

Probabilistic atlas statistical estimation with multimodal datasets and its application to atlas based segmentation

Hao Xu

▶ **To cite this version:**

Hao Xu. Probabilistic atlas statistical estimation with multimodal datasets and its application to atlas based segmentation. Statistics [math.ST]. Ecole Polytechnique X, 2014. English. NNT : . pastel-00969176

HAL Id: pastel-00969176
https://pastel.hal.science/pastel-00969176

Submitted on 8 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Thesis
submitted in partial fulfillment for the degree of
Doctor of Applied Mathematics
in the
Ecole Doctorale de l'Ecole Polytechnique
Centre of Applied Mathematics
by

Hao XU

Probabilistic atlas statistical estimation with multimodal datasets and its application to atlas based segmentation

Rapporteurs:

Pr. Marc	NIETHAMMER	University of North Carolina
Pr. Frédéric	RICHARD	University of Aix-Marseille

Publicly defense on 31st March 2014 before the jury consisting of

Pr. Stéphanie	ALLASSONNIÈRE	Ecole Polytechnique	(Supervisor)
Pr. Emmanuel	BACRY	Ecole Polytechnique	(Examiner)
Pr. Antonin	CHAMBOLLE	Ecole Polytechnique	(Examiner)
Pr. Olivier	COLLIOT	Brain and Spine Institute	(Examiner)
Pr. Frédéric	RICHARD	University of Aix-Marseille	(Rapporteur)
Pr. Bertrand	THIRION	Inria Saclay	(Supervisor)

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisors Stéphanie and Bertrand.

I am deeply grateful to Stéphanie. Thank you for agreeing to supervise me and trust me for three years. Thank you for the continuous support of my Ph.D study and research, for your patience, motivation, enthusiasm, and immense mathematics knowledge. Your guidance helped me in all the time of research and writing of this thesis. Finally, thank you for this fascinating and useful subject.

I would also like to thank Bertrand. Thank you for your knowledge in the domain of medical images. Thank you for preparing the datasets and validating the application of my statistical model in the medical images.

I also wish to express my deep gratitude to all those who enabled me to complete this work. I sincerely thank Marc and Frédéric for agreeing to be the rapporteur of this thesis and for their careful reading of the manuscript. I also wish to express my gratitude to Antonin, Emmanuel and Olivier for agreeing to be examiners in the jury.

I greet all the members of CMAP in which there is a warm and friendly atmosphere. I want to especially thank Christophe for proposing this project to me. I would also like to thank the doctors of CMAP and especially these in the office 2011. It is with great pleasure that I thank those who have accompanied me throughout my thesis: Camille, Chao, Maxime, Laurent, Xavier, Xiaolu, and the horde of new doctors. I want to thank Sylvain for installing the software and others computing supports. I also want to thank the secretaries of the CMAP for their incredible efficiency, their kindness and smile in all circumstances.

I also greet all the members of the Parietal group. The group has been a source of friendships as well as good advice and collaboration. Pierre and Gäel give insightful comments and suggestions. I want to thank Alexandre and Fabien for your help of using python. I thank Alexandre, Virgile and Viviana for accompanying me for the MICCAI conference.

I would like to thank my friend Min for correcting the English errors in my articles and thesis. I want to thank Huiqi, Ke and Zhen for being with me when the work was not progressing well and when I was sad.

I also think back to my parents. Thank you for always supporting in my choice and giving me your faith. I also want to thank my French family the ROUX. I am glad to be one of the family. It is my pleasure to spend the holidays with you which makes me forget all the unpleasant things.

Contents

Acknowledgements **ii**

1 Introduction. **1**
1.1 The human brain and its shape . . . 2
1.1.1 Brain anatomy . . . 2
1.1.2 Brain imaging techniques . . . 4
1.1.3 Computational neuroanatomy . . . 5
1.1.4 Publicly available brain templates . . . 7
1.1.5 Some application of digital brain templates . . . 11
1.2 Segmentation and registration . . . 12
1.2.1 Registration . . . 12
1.2.1.1 Non-rigid registration methods . . . 13
1.2.1.2 Application of brain image registration . . . 15
1.2.1.3 Non-rigid registration algorithms . . . 16
1.2.2 Segmentation . . . 16
1.2.2.1 Methods of segmentation . . . 17
1.3 Template estimation . . . 20
1.4 Algorithms used in this work . . . 22
1.4.1 Gradient descent . . . 22
1.4.2 Stochastic algorithms . . . 23
1.5 Contributions of this work . . . 25
1.5.1 Generative Statistical Model . . . 25
1.5.2 Statistical Learning Procedure . . . 26
1.5.3 Segmentation of new individuals . . . 27

2 Probabilistic Atlas and Geometric Variability Estimation. **29**
2.1 Introduction . . . 30
2.2 The Observation Model . . . 32
2.2.1 Statistical Model . . . 33
2.2.2 Parameters and likelihood . . . 35
2.2.3 Bayesian Model . . . 36
2.3 Estimation . . . 37
2.3.1 Existence of the MAP estimation . . . 37
2.3.2 Consistency of the estimator on our model . . . 38
2.4 Estimation Algorithm using Stochastic Approximation Expectation-Maximization 39

2.4.1 Model factorization . . . 39
2.4.2 Estimation Algorithm . . . 40
Step 1: Simulation step. . . . 40
Step 2: Stochastic approximation step. . . . 41
Step 3: Maximization step. . . . 41
2.4.3 Convergence analysis . . . 42
2.5 Experiments and Results . . . 43
2.5.1 Simulated data . . . 43
2.5.2 Real data . . . 47
2.6 Conclusion and discussion . . . 51
2.7 Proof of Theorem 2.1 . . . 52
2.8 Proof of Theorem 2.3 . . . 54
2.8.1 Proof of assumption (**A1'**). . . . 54
2.8.2 Proof of assumption (**A2**) . . . 56
2.8.3 Proof of assumption (**A3'**) . . . 58

3 Bayesian Estimation of Probabilistic Atlas for Tissue Segmentation 59
3.1 Introduction . . . 60
3.2 Material . . . 61
3.3 Methods . . . 61
3.3.1 Statistical Model. . . . 61
3.3.2 Estimation Algorithm. . . . 63
3.3.3 Segmentation of new individuals. . . . 64
3.4 Experiments and Results . . . 65
3.5 Conclusion and Discussion . . . 69

4 Bayesian Estimation of Probabilistic Atlas for Anatomically-Informed Functional MRI Group Analyses. 71
4.1 Introduction . . . 72
4.2 Methods . . . 73
4.2.1 Statistical Model. . . . 73
4.2.2 Estimation Algorithm. . . . 75
4.3 Experiments and Results . . . 76
4.3.1 Simulated data. . . . 76
4.3.2 In-vivo data. . . . 80
4.4 Conclusion . . . 82

5 Including Shared Peptides for Estimating Protein Abundances: A Significant Improvement for Quantitative Proteomics. 83
5.1 Introduction . . . 84
5.2 Method . . . 84
5.3 Material . . . 88
5.4 Results . . . 90
5.5 Conclusion . . . 93

6 Conclusion and Discussion. 95
6.1 Summary . . . 96
6.2 Large deformations for deformable template estimation . . . 97

6.3 Multicomponent generalization of the models . . . 100
6.4 Other remarks . . . 100
6.4.1 Extension of the multi-modal atlas . . . 100
6.4.2 Kernel choice . . . 101
6.4.3 Algorithm implementation optimization . . . 101
6.4.4 Bias field correction . . . 102

A Definition of the most used similarity measures. 103

Bibliography 107

Dedicated to the XU and ROUX families

Chapter 1

Introduction.

This chapter introduces the context of my PhD project. In particular, I will present key concepts of brain anatomy and computational neuroanatony that underly image registration, segmentation and template estimation, and give an overview of past and present state-of-the-art solutions to this estimation challenge. I will describe some issues that these methods have to deal with and show how we handle them in this work.

1.1 The human brain and its shape

1.1.1 Brain anatomy

The human brain is the most complex organ in the body. It is the center of the nervous system and is protected by the cerebrospinal fluid (CSF). It has a heterogeneous organization, with various sub-structures associated with different functions. It can first be divided into three major parts (Fig.1.1): 1) the cerebrum, the largest part of the brain, that controls movement, sensation, cognition, etc.; 2) the cerebellum, that is involved in coordination of muscle movement and balance; 3) the brainstem, that is involved in relaying information from the cerebrum and cerebellum to the rest of the body.

The great longitudinal fissure separates the cerebrum into two hemispheres. The corpus callosum is the nervous tissue that connects the two hemispheres. The hemispherical surface is folded with various sulci, while the curved surfaces between the sulci are called gyri. The sulci are not formed randomly in the growth of the brain: their spatial organization, shape and time of appearance are reproducible, while their depth and bending are slightly different from person to person. Some of the most important sulci (Fig.1.2) are the central sulcus, lateral sulcus and parieto-occipital sulcus. The shape of sulci and gyri are grossly symmetric across the left and right hemispheres. They separate the hemispheres into four lobes (Fig.1.2): 1) the frontal lobe is involved in abstract thought, problem solving, emotion, etc.; 2) the parietal lobe is involved in sensory combination and comprehension, language, reading and visual functions; 3) The

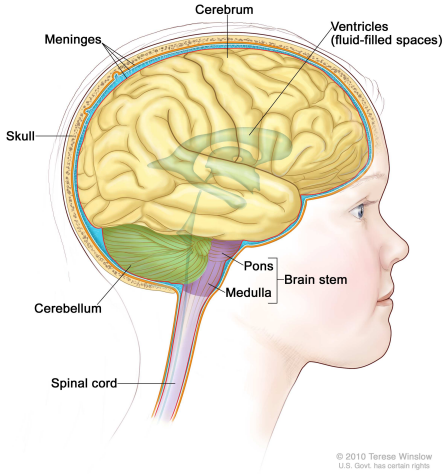


FIGURE 1.1: The brain's macroscopic organization. Image comes from http://www.cancer.gov/cancertopics/pdq/treatment/child-brain-stem-glioma/Patient/page1/AllPages/Print

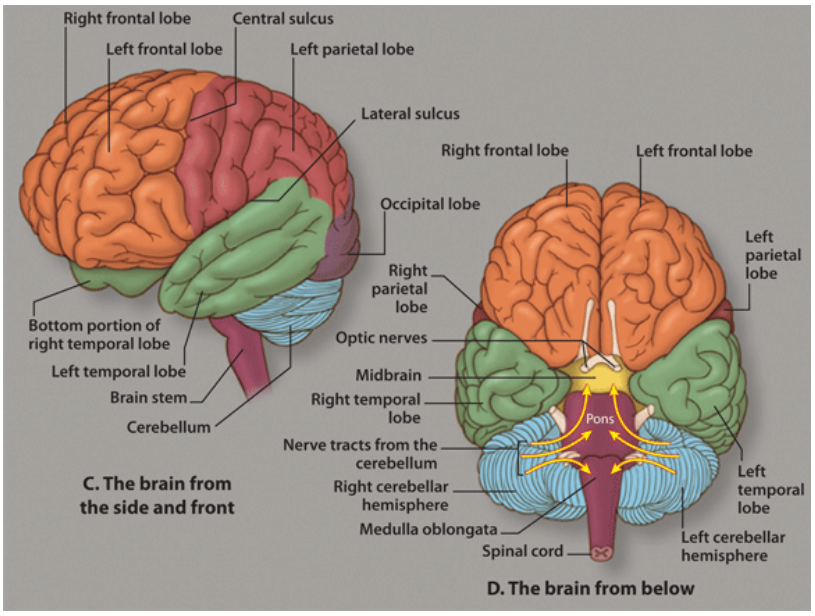


FIGURE 1.2: Large-scale organization of the cerebrum. Image comes from http://forums.psychcentral.com/psychiatric-medications/254913-1-amphetamines-modafinil-armodafinil-their-effectiveness-effects-brain.html

occipital lobe is involved in reading and vision; 4) the temporal lobe is involved in hearing, memory and behavior.

Based on their cellular and molecular content, brain tissues can be split into different types; the main distinction lies between gray matter (GM) and white matter (WM). GM consists of neuronal cell bodies, neuropil, glial cells and capillaries. It is present in the cerebral cortex (surface of the cerebrum), the cerebellar cortex (surface of cerebellum), subcortical nuclei (thalamus, putamen, hippocampus, etc.) and brainstem. It is the part of the central nervous system that performs information processing. WM consists mostly of glial cells and myelinated axons; it forms the bulk of the deep parts of the brain and the superficial parts of the spinal cord; it transmits signals from one region of the GM to another.

One aspect of our work is the segmentation of brain images into these different tissue types.

Part of the human brain complexity lies in its variability: there do not exist two identical brains in the population, and differences between two brains can be large. As described previously, the shape and the location of the main sulci is relatively stable. However, brains also vary from one subject to another for the others structures. Moreover, some patterns (e.g. the presence of some folds) can be absent or present depending on the subject. Such variability presents an ultimate limitation to any geometrical description of a standard brain.

1.1.2 Brain imaging techniques

Brain imaging techniques allow researchers to observe the shape, activity or abnormalities in the organization of the human brain, without neurosurgery. There are a number of imaging techniques in use today in research facilities and hospitals throughout the world. For example, anatomical brain data can be obtained by the three-dimensional tomographic imaging technology such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Diffusion Magnetic Resonance Imaging (dMRI); data on human brain function, on the other hand, are obtained by Positron Emission Tomography (PET) or functional Magnetic Resonance Imaging (fMRI); they reflect brain activity through complex metabolic pathways.

CT scanning (http://en.wikipedia.org/wiki/X-ray_computed_tomography) builds up a picture of the brain based on the differential absorption of X-rays. It uses a series of X-rays of the head taken from many different directions. Images made using X-rays depend on the absorption of the beam by the tissue it passes through. Bone and hard tissue absorb X-rays well, air and water absorb very little and soft tissue is somewhere in between. Thus, CT scans reveal the gross features of the brain but do not resolve its structure well.

MRI (http://en.wikipedia.org/wiki/Magnetic_resonance_imaging) generates images by using the response of brain tissues to radio-frequency excitation in the presence of an intense steady magnetic field. During an MRI scan, a detector records the signal emitted by the hydrogen atoms excited in a fixed magnetic field; these signals are localized through a complex encoding scheme driven by the gradients of the scanners and, after computerized reconstruction, yield 3D brain images. It is particularly effective to observe and characterize brain tissues. Its main features are: 1) compared with other imaging modalities, MRI has the highest contrast for soft tissues; 2) MRI can be used to acquire images with any direction and orientation; 3) MRI enjoys a high resolution (1mm).

fMRI (http://en.wikipedia.org/wiki/Functional_magnetic_resonance_imaging) is an application and further development of MRI: it uses MRI to measure functional responses of the human brain and nervous system. It works by measuring the changes in blood oxygenation and flow that reflect the activity of neurons given that the subject is doing a specified cognitive task. It can be used to produce activation maps showing which parts of the brain are involved in this task. This can also be completed by the study of other electrophysiological techniques such as Electroencephalography (EEG) or Magnetoencephalography (MEG).

dMRI (http://en.wikipedia.org/wiki/Diffusion_MRI) is an MRI method that maps the diffusion process of molecules, mainly water, in the tissues. Molecular diffusion in tissues reflects interactions with many obstacles, such as fibers. Water molecule diffusion patterns can therefore reveal microscopic details about tissue architecture, either in normal or diseased state. It is the only non-invasive imaging method to display the fiber bundles in WM.

PET (http://en.wikipedia.org/wiki/Brain_positron_emission_tomography) can detect the positron emission of the glucose solution containing weak concentrations of a radioactive isotope when the brain absorbs this solution. The brain consumes energy when it works, PET scanning can display the brain regions that consume more glucose. The disadvantage of PET is the use of radioactive substances.

In this work, we will concentrate our study on MRIs and fMRIs, although our modeling can be used for other imaging modalities.

1.1.3 Computational neuroanatomy

The main access to brain structure and function in humans is currently provided by neuroimaging. In particular and as recalled above, Magnetic Resonance Imaging (MRI) provides measures that are used to distinguish the nature of different brain tissues, markers for functional brain subdivisions and markers of the local orientation of brain fibers through three sub-modalities: anatomical, functional and diffusion MRI. These measures are currently at the mm scale and such multi-modal datasets are commonly acquired across the whole brain volume in standard acquisition sessions.

