with the FNR and FPR, respectively, and therefore the obtained results can be directly extrapolated to these measure rates. These new measurements showed that, in general, all methods tended to penalize one of these two coefficients in detriment of the other. Therefore, the rank of methods was clearly distinct, because the measures were only focused on positive or negative outcomes. For instance, PVC was the method from our study that clearly most overestimated WM tissue. Based on the Sensitivity coefficient, we found that PVC was the best ranked method on WM due to the low number of False Negatives. Conversely, it was the worst ranked method when the Specificity was evaluated. In contrast, as stated by the majority of the reviewed studies and our experiments, the Dice coefficient allows a better understanding of the overlap between the segmentation and the ground-truth masks because takes into account both positive and negative outcomes.

The permutation tests return the fraction of times that the mean Dice value of a current method is higher than the rest of the methods in pair-wise significant tests. We have observed that even the ranking of methods returned by each permutation test follows in general the same order produced by simply sorting the mean Dice values, the permutation test allows to disseminate better the differences among methods. Permutation test breaks the linearity of a ranking

based on sorting the mean Dice values by differentiating methods by the relevance of their results (30). Furthermore, a low performance of a method on one or several images of the dataset can decrease notably the mean Dice value and increase the standard deviation, while the permutation test tends to minimize the effect on the images with low performance.

Our results suggest that evaluating the accuracy of the 10 methods on original images of both IBSR datasets introduce an artificial bias, because most of the methods are penalized by the lack of SCSF on IBSR ground-truth. Thus, if we analyze the results on images where SCSF is not considered on the evaluation, SVPASEG is the best ranked method on both GM and WM tissues of IBSR20 images. SVPASEG, SPM5 and SPM8 are ranked in the first group on both tissues with significant mean values. On IBSR18 images, the results are not so clear and change between tissues. On GM, SPM8 and SVPASEG are the best ranked methods while FAST is the only method that is assigned to Rank 1 on WM. Both SPM methods on IBSR20, and FAST on IBSR18 are assigned to Rank 1 after removing SCSF from the evaluation. This fact is especially interesting, because these three methods are the most common used as baseline on comparative studies.

Furthermore, comparing the accuracy of the methods on both datasets can give us a better idea of the robustness of methods to different acquisition artifacts or their independence to changes produced by intensity correction. We observe that on GM, the performance of FAST, SPM5, GAMIXTURE, FCM, and FANTASM is dependent of the characteristics of the dataset. In contrast, ANN, PVC and KNN obtain the lowest performance on both datasets. Interestingly, SPM8 and SVPASEG are the only methods that are grouped to Rank 1 on both datasets. On

WM, FAST is the only method that is assigned to Rank 1 on both datasets, while the rest of methods present a different performance on each dataset. In our opinion, the fact that SVPASEG, SPM8 and FAST have been assigned to Rank 1 in three out of four permutation tests performed without considering SCSF, make these methods suitable for accurate brain tissue measurements.

Most of the methods were sensible to changes in acquisition sequences, intensity inhomogeneities or special attributes of the different datasets. In our opinion, the results of this paper highlight the fact that the brain tissue segmentation problem is still open, because there is not a single method that achieves a very high accuracy on all brain tissues. Although the design of more accurate methods should be the most common choice to follow on future research, it would be also interesting to analyze other alternative frameworks such as fusion processes based on the best segmentation results on each tissue to reduce the inner limitations of each individual brain tissue segmentation method.

The most important limitation of our study is the lack of new IBSR ground-truths that incorporate SCSF voxels as CSF tissue. However, given the limitation on this aspect, we propose to overpass the artificial bias introduced by the IBSR ground-truth SCSF voxel labels, by not considering these voxels into the accuracy measurements.

In conclusion, changes on original ground-truth annotations of IBSR images should be taken into account, especially in comparative studies that include several automatic segmentation methods. On these images, SCSF voxels are labeled as GM, and the inclusion of these voxels in the accuracy measurements can bias the results, due to differences in the amount of SCSF tissue classified as CSF by each method.

## ACKNOWLEDGMENT

S.V. was funded by a FI-DGR2013 grant from the Generalitat de Catalunya. E.R was funded by a grant UdG BR-GR13 from the Universitat de Girona.

## REFERENCES

- Cocosco CA, Kollokian V, Kwan RKS, Pike GB, Evans AC. Brainweb: online interface to a 3D MRI simulated brain database. *Neuroimage* 1997;5:425.
- Awate SP, Tasdizen T, Foster N, Whitaker RT. Adaptive markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification. *Med Image Anal* 2006;10:726-739.
- Akselrod-Ballin A, Galun M, Gomori M, Basri R, Brandt A. Atlas guided identification of brain structures by combining 3D segmentation and SVM classification. In: Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention 2006;4191:209-216.
- Caldairou B, Rousseau F, Passat N, Habas P, Studholme C, Heinrich C. A non-local fuzzy segmentation method: Application to brain MRI. *Comput Anal Images Patterns* 2009;5702:606-613.
- Wels M, Zheng Y, Huber M, Hornegger J, Comaniciu D. A discriminative model-constrained EM approach to 3D MRI brain tissue classification and intensity non-uniformity correction. *Phys Med Biol* 2011;56:3269.
- Tsang O, Gholipour A, Kehtarnavaz N, Gopinath K, Briggs R, Panahi I. Comparison of tissue segmentation algorithms in neuro-image analysis software tools. In: Proceedings of the 30th International Conference of the IEEE in Engineering in Medicine and Biology Society 2008;3924-3928.
- Kasiri K, Kazemi K, Dehghani M, Helfroush M. Atlas-based segmentation of brain MR images using least square support vector machines. In: Proceedings of the 2nd International Conference on Image Processing Theory Tools and Applications 2010;306-310.
- Tohka J, Dinov ID, Shattuck DW, Toga AW. Brain MRI tissue classification based on local markov random fields. *Magn Reson Imaging* 2010;28:557-573.
- Tian G, Xia Y, Zhang Y, Feng D. Hybrid genetic and variational expectation-maximization algorithm for gaussian-mixture-model-based brain MR image segmentation. *IEEE Transactions on Information Technol Biomed* 2011;15:373-380.
- Bazin PL, Pham D. Topology-preserving tissue classification of magnetic resonance brain images. *IEEE Trans Med Imaging* 2007;26:487-496.
- Shiee N, Bazin PL, Ozturk A, Reich DS, Calabresi PA, Pham DL. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *Neuroimage* 2010;49:1524-1535.
- Roy S, Carass A, Bazin PL, Resnick S, Prince JL. Consistent segmentation using a rician classifier. *Med Image Anal* 2012;16:524-535.
- Ortiz AR, Gorri JM, Ramirez J, Salas-Gonzalez D, Llamas-Elvira JM. Two fully-unsupervised methods for MR brain image segmentation using som-based strategies. *Appl Soft Comput* 2013;13:2668-2682.
- Filipek PA, Richelme C, Kennedy DN, Caviness VS. The young adult human brain: an MRI-based morphometric analysis. *Cereb Cortex* 1994;4:344-360.
- Sled JG, Zijdenbos AP, Evans CP. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 1998;17:87-97.
- Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 2001;20:45-57.
- Shattuck DW, Sandor-Leahy DR, Schaper K, Rottenberg DA, Leahy RM. Magnetic resonance image tissue classification using a partial volume model. *Neuroimage* 2001;13:856-876.
- Ashburner J, Friston KJ. Unified segmentation. *Neuroimage* 2005;26:839-851.
- Ashburner J, Barnes G, Chen CC, et al. SPM8 Manual. Wellcome Trust Centre for Neuroimaging Institute of Neurology, UCL 2011.
- Tohka J, Krestyannikov E, Dinov ID, et al. Genetic algorithms for finite mixture model based voxel classification in neuroimaging. *IEEE Trans Med Imaging* 2007;26:696-711.
- Pham DL. Spatial models for fuzzy clustering. *Computer Vision and Image Understanding* 2001;84:285-297.
- Pham DL. Robust fuzzy segmentation of magnetic resonance images. In: Proceedings of the 14th IEEE Symposium on Computer-Based Medical Systems, Bethesda, Maryland, 2001. p 127-131.
- Tian D, Fan L. A brain MR images segmentation method based on SOM neural network. *Bioinform Biomed Eng* 2007;2:686-689.
- De Boer R, Vrooman HA, Van Der Lijn F, et al. White matter lesion extension to automatic brain tissue segmentation on MRI. *Neuroimage* 2009;45:1151-1161.
- Dice LR. Measures of the amount of ecological association between species. *Ecology* 1945;26:297-302.
- Anbeek P, Vincken KL, Van Osch M, Bisschops R, Van der Grond J. Probabilistic segmentation of white matter lesions in MR imaging. *Neuroimage* 2004;21:1037-1044.
- Vrooman HA, Cocosco CA, van der Lijn F, et al. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage* 2007;37:71-81.
- Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans Med Imaging* 1994;13:716-724.
- Menke J, Martinez TR. Using permutations instead of student's t-distribution for p-values in paired-difference algorithm comparisons. Proceedings of the IEEE Joint Conference on Neural Networks, Budapest, 2004. p 1331-1335.

30. Klein A, Andersson A, Ardekani BA, et al. Evaluation of 14 non-linear deformation algorithms applied to human brain MRI registration. *Neuroimage* 2009;46:786–802.
31. Shahvaran Z, Kazemi K, Helfroush HS, Jafarian N, Noorizadeh N. Variational level set combined with markov random field modeling for simultaneous intensity non-uniformity correction and segmentation of MR images. *J Neurosci Methods* 2012; 209:280–289.
32. Zhang T, Xia Y, Feng D. Clonal selection algorithm for gaussian mixture model based segmentation of 3D brain MR images. In: Proceedings of the Second Sino-foreign-interchange Conference on Intelligent Science and Intelligent Data Engineering, Xi'an, China, 2011. p 295–302.
33. Nguyen TM, Wu QM. Robust student's-t mixture model with spatial constraints and its application in medical image segmentation. *IEEE Trans Med Imaging* 2012;31:103–116.

# **Chapter 3**

## **Evaluating the Effects of White Matter Multiple Sclerosis Lesions on the Volume Estimation of 6 Brain Tissue Segmentation Methods.**

In this chapter, we present an study of the impact of MS white matter lesions on the brain tissue measurements of six well-known segmentation techniques. These include straightforward techniques such as Artificial Neural Network (ANN) and fuzzy C-means (FCM) as well as more advanced techniques such as the Fuzzy And Noise Tolerant Adaptive Segmentation Method (FANTASM), FMRIB's Automated Segmentation Tool (FAST), and Statistical Parametric Mapping (SPM) with versions SPM5 and SPM8. This proposed evaluation has been published in the following paper:

Paper published in American Journal of Neuroradiology  
Volume: 36, Pages: 1109-1115, Published: February 2015  
DOI: 10.3174/ajnr.A4262  
Quality Index: 3.59 (Quartile 1)

# Evaluating the Effects of White Matter Multiple Sclerosis Lesions on the Volume Estimation of 6 Brain Tissue Segmentation Methods

S. Valverde,  A. Oliver, Y. Díez, M. Cabezas, J.C. Vilanova, L. Ramió-Torrentà, À. Rovira, and X. Lladó

## ABSTRACT

**BACKGROUND AND PURPOSE:** The accuracy of automatic tissue segmentation methods can be affected by the presence of hypointense white matter lesions during the tissue segmentation process. Our aim was to evaluate the impact of MS white matter lesions on the brain tissue measurements of 6 well-known segmentation techniques. These include straightforward techniques such as Artificial Neural Network and fuzzy C-means as well as more advanced techniques such as the Fuzzy And Noise Tolerant Adaptive Segmentation Method, fMRI of the Brain Automated Segmentation Tool, SPM5, and SPM8.

**MATERIALS AND METHODS:** Thirty T1-weighted images from patients with MS from 3 different scanners were segmented twice, first including white matter lesions and then masking the lesions before segmentation and relabeling as WM afterward. The differences in total tissue volume and tissue volume outside the lesion regions were computed between the images by using the 2 methodologies.

**RESULTS:** Total gray matter volume was overestimated by all methods when lesion volume increased. The tissue volume outside the lesion regions was also affected by white matter lesions with differences up to 20 cm<sup>3</sup> on images with a high lesion load ( $\approx 50$  cm<sup>3</sup>). SPM8 and Fuzzy And Noise Tolerant Adaptive Segmentation Method were the methods less influenced by white matter lesions, whereas the effect of white matter lesions was more prominent on fuzzy C-means and the fMRI of the Brain Automated Segmentation Tool.

**CONCLUSIONS:** Although lesions were removed after segmentation to avoid their impact on tissue segmentation, the methods still overestimated GM tissue in most cases. This finding is especially relevant because on images with high lesion load, this bias will most likely distort actual tissue atrophy measurements.

**ABBREVIATIONS:** ANN = Artificial Neural Network; FANTASM = Fuzzy And Noise Tolerant Adaptive Segmentation Method; FAST = FMRIB Automated Segmentation Tool; FCM = fuzzy C-means; H1 = Hospital Vall d'Hebron, Barcelona, Spain; H2 = Hospital Universitari Dr. Josep Trueta, Girona, Spain; H3 = Clinica Girona, Girona, Spain; WML = white matter lesion

**D**uring the past few years, MR imaging brain tissue segmentation techniques have become important tools in the clinical evaluation and progression of MS because they make it possible to measure the changes in brain atrophy and lesion load.<sup>1–3</sup> However, white matter lesions (WMLs) can significantly affect tissue volume measurements if these lesions are included in the segmentation process.<sup>4–6</sup> Several studies have analyzed the effects of WMLs on brain tissue measurements of common segmentation

techniques such as SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/>)<sup>7</sup> and FMRIB Automated Segmentation Tool (FAST, <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FAST>).<sup>8</sup> Chard et al<sup>5</sup> studied the effect of synthetic lesions on SPM5 segmentations for different WML voxel intensities (from 30% to 90% of normal WM intensity) and lesion loads (from 10 to 20 cm<sup>3</sup>). The authors reported that GM volume was overestimated by  $\approx 2.3\%$ , whereas WM tissue was underestimated by  $\approx 3.6\%$  in scans with 15 cm<sup>3</sup> of simulated lesions. More recently, Battaglini et al<sup>4</sup> also analyzed the effects of different WML intensities and lesion loads on tissue measurements obtained with FAST software. The authors showed again that total GM volume tended to increase with higher lesion loads in segmented images with generated simulated lesions. Gelineau-Morel et al<sup>6</sup> performed a similar study on the effects of simulated and real WMLs but on tissue volume measurements outside lesion regions. The authors reported that on images with simulated lesions, FAST clearly underestimated GM outside lesion regions as long as lesion volume increased and lesion intensities approximated those of GM tissue. The incidence of

Received August 19, 2014; accepted after revision December 20.

From the Computer Vision and Robotics Group (S.V., A.O., Y.D., X.L.), University of Girona, Campus Montilivi, Girona, Spain; Girona Magnetic Resonance Center (J.C.V.), Girona, Spain; Multiple Sclerosis and Neuroimmunology Unit (L.R.-T.), Dr. Josep Trueta University Hospital, Institut d'Investigació Biomèdica de Girona, Girona, Spain; and Magnetic Resonance Unit (M.C., A.R.), Department of Radiology, Vall d'Hebron University Hospital, Barcelona, Spain.

Please address correspondence to Sergi Valverde, Ed. P-IV, Campus Montilivi, University of Girona, 17071 Girona, Spain; e-mail: svalverde@eia.udg.edu

 Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

<http://dx.doi.org/10.3174/ajnr.A4262>

WMLs on real scans was smaller, but FAST still tended to underestimate GM with increasing lesion loads.

On the other hand, various studies have also analyzed the correlation between brain tissue atrophy and MS disability progression.<sup>9,10</sup> These studies showed a brain atrophy decrease rate between 0.3% and 0.5% of change in brain parenchyma per year in patients with MS,<sup>9,10</sup> with a decrease in GM and WM volume of up to 0.4% and 0.2% per year, respectively.<sup>10</sup> This statement along with study results such as those found by Battaglini et al<sup>4</sup> and Gelineau-Morel et al<sup>6</sup> indicates that a portion of brain atrophy could be hidden by the inclusion of WMLs on tissue segmentation.

In this study, we performed a quantitative evaluation of the effects of WMLs on brain tissue volume measurements to analyze the extent to which tissue estimations are affected by changes in WML volume and intensity. In contrast to other similar studies,<sup>4–6</sup> our analysis extended the number of segmentation methods involved, offering a comparative evaluation of the effects of WMLs on the volume measurements of 6 segmentation methods. Furthermore, given the reported correlation between brain atrophy rates and disability progression,<sup>9,10</sup> it can be clinically relevant for the MS community to extend the analysis of the effects of simulated WML to real data of patients with MS; hence, our analysis was focused exclusively on data from the T1-weighted images from patients with clinically confirmed MS.

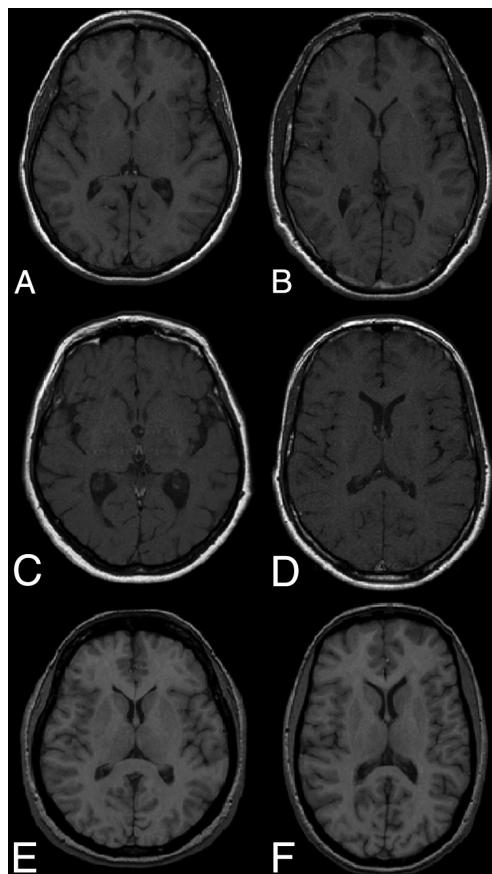
## MATERIALS AND METHODS

### Image Acquisition

The dataset consisted of 30 MR images from patients with clinically confirmed MS at 3 different hospitals (Fig 1). Each patient underwent MR imaging by using the same protocol (T1-weighted, T2-weighted, proton-attenuation-weighted, and FLAIR images), though a different scanner was used at each hospital. Ten patient images from Hospital Vall d'Hebron, Barcelona, Spain, (H1) were acquired on a 1.5T Magnetom Symphony Quantum (Siemens, Erlangen, Germany), with 2D conventional spin-echo T1-weighted (TR, 450 ms; TE, 17 ms), dual-echo proton-attenuation T2-weighted sequences (TR, 3750 ms; TE, 14/86 ms), and FLAIR sequences (TR, 9000 ms; TE, 114 ms; and TI, 2500 ms). Ten patient images from Hospital Universitari Dr. Josep Trueta, Girona, Spain, (H2) were acquired on a 1.5T Intera scanner (R12) (Philips Healthcare, Best, the Netherlands) with 2D conventional spin-echo T1-weighted (TR, 653 ms; TE, 14 ms), dual-echo proton-attenuation T2-weighted (TR, 2800 ms; TE, 16/80 ms), and FLAIR sequences (TR, 8153 ms; TE, 105 ms; and TI, 2200 ms). Ten patient images from Clinica Girona, Girona, Spain, (H3) were acquired on a 1.5T Signa HDxt scanner (GE Healthcare, Milwaukee, Wisconsin) with 3D fast-spoiled gradient T1-weighted (TR, 30 ms; TE, 9 ms; flip-angle, 30°), fast spin-echo T2-weighted (TR, 5000–5600 ms; TE, 74–77 ms), proton-attenuation-weighted (TR, 2700 ms; TE, 11.9 ms), and FLAIR sequences (TR, 9002 ms; TE, 80 ms; and TI, 2250 ms). All images were acquired in the axial view with a section thickness of 3 mm.

### Images of Patients with MS

WML masks were semiautomatically delineated from proton-attenuation-weighted images by using Jim software (Xinapse Systems,



**FIG 1.** T1-weighted images from the 3 hospitals and scanners involved in the study: 1.5T Magnetom Symphony Quantum (Siemens) from H1 (first row), 1.5T Intera (R12) (Philips) from H2 (middle row), and 1.5T Signa HDxt (GE Healthcare) from H3 (last row).

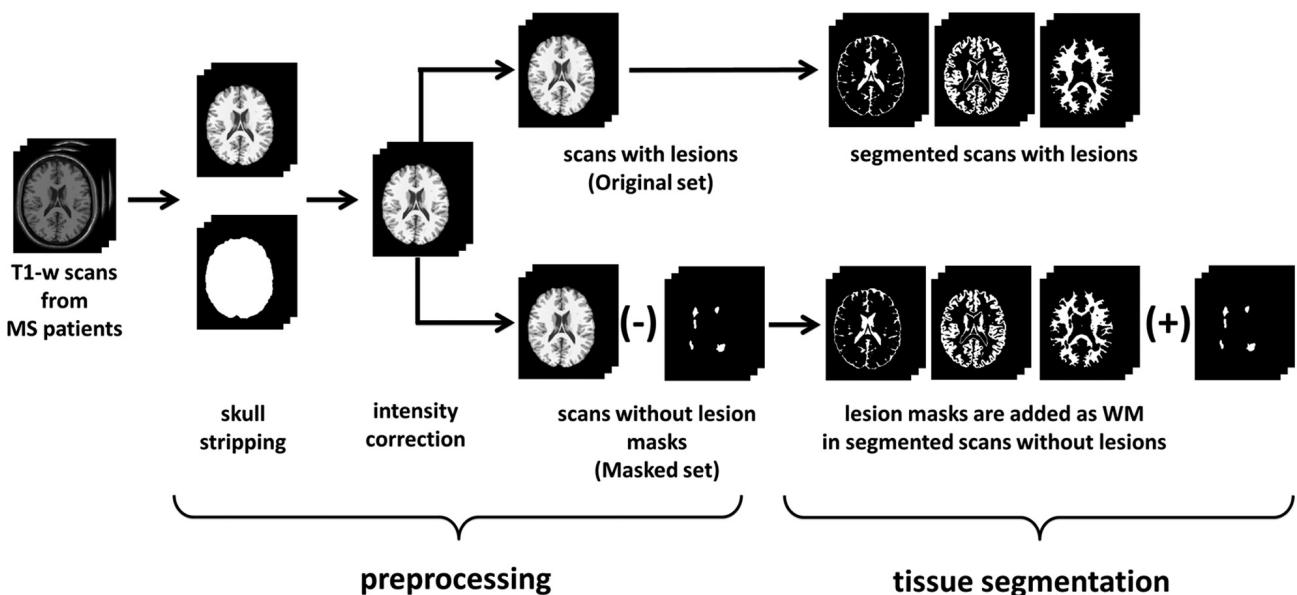
<http://www.xinapse.com/home.php>) by expert radiologists at each hospital. Then, the proton-attenuation-weighted images and lesion masks were coregistered with T1-weighted images by affine registration.<sup>11</sup> The average means and SD lesion volumes for H1, H2, and H3 were  $4.15 \pm 4.35 \text{ cm}^3$  (minimum = 0.11, maximum = 11.22  $\text{cm}^3$ ),  $21.79 \pm 17.79 \text{ cm}^3$  (minimum = 0.18, maximum = 52.45  $\text{cm}^3$ ), and  $4.78 \pm 4.60 \text{ cm}^3$  (minimum = 0.43, maximum = 16.34  $\text{cm}^3$ ).

All T1-weighted patient images were processed following the same pipeline (Fig 2). Internal skull-stripping and intensity-correction options were disabled on SPM5, SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/software/>), and FAST. Instead, to reduce the differences in brain area and signal image intensity produced by different preprocessing tools, we skull-stripped all images by using the Brain Extraction Tool (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET>)<sup>12</sup> and intensity-corrected them by using N3.<sup>13</sup>

As a second step, 2 sets were produced from preprocessed images: an original set that included WMLs as part of current tissue and a masked set in which the WMLs were masked out before tissue segmentation and relabeled as WM after, following the same procedure used by radiologists of the 3 hospitals.

### Segmentation Methods

The set of methods was composed of 6 well-known automatic brain tissue segmentation techniques: Artificial Neural Network (ANN), fuzzy C-means (FCM), Fuzzy And Noise Tolerant Adaptive Segmentation Method (FANTASM), FAST, SPM5, and



**FIG 2.** Our pipeline approach. From the 30 T1-weighted scans of patients with MS, nonbrain parts are stripped and brain voxels are corrected for intensity inhomogeneities. From the same corrected set (original), a new set is generated by removing WML masks from scans before segmentation (masked). The scans of both sets are segmented into 1 of the 3 tissue classes (GM, WM, and CSF). Lesion voxels are added as WM after segmentation on masked images.

SPM8. ANN and FCM were implemented for our study, while the rest of the methods were obtained from available repositories. The ANN method is based on self-organizing maps, also known as Kohonen networks.<sup>14</sup> ANN was implemented for our study by using the Matlab 7.12 environment (MathWorks, Natick, Massachusetts) following the technique proposed by Tian et al.<sup>15</sup> FCM<sup>16</sup> and FANTASM<sup>17</sup> are both based on fuzzy-clustering techniques. FCM implements the classic fuzzy-clustering approach, while FANTASM adds neighboring information to increment the robustness of the method to intensity inhomogeneity artifacts and noise. FCM was also implemented by using the Matlab environment and following the technique described in Pham,<sup>16</sup> in which clusters were initialized according to Bezdek et al.<sup>18</sup> FANTASM is included in the MIPAV toolbox (<http://mipav.cit.nih.gov>). FAST<sup>8</sup> guides the segmentation with spatial information through the optimization of Hidden Markov Random Fields, and the method is included in the fMRI of the Brain Software Library toolbox (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>). SPM5 and SPM8<sup>7</sup> are based on an iterative Gaussian Mixture Model optimization, weighting the probability of belonging to a certain tissue class with a priori spatial information from tissue-probability atlases. However, SPM8 comes with a set of different characteristics to improve registration and tissue segmentation. Both methods are included in the SPM8 toolbox (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8>). All methods were run with default parameters.

### Evaluation

Images from both the original and masked sets were segmented into GM, WM, and CSF tissue classes by using the 6 presented segmentation methods. Then, we computed the normalized tissue volumes as the number of voxels classified as GM, WM, and CSF, respectively, divided by the total number of voxels. Three different analyses were performed on these data. First, we analyzed how lesion voxels were classified by each segmentation

method to establish to what extent the tissue volumes reported by each algorithm on the original and masked images could be expected to be different. Second, we analyzed the direct effect of lesions in the global volume estimation by computing the differences in total tissue volume as the percentage of change between original and masked images. For example, in the case of GM tissue:

$$\% GM = \frac{NGMV_{\text{Original}} - NGMV_{\text{Masked}}}{NGMV_{\text{Masked}}} \times 100,$$

where  $NGMV_{\text{Original}}$  and  $NGMV_{\text{Masked}}$  stand for the normalized gray matter volumes of original and masked images, respectively. Third, we also investigated the indirect effects of lesions in the rest of the tissue volume outside lesion regions. These are tissue volume estimations that incorporate lesions in the segmentation process but do not consider them when the volume is evaluated.

### Statistical Analysis

The correlation among factors (differences in tissue volume, lesion load, and lesion intensity) was calculated by using Pearson linear correlation coefficient ( $r$ ). The significance level  $\alpha$  was set at .05. This level was used both for confidence interval computation and 95% significance hypothesis 2-tailed  $t$  tests. All statistical analyses were calculated by using the Matlab environment.

## RESULTS

### Lesion Classification

Figure 3 depicts the percentage of WML voxels classified either as WM (Fig 3, top) or GM (Fig 3, bottom). Percentages are detailed for each segmentation method and hospital. The amount of WMLs that were classified as GM varied for each method, mostly due to the differences among algorithms. Figure 4 illustrates the differences among methods by showing the output classification performed by each of the 6 segmentation methods.

Observed differences in the percentage of classified WML vox-

els as GM and WM between hospitals can be attributed to each particular scanner acquisition configuration that defines the tissue signal-intensity distributions. The distance between WML and WM mean signal intensities was highest in H3 as computed by each of the 6 methods (range, from  $89.2 \pm 4.45\%$  to  $92.22 \pm 4.45\%$  of WML mean signal intensity with respect to WM) and was lowest in H2 (range, from  $95.3 \pm 1.76\%$  to  $100.34 \pm 6.39\%$ ). As shown in Fig 1, there is a better contrast between GM and WM tissue on the H3 images compared with the H1 and H2 images.

The correlation between the percentage of lesion classification and lesion size was not significant in all cases ( $r < 0.33$ ,  $P > .05$ ). In contrast, the percentage of WML classified as GM or WM and the distance between the mean WML and WM signal intensities showed a moderate correlation in all hospitals ( $r > 0.6$ ,  $P < .01$ ). On the basis of our data, the contrast between tissues computed as the normalized difference between the mean GM and WM signal intensity distributions was correlated with the distance between the WM and WML mean signal intensities ( $r = 0.6$ ,  $P < .001$ ).

#### Differences in Total Tissue Volume Estimation

The mean percentage differences in total tissue volume between the original and masked images are presented in Table 1. All methods overestimated GM tissue in original scans, regardless of

the hospital, but the overestimation was increased in H2 compared with H1 and H3 due to greater lesion volumes in H2. The differences among methods for the same hospital and tissue were also significantly greater in H2 than in H1 and H3. Abnormally low mean and high SD values observed in SPM5 for both GM ( $0.10 \pm 2.68$ ) and WM ( $1.04 \pm 3.01$ ) in H2 were caused by 2 patients who exhibited very high opposite differences between their respective original and masked images, decreasing the overall mean difference and increasing the SD.

Correlation between the differences in total mean tissue volume and lesion size was significant in all hospitals: Lesion size had a direct effect on tissue segmentation. Table 2 shows the Pearson correlation values obtained between differences in tissue volume and lesion size across methods. All methods except SPM5 presented a positive correlation in GM and a negative correlation in WM in H1 and H2. SPM5 correlated in H1 but not in H2, where it was influenced by abnormal values in the 2 images with highest lesion load. In H3, only FCM, FANTASM, and FAST were positively correlated in GM and negatively correlated in WM. The correlation coefficients for ANN, SPM5, and SPM8 in H3 were weak and not significant in GM and WM.

#### Volume Estimation of Tissue Outside Lesion Regions

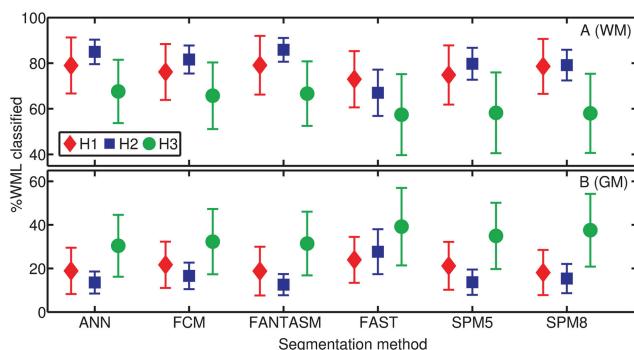
The mean percentage differences in tissue volume outside lesion regions between original and masked images are presented in Table 3. The differences between the images segmented with lesions and images in which the lesions were masked before tissue segmentation were again higher in H2, and the methods still substantially overestimated the GM outside the lesion regions to the detriment of WM, even though analyzed tissues were free of lesion regions. In contrast, only SPM5 and SPM8 reported a noticeable underestimation of GM in H3, also to the detriment of WM.

Differences in tissue volume outside the lesion regions correlated with lesion size for all tissues and hospitals, indicating an effect of lesion size not only on lesion voxels but also on tissue that is not affected by lesions. Table 4 presents the correlation values obtained across methods. In H1, there was a remarkable correlation for ANN, FCM, FANTASM, and FAST in all tissues. The obtained values for

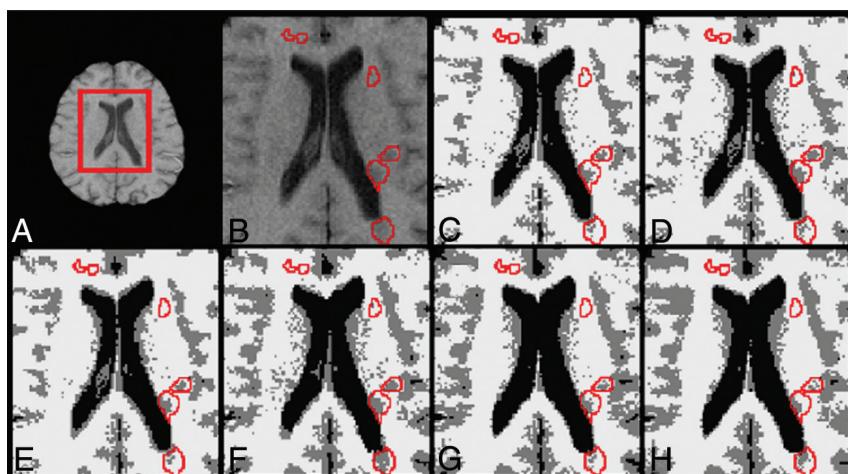
SPM8 were also significant in GM and CSF. In H2, the correlation was significant in ANN, FCM, and FANTASM in all tissues. In H3, only FCM and FAST showed a significant correlation in all tissues, whereas FCM, FAST, SPM5, and SPM8 correlated significantly only in WM. All methods except SPM5 and SPM8 reported a significant correlation for CSF.

#### DISCUSSION

Previous studies have shown that the range of voxel signal intensities composing each of the tissue distributions can be altered by WMLs if these voxels are included in the segmentation process.<sup>4,5</sup> Lesion load and the apparent lesion signal intensity lead to observed changes in tissue segmentation in original images.



**FIG 3.** Percentage of voxels in WML regions having been classified as GM (top) and WM (bottom) for each segmentation method and hospital, H1 (◊), H2 (□) or H3 (○). Reported values are means and SDs.



