A NEW EVALUATION OF THE BRAIN PARENCHYMAL FRACTION: APPLICATION IN MULTIPLE SCLEROSIS GROUP OR LONGITUDINAL STUDIES

J.C. Souplet[1], *C. Lebrun*[2], *N. Ayache*[1], *G. Malandain*[1]

1. INRIA, Asclepios team, 2004 Route des Lucioles BP 93, Sophia-Antipolis, 06902 France
2. CHU Pasteur, Neurology department, 30 voie romaine BP 69, Nice, 06002 France

ABSTRACT

In multiple sclerosis (MS) research, burden of disease and treatments efficacy are mainly evaluated with lesion load and atrophy. The former being poorly correlated with patient's handicap, it is of interest to evaluate accurately the latter. A lot of methods to measure the brain atrophy as the brain parenchymal fraction (BPF) are available in the literature. The BPF needs a precise segmentation of the brain and of the Cerebro-Spinal Fluid. However artefacts like partial volume effects (PVE) can impair this classification. According to some articles, the BPF may also be less precise in longitudinal studies. To address these points, this article proposes a new method to evaluate the BPF thanks to an Expectation-Minimization framework taking into consideration PVE. Modifications of the pipeline are also proposed to improve its reliability in longitudinal study. Experiments have been conducted on simulated pathological images that validate the different measures.

***Index Terms*—** Biomedical Image Processing, Biomedical Magnetic Resonance Imaging, Multiple Sclerosis, Atrophy

1. INTRODUCTION

According to the modified Mc Donald criteria [1], MRI lesions number and location evaluations are mandatory to diagnose Multiple Sclerosis (MS). Lesions load measurements is also used in follow-up studies and pharmaceutical research as surrogate markers. Except for clinically isolated syndromes, clinical studies have shown that T2 lesions load is poorly correlated with patient's handicap [2]. In consequence, other approaches like global atrophy measurement are studied.

Different techniques to quantify brain atrophy in MS are available [3]. One currently used method is the evaluation of the brain parenchymal fraction (BPF) [4]. This method required a precise segmentation of the cerebro-spinal fluid (CSF) compartment or of the brain (gray matter (GM), white matter (WM) and lesions). However some voxels can contain a mixture of two classes (e.g. CSF and GM) because of image resolution: this Partial Volume Effect (PVE) may impair greatly the classification and subsequently the atrophy measurement. For example, sulci introduce a lot of GM/CSF PVE and this effect is one of the main problems to take in consideration to obtain a reliable CSF volume evaluation.

The BPF evaluation presents also other difficulties like variation of the segmentation results caused by the inter- and intra-image inhomogeneities or the skull-stripping step. Moreover the BPF is presented as more appropriate in group study (e.g. MS patients vs. healthy controls) than in longitudinal studies in [3]. Indeed the use of methods like SIENA[1] [5] is presented as a better choice for this kind of study. SIENA uses the local shifts in brain edges to evaluate the percentage brain volume change between two instants.

To improve the robustness of the atrophy measurement result and to address the PVE in CSF volume evaluation, we propose here a novel method to evaluate the BPF. First the segmentation algorithm is presented and is validated with simulated data for which the ground truth is available. Then we propose a method to improve the robustness of the measurement in longitudinal study and we compare our method with SIENA method. The remainder of this article is organized as follows: in Section 2, we describe the method pipeline; sections 3 validates the different steps of the method; perspectives are discussed in Section 4.

2. METHOD

Our new atrophy measurement method is divided in different steps. The different MRI sequences have to be preprocessed. Then a multi channel Expectation Maximization (EM) classification method is applied. From this classification, different segmentations are generated. The computation of these segmentations volume allows to evaluate atrophy.

2.1. MRI sequences coregistration

When diagnosing MS, three MRI sequences are classically acquired: T1, T2 and Proton Density (PD) weighted images. T2 and PD are intrinsically coregistered but this is not the case of T1. As T1 has a higher resolution than T2 and PD, we

[1] http://www.fmrib.ox.ac.uk/fsl/

register T1 on T2 [6] to limit the partial volume effect caused by the resampling.

2.2. Intensity correction

MR images can suffer from bias [7]. We estimate and correct it with the Expectation/Conditional Maximization algorithm proposed in [8].

2.3. Skull-stripping

Classification step is sensitive to a preliminary step, called skull stripping, which aimed at isolating the brain in the image. Indeed, if this step is too restrictive, voxels belonging to the brain may be discarded, or conversely if it is too permissive, part of the meninx may be retained and subsequently misclassified. Different automatic skull stripping methods are available in the literature. As explain in [9], we use a combination of Dugas *et al.* method [10], BET [11] and 3dIntracranial [12] to segment the brain.

