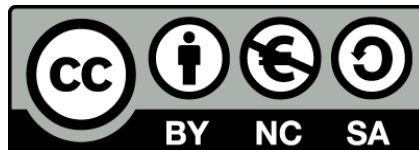UNIVERSITAT DE
BARCELONA

Computational analysis of schizophrenia: Implementation of a multivariate model of anatomical differences
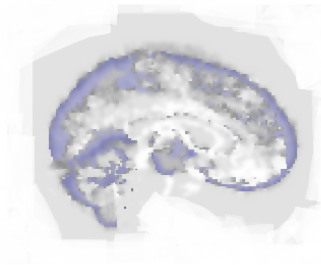
Gemma C. Monté Rubio

University of Barcelona UB

Faculty of Medicine – Biomedicine Doctorate Program

October 2015

Computational analysis of schizophrenia: Implementation of a multivariate model of anatomical differences

Gemma C. Monté Rubio

Supervisors:

Prof. John Ashburner

Wellcome Trust Centre for Neuroimaging, UK.

Dr. Carles Falcón

University of Barcelona, Spain.

Dra. Edith Pomarol-Clotet

FIDMAG Research Foundation, Spain.

A la meva mare Mei,
sense tu saps que no ho hagués aconseguit.
I a la meva filla Ariadna,
aquí tens on he invertit el temps que t'he robat.

Agraïments

Aquesta tesi no només consisteix en el que tot seguit està escrit, més enllà del text simbolitza una part important de la meva vida, i concloure-la es converteix en un acte simbòlic: enllestir un període per començar-ne un de nou. En l'àmbit científic i professional m'ha enriquit moltíssim, però en l'àmbit personal també ha fet molt per mi. Per bé, ja no sóc la mateixa que la va començar.

He d'agrair de tot cor el suport que he rebut durant aquests anys a moltes persones. En primer lloc als meus directors: el Dr. Carles Falcón, que sempre (des que ens vam conèixer, molt abans del doctorat i davant adversitats) ha estat al costat i molt especialment durant la tesi, fent de guia, donant suport i sent un amic en el sentit més ple de la paraula. No sé com podré agrair-te tot el que has fet! Al Prof. John Ashburner, brillant científic i bellíssima persona, qui va apostar per mi de bon començament amb totes les de la llei, i que ha estat al peu del canó fins al darrer moment. John, thanks for believing in me, and also for those moments that we shared with your wonderful family, Hester and little Sophie, we have won a friendship! I la Dra. Edith Pomarol-Clotet, qui sempre ha estat receptiva, donant ànims i suport més enllà del que era establert. Sense el teu suport i el del Dr. Salvador Sarró, no hauria fet l'estada a Londres, entre altres coses!

A més de directora, l'Edith ha estat la meva cap durant cinc anys a la FIDMAG, així com el Prof. Peter Mckenna, qui com ella m'ha aportat molt. I enjoyed our exchange sessions about science and philosophy. A la FIDMAG he viscut un període molt enriquidor del que he tret moltes coses, el millor de tot: les persones. He conegut companys de qui he après de neuroimatge i m'he endut amistats que porto al cor. Eulàlia, Alicia i Sònia, Bàrbara, Jesús, Bene i Ramón, Èlia, Rumi i Ali, Erick, Rai i Quim, Salva gràcies per recolzar-me, i tenir sempre paraules bones i de suport, Pi, Silvia, Ana i sobretot la meva Maria sabeu que us estimo molt.

Acabo d'estrenar etapa a l'IDIBAPS per a la Unitat de Demències del Clínic, amb la gent de la plataforma de Neuroimatge del CEK i de la Fundació Pasqual Maragall. Des del primer moment m'he sentit acollida per un grup de gent integradora, plena de motivació i inquietuds. Gràcies Oriol, Isa i a tots per fer-m'ho tan fàcil. Sobretot vull agrair al meu cap, el Dr. José Luis Molinuevo que m'hagi donat aquest periode per enllestir la tesi. Mil gracias por tu generosidad, tu paciencia y tu calidad humana. I al Dr. Juan Domingo Gispert, gràcies per facilitar tant aquesta tesi i fer-me costat. Els inicis treballant amb tu i el teu equip han estat estimulants i prometedors d'una etapa que començo amb il·lusió.

Fora de l'àmbit professional bons amics m'han donat l'energia, m'han ajudat a buscar la motivació i no m'han deixat caure, persones increïbles que sempre hi son, que sàpiguen que a mi també em tenen: En Giuseppe i l'Elisabeth, dues grans persones que ens acullen sempre amb els braços oberts i ens ho donen tot, sou de la nostra familia! Svetlana, thanks for your warmth and help, our friendship began with this thesis but will go beyond it. Emma, Isabel, Montse, Roser i Mariam sou molt especials! Fefa, repartir tareas del día a día contigo ha sido un respiro. Marifí, Angel Luis i a tots aquells que m'heu encoratjat, gràcies pel vostre suport.

Finalment, com no pot ser de cap altra manera, a la meva família, la meva germana Marina i els meus pares, de qui em sento tan orgullosa que no hi cap en aquestes línies, les millors persones que conec, un model a seguir. Gràcies perquè durant tota la meva vida heu estat una font de recolzament i amor incondicional. També a la Laura, I als meus tiets i cosines, amb qui junts fem pinya davant l'adversitat. I al meu marit i la meva filla, que em donen tot l'amor i em fan mirar de ser cada dia millor persona, gràcies per recolzar-me i no retreure'm ni un segon de tot el temps que us he pres per donar-li a aquesta tesi.

A tots els que m'heu ajudat a arribar a aquest dia: gràcies, perquè per vosaltres és que no he abandonat.

Abbreviations

AUC	Area Under the Curve
BET	Brain Extraction Tool
BF	Bayes Factor
BG	Background tissue
BMI	Body mass index
CSF	Cerebrospinal fluid
CV/LOO-CV	Cross Validation / Leave-one-out Cross Validation
DARTEL	Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra
FAST	FMRIB's Automated Segmentation Tool
FSL	FMRIB Software Library
FWHM	Full-width at half maximum
GM	Grey matter
GM(+)/(-)	False positives / false negatives
GP	Gaussian Processes
GS	Gold standard
LDDMM	Large Deformation Diffeomorphic Metric Mapping
NS	New Segmentation
RMS	Root Mean Squared
SPM	Statistical Parametric Mapping
SVE	Segmentation Validation Engine
TIV	Total Intracranial Volume
US	Unified Segmentation
VBM	Voxel-Based Morphometry
WM	White matter

Contents

1. Introduction

Schizophrenia is a chronic mental disease that is characterised by a distorted interpretation of reality. This disabling brain disorder can be treated, but even though it helps, most patients have to cope with the symptoms throughout all their lives (Patiny, Constant, and Symann 2015; de Paula et al. 2015). In such patients, the clinical picture is often undifferentiated at first presentation, and it is unclear which of them are at most risk of developing chronic schizophrenia or bipolar disorder. Categorization of mental diseases remains controversial, because current diagnoses depend on imprecise categorical distinctions, with arbitrary cutoffs (Thaker 2008). This is a problem for the patient, who spends longer without appropriate treatment with the consequent decline. It is also a problem due to the additional costs, both social and economic, arising from patients who are not correctly treated. Current diagnostics regarding schizophrenia and other mental diseases are based predominantly on phenomenological criteria and are not supported by biomarkers. The identification of relevant disease biomarkers and testing them as novel targets for treatment are of crucial interest for the field (Ivleva et al. 2013).

Chronic schizophrenia has been widely studied, and there are consistent findings that have allowed the characterization of the anatomical pattern associated with this disease. The neuroimaging field has contributed to this knowledge. Structural Magnetic Resonance Imaging (MRI) is a non-invasive imaging modality that allows the study of *in vivo* morphological features in healthy and impaired brain anatomy. Many works in the literature have shown that structural MRI data contribute in research and can aid in clinical assessment (Feinstein et al. 2004; Frisoni et al. 2010). Several studies have contributed knowledge about the pattern of damage associated with the disease (Honea et al. 2005; Bora et al. 2011; Ellison-Wright and Bullmore 2010), and how in these areas abnormalities of brain function and axonal diffusion have also been found (Pomarol-Clotet, Canales-Rodríguez, et al. 2010; Salgado-Pineda et al. 2011).

Morphologically, schizophrenia is characterized by a reduction, of around 2%, in whole brain volume. Locally, larger reductions in regions such as the frontal lobe, in about 50% of VBM studies, and hippocampus, insula, temporal and parietal cortices, in about 20%, are reported (Honea et al. 2005). Other regions are also strongly involved in the disease, such as the medial frontal cortex (Bora et al. 2011), which has been proposed as a key region. Ellison-Wright and Bullmore (Ellison-Wright and Bullmore 2010) performed a meta-

analysis of 42 VBM studies of schizophrenia, where they found that the distribution of GM volume reductions was wider in schizophrenia than for other mental diseases, such as bipolar disorder, affecting frontal, temporal, cingulate, insula and thalamus. Postmortem studies support all these findings (Brown et al. 1986; Pakkenberg 1987; Bogerts et al.). All this information leads to a conclusion that GM is a tissue highly involved in the disease, and that the pattern of brain damage may be more global than neuroimaging findings so far suggest, hence that this is not sufficient.

Finding specific morphometric alterations associated with a particular disease is a widespread goal in neuroimaging research. It has been performed in hundreds of studies, mainly by applying the Voxel-Based Morphometry (VBM) technique (Wright et al. 1995; Ashburner and Friston 2000; Ashburner and Friston 2001) for comparing brain anatomies. In brief, VBM involves preprocessing the original MR images, by segmenting them into different tissue types. These are then aligned into the same anatomical space. Next, a statistical analysis is performed to enable a comparison among populations. VBM has contributed enormously to the field of neuroscience, and in particular to the knowledge about schizophrenia. However, it is a mass-univariate approach, which assumes voxels are independent. The interconnected nature of the brain suggests that this may not be the most biologically plausible assumption to make. Many neuroimaging advances now focus on multivariate analysis frameworks, with many recent developments based on pattern recognition and other machine learning approaches. Powerful machine learning techniques from other fields have been adopted to obtain a more accurate understanding of the different processes that occur in the brain. Having a more accurate model may allow more rapid translation from basic research into clinical applications. This is the role of multivariate techniques in neuroimaging (Ashburner and Klöppel 2011). Such applications yield interesting predictions based on more accurate characterizations of differences between populations of subjects (Schrouff et al. 2013; Sabuncu and Konukoglu 2014). In the last ten years, pattern recognition techniques have been widely applied to structural data, mainly for predicting clinical status at the individual level (Klöppel et al. 2008; Costafreda et al. 2009; Klöppel et al. 2012; Nieuwenhuis et al. 2012; Mourao-Miranda et al. 2012).

Nowadays, brain imaging researchers aim to collect the largest possible number of subjects. Studies attempt to obtain findings that can generalize, for instance differences between populations or the impact of a biomarker, etc. Successful multivariate methods do not necessarily localize differences, but instead aim to capture the patterns of difference that best separate subjects into groups, or predict some continuous variable of interest. Although useful for analyses that attempt to localize differences, the widely used

assumption of independence among the anatomy of different brain regions is not really biologically plausible. If this assumption is removed, greater predictive accuracy may be possible. For example, age or gender differences are not localized to any particular brain region. Instead, there is a pattern of differences that are distributed throughout the brain. Studies based on localizing differences only show the most relevant aspect (like the tip of an iceberg) but they can lose important information pertaining to patterns of relationships among brain regions. A pattern recognition approach attempts to learn a relationship between feature data and their corresponding labels. The algorithm, after learning such a relationship, should be able to predict the label for new data cases.

Many different algorithms for pattern recognition analysis are available (Schrouff et al. 2013). For this thesis, Gaussian processes approach for classification and regression was chosen (Rasmussen and Williams 2006). Gaussian Processes (GP) are kernel-based approaches, set in a Bayesian framework. They achieve similar performance to Support Vector Machines (SVM) for neuroimaging data (Schrouff et al. 2013) with the advantage that they make probabilistic predictions. These supervised algorithms learn the mapping between the input (data features) and its output (labels) from a set of training data. Depending on whether the output is continuous or discrete, it would be respectively a classification or a regression problem. GP processes are more thoroughly presented in section 1.3.

There are many different procedures to model data for classification and regression. Most of these approaches involve pre-processing structural MRI scans in the same way as for a conventional VBM analysis (Wright et al. 1995; Ashburner and Friston 2000) but then applying a pattern recognition technique. These kind of analyses require some form of characterization of inter-subject neuroanatomical variability (more detailed explanation about these characterizations from MRI preprocessed data in sections 1.2 and 1.4). Much of this variance among brain images can be characterized by shape modeling (computational anatomy), where the accuracy of inter-subject registration plays a significant role in terms of the findings and their interpretability (Ashburner and Klöppel 2011). Conclusions from any particular study depend heavily on how the data are modeled, and the assumptions underlying those models. If data are imprecisely modeled or the characterization used does not incorporate key information, this may result in poor predictions. Obtaining informative image features is an area of interest for many using pattern recognition in clinical research.

The use of suboptimal features for pattern recognition limits the accuracy with which predictions may be made. Each feature encodes information about biological processes, mental or neurodegenerative diseases, etc. However, each process or specific disease

involves and alters each bit of brain anatomy in different proportions. No Free Lunch theorem states that prior knowledge about what is likely to be informative should be essential for pattern recognition. This scenario leads to the strong motivation of exploring different features, how they encode information and whether it is possible to generalize about them.

Regarding accuracy in modeling, it is being established step-by-step in the VBM-type preprocessing. In VBM, the outputs from segmentation are often used to drive image registration. If the segmentation does not work accurately, the next step cannot be accurate either. Hence the interest in the accuracy of such automated computational tools is increasing.

Hundreds of studies have used structural MRI data, mostly for VBM (Wright et al. 1995; Ashburner and Friston 2000). Pre-processing structural MRI for VBM usually entails applying a pipeline of different algorithms to model the images. First step involves segmenting the images in order to identify the tissues of interest. These are mainly grey matter (GM) and white matter (WM) tissue classes. Tissue classes are then spatially normalized and smoothed to enable statistical comparisons. Reliable findings depend on reliable registered tissue classes. In VBM, the outputs from segmentation are involved in registration. If segmentation is inaccurate, the next step cannot be accurate either. Hence the accuracy of the final preprocessed data is dependent on each step of the processing.

Several questions arise when a VBM has to be carried out. The first one is about which software to use. In the neuroimaging field, the two most widely used software packages are FSL (Analysis Group, FMRIB. Oxford, UK) and SPM (Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology. London, UK). Both, FSL and SPM, have implemented different approaches for VBM pre-processing. These share some commonalities, but there are also a number of differences. For image segmentation, the SPM package implements a method based on fitting mixtures of Gaussians to the image intensities (Ashburner and Friston 2005). This is combined with an atlas-guided approach, in which a warping from the atlas to each image is performed. In that way, the characteristic space of intensities and the spatial information of the image are both combined. The approach in the FSL package is similar, but also includes a Markov Random Field, which models interactions between voxels and their nearest neighbors (Zhang, Brady, and Smith 2001). There are many relevant differences between these two options, although there are some commonalities. Both SPM and FSL are assembled in a Bayesian framework. In both cases the problem is formulated as MAP estimation. Finally, the model fitting is performed by the Iterative Conditional Modes and the Expectation Maximization

algorithms. These two approaches for brain image segmentation are briefly described, with emphasis on commonalities and differences in section 1.1.

Each of these two packages for VBM-type preprocessing is based on assumptions and, even though their reliability is assumed, results may differ. This leads to the consideration that differences in segmentation and modeling may be biased in some way. A strategy to address this situation involves testing the reliability of each of those software tools. Reliable findings depend on reliably registered tissue classes. This has been explored for a wide range of algorithms for brain image segmentation, and using different strategies on available datasets (Tsang et al. 2008; Helms et al. 2006; Bouix et al. 2007; Lalaoui L. & Mohamadi T. 2013; Valverde et al. 2015). Accurate brain tissue segmentation is crucial for reliable VBM or pattern recognition analyses.

One of the main purposes of the current thesis is the application of improved methodologies for clinical research. Obtaining trustworthy features from structural MRI data, regarding the accuracy in modelling and the information encoded, becomes a need for pattern recognition studies in schizophrenia and other diseases. A step in this direction would contribute to the implementation of these techniques in daily clinical practise.

1.1   Algorithms for brain image segmentation

Most segmentation algorithms give results that are acceptable enough to be used by many researchers around the world. However, tissue class images obtained from different algorithm models can differ in very subtle ways. It would be exceptional to obtain a pair of identical GM maps, if they have been estimated using two different algorithms.

In this thesis, three different implementations for brain image segmentation:

i.   FAST v4.1 (FMRIB's Algorithm Segmentation Tool) implemented in FSL.
ii. Segment: the Unified Segmentation implemented in SPM8.
iii. New Segment: the New Segmentation implemented in SPM8.

1.1.1 Comparative description of the models steps

SPM	*FSL*
1a.	**1b. Brain Extraction**
SPM does not require brain extraction before initializing any segmentation algorithm.	FAST requires skull stripped images. Brain Extraction Tool (**BET**) is used for this purpose. The **Intensity Histogram** is processed to find the image intensity effective range. Then, a rough threshold t is determined, which attempts to distinguish between brain and background, air and skull.
	Then, the brain surface is modeled by a tessellated surface. Beginning with a tessellated sphere placed in the center of gravity, the algorithm deforms it to identify the optimal surface (Smith 2001).

A common approach for two different ways of segmentation

The **Finite Mixture (FM) model** is set within a Bayesian framework, and used to model voxel intensity distributions. The FM model assumes that information about the properties of voxels is spatially independent and it does not take into account any spatial information.

$$P(y_i \mid k) = f(y_i; \theta_k);$$
$$P(c_i = k) = \gamma_k \qquad (1.1)$$

The intensity value of voxel *i* is denoted by y_i, and *k* represents the index of the Gaussian (cluster), which explicitly is defined by their parameters θ_k. Then, γ_k is the mixing coefficient that accounts for the proportion of voxels that corresponds to this k^{th} Gaussian.

In the Bayesian framework, the probability of an entire data set is given by the next expression,

$$P(y \mid \theta, \gamma) = \prod_{i=1}^{I} \left(\sum_{k=1}^{K} P(y_i \mid c_i = k, \theta_k) \cdot P(c_i = k) \right) \qquad (1.2)$$

With the terms introduced in (1.1), this becomes

$$P(y \mid \theta, \gamma) = \prod_{i=1}^{I} \left(\sum_{k=1}^{K} f(y_i; \theta_k) \gamma_k \right) \qquad (1.3)$$

However, not considering spatial information is a downside.To overcome this issue, bias correction models and some spatial properties, or constraints, have to be incorporated. These implementations are based in two different models that are described below.

3a. UNIFIED SEGMENTATION & NEW SEGMENTATION	3b. FAST
Same generative model involves:	The model involves:

3a. UNIFIED SEGMENTATION & NEW SEGMENTATION

Same generative model involves:

- Tissue class segmentation
- Registration
- Bias correction

Using a **Finite Mixture (FM) model**

(1) **Spatial information** is included by incorporating tissue probability maps (TPMs) into the model by modifying the stationary mixing proportions. In the original "Unified Segmentation" implementation, these TPMs are GM, WM and CSF. In the "New Segmentation" implementation, bone, soft tissue and air are added to the above mentioned tissues.

Ignoring registration, the spatial priors for "Unified Segmentation" are given by

$$P(c_i = k \mid \gamma_k) = \frac{\gamma_k b_{ik}}{\sum_{i=1}^{j} \gamma_j b_{ij}} \qquad (1.4a)$$

For the "New Segmentation" of SPM8, priors are defined by:

$$P(c_i = k \mid \gamma_k) = b_{ik} \qquad (1.5a)$$

Registration. These priors are deformable spatial priors. The deformation parameterization in the original "Unified Segmentation" is by a linear combination of cosine transform bases (Ashburner and Friston 1999). However, the "New Segmentation"uses

3b. FAST

The model involves:

- Tissue class segmentation
- Bias correction

Using the **Markov Random Field (MRF)**

(1) **Spatial information** is incorporated by defining a neighborhood system around each voxel. The MRF properties can be described by factorizing according to

$$P(x) > 0, \forall x \in X; P(x_i \mid x_{S-\{i\}}) = P(x_i \mid x_{N_i})$$

$$(1.4b)$$

By multiplying each term (the clique potentials) is possible to get the Gibbs distribution

$$P(x) = Z^{-1} e^{(-U(x))}; Z = \sum_{x \in X} e^{(-U(x))} \qquad (1.5b)$$

The HMRF in FAST can be described by:

- $X = \{X_i, i \in S\}$ which is the hidden MRF, with prior distribution $p(\mathbf{x})$.

- $Y = \{Y_i, i \in S\}$ that is the observable random field, with emission probability distribution $p(y_i \mid x_i)$ for each y_i.

- The parameter set $\theta = \{\theta_l, l \in L\}$ are used to model the distributions.

a many more parameters, and optimisation involves usingsome of the same technology asfor DARTEL (Ashburner 2007), which has about 700,000 parameters. In both Unified and New Segmentation, parameters are introduced into the previous probabilistic equations (1.1) and (1.2).

Incorporating deformable tissue priors, the likelihood equation for the Unified Segmentation becomes

$$P(y \mid \mu,\sigma,\gamma) = \prod_{i=1}^{I} P(y_i \mid \mu,\sigma,\gamma) =$$

$$= \prod_{i=1}^{I} \left(\frac{1}{\sum_{k=1}^{K} \gamma_k b_{ik}(\alpha)} \sum_{k=1}^{K} \frac{\gamma_k b_{ik}(\alpha)}{\left(2\pi\sigma_k^2\right)^{1/2}} \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$

(1.6a)

For the New Segmentation, it becomes

$$P(y \mid \mu,\sigma,\gamma) = \prod_{i=1}^{I} P(y_i \mid \mu,\sigma,\gamma) =$$

$$= \prod_{i=1}^{I} \left(\sum_{k=1}^{K} \frac{b_{ik}(\alpha)}{\left(2\pi\sigma_k^2\right)^{1/2}} \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$

(1.7a)

Optimization involves minimizing the negative logarithm of these probabilities.

$$\varepsilon = -\sum_{i=1}^{I} \log \left(\frac{1}{\sum_{k=1}^{K} \gamma_k b_{ik}(\alpha)} \sum_{k=1}^{K} \frac{\gamma_k b_{ik}(\alpha)}{\left(2\pi\sigma_k^2\right)^{1/2}} \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$

(1.8a), and

$$\varepsilon = -\sum_{i=1}^{I} \log \left(\sum_{k=1}^{K} \frac{b_{ik}(\alpha)}{\left(2\pi\sigma_k^2\right)^{1/2}} \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$

(1.9a)

(2) **Bias correction** assumes that noise is added to the image and then, scaled by a bias, as:

The spatial dependency is modeled by the neighborhood system $p(l) = \gamma_l \Rightarrow p(l \mid x_{N_i})$, so that modifies the FM model expression

$$p(y_i \mid x_{N_i}, \theta) = \sum_{l \in L} p(y_i, l \mid x_{N_i}, \theta) = \sum_{l \in L} f(y_i; \theta_l) p(l \mid x_{N_i})$$

(1.6b)

Assuming a Gaussian distribution of the intensity, for the entire data set, the likelihood is

$$P(x \mid y) = \prod_{i \in S} \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} - \log\left(\sigma_{x_i}\right) \right) \right) \cdot P(X)$$

(1.7b)

(2) **Bias correction**is based on a method by Wells et al. (Wells et al. 1996). The bias field is a multiplicative vector. Then, the effect of the bias

$$y_i = (\mu_i + n_i)/\rho_i \qquad (1.10a)$$

The bias model is included in the FM model by adding a parameterization ($\rho_i(\beta)$; where β is a vector of unknown parameters). It modifies the likelihood equation to

$$\varepsilon = -\sum_{i=1}^{I} \log\left(\rho_i(\beta) \sum_{k=1}^{K} \frac{b_{ik}(\alpha)}{\left(2\pi\sigma_k^2\right)^{1/2}} \exp\left(-\frac{(\rho_i(\beta)y_i - \mu_k)^2}{2\sigma_k^2} \right) \right)$$
$$(1.11a)$$

field at each voxel is expressed as:

$$I_i = I_i^* \times b_i \qquad (1.8b)$$

But it becomes an additive artifact with the logarithm transformation.

$$\log I = \log I^* + \log b \Rightarrow y = y^* + B \quad (1.9b)$$

The intensity distribution, which it is modeled as a FM model, includes the bias field as

$$p(y_i \mid B) = \sum_{j \in L} \left\{ g\left(y_i - b_i; \theta(j) \right) P(j) \right\} \qquad (1.10b)$$

The objective function becomes

$$P(x \mid y) = P(y \mid x) P(X) = \prod_{i \in S} p(y_i \mid x_i) \cdot P(X) =$$
$$= \prod_{i \in S} \left(\sum_{l \in L} g\left(y_i - b_i; \theta_l(j) \right) P(j) \right) \cdot P(X)$$
$$(1.11b)$$

4.a REGULARIZATION

Regularizing penalizes the model to give more plausible bias fields and nonlinear deformations. The probability density of the spatial parameters that parameterize both the bias field and the registration parameters are assumed to be multivariate Gaussians distributions. Regularization of deformations is based on their "bending energy" (Ashburner and Friston 1999).

4.b REGULARIZATION

This approach does not use any regularization.

Model fitting by ICM and EM algorithms

In both softwares, optimization consists of minimizing their respective objective functions. The **Iterated Conditional Modes** (ICM) algorithm (Besag 1986) is used to estimate each parameter, which is subsequently updated by the **Expectation-Maximization** (EM) algorithm (Dempster, Laird, and Rubin 1977; Bishop, Svensen, and Williams 1998; Neal and Hinton 1998). In this way, when a parameter is updated by the EM algorithm, the remaining parameters are held fixed until

its own turn arrives, alternating one parameter each time.

The optimization is split in two steps the Expectation step and the Maximization step:

i. **E-step**: the algorithm calculates the conditional expectation.
ii. **M-step**: it maximizes the probability to obtain the next estimate.

Intuitively, EM is an iterative method that, given the observed data *Y*, alternates among estimating the unknown parameters *(θ)* and the hidden variable (class label *x*). The algorithm computes a distribution over the *X* space, given an estimate *θ* at each iteration, because both are strongly inter-dependent. The objective is to maximize the posterior probability (1.12) of the parameters *θ* given the data *Y*, marginalizing over *X*.

$$\theta^* = \arg\max_{\theta} \sum_{x \in X} P(\theta, x \mid y) \qquad (1.12)$$

For simplicity, the log probabilities are maximised:

$$\theta^* = \arg\max_{\theta} log P(\theta, Y) = \arg\max_{\theta} log \sum_{x \in X} P(y, x, \theta) \quad (1.13)$$

5.a Optimization	5.b Optimization
The estimation process alternates among optimizing each set of parameters (registration, bias correction, and computing the intensity distributions) while keeping the others fixed (ICM).	FAST does not use any kind of spatial information for segmentation. It does the model may not be robust enough in certain cases. For instance, in those where the bias field has an important effect over the data. The EM algorithm attempts to overcome the downside in its own specific HMRF-EM framework, which incorporates the bias field correction step in the same model.
For parameterization of the deformations, the Gauss-Newton strategy is used. In the Unified Segmentation framework, this is relatively straightforward because the matrices involved are relatively small. In the New Segmentation approach, there are many more registration parameters so the multigrid procedure developed for DARTEL is used instead.	An EM solution is sought for the dependent variables: the bias field, the tissue classification and the model parameters. In the E step, a MAP estimate is made of the bias field and the class label. In the M step, the ML allows to calculate the parameters using the estimated bias field and the class labels obtained in the previous E step. The algorithm scheme can be seen in the next table.
A local optimization procedure is used, so it is necessary to establish reasonably good initial starting estimates. These are randomly	

assigned for thecluster parameters. The coefficients for the bias and the nonlinear registration are set to zero (an affine registration is first used to roughly align with the TPM, by optimising the objective function of D'Agostino et al. (Agostino et al. 2004).

The parameters of interest in this model, $\theta = \{\mu, \sigma, \gamma, \alpha, \beta\}$ are optimized by the EM algorithm (Bishop, Svensen, and Williams 1998; Dempster, Laird, and Rubin 1977; Neal and Hinton 1998). Details can be seen in the Unified Segmentation paper (Ashburner and Friston 2005). The general framework is similar in both the Unified and New Segmentation models.

1. Perform image thresholding for initial parameter estimation and segmentation.
2. Estimate the bias field

$$b_i^{(t)} = \frac{[FR]_i}{[F\psi^{-1}1]_i}, \text{ with } 1 = (1, 1, \cdots, 1)^T$$

3. Calculate the likelihood distribution

$$p^{(t)}(y_i|x_i, B) = g^{(t)}(y_i - b_i; \theta(x_i))$$

4. Estimate the class labels by MRF-MAP estimation

$$\mathbf{x}^{(t)} = \arg\max_{\mathbf{x} \in \mathcal{X}} \{P(\mathbf{y}|x, \theta^{(t)}) + P(\mathbf{x})\}$$

5. Calculate the posterior distribution

$$P^{(t)}(\ell|y_i) = \frac{g^{(t)}(y_i; \theta_\ell) \cdot p^{(t)}(\ell|x_{\mathcal{N}_i})}{p(y_i)}$$

6. Update parameters by

$$\mu_\ell^{(t+1)} = \frac{\sum_{i \in \mathcal{S}} P^{(t)}(\ell|y_i)y_i}{\sum_{i \in \mathcal{S}} P^{(t)}(\ell|y_i)}$$

$$(\sigma_\ell^{(t+1)})^2 = \frac{\sum_{i \in \mathcal{S}} P^{(t)}(\ell|y_i)(y_i - \mu_\ell)^2}{\sum_{i \in \mathcal{S}} P^{(t)}(\ell|y_i)}$$

7. $t \leftarrow t + 1$ and repeat from 2 until enough iterations have been performed.

HMRF-EM algorithm for brain MR image segmentation and bias field correction, from Zhang, Brady & Smith, 2001.

1.2   Diffeomorphic image registration by Geodesic Shooting

One of the main purposes in image registration is the measurement and statistical study of variations in brain anatomy. Image registration generally involves finding the set of parameters that best encode the continuous mapping from one shape to another. There are several approaches to do this, but they can broadly be divided into two main categories: the small-deformation and the large-deformation approaches. The main difference is that the first one does not guarantee that topology is preserved, whereas the second does. Large-deformations, or diffeomorphisms, involve transformations that can be inverted. Composing a forward transform with its inverse produces the identity transform. This is relevant, because during the registration process inversions are required (Ashburner 2007). Hence, large-deformation approaches are desirable to get the best parameterization.