**FIG 4.** Classification output returned by each segmentation method on the same image. A, T1-weighted scan. B, Zoomed part of the scan with lesions outlined in red. Brain tissue segmentation outputs also with lesions outlined for ANN (C), FCM (D), FANTASM (E), FAST (F), SPM5 (G), and SPM8 (H). C–H, Segmented GM tissue is represented in gray; WM, in white; and CSF, in black.

**Table 1: Average percentage of change in total tissue volume estimation between original and masked images<sup>a</sup>**

Method	H1			H2			H3		
	GM	WM	CSF	GM	WM	CSF	GM	WM	CSF
ANN	0.33 ± 0.42	-0.23 ± 0.28	0.11 ± 0.11	1.59 ± 1.37	-0.56 ± 0.46	0.78 ± 0.76	0.25 ± 0.31	-0.16 ± 0.28	-0.09 ± 0.09
FCM	0.28 ± 0.37	-0.22 ± 0.29	0.09 ± 0.11	2.28 ± 2.26	-0.90 ± 0.83	0.94 ± 0.90	0.28 ± 0.23	-0.25 ± 0.20	0.08 ± 0.09
FANTASM	0.23 ± 0.26	-0.18 ± 0.21	0.08 ± 0.08	1.34 ± 1.13	-0.49 ± 0.37	0.80 ± 0.73	0.26 ± 0.22	-0.24 ± 0.19	0.07 ± 0.08
FAST	0.29 ± 0.36	-0.29 ± 0.36	0.12 ± 0.13	1.92 ± 1.59	-1.28 ± 1.03	0.47 ± 0.39	0.34 ± 0.28	-0.37 ± 0.31	0.12 ± 0.17
SPM5	0.20 ± 0.30	-0.21 ± 0.20	-0.14 ± 0.54	0.10 ± 2.68	-1.04 ± 3.01	0.53 ± 0.51	0.04 ± 0.17	-0.18 ± 0.36	0.15 ± 0.23
SPM8	0.08 ± 0.09	-0.08 ± 0.08	-0.04 ± 0.18	0.55 ± 0.34	-0.93 ± 0.55	0.54 ± 0.42	0.09 ± 0.15	-0.23 ± 0.25	0.17 ± 0.23

<sup>a</sup> The results are divided by tissue and hospital. Reported values are the means ± SD. Positive values indicate a tissue overestimation on original images compared with masked.

**Table 2: Pearson correlation coefficients between method differences in total volume estimation and WML size<sup>a</sup>**

Method	GM	WM	CSF
H1			
ANN	0.94	-0.90	0.89
FCM	0.93	-0.89	0.83
FANTASM	0.87	-0.80	0.78
FAST	0.97	-0.97	0.96
SPM5	0.58 <sup>b</sup>	-0.89	-0.21 <sup>b</sup>
SPM8	0.92	-0.63	-0.69
H2			
ANN	0.91	-0.88	0.93
FCM	0.92	-0.94	0.92
FANTASM	0.89	-0.87	0.84
FAST	0.95	-0.96	0.82
SPM5	-0.35 <sup>b</sup>	-0.06 <sup>b</sup>	0.72
SPM8	0.76	-0.79	0.57 <sup>b</sup>
H3			
ANN	0.56 <sup>b</sup>	-0.55 <sup>b</sup>	0.88
FCM	0.77	-0.84	0.88
FANTASM	0.74	-0.82	0.85
FAST	0.88	-0.94	0.92
SPM5	-0.06 <sup>b</sup>	-0.03 <sup>b</sup>	0.21 <sup>b</sup>
SPM8	0.56 <sup>b</sup>	-0.48 <sup>b</sup>	0.09 <sup>b</sup>

<sup>a</sup> Correlation was computed for each method and hospital separately. All values were found to be significant (*P* value < .05) unless otherwise noted.

<sup>b</sup> Not significant.

For instance, if a portion of the lesion voxels is classified as WM, the mean overall WM intensity decreases, shifts WM boundaries into darker intensities, and narrows GM tissue distribution.<sup>4,6</sup> Voxels that should have been classified as GM are assigned to WM, increasing the WM volume estimation and decreasing GM volume. If some of the WML voxels are classified as GM, the apparent GM mean intensity increases and the WM tissue distribution narrows. This change occurs because voxels that are theoretically classified as WM are assigned to GM, increasing GM estimation against a lower WM volume estimation.

We compare our results with those in previous studies regarding the effects of WMLs on brain tissue volume measurements. However, given the differences in image data, criterion standards, simulated lesions, and lesion voxel intensities among studies, a direct comparison further than an analysis of trends with similar WML intensities and lesion loads should be carefully performed. Our experiments follow the same trend presented by Battaglini et al,<sup>4</sup> and both studies show that FAST overestimates total GM volume on images segmented with lesions. Similarly, our results also coincide with those found by Chard et al<sup>5</sup> in simulated data, and in both studies, SPM5 overestimated GM tissue on images with lesions. In contrast, our results appear to be inconsistent with those reported by Gelineau-Morel et al.<sup>6</sup> These studies showed a significant correlation between WML intensity and an underesti-

mation of GM volume outside the lesions, especially when the lesions had intensities similar to those of the mean GM. The observed differences are caused by distinct signal-intensity profiles of WMLs in each study. In the case of Gelineau-Morel et al,<sup>6</sup> the WML signal intensities were noticeably more hypointense compared with our data. The probability of voxels to be classified as GM dropped as a result of the influence of hypointense WML intensities in tissue distributions. Part of WML voxels with a signal intensity similar to that of GM were still classified as WM, reducing the signal intensity threshold between GM and WM. As a result, most of the partial volume voxels with signal intensity in the boundary between GM and WM were classified as WM, artificially reducing the overall number of GM voxels.

Our results show that the classification of WML regions is highly dependent on lesion voxel signal intensities and the variation of their signal intensity in terms of the WM signal distribution. Lesion segmentation is clearly determined by this variation because the probability of WML voxels being classified as WM will be higher as long as WML intensities resemble those of WM. However, the signal-intensity contrast among tissues also plays an important role because it can influence the amount of WML voxels that are classified as GM or WM. As long as the contrast among distributions increases, more lesion voxels will be added into the GM distribution. Although the main factor in the observed differences in tissue volume across methods is caused by lesion volume, the percentage of lesion voxels that are classified as GM and WM might also be a remarkable factor in the observed tissue-volume differences, especially in images with high lesion loads. Therefore, the relationship between image quality and lesion load also might have to be considered to explain the differences in tissue volume.

SPM8 was the method with the lowest difference in total tissue volume between original and masked images. In contrast, FAST was the method that was more affected by lesions. In general, all methods overestimated GM in original scans, though values were more significant in H2 than H1 and H3 due to higher lesion loads in H2. In H1 and H3, most of the underestimated WM was shifted into GM. The small percentage of lesions that were segmented as CSF, especially the low lesion volume, limited the impact of WML voxels on the overall CSF tissue distribution of original images.

SPM8 and FANTASM were the methods with the lowest incidence of WML in tissue volume measurements outside lesion regions, while FCM and FAST showed the largest differences among all methods. Lesion volume also explains the limited effect of WML on tissue segmentation outside lesion regions in H1 and H3, compared with images with higher lesion loads such as the H2 images. In H1 and H3, although the behavior differs slightly for

**Table 3: Average percentage change in the volume estimation of tissue outside the lesion regions between original and masked scans<sup>a</sup>**

Method	H1			H2			H3		
	GM	WM	CSF	GM	WM	CSF	GM	WM	CSF
ANN	0.15 ± 0.26	-0.10 ± 0.18	0.07 ± 0.08	0.70 ± 0.61	-0.31 ± 0.24	0.67 ± 0.69	-0.01 ± 0.28	0.04 ± 0.24	-0.12 ± 0.08
FCM	0.09 ± 0.16	-0.07 ± 0.13	0.05 ± 0.08	1.27 ± 1.69	-0.56 ± 0.62	0.82 ± 0.81	0.01 ± 0.03	-0.03 ± 0.05	0.05 ± 0.07
FANTASM	0.06 ± 0.05	-0.05 ± 0.05	0.03 ± 0.05	0.48 ± 0.48	-0.25 ± 0.18	0.68 ± 0.63	0.00 ± 0.04	-0.02 ± 0.05	0.04 ± 0.07
FAST	0.08 ± 0.14	-0.09 ± 0.14	0.07 ± 0.08	0.56 ± 0.87	-0.45 ± 0.64	0.22 ± 0.33	0.02 ± 0.07	-0.06 ± 0.13	0.08 ± 0.16
SPM5	0.06 ± 0.25	0.02 ± 0.13	-0.19 ± 0.54	-0.29 ± 2.61	-0.47 ± 2.91	0.21 ± 0.32	-0.20 ± 0.24	0.23 ± 0.34	0.06 ± 0.15
SPM8	-0.03 ± 0.06	0.09 ± 0.15	-0.10 ± 0.23	0.13 ± 0.30	-0.29 ± 0.33	0.25 ± 0.26	-0.15 ± 0.12	0.14 ± 0.15	0.10 ± 0.20

<sup>a</sup> The results are divided by hospital and tissue. Reported values are the means ± SD. Positive values indicate a tissue overestimation on original images compared with masked.

**Table 4: Pearson correlation coefficients among method differences in volume estimation of tissue outside the lesion regions and WML size<sup>a</sup>**

Method	GM	WM	CSF
H1			
ANN	0.77	-0.74	0.83
FCM	0.82	-0.80	0.71
FANTASM	0.80	-0.73	0.66
FAST	0.86	-0.93	0.97
SPM5	0.11	0.51 <sup>b</sup>	-0.30 <sup>b</sup>
SPM8	-0.57 <sup>b</sup>	0.95	-0.77
H2			
ANN	0.85	-0.92	0.93
FCM	0.71	-0.84	0.94
FANTASM	0.66	-0.82	0.87
FAST	0.33 <sup>b</sup>	-0.46 <sup>b</sup>	0.62 <sup>b</sup>
SPM5	-0.43 <sup>b</sup>	0.18 <sup>b</sup>	0.65 <sup>b</sup>
SPM8	0.16 <sup>b</sup>	-0.37 <sup>b</sup>	0.30 <sup>b</sup>
H3			
ANN	0.07	-0.16 <sup>b</sup>	0.79
FCM	0.50	-0.77	0.89
FANTASM	0.17	-0.57 <sup>b</sup>	0.87
FAST	0.45	-0.73	0.89
SPM5	-0.78 <sup>b</sup>	0.72	0.14 <sup>b</sup>
SPM8	-0.64 <sup>b</sup>	0.72	-0.01 <sup>b</sup>

<sup>a</sup> Correlation was computed for each method and hospital separately. All values were found to be significant (*P* value <.05) unless otherwise noted.

<sup>b</sup> Not significant.

each method, the differences in tissues outside the lesion regions are very small.

The differences outside the lesion regions are especially important because they highlight the bias introduced by WMLs on the estimation of tissue volume that is not pathologically affected. If one compares the results between total tissue volume and tissue volume outside lesion regions, it can be observed that an important part of the overestimated total GM is essentially derived from the same hypointense WML voxels that are classified as GM. Moreover, it is important to highlight the differences in the algorithms. Methods such as FCM and ANN, which only rely on signal intensity, introduce more errors in tissue segmentation compared with methods such as SPM8 and SPM5, which incorporate spatial information. This reinforces the necessity for selecting a segmentation algorithm that does not depend on signal intensity only. However, even though WML voxels have not been considered for computing tissue volume outside the lesion regions, there is still a clear tendency toward overestimating GM. On images with a high lesion load, the observed differences in GM volume outside lesion regions reach values that are equivalent to the yearly expected GM atrophy.<sup>9,10</sup> Following these assumptions, SPM8, FANTASM, and SPM5 are the methods with the lowest reported incidence of

WML on brain tissue volume measurements, especially on images with a high lesion load.

The present study is not free of limitations. The principal limitation is the lack of tissue expert annotations, given that the study incorporated a relatively large number of images from 3 different hospitals and this task was time-consuming. A second limitation of the study is the sensitivity of the tissue segmentation methods to changes in the skull-stripping mask. Errors in the brain mask may lead to the inclusion of blood vessels such as the internal carotid arteries with hyperintense signal intensity, which might bias the tissue distributions. A final limitation of the study is the inherent difficulty of comparing previous studies, given the differences in the scanner protocols used to acquire the images of patients with MS. The differences in the acquisition protocol may cause the observed differences in the lesion intensity profile compared with previous works.<sup>8,10</sup> Our study shows that such an intensity profile introduces variations in GM and WM tissue distributions.

## CONCLUSIONS

The results of this study indicate a direct relationship between the differences in brain tissue volume and changes in lesion load and WML intensity. Of the analyzed methods, SPM8 exhibited the lowest incidence of WMLs in volume estimation, whereas FCM yielded the highest GM overestimation. Furthermore, all methods were affected by WMLs in tissue volume outside the lesion regions. SPM8 and FANTASM exhibited the lowest differences in tissue volume outside the lesion regions, whereas the influence of WMLs outside the lesion regions is more important in methods such as FCM and FAST. The latter results are especially important because even when masking lesions after segmentation to avoid the inclusion of lesion voxels segmented as GM into the volume estimation, the methods tend to overestimate GM tissue on images segmented with lesions. On images with high lesion load, this bias might conceal or falsify part of the GM and WM tissue atrophy.

Disclosures: Sergi Valverde—RELATED: Grant: FI-DGR2013 research grant from the Generalitat de Catalunya. Mariano Cabezas—RELATED: holds a 2014 European Committee for Treatment and Research in Multiple Sclerosis—Magnetic Resonance Imaging in MS research fellowship.

## REFERENCES

- Lladó X, Oliver A, Cabezas M, et al. Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches. *Information Sciences* 2012;186:164–85
- Lladó X, Ganlier O, Oliver A, et al. Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology* 2012;54:787–807

3. Zivadinov R, Bergsland N, Dolezai O, et al. **Evolution of cortical and thalamus atrophy and disability progression in early relapsing-remitting MS during 5 years.** *AJNR Am J Neuroradiol* 2013;34:1931–39
4. Battaglini M, Jenkinson M, De Stefano N. **Evaluating and reducing the impact of white matter lesions on brain volume measurements.** *Hum Brain Mapp* 2012;33:2062–71
5. Chard DT, Jackson JS, Miller DH, et al. **Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes.** *J Magn Reson Imaging* 2010;32:223–28
6. Gelineau-Morel R, Tomassini V, Jenkinson M, et al. **The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis.** *Hum Brain Mapp* 2012;33:2802–14
7. Ashburner J, Friston K. **Unified segmentation.** *Neuroimage* 2005;26:839–51
8. Zhang Y, Brady M, Smith S. **Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm.** *IEEE Trans Med Imaging* 2001;20:45–57
9. Horsfield MA, Rovaris M, Rocca MA, et al. **Whole-brain atrophy in multiple sclerosis measured by two segmentation processes from various MRI sequences.** *J Neurol Sci* 2003;216:169–77
10. Rudick RA, Lee JC, Nakamura K, et al. **Gray matter atrophy correlates with MS disability progression measured with MSFC but not EDSS.** *J Neurol Sci* 2009;282:106–11
11. Rueckert D, Sonoda LI, Hayes C, et al. **Nonrigid registration using free-form deformations: application to breast MR images.** *IEEE Trans Med Imaging* 1999;18:712–21
12. Smith SM. **Fast robust automated brain extraction.** *Hum Brain Mapp* 2002;17:143–55
13. Sled JG, Zijdenbos AP, Evans CP. **A nonparametric method for automatic correction of intensity nonuniformity in MRI data.** *IEEE Trans Med Imaging* 1997;17:87–97
14. Kohonen T. **The self-organizing map.** *Proceedings of the IEEE* 1990;78:1464–80
15. Tian D, Fan L. **A brain MR images segmentation method based on SOM neural network.** In: *Proceedings of the 1st International Conference on Bioinformatics and Biomedical Engineering ICBBE*, Wuhan, People's Republic of China. July 6–8, 2007:686–89
16. Pham DL. **Spatial models for fuzzy clustering.** *Computer Vision and Image Understanding* 2001;84:285–97
17. Pham DL. **Robust fuzzy segmentation of magnetic resonance images.** In: *Proceedings of the Fourteenth IEEE Symposium on Computer-Based Medical Systems*, Bethesda, Maryland. July 26–27, 2001:127–31
18. Bezdek JC, Keller J, Krishnapuram R, et al. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing.* Norwell: Kluwer Academic Publishers; 1999:1650.F89

# **Chapter 4**

## **A white matter lesion-filling approach to improve brain tissue volume measurements.**

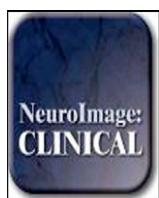
In this chapter, we propose a new technique to fill WM lesions before tissue segmentation. The proposed approach is evaluated in both 1.5T and 3T data. We validate our method comparing its accuracy with other proposed automated lesion filling methods on the same data. Furthermore, the proposed technique has been released for public use both as a standalone program or as SPM8/SPM12 library. This work has been published in the following paper:

NeuroImage: Clinical

Volume: 6, Pages: 86-92, Published: August 2014

DOI: 10.1016/j.nicl.2014.08.016

Quality Index: 2.53 (Quartile 2)



## A white matter lesion-filling approach to improve brain tissue volume measurements



Sergi Valverde<sup>a,\*</sup>, Arnau Oliver<sup>a</sup>, Xavier Lladó<sup>a</sup>

<sup>a</sup>Computer Vision and Robotics Group, University of Girona, Campus Montilivi, Ed. P-IV, Girona 17071, Spain

### ARTICLE INFO

#### Article history:

Received 3 April 2014

Received in revised form 16 August 2014

Accepted 21 August 2014

Available online 23 August 2014

#### Keywords:

Brain

MRI

Multiple sclerosis

Tissue segmentation

White matter lesions

Lesion-filling

### ABSTRACT

Multiple sclerosis white matter (WM) lesions can affect brain tissue volume measurements of voxel-wise segmentation methods if these lesions are included in the segmentation process. Several authors have presented different techniques to improve brain tissue volume estimations by filling WM lesions before segmentation with intensities similar to those of WM. Here, we propose a new method to refill WM lesions, where contrary to similar approaches, lesion voxel intensities are replaced by random values of a normal distribution generated from the mean WM signal intensity of each two-dimensional slice. We test the performance of our method by estimating the deviation in tissue volume between a set of 30 T1-w 1.5 T and 30 T1-w 3 T images of healthy subjects and the same images where: WM lesions have been previously registered and afterwards replaced their voxel intensities to those between gray matter (GM) and WM tissue. Tissue volume is computed independently using FAST and SPM8. When compared with the state-of-the-art methods, on 1.5 T data our method yields the lowest deviation in WM between original and filled images, independently of the segmentation method used. It also performs the lowest differences in GM when FAST is used and equals to the best method when SPM8 is employed. On 3 T data, our method also outperforms the state-of-the-art methods when FAST is used while performs similar to the best method when SPM8 is used. The proposed technique is currently available to researchers as a stand-alone program and as an SPM extension.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

### 1. Introduction

Magnetic resonance imaging (MRI) permits to assess tissue abnormalities *in vivo* and approximate histopathological changes of the multiple sclerosis (MS) disease (Ganiler et al., 2014; Kearney et al., 2014). Several studies have shown that the percentage of change in brain atrophy tends to correlate with the progression of the disease (Pérez-Miralles et al., 2013; Sormani et al., 2014). Moreover, changes in gray matter (GM) atrophy are observed independently from white matter (WM), and hence atrophy measures based on segmentation-based methods are nowadays employed as they allow classifying brain tissues separately (Pérez-Miralles et al., 2013). The performance of different segmentation methods used to quantify brain atrophy or volume estimation has been evaluated deeply in the last 5 years (Klauschen et al., 2009; Derakhshan et al., 2010). However, it is well known that the presence of WM lesions can induce errors on brain tissue volume measurements (Chard et al., 2010; Battaglini et al., 2012; Gelineau-Morel et al., 2012) and non-rigid registration (Sdika and Pelletier, 2009; Diez et al., 2014), if lesions are processed within the images. For instance, if WM lesion voxels are classified as WM, lesion voxels with hypointense signal

intensities are added into the WM tissue distribution, increasing the probability of GM voxels with similar intensity to be misclassified also as WM (Chard et al., 2010).

In the last years, some authors have proposed different techniques to overcome these issues in MS patients by filling WM lesions with intensities similar to those of WM before performing tissue segmentation and image registration. These methods can be divided into two groups: methods which use *local* intensities from the surrounding neighboring voxels of lesions (Sdika and Pelletier, 2009; Battaglini et al., 2012; Magon et al., 2013) and methods which use *global* WM intensities from the whole brain (Chard et al., 2010). In all cases, the performance of these methods is directly related with their ability to minimize the impact of refilled voxels on original tissue distribution, not only due to the addition of these voxels into the tissue distribution, but also due to the effect on the global tissue distributions of filled images.

Within *local* methods, Sdika and Pelletier (2009) have proposed to refill each WM lesion voxel with the mean of its three-dimensional neighboring normal appearance white matter (NAWM) voxels. Battaglini et al. (2012) have suggested refilling each WM lesion voxel with intensities derived from a histogram of NAWM voxels surrounding the two-dimensional lesions. In a recent study, Magon et al. (2013) have proposed to refill each two-dimensional lesion with the intensity from the mean of the surrounding area of the lesion. Regarding *global*

\* Corresponding author.

E-mail address: [svalverde@eia.udg.edu](mailto:svalverde@eia.udg.edu) (S. Valverde).

methods, Chard et al. (2010) have proposed a different approach by using intensities re-sampled from a global WM distribution to refill WM lesion voxels, based on the mean and standard deviation of the total NAWM of the whole image. Both Chard et al. (2010) and Battaglini et al. (2012) methods are available for the community. FSL-L (Battaglini et al., 2012) runs from a computer command-line and does not provide any graphical interface that aids the process. This technique has been integrated into the latest FSL package, and therefore it depends on the whole FSL installation. In the case of LEAP (Chard et al., 2010), the method runs as a stand-alone script also from the command-line and requires the installation and configuration of several external dependencies, which may be difficult to install for non-computer experts.

In this paper we propose a new technique to refill WM lesions which is a compromise between *global* and *local* methods. Hence, for each slice composing the three-dimensional image, we compute the mean and standard deviation of the signal intensity of NAWM tissue. On the one hand, compared to local methods (Battaglini et al., 2012; Magon et al., 2013) which only make use of a limited range of voxel intensities, the fact of using global information from the whole image slice reduces the bias caused by refilled voxels on GM and WM tissue distributions, especially on images with high lesion load. On the other hand, compared to other global methods (Chard et al., 2010), which are based on the mean signal intensity of the NAWM of the three-dimensional image, our method re-computes the mean signal intensity of the NAWM at each two-dimensional slice with the aim of reproducing more precisely the signal variability between MRI slices, especially in low resolution images. In order to easily integrate it into current platforms, the proposed method called SLF is currently available as a stand-alone program and as SPM<sup>1</sup> extension at the SALEM group site (<http://atc.udg.edu/salem/slftoolbox>).

To evaluate the performance of our method, we estimate the deviation in GM and WM tissue volume between a set of healthy images and the same images where artificial WM lesions have been refilled with the proposed technique. To do so, we register WM lesion masks from diagnosed MS patients into two sets of 30 1.5 and 3 T T1-weighted (T1-w) images of healthy subjects, respectively. Afterwards, we simulate realistic lesions on healthy images by replacing the signal intensities of registered lesion voxels with values similar to those of the mean GM/WM interface. Brain tissue volume is computed using both FAST (Zhang et al., 2001) and SPM8 (Ashburner and Friston, 2005) segmentation methods, in order to avoid possible correlations between the filling and segmentation processes. Furthermore, we compare our results with the same images where artificial WM lesions have been segmented as normal tissue, masked-out before tissue segmentation, and refilled using also the methods proposed by Chard et al. (2010); Battaglini et al. (2012), and Magon et al. (2013).

## 2. Materials and methods

### 2.1. Image data

The first set of images is composed of 30 images of healthy subjects (matrix size: 176 × 208 × 176, voxel size: 1 × 1 × 1.25 mm), acquired on a 1.5 T Vision scanner (Siemens, Erlangen, Germany) and obtained from the Open Access Series of Imaging Studies (OASIS) repository<sup>2</sup> (Marcus et al., 2007). Only images from young and middle-aged subjects are selected (age < 50) as they have not been diagnosed with any related pathology. Image references included in the study are as follows: 2, 4, 5, 6, 7, 9, 11, 12, 14, 17, 18, 20, 25, 26, 27, 29, 34, 37, 38, 40, 43, 44, 45, 47, 49, 50, 51, 54, 55, and 57.

The second set of images is composed of 30 images of healthy subjects (matrix size: 256 × 150 × 256, voxel size: 0.92 × 0.92 × 1.20 mm) acquired on a Philips 3 T scanner (Philips Healthcare, Best, NL) and

obtained from the Information eXtraction from Images (IXI) repository maintained by the Imperial College London in London, UK.<sup>3</sup> We selected 30 images acquired from the Hammersmith Hospital. Image references included in the study are as follows: 12, 13, 14, 15, 33, 34, 39, 48, 49, 51, 52, 57, 59, 72, 80, 83, 92, 95, 96, 97, 104, 105, 126, 127, 128, 131, 136, 137, 146, and 159.

### 2.2. Preprocessing

All images are manually reoriented to match the standard MNI space. Skull-stripping is performed using the Brain Extraction Tool (BET) (Smith, 2002), following the optimization workflow suggested by Popescu et al. (2012), with the exception that cerebrospinal fluid tissue has been refilled on skull-stripped images again. This procedure is preferred over other alternatives as it provides the best performance on some lesion-filling methods such as Chard et al. (2010), being also the choice in other recent studies (Popescu et al., 2014). IXI images are corrected from possible intensity non-uniformities and acquisition artifacts using N4, the ITK (Ibáñez et al., 2003) implementation of the N3 package (Sled et al., 1997). N4 is applied on IXI images with default options. Images from the OASIS repository are provided already with N4 applied.

### 2.3. Lesion generation

We use a set of 37 patients with clinically confirmed MS, provided with *initial* and *follow-up* studies (Diez et al., 2014). In these patients, lesions have been annotated semi-automatically on Proton Density-weighted (PD-w) images by a trained technician using JIM software<sup>4</sup> and afterwards co-registered with T1-w images. In order to maintain the independence between the 1.5 and 3 T sets of images, we match randomly 30 patients from the *initial* study into the OASIS images, and we repeat the same procedure with the follow-up study and the IXI image set.

MS lesion masks are registered into healthy images by a non-rigid transformation (Rueckert et al., 1999). To ensure that resulting lesion masks are placed on WM, we remove registered lesion voxels that have not been segmented as WM by both FAST and SPM8 on the healthy image. We computed a Wilcoxon rank sum test to analyze the difference in lesion volumes generated between OASIS and IXI datasets, obtaining that differences were not statistically significant ( $p = 0.162$ ). The obtained mean lesion volume on OASIS images was  $21.1 \pm 20.8$  ml (range from 0.5 to 65 ml), while  $15.4 \pm 16.2$  ml (range from 0.8 to 62 ml) on IXI 3 T images. Note that due to the existing anatomical differences between 1.5 and 3 T image subjects and the enforced WM tissue constraint, the effect of registering the same MS lesion mask into a 1.5 and 3 T image results in two different lesion masks. For instance, the effect of registering lesions from the initial study into the 3 T dataset provided different lesion volumes ( $10.30 \pm 12.10$  ml) and reported statistically significant differences ( $p = 0.007$ ) on the Wilcoxon rank sum tests.

Artificial lesions are simulated by replacing registered lesion voxel intensities with ones between the GM and WM interface, following the same strategy shown in Battaglini et al. (2012). For each original image, GM and WM tissue distributions are computed using only voxels in agreement between FAST and SPM8. WM lesion voxels are filled with random intensities coming from a newly generated normal distribution, with mean equal to the average of the GM and WM mean values and standard deviation equal to the difference between mean WM and GM, divided by 4 (Battaglini et al., 2012). Artificial lesions are refilled with the aim of simulating a profile which clearly separates their signal intensity with healthy tissue. This intensity profile chosen does not

<sup>1</sup> <http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>.

<sup>2</sup> Publicly available at: <http://www.oasis-brain.org>.

<sup>3</sup> Publicly available at <http://biomedic.doc.ic.ac.uk/brain-development/index.php?n>Main.Datasets>.

<sup>4</sup> Xinapse Systems, JIM software webpage, <http://www.xinapse.com/home.php>.

reflect the entire scope of possible real lesions, but allows us to visualize the magnitude of the differences in tissue volume between images with artificial lesions and the same images where lesion have been filled with the proposed method. The intensity profile chosen would not affect any of the methods studied since they do not take into account the artificial lesion intensities.

#### 2.4. Lesion filling

The proposed method aims to combine the *global* approach of Chard et al. (2010) with the similarity between refilled voxel intensities and their surrounding voxels of *local* methods such as Battaglini et al. (2012) and Magon et al. (2013). Basically, for each slice composing the three-dimensional image, lesion voxel intensities are replaced by random intensities of a normal distribution generated from the mean NAWM intensity of the current slice. Fig. 1 summarizes the lesion-filling process graphically.

The proposed algorithm requires two input images: a preprocessed T1-w image (skull-stripped and intensity inhomogeneity corrected) and its corresponding binary WM lesion mask. After testing the performance of the method with different skull-stripping approaches (Smith, 2002; Shattuck et al., 2001), we observed that including this step inside the filling process is not necessary, because the skull-stripping method employed seems to not interfere significantly in the results obtained (Wilcoxon significant rank-sum tests between differences in tissue volume between lesion-filled and original images of both datasets for GM and WM tissue,  $p > 0.13$ ).

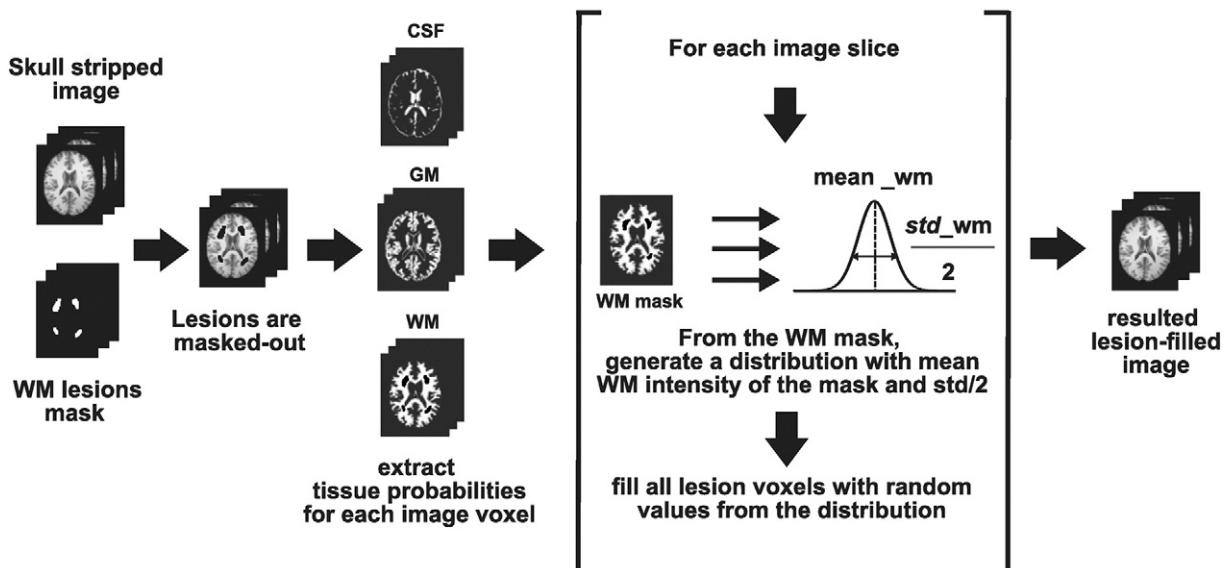
WM lesions are masked out from the T1-w image using the provided lesion mask, in order to avoid the influence of artificial lesions on tissue distributions. The resulting image is used to estimate the probability of each voxel to be classified as CSF, GM, and NAWM, by segmenting tissue with a Fuzzy-C-means approach (Pham, 2001). The Fuzzy-C-means implementation used here follows the algorithm described in Pham (2001), with clusters initialized according to Bezdek et al. (1999). Moreover, input signal intensities are constrained to the mean plus three standard deviations of the signal intensity of the image, in order to avoid outlier signal intensities, such as residual parts of the eyes or neck. From the obtained tissue segmentation output, we compute the

three-dimensional NAWM mask from the image voxels with the highest probability to pertain to the WM cluster.

Finally, the lesion-filling process is achieved as follows: for each axial slice composing the three-dimensional image, we compute the mean and standard deviation of the signal intensity of NAWM tissue. Axial sampling is motivated because after testing the sampling procedure on the coronal, axial and sagittal planes, we found that the best results were obtained when we sampled the axial plane. This was due to the fact that using the axial plane reduced the variability of possible existing WM intensities, when compared to coronal and sagittal sampling. The Fuzzy-C-means approach used to estimate the tissue probabilities is a simple method which in fact does not take into account neither spatial nor neighboring information, and hyper-intense signal intensities such as residual parts of the eyes or the neck produced in the skull-stripping process can bias significantly the clusters. The risk of adding these parts into the WM distribution is minimized in the axial plane because we are reducing it to a certain slice where lesion volume is usually lower than that in central slices. The computed mean and standard deviation values are used to generate a normal distribution with mean equal to the computed NAWM mean intensity and standard deviation equal to half of the computed NAWM standard deviation. Standard deviation is always fixed to half of the WM mean independently of the dataset used. This value was chosen empirically with the aim of balancing the accuracy of the method with both 1.5 and 3 T images. Although a specific tuning of this parameter could provide a better performance on certain cases, we decided to fix it avoiding therefore the number of parameters to tune. Lesion voxel intensities from the current image slice are replaced by random values of the generated distribution. The procedure is repeated until all image slices are completed.

