Cortical thickness variability in multiple sclerosis: the role of lesion segmentation and filling

Palombit A[1], Castellaro M[4,1], Calabrese M[4], Romualdi C[3], Pizzini FB[2], Montemezzi S[2], Grisan E[1*], Bertoldo A[1]

[1] Department of Information Engineering, University of Padova, Padova, Italy; [2] Neuroradiology and Radiology Units, Department of Diagnostics and Pathology, Verona University Hospital, Verona, Italy; [3]Department of Biology, University of Padova, Padova, Italy; [4] Neurology Section, Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy

[*]enrico.grisan@unipd.it

ABSTRACT

Cortical Thickness (CTh) estimation from Magnetic Resonance Imaging (MRI) data of Multiple Sclerosis (MS) patients is influenced at variable extent by the presence of white matter lesions. To overcome this limitation, several methods were developed. In this study, we evaluate the impact on CTh measurements of different lesion corrections obtained combining three lesion segmentations (manual or automatic) with three intensity filling methods at whole brain and regional scale.

Mean relative CTh differences (MRE) after lesion correction with automatic or manually-based methods was used to evaluate the correction effects, analysing also the impact of segmentation and filling with a factorial analysis of variance.

The estimated CTh was remarkably similar between manually-based (gold standard) and fully automatic corrections, with MRE generally well under 2% in all pairwise comparisons and spatial scale.

Although all the segmentation and filling methods showed an overall good agreement in the CTh, estimation, the results suggest that the lesion filling approach provided with FSL library (FMRIB group, Oxford, UK), regardless of the lesion segmentation method used, deliver an underestimate value of CTh, in the order of 1% of MRE, with respect to other corrections.

Index Terms— *Multiple Sclerosis; cortical thickness; MRI; lesion segmentation; lesion filling*

1. INTRODUCTION

Multiple Sclerosis (MS) is a neurological disease which involves an inflammatory state responsible for axonal myelin destruction and cerebral lesions.

Among the MS potential hallmarks, cortical thinning has recently become a significant biomarker [1] of the disease progression.

The cortical thickness (CTh) can be assessed by analysing structural T1-weighted (T1w) MRI images.

Many methods have been proposed to perform this analysis, mainly following volumetric or surface based [2] approaches.

While in healthy subjects, all proposed CTh estimation methods show comparable accuracy [2], in MS patients, the presence of brain lesions poses a challenge for its correct estimation.

In particular, White Matter (WM) lesions typically appear as hypo-intensities on T1w MR images as depicted in Fig. 1 and, consequently, can affect the performance of CTh estimation methods in two ways.

First, all correction methods require the T1w MRI data to be registered to a common image space (typically the same over which the atlas is defined). The non-linear registration process can be easily misguided by the presence of such hypo-intense lesions.

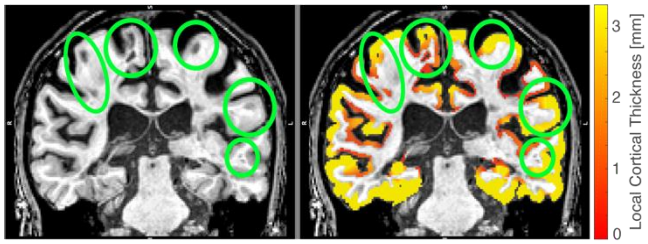

Figure 1. Coronal view of a T1w image in a representative patient (left panel). MS-related lesions appear as hypo-intense areas, hinted in green. The CTh voxel-wise map estimated is superimposed to the T1w image (right panel), highlighting the CTh over- or under-estimation errors due to the presence of lesions.

Secondly, WM lesions next to the cerebral cortex can be easily mistaken with Gray Matter (GM) tissue or cerebro-spinal fluid (CSF) by the CTh estimation algorithms.

One common approach to reduce the bias introduced by the MS lesion presence, consists in the accurate spatial segmentation of WM lesions [3] followed by an intensity filling [4] procedure that replace the intensities within the WM lesion areas with the values of neighbouring normal-appearing WM tissue as represented in Fig. 2.

The Gold Standard (GS) method for MS lesion detection consist, however, in a manual segmentation. Given the subjectivity and the time required by the manual segmentation, a number of automatic alternatives [3] have been proposed.

Preprint submitted to 2017 IEEE International Symposium on Biomedical Imaging.
Received November 7, 2016.

