Comparison of tissue classification models for automatic brain MR segmentation

D. Ribes[1,2], B. Mortamet[1], M. Cuadra Bach[3], C. Jack[4], R. Meuli[5], G. Krueger[1], and A. Roche[1]
[1]Advanced Clinical Imaging Technology, Siemens Medical Solutions-CIBM, Lausanne, Switzerland, [2]Radiology, UNIL, Lausanne, Switzerland, [3]Signal Processing Laboratory (LTS5), EPFL, Lausanne, Switzerland, [4]Mayo Clin, Rochester, MN USA, [5]Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland

Introduction: Structural magnetic resonance imaging (MRI) has potential to be a sensitive tool for providing diagnostic and prognostic biomarkers of degenerative and inflammatory brain diseases [1-3]. Recent work demonstrated the feasibility of creating a disease classifier based on grey matter (GM) posterior probability maps (ppms) to distinguish Alzheimer's disease from normal aging in an individual scan [4]; however, the reliability of such maps can strongly be influenced by the statistical model used for segmentation [5]. We compared the ability of three algorithms (**SPM8** [6], **VBM8** [7] and a variational expectation-maximization tissue classification **VEMTC** [8], developed in-house) to automatically extract white matter (WM), GM and cerebrospinal fluid (CSF) ppms and to differentiate between age and disease. Data was obtained from two databases, the **A**lzeimer's **D**isease **N**euroimaging **I**nitiative (ADNI) [9] and a local clinical MRI scanner (collections further referred as D1 and D2, respectively), comprising of young controls, elderly controls and Alzheimer subjects. Global volumetric measurements and spatially normalized GM ppms extracted by each algorithm were compared.

Materials & Methods: For a detailed description of each tissue classification algorithm we refer the reader to [6-8]. The two main differences summarized in table 1, are: i) **VEMTC** uses the same statistical model as **VBM8** [10], but computes tissue ppms with a numerical scheme known to have better convergence properties [8] ; and ii) **VBM8** and **VEMTC** are pure intensity-based approaches whereas **SPM8** is referred to as a prior-based approach. *Data:* 282 D1/D2 head scans (73±16.5 yrs, 120 normal elderly (CN), 95 Alzheimer (AD)/ 44.6±16.8 yrs, 67 CN) were acquired on 6 different Siemens systems (1.5T and 3.0T) using 3D MPRAGE [11] (GRAPPA factor 3 for D2 scans) and operating under various software and hardware combinations. All images were submitted to successive corrections for the B1 receive field, gradient distortion and intensity non uniformity prior to being passed to **VEMTC**, **SPM8** and **VBM8** algorithms. No further intensity non uniformity correction was applied otherwise default parameters were applied. WM, GM and CSF volumetric measurements were estimated by summing the ppms of each class in subject space (figure 1 top). The extracted GM ppms were spatially normalized using DARTEL [12] and logit transformed. The algorithms were compared on four points: 1) volumetric differences across algorithms; 2) volumetric differences between CN and AD for each algorithm; 3) voxelwise differences in spatially normalized GM ppms between algorithms (figure 1 bottom left); and 4) voxelwise differences between AD and CN for each algorithm (figure 1 bottom right). Voxelwise statistical analyses were performed using **SPM8** mass univariate general linear model with age, disease and a constant baseline as regressors. Corrections for multiple comparisons were performed by controlling the family-wise error rate at $p_{FWE}<0.001$.

	Atlas Priors	Markov Prior
SPM8	✓	✗
VBM8	✗	✓
VEMTC	✗	✓

Table 1. Summary of main differences between **SPM8**, **VBM8** and **VEMTC.**

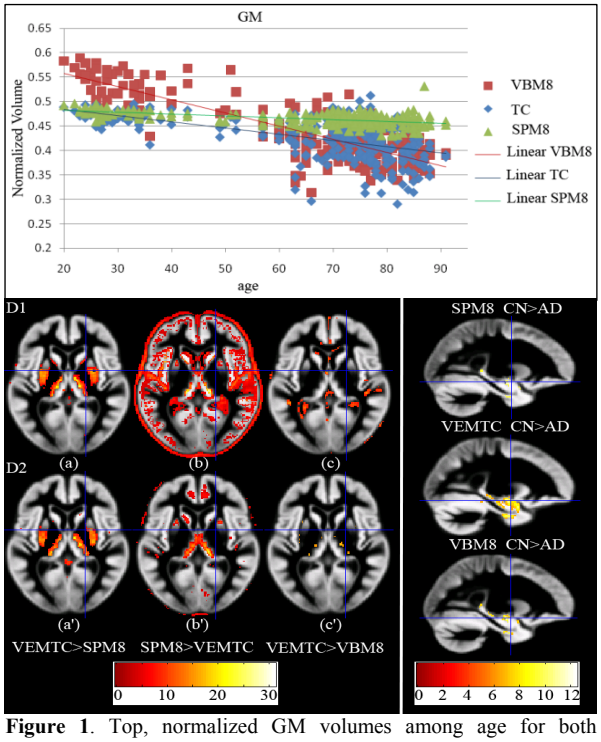

Figure 1. Top, normalized GM volumes among age for both databases obtained by each algorithm. Bottom left, significant GM density differences across algorithms. Bottom right, significant GM density differences between CN and AD (pFWE<0.001).