#### 2.5. Volume analysis

We compute the absolute percentage % difference in normalized gray matter volume (NGMV) and normalized white matter volume (NWMV) between each original and its correspondent lesion-filled images. Normalized volumes are obtained as the ratio of voxels outside lesion regions segmented as GM or WM and the total number of



**Fig. 1.** Proposed algorithm for filling WM lesions. From a preprocessed T1-w image (skull-stripping and intensity inhomogeneity corrected), WM lesions are masked out using the provided WM lesion binary mask. Using a Fuzzy-C-means approach, we estimate the probability of each image voxel to be classified as CSF, GM, and NAWM. For each slice composing the whole image, lesion voxel intensities are replaced by a random intensity derived from a normal distribution with mean and half of the standard deviation of the NAWM tissue intensities of the current slice.

segmented voxels, respectively. For instance, the % difference in NGMV is computed as:

$$\% = \frac{|NGMV_{filled} - NGMV_{orig}|}{NGMV_{orig}} \times 100$$

where  $NGMV_{filled}$  and  $NGMV_{orig}$  values refer to the computed volumes for the lesion-filled and original images, respectively. The higher the performance of the lesion-filling method, the lower the percentage difference between lesion-filled and original images.

In order to analyze possible correlations between the filling process and the segmentation method employed, brain tissue volume is calculated independently on the same subjects using FAST (Zhang et al., 2001) (v.5.0.5) and SPM8 (Ashburner and Friston, 2005) (v.4667) approaches.

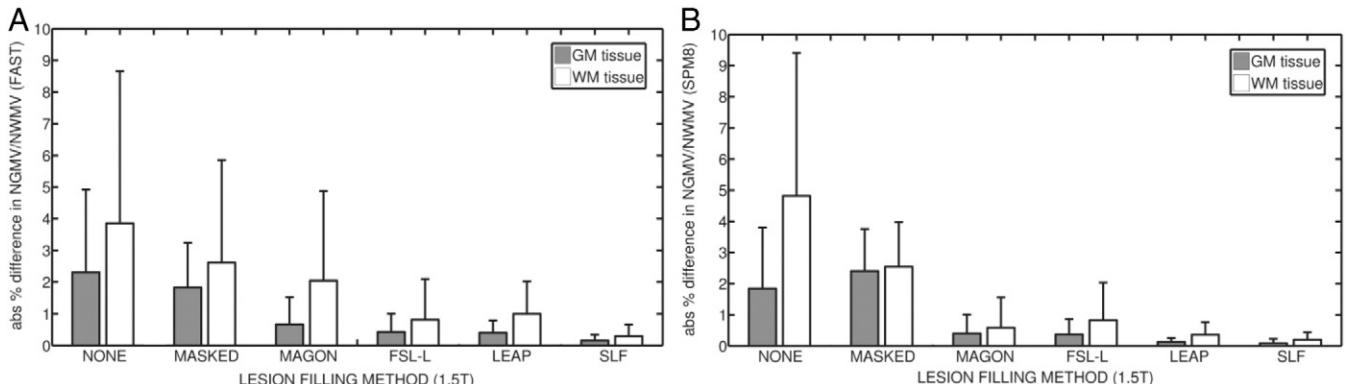
## 2.6. Statistical analysis

We compare the performance of our method with respect to other existing techniques such as the ones proposed by Chard et al. (2010); Battaglini et al. (2012), and Magon et al. (2013). We also add two more sets of images into the comparison: images segmented with artificial lesions and images where WM lesions have been masked out before tissue segmentation. Given the small differences in NGMV and NWMV between original and lesion-filled images, the use of a standard Analysis of the Variance (ANOVA) or a classic *t*-test is impractical here. Instead, we perform a series of permutation tests to determine significant differences in tissue volume between pairs of methods (Menke and Martínez, 2004; Valverde et al., 2014). The permutation tests return the mean  $\mu$  and standard deviation  $\sigma$  of the fraction of times that the difference in NGMV and NWMV for a current lesion-filling method is smaller than the rest of methods with  $p$ -value  $\leq 0.05$ . Afterwards, methods are presented in 3 ranks determined by the mean and standard deviation of the best method and the distance with respect to the mean of the rest of methods (Valverde et al., 2014). In our experiments, we set the number of comparisons between each pair of methods to  $N = 1000$ .

## 3. Results

### 3.1. OASIS dataset (1.5 T data)

Fig. 2 depicts the absolute mean % difference in NGMV and NWMV between the 30 original 1.5 T images and the same images with artificial lesions (NONE), masked-out lesions before segmentation (MASKED), and lesion-filled using Magon et al. (2013) (MAGON), Battaglini et al. (2012) (FSL-L), and Chard et al. (2010) (LEAP), and finally our proposed algorithm SLF.



**Fig. 2.** Absolute % difference in NGMV and NWMV between original and filled images from the OASIS (1.5 T) dataset. (a) Results for images segmented using FAST. (b) Results for images segmented with SPM8. Gray bars represent the absolute mean % difference in NGMV, while white bars represent the absolute mean % difference in NWMV. Lines above each bar represent the standard deviation for each method and tissue.

When FAST is used, SLF reports the lowest absolute mean difference in NGMV ( $0.16 \pm 0.14$ ), followed by LEAP ( $0.40 \pm 0.30$ ) and FSL-L ( $0.43 \pm 0.58$ ) methods. Our proposal also provides the lowest difference in NWMV ( $0.29 \pm 0.36$ ), followed by FSL-L ( $0.81 \pm 1.28$ ). Maximum values in NGMV are found in NONE images, with differences up to  $2.30 \pm 2.62$  in NGMV and  $3.85 \pm 4.81$  in NWMV.

When SPM8 is used, SLF also reports the lowest differences in NGMV ( $0.09 \pm 0.14$ ), followed by LEAP method ( $0.12 \pm 0.13$ ). Our proposed method also performs better than the rest of the methods on NWMV ( $0.20 \pm 0.24$ ), followed by the LEAP method ( $0.36 \pm 0.40$ ). Again, the highest differences in NGMV ( $1.84 \pm 1.97$ ) and NWMV ( $4.82 \pm 4.58$ ) are found in NONE images. Table 1 shows the absolute mean difference in WM volume for all methods where lesion volume has been ranged by size intervals. Results are presented for both SPM8 and FAST segmentation methods.

Table 2 presents the performance of each filling-method after running all possible pair-wise permutation tests. With a significant  $p$ -value of  $\leq 0.05$ , all tests run on images segmented with FAST show the superiority of SLF over the other methods presented. On images segmented with SPM8, all tests show a clear superiority of SLF over the other methods on NWMV, while a similar performance of SLF and LEAP over the other methods on NGMV.

### 3.2. IXI dataset (3 T data)

We also test the performance of our algorithm using 3 T data. As before, Fig. 3 shows the absolute mean % difference in NGMV and NWMV between the 30 original 3 T images and the same images with added lesions (NONE), masked-out lesions before segmentation (MASKED), and lesion-filled methods MAGON, FSL-L, and LEAP, and our proposed approach SLF.

When FAST is used, SLF reports the lowest absolute mean % difference in NGMV ( $0.06 \pm 0.06$ ), followed by LEAP ( $0.09 \pm 0.10$ ). Our method SLF also performs the lowest difference in NWMV ( $0.09 \pm 0.09$ ), followed again by LEAP ( $0.12 \pm 0.08$ ). Maximum values in NGMV are found in NONE images, with differences up to  $1.40 \pm 1.56$  in NGMV and  $1.00 \pm 1.32$  in NWMV.

When SPM8 is used, both LEAP ( $0.04 \pm 0.06$ ) and SLF ( $0.05 \pm 0.05$ ) yield the lowest absolute % mean difference in NGMV. On NWMV, also LEAP ( $0.09 \pm 0.12$ ) and SLF ( $0.08 \pm 0.09$ ) report the lowest absolute mean % difference in volume between original and lesion-filled images. Again, highest differences in NGMV ( $1.84 \pm 1.97$ ) and NWMV ( $4.82 \pm 4.58$ ) are found in NONE images. Table 3 shows the absolute mean difference in WM volume for all methods on IXI images, where lesion volume has been ranged by size intervals. Results are presented for both SPM8 and FAST segmentation methods.

Table 4 shows the performance of each filling-method after running the permutation tests. Tests run on images segmented with FAST show

**Table 1**

Absolute mean difference in NWMV between original and filled images from the 1.5 T OASIS images. Results are presented for both SPM8 and FAST segmentation methods. Lesion volume is ranged by size intervals with  $n = 6$  by interval. Values indicate the mean and standard deviation of the absolute difference in volume ( $\mu \pm \sigma$ ) of each lesion-filling method at a current lesion interval.

Method/ lesion(ml)	0.5–4 ml (n = 6)	4–11 ml (n = 6)	11–20 ml (n = 6)	25–36 ml (n = 6)	>36 ml (n = 6)
<i>SPM8 segmentation method</i>					
NONE	0.47 ± 0.50	1.54 ± 0.95	2.71 ± 0.60	7.09 ± 1.42	10.64 ± 3.10
MASKED	1.56 ± 0.94	2.42 ± 0.70	1.49 ± 0.43	3.16 ± 1.35	3.91 ± 1.76
MAGON	0.03 ± 0.03	0.08 ± 0.07	0.24 ± 0.25	0.32 ± 0.19	1.95 ± 1.25
FSL-L	0.03 ± 0.01	0.10 ± 0.05	0.31 ± 0.15	0.55 ± 0.07	2.38 ± 1.26
LEAP	0.04 ± 0.04	0.10 ± 0.05	0.19 ± 0.05	0.44 ± 0.22	0.92 ± 0.42
SLF	0.03 ± 0.03	0.04 ± 0.03	0.09 ± 0.06	0.23 ± 0.20	0.55 ± 0.23
<i>FAST segmentation method</i>					
NONE	0.21 ± 0.21	0.71 ± 0.38	1.88 ± 0.56	4.55 ± 2.04	8.95 ± 4.36
MASKED	9.52 ± 1.20	8.36 ± 1.30	11.53 ± 4.91	7.42 ± 1.08	5.79 ± 1.92
MAGON	0.08 ± 0.04	0.25 ± 0.22	0.91 ± 0.63	1.28 ± 0.39	6.24 ± 2.74
FSL-L	0.03 ± 0.02	0.05 ± 0.05	0.30 ± 0.21	0.58 ± 0.19	2.13 ± 1.22
LEAP	0.08 ± 0.07	0.34 ± 0.10	0.65 ± 0.13	1.07 ± 0.66	2.50 ± 0.80
SLF	0.07 ± 0.05	0.13 ± 0.09	0.22 ± 0.15	0.36 ± 0.30	0.42 ± 0.16

a significant superiority of SLF over the rest of the methods on NWMV, and a slightly better performance of SLF with respect to LEAP on NGMV, although both methods are clearly superior to the rest of methods presented. When SPM8 is used, tests show a similar performance of SLF and LEAP over the rest of the methods on both NWMV and NGMV.

#### 4. Discussion

Several studies have proposed to use different filling techniques in order to reduce the effects of WM lesions on brain tissue measurements of T1-w images. Up to date, only LEAP (Chard et al., 2010)<sup>5</sup> and FSL-L (Battaglini et al., 2012)<sup>6</sup> are publicly available methods that permit to refill T1-w images given a WM lesion mask. The Lesion Segmentation Toolbox (LST) proposed by Schmidt et al. (2012) also provides a lesion-filling approach based on the work of Chard et al. (2010), but it is dependent of a FLAIR image and an internal lesion-probability map obtained during the lesion segmentation step.

In general, deviation in tissue volume between original and lesion-filled images tends to be higher on 1.5 T OASIS images than on 3 T IXI images. The observed deviation is caused by differences in intensity, slice thickness and dimensionality between datasets. On IXI images, the distance between GM and WM signal intensity distributions is narrower than that of 1.5 T data. Applying the lesion generation algorithm (Battaglini et al., 2012) with identical parameters of those used with 1.5 T images creates simulated lesions whose intensity are noticeably similar to the mean WM, because the standard deviation of the generated lesion distribution is the mean between the GM and WM tissue divided by 4. However, this fact only explains the difference found on images segmented with artificial lesions. In the rest of the methods, the signal intensity of the generated lesions is not interfering with the obtained results since in all cases lesion voxels are replaced before tissue segmentation. On images where lesions have been masked before segmentation (MASKED), the lower deviation in tissue volume of 3 T images can be explained by the increase in the resolution of the images when compared to 1.5 T data, which reduces the effect of masked voxels in tissue distributions. The same reason can be behind the lower deviation found on all four lesion filling methods. By increasing the number of slices, differences produced by the methods on certain slices can be smoothed by tissue segmentation methods. Moreover, the use of a reduced sampling space or a better tuning of the parameters involved in

the WM tissue distribution generated to refill lesion voxels could increase the performance of the presented method. Nevertheless, in all our experiments we decided to fix the standard deviation to 2 for simplicity.

Analyzing the results by dataset, on 1.5 T images from the OASIS dataset, our results show that compared to the available methods, the proposed algorithm SLF reduces significantly the differences in NWMV between original and filled images, independently of the brain tissue segmentation method used to measure the tissue volume. With the same data, SLF also reduces significantly the differences in NGMV when FAST is used. Although our method reports the lowest mean % difference in NGMV when SPM8 method is used, the permutation test clearly shows that differences between SLF and LEAP are not relevant. On 3 T images from the IXI dataset, SLF also yields the lowest mean % differences in NGMV and NWMV, when FAST is used to measure tissue volume. These results are clearly significant in NWMV, but not in NGMV, although our method reports also the lowest difference among all methods. When SPM8 is used, SLF presents a similar performance of that of LEAP, and both methods tie on the results of the significance tests.

Compared with local methods, our algorithm performs quantitatively better than *local* methods on images with high lesion load ( $> 36$  ml). The MAGON method incorporates all neighbor voxels surrounding a WM lesion region to compute a mean intensity which is used to refill all lesion voxels. On images with high lesion load touching GM tissue, including GM voxels can decrease refilled intensities and modify the tissue distribution of filled images. FSL-L overpasses this limitation by building an intensity distribution based only on WM voxels surrounding lesions. However, on large lesion regions, all lesion voxels will be filled with a narrow range of intensities coming from the neighboring voxels that can have a direct incidence on GM and WM tissue distributions. By contrast, lesion volume appears to affect less *global* methods. In our case, the intensity distribution generated to refill lesion voxels will be independent of both the size and the position of lesion. Furthermore, the effect of filled voxels on the global WM tissue distribution is smoothed by the addition of intensities which try to reassemble the global NAWM of the current slice.

Compared with global methods, there are some interesting differences between our method and LEAP. Contrary to *local* methods, *global* methods have to deal with the skull-stripping process before processing images. LEAP incorporates the skull-stripping process as part of the processing pipeline. In addition, LEAP also allows the user to provide a brain

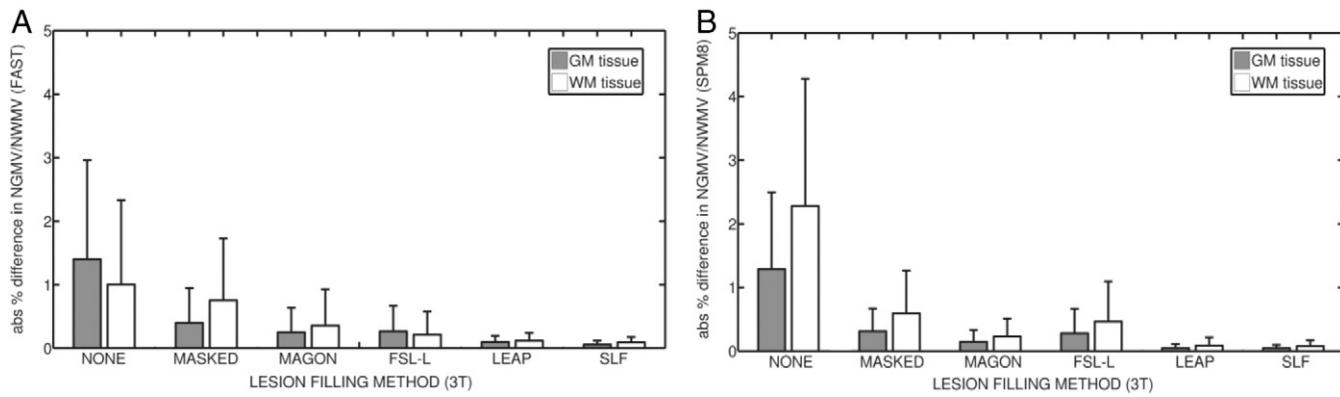
**Table 2**

Permutation tests for obtained absolute % differences in NGMV and NWMV on 1.5 T images. Reported values are mean and standard deviation ( $\mu_0, \sigma_0$ ) of the fraction of times when each method produces significant  $p$ -values ( $p \leq 0.05$ ). (a) Results when using FAST. (b) Results when using SPM8. Positive values indicate that in average, the method outperforms the other methods in pair-wise significant tests. Negative values indicate the contrary. Rank 1:  $(\mu_0 - \sigma_0, \mu_0]$ , rank 2:  $(\mu_0 - 2\sigma_0, \mu_0 - \sigma_0]$ , rank 3:  $(\mu_0 - 3\sigma_0, \mu_0 - 2\sigma_0]$ .

		NGMV		NWMV	
Method	$\mu \pm \sigma$	Method	$\mu \pm \sigma$	Method	$\mu \pm \sigma$
<i>(a) FAST segmentation method (1.5 T)</i>					
Rank 1	SLF	0.83 ± 0.41	SLF	0.83 ± 0.41	
Rank 2	FSL-L	0.33 ± 0.82	FSL-L	0.33 ± 0.82	
	LEAP	0.33 ± 0.82	LEAP	0.33 ± 0.82	
Rank 3	MAGON	-0.17 ± 0.98	MAGON	-0.17 ± 0.98	
	MASKED	-0.23 ± 0.41	MASKED	-0.23 ± 0.41	
	NONE	-0.50 ± 0.84	NONE	-0.50 ± 0.84	
<i>(b) SPM8 segmentation method (1.5 T)</i>					
Rank 1	SLF	0.67 ± 0.52	SLF	0.83 ± 0.41	
	LEAP	0.67 ± 0.52			
Rank 2	MAGON	0.00 ± 0.89	LEAP	0.33 ± 0.82	
	FSL-L	0.00 ± 0.89	MAGON	0.17 ± 0.75	
Rank 3	NONE	-0.67 ± 0.52	FSL-L	0.00 ± 0.89	
	MASKED	-0.67 ± 0.52	MASKED	-0.50 ± 0.84	
			NONE	-0.83 ± 0.41	

<sup>5</sup> <http://www.nmrgroup.ion.ucl.ac.uk/analysis/lesionfill.html>.

<sup>6</sup> [http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/lesion\\_filling](http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/lesion_filling).



**Fig. 3.** Absolute mean % difference in NGMV and NWMV between original and filled images from the IXI (3 T) dataset. (a) Results for images segmented using FAST. (b) Results for images segmented with SPM8. Gray bars represent the absolute mean % difference in NGMV, while white bars represent the absolute mean % difference in NWMV. Lines above each bar represent the standard deviation for each method and tissue.

mask. By contrast, our method does not deal with skull-stripping internally, and the method requires an already skull-stripped image or a brain mask. As noted previously, the skull-stripping method employed seems to not interfere significantly in the results obtained by our method. While setting up each of the different processes involved in the proposed pipeline, we found that, at least with our data, the performance of LEAP decreased or failed in 1.5 T scans when the skull-stripping methods BSE (Shattuck et al., 2001) and BET (Smith, 2002) were used with default options. By contrast, LEAP provided the best results when the optimized method proposed by Popescu et al. (2012) was used. This fact motivated the selection of this skull-stripping method for all the experiments of the study.

Furthermore, on both datasets, we have also compared the differences between our method and LEAP estimating the mean NAWM intensity used as a basis to fill lesion voxels. In most of the images, the global mean NAWM intensity does not differ significantly between fity<sup>7</sup> on LEAP and our Fuzzy-C-means approach. Hence, we can reject the hypothesis that observed differences in 1.5 T images can be caused by the approach employed to compute the NAWM tissue distribution before filling lesion voxels. However, on both lesion-filling approaches, tissue segmentation methods tended to increase the apparent mean WM tissue distribution on 1.5 T images with high lesion load (>40 ml) due to the increase of voxels refilled with intensities higher than the actual mean WM signal intensity. This effect is clearly more visible on LEAP than in our method, especially when FAST is used. The resolution of the OASIS 1.5 T images ( $176 \times 208 \times 176$  slices) is lower than that of IXI 3 T images ( $256 \times 150 \times 256$  slices). On images with low number of slices, each slice has a higher weight into the global tissue distribution. After comparing the a priori WM tissue distribution values estimated by both the LEAP and SLF methods with the already computed WM tissue distributions obtained from healthy images, we found that as lesion size increases, global methods such as LEAP and SLF tend to increase the differences in tissue volume with respect to original images. In both methods, we have observed that the a priori estimated mean intensity of the WM distribution tends to be higher than the actual tissue distribution as computed by FAST and SPM8 on healthy images. As lesion volume increases, the addition of more filled voxels with intensity higher than the actual mean tissue intensity is more prominent, causing a displacement of the mean intensity of the WM distribution returned by the segmentation methods on filled images. Consequently, more voxels bordering GM/WM are segmented as GM and WM tissue volume decreases. In this scenario, the strategy followed by SLF, where WM is sampled independently at each slice, is more robust to the increase of lesions size than a global estimation of the WM tissue (LEAP) because

possible errors introduced by a particular slice are not propagated into the rest of the slices. Contrary to SPM8, which estimates the tissue distributions based on a Gaussian Mixture Model approach of the whole image, FAST builds a network of neighboring relations based on a Markov random field approach, more sensible to changes between slices. The same reason can also be behind the better performance of our method on 3 T when FAST is used. Compared with 1.5 T images, the probability of intensity change between slices is less prominent on 3 T images due to a higher resolution between slices.

Analyzing the possible deviations in tissue volume caused by each tissue segmentation process, we obtained results which suggest that the chosen tissue segmentation method does not affect significantly the performance of our filling-method. Results between the same filled images segmented with FAST and SPM8 differ (<0.1%) in the worst case on both datasets and tissues. By contrast, MAGON, FSL-L and LEAP switch their rank on 1.5 T images, depending on the segmentation method used. On 3 T images, only MAGON and FSL-L appear to switch between ranks when FAST or SPM8 is used, respectively.

The present study is not free from limitations. The most important one is the lack of images of MS patients with brain tissue expert annotations. All images from MS patients taken from Diez et al. (2014) have been only provided with lesion annotations delineated by a trained expert, but not brain tissue annotations. To overpass this limitation, we have registered WM lesions from MS patients into healthy images as performed in Battaglini et al. (2012) and double-checked that registered

**Table 3**

Absolute mean difference in NWMV between original and filled images from the 3 T OASIS images. Results are presented for both SPM8 and FAST segmentation methods. Lesion volume is ranged by size intervals with  $n = 6$  by interval. Values indicate the mean and standard deviation of the absolute difference in volume ( $\mu \pm \sigma$ ) of each lesion-filling method at a current lesion interval.

Method/ lesion(ml)	0.8–3 ml (n = 6)	4–6 ml (n = 6)	6–13 ml (n = 6)	16–21 ml (n = 6)	>21 ml (n = 6)
<i>SPM8 segmentation method</i>					
NONE	$0.68 \pm 0.56$	$0.92 \pm 0.31$	$1.61 \pm 0.85$	$3.37 \pm 0.81$	$5.16 \pm 1.83$
MASKED	$0.07 \pm 0.03$	$0.21 \pm 0.16$	$0.34 \pm 0.22$	$1.07 \pm 0.79$	$1.42 \pm 0.65$
MAGON	$0.05 \pm 0.10$	$0.15 \pm 0.28$	$0.14 \pm 0.15$	$0.47 \pm 0.44$	$0.41 \pm 0.22$
FSL-L	$0.06 \pm 0.06$	$0.06 \pm 0.03$	$0.19 \pm 0.16$	$0.80 \pm 0.80$	$1.32 \pm 0.53$
LEAP	$0.01 \pm 0.01$	$0.03 \pm 0.02$	$0.05 \pm 0.05$	$0.13 \pm 0.15$	$0.22 \pm 0.18$
SLF	$0.03 \pm 0.03$	$0.02 \pm 0.01$	$0.09 \pm 0.12$	$0.09 \pm 0.06$	$0.16 \pm 0.13$
<i>FAST segmentation method</i>					
NONE	$0.14 \pm 0.10$	$0.24 \pm 0.06$	$0.52 \pm 0.34$	$1.27 \pm 0.35$	$2.94 \pm 1.67$
MASKED	$0.07 \pm 0.05$	$0.17 \pm 0.07$	$0.41 \pm 0.27$	$0.95 \pm 0.25$	$2.23 \pm 1.13$
MAGON	$0.05 \pm 0.03$	$0.07 \pm 0.06$	$0.08 \pm 0.04$	$0.59 \pm 0.55$	$1.07 \pm 0.79$
FSL-L	$0.04 \pm 0.02$	$0.03 \pm 0.02$	$0.03 \pm 0.02$	$0.18 \pm 0.20$	$0.77 \pm 0.45$
LEAP	$0.07 \pm 0.05$	$0.03 \pm 0.03$	$0.14 \pm 0.13$	$0.19 \pm 0.16$	$0.29 \pm 0.13$
SLF	$0.03 \pm 0.02$	$0.04 \pm 0.02$	$0.08 \pm 0.06$	$0.20 \pm 0.15$	$0.34 \pm 0.14$

<sup>7</sup> Available at: <http://sourceforge.net/projects/fityk>.

**Table 4**

Permutation tests for obtained absolute % differences in NGMV and NWMV on 3 T images. Reported values are mean and standard deviation ( $\mu_0$ ,  $\sigma_0$ ) of the fraction of times when each method produces significant  $p$ -values ( $p \leq 0.05$ ). (a) Results when using FAST. (b) Results when using SPM8. Positive values indicate that in average, the method outperforms the other methods in pair-wise significant tests. Negative values indicate the contrary. Rank 1: ( $\mu_0 - \sigma_0, \mu_0$ ), rank 2: ( $\mu_0 - 2\sigma_0, \mu_0 - \sigma_0$ ), rank 3: ( $\mu_0 - 3\sigma_0, \mu_0 - 2\sigma_0$ ).

NGMV		NWMV	
Method	$\mu \pm \sigma$	Method	$\mu \pm \sigma$
<i>(a) FAST segmentation method (3 T)</i>			
Rank 1	SLF	0.67 ± 0.52	SLF
	LEAP	0.66 ± 0.51	0.67 ± 0.52
Rank 2	MAGON	0.00 ± 0.8	MAGON
	FSL-L	0.00 ± 0.3	−0.17 ± 0.98
Rank 3	MASKED	−0.50 ± 0.84	MASKED
	NONE	−0.83 ± 0.41	−0.83 ± 0.41
<i>(b) SPM8 segmentation method (3 T)</i>			
Rank 1	LEAP	0.67 ± 0.52	LEAP
	SLF	0.67 ± 0.52	0.67 ± 0.52
	MAGON	0.17 ± 0.98	MAGON
Rank 2	FSL-L	−0.33 ± 0.82	FSL-L
	MASKED	−0.33 ± 0.82	−0.17 ± 0.98
Rank 3	NONE	−0.83 ± 0.41	MASKED
			−0.50 ± 0.84
			NONE
			−0.83 ± 0.41

lesions have replaced voxels segmented as WM by FAST and SPM8. This strategy has a negligible impact on the performance of the filling-methods analyzed in this study, because we assure a priori that generated lesions are on WM, and moreover none of the methods use information from the artificial lesions generated. Furthermore, although we tested the performance of the proposed method with two datasets with different magnetic field strengths, our results are limited to these two different scanners with particular configurations, and hence it is difficult to generalize the results to all 1.5 and 3 T scanners.

In conclusion, the results of this study show that regardless of the lesion size, the SLF method performs consistently well compared to other existing methods such as LEAP, especially on 1.5 T images. Furthermore, the results obtained show that the proposed method can be an effective method for low resolution images. The skull-stripping process does not especially affect the accuracy of the method, which allows integrating it with different preprocessing pipelines. Additionally, volume estimations of lesion filled images processed by our algorithm appear to be not affected by the segmentation method employed. In contrast to other approaches, SLF may be installed by non-computer experts who can easily use it without any parameter tuning. SLF is currently available to researchers as a stand-alone script and as an SPM library extension which facilitates to incorporate the lesion filling process into the expert workflow for tissue volume segmentation.

## Acknowledgments

The authors would like to thank Dr. Ferran Prados, PhD, for helpful discussion. Sergi Valverde holds a FI-GDR2013 grant from the Generalitat de Catalunya.

## References

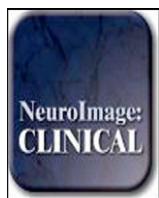
- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *NeuroImage* 26, 839–851. <http://dx.doi.org/10.1016/j.neuroimage.2005.02.01815955494>.
- Battaglini, M., Jenkinson, M., De Stefano, N., 2012. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Human Brain Mapping* 33, 2062–2071. <http://dx.doi.org/10.1002/hbm.2134421882300>.
- Bezdek, J.C., Keller, J., Krishnapuram, R., Pal, N.R., 1999. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing* The Handbooks of Fuzzy Sets Series. 4. Springer, New York.
- Chard, D.T., Jackson, J.S., Miller, D.H., Wheeler-Kingshott, C.A., 2010. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *Journal of Magnetic Resonance Imaging: JMRI* 32, 223–228. <http://dx.doi.org/10.1002/jmri.2221420575080>.
- Derkhshan, M., Caramanos, Z., Giacomini, P.S., Narayanan, S., Maranzano, J., Francis, S.J., Arnold, D.L., Collins, D.L., 2010. Evaluation of automated techniques for the quantification of grey matter atrophy in patients with multiple sclerosis. *NeuroImage* 52, 1261–1267. <http://dx.doi.org/10.1016/j.neuroimage.2010.05.02920483380>.
- Diez, Y., Oliver, A., Cabezas, M., Valverde, S., Martí, R., Vilanova, J.C., Ramió-Torrentà, L., Rovira, A., Lladó, X., 2014. Intensity based methods for brain MRI longitudinal registration. A study on multiple sclerosis patients. *Neuroinformatics* 12, 365–379. [http://dx.doi.org/10.1007/s12007\\_012021-013-9216-z](http://dx.doi.org/10.1007/s12007_012021-013-9216-z).
- Ganiler, O., Oliver, A., Diez, Y., Freixenet, J., Vilanova, J.C., Beltran, B., Ramió-Torrentà, L., Rovira, A., Lladó, X., 2014. A subtraction pipeline for automatic detection of new appearing multiple sclerosis lesions in longitudinal studies. *Neuroradiology* 56 (5), 363–374. <http://dx.doi.org/10.1007/s00234-014-1343-124590302>.
- Gelineau-Morel, R., Tomassini, V., Jenkinson, M., Johansen-Berg, H., Matthews, P.M., Palace, J., 2012. The effect of hypointense white matter lesions on automated gray matter segmentation in multiple sclerosis. *Human Brain Mapping* 33, 2802–2814. <http://dx.doi.org/10.1002/hbm.2140221976406>.
- Ibáñez, L., Schroeder, W., Ng, L., Cates, J., 2003. *The ITK Software Guide: The Insight Segmentation and Registration Toolkit* Kitware Inc.
- Kearney, H., Rocca, M.A., Valsasina, P., Balk, L., Sastre-Garriga, J., Reinhardt, J., Ruggieri, S., Rovira, A., Stippich, C., Kappos, L., Sprenger, T., Tortorella, P., Rovaris, M., Gasperini, C., Montalban, X., Geurts, J.J.G., Polman, C.H., Barkhof, F., Filippi, M., Altmann, D.R., Ciccarelli, O., Miller, D.H., Chard, D.T., 2014. Magnetic resonance imaging correlates of physical disability in relapse onset multiple sclerosis of long disease duration. *Multiple Sclerosis* (Hounds Mills, Basingstoke, England) 20 (1), 72–80. <http://dx.doi.org/10.1177/135245851349224523812283>.
- Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A., 2009. Evaluation of automated brain MR image segmentation and volumetry methods. *Human Brain Mapping* 30, 1310–1327. <http://dx.doi.org/10.1002/hbm.2059918537111>.
- Magon S., Chakravarty M., Lerch J., Gaetano L., Naegelin Y., Stippich C., Kappos L., Redue E.W., Sprenger T. (2013). Lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients. *Multiple Sclerosis*. Copenhagen, Denmark, vol. 19, pp 74–558
- Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience* 19, 1498–1507. <http://dx.doi.org/10.1162/jocn.2007.19.9.149817714011>.
- Menke, J., Martinez, T.R., 2004. Using permutations instead of Student's t distribution for p-values in paired-difference algorithm comparisons. *Proceedings of the IEEE Joint Conference on Neural Networks*.
- Pham, D.L., 2001. Robust fuzzy segmentation of magnetic resonance images. *Proceedings of the Fourteenth IEEE Symposium on Computer-Based Medical Systems CBMS 127–134*.
- Pérez-Miralles, F., Sastre-Garriga, J., Tintoré, M., Arrambide, G., Nos, C., Perkal, H., Río, J., Edo, M.C., Horga, A., Castilló, J., Auger, C., Huerga, E., Rovira, A., Montalban, X., 2013. Clinical impact of early brain atrophy in clinically isolated syndromes. *Multiple Sclerosis* (Hounds Mills, Basingstoke, England) 19 (14), 1878–1886. <http://dx.doi.org/10.1177/135245851348823123652215>.
- Popescu, V., Battaglini, M., Hoogstrate, W.S., Verfaillie, S.C., Sluimer, I.C., Van Schijndel, R.A., Van Dijk, B.W., Cover, K.S., Knol, D.L., Jenkinson, M., Barkhof, F., de Stefano, N., Vrenken, H., 2012. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. *NeuroImage* 61, 1484–1494. <http://dx.doi.org/10.1016/j.neuroimage.2012.03.07422484407>.
- Popescu, V., Ran, N.C.G., Barkhof, F., Chard, D., Wheeler-Kingshott, C., Vrenken, H., 2014. Accurate GM atrophy quantification in MS using lesion-filling with co-registered lesion masks. *NeuroImage*: Clinical 4, 366–373. <http://dx.doi.org/10.1016/j.nic.2014.01.004>.
- Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* 18, 712–721. <http://dx.doi.org/10.1109/42.79628410534053>.
- Sdika, M., Pelletier, D., 2009. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Human Brain Mapping* 30, 1060–1067. <http://dx.doi.org/10.1002/hbm.2056618412131>.
- Schmidt, P., Gaser, C., Arsic, M., Buck, D., Forschner, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V.J., Zimmer, K., Hemmer, B., Muhlau, B., 2012. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage* 59, 43774–43783.
- Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A., Leahy, R.M., 2001. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage* 13, 856–876. <http://dx.doi.org/10.1006/nimg.2000.073011304082>.
- Sled, J.G., Zijdenbos, A.P., Evans, C.P., 1997. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. I. *E.E.E. Transactions on Medical Imaging* 17, 87–97.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Human Brain Mapping* 17, 143–155. <http://dx.doi.org/10.1002/hbm.1006212391568>.
- Sormani, M.P., Arnold, D.L., De Stefano, N., 2014. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. *Annals of Neurology* 75, 43–49. <http://dx.doi.org/10.1002/ana.2401824006277>.
- Valverde, S., Oliver, A., Cabezas, M., Roura, E., Lladó, X., 2014. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. *Journal of Magnetic Resonance Imaging: JMRI* <http://dx.doi.org/10.1002/jmri.2451724459099>.
- Zhang, Y., Brady, M., Smith, S.A., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. *IEEE Transactions on Medical Imaging* 20, 45–57. <http://dx.doi.org/10.1109/42.9064211293691>.