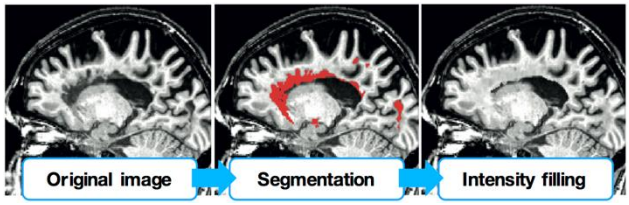

Figure 2. MS lesion correction pipeline: from the original T1w image (left panel) a lesion segmentation procedure is applied providing a spatial lesion map overlaid to T1w image (red areas in the middle panel). panel). Intensity of all voxels inside those lesion areas are then replaced with a normal-appearing one by a suitable filling procedure recovering a normal tissue appearance (right panel).

In this study, we evaluate how CTh estimation is affected by different combinations of lesion segmentation and filling methods. CTh effects due to lesion correction will be assessed both at global or regional brain scale.

2. MATERIALS &METHODS

Data

15 MS patients (12 RR, 3 SP, age 43±5 y, range 36-54 y, M/F: 6/9) were retrospectively selected as a subset of an ongoing study by randomly choosing subjects with a variable amount of lesions. The patients underwent an MRI protocol which included the acquisition of a 3D T1-MPRAGE sequence (TR/TE=8.3/3.7ms, Field of view of 240x240x180 mm, 1 mm^3 isotropic resolution) and a 3D Fluid Attenuated Inversion Recovery (FLAIR, TR/TE=8000/263ms, TI=1650ms, 1 mm^3 isotropic resolution). Data were acquired at the Neuroradiology Unit, Department of Radiology, University Hospital Verona, Italy, using a Philips Achieva 3TX MRI scanner equipped with 8-channel head coil.

The study was approved by the local ethical committee and all patients signed the informed consent.

T1w and FLAIR images were pre-processed with: intensity normalization (*N4BiasFieldCorrection*, ANTs, [5]), skull-stripping (*bet*, FSL, [6]), then FLAIR image was rigidly registered (ANTs [7]) on the T1w image. Finally, as in [8] a brain parcellation of 98 Regions Of Interest (ROI) was obtained with a multi-atlas segmentation approach (*Multi-Atlas Label Fusion*, MALF, [9]).

Lesion segmentation

The reference lesion segmentation was provided by an expert neuroradiologist through manual segmentation of the T1w and FLAIR images of each MS patient. Automatic segmentations were obtained using two automatic methods: *Lesion Segmentation Tool* – LST [10], and Salem Lesion Segmentation - SLS [11].

Lesion filling

Three different methods for replacing the lesion intensities with normal-appearing WM tissue were considered: *lesion_filling* provided by FSL library [12], *Lesion Segmentation Tool* – LST [10] and SLF of the Salem Lesion Filling Toolbox [11].

Cortical thickness estimation

Voxel-wise CTh was estimated using a Diffeomorphic Registration based method (*DiReCT*, ANTs [13]), applied on T1w images corrected with all the combinations of lesion segmentations (manual, LST, SLS) and filling (FSL, LST, SLF). For each patient, a Whole Brain (WB) representative value was obtained averaging the voxel-wise CTh on all the voxels of the cortical ribbon reported as $CTh_{wb,i,j}$, where wb denote the spatial scale, $i \in \{man, LST, SLS\}$ denotes the manual (man) or automatic (LST, SLS) segmentation, $j \in \{FSL, LST, SLF\}$ the filling method used. Similarly a regional estimation was obtained by averaging only the voxels inside single ROIs, represented as $CTh_{r,i,j}$ (where r additionally indexes the ROI as $r = 1, \dots, 98$ pointed).

Statistical Analysis

As first step, we want to assess if different lesion filling on the same T1w image with a fixed lesion segmentation provide consistently different CTh estimates. To this aim, we compute the Coefficient of Variation ($CV_{\%} = \sigma_{inter}/\mu_{inter}$, with σ_{inter} the standard deviation between CTh with different corrections applied and μ_{inter} their mean value) of the CTh estimates both for the WB and ROI level. Then, we want to test if there is any difference among different combinations of three segmentation methods and three filling methods.