Early diffeomorphic registration was based on a framework that models deformations as a "viscous fluid", such that one shape flows to match another (G. Christensen et al. 1995; G. E. Christensen 1994). More recently, approaches for nonlinear image registration are based on the Large Deformation Diffeomorphic Metric Mapping (LDDMM) described by Beg. et al. (Beg et al. 2005). Given a template image $\boldsymbol{\mu}$, an image target $\boldsymbol{f}$ can be represented as a function of $\boldsymbol{\mu}$. The LDDMM algorithm computes a diffeomorphic transformation φ: $\Omega{\rightarrow}\Omega$ between these images (where spatial domain $\Omega{\subseteq}\Re^3$), such that $\boldsymbol{f} = \boldsymbol{\mu}(\varphi^{-1})$. The curve $\varphi = \phi_t$, defined between $t{\in}[0,1]$, parameterizes the path by means of the ordinary differential equation $\dot{\phi}_t = v_t(\phi_t)$, the diffeomorphism $\varphi = \phi_0 = Id$ is the identity and $\varphi = \phi_1$ is defined as the end point of the path. The equation is dependent on the velocity vector field of the flow of the deformation, $v_t: \Omega{\rightarrow}\Re^3$, and on $t{\in}[0,1]$. The diffeomorphism φ is obtained from

$$\varphi = \phi_1 = \int_0^1 v_t(\phi)dt \qquad (1.14)$$

The objective is to estimate a series of velocity fields v_t, over $t{\in}[0,1]$. The optimal curve of v_t is obtained by optimizing a variational problem. The large-deformation diffeomorphic metric mapping is obtained by minimizing

$$\frac{1}{2}\int_0^1\|Lv_t\|^2dt + \frac{1}{2\sigma^2}\left\|\boldsymbol{f} - \boldsymbol{\mu}(\phi_1^{-1})\right\|^2 \quad (1.15)$$

where $\boldsymbol{L}$ is a differential operator. The first term minimizes the squared distance of the deformation, which is the geodesic shortest path for the metric distance. The second term minimizes the difference between the warped template, $\boldsymbol{\mu}(\phi_1^{-1})$, and the individual scan $\boldsymbol{f}$.

For this thesis, the algorithm for image registration used is based on the LDDMM framework, but includes some additional components that enable more efficient registration, both the number of iterations needed to achieve convergence and the amount of memory required to encode the deformations (Ashburner and Friston 2011). This is the Geodesic Shooting toolbox implemented in SPM12. Such an image registration model involves learning the relative shape of a brain in a geometric sense. LDDMM is able to encode relative form using the "initial momentum" formulation (Wang et al. 2007; Younes 2007), which is based on the conservation of momentum. The intermediate and final configurations (deformation) may be determined from the initial conditions, which are the spatial configuration (an identity transform φ_0) and the initial velocity or momentum. It is not necessary to optimize the entire series of velocity fields, but instead it is possible to perform registration by only optimizing an initial velocity, v_0 (development and equations in Ashburner J and Friston K (2011).

This work involves using a concept known as "scalar momentum". This does not have a simple biological interpretation, as it is merely a concept from geometric mechanics. Files encoding scalar momenta are computed from the registration outputs of the Geodesic Shooting toolbox of SPM12. The scalar momentum is defined as

$$\boldsymbol{u_o} = \frac{1}{\sigma^2} \left| \boldsymbol{D}_{\varphi_1} \right| (\nabla \boldsymbol{\mu}) (\boldsymbol{\mu} - \boldsymbol{f} \circ \varphi_1) \qquad (1.16)$$

Where $\boldsymbol{\mu}$ is the template that matches the image $\boldsymbol{f}$, and $\left| \boldsymbol{D}_{\varphi_1} \right|$ is the determinant of the Jacobian of the transformation. The scalar momentum is related to the warped tissue classes $(\boldsymbol{f} \circ \varphi_1)$, which unmodulated are difficult to interpret in a VBM analysis. However, modulated warped tissue classes, $\left| \boldsymbol{D}_{\varphi_1} \right| (\boldsymbol{f} \circ \varphi_1)$, can be more easily interpreted in a VBM as volume differences.

Scalar momenta become more complex when several tissues are aligned simultaneously. In this case, they are not scalar fields. The scalar momentum used in the current study involves grey matter (GM), white matter (WM) and an implicit background class BG = (1- (GM + WM)), essentially can be described as:

$$\left| \boldsymbol{D}_\varphi \right| (\mu_1 - c_1(\varphi))$$

$$\left| \boldsymbol{D}_\varphi \right| (\mu_2 - c_2(\varphi))$$

$$\left| \boldsymbol{D}_\varphi \right| (1 - \mu_1 - \mu_2) - (1 - c_1(\varphi) - c_2(\varphi)) \qquad (1.17)$$

This means that the scalar momenta encode information from the deformation fields, associated with the GM and WM, and also the residual difference between these tissue classes and the template. Further details about the method and the implementation are in

(Ashburner and Friston 2011). Note that in practice, even though there are three tissue components (GM, WM and BG), the scalar momenta are reduced to two components. This is because there is redundancy in the information, so some disk space can be saved using the two component approach.

1.3 Gaussian Process Models

Given a dataset $D = \{x_i, y_i\}$, $i = 1, ..., N$, consisting of pairs of samples $x_i \in \Re^d$ and labels y_i, the objective of a supervised learning algorithm is to learn a function from the data that can accurately predict the corresponding label, $f(x_i) = y_i$, for new samples.

A GP is a generalization of a Gaussian probability distribution, understood as a process that governs the properties of such functions. However, the aim is not only to find a function that fits the data, but also finding that one that better fits the data. Thus the model becomes complete when it is used in a Bayesian framework, using prior probability for each possible function that models the data and giving higher probability to those that are more likely to occur. In this way, over-fitting may be avoided.

In particular, GPs were initially developed for regression (Williams and Rasmussen 1996), and can be conceptualized as a Bayesian extension of linear regression (Bishop 2006). For the inputs in the dataset D, $x = (x_1, x_2, ... , x_N)$, the model can be described as follows:

$$y(w, x) = w^T \phi(x) \qquad (1.18)$$

Where $\phi = (\phi_1, \phi_2, ... , \phi_M)^T$ are the basis functions of the inputs x, and $w = (w_1, w_2, ... , w_M)^T$ is the weights vector, the distribution of w imposes a limitation on the model, $y(w, x)$ is a linear function of the input variables x governed by them. If a prior distribution is assumed for w, given by an isotropic Gaussian distribution,

$$p(w) = N(w|0, \sigma^2) \qquad (1.19)$$

then this distribution induces a continuous Gaussian distribution over $y(x)$.

The aim of this approach is to finally evaluate this function at a particular set of $x = (x_1, x_2, ... , x_N)$; i.e. the training sample described as input in D. Those function values related to the input samples are described in y with components $y(x_1), y(x_2), ... , y(x_N)$. Then, this vector can be expressed as

$$y = \Phi w \qquad (1.20)$$

Where Φ is the matrix that encompasses the basis functions, $\Phi_{nk} = \phi_k(x_n)$. A mean and covariance describe a Gaussian distribution of y. These are given from the prior probability of w due to the relationship between y and w in (1.19), by

$$E[\boldsymbol{y}] = \boldsymbol{\Phi}E[\boldsymbol{w}] \qquad\qquad (1.21)$$

$$cov[\boldsymbol{y}] = E[\boldsymbol{y}\boldsymbol{y}^T] = \boldsymbol{\Phi}E[\boldsymbol{w}\boldsymbol{w}^T]\boldsymbol{\Phi}^T = \frac{1}{\alpha}\boldsymbol{\Phi}\boldsymbol{\Phi}^T = \boldsymbol{K} \qquad (1.22)$$

Where the elements of $\boldsymbol{K}$ come from the available basis functions

$$\boldsymbol{K}_{nm} = \boldsymbol{K}(\boldsymbol{x}_n, \boldsymbol{x}_m) = \frac{1}{\alpha}\boldsymbol{\Phi}(\boldsymbol{x}_n)^T\boldsymbol{\Phi}(\boldsymbol{x}_n). \quad (1.23)$$

The hyper-parameter α represents the precision of the distribution of $\boldsymbol{w}$ and corresponds to the inverse of σ^2. $\boldsymbol{K}$ is the covariance function, also known as kernel function, which becomes directly dependent on the input sample.

In general, a GP is defined as a Gaussian probability distribution over the functions $y(\boldsymbol{x})$ evaluated at a specific set of points (training samples), such that it jointly also fits a Gaussian distribution (Bishop 2006).

In summary, a GP model can be specified by a multivariate Gaussian distribution with a mean vector and covariance matrix:

$$GP \sim N(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')) \qquad\qquad (1.24)$$

Generally, prior information about the mean is not known. By symmetry it is taken to be zero and so, the mean of the prior over the weight values is also assumed to be zero, *p($\boldsymbol{w}$/α)*. The GP view is then completed by estimating the covariance between any pair of input sample by means of the kernel function as

$$E[y(\boldsymbol{x}_n)y(\boldsymbol{x}_m)] = k(\boldsymbol{x}_n, \boldsymbol{x}_m). \qquad\qquad (1.25)$$

In each particular case, this kernel function can be estimated by defining a specific relationship between samples. In the current work, as in the linear regression model where weights follow an isotropic Gaussian distribution, the kernel function is given by the dot product (1.23).

To apply GP to regression, the noise of the observed target values must be considered. It is assumed that a random noise term is added to $y(\boldsymbol{x})$, which follows a Gaussian distribution independent of each observation (Bishop 2006).

GP for classification is an extension of GP for regression. As set out above, GP gives predictions in the continuum, but when doing classification, the aim is to predict discrete categories. If the output in the continuum is transformed by means of an activation function, which gives 0 or 1 depending on the input, the predictions by a GP are adapted to a classification problem. A Gaussian process defined over a function $a(\boldsymbol{x})$ is transformed by using a logistic sigmoid $\boldsymbol{y} = \sigma(\boldsymbol{a})$, to give a non-Gaussian stochastic process over functions $y(\mathbf{x})$ where $y \in (0,1)$, as would be expected for a two-class problem with target $t \in \{0,1\}$.

This section only provides a brief introduction to GP in order to present the context necessary to follow the material in this thesis. The interested reader is referred to Bishop (2006) and Rasmussen and Williams (2006) for a more detailed explanation.

1.4   Data feature representation

As mentioned, the success of machine learning methods not only relies on the approach, it also depends on the features used to fit the model. More powerful data features are essential to achieve optimal performance from machine learning. Different kinds of features can be obtained from the preprocessing commonly used for VBM analyses. Outputs become features that can be used as inputs for multivariate techniques.

In this thesis, an assortment of images from the VBM-type pre-processing with SPM software that may allow interesting characterizations were selected as features. Grey matter (GM) and white matter (WM) tissue classes, also Jacobians, which encode the volume changes from warping native images to the standard space, and different operations and combinations of these images were explored.

Additionally, other relevant types of data were considered The VBM-type pre-processing has improved enormously since the beginnings (Ashburner and Friston 2001). The current methods for aligning the images of different subjects are more sophisticated and accurate than a few years ago (Klein et al. 2009; Ashburner and Miller 2015). Some of them allow different representations of image shapes, which may be useful for pattern recognition. In that respect, scalar momenta are candidate to become interesting features. These images, even though do not have biological interpretacion, may encode key information about inter-subject anatomical differences.

2. Objectives

The main objective of this thesis is to obtain reliable and accurate features from structural MRI data for pattern recognition studies in schizophrenia, which should contribute to the application of improved methodologies in clinical research. To accomplish the main objective, the following specific objectives are posed:

(1) Determining, from among the most used methods for brain tissue segmentation, which algorithm produces the most reliable grey matter (GM) segments with respect to a defined ground truth.

(2) Finding a class of features from structural MRI data that is relatively efficient, irrespective of whether the target is discrete or continuous, for pattern recognition studies. Image data from healthy subjects are used in order to predict demographic variables, such as age and gender. Features that are aimed to be studied are those obtained from the VBM-type of preprocessing: tissue classes scaled and non-scaled by the Jacobians, Jacobians, divergence of velocities and scalar momenta.

(3) Exploring the relationship between features from preprocessed MRI data and smoothing.

(4) Designing how total intracranial volume (TIV) can be introduced in the features from modulated data and studying how it influences the performance of predictions in modulated tissue classes.

(5) Verifying whether or not GM provides the best feature for predicting schizophrenia.

(6) Determining the optimal amount of image smoothing for predicting schizophrenia.

(7) Testing whether the types of feature that best predict demographic variables in healthy subjects are also best for predicting schizophrenia.

2.1 Roadmap

This thesis is divided into three studies, designed to accomplish the objectives presented above.

Study 1

A comparison between three different segmentation algorithms is performed. These algorithms are "Unified Segmentation" (US) and "New Segmentation" (NS) from the SPM8 package, and FAST from the FSL package. A publicly available data set, which includes manually segmented tissue classes by experts, was used. A detailed comparison between algorithms and ground truth was conducted using different methods for testing the reliability of each algorithm with respect to the "gold standard" segmentations. With this study objective (1) can be fulfilled.

Study 2

This study involves applying the Gaussian Processes machine learning approach to a number of sets of features derived from same subjects' preprocessed MRI data. The age, body mass index and gender of all subjects was predicted in order to find an effective feature representation of image data from multiple subjects, such that pattern recognition methods may be made more accurate. Predictive accuracy was explored using a variety of image features, which had been spatially smoothed over a wide range of full-width at half maximum (FWHM). Also a novel method for including total intracranial volume (TIV) into the features was attempted. With this work, objectives (2), (3) and (4) can be evaluated.

Study 3

This study consists in applying conclusions from the previous evaluations to schizophrenia. Our hypothesis was that image features that worked well in study 2 would also work well for predicting schizophrenia. With this study, objectives (5), (6) and (7) were to be answered, as well as accomplish the main objective of this thesis.

3. Study 1: Influence of segmentation model in VBM-type preprocessing

Nowadays, the interest in the accuracy of automated computational neuroanatomy tools is increasing. Hundreds of studies have used structural MRI data, mostly for Voxel-Based Morphometry (VBM) (Wright et al. 1995; Ashburner and Friston 2000; Ashburner and Friston 2001). Pre-processing structural MRI for VBM usually entails applying a pipeline of different algorithms to model the images. First step involves segmenting the images in order to identify the tissues of interest. These are mainly grey matter (GM) and white matter (WM) tissue classes. Tissue classes are then spatially normalized and smoothed to enable statistical comparisons. Reliable findings depend on reliable registered tissue classes. In VBM, the outputs from segmentation are involved in registration. If segmentation is inaccurate, the next step cannot be accurate either. Hence the accuracy of the final preprocessed data is dependent on each step of the processing.

Several questions arise when a VBM has to be carried out. The first one is about which software to use. In the neuroimaging field, the two most widely used software packages are FSL (Analysis Group, FMRIB. Oxford, UK) and SPM (Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology. London, UK). Both, FSL and SPM, have implemented different approaches for VBM pre-processing. These share some commonalities, but there are also a number of differences.

For image segmentation, the SPM package implements a method based on fitting mixtures of Gaussians to the image intensities (Ashburner and Friston 2005). This is combined with an atlas guided approach, in which a warping from the atlas to each image is performed. In that way, the characteristic space of intensities and the spatial information of the image are both combined. The approach in the FSL package is similar, but also includes a Markov Random Field, which models interactions between voxels and their nearest neighbors (Zhang, Brady, and Smith 2001).

There are many relevant differences between these two options, although there are some commonalities. Both SPM and FSL are assembled in a Bayesian framework. In both cases the problem is formulated as MAP estimation. Finally, the model fitting is performed by the Iterative Conditional Modes and the Expectation Maximization algorithms.

3.1   Motivation & Objectives

Two different theoretical frameworks entail different algorithm implementations, leading to subtly different tissue class images arising from FSL and SPM. This study focuses in quantifying those differences. For this purpose, a public dataset has been used. This dataset has been selected because tissue classes segmented by experts are available. These manually segmented tissue classes were used as the "ground truth".

Differences due to methodology and systematic bias inherent to the algorithms can be influenced by additional factors. Poor quality images, artifacts, pathology, and anatomical variability in general, can cause miss-segmentation. Images used in this study have poor quality, exhibit low contrast and relatively large intensity gradients. Segmentation algorithms must face these additional difficulties.

The objective of the current study is to determine which algorithm gives tissue classes more similar to the ground truth. This decision is made by (1) evaluating the influence of the differences between approaches, (2) detecting systematic errors, if they exist, caused by each segmentation algorithm, and (3) quantifying similarities between resulting tissue classes from each segmentation method and the ground truth.

3.2   Material & Methods

This study focused on the comparison between two segmentation approaches, by SPM and FSL. For this purpose, both algorithms were applied to the same public dataset. This dataset also included manually segmented tissue classes that were used as gold standard.

The work is divided into two parts. First, preliminary evaluations were performed in order to define the initial requirements and how they may interfere in the segmentation process. Hence, the effects ofthe manual reorientation necessary for SPM, and the skull stripping essential in FSL, were explored.

Second, the pipelines were conducted and evaluated at each step for finally quantifying differences at the final stage. Each segmentation process was carried out in identical conditions and default settings were used for each software.

3.2.1   Data

The evaluation was performed with 20 normal MR brain scans provided by the Center for Morphometric Analysis at Massachusetts General Hospital. Scans are available at the IBSR

web site[1]. These images have been manually segmented by experts, and the segments are also available at the IBSR web site. Ground truth GM, WM and CSF were compared with the tissue classes estimated by the automated methods. These images involve certain level of difficulty, there are scans that contain considerable acquisition artifacts as well as irregularities.

Data acquisition

The high resolution structural T1-weighted MRI data were acquired ona 1.5 Tesla GE Signa scanner (General Electrics Medical Systems, Milwaukee, Wis.), with the following acquisitions parameters: Matrix size 512 × 512; 180 contiguous axial slices; voxel resolution 0.47×0.47×1 mm[3]; echo (TE), repetition (TR) and inversion (TI) times, (TE/ TR/ TI)= 3.93ms/ 200ms/ 710ms respectively; flip angle 15 degrees).

3.2.2 Software

Comparison was carried out between:

- **FSL 4.1.5** - Two scripts from FSL-VBM v1.1were used: the BET2.1 (fslvbm_1_bet), and the FAST tools (fslvbm_2_template).

- **SPM8** - Segmentation for VBM pre-processing includes two different approaches: "Unified Segmentation" and "New Segmentation".

3.2.3 Preliminary evaluations

Differences between software pipelines at various stages were explored. These stages relate to how data should be prepared: skull stripping that is necessary for FSL segmentation and manual reorientation of the images recommended for SPM.

Manual realignment

Prior to the actual segmentation in SPM, an affine registration between tissue probability priors and the images to segment is performed. The idea behind the manual positioning is to make the segmentation more robust to local optima from this registration. In the "Unified" and "New Segmentation", manual reorientation is recommended. This involves reorienting images so that the anterior commissure (AC) is close to the origin of their "world" coordinate system, approximately with same orientation asthe ICBM templates supplied with SPM. A preliminary test to visualize this effect was explored: the SPM8

---

[1] *Internet Brain Segmentation Repository | www.cma.mgh.harvard.edu/ibsr/*

implementations of "Unified Segmentation" and "New Segmentation" were used to segment non-reoriented data.

Brain Extraction Tool

VBM with FSL was performed using the FSL_VBM v1.1 tool. The pipeline of this tool consists of three steps: **fslvbm_1_bet**, where the brain is extracted from the T1-weighted images; **fslvbm_2_template**, in which GM tissue class is segmented and a GM template is created, and **fslvbm_3_proc**, the non-linear registration of the GM images is performed.

The first step (**fslvbm_1_bet**) conducts brain extraction, and it is done by BET. This tool is provided with several options for skull stripping. Depending on the option, extracted brains may differ. For instance, automatically segmented brain may include part of the neck, eye balls, etc. On the contrary, tissue of the cortex may be wrongly removed. In addition, the accuracy depends on whether the "standard" (runs BET2) or "robust" (runs BET2 twice) forms of the procedure are run. Subsequent GM segmentation may dependon the prior brain extraction.

A preliminary test was conducted to explore the role of BET in a VBM. Two trials were performed with the ISBR dataset:

First trial: Brain extraction of the IBSR dataset

The most used commands available for **fslvbm_1_bet** were explored in the "standard" way. The options explored were:

(1) *fslvbm_1_bet -b:* BET works by default.

(2) *fslvbm_1_bet -N*: This option is more restrictive, and is recommended when images include part of the neck.

(3) *fslvbm_1_bet -N -f 0.5:* This command fixes the fractional intensity threshold to 0.5, instead of default 0.4, which is less restrictive.

(4) *fslvbm_1_bet -N -f 0.5 –R:* Adding –R, BET runs a more robust brain centre estimation.

Second trial: Segmentation validation engine (SVE)

The second trial was conducted by means of the segmentation validation engine[2] (SVE). This is an online resource for validation of brain segmentation methods (Shattuck et al. 2009). The method involves extracting the brain from a dataset that it is provided for this purposewith the tool selected by the user. This dataset is the LPBA40 (Shattuck et al.

---

[2] *Segmentation Validation Engine | http://sve.loni.ucla.edu/archive/*

2008), and consist of 40 T1-weighted MRI volumes stored in NIFTI format. It has been used in several studies, such as the creation of the LONI Probabilistic Brain Atlas. Data were facilitated by the UCLA Laboratory of Neuro Imaging (LONI) for analysis following IRB approved procedures of both NSLIJHS and UCLA. For the current validations, data were not pre-processed.

Procedure: LPBA40 data were skull stripped using the same commands as in the first trial, outcomes were returned and an evaluation metrics was computed. The evaluation of differences was done by a metrics used for comparing binary segmentations of the same structures.

Using different measures from the data, success and error rates and other metrics often used in these kinds of comparisons were estimated. The Jaccard similarity coefficient is one of the most used for quantifying similarities and differences between images, and it was used in the next analyses.

Jaccard similarity coefficient

Given a pair of images, A: automated GM segment and M: manual GM segment, each one with *n* binary attributes, the Jaccard coefficient is a measure of the overlap between *A* and *M*. Each attribute of *A* and *M* can either be 0 or 1. The total number of each combination of attributes for both *A* and *M* are specified as follows:

- M_{11} represents the total number of voxels where *A* and *M* both have value 1.
- M_{01} represents the total number of voxels where *A* is 0 and *M* is 1.
- M_{10} represents the total number of voxels where *A* is 1 and *M* is 0.
- M_{00} represents the total number of voxels where *A* and *M* both have value 0.

Each voxel must fall into one of these four categories, the sum of all them are the total number of voxels *(n)*:

$$M_{11} + M_{01} + M_{10} + M_{00} = n \tag{2.6}$$

The Jaccard similarity coefficient is calculated as

$$J(A,M) = \frac{|A \cap M|}{|A \cup M|} = \frac{|M_{11}|}{|M_{01}| + |M_{10}| + |M_{11}|} \tag{2.7}$$

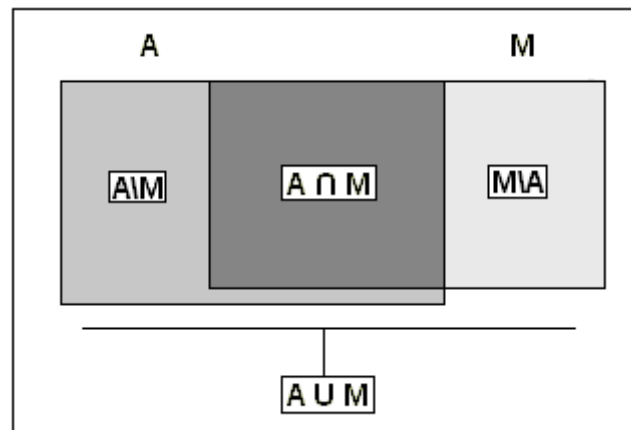*Fig. 3.1 – Pair of images, A and M. Representation of the overlap between images.*
The Jaccard index quantifies similarity between images.

3.2.4 Experimental design

After preliminary tests, the main body of the study was performed. Experimental design is divided into two parts. Part 1 is performed in the native space, and Part 2 is conducted in the stereotaxic space. Figure 3.2 describes an outline of these two parts. Each box represents one step in the processing pipeline.
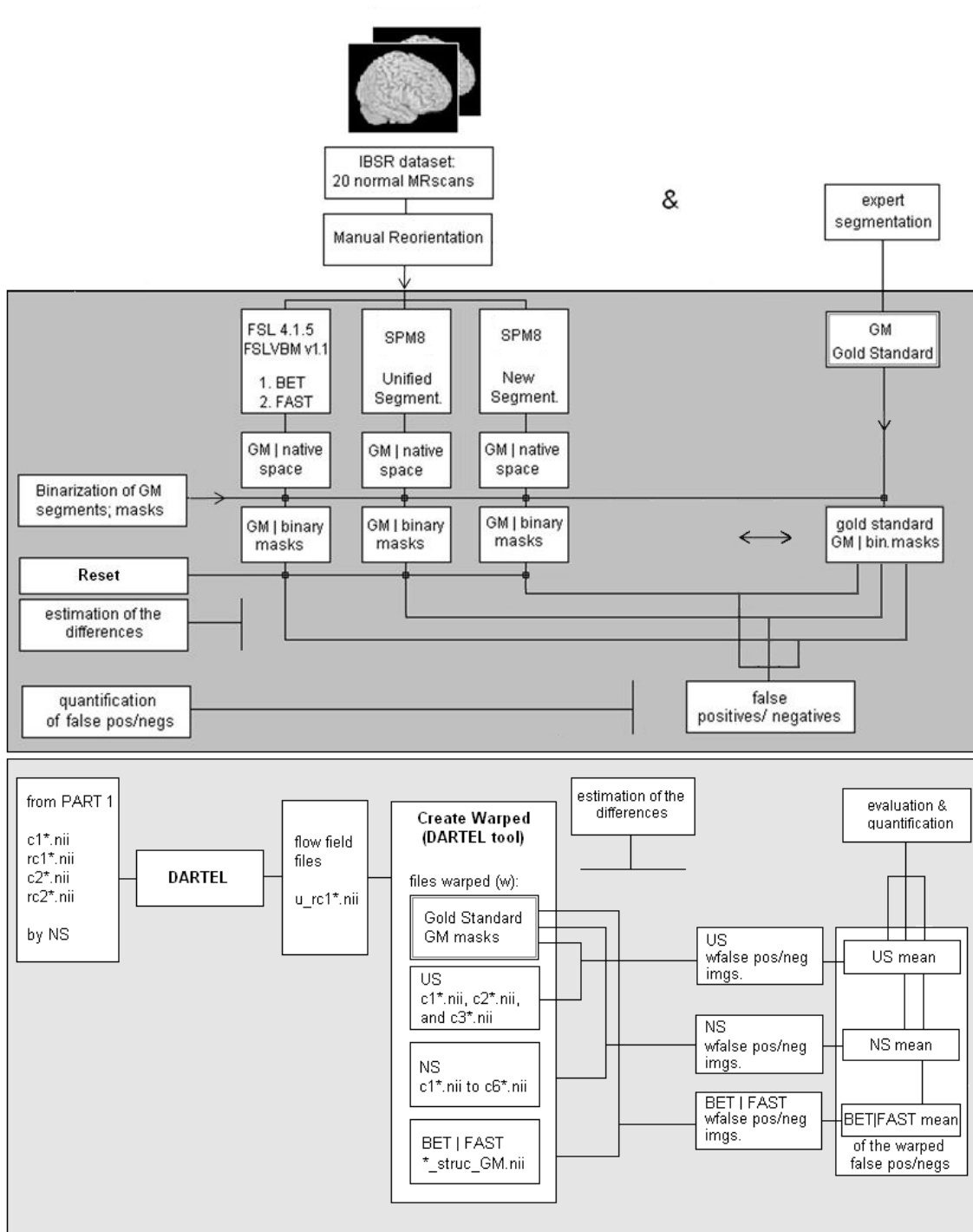
Fig 3.2– Part 1 involves the comparison in native space (dark grey). Part 2 is conducted in the same anatomical space, it involves a registration process (pale grey).

Preliminaries

In the preliminary evaluations, the effects of manual reorientation on SPM's segmentationand the skull stripping necessary for FSL were explored. These evaluations were conducted for establishing initial conditions in the current study.

The IBSR dataset was manually reoriented. This process does not interfere with the FSL processing. Then, the dataset used in all the pipelines with SPM and FSL was the same. Manual reorientation of the IBSR scans was done jointly with their respective manual segmentations. The comparison between automatic segmentations and their associated gold standards was conducted in the original and the reoriented position.

On the other hand, skull stripping with *fslvbm_1_bet –N* was conducted in the pipeline for segmentation using FSL.

Part 1: Native space

The aim of this section is to quantify differences that occur at the individual level due to the processingin the native space. Analyses were carried out using both FSL and SPM with the following tools:

1. BET (*fslvbm_1_bet -N*) + FAST (FSL VBM v1.1 in FSL 4.1.5)
2. "Unified Segmentation" (SPM8)
3. "New Segmentation" (SPM8)
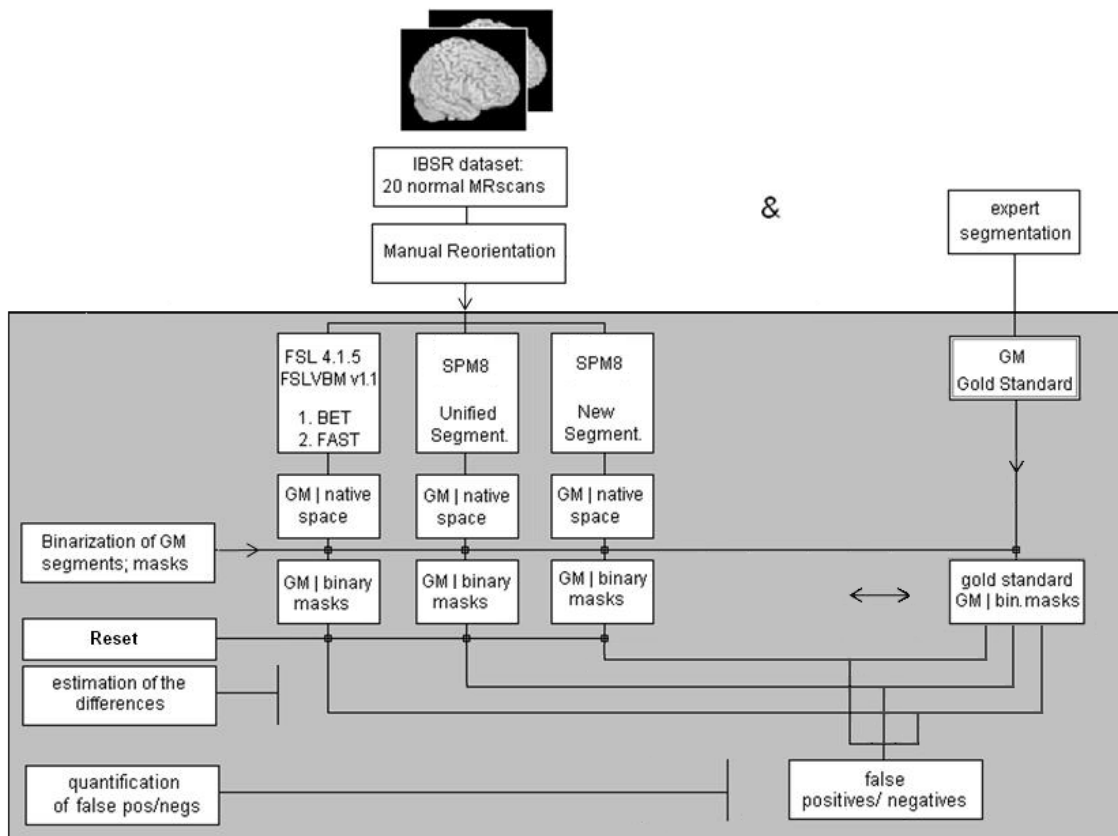
Fig. 3.3 - Design of PART 1 trial.