However, these images display different resolutions and contrasts, suffer from distortion, patient motion during acquisition, which leads to a low signal to noise ratio and, sometimes, patient specific contrasts or artefacts. At the observed resolution, cross-subject variability of the measurements is striking, and naturally yields the question of what features are common across subjects and how to compare different human brain images. Understanding the topographical organization of the most common features means building a reference image, that will be called *template* in the sequel, while allowing some flexibility in the comparison of individual data with this template. The reference image is assumed to characterize the population as it exhibits the common features and provides also a reference frame where individual data can be compared.

Computational anatomy [66] is emerging for solving these problems. It combines the knowledge of human anatomy, mathematical methods and computational techniques in order to design models and develop algorithms to compare images, estimate the template

image and transport the available information from the template to new subjects. The goal of these efforts is to help clinicians to measure and quantify the morphometry of the organs, localize and describe what happens during a time period and provide statistical analysis tools to characterize the differences between a given subject and the template.

Computational neuroanatomy is a branch of Computational Anatomy that considers the human brain. Because of the complex structure of the brain, computational neuroanatomy has to face more challenges, such as tissue and structure segmentation, functional activity detection, etc. Overall, the main challenge, compared to others organs, is clearly the huge variability that can be observed even in a healthy group.

One side of the research in computational anatomy is the statistical analysis that aims at characterizing group difference, as well as classifying and discriminating between populations. It can be split up into three main different direction. The first one is to *determine the shape and contrast or equivalently the image of a standard brain.* Not only should we exhibit the template brain but it is also important to quantify the characteristic geometrical variability of the shape in a specified population. These two elements make it possible to better summarize this population. Therefore, in this work, we will focus on the determination of an *atlas* that involves both the template image and a quantification of its standard geometrical variability. This step is detailed in Section 1.3.

The second one is to *compare individual images with this atlas.* This requires to design the mathematical methods that make it possible to compare the observations with the atlas. This has been at the core of intense researches and is known as registration techniques (see Section 1.2.1). This step enables to warp the observed images in the same frame. As we will see next, it also plays a crucial role in the atlas estimation.

The third one is to *identify the structures in the observed images with those of the template, up to some deformation.* This is know as image segmentation (see Section 1.2.2) and has also been intensively studied. This segmentation step can be used as prior information for registration in order to improve the related estimation scheme, as it better fits the structures of interest [119]. Segmentation can also be the purpose of the registration procedure that carries the segmented structures from one subject onto another [155]. This clearly highlights the link between these two problems and therefore suggests that they should be combined with each other to perform well. This is one idea that we have developed in our work.

Some clear application of the *atlas* estimation, *registration* and *segmentation* issues arose from the clinical needs.

Computational neuroanatomy is the basis of **image-guided surgery** [109] and surgical simulation [40]. It helps surgeons to localize lesions such as tumors or strokes using segmentation algorithm on pre-surgical acquisitions. In the near future, brain imaging together with computational anatomy will provide critical information to define the course of tumor resection surgery by registering the patient pre-surgical image to the current state of the organ. This will help to avoid damaging or removing important brain structures that can be preserved.

Another example lies in **research on brain development**, that tries to analyze normal development trajectories and identify the differences in the different brain structures or regions between controls and patients during growth, aging or disease progression. To quantitatively characterize the differences in the brain structures between the normal individuals and the patients in different developmental periods, these structures have to be extracted from neuroimaging data. This implies the segmentation of brain images. To track the same patient along time, registration methods have to be used to analyze the changes in the structures. Most importantly, the comparison between a new patient to one or several reference images (representing healthy controls and patients with a specific disease) should be possible to classify them, e.g. to diagnose a disease.

1.1.4 Publicly available brain templates

A template is defined as a gray level image with anatomical labels. It can also be defined as a probabilistic image that gives the probability that the voxel belongs to each tissue. An atlas is defined as a (probabilistic) template together with the deformation metric like in [82]. The construction of digital atlases of the human brain is currently a very important topic, for which a lot of efforts have been spent. In the following we will present three brain templates that are widely used by clinicians: the well known three-dimensional Talairach-Tournoux brain template (referred to as Talairach brain Atlas in the literature), the Whole Brain Atlas developed by Harvard Medical School and the Brainweb template developed by McGill University in Canada. Note that the word *atlas* is sometimes used for *template*. In the present work, we highlight the complementary information that the atlas carries compared to the template: the geometrical variability.

Talairach brain atlas :

The Talairach Brain Atlas [131] relies on a coordinate system that is based on the anterior-posterior commissural (AC-PC) line. It relies on a grid system to project the brain images into the three-dimensional space. Talairach Brain Atlas was constructed upon postmortem sections of a 59-year-old French woman. The brain specimen was cut every 2-5mm in sagittal direction, then the cross-section

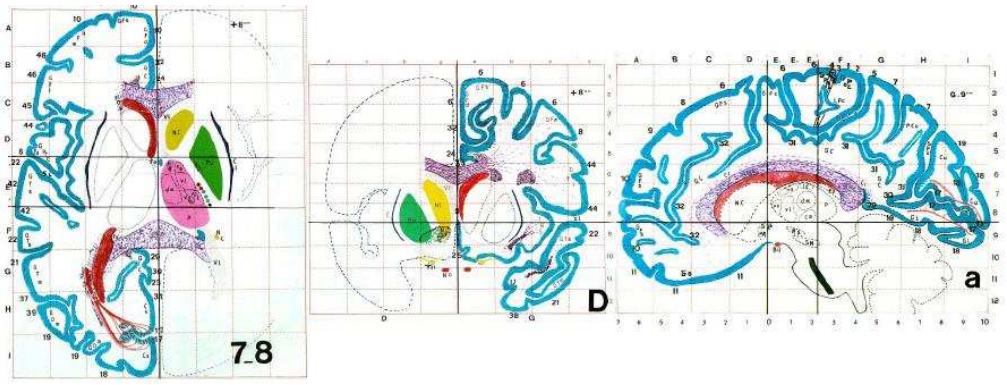


FIGURE 1.3: The axial, sagittal, coronal view of Talairach-Tournoux brain atlas. Image comes from http://psyphz.psych.wisc.edu/web/afni/class_handouts/afni08_talairach/afni08_talairach.jpg/index.html

of each slice was photographed. Neurologists outlined the contours of the brain structure according to the photographs, and then color filled these contours, the same brain structures having the same color or texture expression. The horizontal and coronal section data were created by the interpolation of sagittal cross-section data. Talairach brain atlas divides the reference brain into 8 parts in the X-axis direction (from the left to the right of the brain), 11 parts in the Y-axis direction (from the front to the back of the brain) and 12 parts in the Z-axis direction (from the button to the top of the brain), see Fig.1.3. The Talairach space provides a cube where the brain should fit in a specified orientation. Since this reference is based on the assumption that the brain is symmetric, it contains only one brain hemisphere.

The interest of this template is that, when compared to a new individual, the localization of brain structures is accurate for areas close to AC-PC, for example, the thalamus. However the template accuracy decreases significantly for cortical areas, especially those that are highly asymmetric between the two hemispheres (e.g., the temporal lobe). Nevertheless, the Talairach brain template is widely accepted by the neuroscientists, because it was historically the first computational template. The neuroscientific literature still often refers to *Talairach coordinates*.

There are however several disadvantages. The first one is its low resolution. The second one is that it makes a strong assumption on the symmetry of the whole brain which is not satisfied and is a rough approximation which may lead to misanalysis of patients. The third one is that it is based on a single subject who may not be representative of the whole population.

Whole Brain Atlas :

Johnson and Becker, Brigham and Women's Hospital [75] develop several digital templates of the human brain http://www.med.harvard.edu/aanlib/home.

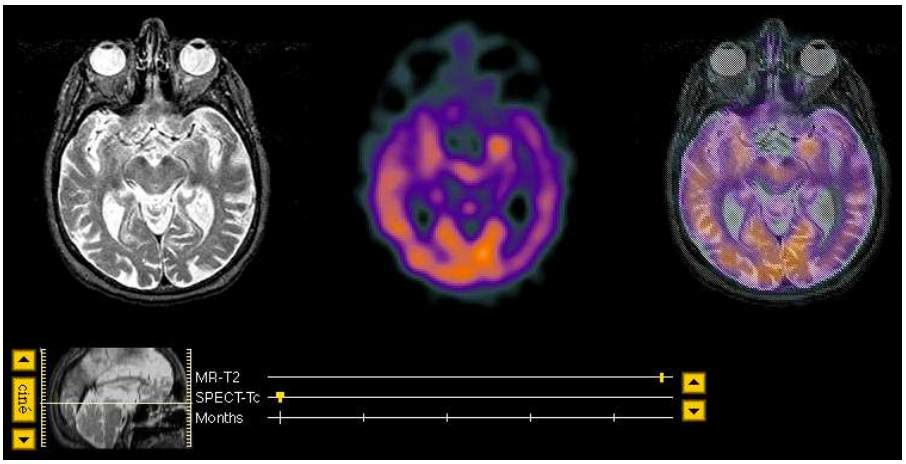


FIGURE 1.4: One slice of the whole brain atlas for Alzheimer's disease. Image comes from http://www.med.harvard.edu/aanlib/cases/case3/mr1-tc1/020.html. T1 MRI, SPECT (an imaging technique that is similar to PET) and coupled T1-SPECT images are shown here. On the slice of T1 MRI, the atrophic hippocampus and amygdala can be seen. These structures subserve memory function, and are the sites of major damage in Alzheimer's disease. On the SPECT image, the dark blue regions in the parietal lobes represent areas of decreased blood flow or perfusion. This reduction in blood flow is due to the functional"disconnection" of this from other brain regions affected by the disease.

html. The normal template is based on 3D brain data using 1.5T MR images from a 25-year-old white man right-handed with a resolution of $0.92 \times 0.92 \times 1.5mm$ leading to $256 \times 256 \times 128$ image size. The template includes gray matter (divided into several lobes and gyris), cerebellum, brainstem structures, corpus callosum, basal ganglia, limbic structures and ventricular system. The normal brain is divided into 100 structures.

The *Whole Brain Atlas* research program not only includes the normal human brains, but also diseased or injured brains. Many diseases are represented by images from one or more modalities, including MR, CT and PET (Fig. 1.4). Some of them include images obtained at different time, therefore the changes can be observed longitudinally.

As for the Talairach brain template, they are constructed from the images of one single person. Although providing a large range of images, all these templates may not be representative of each population. Note that an atlas-based model segmentation method using this atlas has been developed [79].

BrainWeb brain atlas :

Evans, Collins et al. [43] from Montreal Neurologic Institute (MNI) created a probabilistic anatomical template using the MR images of 305 normal subjects. First, they manually defined some landmarks in 241 MR images, then they matched the landmarks to their position on the Talairach atlas. Then, they used a linear

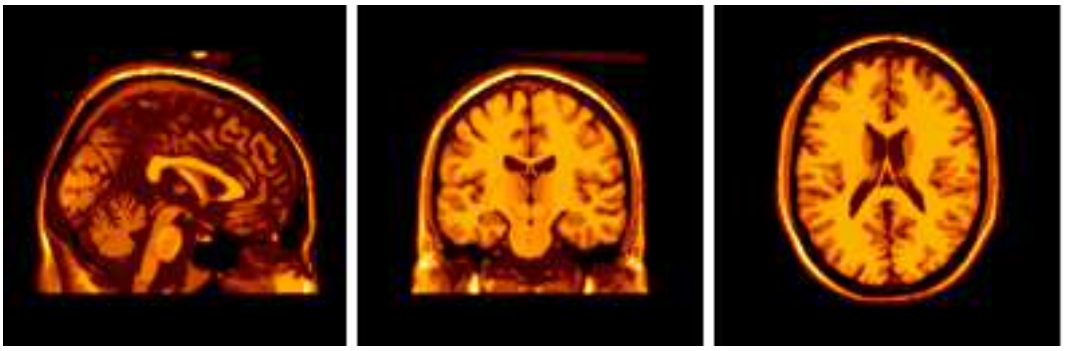


FIGURE 1.5: The coronal, sagittal, axial view of BrainWeb brain atlas. Image comes from http://brainweb.bic.mni.mcgill.ca/brainweb/selection_normal.html

algorithm to match 305 MR images to the average of the 241 brains that had been matched to the Talairach atlas and took the average of 305 mapped images as template.

In addition, they scanned a healthy young subject in stereotactic space 27 times. The images were subsampled and intensity averaged in order to obtain a data with high contrast and high signal to noise ratio. The volume contains $181 \times 217 \times 181$ voxels. The map contains 10 different brain volumetric data sets that define the spatial distribution for these structures, including gray matter, white matter, cerebrospinal fluid, skin, skull, fat, etc.

The digital brain model can be used to simulate the head tomography, or is an ideal tool to test inter-modality registration. This template is devoted to model-based segmentation and coregistration. It is the standard template in many softwares, such as Statistical Parametric Mapping (SPM) [7] [1]. In some other software, like Freesurfer [2] [2] it is used as a coarse alignment step.

The problem of this template is that the linear transform does not match the brains completely to the Talairach space. This makes the MNI brains slightly larger than the Talairach brain. The differences are larger as one gets further from the middle of the brain.

[1] SPM is an open source created by members and collaborators of the Wellcome Trust Centre for Neuroimaging. It deals with pre-processing, image registration, image segmentation, fMRI analysis, EEG/MEG analysis, etc. It typically implements voxel-based methods for brain data analysis. A parametric statistical model is fit in each voxel (it is thus a mass-univariate approach). It uses the estimated parameters of the univariate model to do a statistical test and then creates images that display the statistical information.

[2] Freesurfer is an open source software that provides a tool suite for processing and analyzing human brain MRI images, including skullstripping, image registration, image segmentation, cortical surface reconstruction, cortical thickness estimation, fMRI analysis, etc. The underlying spatial model is surface-based: it warps each individual cortical hemisphere, corrected to have a spherical topology, to a reference sphere. Then, a map of the surface curvature is compared with that of the template brain, sampled on the sphere. The individual spherical model is warped to match the contrast of the template.

Talairach template and the Whole Brain atlas were created from the images of one person, hence they cannot capture relevant characteristic elements of the whole population. On the other hand, the MNI template uses 305 subjects which makes it more representative. However, the method uses linear deformations, which is not accurate enough. Moreover, it is based on landmarks that are always difficult to place exactly; since these landmarks are defined manually, this makes the procedure hardly reproducible. The neuroimaging community is still lacking a standardized procedure to estimate brain atlases. Other questions arise: because of the large variation across subjects, one single template may not be enough. It is also important to consider that the comparison of mean images or templates may not be the right statistic to compare two populations: indeed, two populations may have the same mean image although the differences may appear in the deformation distribution. Last but not least, it only relies upon anatomical images, therefore it might not be the optimal setting to detect functional activity or other contrasts of interest.

1.1.5 Some application of digital brain templates

Teaching of neuroanatomy :

Traditional neural anatomy teaching requires the aid of anatomical charts, books, images and renderings. Due to the very complex structure of the human brain, it is difficult to understand its shape, subcortical structures and the relationships between them. Moreover, brain-structure is, to some extent, subject dependent, which makes the use of a digital brain templates particularly relevant. Some publications and web applications [127, 146] make it possible to observe the brain easily with any translation, rotation or zoom on the region of interest. Moreover, as they are supposed to capture the most common features, they appear useful in brain anatomy understanding.

Surgical planning and reference :

As already mentioned above, digital brain templates can provide accurate and reliable information for surgical planning. It can be used to assess the risk of surgery and choose the best surgical approach. A particular example can be mentioned here with the deep brain stimulation (DBS) [21, 152]. The template is used in order to predefine the coordinate of one point in a subcortical structure where a deep electrode is inserted in order to stimulate a precise territory and cure some disorders. It is important to couple the anatomical and functional images, as this makes it possible to identify some areas that are critical to some essential functions and avoid to lesion those during the surgery.