CTh differences are assessed by CTh Mean Relative Error as: $MRE_{\%} = avg\left(\left[CTh_{wb/r,i,j} - CTh_{wb/r,man,k}\right]/CTh_{wb/r,man,k}\right)$ where the CTh estimated after correction with a segmentation i and filling j is compared with one estimated after a manual segmentation and filling k (GS-corrected).

Statistical pairwise comparisons of CTh are conducted with permutation test (5000 permutations, typical significance of 0.05 when not specified) and False discovery rate (FDR) criterion to account for multiple comparisons (0.05 rate when not specified) while group-wise comparisons adopted a repeated measures Analysis Of Variance (rANOVA).

The full 3x3 factorial design allows us to study the main sources of CTh variability related to the lesion correction applied before CTh estimation. Taking advantage of this, a two-way rANOVA with segmentation and filling as within-subject design factors. CTh differences between differently corrected images are thus evaluated to find any effect of segmentation or filling factors or their interaction.

All the analysis was repeated at whole brain and regional level in Matlab (R2015b, The Mathworks, Natick, MA).

3. RESULTS

CTh estimation variability

The $CV_{\%}$ of whole-brain CTh estimates considering all three filling methods based on a manual segmentation together, was 0.8%. This percentage increased at 1.31% when averaging the $CV_{\%}$ of ROI-wise estimates with the three GS corrections over all the ROIs. Such small $CV_{\%}$ suggests that no substantial CTh variations should be expected with the filling methods evaluated. Similarly, including also all the

Preprint submitted to 2017 IEEE International Symposium on Biomedical Imaging.
Received November 7, 2016.

other lesion corrections combinations that make use of the automatic lesion segmentations, the CV% was 0.77% at whole brain level and 1.33% at ROI level.

CTh filling sensitivity: whole brain

The group-wise comparison found statistically significant differences ($p<0.05$) between CTh estimates obtained from lesion-corrected images with GS methods as well as comparing all correction methods together. However, all MRE values are consistently low, being 1.3 % at maximum, if comparing corrections based on automatic segmentation with GS ones as in Tab. 1. Note that the MRE values between GS corrections, where CTh differences are referenced by the average of the CTh compared (not reported as not of interest in this study) were always lower than MRE observed between non GS corrections to GS ones (Tab. 1).

The MRE among corrections that make use of the same lesion filling, but different segmentations, was negligible (-0.15% to 0.48%) and never statistically significant (Tab. 1) compared to the MRE with different fillings (-1.2% to 1.3%).

Automatic Segmentation	Filling	Manual Segmentation: FSL	Manual Segmentation: LST	Manual Segmentation: SLF
SLS	FSL	0.13 (0.59)	-1.2** (1.0)	-0.68 (1.5)
SLS	LST	1.1** (1.1)	-0.17 (1.1)	0.32 (1.6)
SLS	SLF	1.1* (1.0)	-0.21 (1.1)	0.28 (1.4)
LST	FSL	0.28 (0.57)	-1.0** (0.81)	-0.53 (1.1)
LST	LST	1.2* (1.0)	-0.15 (0.97)	0.33 (0.85)
LST	SLF	1.3** (1.9)	0.008 (1.7)	0.48 (1.1)

TABLE 1. Whole brain CTh differences between corrections methods assessed with MRE (%) with standard deviation over the subjects in brackets. Significant differences between different combinations ($p<0.05$) were marked with "*", while "**" denote highly statistically significant differences ($p<10^{-3}$). SLS+SLF and LST+LST automatic correction pipelines are coherently highlighted in blue and green.

These higher MRE values among different filling methods held in particular when comparing FSL filling-based corrections to LST or SLF.

On top of those observations there was a statistically significant main effect due to the filling factor ($F(1.42, 14) = 18.8$, $p=9\times10^{-5}$) as suggested by 2-way rANOVA test (with Greenhouse-Geisser correction to account for non-sphericity, tested with Mauchly test).

CTh filling sensitivity: ROI

At ROI level, the MRE results are consistent with those observed at the whole brain (see previous paragraph).