Evaluation of GM tissue class

The three segmentation processes were run using the re-oriented IBSR dataset. From all of them, the resulting GM tissue classes were selected and binary masks were created. In line with that, other masks were created from the manually segmented GM tissue classes, which were considered the gold standard (GS) masks.

The comparison between GS masks and the GM masks obtained from the three pipelines, was made in the native space. Thus, masks from automatically segmented tissues were transformed back to the native position. Header matrices were checked to ensure that they were identically positioned.

Differences between the GM segments from each pipeline and their respective GS masks, were calculated using the Jaccard similarity coefficient (described above in the SVE section).

Part 2: Customized anatomical space

In this part, the averaged differences between segmentations from each of the pipelines and the gold standard were computed. For this purpose, data were transformed to the same anatomical space. Next, differences between performances from each pipeline, with respect to the manual gold standard,were calculated and compared. This procedure allows any systematic errors to be detected.
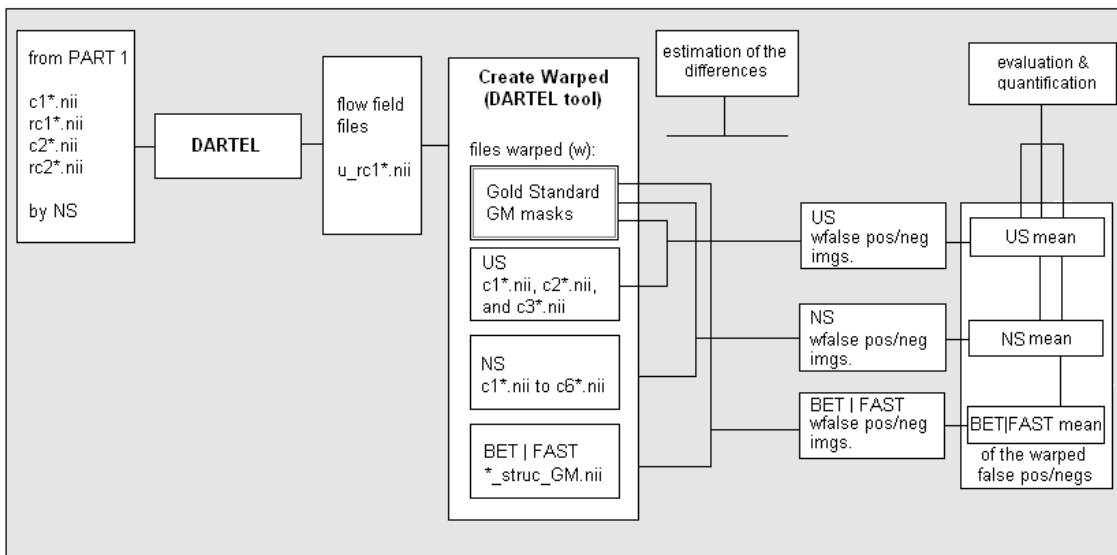
Fig. 3.4 –Design of Part 2

Registration

Registration of the GM data was carried out using Dartel (Ashburner 2007). Normalization parameters were estimated once, from pipeline 3 using "New Segmentation". It allowed the rest of sets to be normalized, because the same data were used in the other pipelines.

When required, the "New Segmentation" algorithm generates additional files that can be used to drive registration to a common anatomical space. These files are rigidly aligned versions of the GM and WM, which have been down sampled to isotropic 1.5mm resolution ("Dartel imported files").

Dartel (Ashburner 2007) was used to estimate accurate spatially normalizing transformations. Registration was driven by simultaneously aligning GM with GM, and WM with WM, using the default regularization and optimization settings. This approach especially improves the alignment of smaller inner structures. The version in a former release (SPM5) has been shown to outperform other approaches to spatial normalization (Klein et al. 2009).

Warping of tissue classes

The files encoding the spatial transformations ("flow fields") were used for warping the data obtained from the three pipelines. Each tissue class (GM, WM and CSF) in the native space was warped with its respective flow field. Also the manual gold standard segments were warped.

Evaluation of the normalized tissue classes

Binary masks were created from the warped tissues following same procedure described in part 1. Then, GM masks were subtracted from the gold standard GM masks, which were also in the same anatomical space. Resulting images consisted of the warped false positive/negative images associated with each subject. A scheme of this process is outlined in figure 3.5.
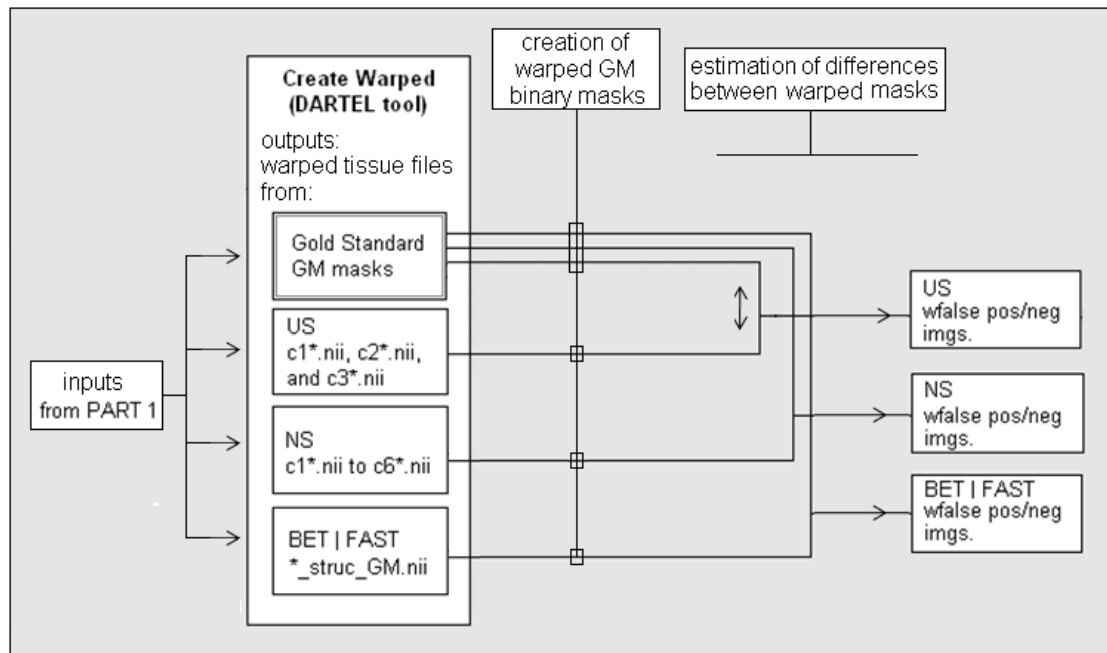
Fig. 3.5 –Outline of the process for creating warped false positive-negative images.

For computing differences:

i. **Statistics.** Voxels from false positive/negative images were counted. With numerical data, a two sample t-test was performed pair by pair, for false-positives and for false-negatives.

ii. **Jaccard similarity coefficient.** Differences between the masks from the automated algorithms and their respective GS masks were calculated using the Jaccard coefficient index, and a table with the results is presented in Table 3.3.

iii. **Mean images.** From the false positive/negative images, a mean image was created for each set of false positives/negatives. These average images encode information about the general performance of each algorithm and if any associated systematic errors.

However, these mean images may be biased due to a local fluctuation caused by a subject. The contribution of one subject to the mean may not be representative of the performance of an algorithm. In order to regularize this effect, the mean images were threshold. The original mean images had voxels from -1 to 1. Voxels with a value 1 and -1 were involved in 100% of the sample. If the value was 0.8, then 80% of the sample contained such voxel in the false positive-negative images, if the value was 0.5, it was the 50% and so on. Differences are more consistent when mean images are filtered. Contributions of 80% and 50% were explored. Cleaner mean images were more representative of the automated segmentation algorithms.

To determine whether the areas that appeared in the mean images were involved,either in false positive or in false negative, thresholded mean images were split, by separating false positives and false negatives. Locations of the areas are detailed in Table 3.7.

iv. Finally, the **density of voxels** at the locations detailed in Table 3.7 were computed, in Table 3.8. Mean images from 50% and 80% of sample contribution were created. It was calculated from mean images that contained only voxelvalues associated with the 50% and 80% of sample contribution. Density was estimated by obtaining the volume of these files, and dividing it by the number of voxels included in the respective mean, 50% or 80%.

3.3   Results

3.3.1   Preliminary evaluations

Manual realignment

Unified Segmentation

Using "Unified Segmentation", the segmentation failed for all of the IBSR data set without reorienting (Fig. 3.6 - top).

New Segmentation

In many cases this segmentation showed very different results compared to those obtained by the Unified Segmentation. New Segmentation resolved the lack of positioning in many cases. The bottom of Fig. 3.6 shows the same subject segmented by New Segmentation, with and without reorientation.
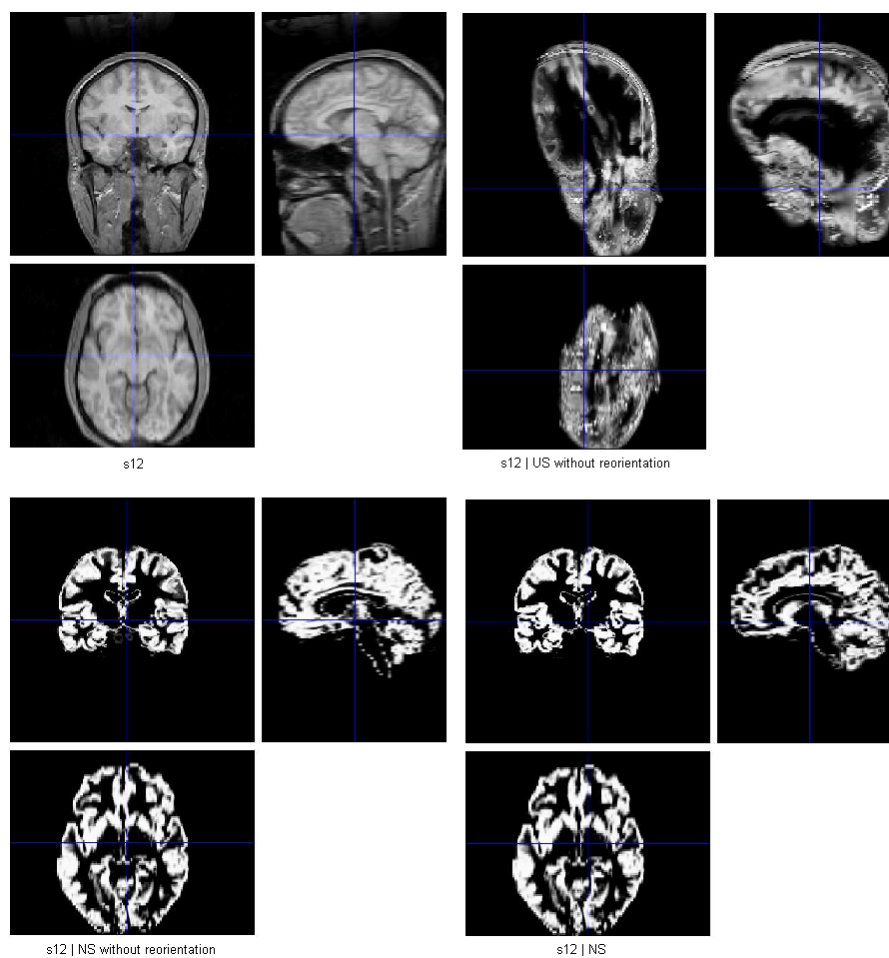
Fig 3.6 – Anatomical image (Ctrl. 12), and its respective GM segments by US without reorienting, and by NS without reorienting and reorienting.

a. GM segment from a reoriented image (ctrl.13)   b. GM segment from a non-reoriented image (ctrl.13)
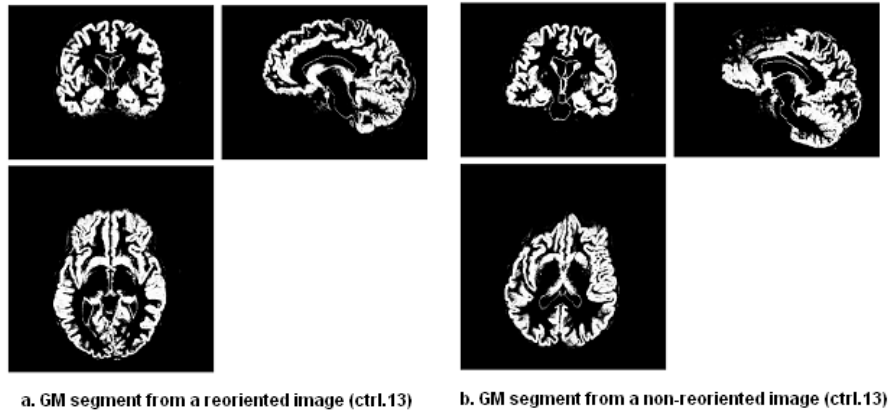
Fig. 3.7- Same subject (Ctrl. 13)) segmented by NS:Reoriented (left), and non-reoriented(right).

Although performance was better than for "Unified Segmentation", lighter deformations were found. The origin of these distortions lies in the affine registration that it is done to initialize the segmentation algorithm. The New Segmentation by itself does not have problems with the orientation, but a poor reorientation entails the risk that the affine registration algorithm could find a local optimum instead of the global optimum.

Actually, when the New Segmentation does not work the effect is pretty clear. It can be tested by introducing -1, to rotate one axis, or by rotating the image π or $\pi/2$ around the z axis (1.5702 radians at the raw).

a. reoriented   1. whole brain   2. detail

b. rotated   1. whole brain   2. detail
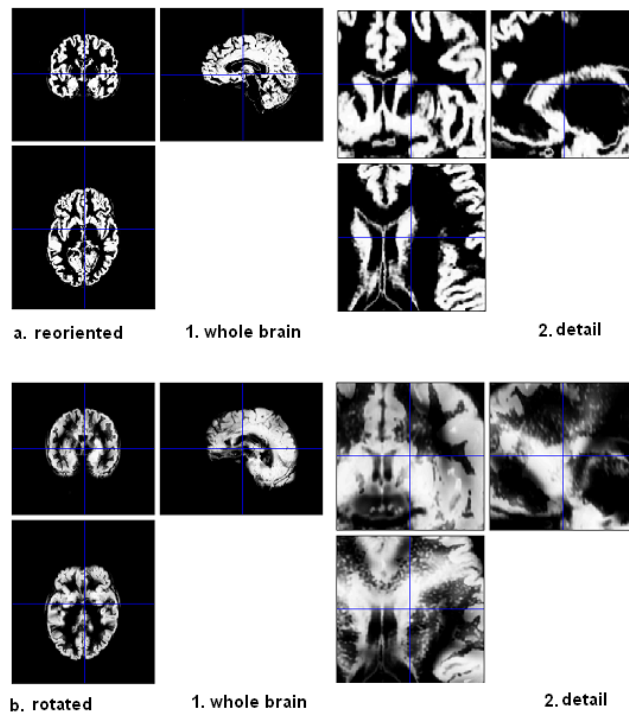
Fig. 3.8 – Same subject (ctrl. 20) segmented by NS: reoriented, and rotated $\pi/2$.

Reorientation prior to running the segmentation seems necessary to avoid problems. Centering the origin at the anterior commissure, positioning the axis within about 3 cm and 15° degrees usually enables a good performance of the algorithm.

Brain ExtractionTool

For the IBSR dataset, the performances of different brain extraction settings were assessed:

(1) *fslvbm_1_bet -b* : Tissue from neck was not eliminated.

(2) *fslvbm_1_bet -N* : Appeared more restrictive than (1). Results did not show extra brain areas, brains seemed well extracted.

(3) *fslvbm_1_bet -N -f 0.5* : Results were similar to (2) but this option seemed more intrusive in the extraction at boundaries.

(4) *fslvbm_1_bet -N -f 0.5 –R :* Results were similar to the both above (2) and (3).
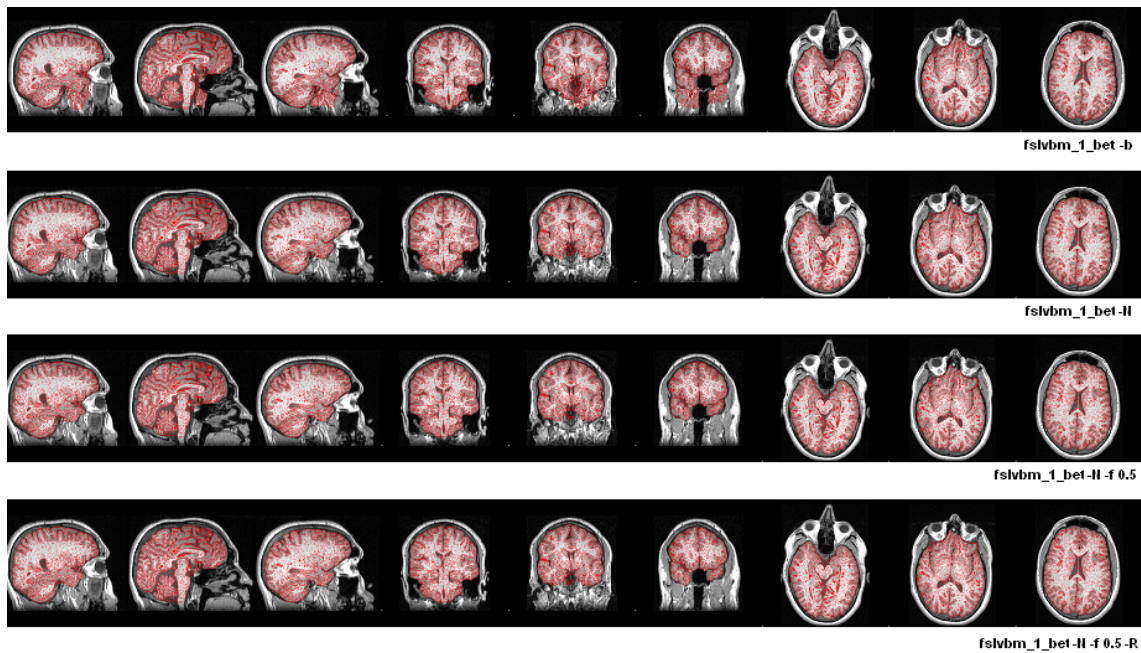
Fig. 3.9- Performances using different fslvbm_1_bet settings. Same IBSR subject (Ctrl. 20).

Last two settings showed very restrictive results after visual inspection, which can entail the risk of removing some tissue that should be included in a VBM analysis. On the other hand, including extra tissue that is not assumed to be GM is not ideal either. This happened using *fslvbm_1_bet –N*, which failed in 7 of the 20 cases. These seven subjects were finally skull stripped by giving the origin of coordinates, where the center of mass was visually estimated. In those cases, instead of "*-N*", another command was used, "*-B*", which attempts to reduce image bias and residual neck voxels.
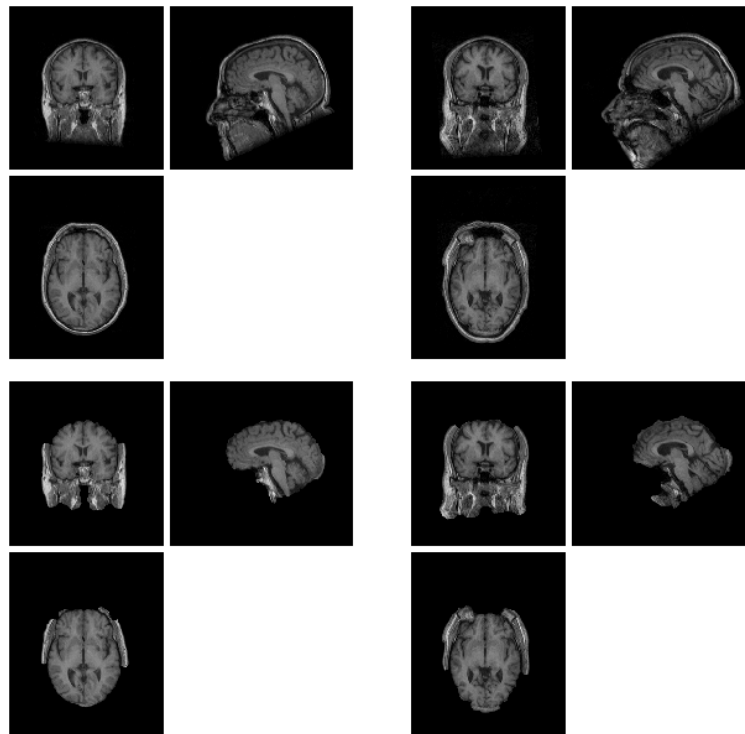
Fig. 3.9- BET by fslvbm_1_bet –N failed in two subjects. Whole brain and skull stripped images.

Segmentation validation engine (SVE)

A binary mask was created from each of the segmented LPBA40 images, and these masks evaluated. Projection maps were computed from the false positive and false negative values for each segmentation result. For the validation process, resulting false positive and false negative were re-mapped to a common atlas space based on the LPBA40. Then, SVE averaged these maps across the 40 subjects.

Dice coefficients, Jaccard indices, sensitivity and specificitywere computed from outputs, and results for each of the BET settings are displayed in the tables and in Figures 3.10 to 3.13.

Areas that appear brighter indicate a larger number of false negative voxels, indicating that the removal of those areas disagreed with the manually-labeled ground truth.
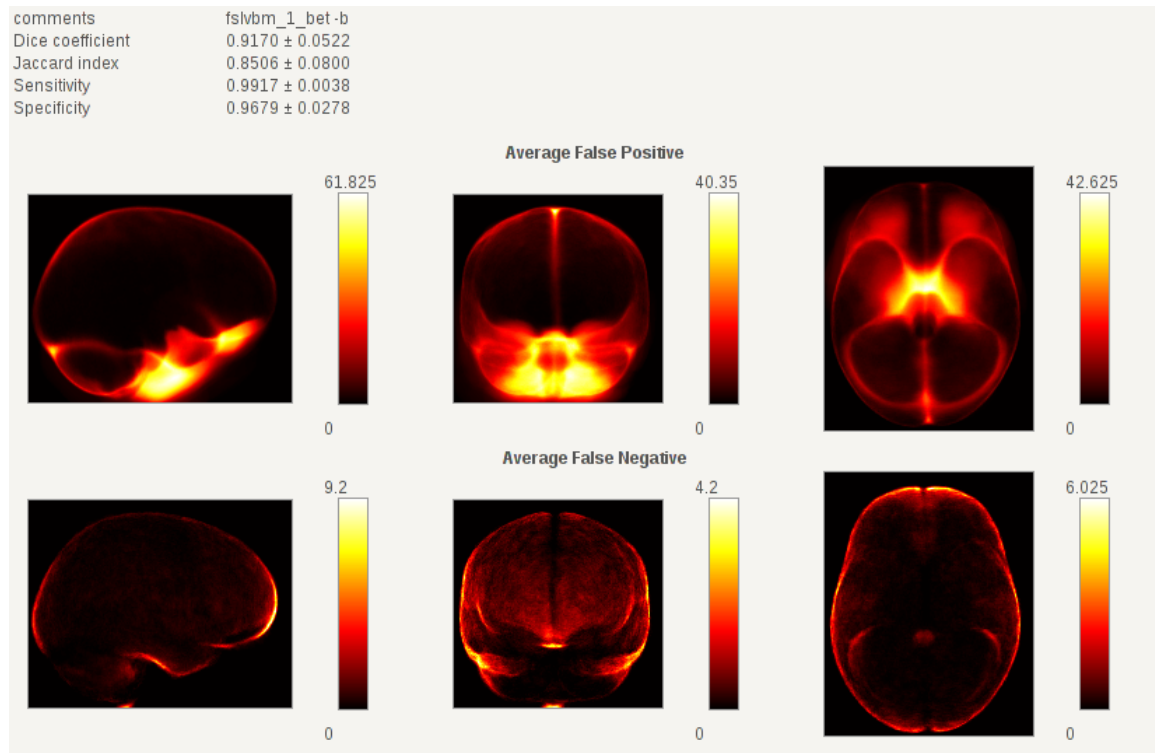
Fig. 3.10- SVE output of the fslvbm_1_bet –b option.
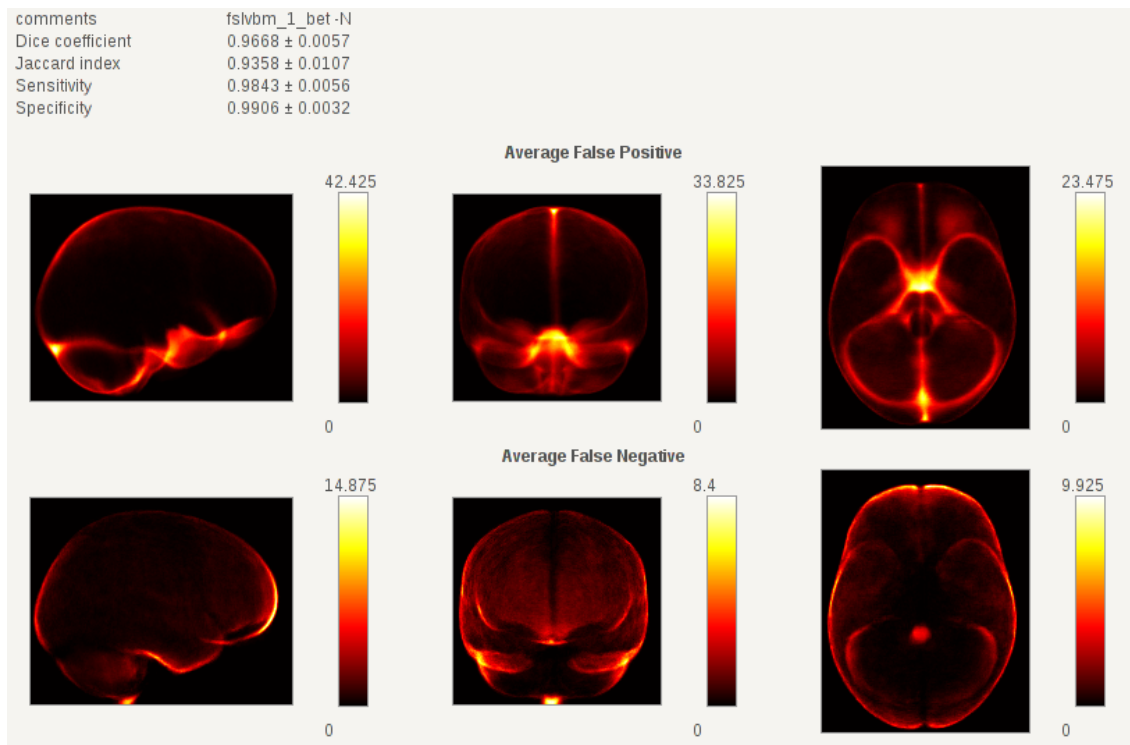
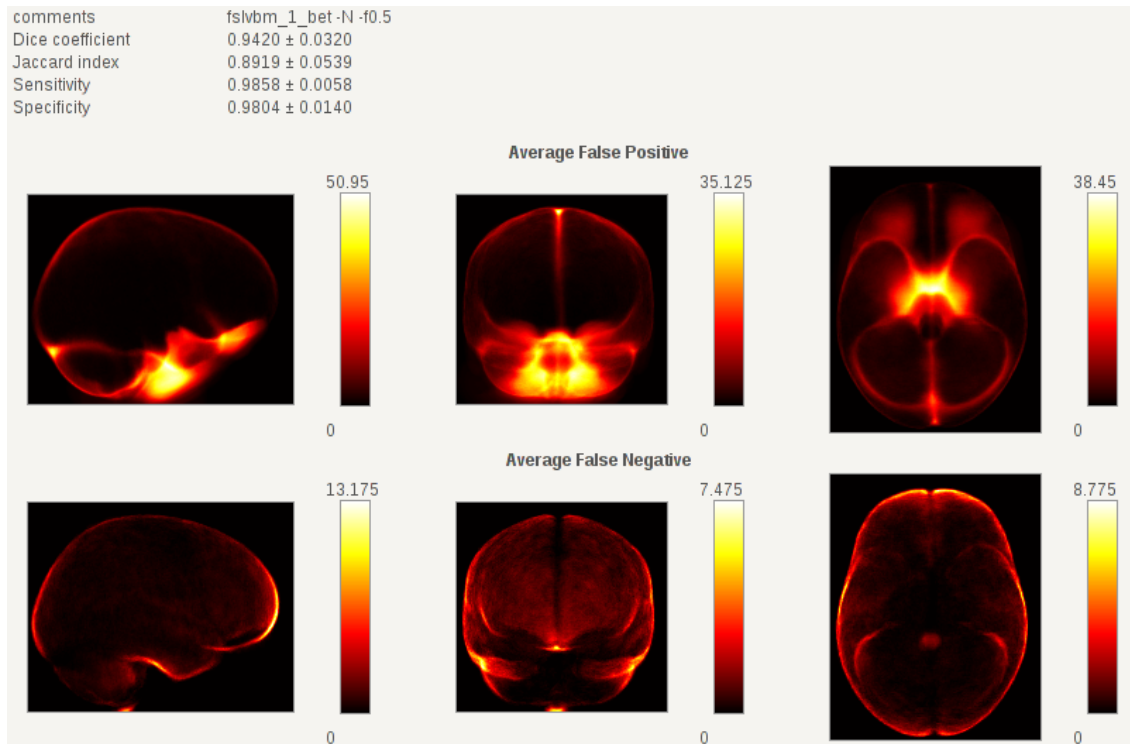

Fig. 3.11- SVE output of the fslvbm_1_bet –N option.

Fig. 3.12 - SVE output of the fslvbm_1_bet –N –f 0.5 option.
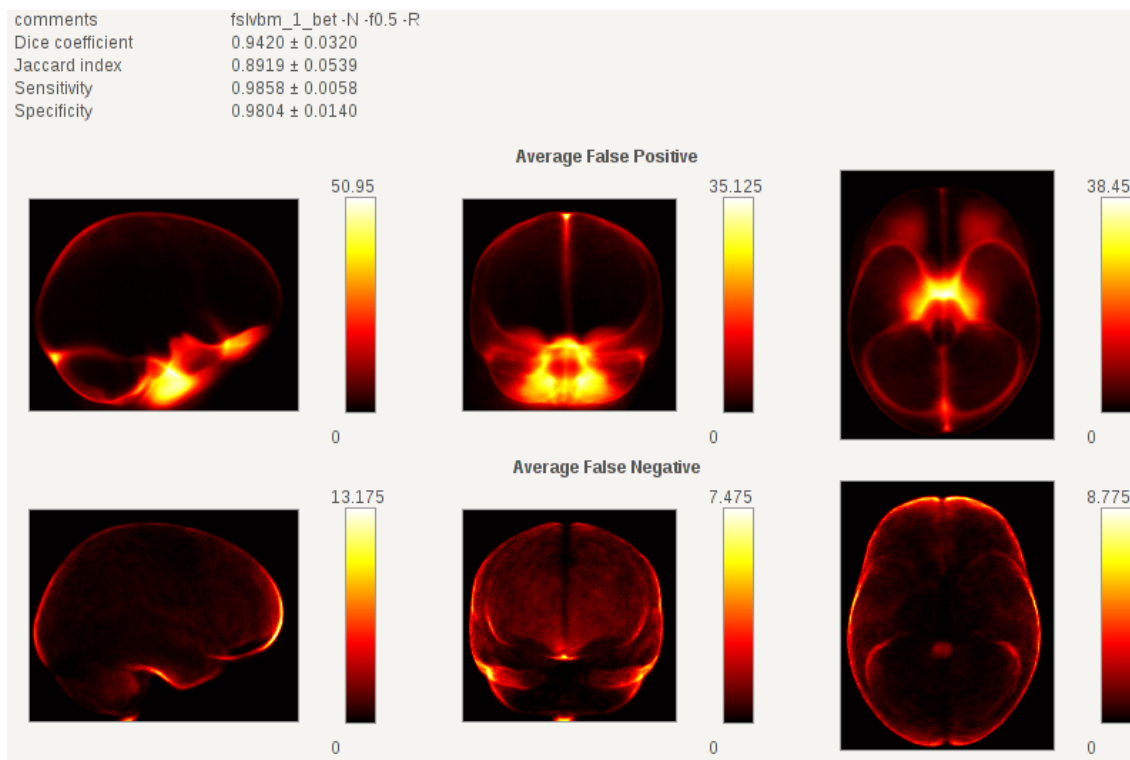

Fig. 3.13- SVE output of the fslvbm_1_bet –N –f 0.5 –R option.