# Chapter 5

## Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling.

In this chapter we present a detailed evaluation of the performance of different pipelines that incorporate fully automated processes such as lesion segmentation, lesion filling and tissue segmentation on MS data. For each automated pipeline, we analyze the percentage of error in tissue segmentation between a set of 70 MS images where WM lesions have been refilled before segmentation and the same images processed different levels of automation from manually masking lesion to fully automated lesion segmentation and filling. This analysis has been published in the following paper:

NeuroImage: Clinical (NICL)  
Volume: 9, Pages: 640-647, Published: October 2015  
DOI: doi:10.1016/j.nicl.2015.10.012  
Quality Index: 2.53 (Quartile 2)



## Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling



Sergi Valverde <sup>a,\*</sup>, Arnau Oliver <sup>a</sup>, Eloy Roura <sup>a</sup>, Deborah Pareto <sup>b</sup>, Joan C. Vilanova <sup>c</sup>, Lluís Ramió-Torrentà <sup>d</sup>, Jaume Sastre-Garriga <sup>e</sup>, Xavier Montalban <sup>e</sup>, Àlex Rovira <sup>b</sup>, Xavier Lladó <sup>a</sup>

<sup>a</sup>Dept. of Computer Architecture and Technology, University of Girona, Spain

<sup>b</sup>Magnetic Resonance Unit, Dept. of Radiology, Vall d'Hebron University Hospital, Spain

<sup>c</sup>Girona Magnetic Resonance Center, Spain

<sup>d</sup>Multiple Sclerosis and Neuro-immunology Unit, Dr. Josep Trueta University Hospital, Spain

<sup>e</sup>Neurology Unit, Multiple Sclerosis Centre of Catalonia (Cemcat), Vall d'Hebron University Hospital, Spain

### ARTICLE INFO

#### Article history:

Received 22 July 2015

Received in revised form 21 October 2015

Accepted 23 October 2015

Available online 28 October 2015

#### Keywords:

Brain

Multiple sclerosis

MRI

Brain atrophy

Automated tissue segmentation

White matter lesions

Lesion filling

### ABSTRACT

Lesion filling has been successfully applied to reduce the effect of hypo-intense T1-w Multiple Sclerosis (MS) lesions on automatic brain tissue segmentation. However, a study of fully automated pipelines incorporating lesion segmentation and lesion filling on tissue volume analysis has not yet been performed. Here, we analyzed the % of error introduced by automating the lesion segmentation and filling processes in the tissue segmentation of 70 clinically isolated syndrome patient images. First of all, images were processed using the LST and SLS toolkits with different pipeline combinations that differed in either automated or manual lesion segmentation, and lesion filling or masking out lesions. Then, images processed following each of the pipelines were segmented into gray matter (GM) and white matter (WM) using SPM8, and compared with the same images where expert lesion annotations were filled before segmentation. Our results showed that fully automated lesion segmentation and filling pipelines reduced significantly the % of error in GM and WM volume on images of MS patients, and performed similarly to the images where expert lesion annotations were masked before segmentation. In all the pipelines, the amount of misclassified lesion voxels was the main cause in the observed error in GM and WM volume. However, the % of error was significantly lower when automatically estimated lesions were filled and not masked before segmentation. These results are relevant and suggest that LST and SLS toolboxes allow the performance of accurate brain tissue volume measurements without any kind of manual intervention, which can be convenient not only in terms of time and economic costs, but also to avoid the inherent intra/inter variability between manual annotations.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Multiple sclerosis (MS) is associated with irreversible brain damage not only in demyelinated plaques, but also in normal-appearing gray matter (GM) and white matter (WM), where recent studies have shown that the rate of tissue loss per year in MS patients ranges from 0.7% to 1.6% in GM, and 0.6% to 0.9% in WM (Filippi et al., 2013; Pérez-Miralles et al., 2013; Sastre-Garriga et al., 2014). Given the correlation between brain atrophy and disease disability, measuring the change in tissue volume is clinically relevant because it allows for optimizing possible treatments and patient management in early stages of the disease (Filippi et al., 2013; Sastre-Garriga et al., 2014; Uher et al., 2014).

Automated tissue segmentation techniques based on magnetic resonance imaging (MRI) such as the Statistical Parametric Mapping (SPM) (Ashburner and Friston, 2005), FAST (Zhang et al., 2001), or SIENA-X (Smith et al., 2002) are currently standard tools to assess brain tissue volume (De Bresser et al., 2011; Valverde et al., 2015a). The reproducibility of these techniques has been analyzed in several studies using scan-rescan measurement tests, reporting mean percentages of error in FAST GM of –0.22% (De Boer et al., 2010), 0.05% (De Boer et al., 2010) and –0.80% (Nakamura et al., 2014) in SPM8 GM, 1.50% (Nakamura et al., 2014) in SIENA-X GM, 0.13% (De Boer et al., 2010) in FAST WM, and 0.25% (De Boer et al., 2010) in SPM WM. However, existing differences for a particular method in different studies may be influenced by the same image data, imaging hardware and acquisition parameters (Clark et al., 2006). Furthermore, several authors have reported that the inclusion of WM lesions in tissue segmentation can affect significantly the accuracy of these techniques (Battaglini et al., 2012; Nakamura and Fisher, 2009; Valverde et al., 2015b), leading to

\* Corresponding author at: Ed. P-IV, Campus Montilivi, University of Girona, 17071 Girona, Spain.

E-mail address: [svalverde@eia.udg.edu](mailto:svalverde@eia.udg.edu) (S. Valverde).

the development of different preprocessing strategies to fill lesion regions with signal intensities similar to WM before tissue segmentation (Battaglini et al., 2012; Chard et al., 2010; Valverde et al., 2014). So far, in all the lesion filling approaches, MS lesions have to be delineated first, usually by their manual annotation, which is a tedious, challenging and time-consuming task (Sanfilipo et al., 2005). This fact and the necessity to analyze quantitatively focal MS lesions in individual (Cabezas et al., 2014) and temporal (Ganiler et al., 2014) studies have been driving in recent years the development of automated new lesion segmentation techniques (García-Lorenzo et al., 2013; Guizard et al., 2015; Lladó et al., 2012).

Although lesion filling techniques have already been applied to assess the progression of GM atrophy of MS patients (Cecarelli et al., 2012; Nakamura et al., 2014; Popescu et al., 2014), still an extensive analysis of the effect of fully automated pipelines, incorporating both automated MS lesion segmentation and posterior lesion filling on tissue segmentation methods has not yet been performed. In this study, we analyze the effect of two publicly available automated pipelines, Salem Lesion Segmentation (SLS) (Roura et al., 2015) and Lesion Segmentation Toolbox (LST) (Schmidt et al., 2012), on the accuracy of the GM and WM volume estimations of a cohort of 70 clinically isolated syndrome (CIS) patients. For each automated pipeline, we evaluate the deviation in GM and WM volume between images where manual expert annotations have been used to refill lesions before tissue segmentation with SPM8 (Ashburner and Friston, 2005), and the same images where lesions have been automatically segmented and either masked or lesion filled before tissue segmentation.

## 2. Materials and methods

### 2.1. Image acquisition

Seventy CIS patients from the same center (Hospital Vall D'Hebron, Barcelona (Spain)) in which the clinical presentation was clearly suggestive of multiple sclerosis underwent MR imaging on the same 3 T Siemens with 12-channel phased-array head coil (Trio Tim, Siemens, Germany). The following pulse sequences were obtained: 1) transverse proton density and T2-weighted fast spin-echo (TR = 2500 ms, TE = 16–91 ms, voxel size = 0.78 × 0.78 × 3 mm<sup>3</sup>); 2) transverse fast T2-FLAIR (TR = 9000 ms, TE = 93 ms, TI = 2500 ms, flip angle = 120°, voxel size = 0.49 × 0.49 × 3 mm<sup>3</sup>); and 3) sagittal 3D T1 magnetization prepared rapid gradient-echo (MPRAGE) (TR = 2300 ms, TE = 2 ms; flip angle = 9°; voxel size = 1 × 1 × 1.2 mm<sup>3</sup>). White matter lesion masks were semi-automatically delineated from either PD-w (46 patients) or FLAIR (24 patients) images using JIM software (Xinapse Systems, <http://www.xinapse.com/home.php>) by an expert radiologist of the same hospital center with more than 10 years of experience. Mean lesion volume was 4.1 ± 4.7 ml (range 0.2–18.3 ml), and 3.65 ± 3.94 ml (range 0.1–18.3 ml) on PD-w and FLAIR images, respectively.

### 2.2. Automated lesion segmentation and filling

Automated lesion segmentation and filling was performed using the T1-w and FLAIR image modalities on two publicly available toolkits implemented for the SPM (<http://www.fil.ion.ucl.ac.uk/spm>) software package:

#### 2.2.1. SLS toolbox

The SLS pipeline (<http://atc.udg.edu/salem/slsToolbox/index.html>) was composed of the following automated steps: T1-w and FLAIR images were first skull-stripped and intensity corrected using the Brain Extraction Tool (BET) (Smith, 2002) with optimized parameter choice as described in Popescu et al. (2012), and the N3 method (Sled et al., 1998), respectively. Corrected T1-w and FLAIR images were then linearly co-registered (12-parameter affine) using internal SPM routines, with normalized mutual information as objective function and trilinear

interpolation with no wrapping. Lesion segmentation was performed by an initial tissue segmentation of the T1-w image to separate lesions from tissue, followed by a thresholding step and a regionwise refinement of the FLAIR image (Roura et al., 2015). The initial parameter used to adjust the detected candidate lesions was set to  $\alpha = 2$ , while the percentage of lesion candidate regions to belong to WM and GM over cerebro spinal fluid (CSF), percentage of neighbor voxels belonging to WM, and candidate size was set to  $\lambda_{ts} = 0.7$ ,  $\lambda_{nb} = 0.6$ , and size = 3 mm<sup>3</sup>. Estimated lesion masks were then automatically filled using the method (Valverde et al., 2014), where candidate region voxels were replaced by random values of a normal distribution generated from the mean normal-appearing WM signal intensity of each two-dimensional T1-w slice. The SLF method was run with default parameters.

#### 2.2.2. LST toolbox

The LST pipeline ([www.applied-statistics.de/lst](http://www.applied-statistics.de/lst)) was composed of the following automated steps: T1-w and FLAIR images were skull-stripped and intensity-corrected using the VBM8 toolbox included also as part of the SPM package. Afterwards, corrected T1-w and FLAIR images were linearly (12-parameter affine) and non-linearly co-registered using also internal SPM8 routines. Lesion segmentation was performed by computing an initial tissue segmentation of the T1-w image to compute a lesion belief map based on the FLAIR and T1-w images (Schmidt et al., 2012). This map was refined iteratively weighting the likelihood of belonging to WM or GM against the likelihood of belonging to lesions until no further voxels were assigned to lesions. The required initial threshold kappa was set to  $k = 0.15$ , while the lesion belief map was set to  $l_{bm} = \text{GM}$ . Estimated lesion masks were then automatically filled using an internal filling method inspired by a previous technique proposed in Chard et al. (2010), where candidate region voxels were replaced by random intensities from a Gaussian distribution generated from the normal-appearing WM intensities and then filtered to reintroduce the original spatial variation in WM.

### 2.3. Tissue volume analysis

All images were processed with both toolboxes and compared independently in order to preserve the differences in the internal routines of each toolbox. First, T1-w images processed by the SLS toolbox (see Table 1(a)) were segmented into GM, WM and CSF volumes using SPM8 after following five different pipeline configurations that differed in the level of manual intervention: 1) Original images were segmented including WM lesions (*Original pipeline*); 2) Expert manual lesion annotations were masked before tissue segmentation and relabeled as WM after (*Expert masked pipeline*); 3) Estimated lesion masks provided by the SLS method were masked before tissue segmentation and relabeled as WM after (*SLS masked pipeline*); 4) Estimated lesion masks provided by the SLS method were filled with the SLF method before tissue segmentation (*SLS filled*); and 5) Expert manual lesion annotations were filled before tissue segmentation and used as ground-truth images (*Expert filled pipeline*). In the case of the pipelines where lesions voxels were masked, either with automatic or manual annotations, lesion masks were used to remove lesion voxels in the T1-w image. Therefore, those voxels were not considered during tissue segmentation and were added to the WM class after it to maintain the actual brain volume of each patient. In contrast, in the lesion filling pipelines, automatic or manual lesion annotations were used to refill the correspondent T1-w image voxels with signal intensities similar to the WM, and lesion voxels were considered as normal-appearing WM in tissue segmentation.

All resultant tissue probability maps were binarized into GM, WM and CSF masks by extracting the maximum probability for each particular tissue. GM and WM tissue volume was computed by multiplying the number of voxels in binary masks by the voxel size (1 × 1 × 1.2 mm<sup>3</sup>). Volume measures were normalized to correct the differences between subjects by dividing the GM and WM volume by

**Table 1**

Evaluation pipelines followed in the present study. The set of T1-w images is processed independently for either the SLS (a) and LST (b) toolboxes. First, T1-w images are preprocessed (skull stripped and intensity corrected) using the routines indicated by each toolbox. Then, the preprocessed images are segmented into CSF, GM and WM tissue using SPM8 after following five different pipelines that differ in the level of manual intervention: 1) images are segmented including WM lesions (*Original pipeline*), 2) Expert manual lesion annotations are masked before tissue segmentation (*Expert masked pipeline*), 3) Estimated lesion masks returned by the same toolbox are masked before tissue segmentation (*SLS/LST masked pipeline*), 4) Estimated lesion masks returned by the same toolbox are filled with the lesion-filling method incorporated by each pipeline (*SLS/LST filled pipeline*), and 5) Expert manual lesion annotations are filled before tissue segmentation and used as ground-truth images (*Expert filled pipeline*).

(a)				
Pipeline	Preprocessing	Lesion segmentation	Lesion filling	Tissue segmentation
1. <i>Original</i>	BET + N3	–	–	SPM8
2. <i>Expert masked</i>	BET + N3	Manual	Expert annotations are masked	SPM8
3. <i>SLS masked</i>	BET + N3	SLS	SLS lesion masks are masked	SPM8
4. <i>SLS filled</i>	BET + N3	SLS	SLS lesion masks are filled by SLF	SPM8
5. <i>Expert filled (GT)</i>	BET + N3	Manual	Expert annotations are filled by SLF	SPM8

(b)				
Pipeline	Preprocessing	Lesion segmentation	Lesion filling	Tissue segmentation
1. <i>Original</i>	SPM8	–	–	SPM8
2. <i>Expert masked</i>	SPM8	Manual	Expert annotations are masked	SPM8
3. <i>LST masked</i>	SPM8	LST	LST lesion masks are masked	SPM8
4. <i>LST filled</i>	SPM8	LST	LST lesion masks are filled by LST	SPM8
5. <i>Expert filled (GT)</i>	SPM8	Manual	Expert annotations are filled by LST	SPM8

the whole brain volume. Then, the percent (%) absolute error in total and normal-appearing GM and WM volume was computed between pipelines: *Original* versus *Expert filled* images, *Expert masked* versus *Expert filled*, *SLS masked* versus *Expert filled*, and *SLS filled* versus *Expert filled*. The absolute error in total and normal-appearing GM and WM volume for each automated pipeline were computed using the following equations:

$$GM_{\{1 \dots 4\}vs5} = \frac{|NGMV_{\{1 \dots 4\}} - NGMV_5|}{NGMV_5} \times 100$$

$$WM_{\{1 \dots 4\}vs5} = \frac{|NWMV_{\{1 \dots 4\}} - NWMV_5|}{NWMV_5} \times 100$$

where  $NGMV_{\{1 \dots 4\}}$  and  $NWMV_{\{1 \dots 4\}}$  refer to the normalized GM and WM tissue volume, and the sub-indexes indicate the pipeline used: (1) *Original*, (2) *Expert masked*, (3) *SLS masked*, (4) *SLS filled* and (5) *Expert filled* pipeline used as ground-truth. Normal-appearing GM and WM volume was computed similarly, but lesion voxels were not considered in normalized GM and WM volume estimations. The procedure was then repeated identically for the LST toolbox (see Table 1(b)).

#### 2.4. Statistical analysis

Statistical analysis was performed using the Matlab software package (<http://es.mathworks.com/products/matlab>). Differences in GM and WM volume of each evaluated pipeline were analyzed using a repeated measures ANOVA model with 3 degrees of freedom for the time variable and 207° for the error, followed by a series of post-hoc pairwise significant t-tests with Bonferroni correction between methods. Moreover, the Pearson's linear correlation coefficient was used to compute the correlation between % differences in GM and WM and lesion volume, and between % differences in GM and WM and the error produced by the automated lesion segmentation methods (Error I type: number of false positive outcomes, and Error II type: false negative outcomes). In all the analysis, we considered data significant at p-values < 0.05.

### 3. Results

#### 3.1. Differences in tissue volume

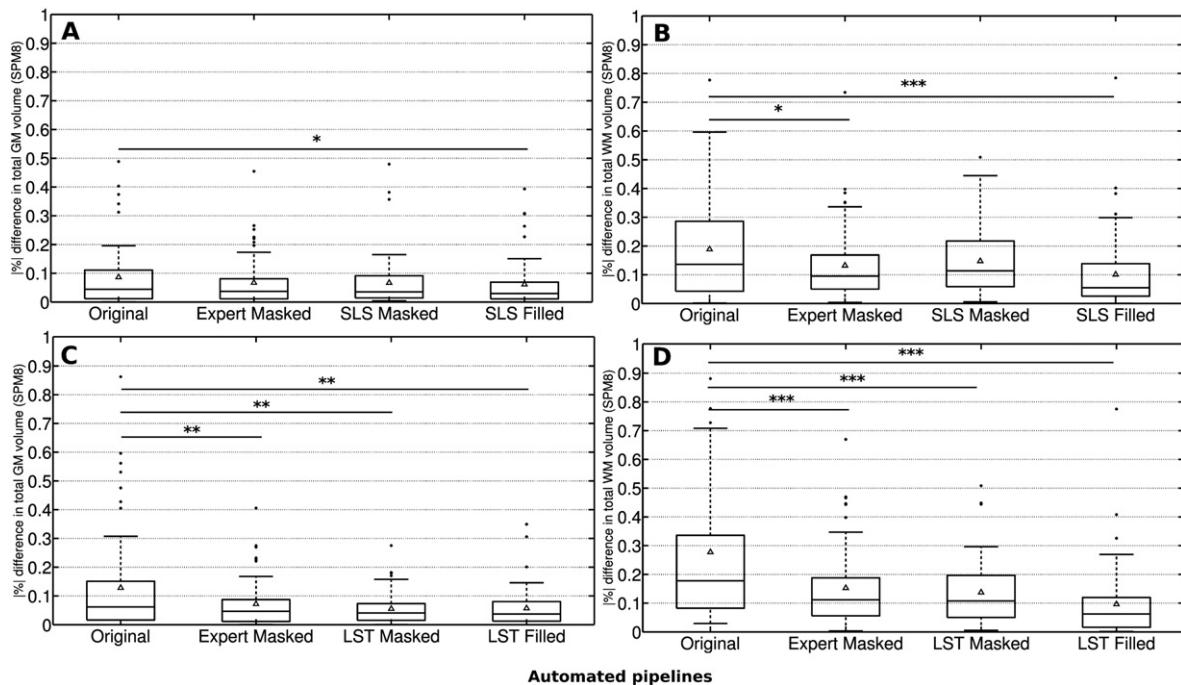
First, we analyzed the differences in total tissue volume between the images processed following each of the SLS pipelines and the images

where expert lesion masks had been filled with the SLF method before tissue segmentation. Automated lesion segmentation and filling reduced significantly the % of error in total GM ( $p < 0.032$ ) on the images processed with the fully automated *SLS filled* pipeline when compared with the same images segmented including lesions (*Original pipeline*) (see Fig. 1A). Similarly, the % differences in total WM were also significantly lower on the *Expert masked* ( $p < 0.040$ ) and *SLS filled* ( $p < 0.002$ ) pipelines when compared with the *Original* images (see Fig. 1B). Differences in total GM and WM between the *SLS masked* and *SLS filled* pipelines were not statistically different.

Regarding the LST toolbox, the mean % of error in GM volume was <0.12% in all the evaluated pipelines and similar to the values reported previously by the SLS, but was significantly higher in the *Original* images ( $p < 0.003$ ) (see Fig. 1C). In WM, the effect of hypo-intense lesions was also significantly higher in the *Original* images ( $p < 0.001$ ) when compared with the rest of the pipelines (see Fig. 1D). As in the SLS, the differences in total GM and WM between *LST masked* and *LST filled* were not significant.

The observed % of error in total GM and WM volume was not only distributed in lesion regions but also in normal-appearing tissue (see Fig. 2). In all the evaluated pipelines but the *Expert masked*, normal-appearing WM was overestimated by the effect of hypo-intense lesion voxels that were still present before tissue segmentation, either because they were not processed intentionally (*Original pipeline*), or as the result of misclassified lesion voxels. Lesion voxels that were classified as WM shifted down the signal intensity threshold between GM and WM and caused the actual GM voxels presenting an intensity profile similar to that of the lesions to be reassigned to WM. Identically; normal-appearing GM was underestimated by the opposite effect of lesion voxels in GM tissue volume. More importantly, in the images processed with the *Original*, *SLS masked*, and *SLS filled* pipelines, the actual % of error in total GM and WM volume was partly canceled between the opposite directions of the errors produced in normal-appearing tissue and the number of remaining lesion voxels that were incorrectly classified as GM (see Fig. 2).

As expected, images where expert lesion masks were masked before segmentation (*Expert masked pipeline*) returned the lowest % of error in normal-appearing GM (see Fig. 2A) and WM (see Fig. 2B) when compared not only with *Original* images ( $p < 0.001$ ), but also with images processed with the *SLS masked* pipeline ( $p < 0.018$ ). The % differences in normal-appearing WM of the images where estimated lesions using SLS were filled were significantly lower than in the same images where lesions were masked ( $p < 0.024$ ). In contrast, differences were

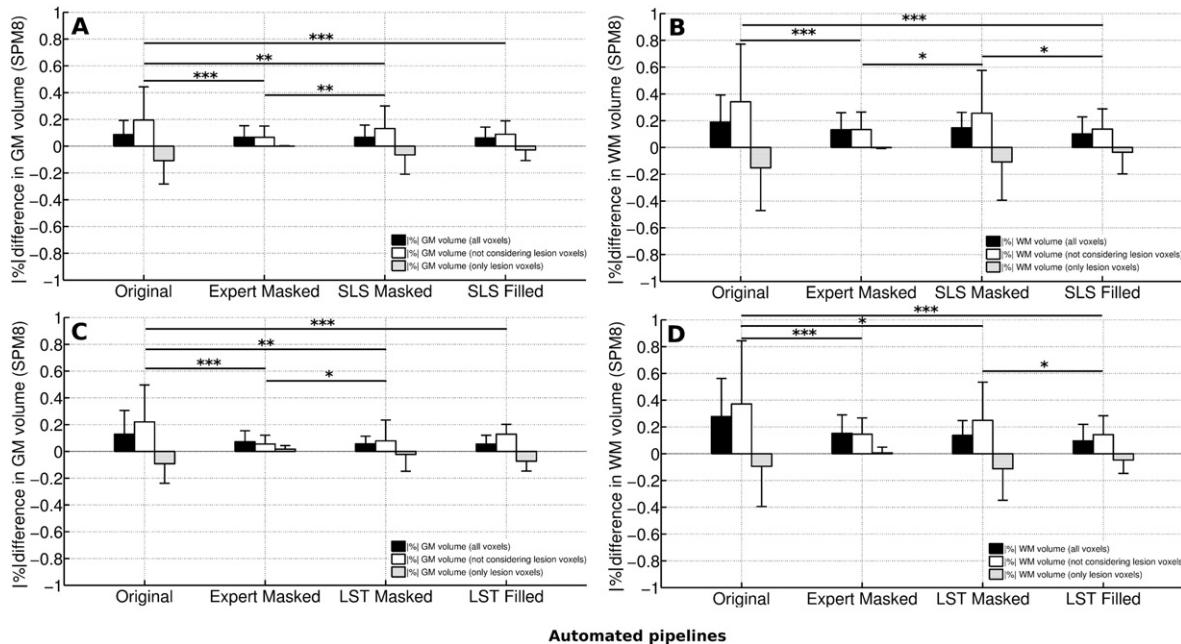


**Fig. 1.** % of absolute error in total GM and WM volume between segmented images where the annotated lesion masks were refilled before tissue segmentation (*Expert filled*) and the same images processed following the *Original*, *Expert masked*, *SLS/LST masked*, and *SLS/LST filled* pipelines. Results for the SLS toolbox are shown in the top row for GM (A) and WM (B), and for the LST toolbox in the bottom row for GM (C) and WM (D). The  $\Delta$  symbol depicts the mean % difference in total GM/WM tissue for each pipeline. Horizontal lines show significant differences between evaluated pipelines with \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

similar for both tissues between the fully automated *SLS filled* and the *Expert masked* pipelines, showing that refilled voxels reduced the effect of hypo-intense lesions in normal-appearing tissue.

Similarly, in LST pipelines part of the % differences in total GM and WM was also partly canceled by the opposite direction of the errors in normal-appearing and the remaining lesion voxels that were

incorrectly classified as GM. As expected, the % differences in normal-appearing GM (see Fig. 2C) and WM (see Fig. 2D) were lower in the *Expert masked* pipeline ( $p < 0.024$ ), due to the null effect of hypo-intense lesions in tissue segmentation. As in SLS, the effect of masking expert lesion masks on the errors in tissue segmentation was similar to that in the automated lesion segmentation and filling. The % differences in



**Fig. 2.** Mean % of absolute error in tissue volume between segmented images where the annotated lesion masks were refilled before tissue segmentation (*Expert filled*) and the same images processed following each of the evaluated pipelines. Results for the SLS toolbox are shown in the top row for GM (A) and WM (B), and for the LST toolbox in the bottom row for GM (C) and WM (D). Differences in tissue volume are split in three regions: those produced when all voxels are considered (black bars), those produced when not considering lesion voxels (white bars), and those produced only in lesion voxels (gray bars). Lesion regions bars are plotted with negative bars to visualize the opposite direction of the errors in lesion voxels with respect to normal-appearing tissue. Vertical lines at each bar depict the % standard deviation difference in tissue volume for each pipeline. Horizontal lines show significant differences in normal-appearance volume between evaluated pipelines with \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

normal-appearing WM of the images where estimated lesions using LST were filled were also significantly lower than in the same images where lesions were masked ( $p < 0.048$ ).

**Fig. 3** depicts for a single patient image, the differences in the overlap of the tissue segmentation classes for each evaluated pipeline and the pertinent *Expert filled* image used as a ground-truth. As expected, the error in normal-appearing tissue (shown in red) was the lowest in the masked pipelines (images F and J), while the number of misclassified lesion voxels (shown in green) was remarkably higher in the *Original* pipelines (images E and I). This fact showed that the inclusion of hypo-intense lesion voxels into the tissue distributions has a clear effect in the misclassification of the normal-appearing tissue between boundaries, and also produces changes in the segmentation of brain structures such as the putamen. In contrast, when compared to these pipelines, the number of misclassified voxels in the automated pipelines incorporating lesion filling (panels H and L) was remarkably lower, although some false negatives were still present in the segmentation due to errors in the automatic lesion segmentation. The number of misclassified voxels was moderately lower in the automated pipelines incorporating lesion filling, when compared with automated pipelines where lesion masks were masked before segmentation (images G and K), although those differences were hardly appreciated in the picture.

When analyzing the % differences in tissue volume between LST and SLS pipelines, we observed that differences in GM between the

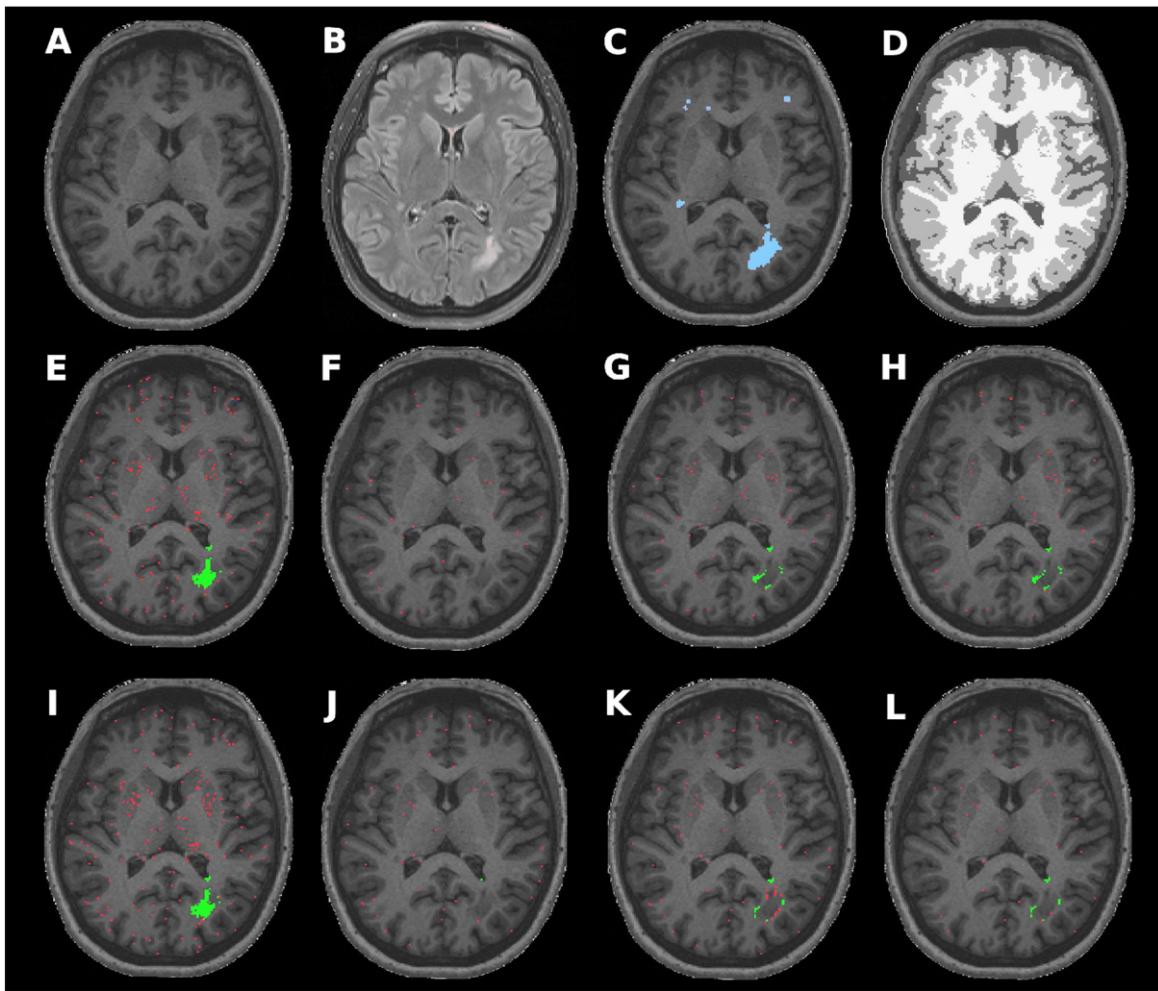
evaluated pipelines were not significant. In contrast, the % differences between *masked* and *filled* pipelines were found significant for total WM between *LST filled* and *SLS masked* ( $p < 0.191$ ), normal-appearing WM between *LST masked* and *SLS filled* ( $p = 0.007$ ), and normal-appearing WM between *LST filled* and *SLS masked* ( $p < 0.002$ ).

Finally, we studied the effect of the image modality used to annotate the expert lesion masks in the overall result. We recomputed the differences in total and normal-appearing GM and WM volume for the two subsets of images where expert masks were annotated using PD-w or FLAIR images. The differences in GM and WM volume between subsets were not statistically different for any of the SLS or LST evaluated pipelines ( $p > 0.42$ ).