2.4. EM classification method

The EM framework is currently used to classify brain MRI voxels. To take into consideration the different artefacts which are present in MS patients brain MRI, we decided to classify voxels into ten classes: WM, GM, CSF, six GM/CSF PVE classes (with different proportions), and an outlier class, that will contain vessels and some MS lesions [10]. The probability density function (PDF) of each class is modelled by Gaussians, μ and σ denoting respectively the mean and standard deviation. Our PVE model approximates the intensity of a voxel which contain a proportion α of tissue x with the intensity I_x and a proportion $(1-\alpha)$ of tissue y by $I_{PVE} = \alpha * I_x + (1-\alpha) * I_y$. PVE classes PDF follows then a Gaussian PDF with a mean of $(\alpha\mu_x + (1-\alpha)\mu_y)$ and a standard deviation of $\sqrt{\alpha^2\sigma_x + (1-\alpha)^2\sigma_y}$.

Following an initialization thanks to an affine registration of the MNI probabilistic atlases [13], our EM framework is then composed of three steps which are iterated:

- the Expectation step which consists in the labelization of all classes (including PVE classes),
- the Maximization step which consists in the estimation of the CSF, GM, WM, Outliers Gaussians parameters by maximizing the likelihood of the whole image,
- the computation of PVE classes' parameters.

After algorithm convergence, the final segmentations are obtained by classifying each voxel to the most probable class. MS lesions are classified mainly in the GM or outlier classes.

2.5. Volume and Brain parenchymal fraction computation

To include PVE segmentation into volumes computation, we generate CSF and brain repartition maps. These maps are not a probabilistic segmentation of a compartment but give the proportion of the considered class (or compartment) in each voxel. CSF and brain repartition maps ($RM_{(CSF)}$, $RM_{(B)}$) are obtained by equations 1 and 2 where $SEG_{(PVE\alpha)}$ represents the segmentation of the PVE class with the proportion α of GM.

$$RM_{(CSF)} = \sum_{i=1}^{6} \frac{7-i}{7} \times SEG_{PVEi} + SEG_{CSF} \quad (1)$$

$$RM_{(B)} = \sum_{i=1}^{6} \frac{i}{7} \times SEG_{PVEi} + SEG_{GM} + SEG_{WM} \quad (2)$$

CSF and brain volumes are then obtained by the addition of all the voxels values of the considered repartition map, multiplied by the volume of one voxel. Thanks to the volume computation, the BPF can be computed (see equation 3).

$$BPF = 100 \times \frac{\text{Brain Volume}}{\text{Brain Volume + CSF Volume}} \quad (3)$$

2.6. Atrophy computation

As defined in the previous section, the BPF allows to compare different groups of population (e.g. MS patients vs. healthy controls). If acquisitions at different timepoints are available, the atrophy is given by the difference of successive BPF. However some part of the pipeline process has to be changed. All the images are co-registered on the T2 sequence of the first timepoint. The intensities of images of the same sequence are equalized. The skull-stripped mask of the first timepoint is used for the following ones. In the EM classification, the classes parameters are computed from the images of all timepoints. Then the obtained parameters are used to give the corresponding segmentation at each timepoint. Since outliers are not considered in BPF, this may bias the atrophy measure: to handle this, we consider as outliers for all the timepoints the union of all outliers detected in each timepoint.

3. VALIDATION

It is not realistic to validate the segmentation of PVE classes, since their PDF have a significant overlap. In consequence, an expert has first realized a qualitative validation by visual inspection of the results on real MRI but no significant error has been identified. Secondly, we realize a quantitative comparison of the obtained segmentations on simulated data with the ground truth.

Table 1. Segmentation comparison criteria

Criteria	Conventional criteria	Generalized criteria
Similarity Index (SI)	$\frac{2Card(Ref \times Seg)}{Card(Ref)+Card(Seg)}$	$\frac{2\sum_i \min(Ref_{(i)}, Seg_{(i)})}{\sum_i Ref_{(i)}+\sum_i Seg_{(i)}}$
Sensitivity (SEN)	$\frac{Card(Ref \times Seg)}{Card(Ref)}$	$\frac{\sum_i \min(Ref_{(i)}, Seg_{(i)})}{\sum_i Ref_{(i)}}$
Specificity (SPE)	$\frac{Card(\overline{Ref \times Seg})}{Card(\overline{Ref})}$	$\frac{\sum_i \min(1-Ref_{(i)}, 1-Seg_{(i)})}{\sum_i (1-Ref_{(i)})}$

3.1. Segmentation comparison criteria

The first step of a validation is to identify comparison criteria. The used classification method gives GM and CSF repartition maps but not binary segmentations (see section 2.5). To compare these maps, we use the approach presented in [14] to define a generalize version of the Similarity Index (SI), of the sensitivity (SEN) and the specificity (SPE). These criteria are given in Table 1 for a segmentation (Seg) with a reference image (Ref). $Ref_{(i)}$ and $Seg_{(i)}$ represents the intensity of the voxel i in the corresponding image. The obtained equations give the same results than the conventional equation on binary images and their values remain between 0 and 1.