The best results were obtained using *fslvbm_1_bet –N*, where the Jaccard indices and Dice coefficients were much higher than in the other cases.

Results obtained from the two lasts settings showed similar results. A detailed inspection of the images showed that there were no large differences between skull-stripped images from these two options. So, adding *–R* did not result in amore restrictive extraction in this data set.

The results largely agreed with those obtained with the ISBR dataset, and both trials suggested that the "*–N*" setting was most suitable.

3.3.2 Experimental design

Part 1: Computing differences in native space

Differences between sets of masks from each pipeline were computed. After subtracting automated GM tissue class masks from GS masks of GM, the resulting images contained only voxels with value -1, 0 or 1.

▫ Voxels where the algorithm assigned more tissue in the automated GM segment than in the gold standard GM segment were considered *false positive* (**1**, colored in white).

▫ Where the gold standard GM segment contains more tissue than the automated segment was considered a *false negative* (**-1**, colored in black).
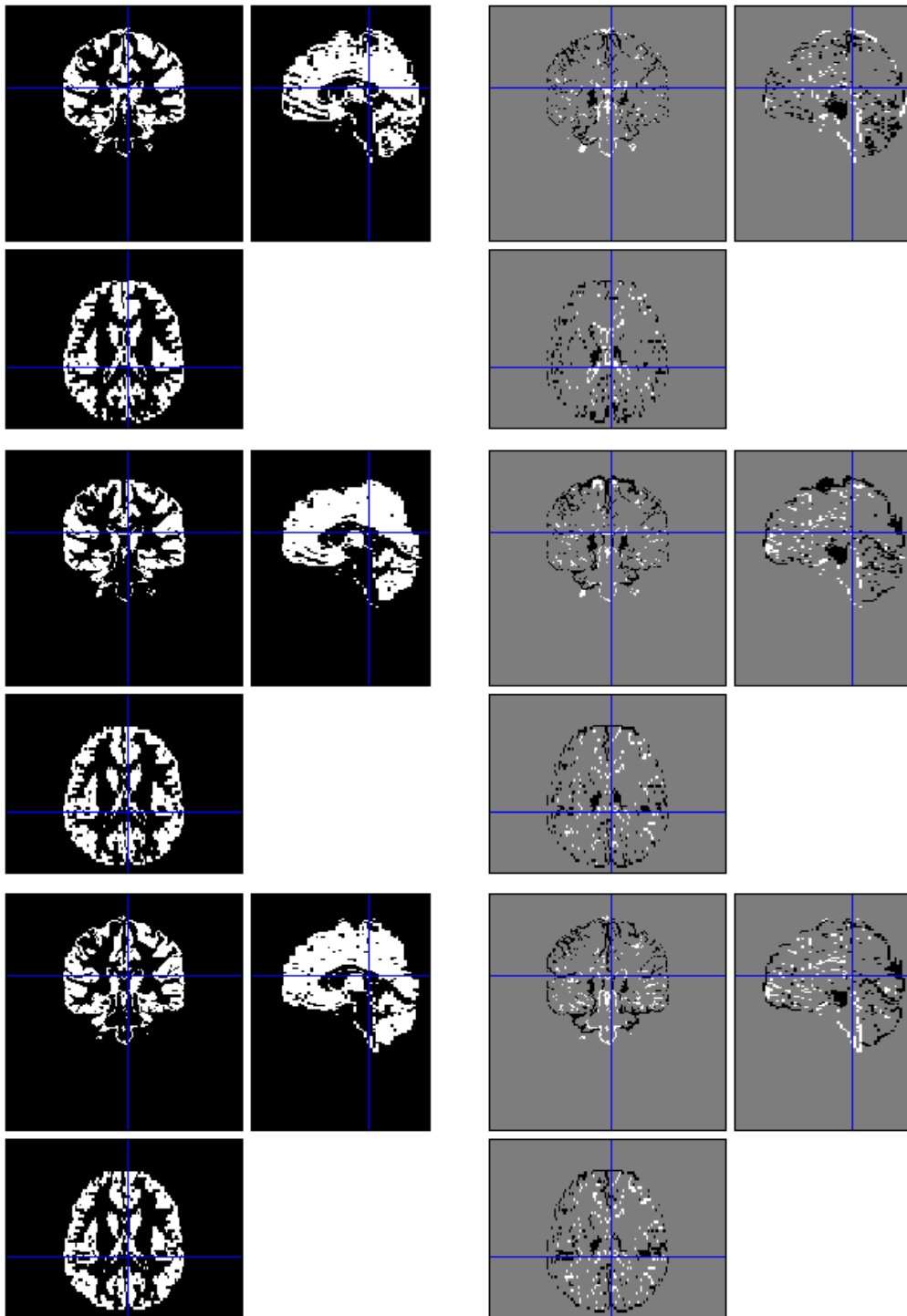
Fig. 3.14 – Same subject binary masks obtained from a.1) FAST, b.1) NS and c.1) US, and their respective difference images, obtained by subtracting these masks to their respective GS mask: a.2) FAST - GS, b.2) NS- GS and c.2) US- GS. (Subject 100_23)
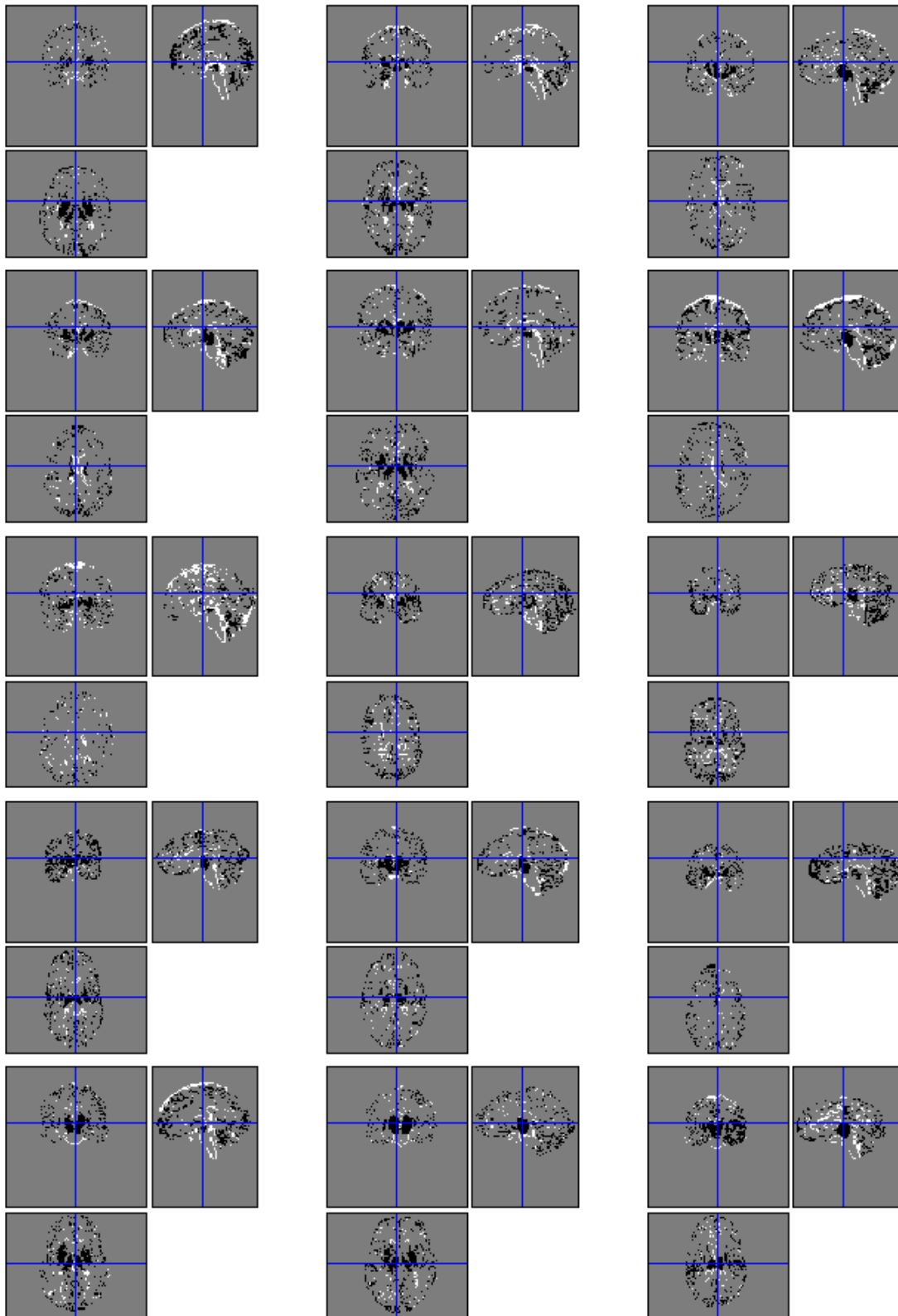
Fig. 3.15- False positive/negative images. Between FAST and the Gold Standard: where the algorithm assigned more tissue than the manual is in white, where the algorithm assigned less tissue than the manual is colored in black.
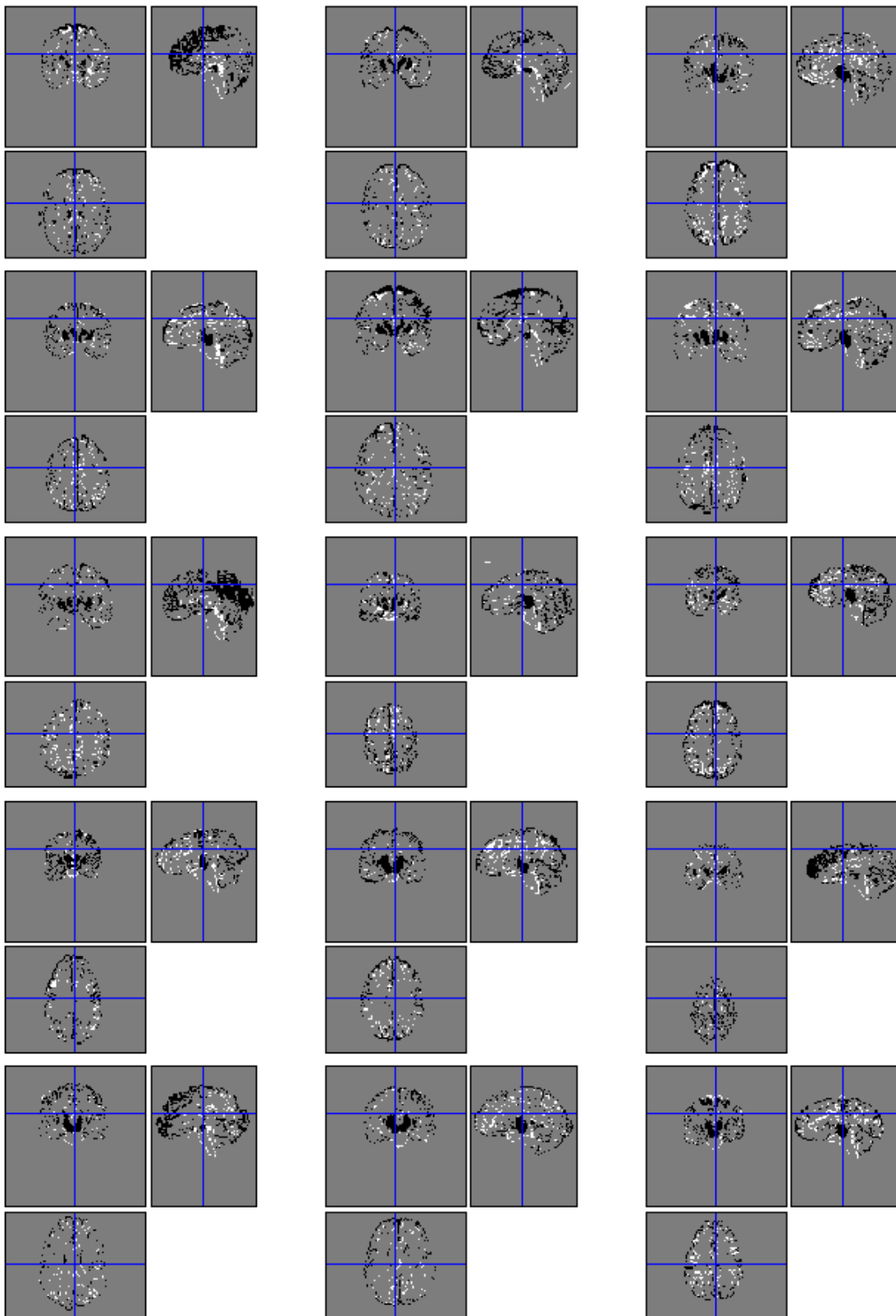
Fig. 3.16- False positive/negative images. Between NS and the Gold Standard: where the algorithm assigned more tissue than the manual is in white, where the algorithm assigned less tissue than the manual is colored in black.
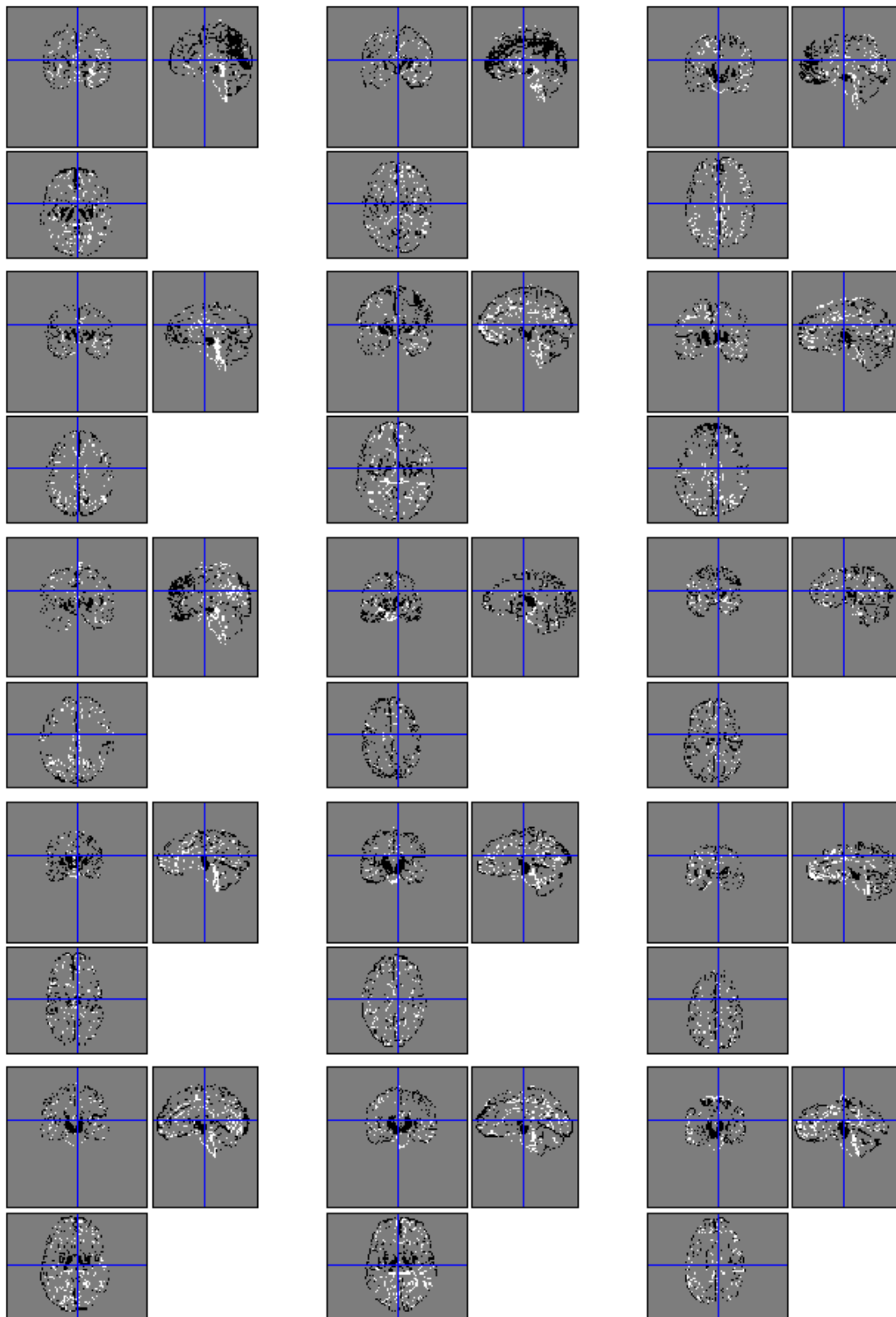
Fig. 3.17- False positive/negative images. Between US and the Gold Standard: where the algorithm assigned more tissue than the manual is in white, where the algorithm assigned less tissue than the manual is colored in black.

Resulting false positive-negative images allowed detecting regions where the algorithms systematically over- or under- determined certain amount of tissue.
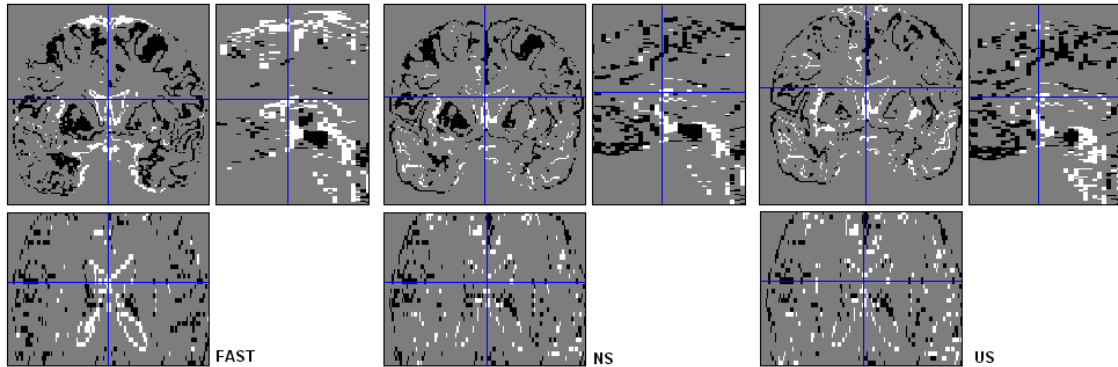

Fig. 3.18- Detail from images of the differences for (a) FAST, (b) NS and (c) US. (Subject 110_3)

The number of voxels equal to 1 (considered as *false positives*), and the number of voxels equal to -1 (*false negatives*) were counted in order to evaluate these differences statistically.

IBSR code	FAST		NS		US	
	1	**-1**	**1**	**-1**	**1**	**-1**
100_23	17226	69347	20711	69055	27824	64359
110_3	24151	77897	24286	69988	35448	57253
111_2	26660	74047	33608	70046	39583	64278
112_2	18706	68152	21164	64466	26309	59408
11_3	23057	93879	26990	95554	36157	86420
12_3	42335	96007	33152	75120	38433	76471
13_3	36761	68324	30156	66796	42143	57850
15_3	19629	79852	21228	66749	17097	79708
16_3	13193	69743	19180	64355	18844	62468
17_3	10187	71695	19769	54872	22141	55191
191_3	14917	67485	21831	64092	26900	64450
1_24	10878	64136	18546	47511	23259	50060
202_3	17022	72303	24090	59173	34674	56883
205_3	15903	67878	24357	61754	33254	66244
2_4	17176	65348	23511	53175	26816	48297
4_8	14218	56574	20222	50567	23917	44962
5_8	20586	82838	25564	62342	25013	62227
6_10	25997	82704	25849	64248	22664	71301
7_8	11046	78196	18683	52370	24067	53624
8_4	19358	68236	24631	60651	28086	58933
average	*19950*	*73732*	*23876*	*63644*	*28631*	*62019*
st.dev.	8258	9740	4437	10415	7122	10413

T-test	1	-1
FAST vs. NS	*0.0688*	*0.0031*
FAST vs. US	*0.0010*	*0.0007*
NS vs. US	*0.0155*	*0.6246*

Table 3.1 - Table shows the number of voxels that overcame the subtraction between each algorithm segmentation and their respective gold standard: the positives (as false positives) and the negatives (as false negative). At the bottom: average, standard deviation, and the t-test between sets.
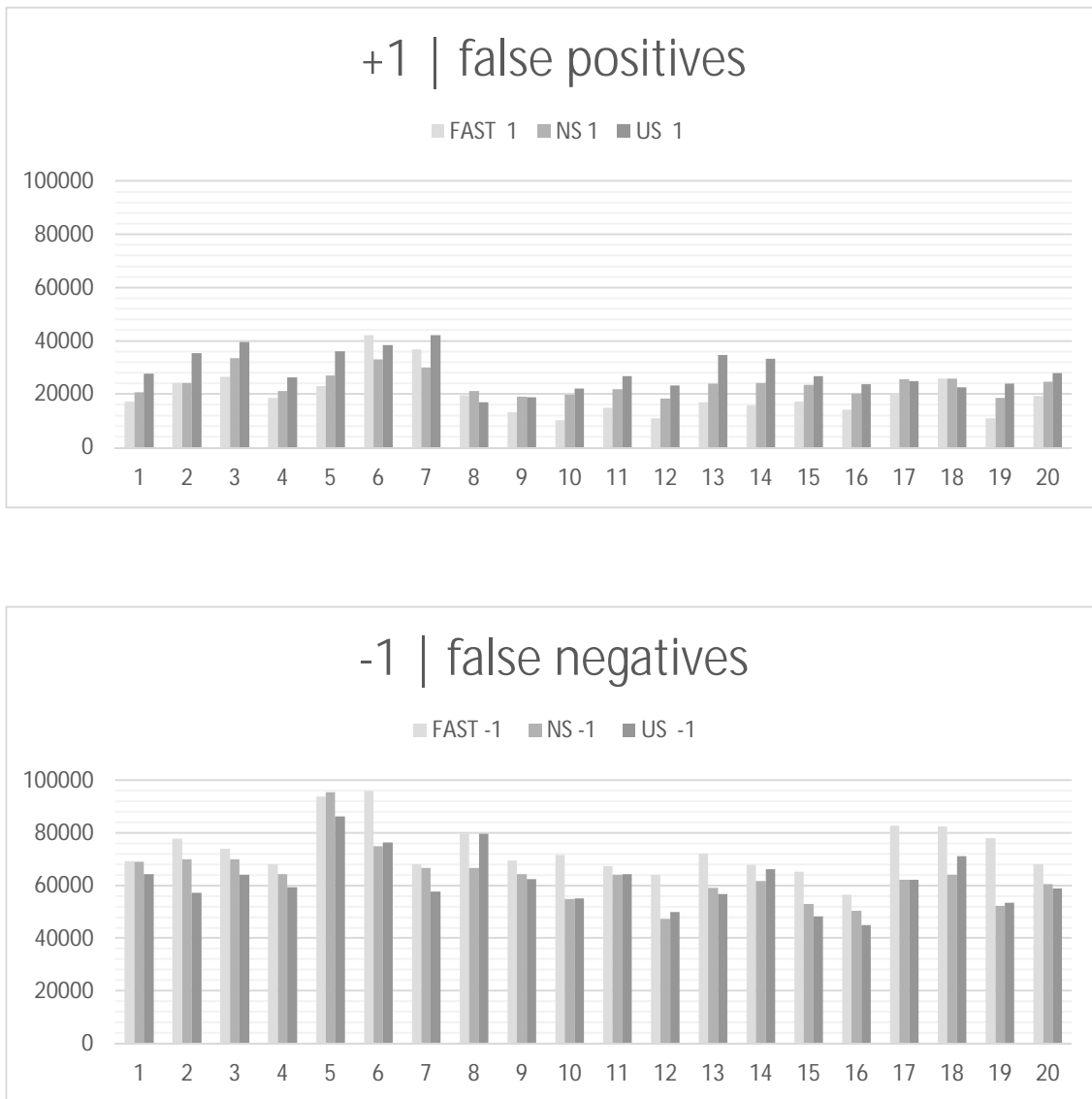
Fig. 3.19–Bar diagram of the number of voxels obtained from the subtraction between the algorithm segment and their respective GS: a) false positives, and b) false negatives.

FAST tended to give fewer false positives than the New Segmentation and Unified Segmentation. These differences became significant when comparing FAST and Unified Segmentation.

However, regarding false negatives the trend was reversed, and FAST gave more false negatives. Differences became statistically significant when comparing FAST with both the New Segmentation and the Unified Segmentation.
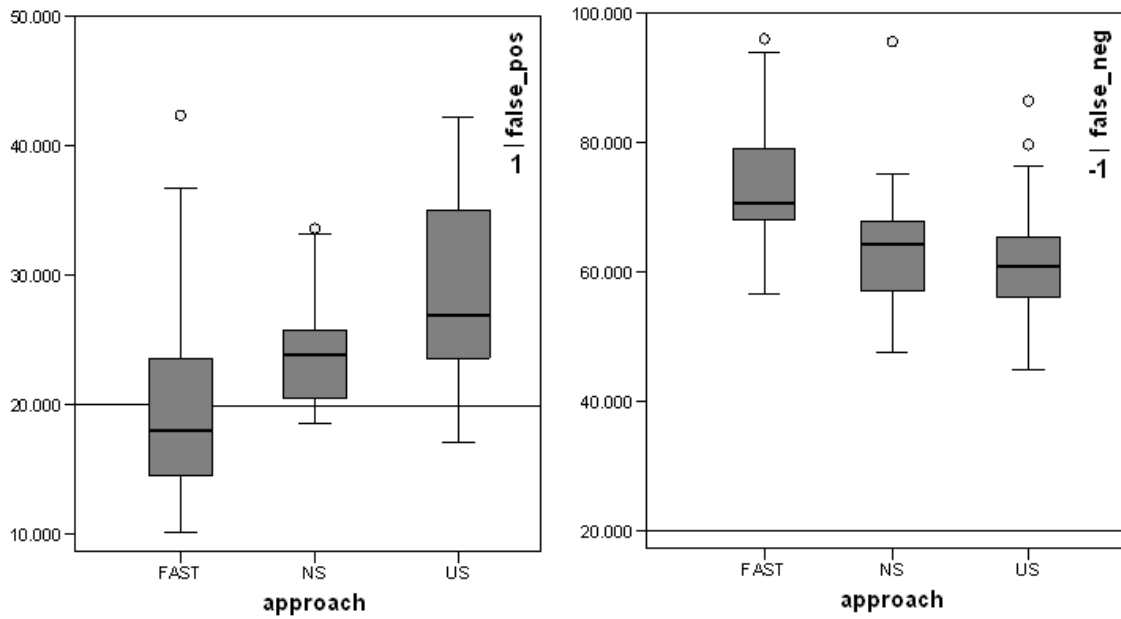
Fig. 3.20 – Box plots: (a) excess of GM by the algorithms with respect to the gold standard (voxels +1 ~ false positives), and (b) lack of GM by the algorithms with respect to the gold standard (voxels -1 ~ false negatives).

Jaccard similarity coefficient

The Jaccard similarity index showed high scores in most of the subjects between the New Segmentation and their respective gold standard. The average of the New Segmentation Jaccard indexes achieved the most similarity with respect to the other algorithms. It was followed closely by the Unified Segmentation, and finally by FAST.

	Jaccard index
New Segmentation	66.22%
UnifiedSegmentation	65.53%
FAST	63.10%

Table 3.2 – Ranking of the Jaccard indexes of the algorithms for this sample.

Next, Jaccard indexes estimation for each subject and algorithm are shown.