Atlas-based segmentation :

Due to the large variability of the human brain, tissue segmentation is particularly difficult. It becomes even more challenging as we have to face the partial volume effect (PVE) issue. In low resolution images, PVE appears at boundaries between tissues, where a given voxel contains several tissue types. This blurs the boundaries between tissues. The atlas information can be used to guide the segmentation using both the template image to indicate the location and shape of the tissues and structures but also using the quantification of the normal geometrical variability to favor particular shapes and spatial organization.

All these applications have led to the need of population based atlas and template estimation. Many template construction methods have been proposed in the past decades. The most common approach has been to choose as the template one image among several observed ones. This image is considered as standard or characteristic of the population under study. The other observations are then mapped to this *template* - via different registration algorithms - and can then be compared in this common coordinate frame. The problem is that the *template* is subject to an arbitrary choice. If it is far from the true population mean - which has to be defined- it does probably not reflect the population correctly and the registration will not show typical variability as it will first account for correcting the bias [27]. This is not satisfactory and has motivated the design of methods for the statistical estimation of templates or atlases from a population of observed images.

Ahead from this, registration techniques and segmentation methods, that are prior steps to atlas estimation, have raised a lot of attention in the last decades. In the following, we briefly recall some principles of these two topics before introducing the core of our subject, *statistical atlas estimation*.

1.2 Segmentation and registration

1.2.1 Registration

Medical image registration has become an efficient tool to compare images from two different subjects. The difference between two objects observed in images correspond to a variability in position, described by rigid transformations, size, described by affine transformations, and more general transformation that change the details of the objects. These different classes of transformations are characterized by different degrees of freedom: six for rigid transformation, twelve for affine transformation, a few tens or

hundred of degrees of freedom when parametric models are used (polynomials, spline) and up to thousands or even infinite when considering more generic models that induce large deformations. Rigid or affine methods are widely used in clinical practice because of their simplicity and reduced computational cost. However, as the variability of the brain consists of changes in its size, shape and internal tissue organization, affine deformations are not sufficient to represent brain variability. Non-rigid deformations have therefore been proposed to overcome this limitation. The research on non rigid deformation methods has made considerable progress in the past years. In the following, we briefly describe some of them.

1.2.1.1 Non-rigid registration methods

Non-rigid registration based on affine transformation and polynomials :

Farneback [55] proposed a non-rigid registration method based on polynomial transformation. It was further developed by Wang et al. [143]. The disadvantage of this method is that it requires polynomial functions of high degree to produce a well behaved transformation.

Non-rigid registration based on physical models :

Non-rigid registration techniques based on physical models are also common. These approaches assume that the registration object is a homogeneous isotropic elastic body. They first create a physical model of the object and then deform this model under external forces to achieve the registration with the reference image. This registration method can ensure the smoothness of the deformation and the preservation of the object topology. Most contributions on these methods have focused on the definition of external forces, while some of them improve the method by adjusting the elastic properties of the physical model. For example, Christensen [38] proposed to replace the elastic model by viscous fluids model, however this method increases the computational complexity, Bro-Nielsen [31] improved the viscous fluids model by using fast convolution methods.

Non-rigid registration based on a smooth function :

Alternative methods based on smoothing functions, such as thin-plate spline, multivariate quadratic equations or transforms of a Gaussian function have also been used in his context. These models involve smooth functions, which are regularized so that they represent a prior on relevant deformations. Meyer et al. [93] use a thin-plate spline interpolation to adjust the position of homologous control points and calculate the similarity of image to the reference image after transformation.

Like the thin-plate spline, B-spline is also widely used in non-rigid transformation. For this transformation, the deformation field is calculated by using B-spline interpolation to adjust the position of the control points. This transformation is usually performed using mutual information similarity measure (see [116, 128], and definition in Appendix A).

Linearized deformations :

Linearized deformation methods [10, 15] define a deformation φ of a domain D as the displacement of each point x in D by a vector $v(x)$, which is written as:

$$\varphi(x) = x + v(x).$$

In this case, the image under deformation is modeled as an elastic body. The model is simple because it only depends on a vector field v. An important drawback is that when the source and target images are interchanged, the obtained transformation is not the inverse of the previous solution, as the invertibility is not ensured.

Large Deformation Diffeomorphic Metric Mapping (LDDMM) :

LDDMM [134] has been proposed to model large deformations and overcome the non invertibility issue. In this case, the deformation is modeled through the composition of a velocity field that evolves over time according to the Lagrange transport equation. This method provides the definition of a mathematical metric on the space of images. This distance is the length of the geodesic that connects them according to this metric and can be used to study the anatomical variability [23, 137]. However, it is computationally costly because the velocity field has to be integrated over time. Recent methods [52, 53] approximate this diffeomorphism by a finite-dimensional vector enabling fast and accurate registration.

Demons and Diffeomorphic demons :

Thirion [133] proposed to perform image registration as a diffusion process inspired by Maxwell's Demons. The diffusion is driven by the force called "Demons force". The Demons algorithm is an efficient algorithm that provides dense correspondences but lacks a sound theoretical justification. Vercauteren et al. [139] propose a non parametric diffeomorphic image registration algorithm based on the initial Thirion's demons algorithm. They provide theoretical properties of the different variants of this algorithm. These algorithms converge quickly, however, they are not based on a sound mathematical model that would provide a clearly defined metric on the image space. Tools can be download in Insight Segmentation and Registration Toolkit (ITK) [6].

A more detailed description of multiple image registration techniques are presented in [126]. Fourteen of them are compared in [80].

1.2.1.2 Application of brain image registration

According to the modality and the patient, non-rigid registration can be used in different frameworks:

Registration of the same modality of one patient (longitudinal study) :

This approach is used to study the evolution of a subject over time. If these images contain some non-rigidly evolving structures such as tumors or stroke lesions, the brain is consequently deformed between one time point and the following ones. The registration map highlights (and depending on the method, quantifies) the differences, which may be used in disease diagnosis or observation of treatment results.

Registration of images from the same modality in different patients (intersubject) :

This is the most common use of registration mappings. Indeed, it makes it possible to compare two subjects by highlighting shape differences. This is now used for statistical analysis of populations that were either compared pairwise (for example [94]) or mapped to a given template. In particular, the statistical analysis can be performed in the space of deformation, which appears to be a linearizing space to analyze shape differences (among many others, we can cite [37, 44, 90, 137]).

This is one aspect that we will consider for our purpose in this work.

Registration of images from different modalities in the same patient :

In this case, the images to be registered come from different modalities of the same patient. This is currently the most frequent use of registration in clinical settings (mostly with rigid-body deformations). It is used to fuse the information from the different modalities into the same image so that the information can be combined in order to increase the accuracy of the analysis.

This is also an aspect that we will use in the following. In particular, considering together anatomical and functional MRIs can be used to highlight some particular regions in the gray matter that are activated during specific cognitive tasks. Constraining this activation to appear only in the gray matter potentially provides more accurate localization and thickness of the gray matter and a better localization of the subcortical structures. On the other hand, as the gray matter is

observed in the anatomical image, this knowledge also increases the accuracy of the support of the brain activity. This coupling will be at the core of our multimodal atlas estimation in Chapter 4.

1.2.1.3 Non-rigid registration algorithms

There are currently many reports about non-rigid registration algorithms in the literature. Most of the non-rigid registration problem can be seen as an optimization problem where the energy to minimize takes into account the difference between one reference image and the deformed image and a regularity term that penalizes "too large" deformations – which has to be defined according to the deformation framework. This can be expressed as

$$\hat{\varphi} = \arg\max_{\varphi \in \mathcal{S}} E(\varphi \cdot B, A, \varphi) = J(\varphi \cdot B, A) + L(\varphi)\,, \tag{1.1}$$

where A and B are the first and the second images, J is the similarity measure, φ is the transformation between the two images, $\mathcal{S}$ the space of admissible deformations and L the regularity term. The different approaches differ from each other by considering different $\mathcal{S}$, J and L.

Different frameworks to construct and constrain φ have been given above as well as the regularity term that includes the expected constraints on φ. For example, in LDDMM and Diffeomorphic demons, φ is a diffeomorpism. $\varphi \cdot B = B \circ \varphi^{-1}$ and L is the kinetic energy of the path generated by φ. The similarity measure J quantifies the similarity between two images and the most used are the correlation [107], gradient cross-correlation [33], mean square error [135], mutual information or normalized mutual information [130] that are described in Appendix A.

Although quite different, all these approaches can be solved the same way as the optimization of the matching energy E. Currently, gradient descent optimization method -and its accelerated versions- appears to be the easiest algorithm. However, other methods such as Newton-Raphson, quasi-Newton, the simplex method, simulated annealing or even genetic algorithms, may be used depending on the energy E.

1.2.2 Segmentation

Medical image segmentation is the key to performing computer image analysis. It has important significance in biomedical research, clinical diagnosis and pathological analysis. The main factors affecting the brain MR image segmentation are: 1) intensity

inhomogeneity: due to the radio frequency magnetic field inhomogeneity, MR equipment itself, the differences between tissues in the brain, etc. 2) Partial Volume Effect (PVE): due to the movement of the patient –given that the acquisition lasts several minutes– and the complex shape of the structures, the boundaries of the target structures are not continuous and the boundaries between the soft tissue can be blurred. 3) Image noise. These issues altogether make segmentation a challenging task.

1.2.2.1 Methods of segmentation

In this section, we recall few examples of image segmentation that we have grouped into four classes based on the way they are performed.

Gray level based approaches :

This class of methods includes the threshold-based method [122], region growing [54], clustering algorithms [62], etc. These methods are based on the gray level image directly. The threshold-based method assumes that the distribution of the target and the background is separable in the image histogram, so one can use a threshold to distinguish between the target and the background. The method is simple, easy to implement and fast. However in brain MR images, due to the intensity inhomogeneity, the distribution of the tissues often overlap. It is impossible to obtain the correct segmentation result by the threshold method. The main idea of the region growing method is to set the initial seed, classify the voxel that are similar to the seeds of a given class. The disadvantage of this method is that the result of the algorithm often depend on the selected seeds. Clustering algorithms include K-means algorithm [87], fuzzy C-means algorithm (FCM) [51], Gaussian mixture model (GMM) [48]. Wells et al. [145] propose a segmentation method using Expectation-Maximization (EM) algorithm, that can automatically segment the tissues, however this method requires the specification of Gaussian distribution for each class. Ahmed et al [8], Pham et al. [102] have proposed methods based on Fuzzy C-Means algorithm that segment and remove the bias field at the same time. However the initial values are sensitive for these clustering algorithm. In addition, most of the proposed models are based on the voxel information and do not consider the smoothness of the tissue. Ashburner and Friston [20] propose a probabilistic framework for joint nonlinear registration, intensity normalization and segmentation of a single image providing tissue probability maps.

Active contour method :

Active contour segmentation models have become an important research topic in the last decades: specifically, parametric active contour model (Snake model) and

geometric active contour model (level set model) are two commonly used active contour models. Here is a brief introduction of these two models, that come from computer vision and thus may fail due to the characteristics of brain images.

(i) Parametric active contour model

The parametric active contour model (Snake) was proposed by Kass et al. [78], further developed by [30] and successfully applied to the image segmentation [41, 92]. The idea is to find a methodology for the minimization of a criterion composed of the internal energy and external energy over the entire shape. Although this model has achieved great success in medical image segmentation, there are still some disadvantages: 1) the results are often dependent on the initialization; 2) the capture range of edge energy is small; 3) the sensitivity to noise is large, the model fails in images with weak boundaries. As the topology of brain MR images are complex, Snake models cannot be used to segment several structures in the brain MR images at the same time. However, they are often used on a single tissue segmentation [147], for example, segment the hippocampus, ventricles or tumor.

(ii) Geometric active contour model

Geometric active contour models can be considered as an improvement of Snake model, proposed by Caselles et al. [144] and Malladi et al. [89]. This model can be solved by the level set method. The level set method can be applied to image denoising and enhancement, image segmentation, image restoration. The level set model [18, 46, 101] improves upon the Snake model, as it allows topological changes (the contour can break up, merge, or disappear during the course of time evolution). However the resulting segmentation is sensitive to the initialization and it is difficult to obtain global optimal solution. Chan-Vese model [36] is a classical model, that assumes the image is divided into two parts with different means and segment the target from the background. Since it uses the gray level information of both the internal and external of the contour, it still works well for the images with large noise or weak boundaries. However the model assumes that the intensity is homogeneous in each class, therefore it often yields a wrong segmentation on brain MR images. Many segmentation models [142, 156] based on local information have been proposed in order to overcome the intensity inhomogeneity in the brain MR images. However the result is often dependent on the initialization.

Graph-based segmentation :

Graph-based segmentation was introduced for image analysis in 1971 [154]. This method converts the image to a weighted graph where the voxels are handled as nodes, then it uses a minimum cut criterion, e.g. minimal spanning tree, to get the best image segmentation. It essentially transforms image segmentation problem into a discrete optimization problem with convex relaxation. This class of methods

include the graph cuts [28], random walker [64], isoperimetric partitioning [65] and so on. Currently, the research of the graph-based segmentation focuses on the following aspects: 1) optimal cut-off criteria [28, 69]; 2) spectral methods for segmentation [45, 113]; 3) fast algorithms [29, 56].

Atlas-based segmentation :

Atlas-based segmentation can been considered as a registration problem, namely that of deforming a brain atlas into a new subject's brain. It relies on a reference image in which some structures have been segmented. Then, a non-rigid registration is applied from the reference to the new subject. The structures of the new subject are segmented by transporting the labels from the reference. Atlas-based segmentation is of particular interest when the information from the gray level intensities is not sufficient. For example, the gray level intensities of subcortical structures may be closer to the white matter (WM) than gray matter (GM) intensity mean. This makes it difficult to segment these structures as GM without any prior. However, atlas-based segmentation manages to segment these structures correctly with the help of a priori information given by the atlas [119].

There are two problems with atlas-based segmentation of brain images. The first one is the choice of the template, as it should be representative of a population and carry an accurate segmentation. The second one is the choice of the registration framework that constrains the deformation and therefore captures different similarity while leaving residuals which may be of interest. This assumes that the template is close to the subject's anatomy. Otherwise, large registration errors may cause important segmentation errors if large anatomical differences exist.

Aljabar et al. [9] proposed two methods for templates selection. The first one uses meta-information. They select the template that is closest in age to the image to be segmented. The second one uses similarity metrics to compare the images. Obviously, the quality of the results depends on the metric used.

Another solution is to use multi-atlas based segmentation. It has been proposed to better deal with the registration errors obtained when using a single template. It also better captures the anatomical variability. Klein et al. [81] and Wu et al. [148] show that using multi-atlas segmentation improves segmentation accuracy. One problem is that it is unclear how many templates should one use for the multi-atals based segmentation. Aljabar et al. [9] shows that it is not necessary to use all the observations in the data set as templates. First, if the number of "templates" is large, registering all "templates" to the image to be segmented increases the computation time. Second, the anatomical structure may vary across the population. For example, if a structure can be represented by two distinct

shapes, using a large number of templates may give a shape that does not represent either of them. In their study, they find that 15-25 templates are sufficient to get a good performance. However, the choice of the 15-25 templates still depends the metric they used.

When doing atlas-based segmentation, the templates selection depend on the metric used to compare images. The choice of the registration framework assumes that the template is close to the subject's anatomy. A statistical estimation of the atlas that contains the template and the metric is a better choice, because it avoids to fix the template or the metric which are closely linked. Now we will focus on atlas construction.

1.3 Template estimation

Two approaches have been used to construct templates from medical images: to take the image closest to the population mean or to estimate the true population mean.

Marsland et al. [91] propose a method to construct the template as the observed image that is the closest to the geometrical mean. They choose the target image that minimizes the sum of distances from this image to the rest of the images and maximizes the sum of mutual information (MI) between them. Park et al. [99] propose a method that deals with the same problem as Marsland et al.'s. Their selected image is chosen based on MI only, which is most robust to the noise inherent in anatomical images. These approaches gives the template with the smallest possible bias but nevertheless this template is still biased, as it is one of the image in the dataset. It is not the real mean geometry image.