The pairwise comparison analysis (Tab. 2) show an increase in MRE values when comparing different filling methods (-1.6% - 1.8%). MRE values (Tab. 2) with the same lesion filling and different segmentations appear to be consistently

Automatic Segmentation	Filling	Manual Segmentation: FSL MRE (%)	FSL ROI	LST MRE (%)	LST ROI	SLF MRE (%)	SLF ROI
SLS	FSL	0.015 (.6)	14	-1.6 (1.3)	81	-1.2 (1.4)	50
SLS	LST	1.3 (.6)	76	-0.33 (.6)	0	0.043 (.9)	10
SLS	SLF	1.3 (.6)	69	-0.42 (.8)	0	-0.058 (1)	0
LST	FSL	0.34 (.3)	31	-1.3 (.9)	91	- 0.9 (1.1)	60
LST	LST	1.5 (.8)	77	-0.22 (.5)	0	0.14 (.7)	0
LST	SLF	1.8 (1)	78	0.047 (.5)	0	0.39 (.7)	10

TABLE 2. ROI-wise CTh comparisons. MRE column report the relative MRE percentage (standard deviation along ROIs in brackets) between each gold standard and automatic segmentation based correction. ROI columns propose the number of cortical areas (up to 98) that exhibit statistically significant CTh differences (FDR-corrected, 0.05 rate) between compared corrections. SLS+SLF and LST+LST automatic correction pipelines are coherently highlighted in blue and green.

smaller (-0.33% to 0.39%) than the pairwise MRE with different fillings (-1.6% to 1.8%).

The same applies for the number of regions which exhibit statistically significant differences (FDR-corrected) among compared methods. Strikingly, FSL filling with any automatic segmentation provide a high number of regions with significant differences compared to the manual segmentation counterpart and considerably more regional differences than any other combination of SLS-LST segmentations with LST-SLF fillings compared.

Moreover, LST and SLF filling, in combination with any segmentation method (SLS or LST), appear to perform corrections that provides a good agreement of CTh estimates with the relative GS correction, in fact most of the ROIs do not exhibit significant statistical differences.

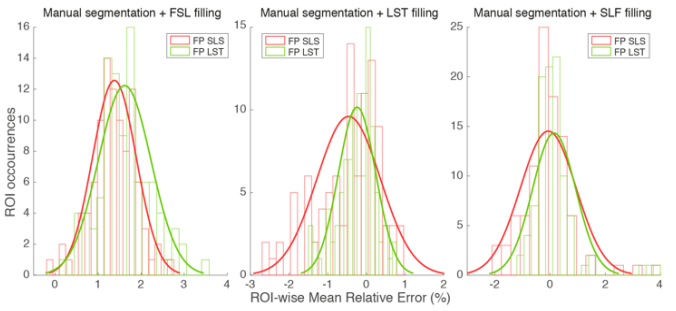

Figure 3. MRE (%) distribution over the ROIs between two literature available automatic corrections (LST as *FP LST*, SLS+SLF as *FP SLS*) and manual segmentation respectively combined with FSL filling (left panel), LST filling (center panel), SLF filling (right panel).

Despite some regional variability, all correction methods are in overall agreement since the maximum observed MRE of 2-3% is quite comparable with the inter-subject CTh regional variability. The MRE distribution over the ROIs, dedicated to automatic corrections (SLS+SLF and LST+LST) against GS ones (manual+SLF and manual+LST) is depicted in Fig. 3.

Filling factor represent a statistically significant main effect in most (92/98, FDR-corrected) of the ROIs as given by 2-way rANOVA test analysis (Greenhouse-Geisser correction

Preprint submitted to 2017 IEEE International Symposium on Biomedical Imaging.
Received November 7, 2016.

accounted for non-sphericity, tested with Mauchly test).