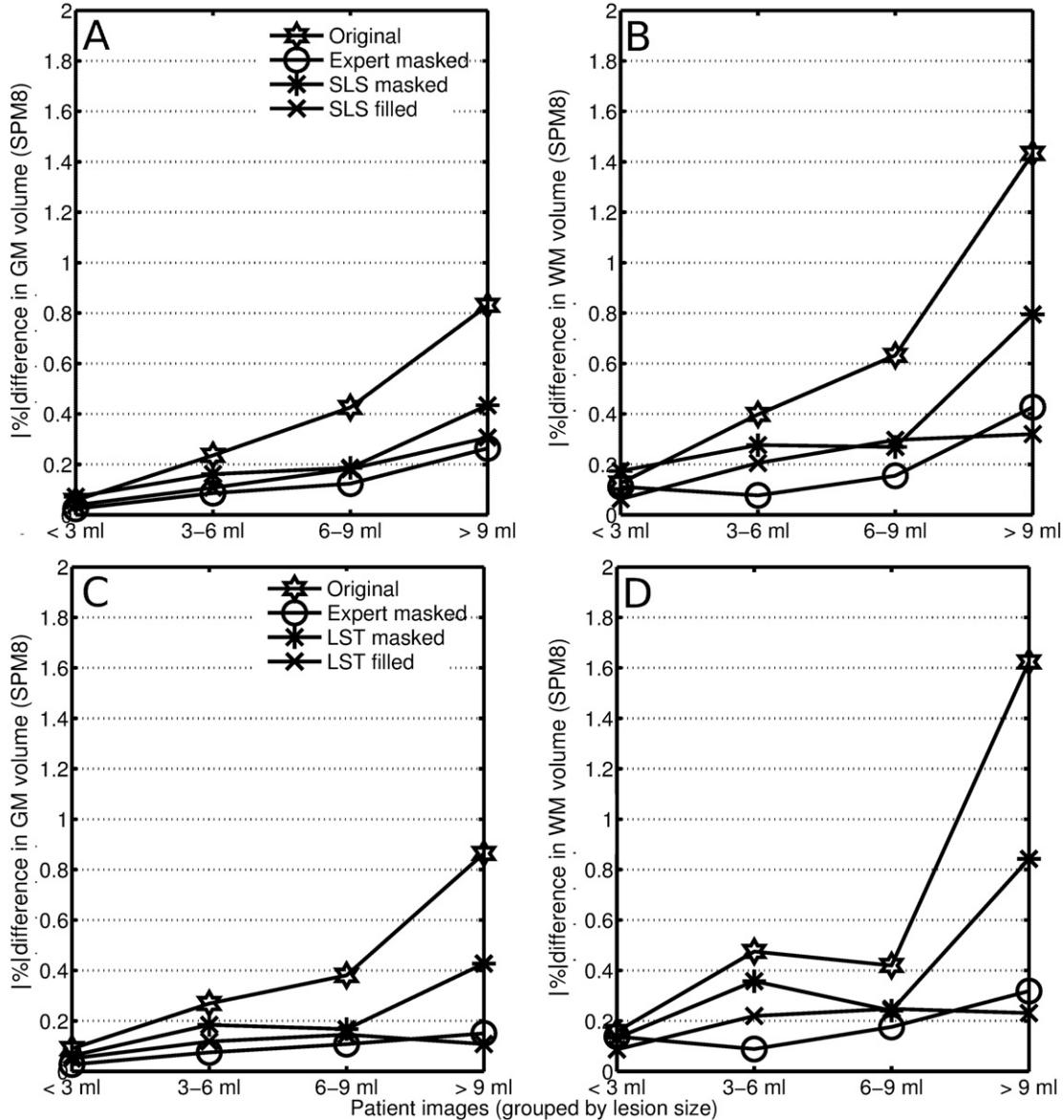
### 3.2. Correlation with lesion volume

We also analyzed the extent to which lesion volume affected the normal-appearing GM and WM volume measurements of each of the evaluated pipelines. Lesion volume strongly correlated with the reported % of error in GM and WM in the *Original*, *Expert masked* and *SLS filled* pipelines ( $r > 0.77$ ,  $p < 0.001$ ), and moderately in the *SLS masked* ( $r > 0.41$ ,  $p < 0.001$ ). However, the effect of lesion volume was different for each evaluated pipeline (**Fig. 4A**).

As expected, the deviation in normal-appearing GM and WM volume was remarkably higher in the images segmented with lesions,



**Fig. 3.** For a single patient image of the dataset, we show the differences in the overlap of the tissue segmentation classes for each evaluated pipeline and the pertinent *Expert filled* image used as a ground-truth. Differences for any tissue class with respect to *Expert filled* are represented in green for lesion voxels, and in red for normal-appearing voxels. First row: input T1-w (A), input FLAIR (B), T1-w with expert annotations highlighted in blue (C), and T1-w output segmentation for the *Expert filled* image with CSF, GM and WM voxels depicted in black, gray, and white, respectively. Second row: for the images processed with the SLS toolkit, differences in any tissue classes for the *Original* (E), *Masked* (F), *SLS masked* (G), and *SLS filled* (H) pipelines. Third row: for the images processed with the LST toolkit, differences in any tissue class for the *Original* (I), *Masked* (J), *LST masked* (K), and *LST filled* (L) pipelines.



**Fig. 4.** Mean % of absolute error in normal-appearing GM and WM volume split by image groups with lesion size in the range (<3 ml, 3–6 ml, 6–9 ml, and >9 ml). Values for each group represent the mean % error between the images processed with the *Expert filled* and each of the evaluated pipelines (Original (\*), Expert masked (○), SLS/LST masked (\*), and SLS/LST filled (\*\*)). Results for the SLS toolbox are shown in the top row for GM (A) and WM (B), and for the LST toolbox in the bottom row for GM (C) and WM (D).

where the % of error in WM was up to 1.46% on images with >9 ml lesion load (see Fig. 4B). The error in WM increased with lesion volume on images where lesions were automatically segmented, but this was remarkably lower on the *SLS filled* images than those that were masked before segmentation (*SLS masked*). On the subset of images with >9 ml, the performance of the *SLS filled* was similar to that of the *Expert masked* pipeline.

Lesion volume also strongly correlated with the observed differences in normal-appearing GM and WM for the *Original* ( $r > 0.78$ ,  $p < 0.001$ ) and *LST masked* ( $r > 0.78$ ,  $p < 0.001$ ) pipelines, and moderately for the *Expert masked* ( $r > 0.36$ ,  $p < 0.001$ ) and *LST filled* ( $r > 0.40$ ,  $p = 0.001$ ). As in the SLS, the error in GM and WM increased with lesion size on images where lesions were automatically segmented or intentionally left, and also increased remarkably in images where automatic lesion masks were masked instead of filled (see Fig. 4D). The % error in normal-appearing GM and WM of the *LST filled* pipeline was similar to that of the *Expert masked*.

### 3.3. Effect of lesion segmentation and filling

The lesion detection accuracy rate (true positives) of the SLS method was  $0.43 \pm 0.21$ , while the Dice similarity coefficient (Dice, 1945) between the estimated and manual annotated masks was  $0.32 \pm 0.17$ . The number of false positive lesion voxels (number of voxels misclassified as lesion), and false negative lesion voxels (number of missed lesion voxels) correlated with the % of error in total GM and WM volume of the *SLS filled* ( $r > 0.60$ ,  $p < 0.001$ ), and *LST filled* pipeline images ( $r > 0.42$ ,  $p < 0.001$ ). This suggested that in these pipelines, the observed error in tissue segmentation was mostly caused by the addition of false positive lesion voxels pertaining to GM that were filled with typical WM signal intensity, and also by the effect of missed hypo-intense WM lesion voxels into tissue distributions. In contrast, the % error in normal-appearing GM and WM in the images processed with the *SLS filled* and *LST filled* pipelines only correlated weakly with the number of false positives. Even some actual GM false positive voxels

were reassigned to WM, still WM voxels that were misclassified as lesion voxels were again reassigned to WM reducing the effect of false positives on the observed errors in normal-appearing tissue volume.

Similarly, the detection accuracy rate of the LST method was  $0.41 \pm 0.20$ , with Dice similarity coefficient of  $0.35 \pm 0.21$ , the number of false positives and false negatives correlated with the % of errors in total GM and WM of the *LST masked* pipeline ( $r > 0.30$ ,  $p = 0.01$ ), and only with the error in total GM of the *SLS masked* pipeline ( $r > 0.52$ ,  $p < 0.001$ ). Moreover, the number of false negative lesion voxels correlated weakly with the % of errors in GM and WM ( $r > 0.40$ ,  $p = 0.001$ ) of both pipelines. Contrary to filled images, actual GM voxels that were incorrectly classified as WM were not considered in tissue volume, reducing the linear correlation between the errors in lesion and tissue segmentation.

We also interchanged the lesion filling methods between the SLS and LST toolboxes and segmented again each set of images with the aim of evaluating the effect of each lesion filling process on the observed % differences in tissue volume. Differences were not statistically different with respect to the original pipelines for both GM and WM volume.

#### 4. Discussion

The effect of lesions on total tissue volume was partly limited due to the canceling effect between the errors produced in normal-appearing tissue and the number of lesion voxels that were segmented as GM (Valverde et al., 2015b). This aspect is relevant because it explains why the observed % of error in total tissue volume was small or not significant between the evaluated pipelines of our study, even within the *Original* images intentionally segmented containing lesions. Furthermore, the % of error in total and normal-appearing WM volume in the images automatically segmented with either the SLS or LST was significantly lower when lesions voxels were filled than when they were masked before segmentation. As also reported in previous studies (Battaglini et al., 2012; Chard et al., 2010; Valverde et al., 2014), our results highlight the necessity to refill WM lesions before tissue segmentation for accurate cross-sectional tissue volume measurements.

However, the accuracy of automated lesion segmentation techniques is still low (Roura et al., 2015; Schmidt et al., 2012). Both automated pipelines overestimated normal-appearing WM (and underestimated GM) mostly by the effect of misclassified lesion voxels. Our results showed a significant but moderate correlation between underestimated total WM and the number of false positives of the *SLS filled* and *LST filled* pipelines. In contrast, the number of false positives correlated weakly with the differences in normal-appearing GM and WM, which might indicate that part of the false positive voxels that were actually WM were correctly reclassified after being filled. The % of error in the *SLS filled* and *LST filled* pipelines also correlated with the number of missed lesion voxels, which in addition to the clear correlation between the errors in tissue segmentation and lesion size, suggests that most of the differences observed in normal-appearing tissue volume were produced by the amount of missed lesion voxels that altered the tissue signal intensity distributions. This aspect suggests that the accuracy of new automatic tissue segmentation pipelines may be increased specially by reducing the number of missed lesion voxels, and in particular when those are hypo-intense in T1-w and should be filled before tissue segmentation. However, this study did not evaluate the methods with RRMS or SPSS image data, because the clinical focus of the study was on the initial CIS phenotype of MS, where paraclinical information is more relevant. In this regard, a further analysis of the accuracy of the evaluated pipelines on images with larger lesion load should be performed.

As expected, the *Expert masked* pipeline reported the lowest error in total and normal-appearing volume, although our results confirmed that masking out lesion voxels before tissue segmentation might not be optimal, as the error in tissue segmentation tends to increase with lesion size (Valverde et al., 2014). More interestingly, the performance of

the fully automated *SLS filled* and *LST filled* pipelines was similar to that of the *Expert masked*, which seems to indicate that upon a certain lesion load, the errors produced by misclassified lesion voxels in the fully automated pipelines were comparable to the masking out error produced by not filling the expert annotations before tissue segmentation.

Within our data, the maximum differences in tissue volume produced by the *SLS filled* and *LST filled* might be lower than the own reproducibility of the SPM method, as stated in previous studies (De Boer et al., 2010; Nakamura et al., 2014). However, a direct comparison between studies has to be contemplated with care, because we did not perform a scan-reposition-rescan analysis of the evaluated pipelines, and consequently the differences in tissue volume produced by automated methods should be added to the inner reproducibility of the tissue segmentation method. Additionally, differences in the pre-processing pipelines between studies should be also contemplated, as shown in previous studies (Boyes et al., 2008; Zheng et al., 2009). The maximum differences in tissue volume produced by the fully automated pipelines also raises the question if the observed differences could be considered negligible when compared with the loss in tissue volume observed in follow-up scans. In this aspect, the differences in tissue volume shown by both the *SLS filled* and *LST filled* are remarkably lower than yearly tissue loss reported in recent clinical studies (Filippi et al., 2013; Sastre-Garriga et al., 2014; Uher et al., 2014). Hence, given the small error introduced by these methods, we recommend the use of either the SLS or LST toolkit.

There are a number of limitations in this work that have to be considered. This study was conducted using single-center data, and hence the applicability in a multi-center study was not determined here. The lack of manual tissue annotations does not allow us to analyze the tissue segmentation accuracy of each of the evaluated pipelines. Gold-standard annotations are time-consuming and have to be delineated by trained experts, a task which unfortunately is not always possible, especially when the number of subjects grows. In this aspect, the results of this study have to be understood under the premise that we are not evaluating the accuracy of the tissue segmentation methods, but the differences with respect to the manual expert pipeline that introduces the lowest error in tissue volume in images containing WM lesions (Battaglini et al., 2012; Valverde et al., 2014). The % of error in GM and WM volume introduced by the evaluated pipelines was small, and it was difficult to scale our findings with previous studies, given the differences in preprocessing and internal routines of each pipeline. Furthermore, in spite of the small error observed, our claims about the effectiveness of the fully automated pipelines have to be prudent, given the lesion load of the cohort of CIS patients of our study. As a future work, we will investigate the effect of images with higher lesion load on automated lesion segmentation, the posterior lesion filling process, and the impact of these automated processes in tissue segmentation methods.

In summary, this study shows that the automated lesion segmentation and filling methods included in the LST and SLS toolboxes reduce significantly the impact of T1-w hypo-intense lesions on the SPM8 tissue segmentation method. Our results show that compared with the evaluated pipelines that require manual expert intervention, the accuracy in tissue segmentation is not affected remarkably on images processed with the fully automated pipelines. This is relevant and suggests that LST and SLS toolboxes allow for performing accurate brain tissue volume measurements without any kind of manual intervention. The possibility of filling MS white matter lesions without manual delineation of lesions is pertinent not only in terms of time and economic costs, but also to avoid the inherent *intra/inter* variability between manual annotations.

#### Acknowledgments

S. Valverde holds a FI-DGR2013 grant from the Generalitat de Catalunya. E. Roura holds a BR-UdG2013 grant. This work has been partially

supported by “La Fundació la Marató de TV3” and by Retos de Investigación TIN2014-55710-R.

## References

- Ashburner, J., Friston, K., 2005. Unified segmentation. *NeuroImage* 26, 839–851.
- Battaglini, M., Jenkinson, M., De Stefano, N., 2012. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum. Brain Mapp.* 33 (9), 2062–2071.
- Boyes, R., Gunter, J.L., Frost, C., Janke, A.L., Yeatman, T., Hill, D., Bernstein, M.A., et al., 2008. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *NeuroImage* 39, 1752–1762.
- Cabezas, M., Oliver, A., Roura, E., Freixenet, J., Vilanova, J.C., Ramió-Torrentà, L., Rovira, A., Lladó, X., 2014. Automatic multiple sclerosis lesion detection in brain MRI by FLAIR thresholding. *Comput. Methods Prog. Biomed.* 115 (3), 147–161.
- Ceccarelli, A., Jackson, J.S., Tahirid, S., Arora, A., Gorky, J., Dell'Osoglio, E., Bakshi, A., et al., 2012. The impact of lesion in-painting and registration methods on voxel-based morphometry in detecting regional cerebral gray matter atrophy in multiple sclerosis. *Am. J. Neuroradiol.* 33 (8), 1579–1585.
- Chard, D.T., Jackson, J.S., Miller, D.H., Wheeler-Kingshott, C.A., 2010. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *J. Magn. Reson. Imaging* 32, 223–228.
- Clark, K.A., Woods, R.P., Rottenberg, D.A., Toga, A.W., Mazziotta, J.C., 2006. Impact of acquisition protocols and processing streams on tissue segmentation of T1-w weighted MR images. *NeuroImage* 29, 185–202.
- De Boer, R., Vrooman, H., Ikram, M., Vernooy, M., Breteler, M., Van der Lugt, A., Niessen, W., 2010. Accuracy and reproducibility study of automatic MRI brain tissue segmentation methods. *NeuroImage* 51 (3), 1047–1056.
- De Bresser, J., Portegies, M.P., Leemans, A., Biessels, G.J., Kappelle, L.J., Viergever, M.A., 2011. A comparison of MR based segmentation methods for measuring brain atrophy progression. *NeuroImage* 54 (2), 760–768.
- Dice, L.R., 1945. Measures of the amount of ecological association between species. *Ecology* 26 (3), 297–302.
- Filippi, M., Preziosa, P., Copetti, M., Riccitelli, G., Horsfield, M.A., Martinelli, V., Comi, G., Rocca, M.A., 2013. Gray matter damage predicts the accumulation of disability 13 years later in MS. *Neurology* 81 (20), 1759–1967.
- Ganlier, O., Oliver, A., Díez, Y., Freixenet, J., Vilanova, J.C., Beltrán, B., Ramió-Torrentà, L., Rovira, A., Lladó, X., 2014. A subtraction pipeline for automatic detection of new appearing multiple sclerosis lesions in longitudinal studies. *Neuroradiology* 56 (5), 363–374.
- 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.
- Guizard, N., Coupé, P., Fonov, V., Manjón, J., Arnold, D.L., Collins, D.L., 2015. Rotation-invariant multi-contrast non-local means for MS lesion segmentation. *NeuroImage* Clin. 8, 376–389.
- Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J.C., Quiles, A., Valls, L., Ramió-Torrentà, L., Rovira, A., 2012. Segmentation of multiple sclerosis lesions in brain MRI: a review of automated approaches. *Inf. Sci.* 186 (1), 164–185.
- Nakamura, K., Fisher, E., 2009. Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. *NeuroImage* 44 (3), 769–776.
- Nakamura, K., Guizard, N., Fonov, V.S., Narayanan, S., Collins, D.L., Arnold, D.L., 2014. Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis. *NeuroImage Clin.* 4, 10–17.
- Pérez-Miralles, F., Sastre-Garriga, J., Tintoré, M., Arrambide, G., Nos, C., Perkal, H., Río, J., et al., 2013. Clinical impact of early brain atrophy in clinically isolated syndromes. *Mult. Scler.* 19 (14), 1878–1886.
- Popescu, V., Battaglini, M., Hoogstrate, W.S., et al., 2012. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. *NeuroImage* 61 (4), 1484–1494.
- Popescu, V., Ran, N.C.G., Barkhof, F., Chard, D.T., Wheeler-Kingshott, C.A., Vrenken, H., 2014. Accurate GM atrophy quantification in MS using lesion filling with co-registered 2D lesion masks. *NeuroImage Clin.* 4 (January), 366–373.
- Roura, E., Oliver, A., Cabezas, M., Valverde, S., Pareto, D., Vilanova, J.C., Ramió-Torrentà, L., Rovira, A., Lladó, X., 2015. A toolbox for multiple sclerosis lesion segmentation. *Neuroradiology* <http://dx.doi.org/10.1007/s00234-015-1552-2>.
- Sanfilipo, M.P., Benedict, R.B., Sharma, J., Weinstock-Guttman, B., Bakshi, R., 2005. The relationship between whole brain volume and disability in multiple sclerosis: a comparison of normalized gray vs. white matter with misclassification correction. *NeuroImage* 26 (4), 1068–1077.
- Sastre-Garriga, J., Tur, C., Pareto, D., Vidal-Jordana, A., Auger, C., Río, J., Huerga, E., Tintoré, M., Rovira, A., Montalban, X., 2014. Brain atrophy in natalizumab-treated patients: a 3-year follow-up. *Mult. Scler.* <http://dx.doi.org/10.1177/1352458514556300>.
- Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förtschler, A., Berthele, A., Hoshi, M., et al., 2012. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage* 59 (4), 3774–3783.
- Sled, J.G., Zijdenbos, P., Evans, C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17 (1), 87–97.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage* 17 (1), 479–489.
- Uher, T., Horáková, D., Bergstrand, N., Tyblová, M., Ramasamy, D.P., Seidl, Z., Vaněcková, M., Krasensky, J., Havrdová, E., Zivadinov, R., 2014. MRI correlates of disability progression in patients with CIS over 48 months. *NeuroImage: Clinical* 6 (January). Elsevier B.V., pp. 312–319.
- Valverde, S., Oliver, A., Lladó, X., 2014. A white matter lesion filling approach to improve brain tissue volume measurements. *NeuroImage Clin.* 6 (January), 86–92.
- Valverde, S., Oliver, A., Cabezas, M., Roura, E., Lladó, X., 2015a. Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. *J. Magn. Reson. Imaging* 41 (1), 93–101.
- Valverde, S., Oliver, A., Díez, Y., Cabezas, M., Vilanova, J.C., Ramió-Torrentà, L., Rovira, A., Lladó, X., 2015b. Evaluating the effects of white matter multiple sclerosis lesions on the volume estimation of 6 brain tissue segmentation methods. *Am. J. Neuroradiol.* <http://dx.doi.org/10.3174/ajnr.A4262>.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57.
- Zheng, W., Chee, M.W., Zagorodnov, V., 2009. Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. *NeuroImage* 48, 73–83.



# Chapter 6

## Automated tissue segmentation of MR brain images in the presence of white matter lesions.

In this chapter, we propose a novel automated pipeline for tissue segmentation of MS patient images containing lesions. The accuracy of the method is evaluated using both the challenge MRBrainS13 database <sup>1</sup> and a 3T MS database of MS patient images. We validate the accuracy of the proposed method with other state-of-the-art techniques. A public version of the method has been released for public use. The proposed pipeline has been described in detail in the next paper and submitted to the Medical Imaging Journal:

Paper published in Medical Image Analysis (MIA)  
Volume: xx, Issue: x, Pages: xx-xx, Published: xxx  
DOI: xxxx  
Quality Index: 3.28 (Quartile 1)

---

<sup>1</sup><http://mrbrains13.isi.uu.nl/>

# Automated tissue segmentation of MR brain images in the presence of white matter lesions

Sergi Valverde<sup>a,\*</sup>, Arnau Oliver<sup>a</sup>, Eloy Roura<sup>a</sup>, Sandra González<sup>a</sup>, Deborah Pareto<sup>b</sup>, Joan C. Vilanova<sup>c</sup>, Lluís Ramió-Torrentà<sup>d</sup>, Àlex Rovira<sup>b</sup>, Xavier Lladó<sup>a</sup>

<sup>a</sup>Research institute of Computer Vision and Robotics, University of Girona, Spain

<sup>b</sup>Magnetic Resonance Unit, Dept of Radiology, Vall d'Hebron University Hospital, Spain

<sup>c</sup>Girona Magnetic Resonance Center, Spain

<sup>d</sup>Multiple Sclerosis and Neuroimmunology Unit, Dr. Josep Trueta University Hospital, Spain

---

## Abstract

During the last years, the increasingly interest of brain tissue volume measurements on clinical settings has lead to the development of a wide number of automated tissue segmentation methods. However, white matter lesions are known to reduce the accuracy of automated tissue segmentation methods, which requires to manually annotate lesions and refill them before segmentation, which is tedious and time-consuming. Here, we propose a new fully automated T1-w/FLAIR tissue segmentation approach designed to deal with images in the presence of WM lesions, which integrates a robust partial volume tissue segmentation with WM outlier rejection and filling, combining intensity, probabilistic and morphological prior maps. We evaluate the accuracy of the method on the MRBrainS13 tissue segmentation challenge database, and also on a set of Multiple Sclerosis (MS) patient images. On both databases, we validate the performance of our method with other state-of-the-art techniques. On the MRBrainS13 data, the presented approach was the best unsupervised ranked method of the challenge (7th position) and clearly outperformed other unsupervised pipelines such as *FAST* and *SPM12*. On MS data, the differences in tissue segmentation between the images segmented with our method and the same images where manual expert annotations were used to refill lesions on T1-w images before segmentation were lower or similar to the best state-of-the-art pipeline incorporating automated lesion segmentation and filling. Our results show that the proposed pipeline quantitatively improved the accuracy of tissue segmentation while it achieved very competitive results on MS images. A public version of the approach is available to download for the neuro-imaging community.

**Keywords:** Brain, MRI, multiple sclerosis, automatic tissue segmentation, white matter lesions

---

## 1. Introduction

Brain tissue volume based on Magnetic Resonance Imaging (MRI) is increasingly being used in clinical settings to assess brain volume in different neurological dis-

eases such as the Multiple Sclerosis (MS) (Giorgio and De Stefano, 2013). In MS, several studies have analyzed the histopathological changes of patients with respect to the progression of the disease, showing that the percentage of change in brain volume tends to correlate with worsening conditions (Pérez-Miralles et al., 2013; Sormani et al., 2014). However, manual segmentation of brain tissue is both challenging and time-consuming because of the large number of MRI slices for each patient which composes the three-dimensional information, and the inher-

---

\*Corresponding author. S. Valverde, Ed. P-IV, Campus Montilivi, University of Girona, 17071 Girona (Spain). e-mail: svalverde@eia.udg.edu. Phone: +34 972 418878; Fax: +34 972 418976.

ent intra/inter-observer variability of manually segmented scans (Cabezas et al., 2011). The development of automated MS tissue segmentation methods that can segment large amounts of MRI data, do not suffer from intra/inter-observer variability and specific changes of the brain such as MS associated lesions and brain atrophy are still an active research field (Klauschen et al., 2009; de Bresser et al., 2011; Valverde et al., 2015a; Vincken et al., 2015).

Different brain tissue segmentation methods have been used in MS so far. General purpose intensity based methods combining intensity and a priori statistical anatomic information such as FAST (Zhang et al., 2001) or SPM (Ashburner and Friston, 2005) are nowadays widely used. However, tissue abnormalities found in MS image patients such as White Matter (WM) lesions reduce the accuracy of these techniques (Chard et al., 2010; Battaglini et al., 2012) usually causing an overestimation of Gray Matter (GM) tissue not only by the effect of hypo-intense WM lesion voxels classified as GM, but also by the effect of these lesion voxels in normal-appearing tissue (Valverde et al., 2015b). In these cases, in-painting lesions on the T1-weighted image (T1-w) with signal intensities of the normal-appearing WM before tissue segmentation may be used to reduce the effects of WM lesions on tissue segmentation (Chard et al., 2010; Battaglini et al., 2012; Valverde et al., 2014). However, MS lesions have to be delineated manually first, which may be tedious, challenging and time-consuming task depending of the characteristics of the image (Lladó et al., 2012).

Regarding this issue, a wide number of automated lesion segmentation techniques have been proposed during the last years (Lladó et al., 2012; García-Lorenzo et al., 2013). Most of these methods integrate other imaging modalities such as T2-weighted, Proton Density (PD) and Fluid Attenuated Inverse Recovery (FLAIR), as these modalities present a high contrast between tissue and lesions (Lladó et al., 2012). More recent techniques include supervised learning classifiers based on spatial decision forest (Geremia et al., 2011), statistical methods (Sweeney et al., 2013), patch-based models (Guizard et al., 2015) or adaptive dictionary learning methods (Deshpande et al., 2015). Further, different unsupervised learning techniques make use of probabilistic models to separate WM lesions from normal-appearing tissue by considering lesions as an outlier class (Harmouche et al., 2015; Tomas-Fernandez and Warfield, 2015; Jain et al.,

2015). Also, other unsupervised techniques make use of the signal intensity of lesions on FLAIR several thresholding methods with post-processing steps to automatically segment lesions (Schmidt et al., 2012; Roura et al., 2015). In contrast, there are fewer studies that have been focused on the tissue segmentation of MS images containing lesions. Those include non-supervised techniques combining intensity, anatomical and morphological maps (Nakamura and Fisher, 2009; Shiee et al., 2010), or supervised methods such as statistical classifiers (Datta and Narayana, 2013), atlas based nearest-neighbor methods (De Boer et al., 2009) and sparse dictionary learning approaches (Roy et al., 2015).

The increasing amount of published studies regarding automated WM lesion segmentation may be given by the particular need of a quantitative analysis of focal MS lesions in individual and temporal studies (Lladó et al., 2012). Recent studies in MS (Chard et al., 2010; Gelineau-Morel et al., 2012; Ceccarelli et al., 2012; Pérez-Miralles et al., 2013; Popescu et al., 2014; Magon et al., 2014; Valverde et al., 2015b) indicate a certain tendency to the use of widely validated segmentation tools such as Siena (Smith et al., 2002), FAST (Zhang et al., 2001) or SPM (Ashburner and Friston, 2005) in combination with automated lesion segmentation and/or lesion-filling approaches, although their application in clinical practice is still not generalized (Giorgio and De Stefano, 2013).

In this paper, we present the Multiple Sclerosis SEGmentation pipeline (*MSSEG*), a novel multi-channel method designed to segment GM, WM and cerebro-spinal fluid (CSF) tissues in images of MS patients. This method is motivated by our previous analysis of the effects of tissue segmentation on MS images (Valverde et al., 2015b), the role of lesion-filling (Valverde et al., 2014), and its combination with automated lesion segmentation on tissue segmentation (Valverde et al., 2015b). Similar to the works of Nakamura and Fisher (2009) and Shiee et al. (2010), our approach utilizes a combination of intensity, anatomical and morphological prior maps to guide the tissue segmentation. However, tissue segmentation is here based on a robust partial volume segmentation where WM outliers are estimated and refilled before segmentation using a multi-channel post-processing algorithm. This post-processing algorithm is partially inspired on MS lesion segmentation algorithm

proposed by Roura et al. (2015), but here we integrate multi-channel support, partial volume segmentation, spatial context, and prior anatomical and morphological atlases. In order to perform quantitative and qualitative evaluations of our approach, we analyze its accuracy with the MRBrainS13 challenge database which includes manual tissue annotations, and also with a set of MS patient images with different lesion burden. We quantitatively compare the performance of our approach with different state-of-the-art techniques that also competed on the MRBrainS13 challenge and/or have been used in recent MS studies. Furthermore, we also analyze the differences in the performance of our approach when using only T1-w or when using the multi-channel approach that includes T1-w and FLAIR. The *MSSEG* method is currently available for download at our research group webpage (<http://atc.udg.edu/nic/msseg/index.html>).

## 2. Methods

The proposed brain tissue segmentation method is composed of five different processes: registration of statistical atlas into T1-w space (Sec. 2.2), tissue estimation (Sec. 2.3), detection and re-assignment of lesion candidate to T1-w (Sec. 2.4), tissue re-estimation, and partial volume re-assignment of tissue maps into CSF, GM and WM (Sec. 2.5). The overall schema of the pipeline is depicted in figure 1. We describe each step in detail in the following subsections.

### 2.1. Notation

To describe our proposed approach, we employ the following notation.  $T$  and  $F$  denote the input images T1-w and FLAIR, respectively.  $P^c$  denotes a probabilistic tissue atlas of particular class  $c = \{csf, csfgm, gm, gmwm, wm\}$ .  $S^{st}$  denotes a morphological brain atlas of a particular parcellated structure  $st$ . For each of the above images,  $T_j$ ,  $F_j$ ,  $P_j^c$  and  $S_j^{st}$  denote an observation at a voxel  $j \in \Omega$ , being  $\Omega$  the image domain.

### 2.2. Tissue prior registration

The MNI-ICBM 152 2009a Nonlinear T1-w average structural template image<sup>1</sup> was first affine registered to

---

<sup>1</sup><http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009>

the native T1-w image space based on a block matching approach (Ourselin et al., 2002), and then non-rigid registered with a fast free-form deformation method (Modat et al., 2010), both using the Nifty Reg package<sup>2</sup>. Obtained transformation parameters were then used to resample the available MNI CSF, GM and WM tissue priors to the T1-w space. The resampled probabilistic tissue maps  $P^{csf}$ ,  $P^{gm}$  and  $P^{wm}$  were extended to build intermediate partial volumes  $P^{csfgm}$  as ( $P^{csf} \geq 0.5 \cap P^{gm} \geq 0.5$ ) and  $P^{gmwm}$  as ( $P^{gm} \geq 0.5 \cap P^{wm} \geq 0.5$ ) and taking the mean value of the two input atlases.

Besides, a morphological brain structure atlas was first parcellated on the original MNI atlas using the hierarchical algorithm proposed by Pohl et al. (2007) on EM-Segmenter<sup>3</sup>, and then manually refined for the selected structures. The resulting atlas ( $S^{st}$ ) consisted of useful structures for tissue segmentation such as cortical GM ( $S^{CORTEX}$ ), ventricles ( $S^{VENT}$ ), basal ganglia ( $S^{BASAL}$ ) and brainstem ( $S^{BRAINSTEM}$ ). The same transformation parameters were also used to resample the morphological atlas to the T1-w space.