IBSR code	Jaccard FAST	Jaccard NS	Jaccard US
100_23	0.7084	0.7011	0.7002
110_3	0.6654	0.6910	0.7069
111_2	0.6636	0.6616	0.6674
112_2	0.6657	0.6735	0.6794
11_3	0.6824	0.6707	0.6785
12_3	0.6107	0.6872	0.6730
13_3	0.6920	0.7102	0.7114
15_3	0.5677	0.6204	0.5747
16_3	0.5841	0.5933	0.6034
17_3	0.5969	0.6491	0.6405
191_3	0.6713	0.6664	0.6522
1_24	0.6315	0.6872	0.6604
202_3	0.6839	0.7125	0.6951
205_3	0.6908	0.6918	0.6549
2_4	0.5997	0.6391	0.6519
4_8	0.6308	0.6420	0.6580
5_8	0.5316	0.6107	0.6127
6_10	0.5509	0.6276	0.6064
7_8	0.5837	0.6799	0.6583
8_4	0.6089	0.6280	0.6260
average	*0.6310*	*0.6622*	*0.6556*
st.Dev.	*0.0521*	*0.0344*	*0.0367*

algorithms	t-test | p
FAST vs. NS	*0.0315*
NS vs. US	*0.5604*
FAST vs. US	*0.0927*

Table 3.3 –Jaccard similarity coefficient estimated for each mask from the segments by FAST. NS and US versus their respective GS. At the bottom. some statistical estimates; t-test became significant at the FAST vs. NS comparison.
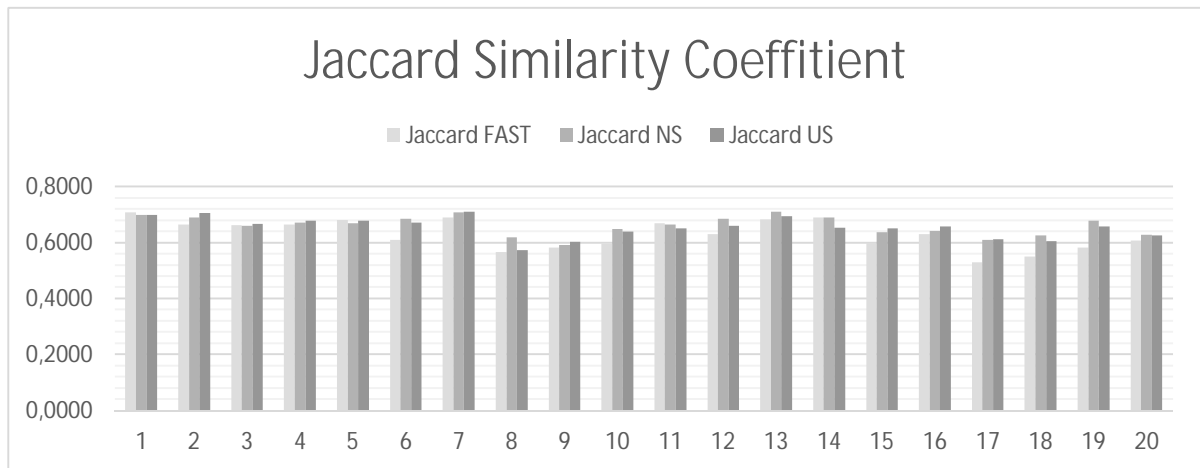
Fig. 3.21–Bar diagram of each estimated Jaccard coefficient

Part 2: Means in standard space from false positive-negative images

Each of the mean images shown in figures 3.22 to 3.23 was created from the 20 warped false positive/negative maps associated with each automated segmentation processes. White voxels represent false positives and black voxels represent false negatives.
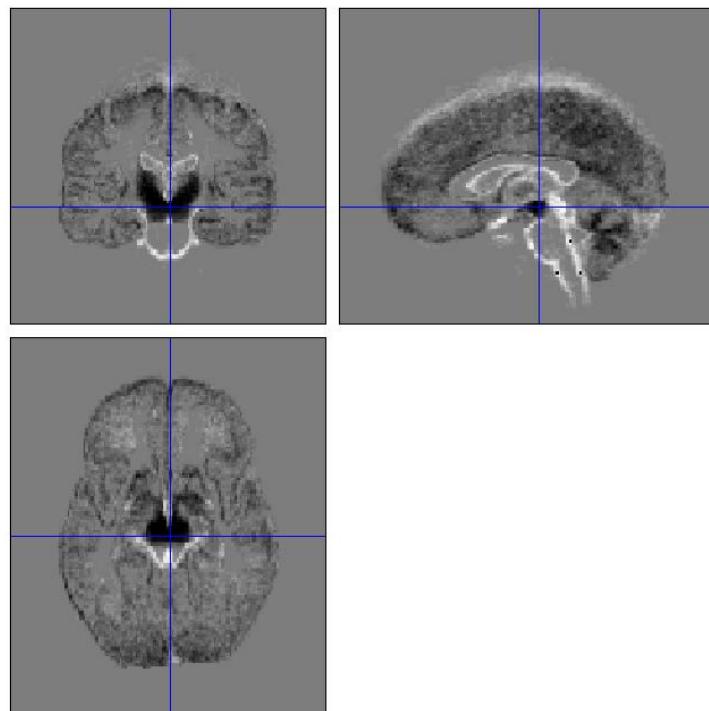

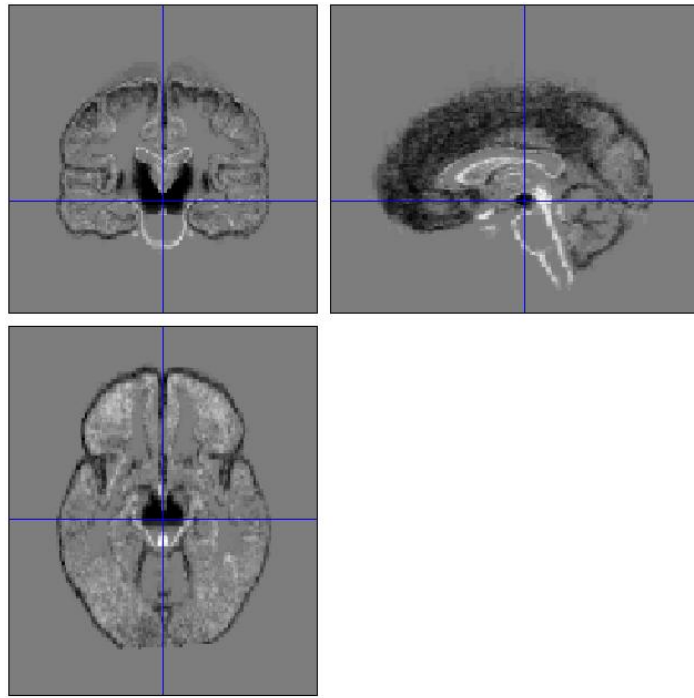*Fig. 3.22 - Mean image of false positive-negative from FAST.*

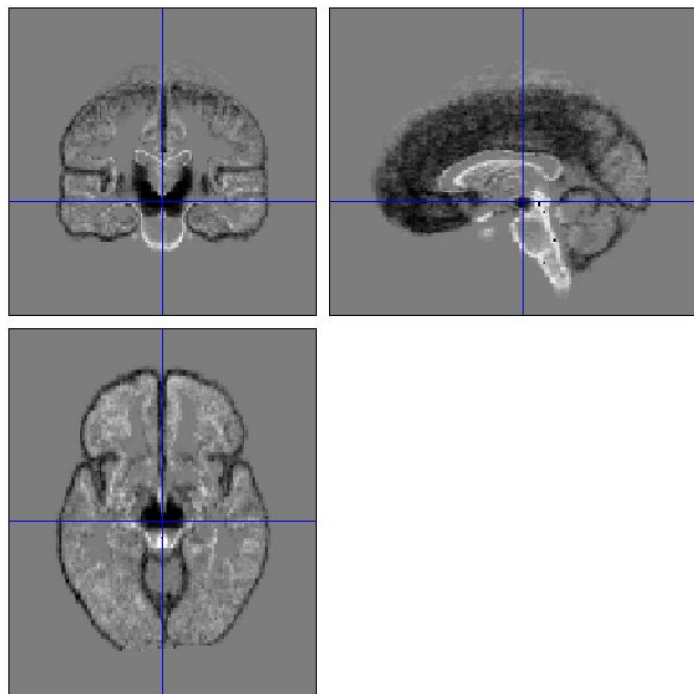*Fig. 3.23 - Mean image of false positive-negative from New Segmentation.*

Fig. 3.24 - Mean image of false positive-negative from Unified Segmentation.

Table 2.4 shows the number of voxels in the mean images. False positives are in column GM(+), and the false negatives in GM(-).

	GM (+)	GM (-)
FAST	76612	278951
NS	97189	227608
US	111381	217696

Table 3.4 – Number of false positive and negative voxels in the mean images.
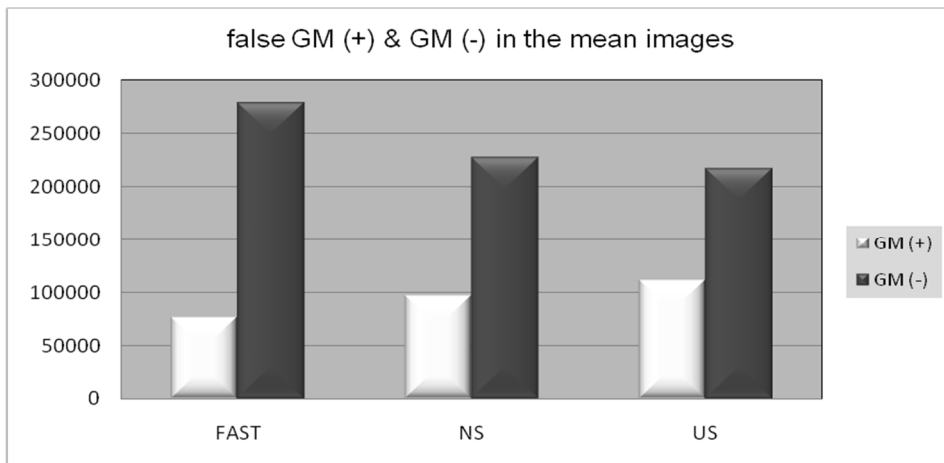
Fig. 3.25 – Bar diagram showing the amount of false positives (GM (+); in white) and false negatives (GM (-); in black)in the mean image obtained from the three pipelines.

Mean images from 50% contribution

In these mean images, voxels (Figure 3.26) show where false-positives occurred in at least 50% of the samples. Differences between the mean from FSL and SPM were noticeable. The areas where the algorithms assumed extra or lack of GM tissue were located in different areas of the brain.

a. FAST

b. New Segmentation

c. Unified Segmentation

Fig. 3.26- Different slides of the three means with 50% sample contribution:a. FAST; b. New Segmentation, and c.Unified Segmentation.

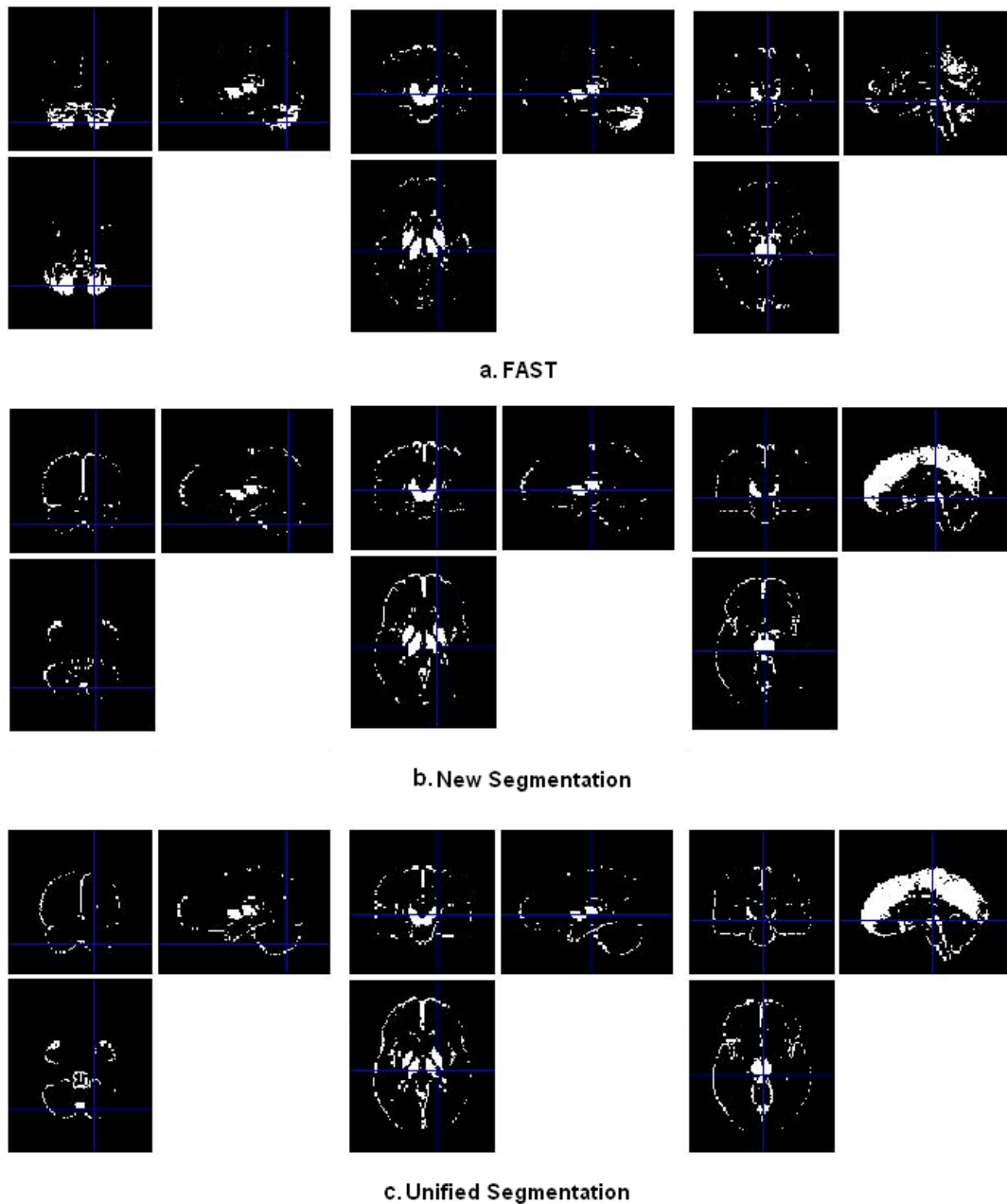Next, mean images were divided into false-positive and false-negative, separately. These images indicate the locations where algorithms over- or under-segmented the GM tissue class (Figures 3.27 and 3.28).
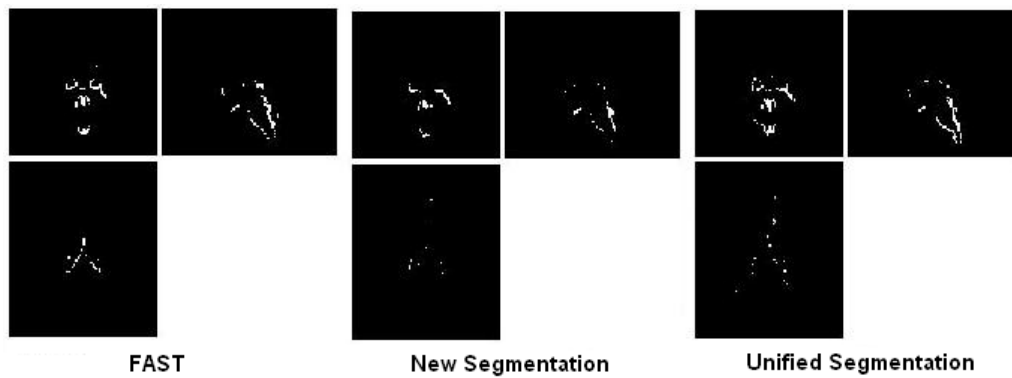
Fig. 3.27- Mean images from 50% of the false positives of the sample.
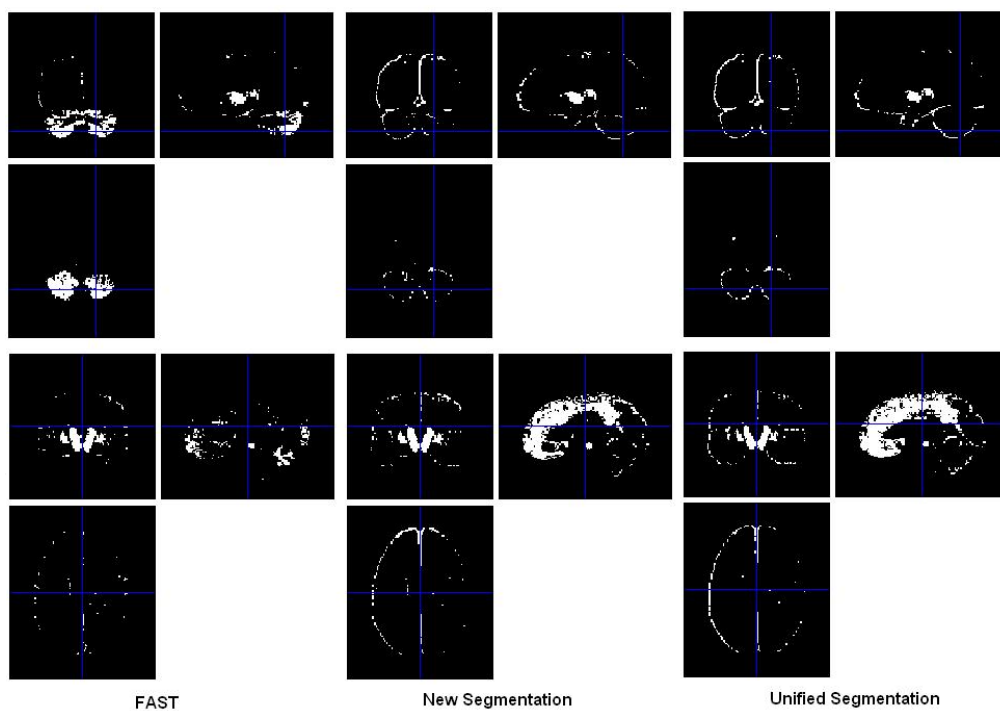
Fig. 3.28- Mean images from 50% of the false negatives of the sample.

False positives did not show substantial differences between algorithms. Interestingly, the location of the areas converged. However, regarding false negatives, the areas where algorithms incorrectly classified GM had slightly different locations.

Mean (50%)	GM (+)	GM (-)
FAST	2790	21995
NS	1969	24727
US	3326	27056

Table 3.5- number of voxels in false GM(+) and GM(-) mean images, contribution 50% of the sample.
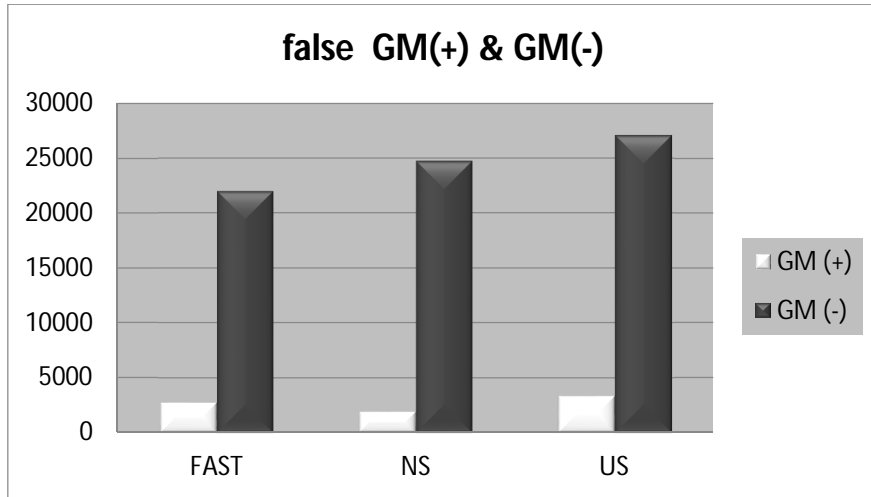
Fig. 3.29 – Table 2.5 represented in a boxplot.

New Segmentation contained the least false positives, followed by FAST and then Unified Segmentation. For false negative voxels, FAST contained the least, followed by New Segmentation, and then Unified Segmentation.

Mean images from 80% contribution

The mean images of false positives/negatives with 80% of sample contribution were created.
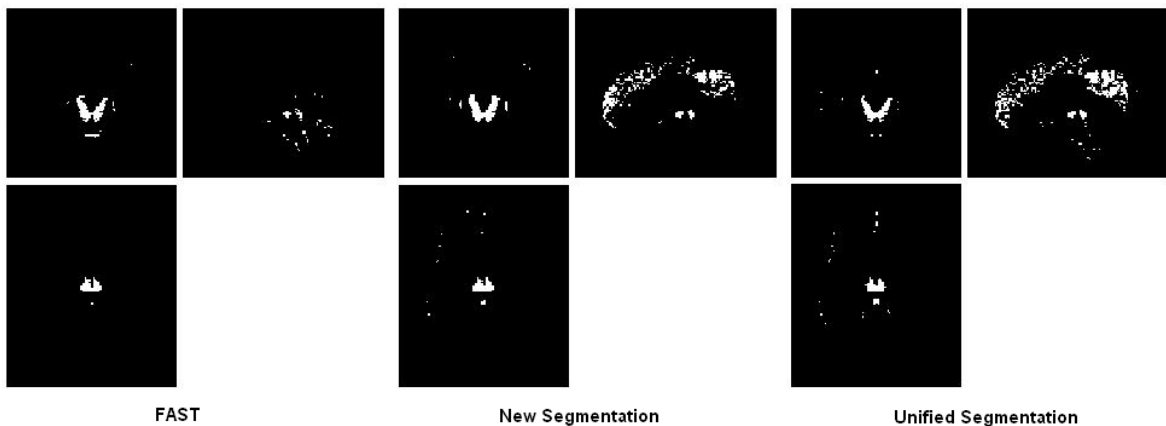
Fig. 3.30 – Mean imagess of 80% of the sample contribution.

The number of voxels decreased considerably. False positives and negatives were separated and showed a similar trend to the previous case (50%). Regarding false positives, there were no substantial differences among algorithms, and their location was fairly consistent. Regions where false negatives occurred tended to vary across algorithms.
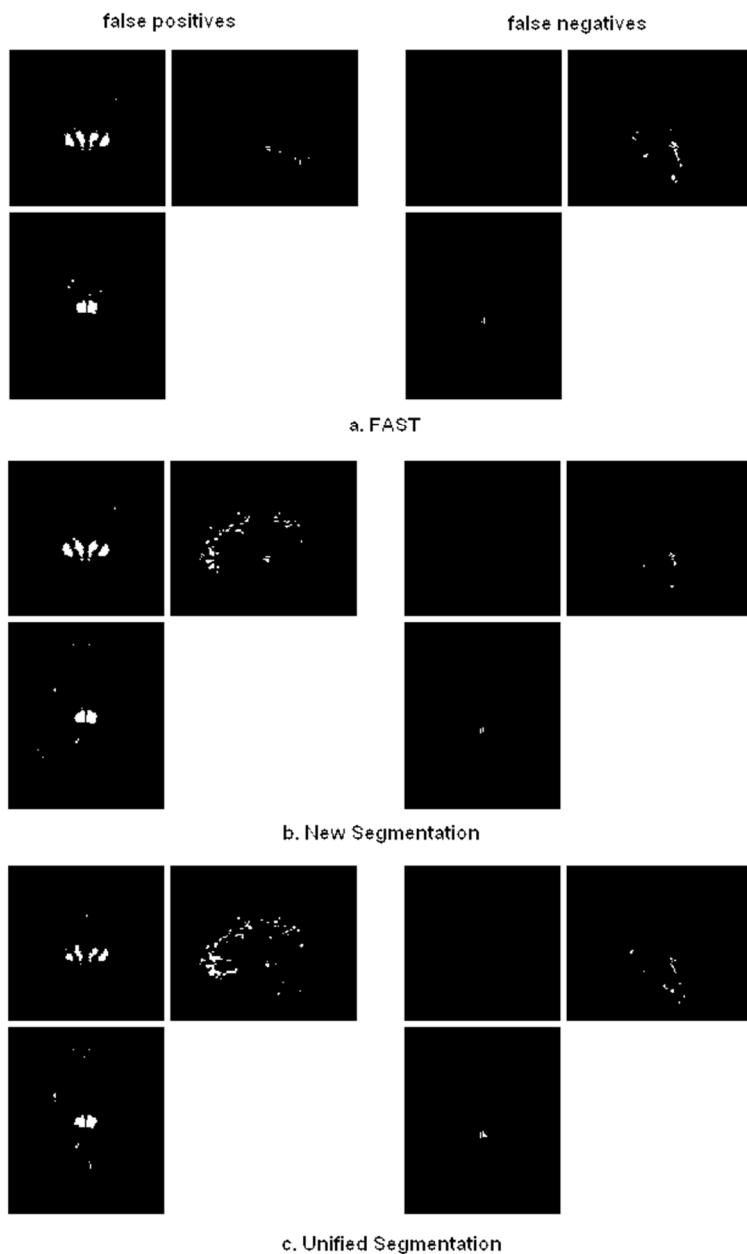
Fig. 3.31 – Mean images from false positives and negatives; Contribution 80% of the sample.

New Segmentation was the algorithm with the fewest false positives, followed by FAST, and then by Unified Segmentation. FAST contained the fewest false negative voxels, followed by Unified Segmentation, and then New Segmentation.

Mean (80%)	GM (+)	GM (-)
FAST	321	3313
NS	188	4313
US	358	3578

Table 3.6- Number of voxels in false GM(+) and GM(-) mean, concerning the contribution of 80% of the sample.

Fig. 3.32- Table 2.6 represented in a boxplot.

Location of false positive and negative

The next table reports the locations of the areas that appeared in 50% and 80% of sample contribution to the mean images.

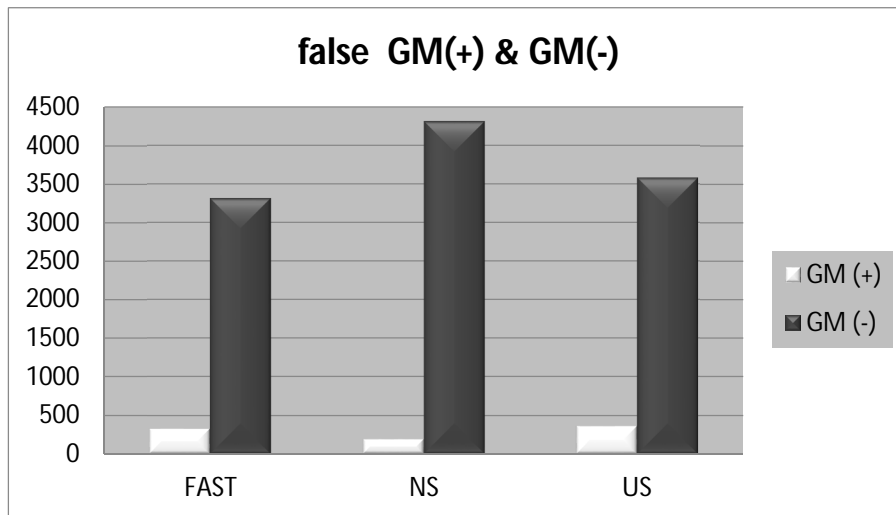| Mean 50%	false positive location	false negative location
FAST	anterior cerebelar artery edge of corpus callosum & ventricles cerebral peduncle, pons, and medula	frontal superior calcarine basal ganglia (caudate, putamen, and thalamus) vermis & cerebellum
New Segm.	edge of corpus callosum & ventricles cerebral peduncle, pons, and medula	edge of external GM surface medial longitudinal fissure basal ganglia (caudate, putamen, and thalamus)
Unif. Segm.	edge of corpus callosum & ventricles cerebral peduncle, pons, and medula	edge of external GM surface medial longitudinal fissure basal ganglia (caudate, putamen, and thalamus)
Mean 80%	false positive location	false negative location
FAST	tin edge of ventricles around post hippocampus cerebral peduncle, pons, and medula	basal ganglia (caudate, putamen, and thalamus) vermis & cerebellum
New Segm.	tin edge of ventricles around post hippocampus corpora quadrigemina	medial longitudinal fissure basal ganglia (caudate, putamen, and thalamus)
Unif. Segm.	tin edge of ventricles around post hippocampus cerebral peduncle, pons, and medula	medial longitudinal fissure basal ganglia (caudate, putamen, and thalamus)

Table 3.7 – Location of the areas reported as false positive and negative associated to the 50% and the 80% of the contribution to the means.

Density of voxels in false positive and negative areas

Volumeof GM in the false positive and false negative images was calculated from the mean files, it is detailed in the table 3.8.

Mean (50%)	vol GM(+) (l)	vol GM(-) (l)
FAST	0.0058	0.0468
NS	0.004	0.0535
US	0.0069	0.0573

Mean (80%)	vol GM(+) (l)	vol GM(-) (l)
FAST	0.0009	0.0098
NS	0.0005	0.0132
US	0.001	0.0105

Table 3.8 – GM volume of false positive and negative in 50% and 80% of contribution.

Density was thought as the volume obtained from the GM that remained in the false positive and false negative images (Table 3.7) divided by the number of voxels counted in the corresponding false positive and false negative images.

Mean (50%)	dens (+)*(10^-8)	dens (-)*(10^-8)
FAST	207.88	212.77
NS	203.14	216.36
US	207.45	211.78

Mean (80%)	dens (+)*(10^-8)	dens (-)*(10^-8)
FAST	280.37	295.80
NS	265.95	306.05
US	279.32	293.46

Table 3.9 – Estimation of density of GM from false positive and negative in 50% and 80% of contribution.
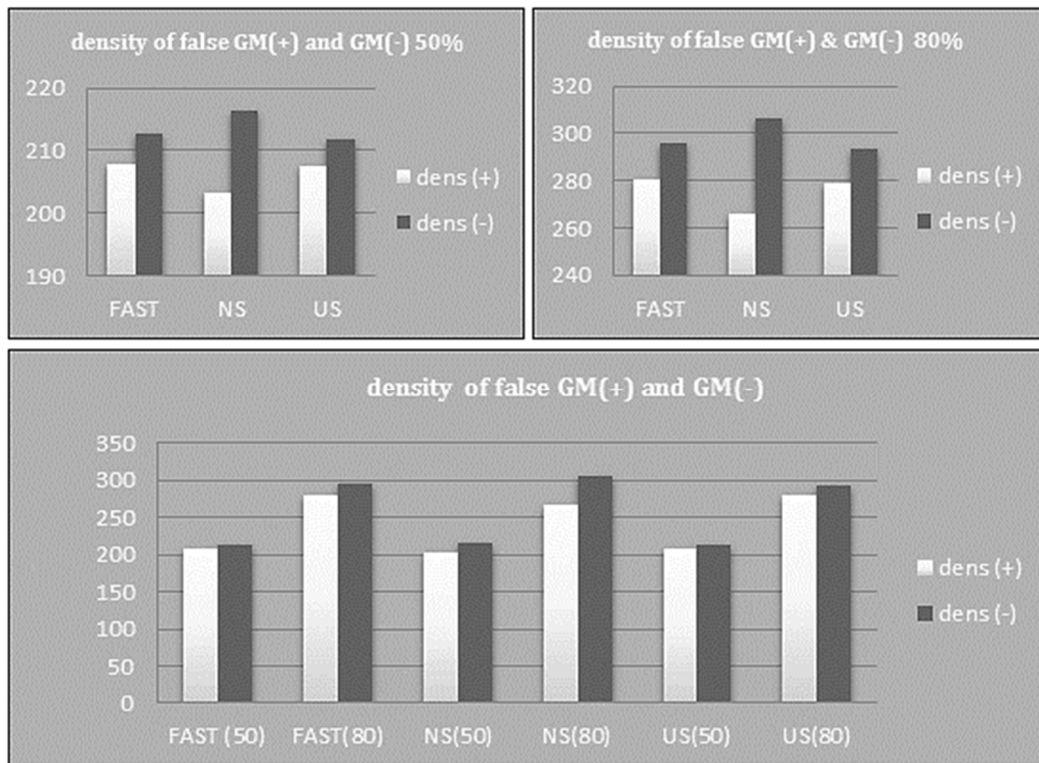
Fig. 3.33 – Bar plots of density from false positive and negative. The trend was the same in both situations. (top-left) 50% mean and (top-right) 80% mean.

The concept of density gave an indication of how many times a voxel was a false positive or negative for each algorithm. In both cases, at 50% and 80% of the sample contribution, the New Segmentation algorithm gave a lower density of false positives. It was followed by the Unified Segmentation and finally FAST. On the other hand, Unified Segmentation showed the lowest density of false negatives. This was followed by FAST, and finally by New Segmentation.

4. Study 2: Feature selection for structural pattern recognition studies

A popular approach for analysis of structural MRI datais the concept of Voxel-Based Morphometry (VBM) (Wright et al. 1995; Ashburner and Friston 2000). First, data are preprocessed by classifying scans into different tissue types, which are aligned into a common anatomical space. Following preprocessing, a statistical analysis is performed (under some assumptions) to enable the comparison among populations. A mass-univariate approach underlies the resulting maps in the SPM framework (Friston 1994). However, variability of brain structural patterns is of a multivariate nature, so multivariate approaches may provide a more veridical framework. Powerful machine learning techniques from other fields have been adopted to obtain a more accurate understanding of the different processes that occur in the brain. Having a more accurate model may allow more rapid translation from basic research into clinical applications. This is the role of multivariate techniques in neuroimaging (Ashburner and Klöppel 2011).

Nowadays, brain imaging researchers aim to collect the largest possible number of subjects. Studies attempt to obtain findings that can generalize, for instance differences between populations or the impact of a biomarker, etc. In general, these require some form of characterization of inter-subject neuroanatomical variability. Capturing much of the inter-subject variance among brain images involves shape modeling (computational anatomy), where the accuracy of inter-subject registration plays a significant role in terms of the findings and their interpretability (Ashburner and Klöppel 2011). Conclusions from any particular study depend heavily on how the data are modeled, and the assumptions underlying those models.