Studholme [129] proposes a method to jointly register all images simultaneously to a common space that is very close to the mean geometry to reduce the bias inherent to the choice of one particular sample. A cost function is optimized with the aim of maximizing the similarity between images, while penalizing displacement from the average shape. However, this requires explicitly choosing a weighting parameter to specify the influence of the penalty term and thus how well the constraint is satisfied.

Bhatia et al. [26] propose a method where an arbitrary image is used just as an intensity reference, after which the similarity between images is maximized using non-rigid registration. To ensure that the image calculated in this fashion is actually the mean, they enforce the constraint that the sum of all transformations, represented by a suitable parametrization, is equal to zero. The algorithm does not require specifying any geometric reference, however an intensity reference has to be chosen to evaluate the similarity during the registration step.

Guimond et al. [67] propose an automatic method to build average anatomical brain atlases. First, they choose an image as the reference and register all images to it using an affine registration to correct the position and global shape differences. Then, they register the images after the affine registration to the reference using an elastic registration and calculate the average of the images and deformations after registration. At last, they apply the average deformation to the average images in order to compute the average atlas. The advantage of this method is that it depends less on the reference image used for its construction. A similar idea has been proposed in [86] where the template is the mean deformation of an hyper template. However it requires a pre-segmentation of the data.

Rueckert et al. [115] propose a statistical deformation models that can be used to construct an average atlas of the anatomy and their variability. They use a non-rigid registration algorithm. The transformation consists of a global transformation and a local transformation. This algorithm leads to good correspondences, in particular, in the subcortical structures.

Joshi et al. [76] propose a method to estimate a template given a collection of observed images. The problem is that the deformation is applied to the observations whereas it should act on the template (ideal image). This makes this estimation different from the computational anatomy model proposed by Grenander and Miller [66]. Moreover, the iterative numerical scheme is very sensitive to noise.

Niethammer et al. [96] propose robust estimation methods for parametric models based on Diffusion Weighted Imaging (DWI). By applying the estimation method to registered DWIs, a DW-atlas is constructed. This allows for the representation of average diffusion information with more flexible diffusion models than the diffusion tensor.

All the methods so far produce deterministic templates, that is to say a deterministic gray level image. On the other hand, probabilistic templates providing tissue probability maps are especially attractive, as they make it possible to take into account the uncertainty on the underlying tissue type, which is related to partial volume effect (PVE) or to perfectible registration. Moreover, most of the time, the template is estimated without the geometric variability which may be analyzed a posterior by PCA or ICA. However, the template estimation requires to choose a deformation framework which somehow freeze the geometric variability a priori. Therefore, the posterior analysis highly depends on the prior choice. This suggests that the template should be learned jointly with the geometric variability of the shapes represented in the population. These two quantities will in the sequel form the atlas of the corresponding population. This is the leading idea of our models.

In [25], a method was proposed to do the segmentation and registration jointly, while creating an average brain template. This approach combines groupwise registration using the Kullback-Leibler divergence and the EM algorithm for segmentation, and thus demonstrates the benefit of their integration. However it does not learn the geometric variability within the estimation procedure, which may reduce the accuracy of the template to match the observations with prior deformations.

Ribbens et al. [110] propose a probabilistic model to segment a heterogeneous data set of brain MR images simultaneously while constructing templates for each mode in the heterogeneous population using an EM algorithm. However, it performs clustering as an additional step, and does not learn the geometric variability of the population.

Sabuncu et al. [153] propose a spherical demons algorithm with geometric variability for registering images and for creating an atlas. The registration was more accurate and this registration could be used to transfer segmentation labels onto a new image. With such an approach, the segmentation is not performed during the estimation.

Glasbey and Mardia [61] made the first step towards the statistical estimation of a complete atlas based on a statistical model. This was improved by Allassonnière et al. [10] who proposed a model to create an atlas containing the geometric variability. As the inputs are scalar images, the template is also estimated as a scalar (gray level) image. As a consequence, the segmentation of the population is not part of the estimation process. Using both kinds of information increases the population classification accuracy, as the model better fits the observations, as for the population classification in [13], the segmentation (tissue classification of voxels) takes advantage of the registration to the template. Although they learn the geometric variability, their template is deterministic. In this work, we will develop these models in order to create a probabilistic atlas.

1.4 Algorithms used in this work

1.4.1 Gradient descent

Optimization is a branch of applied mathematics, the main research question is to find the minimum of a function $f(x)$. The gradient descent method is the oldest optimization method proposed by Cauchy in 1847 [34]. The algorithm starts with an initial value $x^{(0)}$ and calculate the gradient vector $\nabla f(x^{(0)})$. The next value $x^{(1)}$ obtained by moving some distance from $x^{(0)}$ in the direction of the descent. In general,

$$x^{(k+1)} = x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)}),$$

where $\alpha^{(k)}$ is the step-size which can be changed at each iteration.
The advantages of this method are that it has often low computation and storage. As an oldest method, the rate of convergence is slower than many other methods. Another limitation of this method is that sometimes it is difficult to calculate the gradient and to choose $\alpha^{(k)}$. However, this method can be optimized using Newton-Raphson algorithm or FISTA [22].

1.4.2 Stochastic algorithms

Stochastic Approximation Expectation-Maximization (SAEM) :

Dempster et al. [48] proposed the Expectation Maximization (EM) algorithm in order to find the maximum likelihood estimator in the context of incomplete data settings also known as mixed effect models. Given a statistical model consisting of an observation X, missing data Z and the parameters θ, the marginal likelihood of the observed data $L(\theta; X)$ is often intractable as it writes as an integral over the distribution of Z. The EM algorithm seeks to find the maximum likelihood estimate (MLE) of the observed likelihood by iteratively applying the following two steps:

1. Expectation step: Calculate the expected value of the log likelihood function given X at the current parameter $\theta^{(t)}$:

$$Q(\theta|\theta^{(t)}) = E_{Z|X,\theta^{(t)}}[\log L(\theta; X, Z)]$$

2. Maximization step: Find the parameter that maximizes this quantity:

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta|\theta^{(t)})$$

Although widely known, it suffers from the need to evaluate the expectation with respect to the posterior distribution of the missing data given the observations. The Stochastic Approximation Expectation-Maximization (SAEM) [47] is an improvement of the EM algorithm with good theoretical properties [47] that deals with this issue. This is an iterative procedure that consists of three steps:

1. Simulation step: Simulate the missing data Z from the conditional distribution $q(Z|X, \theta^{(t)})$.
2. Stochastic approximation step: A stochastic approximation is done on the sufficient statistics of the model using the simulated value of the missing data.

3. Maximization step: Maximize the log-likelihood with respect to the model parameters given this stochastic approximation.

SAEM is a random method which makes it less depend on the initial condition. Its numerical efficiency has also been demonstrated. However, sometimes, we can not simulate the missing data directly and thus resort to some indirect simulation methods, e.g. Markov Chain Monte Carlo methods.

Markov Chain Monte Carlo (MCMC) Method :

MCMC methods are a class of algorithms for sampling from probability distributions. The Metropolis-Hasting algorithm and the Gibbs sampler [60] are two popular and efficient MCMC algorithms for obtaining a sequence of observations when direct sampling is difficult. In particular the Gibbs sampler is well adapted for high dimensional simulation situations. In the simulation step of the SAEM, we use MCMC methods to simulate the missing data.

Metropolis-Hastings algorithm:
In order to reach the stationary distribution $\pi(Z)$ for a random variable Z. The Metropolis-Hastings algorithm starts with a value $Z^{(0)}$. From $Z^{(t)}$, it generates Z^c using the law proposal $q(Z^c|Z^{(t-1)})$ and accept or reject this value Z^c using a procedure of acceptance-rejection. The new value is denoted $Z^{(t+1)}$.

When the number of dimensions is high, it is difficult to find the distribution q to sample efficiently and stride the whole support of π. Gibbs samplers work better in such situations because they simulate the sample for each dimension one by one.

Gibbs sampler:
Suppose that the random variable $Z = (Z_1, ..., Z_d)$ belongs to $\mathbb{R}^d$ with large d. The Gibbs sampler starts with a value $Z^{(0)} = (Z_1^{(0)}, ..., Z_d^{(0)})$. From $Z^{(t)}$, for $i \in [\![1, d]\!]$, it simulates the candidate $Z_i^{(t)}$ from the conditional distribution

$$f_i(Z_i^{(t)}|Z_1^{(t)}, ..., Z_{i-1}^{(t)}, Z_{i+1}^{(t-1)}, ..., Z_d^{(t-1)}).$$

It is necessary to know the conditional distributions f_i to use the Gibbs sampler. Since only the conditional distributions f_i are used in the simulations, all simulations are univariate even for a high dimensional problem. In the Gibbs sampler, the rate of acceptation is always equal to 1.

When f_i is not tractable, one may combine the Metropolis-Hastings and Gibbs sampler. This is known as Metropolis-Hastings within Gibbs sampler.

1.5 Contributions of this work

1.5.1 Generative Statistical Model

In Chapter 2, we propose a model to create a probabilistic atlas, which we define as a probabilistic template together with a quantification of the population geometric variability. Given a set $(y_i)_{1\leqslant i\leqslant n}$ of images observed on a grid of voxels Λ embedded in a continuous domain $D \subset \mathbb{R}^3$, we denote $x_j \in D$ the location of voxel j. We consider that each image is composed of voxels belonging to one class among K, corresponding to K tissues types. We assume that the signal in the K tissue classes is normally distributed with class dependent means $(\mu_k)_{1\leqslant k\leqslant K}$ and variances $(\sigma_k^2)_{1\leqslant k\leqslant K}$ as proposed in [20].

In order to take into account the geometric variability in shape of the brain along a population, we consider that the unknown class of each voxel is supposed to be the discretization on Λ of a random deformation of probability maps $(P_k)_{1\leqslant k\leqslant K}$. These probability maps correspond to the probability of each voxel to belong to each class in the template domain.

We use small deformations in our model, i.e. a deformation φ of the domain D is represented by the displacement of each point x in D by a vector $v(x)$, and is thus written as:

$$\varphi(x) = x + v(x).$$

The template probability maps P_k and the deformation are defined on the whole domain. In order to reduce the dimension to a finite dimension problem, we define both quantities as a finite linear combinations of given kernels centered at some fixed equi-distributed control points in the domain D.

The previous hypothesis provides a generative statistical model for a sample of gray level images. The parameter of our model contains three parts: a photometric part corresponding to the template parametrization, a geometrical part corresponding to the covariance matrix of the deformation parametrization vector and a part for the class corresponding to the class dependent means and variances.

Medical images are usually high-dimensional, but come in small samples, therefore we propose to work in a Bayesian framework.

We consider the maximum a posteriori (MAP) estimator that maximizes the posterior distribution of the parameters $\hat{\theta}$ given the observation. We prove the existence and the consistency of the MAP.

Theorem 1.1 (Existence of the MAP estimation in section 2.3.1)**.** *For any sample* $y_1, ..., y_n$*, there exists* $\hat{\theta}_n \in \Theta$ *such that*

$$q_B(\hat{\theta}_n|y_1, ..., y_n) = \sup_{\theta\in\Theta} q_B(\theta|y_1, ..., y_n).$$

Theorem 1.2 (Consistency in section 2.3.2)**.** *Under weak assumptions, for any compact set* $K \subset \Theta^B$*,*

$$\lim_{n\to+\infty} \pi(\delta(\hat{\theta}_n, \Theta_*) \geqslant \epsilon \wedge \hat{\theta}_n \in K) = 0,$$

where δ *is the metric inherited from the Euclidean metric on* $\mathbb{R}^{n_t}$ *where* n_t *is the dimension of* Θ *and* Θ_* *is the set of optimal parameters.*

In Chapter 4, we generalize the model to multivariate observations. We use multimodal images as the input and we estimate the probability maps for each modality as the result.

1.5.2 Statistical Learning Procedure

In our model, it is impossible to compute the gradient of classical matching energies with respect to the deformation and thus precludes any algorithm based on alternative gradient descents. This led us to use stochastic algorithm. Thanks to the choice of our statistical model, it belongs to the curved exponential family. Therefore we can calculate the sufficient statistics easily. As we are in an incomplete-data setting, we choose the SAEM coupled with a MCMC method to maximize the likelihood. During the simulation step, we simulate the missing data, i.e. the deformation parameters ($\boldsymbol{\beta}$) and the vector of voxel classes using a Metropolis-Hastings algorithm within Gibbs sampler. This particular MCMC method is well adapted for high dimensional simulations and also in our particular case where the spatial distribution of the classes across the brain volume depends on the deformation.

We also prove the almost sure convergence of the previous estimation algorithm towards the MAP estimator given an n-sample of observations.

Theorem 1.3 (Convergence of our estimation algorithm in section 2.4.3)**.** *Under usual assumptions on the model and algorithm, there exists two compact sets* K *and* $\mathcal{K}_0$ *such that for all* $((c)_0, (\boldsymbol{\beta})_0) \in$ K *and* $s_0 \in \mathcal{K}_0$*, we have* $\lim_{m\to\infty} d(s_m, \mathcal{L}) = 0$ $\bar{\mathbb{P}}_{(c)_0,(\boldsymbol{\beta})_0,s_0}$*-a.s, where* $\mathcal{L}$ *is the set of critical points of the observed likelihood and* $\bar{\mathbb{P}}_{(c)_0,(\boldsymbol{\beta})_0,s_0}$ *is the probability measure associated with the chain* $((c)_m, (\boldsymbol{\beta})_m, s_m)_{m\geqslant 0}$ *starting at* $((c)_0, (\boldsymbol{\beta})_0, s_0)$*.*

The MCMC method can also be used to decide which model fits the data best. In Chapter 5, we propose a statistical framework based on a hierarchical modeling to include

the information brought by the peptides shared by different proteins. We use the Gibbs sampler for analyzing large datasets, and compare with a model based on the analysis of one protein at a time, to prove that our method is more reliable for estimating protein abundances and testing abundance changes.

1.5.3 Segmentation of new individuals

As we present in section 1.1.5, one application of the brain atlas is atlas-based segmentation. In Chapter 3, we propose a segmentation method using the atlas that we estimated.

We first define the gray level template $\hat{I}$ using the estimated means and the estimated probability maps. We also use the estimated covariance as a metric for the space of deformations to constrain the registration map according to the learned distribution. Given a target image y, the template $\hat{I}$ is deformed non-rigidly and registered to the target image by minimizing the classical energy $E(\varphi \cdot y, \hat{I}, \varphi)$. It is done by a gradient descent on the deformation variable denoted by β,

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta} \in \mathbb{R}^{3k_g}}{\operatorname{argmin}} E(\varphi_{\boldsymbol{\beta}} \cdot y, \hat{I}, \varphi).$$

Then, the tissue for each voxel denoted by c_j^* is chosen to be the class that maximizes the posterior probability of that voxel to belong to each class, given this deformation field $\boldsymbol{\beta}^*$,

$$c_j^* = \underset{c_j \in [\![1,K]\!]}{\operatorname{argmax}} \left[\log \left(q(y_j|c_j, \hat{\theta}) q(c_j|\boldsymbol{\beta}^*, \hat{\theta}) \right) \right].$$

This shows very interesting results, both quantitatively and qualitatively on synthetic and real data.

Chapter 2

Probabilistic Atlas and Geometric Variability Estimation.

Computerized anatomical atlases play an important role in medical image analysis. While an atlas usually refers to a standard or mean image also called template, that presumably represents well a given population, it is not enough to characterize the observed population in detail. A template image should be learned jointly with the geometric variability of the shapes represented in the observations. These two quantities will in the sequel form the atlas of the corresponding population. The geometric variability is modeled as deformations of the template image so that it fits the observations. In this paper, we provide a detailed analysis of a new generative statistical model based on dense deformable templates that represent several tissue types observed in medical images. Our atlas contains both an estimation of probability maps of each tissue (called class) and a statistical summary of the deformations. We use a stochastic algorithm for the estimation of the probabilistic atlas given a dataset. Experiments are shown on brain T1 MRI datasets.