### 2.3. Tissue estimation

Brain tissue was estimated following a robust fuzzy-clustering approach similar to the one proposed in Pham (2001), as this method provides a straightforward implementation, fairly robust behavior including spatial context information, applicability to multichannel data, and the ability to model uncertainty within the data (Pham, 2001). In our approach, we designed the method to segment 5 classes in order to preserve better the differences in signal intensity between local regions of the brain and lesion candidates. We also extended the spatial penalizing weights by incorporating the probabilistic tissue priors in the segmentation process similarly to Shiee et al. (2010). Hence we modified the objective function proposed by Pham (2001) in order to incorporate also prior-atlas information as follows:

---

<sup>2</sup><http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg>

<sup>3</sup><https://www.slicer.org/slicerWiki/index.php/EMSegmenter-Overview>

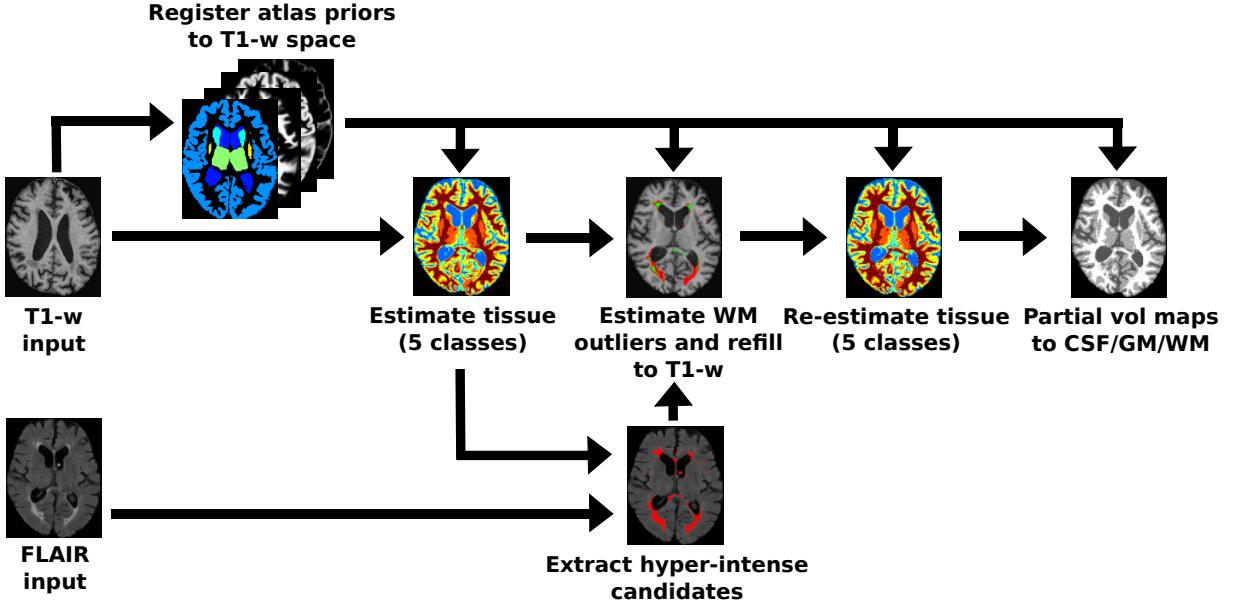


Figure 1: The proposed method *MSSEG* consists of five different steps: 1) Three statistical *a-priori* tissue atlases (CSF, GM and WM) and a brain structure atlas are first registered into the patient space (Sec. 2.2) and then used to 2) guide the tissue segmentation of the input T1-w image (Sec. 2.3). 3) Then, the same output segmentation is employed to detect and reassign candidate regions into WM based on the registered *a-priori* and hyper-intense FLAIR maps if available (Sec. 2.4). The voxel intensities of candidate regions on T1-w are then refilled with normal-appearance WM intensities and 4) tissue is re-estimated (Sec. 2.3). 5) Finally, intermediate volume maps are reassigned into CSF, GM and WM using both neighbor and spatial prior information (Sec. 2.5).

$$\begin{aligned}
J_{MSSEG} = & \sum_{j \in \Omega} \sum_{k=1}^C u_{jk}^q \|T_j - v_k\|^2 + \\
& + \frac{\beta}{2} \sum_{j \in \Omega} \sum_{k=1}^C u_{jk}^q \sum_{l \in N_j^w} \sum_{m \in M_k} u_{lm}^q + \\
& + \frac{\gamma}{2} \sum_{j \in \Omega} \sum_{k=1}^C u_{jk}^q \sum_{l \in N_j^w} \sum_{m \in M_k} P_l^k
\end{aligned} \quad (1)$$

where  $\{k \in C \mid C = \{csf, csfgm, gm, gwm, wm\}\}$ ,  $u_{jk}$  denotes the membership probability of each voxel  $j$  for a particular class,  $v_k$  are the cluster signal intensity centers of each class,  $N_j^w$  is the set of two-dimensional (2D)  $(2w + 1)^2$  or three-dimensional (3D)  $(2w + 1)^3$  neighbors centered on the voxel  $j$ , and  $M_k = \{1, \dots, C\} \setminus \{k\}$ . This approach depends on four parameters to adjust the membership functions: the weighting parameter  $q$  that controls the degree of fuzziness, the spatial constraint parameter  $\beta$  that controls the amount of neighbor information added, the prior belief parameter  $\gamma$  used to control the amount of prior atlas information about each tissue, and finally the window radius of neighbors  $w$ .

Similar to the work of Pham (2001), an iterative algorithm to minimize (1) was derived by evaluating the cen-

troids and the functions that satisfy a zero gradient condition as follows:

---

#### Algorithm 1 Tissue estimation

---

- 1: Obtain the initial estimates of the centroids for each class  $k = \{1, \dots, C\}$ :  

$$v_k = \frac{1}{n} \sum_{j \in \Omega} (T_j^k \mid P_j^k \geq 0.5) \quad n = |(T_j^k \mid P_j^k \geq 0.5)|$$
  - 2: Compute the membership functions  $u_{jk}$
  - 3: Compute the new centroids:  

$$v_k = \frac{\sum_{j \in \Omega} u_{jk}^q T_j}{\sum_{j \in \Omega} u_{jk}^q} \quad k = \{1, \dots, C\}$$
  - 4: Repeat steps 2 and 3 until convergence
- 

Initial centroids  $v_k$  were estimated for each class  $C$  by taking the mean signal intensity of the voxels on the T1-w image with prior-tissue probability  $P_j^k \geq 0.5$ . The membership function  $u_{jk}$  was also adapted to incorporate prior-atlas information and computed as follows:

$$u_{jk} = \frac{(\|T_j - v_k\|^2 + \beta \sum_{l \in N_j} \sum_{m \in M_k} u_{lm}^q + \gamma \sum_{l \in N_j} \sum_{m \in M_k} P_l^m)^{-1/(q-1)}}{\sum_{i=1}^C ((\|T_j - v_i\|^2 + \beta \sum_{l \in N_j} \sum_{m \in M_i} u_{lm}^q + \gamma \sum_{l \in N_j} \sum_{m \in M_i} P_l^m)^{-1/(q-1)})} \quad (2)$$

The five classes tissue segmentation mask  $SEG_j$  was

computed by assigning to each voxel the class with maximum membership as follows:

$$SEG_j = \arg \max_k u_{jk} \quad \forall j \in \Omega \quad (3)$$

The parameters  $q$ ,  $\gamma$  and  $w$  can be tuned manually to increase the performance of the method, but were set to default values  $q = 2$ ,  $\gamma = 0.1$  and  $w = 1$  with 2D that worked well in the majority of cases. In contrast, the  $\beta$  parameter depends on the brightness of the image, the deviation of the signal intensities of voxel class members with respect to their centroid value, and image noise (Pham, 2001). Hence, choosing a proper value for the  $\beta$  parameter was important to obtain optimal or near-optimal performance. In our implementation, we automated the choose of the  $\beta$  parameter by fitting a function of the optimal empirical selection of the parameter with respect to different levels of noise. To do so, we estimated iteratively the sub-optimal  $\beta$  parameter of 10 images of the Brainweb dataset<sup>4</sup> that included different noise level (1-9%) and ground-truth annotations. For each image, we also computed the noise level using the Fast Noise Variance method proposed by Immerkær (1996). Then, the correspondent  $\beta$  parameters and noise levels were used to fit a polynomial function to interpolate the  $\beta$  parameter. For all the evaluated images in the paper, we have automatically approximated the  $\beta$  as a function  $B(x)$  of their noise level  $x$  as  $B(x) = 0.0011x^4 - 0.0015x^3 + 0.0074x^2 - 0.001x + 0.05$ .

#### 2.4. Reassign WM outliers to T1-w

In a three class tissue segmentation approach, lesion regions are usually classified as either GM or WM, given the hypo-intense signal intensity profile of WM lesions. This impedes in some cases to differentiate them from surrounding GM and WM. In contrast, by creating the intermediate classes *CSFGM* and *GMWM*, new local clusters of voxels with similar signal intensities are delimited, increasing the chances that WM lesion regions may be differentiated from normal-appearing GM and WM. Following this assumption, we estimated WM lesion regions by analyzing all local regions not initially segmented as WM based on their the prior probability and the spatial connection to WM.

---

<sup>4</sup><http://brainweb.bic.mni.mcgill.ca/brainweb/>

First, different binary segmentation masks  $M^c$  were computed for each of the classes  $c = \{gmwm, gm, csfgm\}$ . For each mask, all 2D regions of connected components were computed using a flood-fill algorithm with 4-connected neighborhood. We define the set of all 4-connected  $p$  regions given an input binary image as follows:

$$R_p^T \leftarrow \oplus(M_j^c, n), \quad p = \{1, \dots, |R_p^T|\}$$

where the operator  $\oplus$  refers to the connected components function and  $n$  is the number of connected neighbors.

Secondly, a map of hyper-intense region candidates was computed on the FLAIR image following the same strategy shown in Roura et al. (2015). The binary mask  $M^{GM}$  was first used to compute the intensity distribution on the FLAIR image, where GM is typically hyper-intense with respect to CSF and WM, and WM lesions are considered hyper-intense outliers to GM. The mean and standard deviation of the GM distribution was computed using the full-width at half maximum (FWHM) of the main peak of a generated histogram. Then, an initial map of hyper-intense regions voxels  $H^{FLAIR}$  was determined by thresholding the FLAIR image  $F$  as follows:

$$H_j^{FLAIR} = \begin{cases} 1 & \text{if } F_j > \mu + \alpha\sigma \\ 0 & \text{otherwise} \end{cases} \quad \forall j \in \Omega \quad (4)$$

where  $\mu$  and  $\sigma$  were the mean and standard deviation, respectively, of the GM distribution as computed using the FWHM, and  $\alpha$  was a weighting parameter that scaled the minimum signal intensity of outliers. The binary mask  $H^{FLAIR}$  was then used to group the candidate voxels into connected regions using the same method proposed before:

$$R_t^F \leftarrow \oplus(H^{FLAIR}, n), \quad t = \{1, \dots, |R_t^F|\}$$

where  $n$  was set to 3D connected elements ( $n = 6$ ) in order to reduce the amount of 2D false positive regions such as hyper-intense sub-arachnoid tissue.

Given the computed binary masks for each tissue  $M^{gmwm}$ ,  $M^{gm}$  and  $M^{csfgm}$ , the map of hyper-intense voxels on FLAIR  $H^{FLAIR}$ , and its connected components  $R_t^F$ , we used an iterative algorithm to estimate the regions with

high probability to pertain to WM. Lesion filling of selected regions was integrated in the same algorithm. Figure 2 shows in detail each of the steps of the algorithm. Regions not overlapping cortical GM on the morphological prior  $S_{\text{cortex}}^{\text{CORTEX}}$  were only processed if a matched region was also hyper-intense in FLAIR, in order to reduce the amount of false positive regions such as isolated cortical GM segmented regions. Regions not touching cortical GM were filtered based on their prior probability to pertain to WM and their distance to surrounding WM. If half of the voxels of a region had a prior probability to pertain to WM and the region was connected to actual WM, the region was also added to the WM class, and those voxels were refilled as normal-appearing WM into the original T1-w image  $T$  using the same implementation proposed in Valverde et al. (2014). Note that classes were visited from  $gmwm$  to  $csfgm$  in order to add new belief of the actual WM and use it to filter next regions.

If the FLAIR modality is not available,  $H^{\text{FLAIR}}$  is automatically set to zero, disabling the evaluation of  $R_p^H$  regions and henceforth forcing the method to evaluate the next T1 region  $R_p^T$ . If FLAIR is used, all regions that were discarded on the first part of the algorithm or overlapped the cortex, were filtered according to their spatial attributes on the FLAIR image. Each discarded region  $R_q^T$  on the segmented mask  $SEG$  was matched with a particular region in FLAIR  $R_t^H$  based on their overlap ( $R_t^H \mid t = \arg \max(|R_t^H \cap R_q^T|)$ ). Then, matched regions where half of their surrounding neighbors were actually classified as  $gmwm$  or  $wm$  were also added to the WM class and T1-w filled. In all cases, we referred to the neighboring voxels of a region  $N_S$  as the neighbors with one voxel of distance from the region boundaries.

### 2.5. Partial volume maps

Once WM outliers were reassigned to T1-w, the resulting refilled image was used to estimate the brain tissue following the same method described in Sec. 2.3. Afterwards, partial volume maps ( $csfgm$ ) and ( $gmwm$ ) were reassigned to each of the three main classes CSF, GM and WM following a region-wise approach.

Local 2D regions with similar intensity that were classified as  $csfgm$  and  $gmwm$  were estimated using the same connected component algorithm described before. The structural brain atlas  $S$  was then used to reassign regions

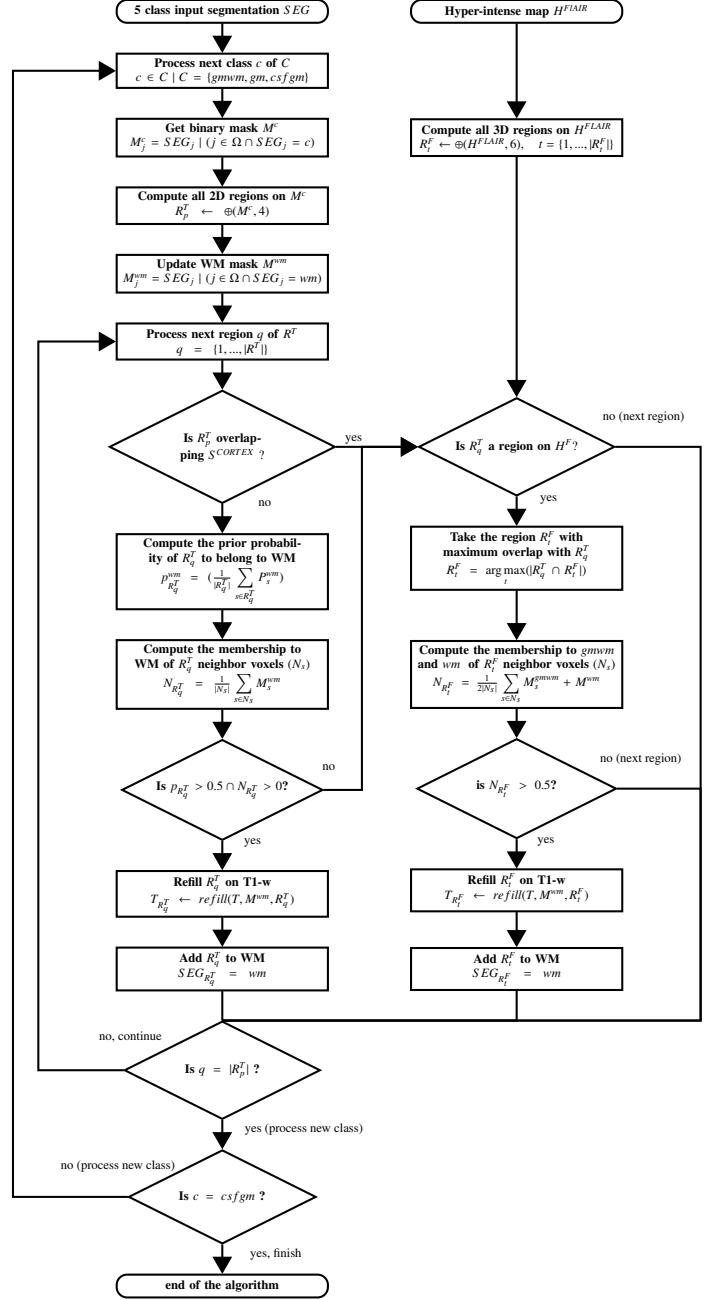


Figure 2: Proposed algorithm to estimate and refill outlier candidate regions into T1-w. The algorithm takes the 5 a class T1-w segmentation and the hyper-intensity map  $H^{\text{FLAIR}}$  if available as inputs. Connected regions of voxels with similar intensity are filtered based on their spatial location probability on tissue and morphological prior atlases. Selected regions are then refilled on the original T1-w image.

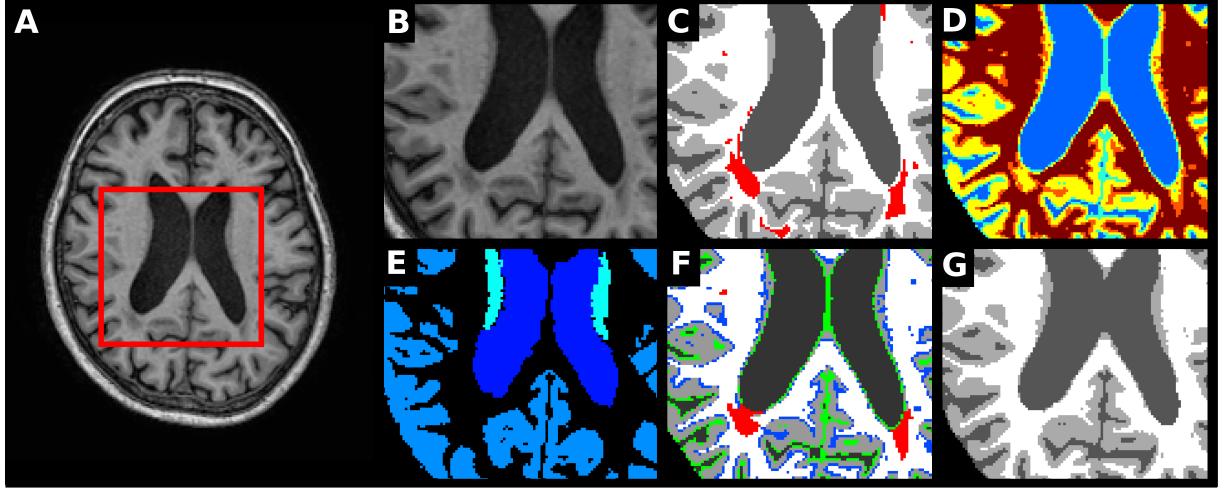


Figure 3: Partial volume assignment during tissue segmentation (Sec. 2.5). A) Original T1-w image. B) Detailed view of the T1-w image. C) Tissue ground-truth with WM lesions highlighted in red. D) Initial 5 class tissue segmentation where *csf*, *csfgm*, *gm*, *gmwm* and *wm* tissues are depicted in blue, cyan, yellow, orange and red, respectively (see Sec. 2.3). E) Morphological brain tissue atlas with parcellated GM regions and ventricles. F) Partial volume *csfgm* and *gmwm* regions (depicted in green and blue, respectively). Previously estimated lesion candidates are depicted in red (see Sec. 2.4). 2D regions with half of their voxels overlapping with the morphological atlas are reassigned to the correspondent tissue. The rest of voxels are reassigned to the neighboring pure class with the most similar signal intensity. G) Final tissue segmentation with partial tissue volume maps re-estimated as CSF (dark gray), GM (light gray) and WM (white).

that at least half of their voxels overlapped with certain structures as follows:

$$SEG_{R_p} = \begin{cases} CSF & \text{if } \left( \frac{1}{|R_p|} \sum_{s \in R_p} S_s^{VENT} \right) > 0.5 \\ GM & \text{if } \left( \frac{1}{2|R_p|} \sum_{s \in R_p} S_s^{Cortex} + S_s^{BASAL} \right) > 0.5 \\ WM & \text{if } \left( \frac{1}{|R_p|} \sum_{s \in R_p} S_s^{BRAINSTEM} \right) > 0.5 \end{cases} \quad (5)$$

for all the regions  $p = \{1, \dots, |R_p|\}$ . The rest of voxels not reassigned previously were reclassified by adding them to the surrounding pure class with the most similar intensity as follows:

$$SEG_j = \arg \min_c \left| T_j - \frac{1}{|N_j|} \sum_{s=1}^{|N_j|} (T_s \mid SEG_s = c) \right| \quad (6)$$

for pure classes  $c = \{csf, gm, wm\}$  and partial volume voxels  $j = \{\forall j \in \Omega \mid SEG_j = csfgm \cup gmwm\}$ . The radius for neighbor voxels was set to 6 in two dimensions. Figure 3 depicts the partial volume re-assignment process for one particular T1-w image.

### 3. Experiments

#### 3.1. MRBrainS database

##### 3.1.1. Data

The public available MRBrainS 2013 database<sup>5</sup> consisted of 20 scans with varying degree of brain atrophy and white matter lesions. These scans were acquired on a 3.0T Philips Achieva MR scanner at the University Medical Center Utrecht (The Netherlands) with the following sequences: 3D T1-w (TR: 7.9ms, TE: 4.5ms), T1-Inverse Recovery (TR: 4416ms, TE: 15ms, and TI: 400ms), and T2-weighted/FLAIR (TR: 11000ms, TE: 125ms, and TI: 2800 ms). Each of the scans was co-registered (Klein et al., 2010) and intensity-corrected (Ashburner and Friston, 2005) before releasing the data. T1, T1-IR, and T2/FLAIR voxel size was  $(0.96 \times 0.96 \times 3.00 \text{ mm}^3)$  after registration (Vincken et al., 2015).

Three experts manually delineated each of the 20 scans into CSF, GM and WM and these annotations were used as the reference standard for the evaluation framework (Vincken et al., 2015). Extended manual annotation containing various brain structures and white matter lesions for 5 scans were provided for training while the remaining 15 scans were blind and had to be skull-stripped and segmented into CSF, GM and WM by participating teams.

<sup>5</sup>available for download at: <http://mrbrains13.isi.uu.nl/>

### 3.1.2. Evaluation:

Segmentation results had to be submitted online for external evaluation based on the following scores for the CSF, GM and WM tissues ( $c$ ):

- Dice similarity coefficient ( $DSC_c$ ) (Dice, 1945) between the manual tissue segmentation ( $GT_c$ ) and the computed segmentation ( $SEG_c$ ) masks:

$$DSC_c = \frac{2 |SEG_c \cap GT_c|}{|SEG_c| + |GT_c|} \times 100 \quad (7)$$

- The modified Hausdorff distance (95th percentile) (Huttenlocher et al., 1993) between the manual tissue segmentation ( $GT_c$ ) points  $p'$  and the computed segmentation points  $p$  in ( $SEG_c$ ) masks:

$$h_c^{95} = \max_{p \in SEG_c} \min_{p' \in GT_c} |p - p'| \quad (8)$$

- The absolute difference in tissue volume ( $AVD_c$ ) between manual tissue segmentation ( $GT_c$ ) and the computed segmentation ( $SEG_c$ ) masks:

$$AVD_c = \left\| \frac{|SEG_c| - |GT_c|}{|GT_c|} \right\| \quad (9)$$

In order to evaluate the performance of our method, we submitted two different segmentation sets using either only the T1-w sequences or both T1-w and FLAIR images. We validated the performance of our approach comparing the obtained scores with other submitted segmentation pipelines.

### 3.1.3. Parameter settings

Skull stripping of input images was performed using a similar approach to other methods participating in the challenge (Jog et al., 2013; Opbroek et al., 2013; Rajchl et al., 2015). The 5 training images were non-rigidly registered to the image space of each of the T1-w (Modat et al., 2010), and the brainmask was generated by simple voting of the registered masks. Afterwards, each mask was refined on the T1-IR image by thresholding hyperintense voxels.

All the parameters of our tissue segmentation method were set to default values ( $q = 2, \gamma = 0.1, w = 1$ ). The  $\beta$  parameter was computed automatically as described in Section 2.3. The  $\alpha$  parameter that scaled the minimum signal intensity on the  $H^F$  mask was set to  $\alpha = 3$ .

## 3.2. MS database

### 3.2.1. Data

This non-public database of images was composed by 24 images of clinically isolated syndrome (CIS) patients acquired on a 3T Siemens MR scanner (Trio Tim, Siemens, Germany) with a 12-channel phased-array head coil (data from Hospital Vall D’Hebron, Barcelona, Spain). The following pulse sequences were obtained: 1) transverse proton density and T2-weighted fast spin-echo (TR=2500 ms, TE=16-91 ms, voxel size=0.78×0.78×3mm<sup>3</sup>); 2) transverse fast T2-FLAIR (TR=9000ms, TE=93ms, TI=2500ms, flip angle=120°, voxel size=0.49×0.49×3mm<sup>3</sup>); and 3) sagittal 3D T1 magnetization prepared rapid gradient-echo (MPRAGE) (TR=2300 ms, TE=2 ms; flip angle=9°; voxel size=1×1×1.2mm<sup>3</sup>). For each scan, T1-w and FLAIR images were first skull-stripped using BET (Smith, 2002) and then intensity-corrected using the N3 method (Sled et al., 1998). Finally, FLAIR images were co-registered into the T1-w space and then re-aligned into the MNI space using SPM12 co-registration tools with the normalized mutual information as objective function and tri-linear interpolation with no wrapping (Ashburner and Friston, 2005). White matter lesion masks were semi-automatically delineated from FLAIR using JIM software<sup>6</sup> by an expert radiologist of the hospital center. Mean lesion volume was 4.30 ± 4.84 ml (range 0.1-18.3 ml).

### 3.2.2. Evaluation

Manual expert annotations of tissues were not available for this database. As validated in previous MS studies (Battaglini et al., 2012; Valverde et al., 2015b,c), WM lesions on original T1-w scans were first refilled with signal intensities similar to normal-appearing WM using the SLF lesion filling method (Valverde et al., 2014). Then, both the original and the refilled images were segmented into CSF, GM and WM tissues using our proposed approach. The performance of our tissue segmentation method was evaluated by computing the absolute difference in tissue volume ( $AVD_c$ ) between the images segmented containing lesions and the same images where WM lesions were refilled before tissue segmentation:

---

<sup>6</sup>Xinapse Systems, <http://www.xinapse.com/home.php>

$$AVD_c = \left\| \frac{|SEG_c| - |GT_c^{fill}|}{|GT_c^{fill}|} \right\| \times 100 \quad (10)$$

where  $SEG_c$  refers to the output segmentation masks of the images segmented containing lesions, and  $GT_c^{fill}$  refers to the output tissue segmentation masks of images where lesions were filled before segmentation and considered as ground-truth.

Several works (DellOglio et al., 2014; Valverde et al., 2015c) have already shown that part of the actual error in tissue segmentation may be partially masked by opposite directions in the differences in total and normal-appearing tissue. In order to add an additional measure estimator of the actual error in tissue segmentation that may not bias by these differences, we also compared for each tissue the percentage of mis-classified voxels  $PMC_c$  between the original  $SEG_c$  and the expert filled  $GT_c^{fill}$  masks:

$$PMC_c = \frac{|\overline{SEG_c} \cap GT_c^{fill}|}{|GT_c^{fill}|} \times 100 \quad (11)$$

In order to analyze the benefits of using the FLAIR image into the proposed approach, we evaluated the performance of our method when using only the T1-w image and when using both T1-w and FLAIR. Furthermore, we validated it with two other automated pipelines widely used in brain tissue segmentation, such as FAST (Zhang et al., 2001) (version FSL 5.0) and SPM (Ashburner and Friston, 2005) (version SPM12 rev 6225), using either original images or after estimating lesions using the automated approach SLS proposed by Roura et al. (2015). On images where lesions were automatically segmented, estimated lesion masks were then filled with the same SLF method (Valverde et al., 2014) before tissue segmentation. Similarly to our approach, we considered the tissue segmentation masks of the expert refilled T1-w images segmented with FAST and SPM12 as the ground-truth for each method. Table 1 summarizes each of the evaluated pipelines and the corresponding process followed to segment the MS images.

### 3.2.3. Parameter settings

The BET skull-stripped process was optimized as proposed by Popescu et al. (2012) without removing CSF

Table 1: Summary of evaluated pipelines and processes used on MS data. On pipelines *FAST only T1* and *SPM12 only T1* images were segmented containing lesions without prior automated lesion segmentation. On pipelines *FAST + SLS* and *SPM12 + SLS*, WM lesions were automatically segmented using the SLS approach (Roura et al., 2015) and estimated lesion masks were afterwards lesion filled using the SLF method (Valverde et al., 2014). On our proposed pipeline using *MSSEG only T1* and *MSSEG T1 + FLAIR* lesion segmentation and filling was part of the same segmentation method. Manual lesion annotations were used to refill T1-w images on pipelines *FAST GT*, *SPM12 GT* and *MSSEG GT* before segmenting the images using *FAST*, *SPM12* and *MSSEG*, respectively.  $AVD_c$  and  $PMC_c$  scores were then computed between pipelines 1 vs 3, 2 vs 3, 4 vs 6, 5 vs 6, 7 vs 9 and 8 vs 9.

Pipeline	Modality	Lesion seg.	Lesion filling	Tissue seg.
1. FAST only T1	T1	none	none	FAST
2. FAST + SLS	T1, FLAIR	SLS (FLAIR)	SLF	FAST
3. FAST GT	T1	expert manual	SLF	FAST
4. SPM12 only T1	T1	none	none	SPM12
5. SPM12 + SLS	T1, FLAIR	SLS (FLAIR)	SLF	SPM12
6. SPM12 GT	T1	expert manual	SLF	SPM12
7. MSSEG only T1	T1	internal	internal	MSSEG
8. MSSEG T1 + FLAIR	T1, FLAIR	internal	internal	MSSEG
9. MSSEG GT	T1	expert manual	SLF	MSSEG

from the brainmask. N3 was run with optimized parameters by reducing the smoothing distance parameter to 30–50 mm (Boyes et al., 2008; Zheng et al., 2009).

The SLF lesion filling method was run with default parameters in all experiments. In the FAST and SPM12 images where we estimated lesion masks automatically, the lesion segmentation method SLS was optimized for 3.0T data identically as shown in Roura et al. (2015).

All the parameters of our proposed method were fixed to default values ( $q = 2, \gamma = 0.1, w = 1$ ) as done in MR-BrainS13 database. The  $\beta$  parameter was computed automatically. The  $\alpha$  parameter that scaled the minimum signal intensity on the  $H^F$  mask was set again to  $\alpha = 3$ .

### 3.3. Statistical significance

The statistical significance of the performance between methods was computed by running a series of permutation tests (Menke and Martinez, 2004; Klein et al., 2009; Diez et al., 2014) between the differences in the scores obtained by each method. These tests permitted to analyze the fraction of times that a particular method with the lowest score was significantly better than the rest of methods with  $p\text{-value} \leq 0.05$ . Methods were then ranked in three different levels according to the difference between the mean score of the best method  $\mu_o \pm \sigma_o$  and the distance with respect to the mean scores of the rest of methods. Hence, Rank 1 contained methods with mean scores  $(\mu_o - \sigma_o, \mu_o]$ , Rank 2

contained those with mean scores ( $\mu_o - 2\sigma_o$ ,  $\mu_o - \sigma_o$ ) and Rank 3 those in the interval ( $\mu_o - 3\sigma_o$ ,  $\mu_o - 2\sigma_o$ ) (Klein et al., 2009; Diez et al., 2014; Valverde et al., 2015a). For all the run tests, we set the number of comparisons between each pair of methods to  $N = 1000$ .

### 3.4. Implementation details

The proposed pipeline was entirely developed in MATLAB (v2014a, The Mathworks Inc, US), except for the registration process that was run using the available NiftyReg package (Ourselin et al., 2002; Modat et al., 2010). The method was configured to run either in CPU or GPU. Experiments were carried out on a GNU/Linux machine with a single Intel core i7 processor at 3.4 Ghz (Intel Corp, US), and a NVIDIA K40 with 12GB of RAM (NVIDIA, US). The average execution time for the proposed method including registration and tissue segmentation was 8 minutes running on the CPU core. Execution time on the GPU was approximately 2 minutes, reducing four times the execution time on the CPU processor.

## 4. Results

### 4.1. MRBrainS13 dataset

Table 2 shows the obtained mean  $DSC_c$ ,  $h_c^{95}$  and  $AVD_c$  scores for our proposed method. We compare the obtained scores with other non-supervised strategies that also participated in the challenge such as FAST, SPM12, or VBM12<sup>7</sup>, and also with respect to the best ranked method proposed by Stollenga and Byeon (2015). The overall rank of methods also included the combined brain (GM + WM) and intracranial (CSF + GM + WM) volumes, which are not shown in the table for simplicity<sup>8</sup>. At the time of submitting our results on the online application, our proposed approach using both T1 and FLAIR (*MSSEG T1+FLAIR*) was the best unsupervised method of the challenge (7th position overall 31 participants), and its accuracy was also very competitive in comparison with several supervised methods that were explicitly trained for the challenge. When only using the T1-w modality

(*MSSEG only T1*), the method was ranked in the 10th position, but still clearly over-performed *FAST* (21th position), and *SPM12* using FLAIR+T1 (17th position), the T1-IR modality (18th position), or the T1-w modality (20th position).

Figure 4 illustrates the different steps performed by our approach. After registering the probabilistic atlases into the subject space (Fig.4 panels E to H), tissue was estimated from the T1-w into 5 different classes (Fig.4 panel I). Then, WM outliers were estimated on the T1-w image by analyzing all the regions not initially segmented as WM with a high probability to pertain to WM based on spatial local probability and prior tissue information (red regions depicted in Fig.4 panel J). If FLAIR was also provided, lesion candidates were also analyzed based on their signal intensity on the FLAIR image and their spatial local probability to pertain to WM (green regions depicted in Fig.4 panel J). Afterwards, lesion candidate regions were refilled into the T1-w image with signal intensities similar to the WM, the refilled T1-w image was re-estimated again and CSFGM and GMWM volumes were reassigned into the three main classes (Fig.4 panels K and L).

### 4.2. MS data

Table 3 (a) depicts the mean % of absolute differences in CSF, GM and WM volume ( $AVD_c$ ) for each of the evaluated methods after segmenting the 24 3T images. Table 3 (b) shows the mean % of mis-classified CSF, GM and WM voxels ( $PMC_c$ ) for each of the evaluated methods.

Methods were ranked based on their scores after running the permutation tests. Table 4 shows the rank of each evaluated method for the  $AVD_c$  (Table 4 (a)) and  $PMC_c$  scores (Table 4 (b)), respectively. The differences in tissue volume were the lowest when methods used also the FLAIR image to estimate the WM outliers. Our proposed approach reported competitive results and again was ranked in the first rank of methods for all tissues. Methods using only T1 yielded a significant higher difference in GM and WM volume and were ranked in the second and third group. The proposed approach using only T1-w was significantly better than the rest of methods in terms of the % of miss-classified GM and WM voxels, but its performance was worst for WM and was ranked in the second group of methods for WM. As shown in Table 4 (a), the % of miss-classified WM voxels was significantly lower in *FAST+SLS* and our proposed method

<sup>7</sup><http://www.neuro.uni-jena.de/>

<sup>8</sup>Overall ranking of methods for all the measurements can be consulted at <http://mrbrains13.isi.uu.nl/results.php>

Table 2: Segmentation results on the 15 test images of the MRBrainS challenge. Mean  $DS C_c$ ,  $h_c^{95}$  and  $AVD_c$  scores for CSF, GM and WM tissue are shown for our proposed method when using only the T1-w modality (*MSSEG only T1*) and when using the T1-w and FLAIR modalities (*MSSEG T1+FLAIR*). The obtained values are compared with the best approach at the time of writing this paper (Stollenga and Byeon, 2015), and also with other unsupervised techniques that also participated in the challenge such as *FAST*, *SPM12* and *VBM12*. The overall ranking of the methods in the challenge is shown in the last column.