There are many different approaches to modeling such characterizations. Successful multivariate methods do not necessarily localize differences, but instead aim to capture the patterns of difference that best separate subjects into groups, or predict some continuous variable of interest. Although useful for analyses that attempt to localize differences, the widely used assumption of independence among the anatomy of different brain regions is not really biologically plausible. If this assumption is removed, greater predictive accuracy may be possible. For example, age or gender differences are not localized to any particular brain region. Instead, there is a pattern of differences that are distributed throughout the brain. Studies based on localizing differences only show the

most relevant aspect (like the tip of an iceberg) but they can lose important information pertaining to patterns of relationships among brain regions. A pattern recognition approach attempts to learn a relationship between feature data and their corresponding labels. The algorithm, after learning such a relationship, should be able to predict the label for new data cases.

Many different algorithms for pattern recognition analysis are available. For this thesis, Gaussian processes for classification and regression have been used (Rasmussen and Williams 2006). Gaussian Processes (GP) are kernel-based approaches, set in a Bayesian framework. They achieve similar performance to Support Vector Machines (SVM) for neuroimaging data (Schrouff et al. 2013) with the advantage that they make probabilistic predictions. These supervised algorithms learn the mapping between the input (data features) and its output (labels) from a set of training data. Depending on whether the output is continuous or discrete, it would be respectively a classification or a regression problem.

4.1 Motivation & Objectives

The motivation for this work is to determine an effective feature representation of image data from multiple subjects, such that pattern recognition methods may be made more accurate. This involves applying GP machine learning approaches to a number of sets of features, derived from the same subjects' scans, in order to predict the ages and genders of the subjects. The hope is that this should allow an effective feature representation to be selected, prior to further work using data from different populations of subjects. When machine learning is applied to relatively small datasets from patient populations, it is important to determine how best to do this beforehand. It would not be good science to try out lots of methods, and selectively report only those that worked the best. This would be an ineffective use of valuable data.

The objective of this study is (1) to find, if possible, the types of feature derived from T1-weighted MR images that are more effective, irrespective of the target for regression and classification. The second objective is (2) to find a good range of spatial smoothing to work with.

4.2   Material & Methods

A relatively large public data set with respective demographic variables was used in this work. Several sets of features were obtained from it, using a VBM-type pre-processing. With them, different analyses were carried out to explore the performance dependency on the features: A GP regression (implemented as a Bayesian ridge regression analysis) to predict age and body mass index (BMI), and a GPclassification analysis to predict gender.

4.2.1   Dataset

The IXI data set, provided by the Biomedical Image Analysis Group[3], consists of a variety of MR images from nearly 600 normal, healthy subjects with their respective demographic information. Only the T1-weighted images were used. MRI data were acquired in three different scanners, two of which were 1.5T and one was 3T. Table 3.1 summarises the demographics of the sample. Subjects whose variables were not available were not included in the analyses. Some "data scrubbing" was performed by identifying variables which were obviously incorrect, and also excluding these from the analysis.

IXI dataset	scanner	sample size (N)	age (years)	gender (male/female)	height (cm)	weight (kg)
Hammersmith Hospital	Philips 3T	181	47.36±16.70	87/94	170.00±9.54	72.84±13.29
Guy's Hospital	Philips 1.5T	313	50.83±15.89	138/175	169.53±8.95	71.79±15.89
Institute of Pschiatry	GE 1.5T	68	42.37±16.60	24/44	168.62±10.36	68.95±12.14
		562	46.85±16.40	249/313	169.38±9.61	71.19±13.77

Table 4.1 - demographics of the sample.

4.2.2   Preprocessing

The T1-weighted images were visually inspected for possible artifacts, and approximately aligned with the SPM template data (via translations). Next, a VBM-type pre-processing was conducted. The default segmentation tool implemented in SPM12 was used for segmenting the images. The SPM12 implementation is an updated version of "New Segment" used in the Study 1. It is based on the algorithm presented in (Ashburner and Friston 2005), but makes use of additional tissue classes, allows multi-channel segmentation (of i.e.T2 and PD-weighted images), and incorporates a more flexible image registration. Changes from the SPM8 New Segment include different regularization for the deformations, some different default settings, as well as re-introducing the re-scaling of

---

[3] available at *http://biomedic.doc.ic.ac.uk/brain-development/index.php?n=Main.Datasets*

the tissue probability maps (which was in the old segment, but not the new). In addition, the SPM12 tissue probability maps were re-generated using the T2-weighted and PD-weighted scans from the same IXI dataset. (For a more extensive description of the differences, see the appendix of (Malone et al. 2015).

Following tissue segmentation, inter-subject registration was performed using the Geodesic Shooting Toolbox, also implemented in SPM12. The approach is described in Ashburner and Friston (2011), and is superficially similar to Dartel (Ashburner 2007). Evaluations show that this model achieves more robust solutions in situations where larger deformations are required (Ashburner and Friston 2011).

4.2.3 Data for structural feature representation

Preprocessing outputs were used as features. Each feature representation encodes a different kind of information about the original image data. The field of view of the feature data covers the whole brain, and the features that have been used are listed in figure 4.1.
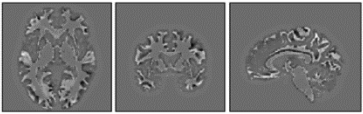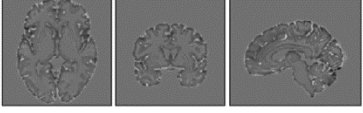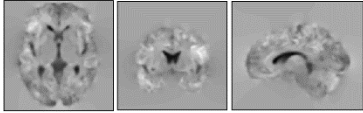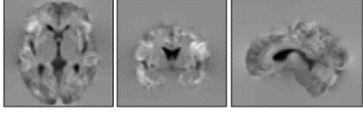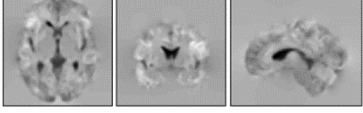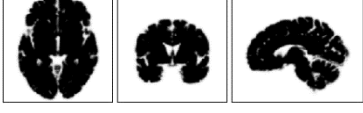
Data selection for feature representation		
Images from preprocessing	Scalar momenta	
	Divergence of velocities	
	Jacobians	
	Logarithm of Jacobians	
	Grey matter (GM)	
	White matter (WM)	
	Background tissue (BG)	
	Jacobian-scaled GM	
	Jacobian-scaled WM	
	Jacobian-scaled BG	

Fig. 4.1 - Resulting images from the pre-processing and those obtained from operations, like the log. of the Jacobians and the Background images, BG= 1-(GM+WM). All these images were tested as features.

Smoothing

Spatial smoothing of varying degrees was applied to the raw feature images. This was intended to reduce noise and finer grain anatomical variability. Exploring the optimal degree of smoothing is a relevant factor. If there is too much blurring, the image may lose relevant information. On the other hand, if there is not enough blurring, differences may be not be detected. The effect of smoothing was explored in order to find the optimal

amount of smoothing for each feature type. For this purpose, all data were filtered using a range of smoothing from 0mm to 20mm FWHM. The 10 feature representations were smoothed using the 21 measures of smoothing. The 210 sets were separately studied as inputs in the pattern recognition approach.

Kernel matrices

In the Gaussian processes representation, from the viewpoint of linear regression, feature data are introduced into the model in the form described in equation (1.23), becoming the covariance function, or kernel function.

An image can be thought of as a vector in a very high-dimensional space (Fig. 4.2 (left)). If a brain image would have only three voxels, it could be visualized as a point in a 3D space. In practice though, the number of dimensions needed to represent a real image in this way makes it impossible to visualize. For binary classification problems, it is possible to define a hyper-plane dividing regions of high-dimensional space, such that one class falls in one region and the other class in the other. Class membership is then assigned to new data, depending which side of this hyper-plane it falls. Similar principals apply to regression, where a continuous label is predicted instead (Ashburner and Klöppel 2011).

$$X_1 = (x_{11}, x_{12}, ..., x_{1k})$$
$$X_2 = (x_{21}, x_{22}, ..., x_{2k})$$
$$...$$
$$X_n = (x_{n1}, x_{n2}, ..., x_{nk})$$
$$\left. \right\} \{X_1, X_2, ..., X_n\}$$

Fig. 4.2 – Representation of a 3D image in a feature vector.

A sample can be thought as a set of *N* vectors each one with *k* components, where *k* is the number of voxels (Fig. 4.2 (right)). The kernel matrix obtained by computing **XX**^{T} is an *N×N* matrix that encodes the similarities among the images.
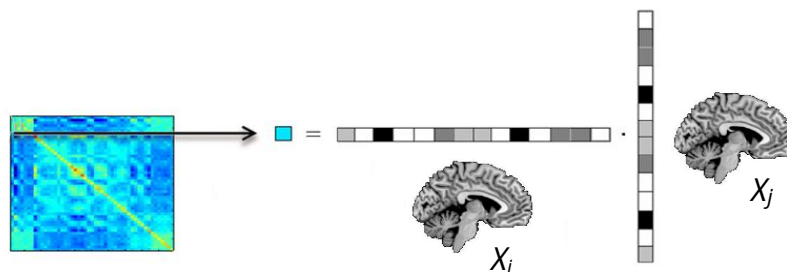
Fig. 4.3 - Representation of how a kernel matrix is created from images using the dot product.

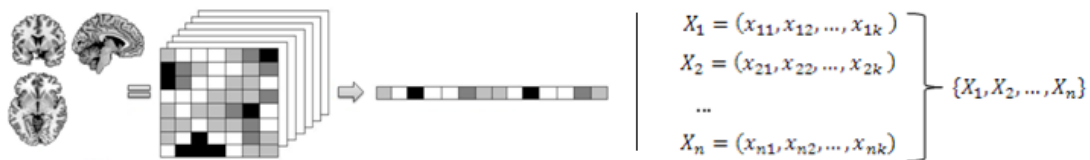Thus, the *k* dimensions encoded in each image are reduced to *N*. When spatial smoothing is used, this may be conceptualized as constructing a kernel matrix from $\mathbf{X}\Sigma\mathbf{X}^T$, where Σ is a Toeplitz matrix that enacts the spatial smoothing.

In the current study, kernels were constructed from the 210 feature sets described above. Several new kernels were also studied, which were constructed by summing some of the original kernel matrices together. Separately each tissue class encodes information about itself regarding the target. The tissue class kernel matrices were combined as follows:

Data selection for feature representation	
Combination of images	GM + WM
	GM + WM + BG
	Jacobian-scaled GM + Jacobian-scaled WM
	Jacobian-scaled GM + Jacobian-scaled WM + Jacobian-scaled BG

Table 4.2 – multi-kernels used as features.

This was done over all levels of smoothing, so 84 new kernels were added to the initial number of 210 kernels.

Adding a variable to the model: Modulated data adjusted for Total Intracranial Volume

In this thesis, a procedure is proposed for dealing with total intracranial volume (TIV). Usually, when modulated data are used in a VBM analysis, TIV is either modeled as a confounding covariate, or is used to normalize the preprocessed data by dividing regional tissue volumes by the TIV. It is useful to include it if the aim is to observe local differences between groups (Mechelli et al. 2005). Modulated data are corrected for regional expansion and contraction incurred during the warping process, such that the original regional tissue volumes are preserved. In this work, we normalise the modulated tissue maps by dividing them by the TIV. This provides an additional set of features. For convenience, the kernel matrices - rather than the features themselves - were normalized (because $(\mathbf{x}_1/s_1)\cdot(\mathbf{x}_2/s_2) = (\mathbf{x}_1\cdot\mathbf{x}_2)/(s_1 \cdot s_2)$, where s_1 and s_2 are TIVs). This process generated an additional 105 kernel matrices. Results obtained from kernels that accounted for TIV were compared to those that do not account for it, modulated and non-modulated.

The kernel matrices were used as inputs in a pattern recognition algorithm for regression when the labels were continuous (e.g. age), and for classification when these were discrete (gender). Gaussian process models were used to make the predictions. Specifically, the ridge regression model described above was used for predicting age and body mass index (BMI = weight (kg)/height (m^2)). The classification model based on (Rasmussen and

Williams 2006) was used for predicting gender. Both algorithms (GP for regression and classification) are available as part of SPM12.

4.3 Performance evaluation

Generalization performance was assessed using a *k*-fold cross-validation (CV) strategy, which allows most of the sample to be used during the training stage. The original sample is randomly partitioned into *k* equal sized subsamples. One of these subsamples is retained to be used as validation data for testing the model, and the remaining subsamples (*k-1*) are used for training the model. The cross-validation process is repeated *k* times (*folds*) with each of the *k* subsamples used once as test data. In those cases when *k* is equal to the number of observations the *k*-fold CV becomes a leave-one-out cross-validation (LOO-CV)[4]. This study used a 10-fold cross validation (CV), with the same subdivision into folds for all kernel matrices.

For Regression

For regression, the model assigns a predicted value of the corresponding target. It also provides estimates of the uncertainty in the form of standard deviations, but these were not used. The root mean squared (RMS) error was estimated for each model (*i* = {age, BMI}, features *k*= {GM, WM, scalar momentum, jacobians, etc...} and smoothing *s*= {0, 1, ..., 20}) as follows:

$$RMSerror_{iks} = \sqrt{\frac{\sum_{j=1}^{N}(m_j - t_j)^2}{N}} \qquad (4.1)$$

Where m_j corresponds to the real value, t_j is the predicted value, and N is the number of observations. The RMS error gives a measure of how well the model can predict the real data, and allows a comparison to be made between feature sets. RMS error was plotted against the degree of smoothing, for kernel matrices of all feature types.

Moreover, the log-marginal likelihood[5] was considered. Predictive results are dependent on the values of the hyper-parameters that define the behavior of the model. The optimal choice for these parameters may be obtained by finding values that maximize the marginal likelihood of the model. Then, the greater the log-marginal likelihood, the better is the model fitting (Rasmussen and Williams 2006). Log-marginal likelihoods, as well as the RMS error, are plotted in figures 4.4 and 4.6 for each feature set, over the range of

---

[4] *https://en.wikipedia.org/wiki/Cross-validation_(statistics)*
[5] *https://en.wikipedia.org/wiki/Marginal_likelihood*

smoothing. For the best data feature, the plot of the real target values versus the predicted values are shown in figures 4.5 and 4.7.

For Classification

For classification, instead of obtaining a predicted variable (plus standard deviation) for each observation, a probabilistic label of belonging to one class or another is predicted. Due to the nature of the outcomes, other measures were considered to estimate the performance.

The area under the curve (AUC) of the ROC curve was calculated. The ROC curve shows the performance of the classification and is obtained using the ratio of true positives versus the ratio of false positives as the discrimination threshold varies (threshold above which it is said that a case is positive or not). The area under this curve is a measure of how well the classifier has performed, being a summary of the performance of the classifier across all decision thresholds[6]. When a classifier makes the perfect discrimination the AUC is 1, when a binary classifier is guessing at chance-level, it would achieve an AUC of around 0.5. Also, the log-marginal likelihood was computed. Log-marginal likelihood and AUC wereplotted for all the feature datasets versus the smoothing in figure 4.8. For the best data feature, the ROC curve is shown in figure 4.9.

Evaluation

Results obtained from using the feature datasets were plotted. For the log-marginal likelihood and the AUC, the higherthe curve, the better the model fit. For the RMS error, the lower the value, the better the model fit.

Plots of results from all features together would be confusing and difficult to interpret. Therefore, depending on the origin of the data, plots have been represented by grouping features. Results from all those features that are implicitly part of the registration process, i.e. Jacobians, log-Jacobians, divergence of velocities, and scalar momentum, were presented together. Results from GM, WM, BG, and combinations of these were grouped, with results from modulated features shown separately from those without modulation. To make comparisons easier, all the plots related to the same target measure are shown with the same axis scale.

In order to conclude which the most efficient feature is, the optimal smooth-feature was selected from their respective set of features according the log-marginal likelihood. RMS

---

[6] *https://en.wikipedia.org/wiki/Receiver_operating_characteristic#Area_under_curve*

error for regression or the AUC for classification were also considered when assigning rankings.

For comparing one model versus another, the Bayes Factor was used (Kass and Raftery 1995; Jeffreys 1961). Given a pair of models with the same number of hyper-parameters, the plausibility of the two different models *M1* and *M2* may be (approximately) assessed by the Bayes factor as

BF = 2·log(K) = 2·log(marginal likelihood(M1))−2·log(marginal likelihood(M2))   (4.2)

The BF value can be interpreted by means of the scale defined by Kass and Raftery (1995). This scale varies from <0 to >10, and is divided into blocks. An increase in the BF provides greater evidencefor one model against other.

2·log(K)	K	Strength of evidence
0 to 2	1 to 3	not worth more than a bare mention
2 to 6	3 to 20	positive
6 to 10	20 to 150	strong
>10	>150	very strong

Table 4.3 – Scale provided by Kass and Raftery for interpreting the Bayes factor[7].

These representations of estimates, the log-likelihood and the RMS error/AUC, provide complete and robust information about the performance from each feature dataset, and allow the comparison between them from different, but complementary, perspectives.

---

[7] *https://en.wikipedia.org/wiki/Bayes_factor*

4.4   Results

Regarding the feature datasets, in all cases the log-marginal likelihood and the RMS error/AUC are in broad agreement with respect to the scores. An interesting aspect, mainly related to the RMS error and the AUC, is that the curves were smooth and the best scores covered several mm of smoothing (Figures 4.4, 4.6 and 4.8).

4.4.1   GP Regression

Age

Interestingly, the best performance was obtained from the scalar momentum, which provided the most efficient feature at 11mm of FWHM. This was closely followed by three different combinations of kernels from tissue types that provided good features for predicting age. Results from the four best scores are summarized in table 3.3.

feature	smooth (mm FWMH)	log-marg. likelihood	RMS error (years)
scalar momenta	11	-1882.0	6.2
modGM + modWM + modBG - acc. for TIV	8	-1883.6	6.2
GM+WM+BG	11	-1884.0	6.2
GM+WM	11	-1886.0	6.2

Table 4.4 – Best scores in age regression. The average age of all subjects is 46.83 years, and standard deviation is ± 16.41 year The RMS error is markedly below this.

Remarkably, the RMS error is far below the standard deviation associated with guessing that all subjects are of average age, which is considerably greater (±16.41 years).

Log-marginal likelihood (on the left side) and the corresponding RMS error (on the right) associated with the same measurements are plotted in figure 4.4. When marginal likelihoods from best features are compared by means of the Bayes factor, differences between models are greater than 2 in all cases. Comparison between scalar momenta with the second in the ranking is 3.2 according the Kass and Raftery's scale, indicating that differences are positive. Difference estimated with respect to the GM+WM achieved 8 points in the scale, indicating strong evidence that scalar momenta outperformed this combination of kernels.

Log-marginal likelihood   **&**   **RMS error**

Fig 4.4 -AGE: (a) Scalar momenta, Jacobians, log-Jacobians and divergence of velocities.(b) GM, WM, BG and combinations:GM+WM and GM+WM and GM+WM+BG. (c) Scaled GM, WM, BG and combinations of scaled GM+WM and scaled GM+WM+BG.(d) Scaled data adjusted for TIV.

When plots are observed, it is noticeable that differences in log-marginal likelihood for each feature markedly varied with the smoothing. Bayes factor differences overtook the strong evidence from 0 to 20 in most of the cases. With smoothing, the feature from scalar momentum outperformed the rest and good results were generally obtained from kernel matrices created from the sum of kernel matrices from GM, WM and BG.

Without any smoothing, the logarithms of the Jacobians provided very effective features. A surprising result was that modulated GM was not a very effective feature for making age predictions.

A combination of kernel matrices from GM, WM and BG modulated provided better performance when accounting for TIV than those without TIV. The Bayes factor gave very strong evidence for this.

GM encoded more information about age than WM or BG, although its performance was not as good as expected. Interestingly, combining kernel matrices from GM, WM and BG provided better age predictions than using any of these kernel matrices alone.

The best performing feature set for predicting age was the scalar momentum. Figure 4.5 shows real ages plotted against predicted ages, and the best fit through this plot.
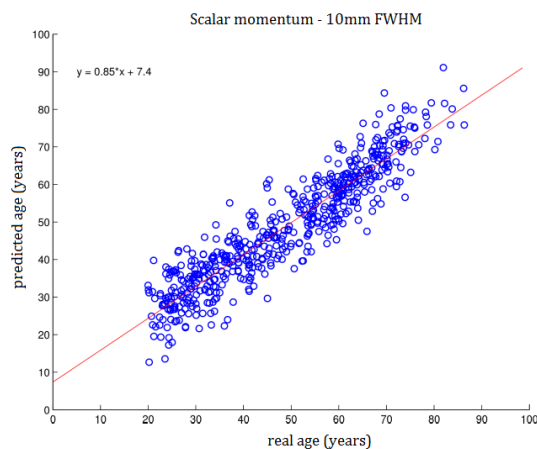
Fig. 4.5- Real versus predicted ages and linear fitting associated to the most effective feature.

BMI

Results showed that WM was a relatively effective feature for predicting BMI. In addition, the combination of the WM with GM and BG reached high scores. The scalar momenta are also in the ranking as one of the best features that better encoded the BMI pattern (Table 4.5). It is worth noting that the RMS error ($3.4 kg/m^2$) associated with the predictions had a similar order of magnitude to the standard deviation, $\pm 3.8 kg/m^2$.

feature	smooth (mm FWMH)	log-marg. likelihood	RMS error (kg/m^2)
WM	14	-1353.0	3.4
GM+WM	13	-1356.0	3.4
scalar momenta	16	-1357.0	3.4
GM+WM+BG	11	-1358.0	3.4

Table 4.5– Best scores in BMI regression. Average and standard deviation of BMI was 24.9±3.8 kg/m². RMS error is much lower than the standard deviation.

Plots of log-marginal likelihood (on the left) and the corresponding RMS error (on the right) are shown in figure 4.6.

Modulated data (with and without TIV), Jacobians and divergence of velocities did not provide good features for predicting BMI.

With smoothing, scalar momenta gave similar results to WM. From about 10mm of FWHM, scalar momentum seemed to be a good feature for predicting BMI. However, in comparison to the WM, there was strong evidence according the Bayes factor that WM was a better model than scalar momentum. Same occurred with the combination of GM and WM. Regarding the sum of tissues GM, WM and BG, the Kass Raftery scale stated that there was very strong evidence that WM was a better model.

Performance of WM remarkably increased with the smoothing and provided very effective feature for predicting BMI. Regarding the kernels from combining WM with GM and with GM+BG, their performance showed very strong evidence of being poorer models than only using WM, according to the Kass and Raftery scale for the Bayes Factor. Contribution of GM and BG tissues to WM were noisier than using only WM for predicting BMI.
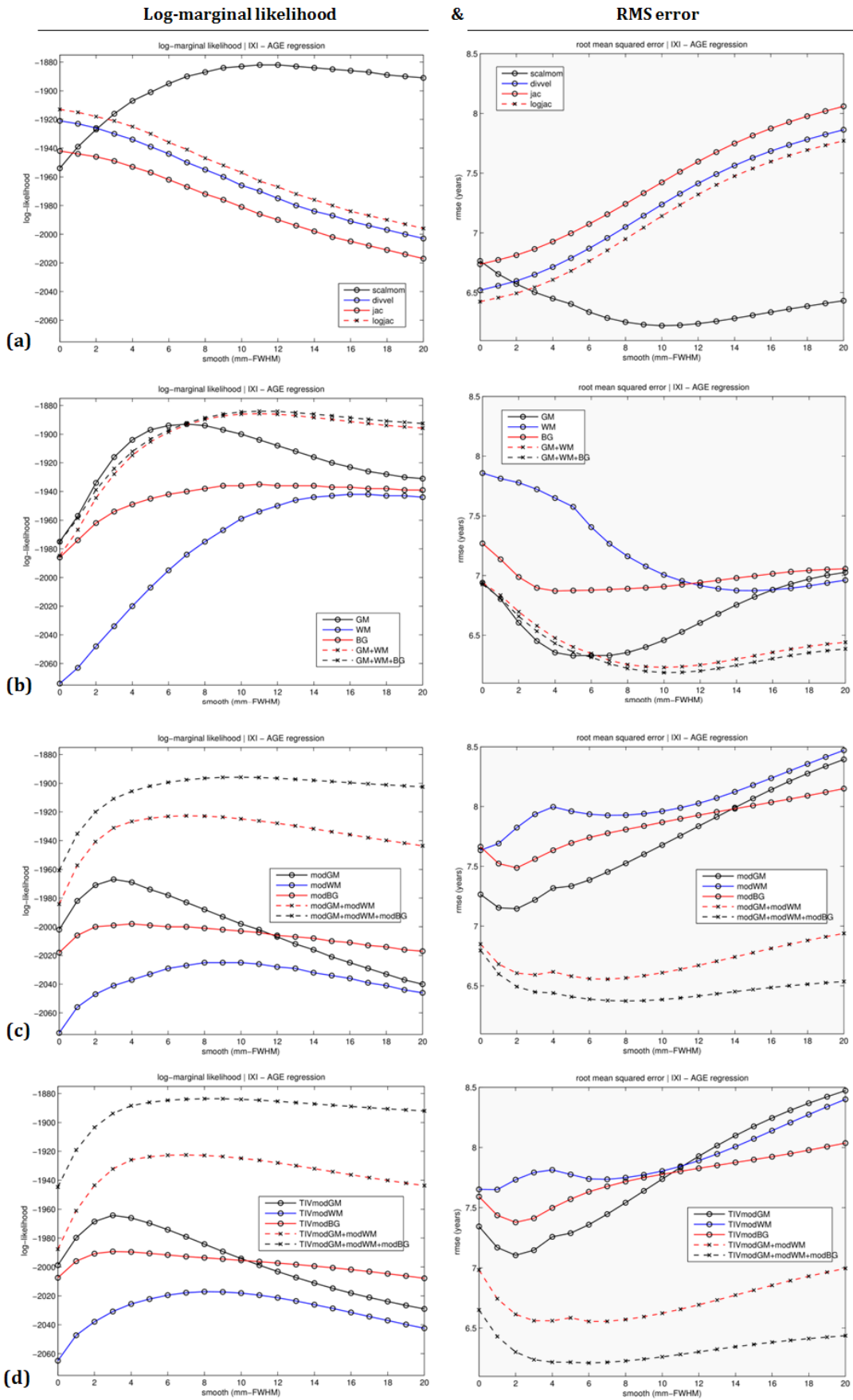
Log-marginal likelihood   **&**   **RMS error**

Fig 4.6 – BMI: (a) Scalar momenta, Jacobians, log-Jacobians and divergence of velocities. (b) GM, WM, BG and combinations: GM+WM and GM+WM and GM+WM+BG. (c) Scaled GM, WM, BG and combinations of scaled GM+WM and scaled GM+WM+BG. (d) Scaled data adjusted for TIV.

Combinations aside, WM and scalar momenta werefeatures that better predicted BMI. Figure 4.7 shows plots of real BMI with respect to predicted BMI, and the linear regression fit for both features.
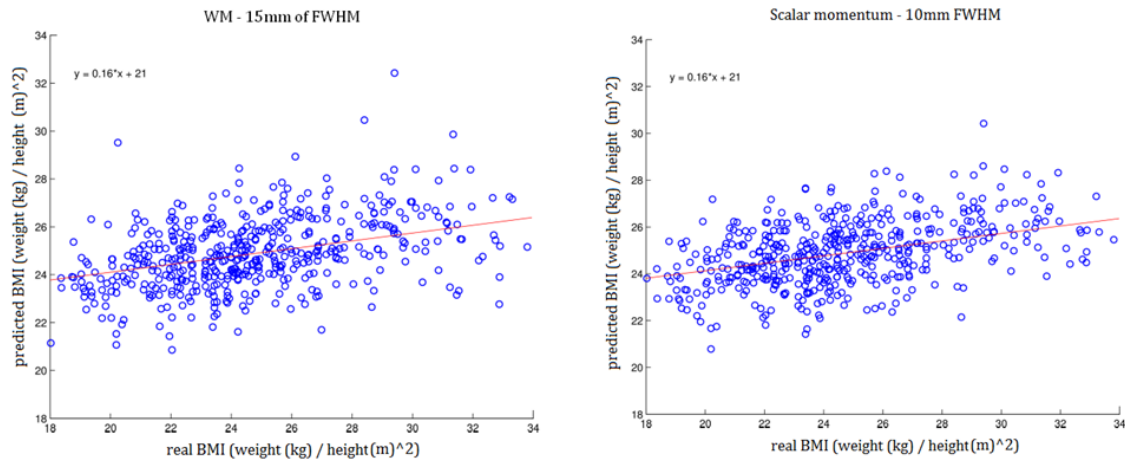
Fig. 4.7–WM and scalar momentum provided best feature representations for predicting BMI.

4.4.2 GP Classification

Gender

Scalar momentum was the best feature for predicting gender. Also, all the possible combinations of GM, WM and BG exhibited high performance. Results are detailed in Table 4.6.

feature	smooth (mm FWMH)	log-marg. likelihood	AUC
scalar momenta	15	-167.96	0.98
modGM + modWM + modBG - acc. for TIV	15	-170.00	0.97
modGM + modWM + modBG	15	-170.00	0.97
GM+WM+BG	15	-177.21	0.97

Table 4.6 – Best scores in Gender classification.