2.1 Introduction

In neuroimaging, brain atlases are useful for both segmentation and registration tasks as they enable to transport known information to a new patient image to perform qualitative and quantitative comparisons. What is often referred to as an atlas actually corresponds to a mean image or *template*. The problem of estimating such an image given a population has become a central issue in medical imaging during the past decade. Many different methods have been proposed for template estimation (see [121, 124] among others); they work either on gray level images, segmented data or shapes summarized by a set of landmarks. Probabilistic templates, that represent the probability of observing different tissues at each location, are especially attractive [63, 84], as they make it possible to take into account the uncertainty on the underlying tissue type, which is related to partial volume effect (PVE) or to perfectible registration. In many template construction methods, pre-segmentation or pre-registration are required. In this paper, we aim at creating a probabilistic atlas, which we define as a probabilistic template together with a quantification of the population geometric variability. However this estimation requires only a coarse pre-registration and no pre-segmentation.

Atlas learning encompasses the two most fundamental problems in image analysis, namely *segmentation* and *registration*, as these are the basis of template estimation and population analysis. Concerning the segmentation issue, it is important to use automated segmentation for the sake of efficiency and reproducibility. Many different methods have already been proposed for segmentation, such as level set methods [50], model-based segmentation [72], template-based approaches [123, 155] among many others. In many cases, segmentation is coupled with registration. Indeed, performing registration and segmentation jointly is generally more effective than performing them sequentially [68, 106, 149]. An accurate segmentation increases the precision of subsequent registration steps. On the other hand, transporting a segmentation from a template to a subject requires an accurate registration procedure. The accuracy of the registration depends on the class of deformations that are considered. Indeed, one may prefer smooth deformations that capture only the global shape changes rather than local details. On the opposite side, when local shape features are meaningful, one has to adapt the class of deformations so that they deal with variable geometric and regularity feature. Part of this choice has to be made by the user and depends on the data. However, the complexity of the deformation set has also to be constrained by the observations themselves. Some deformation models provide a metric on the space of shapes through a metric on the deformation set [134] which describes geometrically the data. Another viewpoint is to propose a probabilistic approach where the probability distribution of the deformation highlights the characteristic deformations in a population of interest.

Both approaches can actually be related, in particular when considering that, assuming a proper parametrization, the deformations are multivariate normally distributed, as the covariance matrix yields a natural metric to compute distances between deformations [10]. More generally, this amount to using a Mahalanobis distance on deformations, which is an acceptable choice even without a Gaussian hypothesis.

As pointed in [15], estimating this probability distribution together with the template (gray level in [15]) increases the population classification accuracy, as the model better fits the observations. As for the population classification in [15], the segmentation (tissue classification of voxels) takes advantage of the registration to the template. In the sequel, the probabilistic template together with the geometric variability will be called *atlas*.

Several solutions have been proposed previously to deal with one or the other part of atlas or template estimation; we now discuss the closest works to ours. First, a problem with average templates construction is that they do not include the nonlinear deformation to align the corresponding structures. In [83], to solve this problem, a generative model was proposed to create a template using mesh-based representations endowed with a deformation model. This method computes estimates of the deformation field and the most compact mesh representation using an Expectation-Maximization (EM) algorithm. However they require the pre-segmentation of the training image. In [25], a method was proposed to do the segmentation and registration jointly, while creating an average brain template. This approach combines groupwise registration using the Kullback-Leibler divergence and the EM algorithm for segmentation, and thus demonstrates the benefit of their integration. However it does not learn the geometric variability within the estimation procedure, which may reduce the accuracy of the template to match the observations with prior deformations. In [110], a probabilistic model was proposed to segment a heterogeneous data set of brain MR images simultaneously while constructing templates for each mode in the heterogeneous population using an EM algorithm. However, it performs clustering as an additional step, and does not learn the geometric variability of the population. In [15], a model was proposed to create an atlas containing the geometric variability. As the inputs are scalar images, the template is also estimated as a scalar (gray level) image. As a consequence, the segmentation of the population is not part of the estimation process. In [153], a spherical demons algorithm with geometric variability was proposed for registering images and for creating an atlas. The registration was more accurate and this registration could be used to transfer segmentation labels onto a new image. However, the segmentation was not performed during the estimation.

In this paper, we propose to include all the aspects of the atlas estimation procedures described previously, which can improve both the estimated template image, the estimated geometric variability and the segmentation of individual data. Moreover, we propose to perform this estimation using a joint segmentation-registration. For this purpose, we propose to model the observations (gray level images) by a generative statistical model, the parameters of this model being our atlas, i.e. a probabilistic template and the geometric variability. We generalize the model proposed in [20] and use the algorithm in [15] for the estimation part. We also learn the geometry as the metric on the space of deformations, which reduces the possible deformations to those that are common in the population.

This takes the form of a multivariate zero mean normal distribution on the deformations, where the main parameter is the covariance matrix, which is not constrained to have a particular structure (e.g. diagonal or sparse). This captures the long distance correlations of the deformations.

To estimate the model parameters, we use a stochastic algorithm that has demonstrated good performances on real data in [14, 111] and has theoretical convergence properties [15]. We get as final output an estimation of both the probabilistic template and the geometric variability. Although the individual deformation and segmentation are not parameters of the model, the algorithm can be used to return individual deformations and segmentations of the individual images. Additional parameters are also learned by this procedure, such as the means and variances of each tissue of gray level distribution.

As a quantitative evaluation of our method, we test our algorithm on synthetic data for which we know the ground truth. We obtain high Jaccard indices on training and test data. We perform two tests to evaluate our method on real data. At first, we tested our algorithm on 8 patients on an anatomical brain MRI dataset for which a manual segmentation is available as a quantitative segmentation evaluation.

The rest of this paper is organized as follows. In Sections 2.2, 2.3 and 2.4, we present the model, the estimation and the algorithm in detail. Section 2.5 yields the experimental results on simulated and real data. In appendix, we prove the existence of the solution and the convergence of the estimation algorithm.

2.2 The Observation Model

In this section, we present our statistical model, the selected set of deformations and the parametric template that we consider for the sake of computational tractability. Then,

we introduce the parameters of interest and the Bayesian framework, i.e. we introduce priors to address the known issue that medical images datasets most often comprises very few samples.

2.2.1 Statistical Model

We consider here n individual MR images from n patients. This set $(y_i)_{1\leqslant i\leqslant n}$ of images are observed on a grid of voxels Λ embedded in a continuous domain $D \subset \mathbb{R}^3$. We denote $x_j \in D$ the location of voxel $j \in \Lambda$. We consider that each image is composed of voxels belonging to one class among K, corresponding to K tissues types. We assume that the signal in the K tissue classes is normally distributed with class dependent means $(\mu_k)_{1\leqslant k\leqslant K}$ and variances $(\sigma_k^2)_{1\leqslant k\leqslant K}$ as proposed in [20]. Therefore the probability of observing a data with intensity y_i^j for the ith image in the jth voxel given that it belongs to the kth class ($c_i^j = k$) is defined as follows:

$$\mathbb{P}(y_i^j | c_i^j = k, \mu_k, \sigma_k^2) \sim \mathcal{N}(y_i^j; \mu_k, \sigma_k^2), \tag{2.1}$$

where $\mathcal{N}(.; \mu, \sigma^2)$ is the normal density with mean μ and variance σ^2. This expression results from the assumption that given the class, the voxels are assumed to have independent gray level. This assumption is not satisfied in real life experiment as the noise of the observation depends on the tissue type. However, this common assumption is a first approximation that is useful for the sake of estimation.

In order to take into account the geometric variability in shape of the brain along a population, we consider that there exists a random deformation from the template to the subject that acts as follows: the *unobserved* classes of the voxels of the data y are assumed to follow the probability distribution given by the discretization on Λ of the warped probabilistic template. This template is defined by the probability maps $(P_k)_{1\leqslant k\leqslant K}$ that yield the probability of each voxel to belong to each class in the template domain. In other words, the probability maps are deformed to match the observation y (in a sense that will be detailed below) ; then they are discretized on Λ to provide, at each voxel, a voxel-dependent discrete probability measure for this point that gives the probability of each voxel to belong to each class.

As the deformation is not observed (and is actually a mathematical tool for population analysis), we assume that these deformations from the template maps to each subject are also *unobserved and random*. We define them through a random field $z : \mathbb{R}^3 \to \mathbb{R}^3$ such that for $j \in \Lambda$ the prior probability of a voxel j from subject i to be in the k^{th} class

is given by:

$$\mathbb{P}(c_i^j = k) = P_k(x_j - z(x_j))\,. \tag{2.2}$$

We consider here the linearized deformation model which defines a deformation φ of the domain D as the displacement of each point x in D by a vector $v(x)$, and is thus written as:

$$\varphi(x) = x + v(x).$$

As we consider linearized deformations, we approximate φ^{-1} by $\varphi^{-1}(x) = x - v(x)$ at the first order. This makes it possible to apply the deformation to an image, here P_k, as

$$\mathbb{P}(c_i^j = k) = \varphi_i \cdot P_k(x_j) = P_k(x_j - z_i(x_j)).$$

As defined above, the deformation is an infinite dimensional object. While such a dense representation is theoretically sound, for sake of computation, we consider a subspace of deformations that will be parameterized. We assume that the deformation is controlled by the displacement of some given control points belonging to D. This reduces the problem to finite dimension. We define the deformation field as a finite linear combinations of a given kernel K_g centered at some fixed equi-distributed control points in the domain D: $(x_g)_{1\leqslant g\leqslant k_g}$ with parameter $\boldsymbol{\beta} \in (\mathbb{R}^3)^{k_g}$

$$\forall x \in D, z_\beta(x) = (\mathbf{K_g}\boldsymbol{\beta})(x) = \sum_{k=1}^{k_g} K_g(x, x_g)\boldsymbol{\beta}(k)\,, \tag{2.3}$$

where K_g is chosen as a radial Gaussian Kernel in our experiments.

As for the deformation model, the templates $P_k : \mathbb{R}^3 \to [0,1], \forall k \in [\![1, K]\!]$, which are the tissue probability maps, should be defined on the whole domain D. However, in order to reduce their dimensions to allow for numerical computation, we pick a fixed set of control points $(p_l)_{1\leqslant l\leqslant k_p}$ that may be different from the geometric ones and parametrize the templates by the coefficients $\alpha_k \in [0,1]^{k_p}$, which satisfy $\forall l \in [\![1, k_p]\!], \sum_{k=1}^{K} \alpha_k^l = 1$. Then, we write

$$\forall x \in D, P_k(x) = \mathbf{K_p}\alpha_k(x) = \sum_{l=1}^{k_p} K_p(x, p_l)\alpha_k^l, \tag{2.4}$$

where $K_p(x, p_l) = 1$ if p_l is the nearest neighbor of x among the set of points $(p_j)_j$ and 0 otherwise.

Remark 1. The unobserved parameter $\boldsymbol{\beta}$ appears in the indicator function of the kernel. This makes it impossible to compute the gradient of classical matching energies with respect to $\boldsymbol{\beta}$ and thus precludes any algorithm based on alternative gradient descents

(as e.g. in [77, 153]). However, this indicator functions make it possible to deal with the constraint on α which would appear much harder with other smoother kernels and can be handled easily with our estimation algorithm.

The previous hypothesis provides a generative statistical model for a sample of gray level images. The random variables are the deformation vector $\boldsymbol{\beta}$, the class of each voxel c and the parameters that characterize the gray levels of the tissues $(\mu_k, \sigma_k^2)_k$. The probability distributions of the former two elements are given by Equation (2.1) and (4.1). We assume that the deformation vector follows a normal distribution with mean zero and full covariance matrix. The hierarchical model is given by: $\forall i \in [\![1, n]\!],\ \forall j \in \Lambda$

$$\begin{cases} \boldsymbol{\beta}_i \sim \mathcal{N}(0, \Gamma_g) | \Gamma_g; \\ c_i^j \sim \sum_{k=1}^{K} \delta_k P_k(x_j - z_{\boldsymbol{\beta}_i}(x_j)) | \boldsymbol{\beta}_i; \\ y_i^j \sim \mathcal{N}(\mu_k, \sigma_k^2) | c_i^j = k, \mu_k, \sigma_k^2, \end{cases} \tag{2.5}$$

where δ_k is a Dirac measure on k. The covariance matrix Γ_g is not assumed to have any particular pattern of zeros. This makes it possible to model local and global correlations between control point moves, in particular, very correlated displacements can be captured such as translation of a large area of the images. The zero mean is a relevant assumption as the population is assumed to be distributed in an ellipsoid around this mean image.

2.2.2 Parameters and likelihood

Given this statistical model, the parameters to estimate are the covariance matrix Γ_g of the deformation coefficient (Equation (2.3)), $(\alpha_k)_{1 \leqslant k \leqslant K}$ the coefficients that define the templates (Equation (2.4)), $(\mu_k)_{1 \leqslant k \leqslant K}$ and $(\sigma_k^2)_{1 \leqslant k \leqslant K}$ the class dependent means and variances. Let $\theta_g = \Gamma_g$, $\theta_p = ((\alpha_k)_{1 \leqslant k \leqslant K})$ and $\theta_c = ((\mu_k)_{1 \leqslant k \leqslant K}, (\sigma_k^2)_{1 \leqslant k \leqslant K})$. We assume that $\theta = (\theta_g, \theta_p, \theta_c)$ belongs to the parameter space Θ defined as the open set

$$\Theta = \{\theta = ((\alpha_k)_{1 \leqslant k \leqslant K}, (\mu_k)_{1 \leqslant k \leqslant K}, (\sigma_k^2)_{1 \leqslant k \leqslant K}, \Gamma_g) | \alpha_k \in]0, 1[^{k_p}, \sigma_k^2 > 0, \mu_k \in \mathbb{R}, \Gamma_g \in \Sigma^+_{3k_g,*}(\mathbb{R})\} \tag{2.6}$$

Here $\Sigma^+_{3k_g,*}(\mathbb{R})$ is the set of strictly positive symmetric matrices of dimension $3k_g \times 3k_g$.

We can notice that due to the unobserved variables β and c, the observed likelihood is an integral over these random variables. This writes

$$q(y|\theta) = \int \int q(y|c, \theta_c) q(c|\boldsymbol{\beta}, \theta_p) q(\boldsymbol{\beta}|\theta_g) dc d\boldsymbol{\beta} \tag{2.7}$$

where the conditional distributions are given by our model

$$q(\boldsymbol{\beta}|\theta_g) = \exp(-\frac{1}{2}\boldsymbol{\beta}^T\Gamma_g^{-1}\boldsymbol{\beta})(2\pi)^{-\frac{3}{2}kg}|\Gamma_g|^{-\frac{1}{2}} \tag{2.8}$$

$$q(c|\boldsymbol{\beta},\theta_p) = \sum_{k=1}^{K}\delta_k P_k(x_j - z_{\boldsymbol{\beta}_i}(x_j)) \tag{2.9}$$

$$q(y|c,\theta_c) = \prod_{j=1}^{|\Lambda|}(2\pi\sigma_{c_j}^2)^{-1/2}\exp\left(-\frac{(y^j-\mu_{c_j})^2}{2\sigma_{c_j}^2}\right) \tag{2.10}$$

where $|\Lambda|$ is the number of voxels.
For sake of simplicity, all the likelihood functions will be denoted by q and the variables specified as arguments of this function q.

2.2.3 Bayesian Model

Medical images are typically high-dimensional, but usually come in small samples. To deal with the data scarcity issue, we choose to regularize the statistical model and we propose to work in a Bayesian framework. As presented in [10], we use standard conjugate priors for each parameter, i.e. an inverse-Wishart ν_g in dimension $3kg \times 3kg$ on Γ_g, a Gaussian ν_m on μ_k and inverse-Wishart ν_p in dimension 1 on σ_k^2 with fixed hyper-parameters. All priors are assumed independent. These priors makes it possible to regularize when needed the estimated parameters but, when the number of observations increases, the relative prior weight decreases.