Results: Qualitative results, showed smaller variations in GM global volume across subjects for **SPM8** (see figure 1 top). Significant differences between algorithms for CSF and GM volumes were found for D1 and D2. Age and disease showed a significant effect on volume differences. Significant CSF and GM volume differences were found between AD and CN only for **VEMTC** and **VBM8**. In figure 1.a and 1.a', both **VEMTC** and **VBM8** output significantly higher GM probabilities than **SPM8** in the central nuclei, for both databases (**VBM8**>**SPM8** not shown). In contrast, **SPM8** output significantly higher GM probabilities than both **VEMTC** and **VBM8** in the cortex, for the database D1. In D2, fewer numbers of cortical regions were found to have higher GM probabilities for **SPM8** compared to **VEMTC** and **VBM8** (figure 1.b and 1.b', **VBM8**>**SPM8** not shown). In figure 1.c and 1.c', significant differences in GM probabilities between **VBM8** and **VEMTC** were observed at the interface between brain and non brain tissues (Skull, veins) and between GM and CSF (e.g. ventricles). Finally, in figure 1 bottom right, higher GM probabilities were observed in the hippocampal region when testing CN>AD for **VEMTC**, **SPM8** and **VBM8**, yet with different significance levels.

Discussion: We mainly suspect differences between **SPM8, VBM8** and **VEMTC** arise from the different segmentation priors. The **SPM8** algorithm uses atlas-based priors built from a group of healthy subjects [12]. We hypothesize that this forces a more homogeneous GM distribution than with pure intensity-based approaches; an effect that may amplify with age (reduced similarity to template) and may explain the lower observed differences between CN and AD volumetric measurements using **SPM8**. When comparing **SPM8** with either **VBM8** or **VEMTC**, differences in the GM probabilities are mostly located in regions strongly affected by partial voluming (e.g. central nuclei) indicating that **SPM8** GM probabilities are dominated by prior information in those regions. Our Results suggest that both **VEMTC** and **VBM8** are more sensitive than **SPM8** to age-related and atrophic GM changes. Future work will determine whether the volumetric differences between **VEMTC** and **VBM8** are due to the different numerical schemes used to compute respective ppms [8], or to the morphological post-processing applied by **VBM8** [7]. We conclude that different segmentation algorithms lead to differences in tissue volume and probability estimations. Moreover, pure intensity driven approaches may be better at detecting disease-related tissue changes. Nevertheless, **SPM8** has already been proved to be able to detect atrophy between two groups of subjects [3] and useful for disease classification [4]. Future studies should evaluate volumetric and GM concentration differences with regard to the priors used, and also sensitivity and specificity of disease classifiers with regard to the algorithms and priors used.

References: [1] Chertkow *et al.*, Can. J. Neurol. Sci. 34(1):577-83 (2007);[2] Nestor *et al.*, Brain 131: 2443–54 (2008);[3] Whitwell *et al*. AONE, V64 i8. 1130 (2007);[4] Klöppel *et al.*, Brain 131 (3): 681-689 (2008); [5] Bach *et al.*, IEEE Trans. Med. Imag, 24(12):1548 – 1565 (2005); [6] Ashburner *et al.*, NeuroImage 26(3):839-851 (2005);[7]http://dbm.neuro.uni-jena.de/vbm8/VBM8-Manual.pdf; [8] Ribes, D. *et al.*, On the convergence of EM-like algorithms for image segmentation using Markov random fields. Submitted (2010). [9]Alzheimer's Disease Neuroimaging Initiative: www.loni.ucla.edu\ADNI; [10] Zhang *et al.*, IEEE Trans. Med. Imag, 20(1):47-57 (2002) [11] http://www.loni.uclaedu/ADNI/Research/Cores/ [12] Ashburner *et al.* NeuroImage 38(1):95-113 (2007).

Acknowledgement: data collection and sharing for this study was founded by ADNI (PI: Michael Weiner; NIH grant U01 AG24904). This work was supported by CIBM of the UNIL, UNIGE, HUG, CHUV, EPFL and the Leenaards and Jeantet Foundations.