From the listed features, good scores in performance covered a wide range of smoothing degrees. The AUC reached 0.971 at 6mm of FWHM and slightly varied to reach the maximum at 15mm, 0.975. The scalar momenta achieved higher performance from 10mm smoothing upwards. Plots of log-marginal likelihood (on the left) and the corresponding AUC (on the right) are shown in figure 4.8.
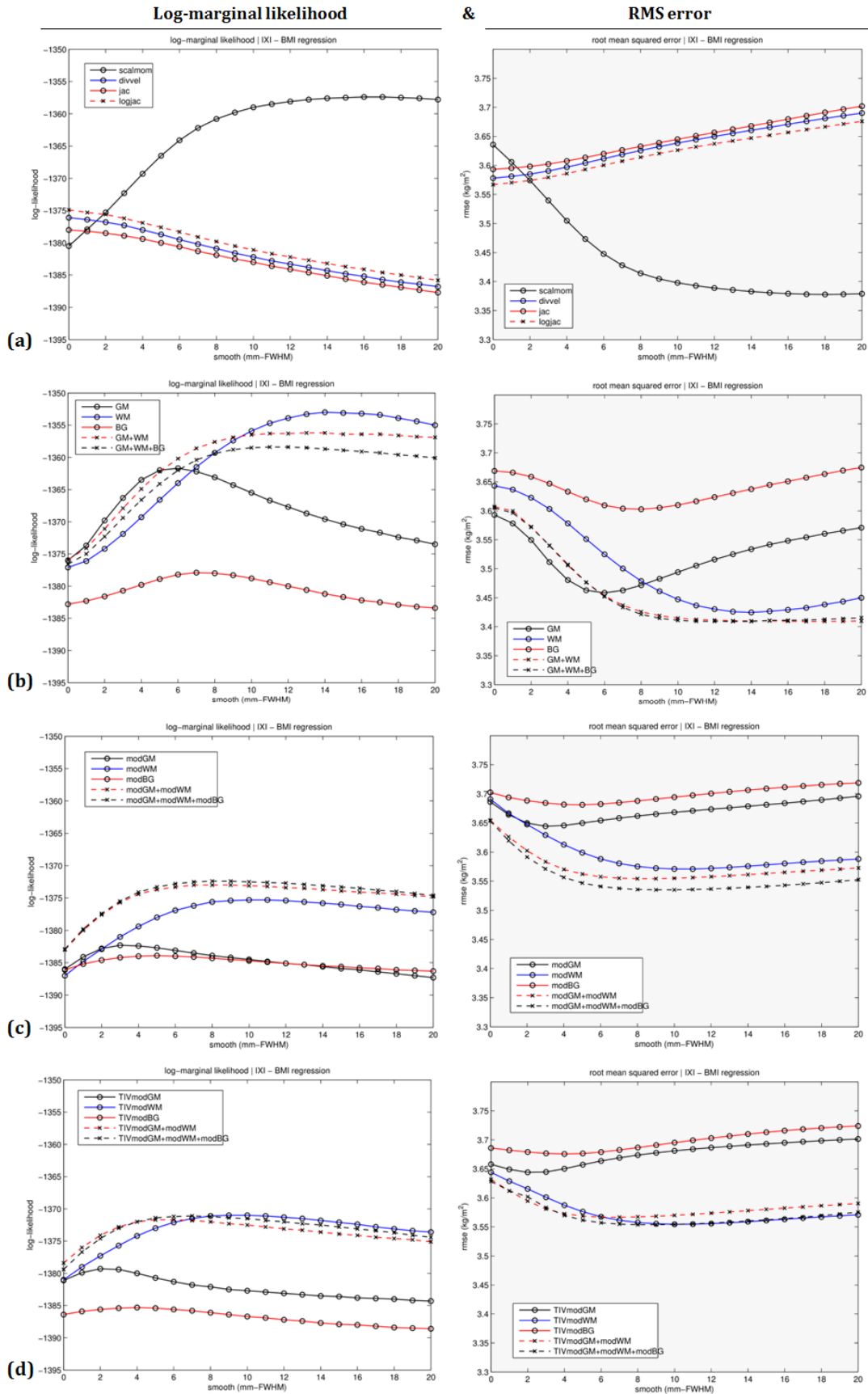
Log-marginal likelihood   **&**   **AUC**
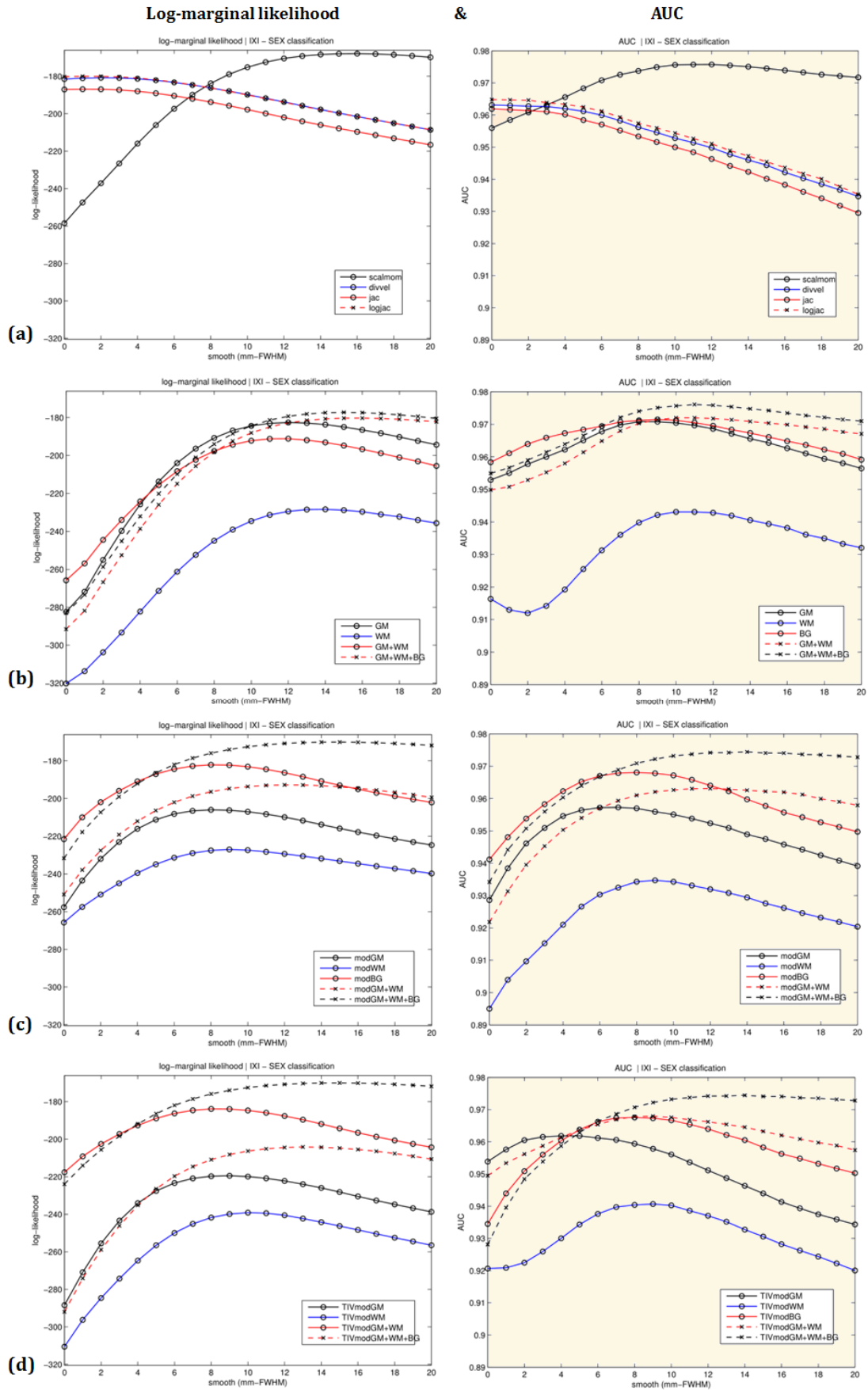
Fig 4.8 –GENDER: (a) Scalar momenta, Jacobians, log-Jacobians and divergence of velocities. (b) GM, WM, BG and combinations: GM+WM and GM+WM and GM+WM+BG. (c) Scaled GM, WM, BG and combinations of scaled GM+WM and scaled GM+WM+BG. (d) Scaled data adjusted for TIV.

WM was the worst feature for predicting gender, whereas BG and GM appeared competitive with the combination of GM, WM and BG.

With smoothing, all tissues and their combinations were better features. The sum of kernel matrices constructed from the three tissue types, with and without modulation, were very effective features for classifying gender.

On the other hand, Jacobians and divergence of velocities were good features. Better scores were reached at lower FWHM sizes, although the dependence of performance on the smoothing was relatively low.

Scalar momentum was the feature that best predicted gender. The ROC curve shows the performance of the classification at 10mm of FWHM, with an AUC of 0.98.
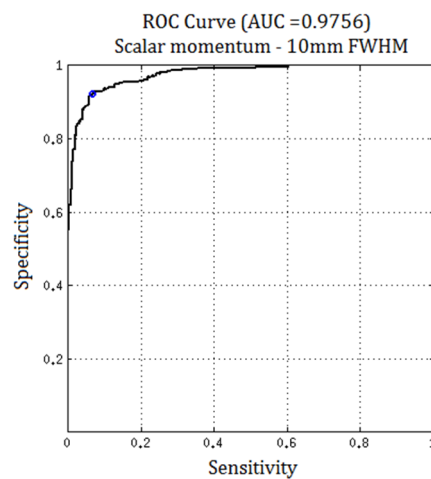
Fig 4.9 – ROC curve and AUC associated to the prediction of gender from scalar momentum at 10mm of FWHM.

5. Study 3: Application to Schizophrenia

One of the main purposes of the current thesis is the application of improved methodologies for clinical research. Finding specific morphometric alterations associated with a particular disease is a widespread goal in neuroimaging research. It has been done in hundreds of studies, mainly from applying Voxel-Based Morphometry (VBM) (Ashburner 2009). Mass-univariate approaches, such as VBM, allow volumetric differences in tissue classes among populations to be localized. However, these methods assume that voxels are independent, which may not be the most biologically plausible assumption to make. A number of works have identified significant amounts of covariance among tissue volumes in different brain regions (Westman et al. 2011; Modinos et al. 2012). Recent developments involve using a multivariate analysis framework, such as pattern recognition approaches. These often provide greater accuracy for characterizing differences between populations of subjects (Schrouff et al. 2013). As shown in study 2, these approaches generally involve pre-processing anatomical MRI scans in the same way as for a conventional VBM study, but then applying a pattern recognition analysis.

In this study, some methodological aspects regarding pattern recognition and features from MRI data, which were previously explored in study 2, have been applied to Schizophrenia.

Chronic Schizophrenia is a widely studied disease that is still an object of study, and provides a clinically useful setting for applying methodological aspects that have been explored in previous chapters. Nowadays, there are consistent findings that have allowed the characterization of the anatomical pattern associated with this disease. Schizophrenia is characterized by a reduction, of around 2%, in whole brain volume. Locally, larger reductions in regions such as the frontal lobe, in about 50% of VBM studies, and hippocampus, insula, temporal and parietal cortices, in about 20%, are reported (Honea et al. 2005). Other regions are also strongly involved in the disease, such as the medial frontal cortex (Bora et al. 2011), which has been proposed as a key region. Ellison-Wright and Bullmore (Ellison-Wright and Bullmore 2010) performed a meta-analysis of 42 VBM studies of schizophrenia, where they found that the distribution of GM volume reductions was wider in schizophrenia than for other mental diseases, such as bipolar disorder, affecting frontal, temporal, cingulate, insula and thalamus. Postmortem studies support all these findings (Brown et al. 1986; Pakkenberg 1987; Bogerts et al.). All this information

leads to a conclusion that GM is a tissue highly involved in the disease, and that the pattern of brain damage may be more global than neuroimaging findings so far suggest.

5.1   Motivation & Objectives

The current work consists in applying GP classification to Schizophrenia using:

(i)   GM tissue class, which encodes key information about the disease, and

(ii)   Scalar momentum, which was found in study 2 to be a very efficient feature for predicting structural patterns of differences.

Subsequent performance results from the different feature sets were compared.

The objective is (1) to assess whether scalar momentum provides a good feature, (2) to test whether GM is really one of the best data features for predicting schizophrenia, and (3) to determine the optimal degree smoothing to apply to image data in order to predict schizophrenia.

5.2   Material & Methods

In this Study, anatomical MRI data from patients with schizophrenia and healthy subjects were pre-processed. A conventional VBM analysis was carried out to see GM volume differences associated with the sample. Next, the same sample was used for predicting disease using GP classification. Two different data features were used: GM and scalar momentum. For exploring the behavior regarding the degree of smoothing, several sizes of FHWM were used.

5.2.1   Participants

The sample of 111 patients with schizophrenia was recruited from Benito Menni CASM (40.94±8.75 years old; 33women), and the 111 healthy subjects came from a database of controls recruited from hospital staff and their acquaintances (40.14±10.11 years old; 33 women).

All participants were right-handed Spanish Caucasian. Pre-morbid IQ was estimated using the Word Accentuation Test (Del Ser et al. 1997; Gomar et al. 2011), which involves pronunciation of words whose accents are removed (score for healthy subjects: 23.07±4.54 and patients: 20.22±5.56).

Patients met DSM-IV criteria for schizophrenia, based on interviews by two psychiatrists and a review of case-notes, and were required to have a current IQ in the normal range

(i.e. 70+), as measured using four subtests of the Wechsler Adult Intelligence Scale III (WAIS-III) (vocabulary, similarities, block design, and matrix reasoning). All were scanned when in a relatively stable condition and all were taking antipsychotic medication.

Patients were included if they (a) were between18 and 65 years old, (b) had no history of brain trauma or neurological disease, and (c) had not shown alcohol/substance abuse within the 12 months prior to participation. Controls met the same criteria as the patients, and were questioned and excluded if they reported a history of mental illness and/or treatment with psychotropic medication.

Participants gave written informed consent, and the study was approved by the local hospital ethics committee. These data have been used previously in other studies (Pomarol-Clotet, Fatjó-Vilas, et al. 2010; Sans-Sansa et al. 2013; Sarró et al. 2013).

All subjects underwent structural MRI scanning using the same 1.5 T GE Signa scanner (General Electric Medical Systems, Milwaukee, WI, USA), located at the Sant Joan de DeuHospital in Barcelona (Spain). High-resolution structural T1-weighted MRI data were acquired with the following acquisition parameters: matrix size 512×512; 180 contiguous axial slices; voxel resolution 0.47×0.47×1 mm^3; echo (TE), repetition (TR) and inversion (TI) times were 3.93 ms, 2000 ms and 710 ms, respectively; flip angle 15°.

5.2.2 MRI preprocessing

After visual inspection for artifacts and anatomical anomalies, datapre-processing was carried out. The same procedure as in Study 2 was conducted using the SPM12 software (http://www.fil.ion.ucl.ac.uk/spm; The FIL methods group; London. UK).

Images were segmented into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) using the "Segmentation" tool. For registration, the Geodesic Shooting Toolbox was used (Ashburner and Friston 2011).

VBM

For the VBM analysis, warped GM tissue classes were scaled by the Jacobian. The modulated data were smoothed by 10 mm FWHM.

Statistical analysis was performed by fitting a voxel-wise general linear model (GLM), with corrections for multiple dependent comparisons using random field theory (Worsley et al. 1996; Worsley 2003) . A two-sample t-test was performed for comparing healthy subjects against schizophrenia patients. The design involved age and gender as covariates, and TIV as a "global". Spatial locations of differences between groups were identified using a FWE

corrected statistical threshold of p<0.01. An extent threshold of k = 500 voxels was applied.

GP classification

Warped GM tissue class and scalar momentum were used for predicting disease. The smoothing applied was in the range of the optimal values in Study 2. For both data features, smoothing from 10mm to 16mm FWHM were applied.

Predictions of the diagnostic group were carried out using the same GP classifier as in study 2 (available within the SPM12 software). The feature datasets were used as inputs, and transformed into linear kernel matrices using the dot product to become the corresponding covariance functions.

5.3   Results

5.3.1   VBM

VBM analysis detected most of the reported areas in the literature associated with schizophrenia (Honea et al. 2005). Only one bilaterally distributed cluster was found (Figure 5.1). It appeared symmetric and was centered at the left rectus [k= 148961, p_{FWE}<0.000, T= 9.84, (0, 39, -15)]. The cluster included two main peaks at the left anterior cingulate [p_{FWE}<0.000, T= 9.77, (2, 48, 14)] and the right cingulum middial [p_{FWE}<0.000, T= 9.11, (9, 39, 34)]. Locations are detailed in Table 5.1.

Distribution	Brain regions	MNI coordinates			Cluster size	T-score
		x	y	z		
cluster 1	Rectus L & R	0	39	-15	148961	9.84
	Cingulum Ant L & R					
	Cingulum mid L & R					
	Frontal sup medial L & R					
	Temporal mid L & R					
	Insula L & R					
	Frontal mid L & R					
	Temporal sup L & R					
	Frontal sup L& R					
	Rolandic oper L & R					
	Temporal inf L & R					
	Temporal pole sup L & R					
	Supramarginal L & R					
	Precentral L & R					
	Postcentral L & R					
	Olfactory L & R					
	Lingual L & R					
	Hippocampus L & R					
	ParaHippocampal L & R					
	Heschl L & R					
	Fusiform L & R					
	Thalamus L & R					
	Putamen R					
	Amygdala L & R					
	Calcarine R					
	Frontal sup orb L & R					
	Frontal mid orb L & R					
	Frontal med org L & R					
	Angular L					

Table 5.1 – Locations that were involved in cluster 1.
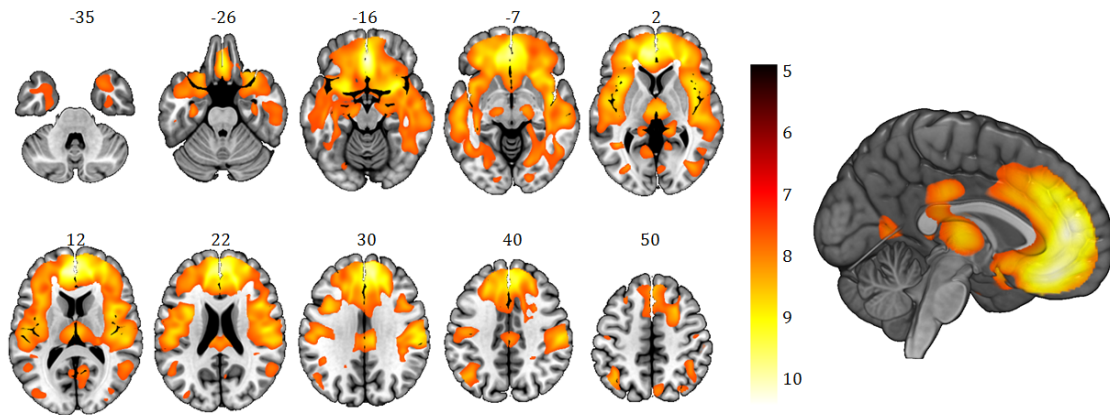
Fig. 5.1 – Statistical parametric map of the pattern of differences from a VBM comparison between healthy subjects and schizophrenia patients, corrected FWE p<0.01.

5.3.2 GP classification

Scalar momentum and GM provided very effective features. However, regarding the log-marginal likelihood, scalar momentum achieved the best performance, with a smoothing of 12mm FWHM. The corresponding AUC was 0.89. GM also was a good feature, and performance reached same AUC as scalar momentum. The best log-marginal likelihood score for GM was found at 11mm FWHM. However, the model fit was poorer than for the scalar momentum. Results from all the observations are detailed in Table 5.2.

smoothing (mm of FWHM)	GM		Scal mom	
	log-marginal likelihood	AUC	log-marginal likelihood	AUC
10	-111.85	0.88	-108.23	0.88
11	**-111.73**	**0.89**	-107.93	0.89
12	-111.86	0.89	**-107.84**	**0.89**
13	-112.23	0.88	-107.91	0.88
14	-112.54	0.88	-108.10	0.88
15	-112.94	0.88	-108.35	0.88
16	-113.39	0.87	-108.63	0.88

Table 5.2 – Marginal likelihood and AUC of the classification from GM and scalar momentum using a range of smoothing comprised between 10 and 16 mm of FWHM.

Estimation of the Bayes Factor concluded that there was strong evidence that scalar momentum provided a better model, scale strength of evidence was 7 according the Kass and Raftery (Kass and Raftery 1995).

Plots of the log-marginal likelihood and the AUC from all the performances conducted using GM and scalar momentum are shown in figure 5.2.
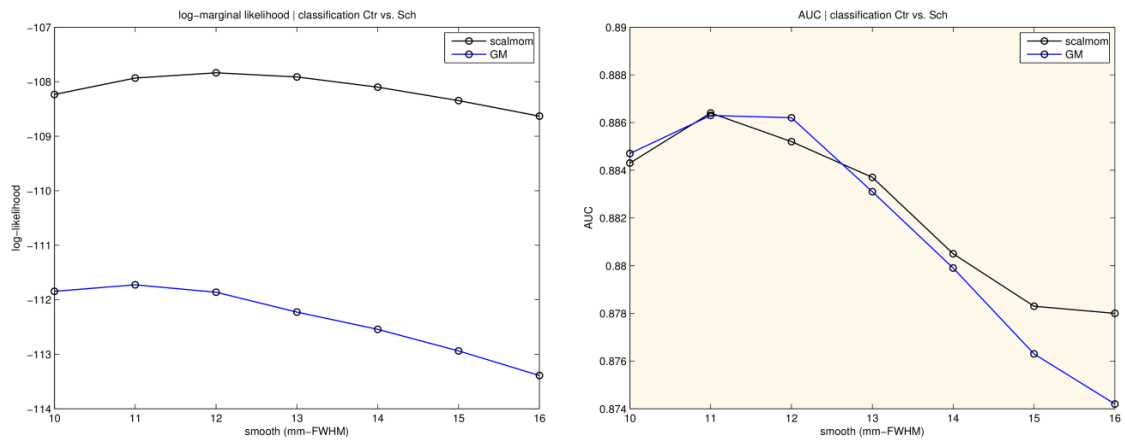
Fig 5.2 – Performance of the predictions about disease status. Controls versus Schizophrenia: (left) log-marginal likelihood and (right) Area under thecurve (AUC) obtained from each classification process. Estimations were performed using scalar momenta (black) and GM (blue) using different degrees of smoothing.

6. Discussion

The effect of the preprocessing steps is a sensitive issue in the classification and regression output. It is a wide field of exploration; few works have focussed on this aspect, in particular regarding fMRI data (LaConte et al. 2003). This thesis has contributed in this direction, by first exploring the reliability of GM segmentation and as well as studying different feature representations from MRI data. Also, a contribution about how TIV can be introduced successfully into the kernel has been presented. Results have shown that scalar momentum has become an excellent feature that allows characterising schizophrenia better than GM.

Before registration, it was necessary to find reliable tissue classes. For this purpose, in the study 1 three of the most frequently used segmentation pipelines, from the SPM and FSL software packages, were compared. The challenge in tissue segmentation lies in having a robust approach. Theoretical approaches have been presented in parallel to detect differences and similarities and outcomes from segmentations at each step have been empirically explored.

Some previous works compared segmentation algorithms. In Tsang et. al. (Tsang et al. 2008) they compare the same software as in the current thesis, although they used a slightly older SPM release (SPM5), and they used the same dataset. Although there were differences in methodology, other aspects were similar. The general conclusion was that SPM5 gave better specificity and sensitivitythan FSL. Fellhauer et al. (Fellhauer et al. 2015) compared SPM, FSL and FreeSurfer[8]. As expected, the best performances were obtained with good quality images. With especially noisy data, the best segmentation results were obtained using the algorithms in SPM8 and SPM12.

These results are in agreement with what was found in the current work. A relevant remark to bear in mind is that all these works, and many others, use different methodologies for testing segmentations. In this work, an additional contribution is that the methodology involves new approaches for exploring data, even though strategies, the dataset and commonly used tools (i.e. Jaccard index) are similar. On the other hand, the

---

[8] *freesurfer.net*

methodology also stops at each step, in native and normalized spaces, and explores how manual reorientation and skull stripping affects the segmentation process.

From preliminary evaluations, we found that manual positioning was necessary in SPM prior to running the "Unified Segmentation", which is in agreement with findings from Tsang et al. (2008). In the "New Segmentation" procedure it is also necessary. Although the estimated deformations are not extremely badly behaved, they may not be quite accurate enough. The issue lies in the initial affine registration; therefore, positioning in the anterior commissure is necessary to avoid the chance that the algorithm finds a local optimum. Concerning BET, special attention must be focused on the outputs before running the next step. A recommended option is *fslvbm_1_bet* with the *–N* command; this works successfully and overcomes a wide range of downsides; however, it does not guarantee that stripping will be accurate in 100% of the cases. If the stripping is not accurate, other more restrictive options are recommended.

In the main body of the study, where false-positive and -negative images were counted, FAST was the software with the highest amount of false negatives, and differences with respect to NS, and also to US, were statistically significant. Regarding false positives, US was the algorithm with highest number of them, statistically significant differences were found with FAST and also with NS.

On the other hand, FAST tended to be more specific with respect to the other algorithms, and US was found more sensitive than the rest. Interestingly, NS was the algorithm that seemed more balanced regarding those aspects, it is not statistically significant with respect to US' sensitivity, and regarding specificity, differences are not significant with respect to FAST. Moreover, NS obtained the highest Jaccard index, and the last in the ranking was FAST. The results indicate that NS is the most reliable of the three segmentation algorithms. In summary, findings point to NS being the most balanced of these three algorithms. Hence that for the study 2, the New Segmentation approach was applied.

There are some limitations in study 1. First, at the time of the evaluations, the versions used were the most recent. However, this is not a static field, and software versions are updated regularly. For study 2, a more recent version of the "New Segmentation" algorithm was used (from SPM12) compared to study 1 (SPM8). The NS method used in study 2 involved some improvements with respect to the older version. In fact, the modifications addressed some of the issues mentioned in this chapter.

Another limitation is related to the IBSR dataset: only one was used, and the images contained noise and artifacts such that the dataset became a challenge for the algorithms.

Additionally, the ground truth segmentations of GM available in the IBSR may bias the results, this is because sulcal CSF is segmented as GM. It involves that many voxels that were distributed all around the GM images as false negatives, are definitely wrong labeled. Because these certainly belong to the sulcal CSF and not to the GM. The inclusion of this voxels in the accuracy measurements definitely bias the results, this is reported and quantified by Valverde et al. (Valverde et al. 2014). NS and US were compared among other 10 brain tissue segmentation methods that also included FAST. Also the IBSR dataset was used for the comparison. They concluded that SPM offers the methods that appeared more suitable for GM tissue segmenation than the other. Remarkably, their findings were in concordance with ours, and stated that the ground truth segmentation may bias the performance by the algorithms (Valverde et al. 2014).

Regarding study 2, the GP machine learning approach has been used to explore different features from structural MRI data. Many features from structural data have been studied in this thesis. This study looked at the effectiveness of a range of features derived from structural data. T1-weighted images were preprocessed in a VBM style, "New Segmentation" and the "Geodesic shooting toolbox" (Ashburner and Friston 2011) of SPM12 were used. Different outcomes from these preprocessing procedures were used as features for predicting age, gender and body mass index (BMI). A wide range of images involved in the transformation, like Jacobians, the logarithm of Jacobians, the divergence of velocities and scalar momenta were explored. Also, more commonly used data, such as spatially normalized tissue classes, with and without Jacobian scaling ("modulation"), were evaluated. Pattern recognition evaluations were performed independently for each kernel at each degree of smoothing. For completeness, Gaussian smoothing from 0 to 20 mm FWHM was applied to the feature data to assess its effects on predictive accuracy.

The main conclusion was that scalar momentum provided an effective and robust feature, irrespective of the target of prediction, whether it was age, gender or BMI. In all cases, this feature obtained a performance very similar to the tissue that biologically encoded more information about the object of study. Best accuracies are generally assumed to be achieved when using the tissue that is more closely related to the biological process. Findings are in agreement with a previous work by Marquand et al. (Marquand et al. 2013) where they showed that the scalar momentum are a good feature for improving classification performance usign structural MRI data.

Allometry between brain and body allows empirical scaling factors to be recognized. An interesting example of this was white matter (WM) for predicting BMI, which was found to provide the best feature. Previous findings in the literature have indeed established that

there is a relationship between BMI and WM (Segura and Jurado Luque 2009; Seitz et al. 2015; Ou et al. 2015).

However, information about what tissue is likely to encode much of some biological process is generally not known a priori. In those cases, scalar momentum can be presented as the best option to start with. This is valuable for several reasons: First for obtaining a competitive predictive performance; second, scalar momentum can be used as a strategy to find the tissue that encodes more of the underlying biological information.

Performance from scalar momentum was not highly dependent on the degree of smoothing, achieving good scores from about 4 mm up to 20 mm FWHM, although the best behavior was from smoothness above 10 mm. However, the optimal smoothing may depend on the number of training images. For this thesis, a sample of 562 subjects was used, but less smoothing would probably be needed with more data. Interestingly, without any smoothing, the logarithm of Jacobians seemed to be the most useful of all features.

One surprising result was that modulated GM was not a very effective feature for making age predictions. Brain ageing has been found to follow a specific pattern in which GM volume plays a relevant role with respect the other tissues. It was found that GM increases from birth until the age of four and then decreases until the 70s (Pfefferbaum et al. 1994). More recent studies have found that GM decreases linearly with age, while WM did not (Good et al. 2001). These, among other studies, justify that GM were used as feature in a number of studies, but the current work suggests that the results of those studies could have been improved. Information from Jacobian-scaled GM only became effective when combined with Jacobian-scaled information from the rest of the image.

Another aspect in this study that is worth mentioning is that TIV was introduced in the prediction model. Kernels from modulated data were divided by the product of TIV associated to each pair of subjects. Such an operation should be equivalent to dividing each subject image for TIV, however, in this way calculations were extremely simplified. Results suggested that accounting for TIV increased the performance from modulated data. This fact makes sense, like in VBM analyses (Mechelli et al. 2005), because the data account for real differences in volume and not for contractions and expansions due to registration transformations to the standard space.

A limitation of this study could be the size of the sample; 562 subjects is not a small number of inputs, but for pattern recognition studies, a larger number of samples is desirable. The behavior of the smoothing would vary with a larger number of subjects, so possibly a lower amount of blurring is required.

In the study 3, a VBM analysis was performed and the results confirmed that grey matter (GM) is strongly involved in schizophrenia. The same pattern of damage as reported in the literature was found. These findings confirmed that GM contains a large amount of information about schizophrenia. After that, scalar momentum was used as feature for predicting schizophrenia, and the performance was compared to that obtained from using GM. A range of smoothing of between 10 mm and 16 mm full-width-at-half-maximum (FWHM) was considered.

The classification of schizophrenia using MRI has been studied extensively, and several papers using a variety of pattern recognition approaches have appeared. In most of them Support Vector Machines (SVM) were used; in some cases, structural GM data were used to study first-episode schizophrenia (Pina-Camacho et al. 2015; Mourao-Miranda et al. 2012), and schizophrenia in general (Nieuwenhuis et al. 2012), with respect to data of relatives. Sometimes this also involved using white matter (WM) as a feature (Fan et al. 2008). Regarding the machine learning approach, Support Vector Machines (SVM) are generally used more widely than Gaussian Processes (GP), although GP have been applied with GM structural data to other diseases, such as Alzheimer's disease (Young et al. 2012).

Only in few cases where MRI data were used, the involved features were not GM or WM. For instance, Jacobians were used with SVM for characterizing a schizophrenia pattern (Yushkevich et al. 2005). However, using scalar momentum instead of GM or any other tissue class is not wide-spread practice.

This study contributes not only by showing that GP achieve a similar performance to SVM for neuroimaging data (Schrouff et al. 2013; Challis et al. 2015), but also shows that GP models make probabilistic predictions. Moreover, scalar momentum provides an excellent feature for pattern recognition studies, which is highly relevant. This work shows that this feature is competitive with the single tissue that encodes most about the biological process being studied, sometimes achieving better performance than expected from the best single tissue type.

GM has been widely proposed to be the best candidate to explain differences between schizophrenia patients and healthy subjects. In this study, this tissue was found to be a good feature for predicting schizophrenia, giving as expected a competitive performance. In fact, the AUC of GM and scalar momentum was 0.89. However, the Bayes Factor (Kass and Raftery 1995) provided strong evidence that the scalar momentum could be a better feature than GM. These results were in agreement with what was hypothesized from the conclusions in study 2.

On the other hand, the smoothing degree that initially was supposed to be the best for GM (~15mm), was finally centered around 11mm FWHM, which was lower than expected. For scalar momentum, a kernel of 12 mm FWHM was applied, which was within in the expected range.

The aim of this work is not clinical; rather, the aim is to explore whether scalar momentum provides a feature that performs well, irrespective of the target when using real clinical data. The outcome of this exploration suggests that Gaussian processes and scalar momentum can contribute to clinical research on mental and neurodegenerative diseases. This line of work could aid routine clinical practice in delimiting differential diagnoses, and in reinforcing clinical treatments, which sometimes are not clear.