4. DISCUSSION

Generally, MS lesion presence is expected to increase the CTh estimate in the immediate neighbouring of a lesion.
In fact, as in T1w images MS lesions typically appear iso to ipo-intense to GM areas, they may be incorrectly considered part of the cortical ribbon, thus increasing the local CTh.
To tackle this effect, the lesion correction aims at recovering a normal appearing tissue intensity, removing (at least locally) this source of bias in the CTh estimation procedure.
As highlighted by the low $CV_{\%}$ observed, the CTh estimation variability due to different segmentation and filling methods, is overall limited and hardly increased when considering averaged regional estimations instead of whole brain CTh.
The role of lesion filling is dominant in explaining the differences on the CTh estimation since it accounted for significantly more MRE variability than the segmentation used. This was confirmed by the 2-way rANOVA test.
To note is that, this observation holds similarly for the whole brain CTh and the regional CTh. In particular, all MRE values concerning FSL filling corrections were significantly higher, regardless of segmentation and filling compared to it. Moreover, SLF and LST filling with manual segmentation provided lower CTh estimates than FSL filling coupled with any automatic segmentation. This suggest that generally FSL filling provide a correction which leads to a CTh underestimation both globally and regionally. This is also consistent to FSL filling algorithm behaviour which corrects all lesions with white matter T1w intensity.
The limited MRE variance over the MS subjects as well as the ROIs, similar among all pairwise correction comparisons, suggest a limited dependency over the specific lesion correction. Proposed automatic correction pipelines LST+LST [10], SLS+SLF [11] provided very low MRE compared to their the respective GS counterparts (manual+LST, manual+SLF). The MRE distribution over the ROIs (Fig. 3) of LST+LST and SLS+SLF methods exhibit consistently similar variability and shape when compared to GS counterparts. A positive consistent MRE bias around 1% appear in most of the ROIs when FSL filling is compared against, even thought with MRE overall limited under 2 to 3 % in most regions.

5. CONCLUSIONS

In this study, we evaluated how different combinations of lesion segmentation and filling methods affects the cortical thickness estimation in presence of MS-related lesions.
Before estimating CTh, T1w images were processed combining a segmentation and a lesion filling method out of three possible lesion segmentation with three lesion-filling methods from which CTh differences related to the applied correction were assessed. The major findings of this study are that the CTh estimated after lesion filling on manually segmented lesion areas are consistently similar to fully automatic correction methods.
The observed mean relative errors between CTh after automatic corrections compared to gold standard ones ranged from –1.6 % to 1.8 % in most of the regions, as well as at whole brain. Thus, the minimal expected CTh variability as solely given by the lesion correction, is in the order of magnitude of 2% in terms of MRE. Overall, the intensity filling step of the lesion correction was very significant as it accounted for most of the CTh differences observed regardless of the segmentation used.

6. REFERENCES

[1] M. Calabrese, et al., "Regional Distribution and Evolution of Gray Matter Damage in Different Populations of Multiple Sclerosis Patients.," *PLoS One*, vol. 10, no. 8, p. e0135428, 2015.

[2] M. J. Clarkson, et al.,, "A comparison of voxel and surface based cortical thickness estimation methods," *Neuroimage*, vol. 57, no. 3, pp. 856–865, 2011.

[3] D. García-Lorenzo, et al., "Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging.," *Med. Image Anal.*, vol. 17, no. 1, pp. 1–18, Jan. 2013.

[4] S. Magon, et al., "White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study," *BMC Neurosci.*, vol. 15, no. 1, p. 106, 2014.

[5] N. J. Tustison, et al., "N4ITK: improved N3 bias correction.," *IEEE Trans. Med. Imaging*, vol. 29, no. 6, pp. 1310–20, Jun. 2010.

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

[7] B. B. Avants, N. J. Tustison, G. Song,et al., "A reproducible evaluation of ANTs similarity metric performance in brain image registration.," *Neuroimage*, vol. 54, no. 3, pp. 2033–44, Feb. 2011.

[8] M. Calabrese, et al., "Epilepsy in multiple sclerosis: The role of temporal lobe damage.," *Mult. Scler.*, Jun. 2016.

[9] Hongzhi Wang, et al., "Multi-Atlas Segmentation with Joint Label Fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 611–623, Mar. 2013.

[10] P. Schmidt, et al., "An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis," *Neuroimage*, vol. 59, no. 4, pp. 3774–3783, Feb. 2012.

[11] E. Roura, et al., "A toolbox for multiple sclerosis lesion segmentation.," *Neuroradiology*, vol. 57, no. 10, pp. 1031–43, Oct. 2015.

[12] M. Battaglini, M. Jenkinson, and N. De Stefano, "Evaluating and reducing the impact of white matter lesions on brain volume measurements.," *Hum. Brain Mapp.*, vol. 33, no. 9, pp. 2062–71, Sep. 2012.

[13] S. R. Das, B. B. Avants, M. Grossman, and J. C. Gee, "Registration based cortical thickness measurement.," *Neuroimage*, vol. 45, no. 3, pp. 867–79, Apr. 2009.

Preprint submitted to 2017 IEEE International Symposium on Biomedical Imaging.
Received November 7, 2016.