3.2. Segmentation comparison results

The method segmentations have been validated thanks to BrainWeb[1] simulated images. Images from the moderate MS lesions brain anatomical model have been generated (Noise: 3%, Intensity non-uniformity: 20%). The ground truth is made of images with a slice thickness of 1 mm: from it, we also generate ground truth images with a slice thickness of 3 mm that are then repartition maps. We evaluate then our method with images generated by BrainWeb with slice thicknesses of both 1 and 3 mm. The results of this evaluation are given in table 2.

Table 2. Segmentations comparison results for Brainweb MS anatomical model.

Slide thickness	Tissues	SI	SEN	SPE
MS 1mm	CSF	0.81	0.79	0.99
MS 1mm	Brain	0.98	0.98	0.99
MS 3mm	CSF	0.82	0.81	0.99
MS 3mm	Brain	0.95	0.98	0.97

The obtained results yield correct segmentations ($SI > 0.8$). The sensibility values are correct ($SEN > 0.79$) even if they seem to indicate a slight under segmentation. The specificity values ($SPE > 0.97$) indicate that there are few false positive voxels. The CSF results are slightly weaker than Brain results. This can be explained by the fact that CSF compartment compared to gray matter (GM) or white matter (WM) represents the smallest volume. This explains that any misclassified voxels will introduce a larger relative error to CSF classification than to Brain (GM+WM) classification.

[1] http://www.bic.mni.mcgill.ca/brainweb/

Table 3. Obtained relative volume errors

	CSF	GM	WM	Brain
MS 1mm	-4 %	+9 %	-11 %	+1 %
MS 3mm	-5 %	+16 %	-9 %	+5 %

3.3. Volume measurements validation

Table 3 gives the relative error in CSF and brain volume measurements. GM and WM volume estimation errors are relevant but the error on the brain volume is acceptable. Moreover we have a good estimation of the CSF volume. Indeed the average CSF volume error is equal to 4.5% whereas the CSF volume is the smallest compared to GM or WM or brain volumes.

3.4. Atrophy measurements validation

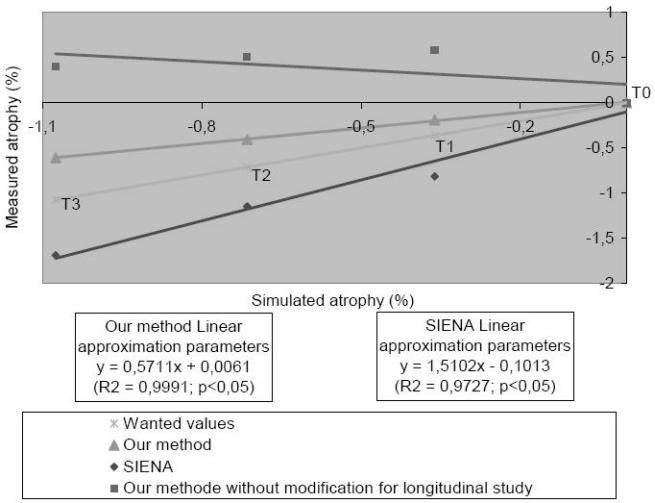

Fig. 1. Measured atrophy vs simulated atrophy, with respect to the initial timepoint T0.

A simulation of normal aging atrophy which uses BrainWeb images has been realized by Camara *et al.* in [15]. We use this simulation to validate our global atrophy measurement which is obtained by the difference of the BPF (see equation 3) between two timepoints. Figure 1 shows the measured atrophy versus the simulated atrophy for our method without modification for longitudinal study, for our method

with modification and for SIENA. To realize this experiment SIENA have been use with its default parameters.

We can observe that our method without modification for longitudinal study yield incorrect results. Our method with longitudinal study modifications underestimates atrophy while SIENA overestimates it (a perfect measure will yield a slope equal to one), our measures being closest to the simulated atrophy. Both methods exhibit a good correlation with the simulated atrophy, ours being a little higher ($R^2 = 0.99, p < 0.05$). It should also be noticed that the linear approximation resulting from our measures goes through the origin (the point $(0, 0)$), which is expected, while SIENA does not. This could suggest a slight superiority of our method with respect to SIENA, but this has to be confirmed with further experiments.

4. CONCLUSION AND FUTURE WORK

The proposed method allows to obtain repartition maps of the different brain compartments. From these repartition maps which give the proportion of the different compartments in each voxel, the volume of CSF and of the brain (GM+WM) can be computed. These volume evaluations are robust to artefacts like PVE and MS lesions in the images thanks to the inclusion of PVE classes and of an outlier class in the classification process. From these robust and reliable volumes computation, a BPF can be computed. In the case of longitudinal study, the difference of the two obtained BPF give an atrophy value which has been shown to be strongly correlated to the real brain atrophy (on simulated data). In this case, the BPF was not less precise than SIENA.