Method	$DS C_c$			$h_c^{95}$			$AVD_c$			<b>Rank</b>
	CSF	GM	WM	CSF	GM	WM	CSF	GM	WM	
<b>Best method</b>	$83.72 \pm 2.63$	$84.82 \pm 1.37$	$88.33 \pm 0.89$	$2.14 \pm 0.36$	$1.70 \pm 0.01$	$2.08 \pm 0.33$	$7.09 \pm 4.01$	$6.77 \pm 3.28$	$7.05 \pm 5.22$	1
<b>FAST only T1</b>	$69.95 \pm 2.81$	$78.66 \pm 2.24$	$85.98 \pm 2.58$	$3.41 \pm 0.25$	$4.35 \pm 1.13$	$3.65 \pm 0.85$	$11.83 \pm 10.38$	$8.65 \pm 6.34$	$11.47 \pm 6.24$	21
<b>SPM12 only T1</b>	$70.69 \pm 3.75$	$80.34 \pm 2.37$	$85.58 \pm 1.73$	$5.34 \pm 1.47$	$2.93 \pm 0.25$	$3.06 \pm 0.08$	$23.24 \pm 16.04$	$6.95 \pm 6.57$	$5.99 \pm 3.95$	20
<b>SPM12 T1+FLAIR</b>	$74.03 \pm 3.42$	$81.17 \pm 2.24$	$86.03 \pm 1.48$	$4.59 \pm 0.61$	$2.90 \pm 0.15$	$3.00 \pm 0.07$	$10.07 \pm 4.86$	$10.59 \pm 8.34$	$5.21 \pm 3.88$	17
<b>SPM12 T1-iR</b>	$78.25 \pm 3.78$	$79.41 \pm 2.15$	$83.54 \pm 2.14$	$4.01 \pm 0.63$	$3.01 \pm 0.29$	$3.60 \pm 0.27$	$10.47 \pm 5.69$	$7.23 \pm 6.50$	$6.34 \pm 4.61$	18
<b>VBM12</b>	$74.56 \pm 2.70$	$82.29 \pm 1.49$	$87.95 \pm 1.71$	$3.03 \pm 0.14$	$3.20 \pm 0.32$	$2.32 \pm 0.42$	$6.80 \pm 4.57$	$5.91 \pm 3.91$	$6.06 \pm 4.42$	8
<b>MSSEG only T1</b>	$80.18 \pm 2.67$	$82.06 \pm 1.68$	$87.05 \pm 1.46$	$2.81 \pm 0.21$	$3.33 \pm 0.21$	$2.91 \pm 0.42$	$7.18 \pm 3.33$	$6.15 \pm 3.51$	$6.20 \pm 5.45$	10
<b>MSSEG T1+FLAIR</b>	$80.16 \pm 2.67$	$82.20 \pm 1.60$	$87.33 \pm 1.35$	$2.81 \pm 0.21$	$3.18 \pm 0.15$	$2.88 \pm 0.39$	$7.21 \pm 3.31$	$5.99 \pm 3.43$	$5.95 \pm 5.44$	7

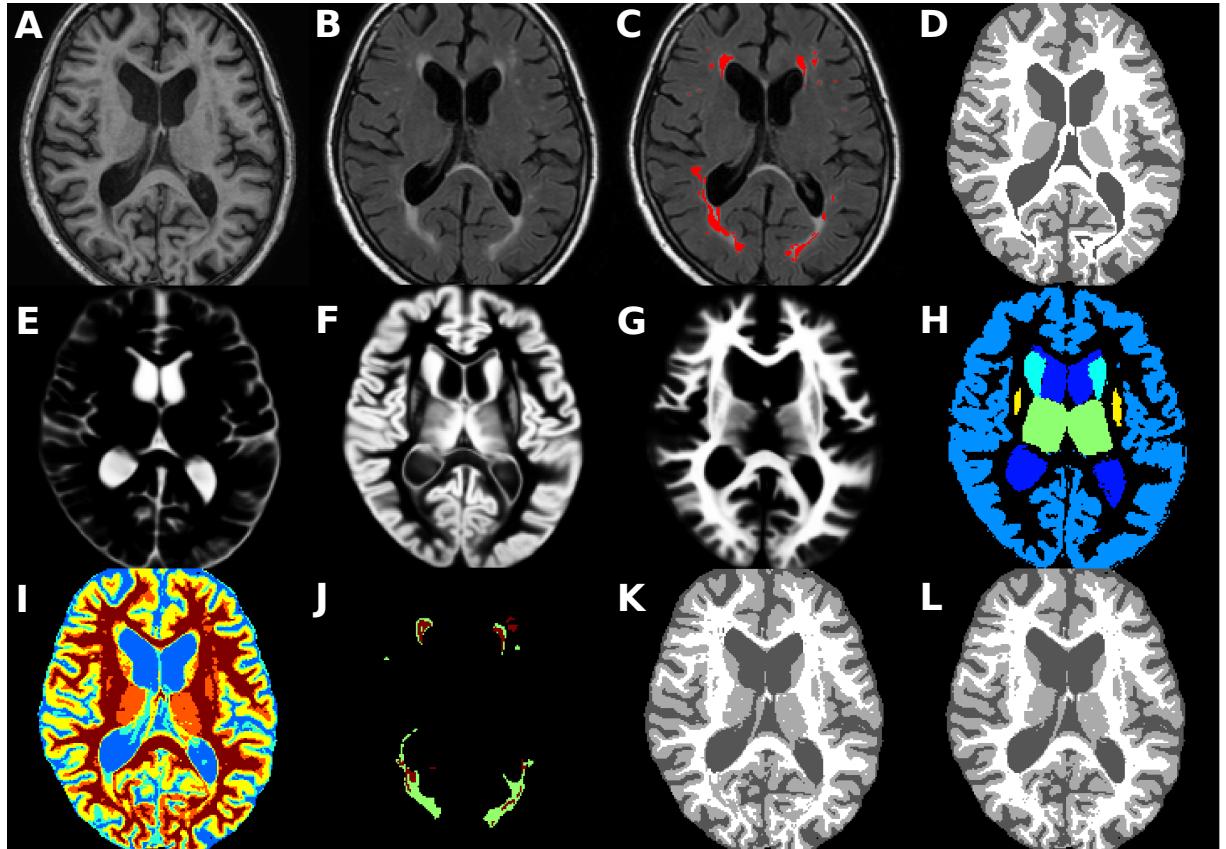


Figure 4: Automated tissue segmentation of the *MSSEG* method on the *second subject* of the training set of the MRBrainS13 database. A) Original T1-w image. B) Original FLAIR image. C) FLAIR image with manual annotated WM lesions depicted in red. D) Provided ground-truth for training purposes. Registered CSF, GM and WM prior atlas to the subject space (E, F and G, respectively). H) Morphological brain structural atlas registered to the subject space. I) First partial volume segmentation with *csf* depicted in blue, *csfgm* in cyan, *gm* in yellow, *gmwm* in orange and *wm* in red. J) Obtained WM outliers extracted from either T1-w (depicted in red) and FLAIR (depicted in green). K) Final tissue segmentation using only the T1-w image, with *CSF* in dark gray, *GM* in light gray and *WM* in white. L) Final tissue segmentation when using both T1 and the FLAIR images.

Table 3: Mean % of absolute difference in CSF, GM and WM volume between the 24 3T tissue masks where expert annotations were refilled before segmentation and the same images segmented including white matter lesions. For each method, reported values are the mean and standard deviation  $\mu \pm \sigma$  for the (a)  $AVD_c$  and (b)  $PMC_c$  scores obtained along the entire database.

(a) Differences in  $AVD_c$

Method	Dif CSF (%)	Dif GM (%)	Dif WM (%)
<b>FAST only T1</b>	$0.07 \pm 0.13$	$0.33 \pm 0.45$	$0.42 \pm 0.56$
<b>FAST + SLS</b>	$0.04 \pm 0.07$	$0.08 \pm 0.12$	$0.11 \pm 0.16$
<b>SPM12 only T1</b>	$0.31 \pm 0.46$	$0.27 \pm 0.45$	$0.56 \pm 0.69$
<b>SPM12 + SLS</b>	$0.22 \pm 0.22$	$0.13 \pm 0.23$	$0.20 \pm 0.32$
<b>MSSEG only T1</b>	$0.13 \pm 0.20$	$0.21 \pm 0.26$	$0.42 \pm 0.54$
<b>MSSEG T1+FLAIR</b>	$0.04 \pm 0.05$	$0.06 \pm 0.05$	$0.13 \pm 0.11$

(b) Differences in  $PMC_c$

Method	CSF (%)	GM (%)	WM (%)
<b>FAST only T1</b>	$0.08 \pm 0.11$	$0.09 \pm 0.12$	$0.53 \pm 0.69$
<b>FAST + SLS</b>	$0.06 \pm 0.06$	$0.14 \pm 0.16$	$0.25 \pm 0.30$
<b>SPM12 only T1</b>	$0.16 \pm 0.32$	$0.25 \pm 0.33$	$0.73 \pm 0.86$
<b>SPM12 + SLS</b>	$0.22 \pm 0.31$	$0.24 \pm 0.29$	$0.41 \pm 0.43$
<b>MSSEG only T1</b>	$0.02 \pm 0.03$	$0.03 \pm 0.05$	$0.46 \pm 0.58$
<b>MSSEG T1+FLAIR</b>	$0.04 \pm 0.04$	$0.14 \pm 0.13$	$0.27 \pm 0.29$

*MSSEG T1+FLAIR* when compared with the rest of evaluated pipelines.

#### 4.2.1. WM outlier rejection

Finally, we evaluated the performance of the proposed WM outlier rejection algorithm with respect to the rest of pipelines. A traditional comparison respect of the number of true-positive and false-positive WM lesion voxels of each pipeline is not appropriate here, given that our approach only processed WM lesion candidates that were not initially classified as WM. In contrast, the error in the expected WM lesion volume segmented can be a good indicator of the performance of each method segmenting WM lesions as WM.<sup>9</sup> Figure 5 shows the % of absolute difference in WM lesion volume between each of the evaluated pipelines and their correspondent  $GT_c^{fill}$  images. As expected, methods yielded the lowest differences in WM lesion volume when used also the FLAIR modality to estimate WM lesions. *MSSEG+FLAIR* showed the lowest differences in WM lesion volume of all evaluated methods.

<sup>9</sup>Aquest punt s'ha de revisar. cal parlar-ne.

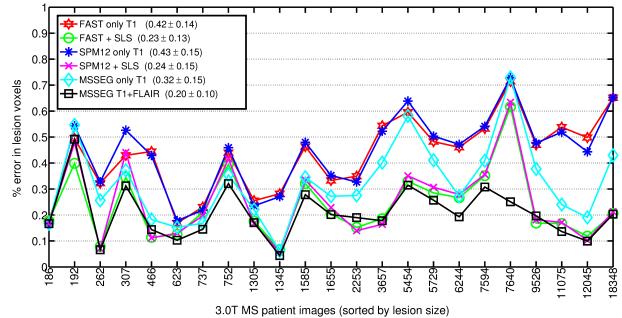


Figure 5: % of absolute difference in WM lesion volume for each of the evaluated pipelines on the 3T MS database. Figure legends also depict mean and standard deviation  $\mu \pm \sigma$  for the entire set of images. Images are sorted by lesion size (number of lesion voxels).

## 5. Discussion

In this paper we have presented a new automated brain tissue segmentation pipeline for MS patient images that combines multi-channel intensities, anatomical and morphological prior maps at different levels to estimate brain tissue in the presence of WM lesions. The current method integrates a WM outlier estimation and refilling algorithm which is applied intermediately in order to reduce the effect of WM lesions on tissue segmentation. As shown by the presented results, the proposed technique yields competitive and consistent results in both general and MS specific databases without parameter tweaking. Furthermore, although we did not analyze explicitly the execution times of each of the evaluated algorithms, the proposed method takes advantage of new affordable processors such as GPUs that reduce up to four times the execution time to register and segment tissue when compared to general purpose CPUs.

The MRBrainS challenge permitted to evaluate the efficacy of our method and to validate it with other state-of-the-art tissue segmentation methods. Although the challenge was not focused on the MS disease, methods were evaluated by comparing them with respect manual expert annotations of tissues and WM lesions, which provided a quantitative measure of the accuracy of the method. The overall results showed that supervised methods obtained the best results of the challenge, taking advantage of the inherent capability to fit the database characteristics. At

Table 4: Permutation tests results for evaluated methods on the 3T MS database. (a) Final rank based on the absolute % difference in CSF, GM and WM volume between methods. (b) Final rank based on % of miss-classified CSF, GM and WM voxels between methods. Reported values are mean and standard deviation ( $\mu_o, \sigma_o$ ) of the fraction of times when each method produces significant p-values ( $p \leq 0.05$ ). Positive values indicate that in average, the method out-performed the other methods in pair-wise significant tests. Negative values indicate the contrary. Rank 1:  $(\mu_o - \sigma_o, \mu_o]$ , Rank 2:  $(\mu_o - 2\sigma_o, \mu_o - \sigma_o]$ , Rank 3  $(\mu_o - 3\sigma_o, \mu_o - 2\sigma_o]$ . All permutation tests were run with 1000 random iterations.

(a) Evaluated methods ranked by the absolute % of CSF, GM and WM volume of 3T data.						
Rank	Method (CSF)	$\mu \pm \sigma$	Method (GM)	$\mu \pm \sigma$	Method (WM)	$\mu \pm \sigma$
<b>Rank 1</b>	<b>MSSEG T1+FLAIR</b>	0.5 ± 0.55	<b>MSSEG T1+FLAIR</b>	0.5 ± 0.55	FAST + SLS	0.67 ± 0.52
	FAST only T1	0.5 ± 0.55	FAST + SLS	0.5 ± 0.55	<b>MSSEG T1+FLAIR</b>	0.5 ± 0.55
	FAST + SLS	0.5 ± 0.55	SPM12 + SLS	0.33 ± 0.52		
<b>Rank 2</b>	<b>MSSEG only T1</b>	-0.43 ± 0.64	<b>MSSEG only T1</b>	-0.17 ± 0.75	SPM12 + SLS	0.14 ± 0.72
	SPM12 + SLS	-0.5 ± 0.55	SPM12 only T1	-0.5 ± 0.55	FAST only T1	-0.24 ± 0.62
	SPM12 only T1	-0.57 ± 0.5				
<b>Rank 3</b>		FAST only T1		-0.67 ± 0.52	<b>MSSEG only T1</b>	-0.47 ± 0.52
					SPM12 only T1	-0.59 ± 0.49

(b) Evaluated methods ranked by the absolute % of miss-classified CSF, GM and WM of 3T data.						
Rank	Method (CSF)	$\mu \pm \sigma$	Method (GM)	$\mu \pm \sigma$	Method (WM)	$\mu \pm \sigma$
<b>Rank 1</b>	<b>MSSEG only T1</b>	0.83 ± 0.41	<b>MSSEG only T1</b>	0.83 ± 0.41	<b>MSSEG T1+FLAIR</b>	0.67 ± 0.52
	<b>MSSEG T1+FLAIR</b>	0.44 ± 0.8	FAST only T1	0.5 ± 0.84	FAST + SLS	0.67 ± 0.52
<b>Rank 2</b>					SPM12 + SLS	-0.17 ± 0.75
<b>Rank 3</b>					<b>MSSEG only T1</b>	-0.17 ± 0.75
					FAST only T1	-0.17 ± 0.75
	FAST + SLS	0 ± 0.89	FAST + SLS	-0.17 ± 0.75	SPM12 only T1	-0.83 ± 0.41
	SPM12 only T1	-0.27 ± 0.43	<b>MSSEG T1+FLAIR</b>	-0.17 ± 0.75		
	FAST only T1	-0.33 ± 0.82	SPM12 only T1	-0.33 ± 0.52		
	SPM12 + SLS	-0.67 ± 0.52	SPM12 + SLS	-0.66 ± 0.51		

the time of writing this paper, *MSSEG T1+FLAIR* was ranked in the 7th position out of 31 participants, being the best non-supervised strategy followed by the VBM12 approach. As shown by the differences at each of the obtained scores, the FLAIR modality appeared to be useful to improve the accuracy of the method when compared with *MSSEG T1*, which was ranked 10th. The performance of *MSSEG* was superior in all tissues when compared to general purpose methods such as *FAST* (ranked 21th) and *SPM12* (best ranked 17th), even if those used both image modalities. However, final ranking of methods should be taken with care, given the differences in the skull-stripping processes between methods. Differences in the boundaries of the estimated skull masks may be behind the remarkable differences in CSF between methods, altering also the intra-cranial cavity measurements and consequently the overall score of each of the methods.

In MS data, the performance of our method was similar or better to the best pipeline incorporating an state-of-the-art method for lesion segmentation and filling, validating

the overall capability of the proposed method to reduce the effects of WM lesions on tissue segmentation. *MSSEG T1+FLAIR* and *FAST+SLS* were ranked in the first group of methods with error differences in tissue volume below 0.15% in all tissues. Pipelines using only T1-w showed a similar or lower % of miss-classified CSF and GM voxels than those using both FLAIR and T1-w. In contrast, the % of miss-classified WM voxels (Table 4 (a)) and the differences in reassigned lesion volume (Fig. 5) were significantly higher on the former, showing that these methods tended to overestimate GM and underestimate WM by the effect of WM lesions. In this aspect, our results are consistent with previous studies also analyzing the effects of WM lesions on tissue segmentation (Battaglini et al., 2012; Gelineau-Morel et al., 2012; Valverde et al., 2015b,c).

Differences in the  $AVD_c$  between *MSSEG T1+FLAIR* and *MSSEG only T1* on the MRBrains13 data were similar than those reported in MS data, showing that in general the inclusion of the FLAIR modality reduced the overall error in tissue volume on all the analyzed databases. On

MS data, the % of miss-classified CSF and GM voxels was significantly lower on the *MSSEG only T1*, but significantly higher in WM, evidencing that *MSSEG only T1* tended to overestimate WM, while the error in *MSSEG T1+FLAIR* was similar in both GM and WM. In addition, the presented results show that the % difference in total WM and lesion volume was significantly lower on the *MSSEG T1+FLAIR* in comparison with *MSSEG only T1*. Hence, we would recommend to use both T1-w and FLAIR modalities when possible. However, the accuracy of the *MSSEG only T1* pipeline was still superior than *FAST* and *SPM12* when compared with ground-truth annotations of the MRBrainS13 database. This suggests that at least with the available data, the improvement in tissue segmentation was not only caused by the addition of the FLAIR modality, but also by the combination of intensity, anatomical and morphological priors.

This study however endorses some limitations. The lack of a database consisting of MS images with manual annotations of tissue, limits our analysis to the differences in tissue volume with respect to images where expert lesion annotations were lesion filled before tissue segmentation. However, the previous analysis has been shown in previous studies to be effective to evaluate the effects of WM lesions on tissue segmentation (Battaglini et al., 2012; Valverde et al., 2015b,c). Furthermore, the mean lesion sizes of the MS cohorts do not allow to investigate better the performance of the proposed method in the presence of images with higher lesion load. As a future work, we believe that an additional study on MS with manual tissue annotated masks and higher lesion load would be helpful not only to analyze the benefits of the proposed algorithm in MS images, but also to investigate the benefits of adding other image channels such as T2 or PD. Furthermore, although the method was designed for cross-sectional data, we are sensible to the fact that the current approach may be benefited by the possibility to evaluate longitudinal changes in tissue volume.

## 6. Conclusion

In this paper, we have proposed the Multiple Sclerosis SEGmentation pipeline (*MSSEG*), a new MRI brain tissue segmentation method designed to deal with MS patient images containing lesions. Our proposed approach incorporates robust partial volume tissue segmentation

with outlier rejection and filling, combining intensity, probabilistic and morphological prior maps in a novel-way. When combining T1-w and FLAIR modalities, our method have shown very competitive results on the MR-BrainS13 database, ranked on the 7th position out of 31 participant strategies and being the best non-supervised approach so far. In MS data, differences in tissue volume were lower or similar to the best available pipeline composed of *FAST* and a state-of-the-art method for lesion segmentation and filling. In all the experiments the inclusion of the FLAIR modality into the proposed method reduced the effect of WM lesions on tissue segmentation, which suggests that this modality should be used when available. In conclusion, our results show that at least with the presented data, *MSSEG* improves the measurement of brain tissue volume on images containing WM lesions. Hence, we strongly believe that the neuro-image community can be benefited by its use in future settings.

## Acknowledgements:

S. Valverde holds a FI-DGR2013 grant from the Generalitat de Catalunya. E. Roura holds a BR-UdG2013 grant. This work has been partially supported by "La Fundació la Marató de TV3", by Retos de Investigación TIN2014-55710-R, and by the MPC UdG 2016/022 grant. The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

## References

- Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26:839–851.
- Battaglini, M., Jenkinson, M., Stefano, N. D., and De Stefano, N. (2012). Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Human Brain Mapping*, 33(9):2062–71.
- Boyes, R. G., Gunter, J. L., Frost, C., Janke, A. L., Yeatman, T., Hill, D. L. G., Bernstein, M. A., Thompson, P. M., Weiner, M. W., Schuff, N., Alexander, G. E., Killiany, R. J., DeCarli, C., Jack, C. R., and Fox, N. C. (2008). Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *NeuroImage*, 39:1752–1762.

- Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., and Bach Cuadra, M. (2011). A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158–e177.
- Ceccarelli, A., Jackson, J. S., Tauhid, S., Arora, A., Gorky, J., Dell’Oglio, E., Bakshi, A., Chitnis, T., Khouri, S. J., Weiner, H. L., Guttmann, C. R. G., Bakshi, R., and Neema, M. (2012). The impact of lesion inpainting and registration methods on voxel-based morphometry in detecting regional cerebral gray matter atrophy in multiple sclerosis. *American Journal of Neuroradiology*, 33(8):1579–85.
- Chard, D. T., Jackson, J. S., Miller, D. H., and Wheeler-Kingshott, C. A. M. (2010). Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *Journal of Magnetic Resonance Imaging*, 32:223–228.
- Datta, S. and Narayana, P. a. (2013). A comprehensive approach to the segmentation of multichannel three-dimensional MR brain images in multiple sclerosis. *NeuroImage: Clinical*, 2:184–96.
- De Boer, R., Vrooman, H. a., van der Lijn, F., Vernooij, M. W., Ikram, M. A., van der Lugt, A., Breteler, M. M. B., and Niessen, W. J. (2009). White matter lesion extension to automatic brain tissue segmentation on MRI. *NeuroImage*, 45(4):1151–61.
- de Bresser, J., Portegies, M. P., Leemans, A., Biessels, G. J., Kappelle, L. J., Viergever, M. a., Bresser, J. D., and Jan, G. (2011). A comparison of MR based segmentation methods for measuring brain atrophy progression. *NeuroImage*, 54(2):760–8.
- DellOglio, E., Ceccarelli, A., Glanz, B. I., Healy, B. C., Tauhid, S., Arora, A., Saravanan, N., Bruha, M. J., Vartanian, A. V., Dupuy, S. L., Benedict, R. H., Bakshi, R., and Neema, M. (2014). Quantification of Global Cerebral Atrophy in Multiple Sclerosis from 3T MRI Using SPM: The Role of Misclassification Errors. *Journal of Neuroimaging*, 25(2):191–199.
- Deshpande, H., Maurel, P., and Barillot, C. (2015). Classification of Multiple Sclerosis Lesions using Adaptive Dictionary Learning. *Computerized Medical Imaging and Graphics*.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- Diez, Y., Oliver, A., Cabezas, M., Valverde, S., Martí, R., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., and Lladó, X. (2014). Intensity based methods for brain MRI longitudinal registration. A study on multiple sclerosis patients. *Neuroinformatics*, 12(3):365–379.
- García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L. (2013). Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical Image Analysis*, 17(1):1–18.
- Gelineau-Morel, R., Tomassini, V., Jenkinson, M., Johansen-berg, H., Matthews, P. M., and Palace, J. (2012). The Effect of Hypointense White Matter Lesions on Automated Gray Matter Segmentation in Multiple Sclerosis. *Human Brain Mapping*, 2814(June 2011):2802–2814.
- Geremia, E., Clatz, O., Menze, B. H., Konukoglu, E., Criminisi, A., and Ayache, N. (2011). Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage*, 57(2):378–390.
- Giorgio, A. and De Stefano, N. (2013). Clinical use of brain volumetry. *Journal of Magnetic Resonance Imaging*, 37(1):1–14.
- Guizard, N., Coupé, P., Fonov, V. S., Manjón, J. V., Arnold, D. L., and Collins, D. L. (2015). Rotation-invariant multi-contrast non-local means for MS lesion segmentation. *NeuroImage: Clinical*, 8:376–389.
- Harmouche, R., Subbanna, N. K., Collins, D. L., Arnold, D. L., and Arbel, T. (2015). Based on Modeling Regional Intensity Variability and Local Neighborhood Information. *IEEE Transactions on Biomedical Engineering*, 62(5):1281–1292.

- Huttenlocher, D., Klanderman, G., and Rucklidge, W. (1993). Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863.
- Immerkær, J. (1996). Fast Noise Variance Estimation. *Computer Vision and Image Understanding*, 64(2):300–302.
- Jain, S., Sima, D. M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., De Mey, J., Barkhof, F., Steenwijk, M. D., Daams, M., Maes, F., Van Huffel, S., Vrenken, H., and Smeets, D. (2015). Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage: Clinical*, 8:367–375.
- Jog, A., Roy, S., Prince, J., and Carass, A. (2013). Mr brain segmentation using decision trees. In *Proceedings of the MICCAI Workshops, Challenge on MR Brain Image Segmentation (MRBrainS 13), The Midas Journal*.
- Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., and Lundervold, A. (2009). Evaluation of automated brain MR image segmentation and volumetry methods. *Human brain mapping*, 30(4):1310–27.
- Klein, A., Ardekani, B., Andersson, J., Ashburner, J., Avants, B., Chiang, M., Christensen, G., Collins, D., Gee, J., Hellier, P., and Others (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802.
- Klein, S., Staring, M., Murphy, K., Viergever, M. a., and Pluim, J. P. W. (2010). Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205.
- Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J. C., Quiles, A., Valls, L., Ramió-Torrentà, L., and Rovira, A. (2012). Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Information Sciences*, 186(1):164–185.
- Magon, S., Gaetano, L., Chakravarty, M. M., Lerch, J. P., Naegelin, Y., Stippich, C., Kappos, L., Radue, E.-W., and Sprenger, T. (2014). White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. *BMC Neuroscience*, 15(1):106.
- Menke, J. and Martinez, T. (2004). Using permutations instead of student’s t distribution for p-values in paired-difference algorithm comparisons. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 1331–1335 vol.2.
- Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., and Ourselin, S. (2010). Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–84.
- Nakamura, K. and Fisher, E. (2009). Segmentation of brain magnetic resonance images for measurement of gray matter atrophy in multiple sclerosis patients. *NeuroImage*, 44(3):769–76.
- Opbroek, A. V., der Lijn, F. V., and Bruijne, M. D. (2013). Automated brain-tissue segmentation by multi-feature svm classification. In *Proceedings of the MICCAI Workshops, Challenge on MR Brain Image Segmentation (MRBrainS 13), The Midas Journal*.
- Ourselin, S., Stefanescu, R., and Pennec, X. (2002). Robust registration of multi-modal images: Towards real-time clinical applications. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI’02*, pages 140–147. Springer.
- Pérez-Miralles, F., Sastre-Garriga, J., Tintoré, M., Arramíbide, G., Nos, C., Perkal, H., Río, J., Edo, M. C., Horga, a., Castilló, J., Auger, C., Huerga, E., Rovira, a., and Montalban, X. (2013). Clinical impact of early brain atrophy in clinically isolated syndromes. *Multiple sclerosis (Hounds Mills, Basingstoke, England)*, 19(14):1878–86.
- Pham, D. L. (2001). Spatial Models for Fuzzy Clustering. *Computer Vision and Image Understanding*, 297:285–297.
- Pohl, K. M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R. W., Kikinis, R., Grimson, W. E. L., Shenton,

- M. E., and Wells, W. M. (2007). A hierarchical algorithm for MR brain image parcellation. *IEEE transactions on medical imaging*, 26(9):1201–1212.
- Popescu, V., Battaglini, M., Hoogstrate, W. S., Verfaillie, S. C. J., Sluimer, I. C., van Schijndel, R. A., van Dijk, B. W., Cover, K. S., Knol, D. L., Jenkinson, M., Barkhof, F., de Stefano, N., Vrenken, H., Barkhof, F., Montalban, X., Fazekas, F., Filippi, M., Frederiksen, J., Kappos, L., Miller, D., Palace, J., Polman, C., Rocca, M., Rovira, A., and Yousry, T. (2012). Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. *NeuroImage*, 61:1484–1494.
- Popescu, V., Ran, N. C. G., Barkhof, F., Chard, D. T., Wheeler-Kingshott, C., and Vrenken, H. (2014). Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks. *NeuroImage: Clinical*, 4:366–73.
- Rajchl, M., Baxter, J. S., McLeod, a. J., Yuan, J., Qiu, W., Peters, T. M., and Khan, A. R. (2015). Hierarchical max-flow segmentation framework for multi-atlas segmentation with Kohonen self-organizing map based Gaussian mixture modeling. *Medical Image Analysis*, 000:1–12.
- Roura, E., Oliver, A., Cabezas, M., Valverde, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., and Lladó, X. (2015). A toolbox for multiple sclerosis lesion segmentation. *Neuroradiology*, 57(10):1031–1043.
- Roy, S., He, Q., Sweeney, E., Carass, A., Reich, D. S., Prince, J. L., and Pham, D. L. (2015). Subject-Specific Sparse Dictionary Learning for Atlas-Based Brain MRI Segmentation. *IEEE J. Biomed. Heal. informatics*, 19(5):1598–609.
- Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förtschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., Zimmer, C., Hemmer, B., and Mühlau, M. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage*, 59(4):3774–3783.
- Shiee, N., Bazin, P.-L., Ozturk, A., Reich, D. S., Calabresi, P. A., and Pham, D. L. (2010). A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage*, 49(2):1524–1535.
- Sled, J. G., Zijdenbos, a. P., and Evans, a. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1):87–97.
- Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17:143–155.
- Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., and De Stefano, N. (2002). Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis . *NeuroImage*, 17(1):479–489.
- Sormani, M. P., Arnold, D. L., and De Stefano, N. (2014). Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. *Annals of neurology*, 75(1):43–9.
- Stollenga, M. and Byeon, W. (2015). [M] Parallel Multi-Dimensional LSTM, With Application to Fast Biomedical Volumetric Image Segmentation. In *Neural Information Processing Systems Conference*.
- Sweeney, E. M., Shinohara, R. T., Shiee, N., Mateen, F. J., Chudgar, A. a., Cuzzocreo, J. L., Calabresi, P. a., Pham, D. L., Reich, D. S., and Crainiceanu, C. M. (2013). OA-SIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *NeuroImage: Clinical*, 2:402–13.
- Tomas-Fernandez, X. and Warfield, S. K. (2015). A Model of Population and Subject (MOPS) Intensities with Application to Multiple Sclerosis Lesion Segmentation. *IEEE transactions on medical imaging*, 0062(c):1–15.
- Valverde, S., Oliver, A., Cabezas, M., Roura, E., and Lladó, X. (2015a). Comparison of 10 brain tissue segmentation methods using revisited IBSR annotations. *Journal of Magnetic Resonance Imaging*, 41(1):93–101.

Valverde, S., Oliver, A., Dez, Y., Cabezas, M., Vilanova, J., Rami-Torrentà, L., Rovira, A., and Lladó, X. (2015b). Evaluating the effects of white matter multiple sclerosis lesions on the volume estimation of 6 brain tissue segmentation methods. *American Journal of Neuroradiology*, 36(6):1109–1115.

Valverde, S., Oliver, A., and Lladó, X. (2014). A white matter lesion-filling approach to improve brain tissue volume measurements. *NeuroImage: Clinical*, 6:86–92.

Valverde, S., Oliver, A., Roura, E., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Sastre-Garriga, J., Montalban, X., Rovira, A., and Lladó, X. (2015c). Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling. *NeuroImage: Clinical*, 9:640–647.

Vincken, K. L., Kuijf, H. J., Breeuwer, M., Bouvy, W. H., Bresser, J. D., Alansary, A., Bruijne, M. D., Carass, A., El-baz, A., Jog, A., Katyal, R., Khan, A. R., Lijn, F. V. D., Mahmood, Q., Mukherjee, R., Opbroek, A. V., Paneri, S., Persson, M., Rajchl, M., Sarikaya, D., Silva, C. A., Vrooman, H. A., Vyas, S., Wang, C., Zhao, L., Biessels, G. J., and Viergever, M. A. (2015). MR-BrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans. *Computational Intelligence and Neuroscience*.

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20:45–57.

Zheng, W., Chee, M. W. L., and Zagorodnov, V. (2009). Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. *NeuroImage*, 48:73–83.



# Chapter 7

## Main results and discussion

On previous chapters, we have elaborated on how automated MRI tissue segmentation techniques are increasingly being used as the standard tools to assess brain tissue volume. However, automated tissue segmentation is still a challenging task in MS, due to tissue abnormalities found in MS image patients such as WM lesions that are known to reduce the accuracy of tissue segmentation methods. When expert manual annotations of WM lesions are available, lesion filling has been shown to be an effective method to reduce the effects of these lesions on tissue segmentation. In this aspect, manual annotations are time-consuming and prone to variability between experts, which in combination of the necessity of disseminating focal lesions in space and time has led to the development of a wide number of automated lesion segmentation of MS lesions. Therefore, a solid understanding of the effects of MS lesions on automated pipelines containing linked processes such as lesion segmentation, lesion filling and tissue segmentation is important.

As said in section 1.3, each of these processes cover part of the necessary knowledge to fulfill in order to tackle the problem of automated brain tissue segmentation of images containing lesions. This chapter provides a comprehensive discussion of the main results obtained on previous chapters, analyzing each of these necessary processes until the development of a fully automated tissue segmentation method for MS images.