More formally we have

$$(\Gamma_g, \mu_k, \sigma_k^2) \sim \nu_g \otimes \nu_m \otimes \nu_p;$$

where

$$\begin{cases} v_g(\Gamma_g) \propto \left(\exp(-\frac{1}{2}\langle\Gamma_g^{-1},\Gamma_g^0\rangle)\frac{1}{\sqrt{|\Gamma_g|}}\right)^{a_g} d\Gamma_g, a_g \geqslant 6k_g+1, \\ v_m(\mu_k) \propto \exp\left(-\frac{(\mu_k-m_u)^2}{2\sigma_\mu^2}\right) d\mu_k \\ v_p(\sigma_k^2) \propto \left(\exp\left(-\frac{\sigma_0^2}{2\sigma_k^2}\right)\frac{1}{\sqrt{\sigma_k^2}}\right)^{a_p} d\sigma_k^2, a_p \geqslant 3. \end{cases}$$

Note that for two matrices A, B we have $\langle A, B\rangle = tr(A^T B)$ the Frobenius inner product on matrices.
Our Bayesian model can be represented by Fig. 2.1 where the dependencies are highlighted.

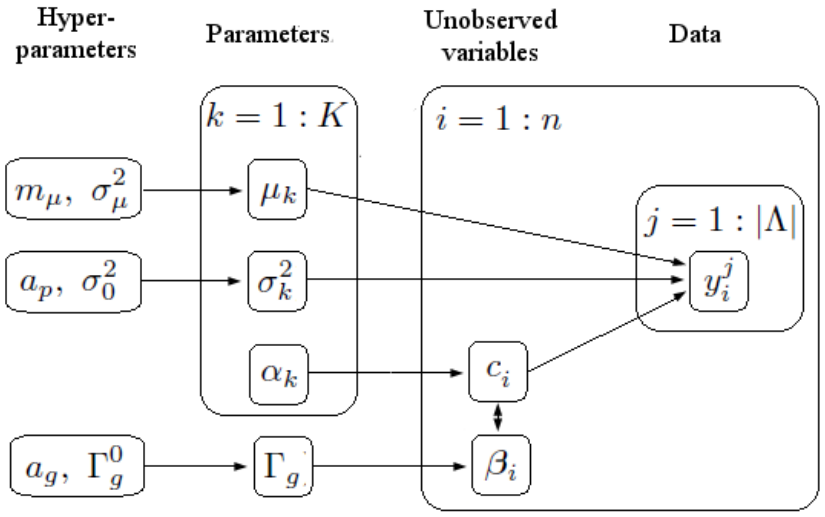


FIGURE 2.1: Generative model that relates the atlas to the observed images.

2.3 Estimation

2.3.1 Existence of the MAP estimation

Given the complete statistical model, we can learn the parameters that best fit the observations. Although real data never follow any parametric model, we try to approximate their generation so that we better understand the common and specific features of a given population. For this purpose, we consider the maximum a posteriori (MAP) estimator: $\hat{\theta}_n = \arg\max_{\theta\in\Theta} q_B(\theta|y_1, ..., y_n)$ where q_B denotes the posterior distribution of the parameters given the n observations $y_1, ..., y_n$.
The following theorem proves here that given a n sample of observations, the maximum a posteriori estimator exists at finite distance in the parameter space.

Theorem 2.1 (Existence of the MAP estimation)**.** *For any sample $y_1, ..., y_n$, there exists $\hat{\theta}_n \in \Theta$ such that*

$$q_B(\hat{\theta}_n|y_1, ..., y_n) = \sup_{\theta\in\Theta} q_B(\theta|y_1, ..., y_n).$$

The proof is postponed to appendix.

Remark 2. Note that one could rely on the prior distribution to prove this property for the means $(\mu_k)_{1\leqslant k\leqslant K}$. However, as we are dealing with a Bayesian model, we introduce priors on all parameter to keep the coherence of the model. Nonetheless, it would be possible to remove the prior on these parameters thanks to the proof in the appendix. Concerning the priors, for the covariance matrix, the prior is informative as we choose the usual kernel matrix used for registration issues. The prior on the means are non-informative as the gray level of the observations change drastically when the acquisition protocols change.

2.3.2 Consistency of the estimator on our model

We are interested now in the consistency property of the MAP estimator without making strong assumptions on the distribution of the observations $y_1, ..., y_n$. This means that we do not assume that the observations are generated by the model described above. We denote the distribution governing the observations by π and seek to prove the convergence of the MAP estimator to the set Θ_* of model distributions *closest* to π:

$$\Theta_* = \{\theta_* \in \Theta | E_\pi(\log q(y|\theta_*)) = \sup_{\theta \in \Theta} E_\pi(\log q(y|\theta))\}$$

However, this consistency only holds for bounded variances $\forall\ c \in [\![1, K]\!]$, $\sigma_c^2 > \sigma_{min}^2$. This assumption on the admissible set is not restrictive as we have proven that the MAP estimator exists out of the boundaries. Let

$$\Theta^B = \{\theta = ((\alpha_k)_{1\leqslant k\leqslant K}, (\mu_k)_{1\leqslant k\leqslant K}, (\sigma_k^2)_{1\leqslant k\leqslant K}, \Gamma_g) | \alpha_k \in]0,1[^{k_p}, \sigma_k^2 > \sigma_{min}^2, \mu_k \in \mathbb{R}, \Gamma_g \in \Sigma^+_{3k_g,*}(\mathbb{R})\} \tag{2.11}$$

$$\Theta_*^B = \{\theta_* \in \Theta^B | E_\pi(\log q(y|\theta_*)) = \sup_{\theta \in \Theta^B} E_\pi(\log q(y|\theta))\}$$

Theorem 2.2 (Consistency)**.** *Assume that Θ_*^B is non empty. Then for any compact set $K \subset \Theta^B$,*

$$\lim_{n\to+\infty} \pi(\delta(\hat{\theta}_n, \Theta_*^B) \geqslant \epsilon \wedge \hat{\theta}_n \in K) = 0,$$

where δ is the metric inherited from the Euclidean metric on $\mathbb{R}^{n_t}$ where n_t is the dimension of Θ.

Proof. The theorem is an application of Wald's consistency Theorem in [136]. We only need to verify that $y \to \log q(y|\theta)$ is π *a.s.* upper semi-continuous and that for any $\theta \in \Theta$, there exists an open set $U \ni \theta$ such that $E_\pi(\sup_{\theta' \in U} \log^+(q(y|\theta'))) < \infty$ (where $\log^+$ is the positive part of log). In our setting, for any $\theta = (\alpha_k, \mu_k, \sigma_k^2, \Gamma_g) \in \Theta^B$, we denote $U = \{]0,1[^{k_p}, \mathbb{R},]\sigma_{min}^2, +\infty[, \Sigma^+_{3k_g,*}(\mathbb{R})\}$, so that

$$\begin{aligned}
\sup_{\theta' \in U} \log(q(y|\theta')) &\leqslant \sup_{\theta' \in U} \log(\sum_{k=1}^K q(y|c, \theta')) \\
&\leqslant \sup_{\theta' \in U} \log \left(\sum_{k=1}^K \left(\frac{1}{\sqrt{2\pi\sigma_k^2}} \right)^{n|\Lambda|} \right) \\
&\leqslant \log \left(K \left(\frac{1}{\sqrt{2\pi\sigma_{min}^2}} \right)^{n|\Lambda|} \right) < \infty \quad ,
\end{aligned}$$

where σ^2_{min} is the lower bound of σ^2_k. □

Remark 3. Note that we only proved a limited consistency result as we have no guaranty that Θ^B_* is not empty. However, looking at our model, generalized from the Bayesian Mixed Effect Templates introduced in [10], we expect the same result. Future work will generalize the proof of consistency in [10] to the present model. We focus here on the convergence property of the estimation algorithm that we present in the following section.

2.4 Estimation Algorithm using Stochastic Approximation Expectation-Maximization

We now present the estimation algorithm that we use to reach the MAP estimate of the parameters. We assume now that we observe a fixed number n of gray level images taken from a homogeneous population.

2.4.1 Model factorization

Despite the complex dependencies of the random variables in our statistical model, it belongs to the curved exponential family. That is to say, the complete likelihood q writes as:

$$q(y, c, \boldsymbol{\beta}, \theta) = \exp[-\psi(\theta) + \langle S(c, \boldsymbol{\beta}), \phi(\theta)\rangle] = \exp(L(\theta, S)) \tag{2.12}$$

where ψ, ϕ are two Borel functions depending on the parameters, $S(c, \boldsymbol{\beta})$ is a vector of sufficient statistics and the scalar product is the usual Euclidean one. For sake of simplicity, we have omitted the dependency with respect to the observations that are handled as a fixed input to the estimation process. The function L is called the complete log-likelihood.

Thanks to Equation (2.12), we can show that the following matrix-valued functions are the sufficient statistics of the model: $\forall k \in [\![1, K]\!]$, $\forall l \in [\![1, k_p]\!]$,

$$\begin{aligned}
S_{0,k}(c, \boldsymbol{\beta}) &= \sum_{i=1}^{n}\sum_{j=1}^{|\Lambda|} \mathbb{1}_{c_i^j = k}, & S_{1,k}(c, \boldsymbol{\beta}) &= \sum_{i=1}^{n}\sum_{j=1}^{|\Lambda|} \mathbb{1}_{c_i^j = k} y_i^j, \\
S_{2,k}(c, \boldsymbol{\beta}) &= \sum_{i=1}^{n}\sum_{j=1}^{|\Lambda|} \mathbb{1}_{c_i^j = k} (y_i^j)^2, & S_3(c, \boldsymbol{\beta}) &= \sum_{i=1}^{n} \boldsymbol{\beta}_i \boldsymbol{\beta}_i^T, \\
S_{4,k,l}(c, \boldsymbol{\beta}) &= \sum_{i=1}^{n}\sum_{j=1}^{|\Lambda|} \mathbb{1}_{c_i^j = k} \mathbb{1}_{\|x_j - x_l - K_g \boldsymbol{\beta}_i(x_j)\|_1 < \frac{1}{2}}.
\end{aligned} \tag{2.13}$$

We denote $S(c,\boldsymbol{\beta}) = (S_{0,k}(c,\boldsymbol{\beta}), S_{1,k}(c,\boldsymbol{\beta}), S_{2,k}(c,\boldsymbol{\beta}), S_3(c,\boldsymbol{\beta}), S_{4,k,l}(c,\boldsymbol{\beta}))_{\forall k\in[\![1,K]\!],\ \forall l\in[\![1,k_p]\!]}$ the vector of sufficient statistics and define the sufficient statistic space as

$$\begin{aligned}\mathcal{S} = \{(S_{0,k}, S_{1,k}, S_{2,k}, S_3, S_{4,k,l}) \mid \forall k \in [\![1,K]\!],\ \forall l \in [\![1,k_p]\!],\\ S_{0,k} \in \mathbb{R}_*^+, S_{1,k} \in \mathbb{R}_*^+, S_{2,k} \in \mathbb{R}_*^+, S_3 + a_g\Gamma_g^0 \in \mathrm{Sym}_{3kg}^+, S_{4,k,l} \in R_*^+)\}. \quad (2.14)\end{aligned}$$

The set $\mathcal{S}$ can be viewed as an open set of $\mathbb{R}^s$ with $s = K + K + K + \frac{3kg(3kg+1)}{2} + Kk_p$.

Remark 4. Note that the sufficient statistics $S_{0,k}$, $S_{1,k}$, $S_{2,k}$ and $S_{4,k}$ cannot vanish. Indeed, if for one class k_0, $S_{0,k_0} = 0$, this particular class would be empty which means that there are no voxel belonging to this tissue class. We assume that we actually know the number of expected tissues in the gray level images so that this never happens.

The second property of our model is that there exists $\hat{\theta}$ such as $\max_{\theta\in\Theta} L(\theta, S) = \hat{\theta}(S)$. Indeed μ_k, σ_k^2, α_l^k and Γ_g are explicitly expressed with the above sufficient statistics as follows: $\forall k \in [\![1,K]\!]$, $\forall l \in [\![1,k_p]\!]$,

$$\begin{aligned}\hat{\mu}_k(S) &= \frac{S_{1,k}(c,\boldsymbol{\beta})}{S_{0,k}(c,\boldsymbol{\beta})}, & \hat{\sigma}_k^2(S) &= \frac{1}{n+a_p}\left(n\left(\frac{S_{2,k}(c,\boldsymbol{\beta})}{S_{0,k}(c,\boldsymbol{\beta})} - \frac{S_{1,k}(c,\boldsymbol{\beta})^2}{S_{0,k}(c,\boldsymbol{\beta})^2}\right) + a_p\sigma_0^2\right),\\ \hat{\Gamma}_g(S) &= \frac{S_3(c,\boldsymbol{\beta}) + a_g\Gamma_g^0}{n * |\Lambda| + a_g}, & \hat{\alpha}_k^l(S) &= \frac{S_{4,k,l}(c,\boldsymbol{\beta})}{\sum_{k'=1}^{K} S_{4,k',l}(c,\boldsymbol{\beta})}.\end{aligned} \quad (2.15)$$

These equations are well defined thanks to Remark 4. This also justifies the fact that the coefficients α_k^l belongs to $]0,1[$ for all $1 \leqslant k \leqslant K$ and for all $1 \leqslant l \leqslant k_p$.

2.4.2 Estimation Algorithm

As we are in an incomplete-data setting, a natural way to maximize a likelihood is to use the EM algorithm or an algorithm derived from EM. We choose the Stochastic Approximation EM (SAEM) coupled with a Markov Chain Monte Carlo (MCMC) method thanks to its good theoretical [15] and numerical [14] performances in such settings.

We detail here the $m+1^{th}$ iteration of the SAEM-MCMC algorithm which consists of three steps:

Step 1: Simulation step. The missing data, i.e. the deformation parameters $(\boldsymbol{\beta}) = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_n)$ and the vector of classes $(c) = (c_1, \cdots, c_n)$, are drawn using the

transition probability of a ergodic Markov chain Π_θ having the posterior distribution $q_{post}(\cdot|y,\theta)$ as its stationary distribution:

$$((\boldsymbol{\beta})_{m+1},(c)_{m+1}) \sim \Pi_{\theta_m}(((\boldsymbol{\beta})_m,(c)_m),\cdot)$$

where we choose Π_θ to be a Metropolis-Hastings within Gibbs sampler. This particular MCMC method is well adapted for high dimensional simulation and also in our particular case where the distribution of the class depends on the deformation. The Gibbs sampler works coordinate by coordinate. Since we cannot sample from the posterior distribution of one coordinate of the vector $((\boldsymbol{\beta}),(c))$ given the others, we use a Metropolis-Hastings step inside these loops. Therefore, we simulate the coordinates one by one. We choose as the proposal of the Metropolis-Hastings method to use the probability distribution of this coordinate given the others coming from the model distributions 2.5. This way, one can estimate deformations that improve the segmentation and segmentations that improve the registration. With this choice, it is easy to calculate the acceptance rates (see in Algorithm 1). Note that it would be possible to choose others priors, however paying attention to the computational cost of the acceptance rates. Since we have a couple of missing data, we first simulate each coordinate of $(\boldsymbol{\beta})$ knowing others coordinates of $(\boldsymbol{\beta})$ and (c), then simulate each coordinate of (c) knowing others coordinates of (c) and the new $(\boldsymbol{\beta})$. The detailed steps of the whole algorithm is given in Algorithm 1 in particular, the hybrid Gibbs sampler steps are precise.

Step 2: Stochastic approximation step. A stochastic approximation is done on the sufficient statistics using the simulated value of the missing data:

$$s_{m+1} = s_m + \Delta_m[S((c)_{m+1},(\boldsymbol{\beta})_{m+1}) - s_m]$$

where $\Delta = (\Delta_m)_m$ is a decreasing sequence of positive step-sizes.

Step 3: Maximization step. The parameters are updated using the previous formula (2.15) where the sufficient statistics are replaced by their stochastic approximations.

$$\theta_{m+1} = \arg\min_{\theta\in\Theta} \hat{\theta}(s_{m+1}).$$

The initial values $(\boldsymbol{\beta})_0$, $(c)_0$, s_0 and θ_0 are arbitrarily chosen (see Algorithm 1).