Because of the lack of simulated image with MS brain atrophy, the next step will consist in comparing of the atrophy measurements of our method against other methods (e.g. SIENA) on real MS patient MR images. To that end, we are currently collecting MR images from a multi-center study. For the moment, we do not have enough patients with several acquisitions to present statistically significant results. Preliminary results on a few patients are promising. Obtained segmentations were qualitatively validated by experts and the obtained volumes and atrophy measurements are realistic according to the literature [16, 3].

Acknowledgments

We are very thankful to Oscar Camara of the Centre for Medical Image Computing (UCL) for providing us with simulated atrophy images.

5. REFERENCES

[1] C.H. Polman, S.C. Reingold, G. Edan, M. Filippi, H-P. Hartung, L. Kappos, F.D. Lublin, L.M. Metz, H.F. McFarland, P.W. O'connor, M. Sandberg-Wollheim, A.J. Thompson, B.G. Weinshenker, and J.S. Wolinsky, "Diagnostic criteria for multiple sclerosis: 2005 revisions to the "Mc Donald Criteria"," *Ann Neurol*, vol. 58, no. 6, pp. 840–6, 2005.

[2] J. Grimaud, N. Pageot, and M. Rovaris, "Parallels between clinical aspects and MRI in multiple sclerosis," *Rev Neurol (Paris)*, vol. 157, no. 8-9 Pt 1, pp. 884–90, 2001.

[3] N. De Stefano, M. Battaglini, and S.M. Smith, "Measuring brain atrophy in multiple sclerosis," *J Neuroimaging*, vol. 17 Suppl 1, pp. 10, 2007.

[4] R.A. Rudick, E. Fisher, J.C. Lee, J. Simon, and L. Jacobs, "Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting ms.," *Neurology*, vol. 53, no. 8, pp. 1698–704, 1999.

[5] S.M. Smith, Y. Zhang, M. Jenkinson, J. Chen, P.M. Matthews, A. Federico, and N. De Stefano, "Accurate, robust, and automated longitudinal and cross-sectional brain change analysis," *Neuroimage*, vol. 17, no. 1, pp. 479–89, 2002.

[6] S. Ourselin, A. Roche, S. Prima, and N. Ayache, "Block matching: A general framework to improve robustness of rigid registration of medical images," in *Proc. of MICCAI'00*. 2000, pp. 557–566, Springer.

[7] J.G. Sled, A.P. Zijdenbos, and A.C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans Med Imaging*, vol. 17, no. 1, pp. 87–97, 1998.

[8] S. Prima, N. Ayache, Tom Barrick, and Neil Roberts, "Maximum likelihood estimation of the bias field in MR brain images: Investigating different modelings of the imaging process," in *Proc. of MICCAI'01*. 2001, pp. 811–819, Springer.

[9] J.C. Souplet, C. Lebrun, P. Clavelou, W. Camu, S. Chanalet, N. Ayache, and G. Malandain, "A comparative study of skull stripping methods in relapsing-remitting multiple sclerosis: Emergence of a new automatic segmentation algorithm," in *ECTRIMS*, Prague, Czech Republic, 2007.

[10] G. Dugas-Phocion, M. Ángel G. Ballester, G. Malandain, C. Lebrun, and N. Ayache, "Improved EM-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI," in *Proc. of MICCAI'04*. 2004, Springer LNCS 3216.

[11] S.M. Smith, "Fast robust automated brain extraction," *Hum Brain Mapp*, vol. 17, no. 3, pp. 143–55, 2002.

[12] B.D. Ward, "Intracranial segmentation," Tech. Rep., Biophysics Research Institute Medical College of Wisconsin, 1999.

[13] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans, "Multimodality Imaging - Design and Construction of a Realistic Digital Brain Phantom," *IEEE Trans Med Imaging*, vol. 17, no. 3, pp. 463–468, 1998.

[14] W.R. Crum, O. Camara, and D.L.G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans Med Imaging*, vol. 25, no. 11, pp. 1451–61, 2006.

[15] O. Camara, M. Schweiger, R.I. Scahill, W.R. Crum, B.I. Sneller, J.A. Schnabel, G.R. Ridgway, D.M. Cash, D.L.G. Hill, and N.C. Fox, "Phenomenological model of diffuse global and regional atrophy using finite-element methods," *IEEE TMI*, vol. 25, no. 11, pp. 1417–30, 2006.

[16] V.M. Anderson, N.C. Fox, and D.H. Miller, "Magnetic resonance imaging measures of brain atrophy in multiple sclerosis," *J Magn Reson Imaging*, vol. 23, no. 5, pp. 605–18, 2006.