### 7.1 Effect of WM lesions on tissue segmentation

A wide number of automated tissue segmentation methods have been proposed in the literature so far. On chapter 2, we evaluated the accuracy of ten approaches using the two public available databases of healthy subjects IBSR18 and IBSR20. With the aim to include a wide set of different segmentation approaches and available tools,

the analysis included well-known intensity based algorithms such as ANN, FCM, and KNN, and also public available toolboxes such as FAST, SPM5, SPM8, PVC, GAMIXTURE, SVPASEG, and FANTASM. Results were presented before and after correcting CSF masks, as it was shown that available annotations ignored sulcal CSF tissue on original masks. When sulcal CSF was corrected, SVPASEG, SPM8 and FAST yielded the highest accuracy on both databases. However, most of the methods were variant to changes in acquisition sequences, intensity inhomogeneities or special attributes of the different databases, which highlighted the fact that brain tissue segmentation problem is still open, because there was not a single method that achieved a significant higher accuracy on all tissues.

Afterwards, six of these methods (ANN, FCM, FANTASM, FAST, SPM5 and SPM8) were evaluated on MS data from different hospital centers and scanners in order to analyze the extent to which tissue volume estimations were affected by changes in WM lesion volume and intensity. Our results showed that SPM8 was the method with the lowest differences in total volume, while FANTASM and again SPM8 were the methods where the incidence of WM lesions was the lowest on normal-appearing tissue. In general, differences in tissue volume were lower on methods combining morphological prior information such as SPM5 and SPM8, or spatial constraints such as FANTASM and FAST. In contrast, these differences were higher on simpler intensity based algorithms that lacked of spatial correction such as FCM and ANN. This fact and the higher performance on healthy data of former methods, stress the necessity of adding morphological prior information and/or spatial constraints in automated brain tissue segmentation, not only to overcome inherent MRI artifacts but also as an important component to deal with WM lesions.

The main factor in the observed differences in tissue volume across methods was caused by lesion volume. Furthermore, WM lesion voxels tended to be classified as GM on images where the variation between lesion signal intensity and the mean signal intensity of normal-appearing WM was higher, which indicates a direct relationship between the differences in brain tissue volume and changes in lesion load and WM lesion intensity. However, lesion voxels had also a direct effect on the observed differences in GM and WM outside lesion regions. As already commented on Chapter 3, these differences are especially important because they highlight the bias introduced by WM lesions on the estimation of tissue volume that is not pathologically affected. Our analysis showed that even lesion voxels were not considered to compute brain volume, still methods tended to overestimate GM, specially on images with higher lesion load. Observed differences in normal-appearing tissue volume were important, because although lesion voxels could be reassigned to WM after segmentation, if these lesions were present in image segmentation part of the bias is still present. Furthermore, differences in total tissue volume may be canceled between the errors produced in the same lesion regions and the effect of these voxels

in normal-appearing tissue. This fact clearly shows the necessity to process WM lesion before segmentation.

## 7.2 Effect of lesion filling in tissue segmentation

In the last years, different techniques have been proposed to reduce the bias introduced by WM lesions on brain tissue volume measurements of MS images, mostly by in-painting WM lesions on T1-w with signal intensities similar to normal-appearing WM. After reviewing the related available literature on Chapter 4, we classified existing methods by those that filled WM lesions using the *local* intensities from the surrounding neighboring voxels of lesions, and those that used *global* WM intensities from the whole brain to fill WM lesions. Although all these methods had been already validated separately, we performed a general comparison of all the available techniques in order to analyze their accuracy on the same 1.5T and 3T data and also to investigate its performance with different tissue segmentation techniques such as FAST and SPM8.

This analysis served as a basis to propose a new technique to refill WM lesions which was a compromise between *global* and *local* methods. In contrast to other existing techniques, the proposed method filled lesion voxels intensities with random values of a normal distribution generated from the mean WM signal intensity of each two-dimensional slice. Our results showed that when compared to other methods, our approach yielded the lowest deviation in GM and WM volume on 1.5T and 3T data when FAST was used. When SPM8 was used, the performance of our method was also very competitive, yielding the lowest differences or similar to the best method in GM and WM. In contrast to the rest of pipelines, differences in tissue volume between the same images filled with our method and afterwards segmented with either FAST or SPM8 were very low (< 0.1%), which indicates that the proposed strategy was equally efficient independently of the tissue segmentation chosen.

The proposed algorithm performed significantly better than local methods on images with higher lesion load. In contrast to *global* methods, *local* methods may be limited by the range of similar intensities coming from the neighboring voxels, which on images with a large lesions may be introducing a bias on GM and WM tissue distributions by the addition of a considerable number of voxels with similar intensities. Furthermore, the performance our approach was also better on images with high lesion load when compared with *local* methods, specially on images with lower resolution such as 1.5T data, most probably because our method estimated the mean global NAWM intensity for each slice independently, being more sensible to reproduce possible changes in the intensities between slices.

### 7.3 Effect of automating lesion segmentation and filling on tissue segmentation

As already said earlier, lesion filling has been shown to be an effective method to reduce the effects of these lesions on tissue segmentation. However, in all the lesion-filling approaches including ours, MS lesions have to be known *a priori*, which requires to delineate lesions manually. This was a clear limitation in terms of fully automatizing brain tissue on the presence of MS lesions, which motivated the evaluation of the effect of automated lesion segmentation on tissue segmentation. Although different automated tissue segmentation methods have been proposed, most of them are based on supervised learning, which require to explicitly train them usually with a large amount of labeled data. Labeled data may be not available, which had pointed out the interest of the community in unsupervised methods that can operate without prior data. As shown in Chapter 5, we compared two fully unsupervised pipelines that combined automated lesion segmentation and filling as a first step to understand the effect of fully processed images in tissue segmentation.

Given the performance shown in our previous studies and its widely use in clinical studies, SPM8 was used as a reference tissue segmentation method to measure tissue volume on a set of 70 3T images of CIS patients. On these images, available manual expert annotations were employed to refill WM lesions before segmentation using the filling method of the pipeline, and were considered as ground-truth. Afterwards, we evaluated the differences in GM and WM volume between the set of filled images using manual annotations and the same images processed using different variations of the SLS and LST toolkits that differed in the level of manual intervention. Evaluating different pipelines with distinct levels of automation permitted us to analyze the effect of each of the automated process involved in the observed differences in total and normal-appearing tissue volume.

As suggested in Chapter 3, this new analysis showed that the effect of lesions in total tissue volume was limited due to a canceling effect between the errors produced in the same lesion regions, and the effect of these voxels in normal-appearing tissue. In all the pipelines that incorporated automated lesion segmentation, most of the observed differences in normal-appearing tissue where produced by the effect of false positive lesion voxels that were already segmented without refilling them. In contrast, it did not exist a relevant correlation between the number of false positive lesion voxels and the observed differences in normal-appearing GM and WM, which suggests that most of these miss-classified voxels were actually WM before refilling them. The relationship between errors in automated lesion and tissue segmentation suggest the importance of not only to keep reducing the number of missed lesions, but also stress the necessity of contextual spatial information of lesion regions in

order to confine them in WM and hence reduce the effect of miss-classified voxels on tissue segmentation.

Also as suggested in Chapter 4, our results showed that masking-out lesion voxels before tissue segmentation might not be optimal, as leaving lesion voxels out of the tissue distributions appears to increase the differences in tissue volume with respect to lesion filled images, even if these voxels are re-assigned to WM afterwards. However, although not optimal, masking lesion before segmentation has been found a valid alternative to reduce the effects of WM lesions in clinical settings, and only in the recent years, lesion filling techniques are been already applied on clinical settings. In this aspect, our results show that at least with the evaluated data, the differences in tissue volume between images where expert lesion masks have been masked-out and the same images where lesions have been automatically segmented and filled are similar on images with low lesion load ( $< 10ml$ ). In contrast, within our data differences in tissue volume tend to increase with lesion load on masked-out images, while the increase of the error is more moderated on the fully-automated images. However, given the available data and maximum lesion load considered in our analysis ( $< 20ml$ ), these findings should be considered with care.

In any case, our analysis points out the fact that automated lesion segmentation and filling methods reduced significantly the impact of WM lesions on tissue segmentation method, and with a similar performance to the pipelines that required manual expert intervention. These results are relevant and validate that each of these automated processes can be useful not only in terms of time and economic costs, but also as active processes in fully automated tissue segmentation in the presence of WM lesions.

## **7.4 Fully automated tissue segmentation of images containing WM lesions**

Previous sections have stress the necessity to deal with MS lesions before tissue segmentation, showing several general insights that can be useful for automated tissue segmentation of images containing lesions. The obtained results of the different evaluated methods on chapter 2 and 3 have pointed out the superiority of methods that were benefited by morphological prior information or spatial constraints in automated brain tissue segmentation. More importantly, results on Chapter 3 have evidenced the effect of WM lesion on tissue segmentation and the necessity to deal with MS lesions in order to reduce not only the bias produced by the same lesions but also the effect of these lesion voxels in normal-appearing tissue. In this scenario, we have proposed a new lesion filling technique that was very competitive with dif-

ferent databases and tissue segmentation methods, as shown on Chapter 4. Finally, we have shown on Chapter 5 that the addition of unsupervised lesion segmentation and filling into already existing tissue segmentation pipelines reduced significantly the error in tissue volume when compared with previous pipelines where lesions were segmented as normal tissue.

Following these insights, we have developed a novel multi-channel method designed to segment tissues in images of MS patients. As explained in chapter 6, this approach makes use of a combination of intensity, anatomical and morphological prior maps to guide the tissue segmentation. Tissue segmentation has been tackled based on a robust partial volume segmentation where WM outliers have been estimated and refilled before segmentation using a multi-channel post-processing algorithm also integrating multi-channel support, partial volume segmentation, spatial context, and prior anatomical and morphological atlases. Furthermore, the proposed method takes advantage of new affordable processors such as GPUs that reduce up to four times the execution time to register and segment tissue when compared to general purpose CPUs, which make useful for studies containing a large number of subjects to analyze.

The proposed method has been quantitatively and qualitatively evaluated using different databases of images containing WM lesions. In order to analyze the extent to which T1-w and FLAIR modalities intervened in the obtained accuracy, the proposed method was run in all experiment using only T1-w or using both T1-w and FLAIR image sequences. As shown by the presented results, the proposed technique yielded competitive and consistent results in both general and MS specific databases without parameter tweaking. In the MRBrainS tissue segmentation challenge, our method combining both T1-w and FLAIR was the best non-supervised technique of the challenge, being ranked in the 7th position out of 31 participant methods. When only the T1-w modality was used, still the accuracy of the proposed method was clearly superior to general purpose methods such as FAST (ranked 21th) and SPM12 (best ranked 17th), even if those used both image modalities. In MS data, the performance of our method combining T1-w and FLAIR sequence was similar or better to the best evaluated pipeline incorporating lesion segmentation and filling. Obtained differences in tissue volume between images processed with the proposed algorithm and the same images where lesions were filled using expert lesion annotations were lower than 0.15% on all tissues, validating the overall capability of the proposed method to reduce the effects of WM lesions on tissue segmentation.

In general, our results showed that the percentages of error in tissue volume of our proposed approach were low and similar in both databases. The percentages of error were the lowest when the FLAIR modality was used, which evidences that this image sequence has a direct effect on the efficiency of the algorithm, and consequently it should be used when available. However, the accuracy of the method using only

the T1-w modality was also superior to other general purpose strategies, which also evidences that the improvement in tissue segmentation was not only generated by the addition of the FLAIR modality, but also by the combination of intensity, anatomical and morphological priors, and the use of an specific outlier algorithm with integrated lesion filling.

## 7.5 Related collaborations

No entenc molt bé com enfocar aquest punt aquí.

- Que haig de posar que he fet a la resta de papers?
- Comentar sobre colabracions reals? lesion filling a registre?



# Chapter 8

## Conclusions

This thesis synthesizes our work during the last three years. Following the same objectives and stages proposed in the Introduction chapter, we summarize the main conclusions of this thesis:

- We analyzed the state of the art of tissue segmentation methods. This first stage aimed to review different proposed tissue segmentation techniques in order to understand their advantages and drawbacks. As part of the resulting analysis published in the *Journal of Magnetic Resonance Imaging* in January of 2014, **our results showed a higher accuracy on several methods that incorporated morphological prior information and/or spatial constraints such as FAST, SPVASEG and SPM8**. These methods were also less prone to changes in acquisition sequences and intensity inhomogeneities.
- We studied the effect of WM lesions on tissue segmentation of MS patient images. The second stage to cover was focused on the analysis of the effects of WM lesions on the tissue distributions. Six of the analyzed methods on Chapter 2 were evaluated on multi-center 1.5T MS data from different scanners. Related to the previous stage, our results stressed **the necessity of adding morphological prior information and/or spatial constraints in automated brain tissue segmentation, not only to overcome inherent MRI artifacts but also as an important component to deal with WM lesions**. Furthermore, our analysis of the effects of WM lesions on tissue volume showed that **the inclusion of WM lesions on tissue segmentation not only biased the total tissue volume measurements by the addition of miss-classified lesion voxels, but also by the effect of these lesions in observed differences in normal-appearing tissue volume**.

The entire analysis was published in the *American Journal of Neuroradiology* in February 2015.

- We proposed a new technique to reduce the effects of WM lesions on tissue segmentation of MS patient images. The third stage required first to compare the accuracy of different proposed lesion filling techniques in the literature with the aim to afterwards propose a new technique to reduce the effects of WM lesions on tissue segmentation. **The proposed lesion filling method was shown effective in different data and independently of the tissue segmentation method used afterwards.** The proposed approach outperformed the rest of methods on both 1.5T and 3T data when FAST was used, while its performances was similar or lower to the best available strategy when SPM8 was used. The proposed lesion filling method was published in the *NeuroImage: Clinical journal* in August of 2014. **Furthermore, we released a public version on the proposed method that can be freely downloaded from our research team web page<sup>1</sup>.**
- We analyzed the effect of automated WM lesion segmentation and filling on the tissue segmentation. During the fourth stage proposed, we analyzed the accuracy of two state-of-the-art automated pipelines that incorporate unsupervised lesion segmentation, lesion filling and tissue segmentation on MS data. As shown in published paper in the *NeuroImage: Clinical journal* in October of 2015, our analysis showed that **pipelines that incorporated automated lesion segmentation and filling were capable to reduce significantly the impact of WM lesions on tissue segmentation, performing similarly to the pipelines that required manual expert intervention.**
- Finally, we proposed a new fully automated tissue segmentation method for MS patient images containing lesions. The main goal of the thesis was to propose a fully automated tissue segmentation method capable to deal with images of MS patients. As shown in Chapter 6, the proposed method incorporated all the major insights obtained from previous stages with the aim of provide a robust fully automated tissue approach for accurate brain volume measurements. Our results showed that **when compared with existent tissue segmentation methods, the presented approach yielded a higher accuracy in tissue segmentation while the influence of MS lesions on tissue segmentation was lower or similar to the best state-of-the-art pipeline incorporating automated lesion segmentation and filling.** This work has been submitted for publication in the *Medical Image Analysis*

---

<sup>1</sup>The latest version on the proposed lesion filling method can be download from <http://atc.udg.edu/nic/slftoolbox/index.html>

*journal* in February 2016. As part of the goal, we also released a public version on the proposed method that can be freely downloaded from our research team web page<sup>2</sup>.

## 8.1 Future work

Unfortunately, there are several aspects that have been not investigated during this thesis. One of the main limitations on several stages has been the lack of 3T images with high lesion load. As pointed out in Chapters 5 and 6, the low mean lesion load of the cohorts analyzed has not allowed to investigate better the performance of the analyzed pipelines in the presence of images with higher lesion load. In the case of the proposed tissue segmentation method, we believe that an additional analysis of the performance of the method with images containing lesions with higher lesion load would be helpful not only to analyze the robustness of the proposed algorithm, but also to investigate the benefits of adding other image channels such as T2 or PD.

Secondly, although the proposed tissue segmentation method has been designed for cross-sectional data, there is a increasingly clinical interest in the measurements of longitudinal changes in tissue volume. In this aspect, we believe that the proposed method would be also adapted to longitudinal changes by re-adapting the pipeline with prior registering of time point images before tissue segmentation. This is in fact one of the goals that our team wants to tackle first within the research framework of the BiomarkEM.cat project, in order to release suitable tools that can be used in clinical settings.

---

<sup>2</sup>A public version of the method can be download from <http://atc.udg.edu/nic/msseg/index.html>



# Bibliography

- [1] Julio Acosta-Cabronero, Guy B. Williams, João M S Pereira, George Pengas, and Peter J. Nestor. The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. *NeuroImage*, 39(4):1654–1665, 2008.
- [2] Ayelet Akselrod-Ballin, Meirav Galun, Moshe John Gomori, Ronen Basri, and Achi Brandt. Atlas guided identification of brain structures by combining 3D segmentation and SVM classification. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 9(Pt 2):209–16, 2006.
- [3] J B Arnold, J S Liow, K a Schaper, J J Stern, J G Sled, D W Shattuck, a J Worth, M S Cohen, R M Leahy, J C Mazziotta, and D a Rottenberg. Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage*, 13(5):931–943, 2001.
- [4] John Ashburner and Karl J. Friston. Unified segmentation. *NeuroImage*, 26:839–851, 2005.
- [5] Marco Battaglini, Mark Jenkinson, Nicola De Stefano, and Nicola De Stefano. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Human Brain Mapping*, 33(9):2062–71, 2012.
- [6] S. Bricq, Ch. Collet, and J.P. Armspach. Unifying framework for multimodal brain mri segmentation based on hidden markov chains. *Medical Image Analysis*, 12(6):639 – 652, 2008. *[ce:title]Special issue on information processing in medical imaging 2007[ce:title]*.
- [7] Per Brodal. *The central nervous system*. 2010.
- [8] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158–e177, 2011.

- [9] Mariano Cabezas, Arnau Oliver, Eloy Roura, Jordi Freixenet, Joan C. Vilanova, Llu??s Rami??-Torrent??, ??lex Rovira, and Xavier Llad?? Automatic multiple sclerosis lesion detection in brain MRI by FLAIR thresholding. *Computer Methods and Programs in Biomedicine*, 115(3):147–161, 2014.
- [10] Mariano Cabezas, Arnau Oliver, Sergi Valverde, Brigitte Beltran, Jordi Freixenet, Joan C. Vilanova, Llu??s Rami??-Torrent??, ??lex Rovira, and Xavier Llad?? BOOST: A supervised approach for multiple sclerosis lesion segmentation. *Journal of Neuroscience Methods*, 237:108–117, 2014.
- [11] Beno??t Caldairou, Nicolas Passat, Piotr A Habas, Colin Studholme, and Fran??ois Rousseau. A non-local fuzzy segmentation method: application to brain mri. *Pattern Recognition*, 44(9):1916–1927, 2011.
- [12] A Ceccarelli, J S Jackson, S Tauhid, A Arora, J Gorky, E Dell’Oglio, A Bakshi, T Chitnis, S J Khouri, H L Weiner, C R G Guttmann, R Bakshi, and M Neema. The impact of lesion in-painting and registration methods on voxel-based morphometry in detecting regional cerebral gray matter atrophy in multiple sclerosis. *American Journal of Neuroradiology*, 33(8):1579–85, September 2012.
- [13] Antonia Ceccarelli, Maria a Rocca, Mohit Neema, Vittorio Martinelli, Ashish Arora, Shahamat Tauhid, Angelo Ghezzi, Giancarlo Comi, Rohit Bakshi, and Massimo Filippi. Deep gray matter T2 hypointensity is present in patients with clinically isolated syndromes suggestive of multiple sclerosis. *Multiple sclerosis (Hounds Mills, Basingstoke, England)*, 16(1):39–44, 2010.
- [14] D T Chard, G J M Parker, R Kapoor, A J Thompson, and D H Miller. Brain atrophy in clinically early relapsing remitting multiple sclerosis. 2002.
- [15] Declan T. Chard, Jonathan S. Jackson, David H. Miller, and Claudia A M Wheeler-Kingshott. Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. *Journal of Magnetic Resonance Imaging*, 32:223–228, 2010.
- [16] Harvey E. Cline, William E. Lorensen, Ron Kikinis, and Ferenc Jolesz. Three-Dimensional Segmentation of MR Images of the Head Using Probability and Connectivity. *Journal of Computer Assisted Tomography*, 14(6):1037–1045, 1990.
- [17] Alastair Compston and Alasdair Coles. Multiple sclerosis. *Lancet*, 359(9313):1221–31, 2002.
- [18] Alastair Compston and Alasdair Coles. Multiple sclerosis. *Lancet*, 372(9648):1502–17, 2008.

- [19] Devic M Confavreux C, Aimard G. Course and prognosis of multiple sclerosis assessed by the computerized data processing of 349 patients. *Brain*, 103(2):281–300, 1980.
- [20] Renske De Boer, Henri A Vrooman, Fedde Van Der Lijn, Meike W Vernooij, M Arfan Ikram, Aad Van Der Lugt, Monique M B Breteler, and Wiro Niessen. White matter lesion extension to automatic brain tissue segmentation on mri. *NeuroImage*, 45(4):1151–1161, 2009.
- [21] N. De Stefano, A. Giorgio, M. Battaglini, M. Rovaris, M. P. Sormani, F. Barkhof, T. Korteweg, C. Enzinger, F. Fazekas, M. Calabrese, D. Dinacci, G. Tedeschi, A. Gass, X. Montalban, A. Rovira, A. Thompson, G. Comi, D. H. Miller, and M. Filippi. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. *Neurology*, 74(23):1868–1876, 2010.
- [22] Hrishikesh Deshpande, Pierre Maurel, and Christian Barillot. Classification of Multiple Sclerosis Lesions using Adaptive Dictionary Learning. *Computerized Medical Imaging and Graphics*, 2015.
- [23] Yago Diez, Arnau Oliver, Mariano Cabezas, Sergi Valverde, Robert Martí, Joan Carles Vilanova, Lluís Ramió-Torrentà, Àlex Rovira, and Xavier Lladó. Intensity based methods for brain MRI longitudinal registration. A study on multiple sclerosis patients. *Neuroinformatics*, 12(3):365–379, 2014.
- [24] Robert R Edelman and Steven Warach. Magnetic Resonance Imaging. *New England Journal of Medicine*, 328(10):708–716, 1993.
- [25] Massimo Filippi, Paolo Preziosa, Massimiliano Copetti, Gianna Riccitelli, Mark a Horsfield, Vittorio Martinelli, Giancarlo Comi, and Maria a Rocca. Gray matter damage predicts the accumulation of disability 13 years later in MS. *Neurology*, 81(20):1759–67, nov 2013.
- [26] Massimo Filippi and Maria A Rocca. MR imaging of multiple sclerosis. *Radiology*, 259(3):659–81, 2011.
- [27] Elizabeth Fisher, Jar-Chi Lee, Kunio Nakamura, and Richard A Rudick. Gray matter atrophy in multiple sclerosis: A longitudinal study. *Annals of Neurology*, 64(3):255–265, sep 2008.
- [28] Leonora K Fisniku, Declan T Chard, Jonathan S Jackson, Valerie M Anderson, Daniel R Altmann, Katherine A Miszkiel, Alan J Thompson, and David H Miller. Gray Matter Atrophy Is Related to Long- Term Disability in Multiple Sclerosis. *Ann Neurol*, 64:247–254, 2008.

- [29] Onur Ganiler, Arnau Oliver, Yago Diez, Jordi Freixenet, Joan C. Vilanova, Brigitte Beltran, Llu??s Rami??-Torrent??, ??lex Rovira, and Xavier Llad?? A subtraction pipeline for automatic detection of new appearing multiple sclerosis lesions in longitudinal studies. *Neuroradiology*, 56(5):363–374, 2014.
- [30] Daniel Garc??a-Lorenzo, Simon Francis, Sridar Narayanan, Douglas L Arnold, and D Louis Collins. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Medical Image Analysis*, 17(1):1–18, January 2013.
- [31] Ezequiel Geremia, Olivier Clatz, Bjoern H. Menze, Ender Konukoglu, Antonio Criminisi, and Nicholas Ayache. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage*, 57(2):378–390, 2011.
- [32] Guido Gerig, John Martin, Ron Kikinis, Olaf Kubler, Martha Shenton, and Ferenc A. Jolesz. Unsupervised tissue type segmentation of 3D dual-echo MR head data. *Image and Vision Computing*, 10(6):349–360, 1992.
- [33] Tal Geva. Magnetic resonance imaging: historical perspective. *Journal of Cardiovascular Magnetic Resonance*, 8(4):573–580, 2006.
- [34] Antonio Giorgio and Nicola De Stefano. Clinical use of brain volumetry. *Journal of Magnetic Resonance Imaging*, 37(1):1–14, 2013.
- [35] Nicolas Guizard, Pierrick Coup??, Vladimir S. Fonov, Jose V. Manj??n, Douglas L. Arnold, and D. Louis Collins. Rotation-invariant multi-contrast non-local means for MS lesion segmentation. *NeuroImage: Clinical*, 8:376–389, 2015.
- [36] Rola Harmouche, Nagesh K Subbanna, D Louis Collins, Douglas L Arnold, and Tal Arbel. Based on Modeling Regional Intensity Variability and Local Neighborhood Information. 62(5):1281–1292, 2015.
- [37] Zujun Hou. A review on MR image intensity inhomogeneity correction, 2006.
- [38] Saurabh Jain, Diana M. Sima, Annemie Ribbens, Melissa Cambron, Anke Maertens, Wim Van Hecke, Johan De Mey, Frederik Barkhof, Martijn D. Steenwijk, Marita Daams, Frederik Maes, Sabine Van Huffel, Hugo Vrenken, and Dirk Smeets. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage: Clinical*, 8:367–375, 2015.
- [39] T Kapur, W E Grimson, W M Wells, and R Kikinis. Segmentation of brain tissue from magnetic resonance images. *Medical image analysis*, 1(2):109–27, 1996.

- [40] Jong Min Lee, Uicheul Yoon, Sang Hee Nam, Jung Hyun Kim, In Young Kim, and Sun I. Kim. Evaluation of automated and semi-automated skull-stripping algorithms using similarity index and segmentation error. *Computers in Biology and Medicine*, 33(6):495–507, 2003.
- [41] Xavier Lladó, Onur Ganiler, Arnau Oliver, Robert Martí, Jordi Freixenet, Laia Valls, Joan C. Vilanova, Lluís Ramió-Torrentà, and Àlex Rovira. Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology*, 54:787–807, 2012.
- [42] Xavier Lladó, Arnau Oliver, Mariano Cabezas, Jordi Freixenet, Joan C. Vilanova, Ana Quiles, Laia Valls, Lluís Ramió-Torrentà, and Àlex Rovira. Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Information Sciences*, 186(1):164–185, March 2012.
- [43] F. D. Lublin and S. C. Reingold. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology*, 46(4):907–911, April 1996.
- [44] Stefano Magon, Laura Gaetano, M Mallar Chakravarty, Jason P Lerch, Yvonne Naegelin, Christoph Stippich, Ludwig Kappos, Ernst-Wilhelm Radue, and Till Sprenger. White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. *BMC Neuroscience*, 15(1):106, January 2014.
- [45] D Mahapatra. Analyzing Training Information From Random Forests for Improved Image Segmentation. *Image Processing, IEEE Transactions on*, 23(4):1504–1512, 2014.
- [46] J.L. Marroquin, B.C. Vemuri, S. Botello, E. Calderon, and A. Fernandez-Bouzas. An accurate and efficient bayesian method for automatic segmentation of brain mri. *Medical Imaging, IEEE Transactions on*, 21(8):934 –945, aug. 2002.
- [47] Dzung L Pham. Spatial Models for Fuzzy Clustering. *Computer Vision and Image Understanding*, 297:285–297, 2001.
- [48] Chris H Polman, Stephen C Reingold, Brenda Banwell, Michel Clanet, Jeffrey a Cohen, Massimo Filippi, Kazuo Fujihara, Eva Havrdova, Michael Hutchinson, Ludwig Kappos, Fred D Lublin, Xavier Montalban, Paul O’Connor, Magnhild Sandberg-Wollheim, Alan J Thompson, Emmanuelle Waubant, Brian Weinshenker, and Jerry S Wolinsky. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of neurology*, 69(2):292–302, 2011.

- [49] V. Popescu, M. Battaglini, W. S. Hoogstrate, S. C J Verfaillie, I. C. Sluimer, R. A. van Schijndel, B. W. van Dijk, K. S. Cover, D. L. Knol, M. Jenkins, F. Barkhof, N. de Stefano, H. Vrenken, F. Barkhof, X. Montalban, F. Fazekas, M. Filippi, J. Frederiksen, L. Kappos, D. Miller, J. Palace, C. Polman, M. Rocca, A. Rovira, and T. Yousry. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. *NeuroImage*, 61(4):1484–1494, 2012.
- [50] V. Popescu, N. C G Ran, F. Barkhof, D. T. Chard, C. A. Wheeler-Kingshott, and H. Vrenken. Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks. *NeuroImage: Clinical*, 4:366–373, 2014.
- [51] Martin Rajchl, John S.H. Baxter, a. Jonathan McLeod, Jing Yuan, Wu Qiu, Terry M. Peters, and Ali R. Khan. Hierarchical max-flow segmentation framework for multi-atlas segmentation with Kohonen self-organizing map based Gaussian mixture modeling. *Medical Image Analysis*, 000:1–12, 2015.
- [52] E. Roura, T. Schneider, M. Modat, P. Daga, N. Muhlert, D. Chard, S. Ourselin, X. Lladó, and C. A. M. Wheeler-Kingshott. Multi-channel registration of fa and t1w images in the presence of atrophy: application to multiple sclerosis. *Functional Neurology*, 30(4), 2015.
- [53] Eloy Roura, Arnau Oliver, Mariano Cabezas, Sergi Valverde, Deborah Pareto, Joan C Vilanova, Lluís Ramió-Torrentà, Àlex Rovira, and Xavier Lladó. A toolbox for multiple sclerosis lesion segmentation. *Neuroradiology*, 57(10):1031–1043, 2015.
- [54] Eloy Roura, Arnau Oliver, Mariano Cabezas, Joan C. Vilanova, ??lex Rovira, Lluís Ramió-Torrentà, and Xavier Lladó. MARGA: Multispectral Adaptive Region Growing Algorithm for brain extraction on axial MRI. *Computer Methods and Programs in Biomedicine*, 113(2):655–673, 2014.
- [55] Snehashis Roy, Aaron Carass, Pierre-Louis Bazin, Susan Resnick, and Jerry L. Prince. Consistent segmentation using a rician classifier. *Medical Image Analysis*, 16(2):524 – 535, 2012.
- [56] Richard a Rudick, Jar-Chi Lee, Kunio Nakamura, and Elizabeth Fisher. Gray matter atrophy correlates with MS disability progression measured with MSFC but not EDSS. *Journal of the neurological sciences*, 282(1-2):106–11, jul 2009.
- [57] Paul Schmidt, Christian Gaser, Milan Arsic, Dorothea Buck, Annette Förtschler, Achim Berthele, Muna Hoshi, Rüdiger Ilg, Volker J Schmid, Claus Zimmer, Bernhard Hemmer, and Mark Mühlau. An automated tool for detection of

- FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage*, 59(4):3774–3783, February 2012.
- [58] Michaël Sdika and Daniel Pelletier. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Human Brain Mapping*, 30(4):1060–1067, 2009.
- [59] D W Shattuck, S R Sandor-Leahy, K a Schaper, D a Rottenberg, and R M Leahy. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13(5):856–876, 2001.
- [60] A. Simmons, P. S. Tofts, G. J. Barker, and S. R. Arridge. Sources of intensity nonuniformity in spin echo images at 1.5 T. *Magnetic Resonance in Medicine*, 32(1):121–128, 1994.
- [61] J G Sled, A P Zijdenbos, and A C Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998.
- [62] Stephen M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, 2002.
- [63] G H Sperber. Clinically Oriented Anatomy. *Journal of anatomy*, 208(3):393, 2006.
- [64] Lawrence Steinman, M.D. Multiple Sclerosis: A Coordinated Immunological Attack against Myelin in the Central Nervous System. *Cell*, 85(3):299–302, may 1996.
- [65] Elizabeth M Sweeney, Russell T Shinohara, Navid Shiee, Farrah J Mateen, Avni a Chudgar, Jennifer L Cuzzocreo, Peter a Calabresi, Dzung L Pham, Daniel S Reich, and Ciprian M Crainiceanu. OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *NeuroImage: Clinical*, 2:402–13, jan 2013.
- [66] Jussi Tohka, Ivo D. Dinov, David W. Shattuck, and Arthur W. Toga. Brain mri tissue classification based on local markov random fields. *Magnetic Resonance Imaging*, 28(4):557 – 573, 2010.
- [67] Xavier Tomas-Fernandez and Simon Keith Warfield. A Model of Population and Subject (MOPS) Intensities with Application to Multiple Sclerosis Lesion Segmentation. *IEEE transactions on medical imaging*, 0062(c):1–15, jan 2015.

- [68] Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, Yuanjie Zheng, Alexander Egan, Paul A. Yushkevich, and James C. Gee. N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010.
- [69] Annegreet van Opbroek, Fedde van der Lijn, and Marleen de Bruijne. Automated brain-tissue segmentation by multi-feature svm classification. In *Proceedings of the MICCAI Workshopsâ€The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainSâ€13)*, 2013.
- [70] H Vrooman, Fedde van der Lijn, and W Niessen. Auto-knn: brain tissue segmentation using automatically trained knearest-neighbor classification. In *Proceedings of the MICCAI Workshopsâ€The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainSâ€13)*, 2013.
- [71] Zhao Yi, Antonio Criminisi, Jamie Shotton, and Andrew Blake. Discriminative, semantic segmentation of brain tissue in mr images. In Guang-Zhong Yang, David Hawkes, Daniel Rueckert, Alison Noble, and Chris Taylor, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, volume 5762 of *Lecture Notes in Computer Science*, pages 558–565. Springer Berlin / Heidelberg, 2009.
- [72] Y Zhang, M Brady, and S Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20:45–57, 2001.