Algorithm 1 SAEM-MCMC Algorithm (with no reprojection)

Require: $c = (c)_0$, $\boldsymbol{\beta} = (\boldsymbol{\beta})_0$, θ_0, s_0, Δ

`Stochastic Approximation Expectation-Maximization`

for $m = 0$ to iters **do**

`Simulation step using Gibbs sampler:`

for $i = 1$ to n **do**

for $p = 1$ to $3k_g$ **do**

`Metropolis-Hastings procedure`

$$b \sim \mathcal{N}\left(\frac{\sum_{q \neq p} R_{p,q}\boldsymbol{\beta}_i^q}{R_{p,p}}, \frac{1}{R_{p,p}}\right)$$

`Compute` $r_p(\boldsymbol{\beta}_i^p, b; \boldsymbol{\beta}_i^{-p}, c, \theta_m) = \left[\frac{q(c|\boldsymbol{\beta}_{i,b\to p})}{q(c|\boldsymbol{\beta}_i)} \wedge 1\right]$

`With probability` $r_p(\boldsymbol{\beta}_i^p, b; \boldsymbol{\beta}_i^{-p}, c, \theta_m)$, `update` $\boldsymbol{\beta}_i^p : \boldsymbol{\beta}_i^p \leftarrow b$

end for

`Update` $\boldsymbol{\beta}_{i,m+1} \leftarrow \boldsymbol{\beta}_i$

for $j = 1$ to $|\Lambda|$ **do**

$$C \sim \sum_{k=1}^{K} \delta_k P_k(x_j - z_{\boldsymbol{\beta}_{i,k+1}}(x_j))$$

`Compute` $r_j(c_i^j, C; c_i^{-j}, \theta_m) = \left[\frac{q(y|c_{i,C\to j}, \theta_m)}{q(y|c_i, \theta_m)} \wedge 1\right]$

`With probability` $r_j(c_i^j, C; c_i^{-j}, \theta_m)$, `update` $c_i^j : c_i^j \leftarrow C$

end for

`Update` $c_{i,m+1} \leftarrow c_i$

end for

`Stochastic approximation step:`

$s_{m+1} = s_m + \Delta_m[\mathcal{S}((c)_{m+1}, (\boldsymbol{\beta})_{m+1}) - s_m]$

`Maximization step:`

$\theta_{m+1} = \arg\min_{\theta\in\Theta} \hat{\theta}(s_{m+1})$.

end for

2.4.3 Convergence analysis

We prove the almost sure convergence of the previous estimation algorithm towards the MAP estimator given a n-sample of observations. This proof requires to add an intermediate step in the estimation algorithm. This consists in projecting the sufficient statistics on increasing compact subsets when the stochastic approximation reaches a too large value. We refer to [17] for more details about this usual additional step. Note that in practice, no projection has been required in our experiments.

Let us first define some quantities that are required in the following Theorem.

Definition 1. Let $\mathcal{S}$ be the open subset of $\mathbb{R}^s$ defined by Equation (2.14). We define the mean field $h : \mathcal{S} \to \mathbb{R}^s$ as $h(s) = \int\int_{\mathbb{R}^{3k_g}} H_s(c, \boldsymbol{\beta}) q_{post}(c, \boldsymbol{\beta}|\mathbf{y}, \hat{\theta}(s)) dc d\boldsymbol{\beta}$ where $H_s(c, \boldsymbol{\beta}) = S(c, \boldsymbol{\beta}) - s$. Let also $w : \mathcal{S} \to [0, \infty[$, $w(s) = -\mathtt{l}(\hat{\theta}(s))$ be the corresponding Lyapunov function where $\mathtt{l}$ is the incomplete data log-likelihood: $\mathtt{l}(\theta) = \log\int\int_{\mathbb{R}^{3k_g}} q(y, c, \boldsymbol{\beta}, \theta) dc d\boldsymbol{\beta}$. Let $\mathcal{L} \triangleq \{s \in \mathcal{S}, \langle \nabla w(s), h(s)\rangle = 0\}$ be the set of critical points of the observed likelihood.

Theorem 2.3 (Convergence of our estimation algorithm for model 2.5)**.** *Assume that there exists an* $M_0 > 0$ *such that* $\mathcal{L} \subset \{s \in \mathcal{S}, w(s) < M_0\}$. *Assume also that the sequences* $\boldsymbol{\Delta} = (\Delta_m)_{m\geqslant 0}$ *and* $\boldsymbol{\varepsilon} = (\varepsilon_m)_{m\geqslant 0}$ *are non-increasing, positive and satisfy:* $\sum_{m=0}^{\infty} \Delta_m = \infty$, $\lim_{m\to\infty} \varepsilon_m = 0$ *and* $\sum_{m=1}^{\infty} \{\Delta_m^2 + \Delta_m \varepsilon_m^a + (\Delta_m \varepsilon_m^{-1})^p\} < \infty$, *where* $a \in]0,1[$ *and* $p \geqslant 1$. *Then there exists a compact set* $\mathrm{K} \subset \mathcal{Z}$ *where* $\mathcal{Z} = [\![1, K]\!]^{|\Lambda|} \times \mathbb{R}^{3k_g}$ *and there exists another compact subset* $\mathcal{K}_0 \subset \mathcal{W}_{M_0} \triangleq \{s \in \mathcal{S},\ w(s) \leqslant M_0\}$ *such that for all* $((c)_0, (\boldsymbol{\beta})_0) \in \mathrm{K}$ *and* $s_0 \in \mathcal{K}_0$, *we have* $\lim_{m\to\infty} d(s_m, \mathcal{L}) = 0$ $\bar{\mathbb{P}}_{(c)_0,(\boldsymbol{\beta})_0,s_0}$*-a.s, where* $\bar{\mathbb{P}}_{(c)_0,(\boldsymbol{\beta})_0,s_0}$ *is the probability measure associated with the chain* $((c)_m, (\boldsymbol{\beta})_m, s_m)_{m\geqslant 0}$ *starting at* $((c)_0, (\boldsymbol{\beta})_0, s_0)$.

We prove that the stochastic approximation sequence generated by our model and algorithm satisfies Assumptions (**A1'**), (**A2**) and (**A3'**) defined in [15]. The proof is postponed to appendix.

2.5 Experiments and Results

We first test our algorithm on simulated data to check that it reaches our objectives, that is to say (1) recover the template image as probability maps, (2) estimate a relevant covariance matrix of deformations, (3) achieve a good estimate of the mean and variance of each class and (4) segment the observations. Then we test on real data and compare with the segmentations provided by SPM8 [7] , FAST [158] in FSL [1] and DARTEL [19] algorithms. The segmentation method in SPM8 can be used for bias correction, spatially normalizing or segmenting the data, it uses the same model as in [20]. FAST segments a 3D image of the brain into different tissue types. The underlying method is based on a hidden Markov Random Field model and an associated EM algorithm. DARTEL is an algorithm for diffeomorphic image registration that registers images by computing a flow field, which can be exponentiated to generate both forward and backward deformations.

As the SAEM algorithm is an iterative procedure, we run until 250 iterations which reaches numerical convergence. We control the convergence visually on the template and numerically looking at the convergence curve of the variances. For the initialization of our algorithm, we choose $(\boldsymbol{\beta})_0 = 0$ and the initial random classification $(c)_0$.

2.5.1 Simulated data

In the simulated data experiment, a $24 \times 24 \times 3$ image of 4 classes is used as the reference image where the values of each class are $\{1, 2, 3, 4\}$. We generate 20 images

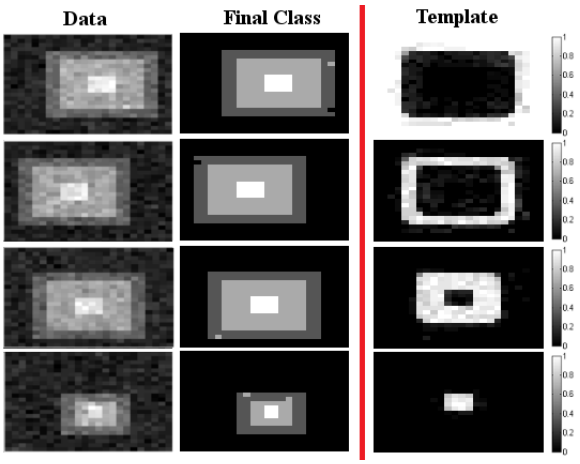


FIGURE 2.2: Experiments on simulated data. The first two columns correspond to one slice of four data images and their final segmentations. The last column correspond to the first slice of the probabilistic template, each row corresponds to a class. The white color corresponds to high probability and the black color to low one.

with translations and zooms and add an independent Gaussian noise with zero mean and standard deviation 0.2 to the deformed image.

We take 64 fixed control points for the deformation model given in Equation (2.3), i.e. one control point in each $3 \times 3 \times 3$ cube and all the points in the image as the control points for the template model given in Equation (2.4) to obtain a complete probabilistic atlas. We choose $(0.3 \times 12)^2$ as the parameter of K_g, where 0.3 is the value that gives the best visual result as in [10] and 12 is a half of the largest dimension size. For the hyper-parameters, we use $a_g = 0.5$, $\Gamma_g^0 = Id$, $a_p = 0.1$ and $\sigma_0^2 = 1$. The values of a_g and a_p are especially small in practice despite the constraints of theoretical definition of the priors. Note that in Equation (2.15), they weight the priors against the data-derived terms weighted by the number of observations. Small values down-weight the priors and increase the importance of data in the estimates. Although the prior laws on Γ_g^0 and σ_0^2 are improper, the posterior laws are well defined. The main purpose of the regularization is to make Γ_g symmetric positive definite. Although it would be possible to use an informative prior on Γ_g^0 as the one used in [10], the results are similar as long as the deformations are well captured.

The results are shown in Fig. 2.2. In the first column, each row corresponds to one slice of three exemplars of the dataset. The final estimated segmentation for each individual is shown in the second column. The most important aspect is that we get the probabilistic template in the third column, each row corresponds to one class. Each voxel belongs to one class with high probability (white) and low probability (black).

	Class 1	Class 2	Class 3	Class 4
Training data	99.8%	98.6%	99.2%	99.4%

TABLE 2.1: Experiments on synthetic data. The Jaccard Index for the training data average across 20 subjects.

Our probabilistic maps are sharp. Most voxels in each class have a probability larger than 0.9. Only few voxels on the boundary of two classes have a non zero probability to belong to two classes. This particularly sharp template demonstrates that the deformations and segmentations have both been well captured through the simulation process. The other parameters are also well estimated upholding the theoretical convergence of our algorithm. The fourth exemplar has a large deformation compared to the others. Thanks to the coupled classification-registration, we can see that our algorithm manages to capture this large deformation and yields the corresponding classification. We calculate the Jaccard index for each class (Table 2.1) which demonstrates that the segmentation done during the atlas estimation is accurate.

In our model, we have a high dimensional parameter Γ_g that is associated with the atlas estimation and imposes to increase the number of observations to get an accurate estimate. To see whether its estimation improves the results, we also run our algorithm without estimating Γ_g. To compare different situations, we fixed different values of Γ_g, $\Gamma_g = 0.5Id,\ Id,\ 2Id$ and $4Id$. We show the estimated probabilistic templates for different values of Γ_g in Fig. 2.3. Each column corresponds to $\Gamma_g = 0.5Id,\ Id,\ 2Id$ and $4Id$, each row corresponds to one class. The voxel belongs to one class with high probability (white) and low probability (black). We can see that the shape of the template do not fit the data as well for $\Gamma_g = Id$ and $4Id$. It seems that we get a better template for $\Gamma_g = 0.5Id$. Compared to our estimated probabilistic template in Fig. 2.2, our maps are sharper than that obtained with any fixed Γ_g and the shape of the template fits better the data.

The segmentation results with these fixed Γ_g are shown in Fig. 2.4. In the first column, each row corresponds to one slice of one exemplar of the dataset. The second to the fifth columns show the final estimated segmentation for each individual for $\Gamma_g = 0.5Id,\ Id,\ 2Id$ and $4Id$. The main problems appear on the fourth exemplar that has a large deformation. None of these values of Γ_g manages to segment this observation well. Despite it simplicity, this example shows the importance of constraining the deformations to relevant ones with respect to the population. This result confirms the classification performances presented in [15] in similar context.

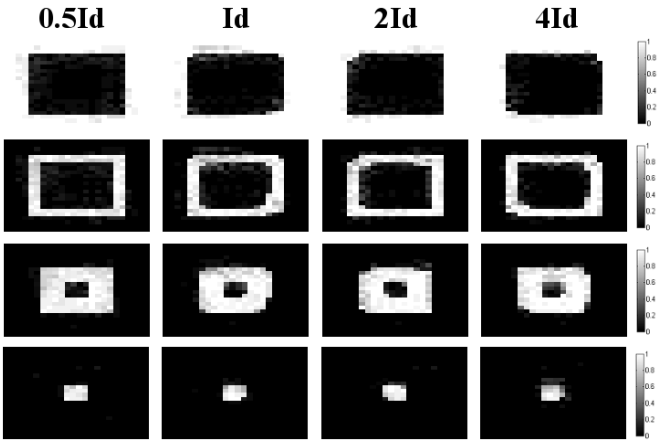


FIGURE 2.3: The estimated probabilistic templates for the model with fixed Γ_g. Each column corresponds to $\Gamma_g = 0.5Id,\ Id,\ 2Id$ and $4Id$, each row corresponds to one class. The voxel belongs to one class with high probability (white) and low probability (black).

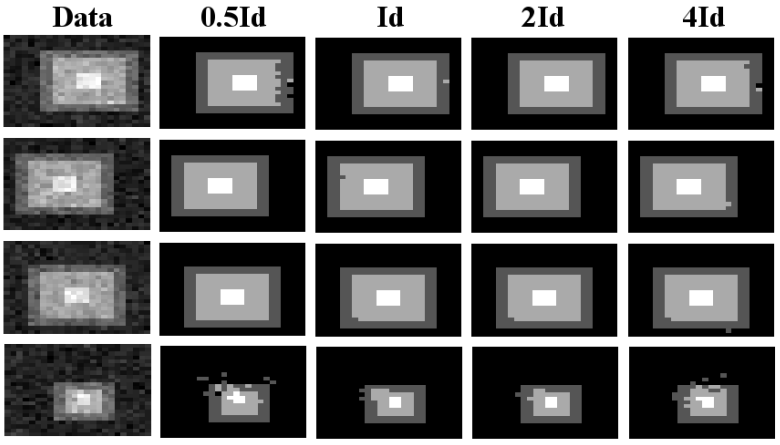


FIGURE 2.4: The segementation results for the model with fixed Γ_g. In the first column, each row corresponds to one slice of one exemplar of the dataset. The second to the fifth columns show the final estimated segmentation for each individual for $\Gamma_g = 0.5Id,\ Id,\ 2Id$ and $4Id$.

As a quantitative evaluation, we calculate the Jaccard index for each class for different values of Γ_g (Table 2.2). We can see that we get a better value of Jaccard index for class 1 and class 2 when the value of Γ_g increase. However, for class 3 and class 4, the value of Jaccard index increases first and then decreases when the value of Γ_g becomes too large. By considering both the template and the Jaccard index, we get a better template for $\Gamma_g = 0.5Id$, however we get a poor Jaccard index. From the first row of Table 2.1 and Table 2.2, we can see that our model always gets a better Jaccard index than the model with fixed values of Γ_g.

In summary, our model always gets a better result than the model with fixed values of

Γ_g	Class 1	Class 2	Class 3	Class 4
0.5Id	99.4%	96.3%	98.0%	96.6%
Id	99.6%	97.2%	98.3%	97.2%
2Id	99.6%	97.7%	**98.9%**	**98.8%**
4Id	**99.7%**	**98.1%**	98.8%	97.2%

TABLE 2.2: Experiments on synthetic data. The Jaccard Index for the training data using the model with fixed Γ_g. Compared with the first row of Table 2.1, our model with estimated Γ_g gets the best result.

Γ_g. However, it would be difficult to choose the optimal value of Γ_g if we want to fix it. Furthermore, we do not know the dependence between the motion of control points, which may be very complex. Therefore, this justifies to estimate the high dimensional parameter Γ_g in our model. Another argument is that the atlas estimation has to be performed only once in each population. Therefore, it could be interesting to spend some time to get an accurate estimation so that the following tasks based on these parameters reach a better performance.