A limitation of this work is the size of the sample, which is smaller than in study 2; the current classification involved 222 subjects. Further analyses should be performed with larger samples of healthy subjects and patients. Several interesting designs could be explored to study the scope of using scalar momentum in depth, for instance by attempting to discriminate between two different diseases that share some common aspectsinstance. Examples of these could be schizophrenia, bipolar, and schizoaffective disorders. This approach should be tested in neurodegenerative diseases, like Alzheimer's disease (AD), and also Pre-clinical and Prodromic AD.

In this study, scalar momentum provided a very efficient feature for predicting schizophrenia, and improved the performance obtained by using GM. Initially, the tissue class was considered the most relevant feature and the fact that scalar momentum outperformed the results from GM confirms that scalar momentum encodes relevant information about the biological course of the disease.

In summary, the aim of study 2 was to explore and optimize a multivariate model of anatomical differences, and to apply the observations to schizophrenia in the next study. In study 3, scalar momentum provided very efficient feature; the feature achieved better than GM, which was initially the best candidate to explain differences between schizophrenia patients and healthy subjects. The most relevant aspect from these studies relies on this particular feature that outperformed the rest independently of the target.

The main result of this thesis was that scalar momentum provided very efficient and robust feature in all cases. This feature achieved a performance that was very similar, and even better than the tissue that biologically encoded more information about the target. Interestingly, BMI was predicted succesfully using WM; scalar momentum provided a very competitive performance with this tissue class. However, WM outperformed all the features used. On the contrary and against expectations, GM was not the feature that

better predicted age. Scalar momentum provided the best feature to predict age, and GM was not among the best features.

Scalar momenta were obtained using the Geodesic Shooting Toolbox (Ashburner and Friston 2011), implemented in SPM12. This is a tool for modeling shapes of the brain, based on the diffeomorphic framework (M. Miller et al. 1997; Grenander and Miller 1998; M. I. Miller 2004). The preprocessing provides a potentially more accurate shape model. In particular, scalar momentum seemed to encode relevant information about brain anatomy and can be interpreted as the ensemble of the warped tissue types involved in the preprocessing, i.e. GM, WM and BG, even though scalar momentum has no straightforward biological interpretation (Ashburner and Klöppel 2011).

A variety of spatial normalization approaches, along with their settings, could have been explored in order to assess which gave more effective feature representations. Previous work by Klein et al. (2009) compared a number of widely used nonlinear registration algorithms, and found that Dartel (Ashburner 2007) was one of the more accurate registration tools. This paper did not assess the Geodesic Shooting toolbox (Ashburner and Friston 2011) of SPM12, which was released later. More recent evaluations (Ashburner and Friston 2011), using some of the same data as those of the Klein et al. paper (Klein et al. 2009), have shown that Geodesic Shooting slightly outperforms Dartel.

6.1 Outlook

Further works might be conducted based on the current work. Regarding segmentation, new versions of the algorithms should be explored; it is desirable that all aspects that arose from the comparisons would be addressed in future work. Other work arises from the way the TIV was added to the kernel; this application could be used with other variables, for instance to include them as nuisance variables in the model. Also, considering the surprising effectiveness of the scalar momentum, studying registration, the model and its outcomes in depth, could become an interesting line of work. Finally, considering the scope of the findings, the presented methodology may be evaluated in other psychiatric and also neurodegenerative diseases. Pattern recognition techniques may strongly contribute to neuroscience, and more importantly could be used in clinical practice for an improvement of those aspects that still remain controversial.

7. Conclusions

In order to answer the issues that initially were posed, three different studies have been designed. Throughout these works, the most relevant steps that may interfere with the process to accomplish the main objective were identified and explored; eventually, we obtained the following conclusions:

1. A comparison between segmentation pipelines, -FAST from FSL, US and NS from SPM-, was performed. NS was found to be the most balanced algorithm regarding the number of false-positives and -negatives and the reliability of the GM class w.r.t. the gold standard. Manual reorientation was confirmed to be necessary in SPM, although it became less relevant in NS than in US. Regarding FAST, the optimal command to perform skull stripping was *fslvbm_1_bet –N*, although this was not accurate when facing some particularly problematic images.

2. The GP machine learning approach was used to test different features from structural MRI data. VBM-type preprocessing was conducted and outputs were used as features. Scalar momentum was found to be one of the most effective features for predicting age, gender and BMI. Against expectations, Jacobian scaled GM was not found to be a particularly good feature for predicting age. Interestingly, WM proved to be very effective for predicting BMI.

3. The performance of the predictions was highly dependent on the smoothing. Jacobians and divergence of velocities provided better features with absent or lower smoothing. Tissue classes, modulated and non-modulated, gave better performances with a higher amount of smoothing. Positively, scalar momentum did not play a strongly subordinate role to the smoothing.

4. The novel strategy - introducing TIV into the kernels for GP prediction from modulated tissues - resulted in a better performance than using the same kernels without accounting for TIV.

5. Scalar momentum with 12mm FWHM of smoothing obtained the best log-marginal likelihood fit when classifying controls and schizophrenia patients. The Bayes Factor provided strong evidence that scalar momentum is a better feature than GM, even though both features achieved the same AUC (0.89).

6.   The final conclusion from this thesis is that multivariate pattern recognition analyses using scalar momentum as feature provide an excellent strategy for classifying schizophrenia. This approach might potentially be extended to other psychiatric and neurodegenerative diseases, both in research and as an aid to differential diagnosis in routine clinical practice.

8. References

Agostino, Emiliano D, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. 2004. "Non-Rigid Atlas-to-Image Registration by Minimization of Class-Conditional Image Entropy." *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, no. March: 745–53. doi:10.1007/978-3-540-30135-6_91.

Ashburner, and K J Friston. 2000. "Voxel-Based Morphometry--the Methods." *NeuroImage* 11 (6 Pt 1): 805–21. doi:10.1006/nimg.2000.0582.

Ashburner, and Karl J. Friston. 2011. "Diffeomorphic Registration Using Geodesic Shooting and Gauss-Newton Optimisation." *NeuroImage* 55 (3): 954–67. doi:10.1016/j.neuroimage.2010.12.049.

Ashburner, J. 2007. "A Fast Diffeomorphic Image Registration Algorithm." *NeuroImage* 38 (1): 95–113. doi:10.1016/j.neuroimage.2007.07.007.

Ashburner, J, and K J. Friston. 2001. "Why Voxel-Based Morphometry Should Be Used." *NeuroImage* 14 (6): 1238–43. doi:10.1006/nimg.2001.0961.

Ashburner, J, and Karl J Friston. 2005. "Unified Segmentation." *NeuroImage* 26 (3): 839–51. doi:10.1016/j.neuroimage.2005.02.018.

Ashburner, J, and Karl J. Friston. 1999. "Nonlinear Spatial Normalization Using Basis Functions." *Human Brain Mapping* 7 (4): 254–66. doi:10.1002/(SICI)1097-0193(1999)7:4<254::AID-HBM4>3.0.CO;2-G.

Ashburner, J, and Stefan Klöppel. 2011. "Multivariate Models of Inter-Subject Anatomical Variability." *NeuroImage* 56 (2): 422–39. doi:10.1016/j.neuroimage.2010.03.059.

Ashburner, J. 2009. "Computational Anatomy with the SPM Software." *Magnetic Resonance Imaging* 27 (8): 1163–74. doi:10.1016/j.mri.2009.01.006.

Ashburner, J., and M.I. Miller. 2015. *Brain Mapping. Brain Mapping.* doi:10.1016/B978-0-12-397025-1.00301-8.

Beg, M. Faisal, Michael I. Miller, Alain Trouvé, and Laurent Younes. 2005. "Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms." *International Journal of Computer Vision* 61 (2): 139–57. doi:10.1023/B:VISI.0000043755.93987.aa.

Besag, Julian. 1986. "On the Statistical Analysis of Dirty Pictures." *Journal of the Royal Statistical Society. Series B (...* 48 (3): 259–302. doi:10.1080/02664769300000059.

Bishop, Christopher M. 2006. *Bishop - Pattern Recognition and Machine Learning.* doi:10.1117/1.2819119.

Bishop, Christopher M., Markus Svensen, and Christopher K. I. Williams. 1998. "GTM: The Generative Topographic Mapping." *NEURAL COMPUTATION* 10 (1): 215–34. doi:10.1162/089976698300017953.

Bogerts, B, P Falkai, M Haupts, B Greve, S Ernst, U Tapernon-Franz, and U Heinzmann. "Post-

Mortem Volume Measurements of Limbic System and Basal Ganglia Structures in Chronic Schizophrenics. Initial Results from a New Brain Collection." *Schizophrenia Research* 3 (5-6): 295–301. http://www.ncbi.nlm.nih.gov/pubmed/2282334.

Bora, Emre, Alex Fornito, Joaquim Radua, Mark Walterfang, Marc Seal, Stephen J. Wood, Murat Yücel, Dennis Velakoulis, and Christos Pantelis. 2011. "Neuroanatomical Abnormalities in Schizophrenia: A Multimodal Voxelwise Meta-Analysis and Meta-Regression Analysis." *Schizophrenia Research*. doi:10.1016/j.schres.2010.12.020.

Bouix, Sylvain, Marcos Martin-Fernandez, Lida Ungar, Motoaki Nakamura, Min-Seong Koo, Robert W McCarley, and Martha E Shenton. 2007. "On Evaluating Brain Tissue Classifiers without a Ground Truth." *NeuroImage* 36 (4): 1207–24. doi:10.1016/j.neuroimage.2007.04.031.

Brown, R, N Colter, J A Corsellis, T J Crow, C D Frith, R Jagoe, E C Johnstone, and L Marsh. 1986. "Postmortem Evidence of Structural Brain Changes in Schizophrenia. Differences in Brain Weight, Temporal Horn Area, and Parahippocampal Gyrus Compared with Affective Disorder." *Archives of General Psychiatry* 43 (1): 36–42. doi:10.1001/archpsyc.1986.01800010038005.

Challis, Edward, Peter Hurley, Laura Serra, Marco Bozzali, Seb Oliver, and Mara Cercignani. 2015. "Gaussian Process Classification of Alzheimer's Disease and Mild Cognitive Impairment from Resting-State fMRI." *NeuroImage* 112 (May): 232–43. doi:10.1016/j.neuroimage.2015.02.037.

Christensen, G.E. 1994. "Deformable Shape Models for Anatomy." *Electrical Engineering D.Sc.* Dissertati.

Christensen, GE, RD Rabbit, MI Miller, SC Joshi, U Grenander, T Coogan, and DC Van Essen. 1995. "Topological Properties of Smooth Anatomic Maps." *Information Processing in Medical Imaging* 3: 101–12.

Costafreda, Sergi G, Carlton Chu, John Ashburner, and Cynthia H Y Fu. 2009. "Prognostic and Diagnostic Potential of the Structural Neuroanatomy of Depression." *PloS One* 4 (7): e6353. doi:10.1371/journal.pone.0006353.

de Paula, André Luiz Damião, Jaime Eduardo Cecílio Hallak, João Paulo Maia-de-Oliveira, Rodrigo A Bressan, and João Paulo Machado-de-Sousa. 2015. "Cognition in at-Risk Mental States for Psychosis." *Neuroscience and Biobehavioral Reviews*, September. doi:10.1016/j.neubiorev.2015.09.006.

Del Ser, T, J I González-Montalvo, S Martínez-Espinosa, C Delgado-Villapalos, and F Bermejo. 1997. "Estimation of Premorbid Intelligence in Spanish People with the Word Accentuation Test and Its Application to the Diagnosis of Dementia." *Brain and Cognition* 33 (3): 343–56. doi:10.1006/brcg.1997.0877.

Dempster, A, N Laird, and D Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1): 1–38. http://links.jstor.org/sici?sici=0035-9246(1977)39:1<1:MLFIDV>2.0.CO;2-Z\npapers2://publication/uuid/B5CEE87B-34E6-4056-B037-31444FAD60C1.

Ellison-Wright, Ian, and Ed Bullmore. 2010. "Anatomy of Bipolar Disorder and Schizophrenia: A Meta-Analysis." *Schizophrenia Research* 117 (1): 1–12. doi:10.1016/j.schres.2009.12.022.

Fan, Yong, Raquel E Gur, Ruben C Gur, Xiaoying Wu, Dinggang Shen, Monica E Calkins, and Christos Davatzikos. 2008. "Unaffected Family Members and Schizophrenia Patients Share Brain Structure Patterns: A High-Dimensional Pattern Classification Study." *Biological Psychiatry* 63 (1): 118–24. doi:10.1016/j.biopsych.2007.03.015.

Feinstein, a, P Roy, N Lobaugh, K Feinstein, P O'Connor, and S Black. 2004. "Structural Brain Abnormalities in Multiple Sclerosis Patients with Major Depression." *Neurology* 62: 586–90. doi:10.1212/01.WNL.0000110316.12086.0C.

Fellhauer, Iven, Frank G Zöllner, Johannes Schröder, Christina Degen, Li Kong, Marco Essig, Philipp A Thomann, and Lothar R Schad. 2015. "Comparison of Automated Brain Segmentation Using a Brain Phantom and Patients with Early Alzheimer's Dementia or Mild Cognitive Impairment." *Psychiatry Research* 233 (3): 299–305. doi:10.1016/j.pscychresns.2015.07.011.

Frisoni, Giovanni B, Nick C Fox, Clifford R Jack, Philip Scheltens, and Paul M Thompson. 2010. "The Clinical Use of Structural MRI in Alzheimer Disease." *Nature Reviews. Neurology* 6 (2): 67–77. doi:10.1038/nrneurol.2009.215.

Friston, Karl J. 1994. "Functional and Effective Connectivity in Neuroimaging: A Synthesis." *Human Brain Mapping* 2 (1-2): 56–78. doi:10.1002/hbm.460020107.

Gomar, Jesús J., Jordi Ortiz-Gil, Peter J. McKenna, Raymond Salvador, Bibiana Sans-Sansa, Salvador Sarró, Amalia Guerrero, and Edith Pomarol-Clotet. 2011. "Validation of the Word Accentuation Test (TAP) as a Means of Estimating Premorbid IQ in Spanish Speakers." *Schizophrenia Research* 128 (1-3): 175–76. doi:10.1016/j.schres.2010.11.016.

Good, C D, I S Johnsrude, J Ashburner, R N Henson, K J Friston, and R S Frackowiak. 2001. "A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains." *NeuroImage* 14 (1 Pt 1): 21–36. doi:10.1006/nimg.2001.0786.

Grenander, Ulf, and Michael I Miller. 1998. "Computational Anatomy: An Emerging Discipline." *Quart. App. Math.* 56 (June 1997): 617–94.

Helms, Gunther, Kai Kallenberg, and Peter Dechent. 2006. "Contrast-Driven Approach to Intracranial Segmentation Using a Combination of T2- and T1-Weighted 3D MRI Data Sets." *Journal of Magnetic Resonance Imaging* 24 (4): 790–95. doi:10.1002/jmri.20692.

Honea, Robyn, Tim J. Crow, Dick Passingham, and Clare E. Mackay. 2005. "Regional Deficits in Brain Volume in Schizophrenia: A Meta-Analysis of Voxel-Based Morphometry Studies." *American Journal of Psychiatry.* doi:10.1176/appi.ajp.162.12.2233.

Ivleva, E., A.S. Bidesi, M.S. Keshavan, G.D. Pearlson, S.A. Meda, D. Dodig, A.F. Moates, et al. 2013. "Gray Matter Volume as an Intermediate Phenotype for Psychosis: Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP)." *Am J Psychiatry* 170: 1285–96.

Jeffreys, Harold. 1961. *Theory of Probability. Theory of Probability*. Vol. 2. http://ocw.mit.edu/OcwWeb/Mathematics/18-175Spring-2007/LectureNotes/Index.htm.

Kass, Robert E, and Adrian E Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90 (430): 773–95. doi:10.2307/2291091.

Klein, Arno, B.A. Ardekani, Jesper Andersson, John Ashburner, Brian Avants, M.C. Chiang, G.E. Christensen, et al. 2009. "Evaluation of 14 Nonlinear Deformation Algorithms Applied to Human Brain MRI Registration." *Neuroimage* 46 (3): 786–802. doi:10.1016/j.neuroimage.2008.12.037.

Klöppel, Stefan, Ahmed Abdulkadir, Clifford R. Jack, Nikolaos Koutsouleris, Janaina Mourão-Miranda, and Prashanthi Vemuri. 2012. "Diagnostic Neuroimaging across Diseases." *NeuroImage* 61 (2): 457–63. doi:10.1016/j.neuroimage.2011.11.002.

Klöppel, Stefan, Cynthia M Stonnington, Carlton Chu, Bogdan Draganski, Rachael I Scahill, Jonathan D Rohrer, Nick C Fox, Clifford R Jack, John Ashburner, and Richard S J Frackowiak. 2008.

"Automatic Classification of MR Scans in Alzheimer's Disease." *Brain : A Journal of Neurology* 131 (Pt 3): 681–89. doi:10.1093/brain/awm319.

LaConte, Stephen, Jon Anderson, Suraj Muley, James Ashe, Sally Frutiger, Kelly Rehm, Lars Kai Hansen, et al. 2003. "The Evaluation of Preprocessing Choices in Single-Subject BOLD fMRI Using NPAIRS Performance Metrics." *NeuroImage* 18 (1): 10–27. doi:10.1006/nimg.2002.1300.

Lalaoui L., and Mohamadi T. 2013. "A Comparative Study of Image Region-Based Segmentation Algorithms." *International Journal of Advanced Computer Science and Applications(IJACSA)* 4 (6). doi:10.14569/IJACSA.2013.040627.

Malone, Ian B, Kelvin K Leung, Shona Clegg, Josephine Barnes, Jennifer L Whitwell, John Ashburner, Nick C Fox, and Gerard R Ridgway. 2015. "Accurate Automatic Estimation of Total Intracranial Volume: A Nuisance Variable with Less Nuisance." *NeuroImage* 104: 366–72. doi:10.1016/j.neuroimage.2014.09.034.

Marquand, Andre F, Maurizio Filippone, John Ashburner, Mark Girolami, Janaina Mourao-Miranda, Gareth J Barker, Steven C R Williams, P Nigel Leigh, and Camilla R V Blain. 2013. "Automated, High Accuracy Classification of Parkinsonian Disorders: A Pattern Recognition Approach." *PloS One* 8 (7): e69237. doi:10.1371/journal.pone.0069237.

Mechelli, Andrea, Cathy J Price, Karl J Friston, and John Ashburner. 2005. "Voxel-Based Morphometry Applications of the Human Brain : Methods and." *Imaging* 1: 1–9. doi:10.2174/1573405054038726.

Miller, M, a Banerjee, G Christensen, S Joshi, N Khaneja, U Grenander, and L Matejic. 1997. "Statistical Methods in Computational Anatomy." *Statistical Methods in Medical Research* 6 (3): 267–99. doi:10.1191/096228097673360480.

Miller, Michael I. 2004. "Computational Anatomy: Shape, Growth, and Atrophy Comparison via Diffeomorphisms." In *NeuroImage*, 23:S19–33. doi:10.1016/j.neuroimage.2004.07.021.

Modinos, Gemma, William Pettersson-Yeo, Paul Allen, Philip K. McGuire, André Aleman, and Andrea Mechelli. 2012. "Multivariate Pattern Classification Reveals Differential Brain Activation during Emotional Processing in Individuals with Psychosis Proneness." *NeuroImage* 59 (3): 3033–41. doi:10.1016/j.neuroimage.2011.10.048.

Mourao-Miranda, J, A A T S Reinders, V Rocha-Rego, J Lappin, J Rondina, C Morgan, K D Morgan, et al. 2012. "Individualized Prediction of Illness Course at the First Psychotic Episode: A Support Vector Machine MRI Study." *Psychological Medicine* 42 (5): 1037–47. doi:10.1017/S0033291711002005.

Neal, Radford M., and Geoffrey E. Hinton. 1998. "A View of the EM Algorithm That Justifies Incremental, Sparse, and Other Variants." In *Learning in Graphical Models*, 355–68. doi:10.1007/978-94-011-5014-9_12.

Nieuwenhuis, Mireille, Neeltje E M van Haren, Hilleke E. Hulshoff Pol, Wiepke Cahn, René S. Kahn, and Hugo G. Schnack. 2012. "Classification of Schizophrenia Patients and Healthy Controls from Structural MRI Scans in Two Large Independent Samples." *NeuroImage* 61 (3): 606–12. doi:10.1016/j.neuroimage.2012.03.079.

Ou, Xiawei, Aline Andres, R T Pivik, Mario A Cleves, and Thomas M Badger. 2015. "Brain Gray and White Matter Differences in Healthy Normal Weight and Obese Children." *Journal of Magnetic Resonance Imaging : JMRI*, April. doi:10.1002/jmri.24912.

Pakkenberg, B. 1987. "Post-Mortem Study of Chronic Schizophrenic Brains." *The British Journal of*

Psychiatry: The Journal of Mental Science 151 (December): 744–52.
http://www.ncbi.nlm.nih.gov/pubmed/3502800.

Patiny, Florentine, Eric Constant, and Sophie Symann. 2015. "A Comparative Study between
Cognitive Impairments of Adults with Schizophrenia and Children with Psychotic Spectrum
Disorders: A Literature Review." *Psychiatria Danubina* 27 (Suppl-1): 250–54.
http://www.ncbi.nlm.nih.gov/pubmed/26417774.

Pfefferbaum, A, D H Mathalon, E V Sullivan, J M Rawles, R B Zipursky, and K O Lim. 1994. "A
Quantitative Magnetic Resonance Imaging Study of Changes in Brain Morphology from
Infancy to Late Adulthood." *Archives of Neurology* 51 (9): 874–87.
http://www.ncbi.nlm.nih.gov/pubmed/8080387.

Pina-Camacho, Laura, Juan Garcia-Prieto, Mara Parellada, Josefina Castro-Fornieles, Ana M
Gonzalez-Pinto, Igor Bombin, Montserrat Graell, et al. 2015. "Predictors of Schizophrenia
Spectrum Disorders in Early-Onset First Episodes of Psychosis: A Support Vector Machine
Model." *European Child & Adolescent Psychiatry* 24 (4): 427–40. doi:10.1007/s00787-014-
0593-0.

Pomarol-Clotet, E, E J Canales-Rodríguez, R Salvador, S Sarró, J J Gomar, F Vila, J Ortiz-Gil, Y Iturria-
Medina, A Capdevila, and P J McKenna. 2010. "Medial Prefrontal Cortex Pathology in
Schizophrenia as Revealed by Convergent Findings from Multimodal Imaging." *Molecular
Psychiatry* 15 (8): 823–30. doi:10.1038/mp.2009.146.

Pomarol-Clotet, E., M. Fatjó-Vilas, P. J. McKenna, G. C. Monté, S. Sarró, J. Ortiz-Gil, C. Aguirre, et al.
2010. "COMT Val158Met Polymorphism in Relation to Activation and de-Activation in the
Prefrontal Cortex: A Study in Patients with Schizophrenia and Healthy Subjects." *NeuroImage*
53 (3): 899–907. doi:10.1016/j.neuroimage.2010.04.018.

Rasmussen, C.E., and C.K.I. Williams. 2006. *Gaussian Processes for Machine Learning. International
Journal of Neural Systems*. Vol. 14. http://www.ncbi.nlm.nih.gov/pubmed/15112367.

Sabuncu, Mert R., and Ender Konukoglu. 2014. "Clinical Prediction from Structural Brain MRI Scans:
A Large-Scale Empirical Study." *Neuroinformatics*, 31–46. doi:10.1007/s12021-014-9238-1.

Salgado-Pineda, P, E Fakra, P Delaveau, P J McKenna, E Pomarol-Clotet, and O Blin. 2011.
"Correlated Structural and Functional Brain Abnormalities in the Default Mode Network in
Schizophrenia Patients." *Schizophrenia Research* 125 (2-3): 101–9.
doi:10.1016/j.schres.2010.10.027.

Sans-Sansa, B., P. J. McKenna, E. J. Canales-Rodríguez, J. Ortiz-Gil, L. López-Araquistain, S. Sarró, R.
M. Dueñas, J. Blanch, R. Salvador, and E. Pomarol-Clotet. 2013. "Association of Formal Thought
Disorder in Schizophrenia with Structural Brain Abnormalities in Language-Related Cortical
Regions." *Schizophrenia Research* 146 (1-3): 308–13. doi:10.1016/j.schres.2013.02.032.

Sarró, Salvador, Edith Pomarol-Clotet, Erick J Canales-Rodríguez, Raymond Salvador, Jesús J Gomar,
Jordi Ortiz-Gil, Ramón Landín-Romero, Fidel Vila-Rodríguez, Josep Blanch, and Peter J
McKenna. 2013. "Structural Brain Changes Associated with Tardive Dyskinesia in
Schizophrenia." *The British Journal of Psychiatry: The Journal of Mental Science* 203: 51–57.
doi:10.1192/bjp.bp.112.114538.

Schrouff, J., M. J. Rosa, J. M. Rondina, A. F. Marquand, C. Chu, J. Ashburner, C. Phillips, J. Richiardi, and
J. Mourão-Miranda. 2013. "PRoNTo: Pattern Recognition for Neuroimaging Toolbox."
Neuroinformatics 11 (3): 319–37. doi:10.1007/s12021-013-9178-1.

Segura, B, and M A Jurado Luque. 2009. "Metabolic Syndrome and Ageing: Cognitive Impairment

and Structural Alterations of the Central Nervous System. [Spanish]\nSindrome Metabolico Y Envejecimiento: Deficit Cognitivo Y Alteraciones Estructurales Del Sistema Nervioso Central." *Revista de Neurologia* 49 (8): 417–24. http://www.neurologia.com/pdf/Web/4908/bc080417.pdf\nhttp://ovidsp.ovid.com/ovidw eb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed9&AN=2009562804\nhttp://openurl. ac.uk/athens:_edu//lfp/LinkFinderPlus/Display?sid=OVID:Embase&id=pmid:&id=&issn=021 0-0010&isbn=&v.

Seitz, Jochen, Martin Walter, Verena Mainz, Beate Herpertz-Dahlmann, Kerstin Konrad, and Georg von Polier. 2015. "Brain Volume Reduction Predicts Weight Development in Adolescent Patients with Anorexia Nervosa." *Journal of Psychiatric Research* 68 (September): 228–37. doi:10.1016/j.jpsychires.2015.06.019.

Shattuck, David W, Anand A Joshi, Dimitrios Pantazis, Eric Kan, Rebecca A Dutton, Elizabeth R Sowell, Paul M Thompson, Arthur W Toga, and Richard M Leahy. 2009. "Semi-Automated Method for Delineation of Landmarks on Models of the Cerebral Cortex." *Journal of Neuroscience Methods* 178 (2): 385–92. doi:10.1016/j.jneumeth.2008.12.025.

Shattuck, David W., Mubeena Mirza, Vitria Adisetiyo, Cornelius Hojatkashani, Georges Salamon, Katherine L. Narr, Russell A. Poldrack, Robert M. Bilder, and Arthur W. Toga. 2008. "Construction of a 3D Probabilistic Atlas of Human Cortical Structures." *NeuroImage* 39 (3): 1064–80. doi:10.1016/j.neuroimage.2007.09.031.

Smith, Stephen M. 2001. "BET: Brain Extraction Tool FMRIB Technical Report TR00SMS2b." Vol. i. doi:10.1002/pssb.201300062.

Thaker, Gunvant K. 2008. "Neurophysiological Endophenotypes across Bipolar and Schizophrenia Psychosis." *Schizophrenia Bulletin* 34 (4): 760–73. doi:10.1093/schbul/sbn049.

Tsang, On, Ali Gholipour, Nasser Kehtarnavaz, Kaundinya Gopinath, Richard Briggs, and Issa Panahi. 2008. "Comparison of Tissue Segmentation Algorithms in Neuroimage Analysis Software Tools." *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2008: 3924–28. doi:10.1109/IEMBS.2008.4650068.

Valverde, Sergi, Arnau Oliver, Mariano Cabezas, Eloy Roura, and Xavier Lladó. 2014. "Comparison of 10 Brain Tissue Segmentation Methods Using Revisited IBSR Annotations." *Journal of Magnetic Resonance Imaging* 101: 93–101. doi:10.1002/jmri.24517.

Wang, Lei, Faisal Beg, Tilak Ratnanather, Can Ceritpglu, Laurent Younes, John C. Morris, John G. Csernansky, and Michael I. Miller. 2007. "Large Deformation Diffeomorphism and Momentum Based Hippocampal Shape Discrimination in Dementia of the Alzheimer Type." *IEEE Transactions on Medical Imaging* 26 (4): 462–70. doi:10.1109/TMI.2006.887380.

Wells, W M, W L Grimson, R Kikinis, and F a Jolesz. 1996. "Adaptive Segmentation of MRI Data." *IEEE Transactions on Medical Imaging* 15 (4): 429–42. doi:10.1109/42.511747.

Westman, Eric, Andrew Simmons, Yi Zhang, J-Sebastian Muehlboeck, Catherine Tunnard, Yawu Liu, Louis Collins, et al. 2011. "Multivariate Analysis of MRI Data for Alzheimer's Disease, Mild Cognitive Impairment and Healthy Controls." *NeuroImage* 54 (2): 1178–87. doi:10.1016/j.neuroimage.2010.08.044.

Williams, C K I, and C E Rasmussen. 1996. "Gaussian Processes for Regression." In *Advances in Neural Information Processing Systems 8*, 8:514–20. http://eprints.aston.ac.uk/651/.

Worsley, K J. 2003. "Detecting Activation in fMRI Data." *Statistical Methods in Medical Research* 12

(5): 401–18. doi:10.1191/0962280203sm340ra.

Worsley, K. J., S. Marrett, P. Neelin, a. C. Vandal, Karl J. Friston, and Alan C. Evans. 1996. "A Unified Statistical Approach for Determining Significant Signals in Images of Cerebral Activation." *Human Brain Mapping* 4 (1): 58–73. doi:10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O.

Wright, I C, P K McGuire, J B Poline, J M Travere, R M Murray, C D Frith, R S Frackowiak, and K J Friston. 1995. "A Voxel-Based Method for the Statistical Analysis of Gray and White Matter Density Applied to Schizophrenia." *NeuroImage* 2 (4): 244–52. doi:10.1006/nimg.1995.1032.

Younes, L. 2007. "Jacobi Fields in Groups of Diffeomorphisms and Applications." *Quarterly of Applied Mathematics.* 650 (1): 11–134.

Young, J., M. Modat, M. J. Cardoso, J. Ashburner, and S. Ourselin. 2012. "Classification of Alzheimer's Disease Patients and Controls with Gaussian Processes." *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, no. Imaging (ISBI), 2012: 1523–26. doi:10.1109/ISBI.2012.6235862.

Yushkevich, Paul, Abraham Dubb, Zhiyong Xie, Ruben Gur, Raquel Gur, and James Gee. 2005. "Regional Structural Characterization of the Brain of Schizophrenia Patients." *Academic Radiology* 12 (10): 1250–61. doi:10.1016/j.acra.2005.06.014.

Zhang, Y, M Brady, and S Smith. 2001. "Segmentation of Brain MR Images through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm." *IEEE Transactions on Medical Imaging* 20 (1): 45–57. doi:10.1109/42.906424.