2.5.2 Real data

The proposed method was also tested on real MRI data, derived from manual annotations that are publicly available at the Internet Brain Segmentation Repository (IBSR) [3]. Eight images are available. Each image has a size of $160 \times 160 \times 128$ with resolution $0.9735 \times 0.9735 \times 1.5\ mm^3$. The images were considered to have 3 tissue classes: gray matter (GM), white matter (WM) and CSF+background. Each tissue class follows a Gaussian distribution. The variances are class dependent rather than homogeneous.

We take 800 fixed control points for the deformation model given in Equation (2.3), corresponding to one control points in each $16 \times 16 \times 16$ cube and $80 \times 80 \times 64$ points in the image as the control points for the template model given in Equation (2.4), corresponding to one control point in each $2 \times 2 \times 2$ cube. We choose $(0.3 \times 80)^2$ as the parameter of K_g as for synthetic images. For the hyper parameters, we choose $a_g = 0.5$, $\Gamma_g^0 = Id$, $a_p = 0.1$ and $\sigma_0^2 = 1$ for the same reasons as above. For comparison purpose, we always present the same image slice z=10 for all methods in these experiments.

In the first experiment, we run our algorithm with 8 patient images as training data. These images are provided with their segmentation, allowing for the validation of our online segmentation of the training images.

At first, we compare our estimated template with DARTEL template that uses SPM's segmentation as input and the average template without deformation. The first three

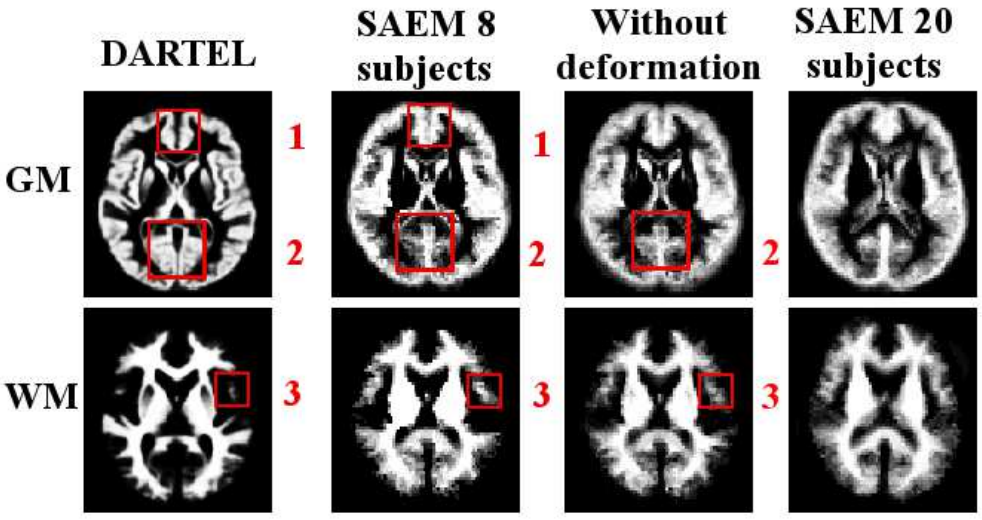


FIGURE 2.5: The template obtained by DARTEL (first column), our method with 8 subjects (second column), the average template without deformation (third column) and 20 subjects (fourth column). GM for the first row and WM for the second. DARTEL template shows a wide line of CSF in zone 1 and a large pattern of the V1 in zone 2 which are thin in the data images. Our probabilistic maps capture the presence of WM in zone 3, however DARTEL only detects a small region. The shape of the V1 in zone 2 for our template fits better the data than the one for the average template. The presence of WM in zone 3 for the average template is fuzzy. Our estimated template with 8 subjects is good. With 20 subjects, we get more details on the boundary of two types.

columns in Fig. 2.5 show one slice of DARTEL template (first column), our probabilistic template with 8 subjects (second column) and the average template without deformation (third column), the first and second rows correspond to GM and WM respectively. Because of the smoothing step that creates regular contours, the DARTEL template is smoother than our SAEM template. Moreover, the anatomical prior template used in DARTEL makes the output very contrasted (almost binary). For DARTEL template, zone 1 shows wide CSF digitations and zone 2 shows large primary visual cortex (V1) pattern, which are much thinner in the data. Also notice that DARTEL requires a pre-segmentation of the data and does not provide the geometric distribution of the population. Our model only takes into account the training data and is thus free from these biases. The weakly contrasted template may also be an advantage as it explains the uncertainty on voxels coming from both the PVE and the registration level of details. Another bias is shown in zone 3 where our probabilistic maps capture the presence of the cerebral white matter. Thanks to the deformation estimated along the atlas estimation, the presence of the cerebral white matter in zone 3 is sharper in our template than the one in the average template and the shape of the V1 area in zone 2 better fits the data (first column in Fig. 2.6). The demonstrates that incorporating the deformation metric improves the atlas estimation.

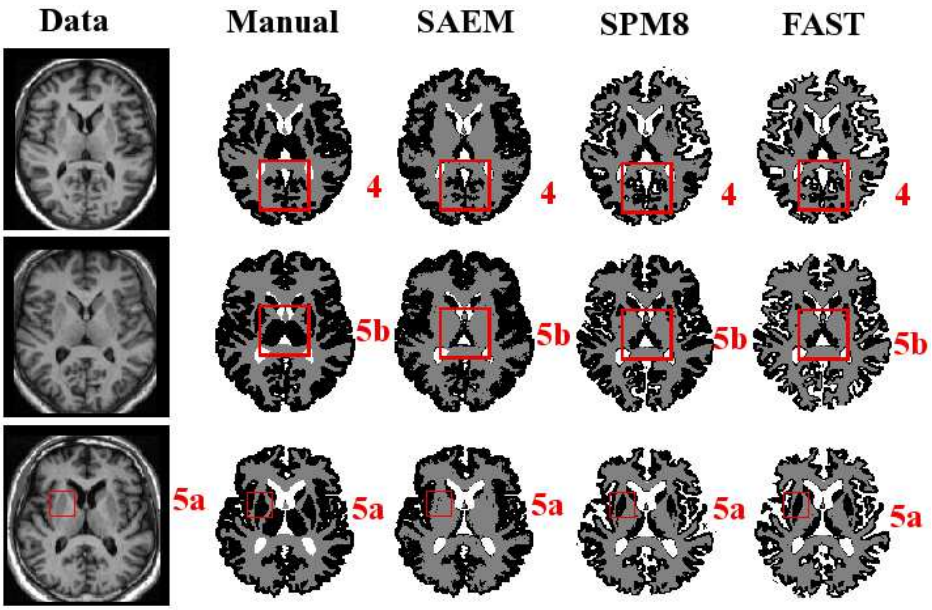


FIGURE 2.6: Experiments on real data. Each column corresponds to the one slice of 3 data images, the manual segmentation and the segmentation obtained by our method, SPM8 and FAST. Our methods shows each fold of the V1 (zone 1). Our method does not manage to segment the subcortical structures (zone 5a), others segment successfully with the strong prior (however not entirely, see zone 5b).

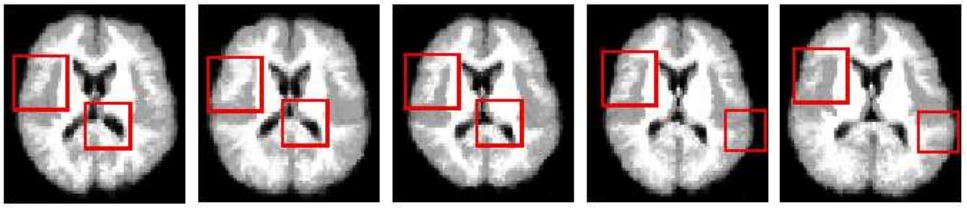


FIGURE 2.7: Five simulated images using the estimated template with 8 subjects. The deformations of the ventricles are realistic as well as the cortex foldings which look like some training ones. Moreover, the cortex thickness changes.

In order to evaluate the estimated geometric variability, we use our generative model to resample some images that should be representative of the population (Fig. 2.7). Our model manages to capture the global and local deformations. The second brain has a more round shape, the forth one a more elliptical shape and the last one is larger than the others. This global shape changes are therefore well captured. The deformations of the ventricles are realistic as well as the cortex foldings which look like some training ones. Moreover, the cortex thickness changes (highlighted in red). This shows that even with a small sample, we manage to capture the population geometric variability accurately.

The segmentation results are shown in Fig. 2.6. Each row of the first column corresponds to the same slice (128×160 voxels) of three training images. The manual segmentation

for each individual is shown in the second column. Our algorithm gives the final classification for each exemplar in the third column. We also get the segmentation from SPM8 and FAST algorithm in the last two columns. SPM8 and FAST outputs are probabilistic, hence we define the deterministic tissue class of each voxel as the tissue that has the maximum probability in this voxel. However, the SPM8 tissue probability maps are sharp. Therefore, there are almost no difference between the probability maps and the binary result (deterministic) that we used in our comparisons. These methods assume that there exists three classes, the class of CSF is considered as CSF+background in our case. Our segmentation looks accurate as it shows the calcarine sulcus folding in zone 4. Moreover, thanks to the class dependent variance, which is estimated along the algorithm iterations, there is no misclassification of voxels that creates holes with both SPM8 and FAST. This can be seen both in zone 4 on both sides of the cortex folds. Furthermore, the segmented cortical thickness by both SPM8 and FAST is much smaller than that given by the manual segmentation. This may come from two different aspects of these algorithms. First, they rely on a template that is not estimated with the observations and therefore may create a bias on the cortical thickness. Moreover, the registration is not done simultaneously with the segmentation. This may also create this bias as the deformation is crucial as already noticed for the synthetic examples. The FAST tissue probability maps are fuzzy, either registration or segmentation is not well done, therefore the resulting uncertainty is very large.

Our method fails to segment the subcortical structures. The voxels belonging to these structures have values between those of the GM and WM means in the training set. Therefore, they are either classified by GM or by WM. For example, the putamen's gray level is closer to WM than GM mean in zone 5a, therefore it is misclassified by our method. On the other hand, in zone 5b, as the gray level of the thalamus proper reaches a value closer to GM, our algorithm performs better. We can notice that both SPM8 and FAST capture these structures (however not entirely, see zone 5b). This is made possible thanks to their prior templates used for segmentation which contain these structures and thus guide the segmentation around these positions. In our model, there is no informative anatomical prior set on the template nor on the segmentations. Hence, the algorithm fails to fully classify these parts as GM.

To quantify the visual performance, we calculate the Jaccard index for each class for different methods (Table 2.3). We perform much better for the gray matter, as we succeed in segmenting the cortex with the right thickness whereas SPM8 and FAST shrink it. However the Jaccard index for the white matter is a little worse due to the misclassified subcortical structures. To compensate for the misclassified subcortical structures, we try to use DARTEL template as the informative prior on the probability maps $(\alpha_k)_{1\leqslant k\leqslant K}$. However the gray level plays a greater role than the prior in the process, we lose the

	FAST	SPM	SAEM
GM	58.6%	65.2%	**79.9%**
WM	**76.6%**	76.0%	68.9%

TABLE 2.3: Experiments on real data. The Jaccard Index for different methods average across 8 subjects. Our method gives a much higher value of Jaccard index for GM. However a little worse for WM, it is because our method does not manage to segment the subcortical structures as GM which is even difficult to segment manually. FAST and SPM8 use an anatomical prior, therefore they segment successfully these structures.

prior gradually and still fail to classify these structures.

As a last experiment, we create the probabilistic atlas with 20 images from Open Access Series of Imaging Studies (OASIS) [5]. There are 416 subjects aged from 18 to 96 with resolution $1 \times 1 \times 1\ mm^3$ in the OASIS dataset. The 20 images aged from 28 to 64 are chosen randomly. For the pre-processing, we use BET [73] to remove the non-brain tissue from the images of the OASIS database. The fourth column in Fig. 2.5 shows one slice of our probabilistic maps with 20 subjects. The template obtained with 20 subjects captures more details on the boundary of two types than the one with 8 subjects. This appears in particular on both right and left cerebral cortex areas, where the 8 subjects template classifies voxels only belonging to GM whereas the 20 subjects template captures the presence of WM voxels. Although our model has high dimensional parameters, we obtain a reasonable estimate with 20 images. The computation time is about 10 days for 8 images and almost a month for 20 images. Since our algorithm can be parallelized for the simulation step, we are working on a parallel C++ version of our code to make it possible to increase the training set and decrease the computation time. With the parallel version, it should cost about 1 day for 8 images and 3 days for 20 images.

2.6 Conclusion and discussion

In this study, we proposed a statistical model and used a stochastic algorithm to perform a probabilistic atlas estimation. This model opens the way to performing registration and segmentation simultaneously along the probabilistic atlas estimation. We also provide a proof of the convergence of the estimation procedure toward a critical point of the observed likelihood. Our algorithm has several advantages. First, the probabilistic atlas contains both the templates and the geometric variability of the population. Second, we do not need any pre-registration to perform the segmentation which is automatically obtained as an output. The experiments show that the proposed approach compares well with state-of-the art tools.

Our experiments also show that our model does not manage to segment the subcortical structures. This is easily explained by the fact that the image gray level provides ambiguous information in these regions, and that the segmentation is an ill-posed problem in the absence of prior information. One possible solution is thus to use the anatomical prior as SPM8 and FAST. Another solution is to use multimodal registration and segmentation. Multimodal images make it possible to take advantage of the different information given by different imaging modalities. In a recent generalisation of this model [150] (see Chapter 4), we manage to segment these structures using T1- and functional MRI. Another improvement would be to consider diffeomorphic deformations as in [134] or [140]. This control on the deformations would help to enforce anatomical constraints. However, one should keep a parametric description in order to be able to sample these deformations easily.

2.7 Proof of Theorem 2.1

From Equation (2.7), we have $q(y|\theta) = \sum_{k=1}^{K} q(y|c=k,\theta_c) \int q(c|\boldsymbol{\beta},\theta_g) q(\boldsymbol{\beta}|\theta_p) d\boldsymbol{\beta}$.
Since the right hand side term of Equation (2.9) is bounded by 1 (as it is a probability distribution),

$$\begin{aligned} q(y|\theta) &\leqslant \sum_{k=1}^{K} q(y|c=k,\theta_c) \int q(\boldsymbol{\beta}|\theta_p) d\boldsymbol{\beta} \\ &\leqslant \sum_{k=1}^{K} q(y|c=k,\theta_c) \\ &= \sum_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma_k^2}^{n|\Lambda|}} \exp\left(-\frac{\sum_{i=1}^{n}\sum_{j=1}^{|\Lambda|}(y_i^j - \mu_k)^2}{2\sigma_k^2} \right) \end{aligned}$$

where n is the number of images and $|\Lambda|$ is the number of voxels on the grid Λ.
We denote

$$f(\sigma_c^2) = \frac{1}{\sqrt{2\pi\sigma_c^2}^{n|\Lambda|}} \exp\left(-\frac{S_c^2}{2\sigma_c^2} \right)$$

where $S_c = \sum_{i=1}^{n}\sum_{j=1}^{|\Lambda|}(y_i^j - \mu_c)^2$. We want to bound f on $\mathbb{R}_*^+$, let f' be its derivative:

$$f'(\sigma_c^2) = \frac{1}{\sqrt{2\pi}^{n|\Lambda|}} (\sigma_c^2)^{-\frac{n|\Lambda|}{2}-1} \exp\left(-\frac{S_c^2}{2\sigma_c^2} \right) \left(\frac{S_c^2}{2\sigma_c^2} - \frac{n|\Lambda|}{2} \right).$$

For $\tilde{\sigma}_c^2 = \frac{S_c^2}{n|\Lambda|}$, $f'(\tilde{\sigma}_c^2) = 0$ and $f''(\tilde{\sigma}_c^2) < 0$, so that for all $\sigma_c^2 > 0$,

$$f(\sigma_c^2) \leqslant f(\tilde{\sigma}_c^2) = \frac{1}{\sqrt{2\pi}^{n|\Lambda|}} \left(\frac{S_c^2}{n|\Lambda|}\right)^{-\frac{n|\Lambda|}{2}} \exp\left(-\frac{n|\Lambda|}{2}\right).$$

Therefore

$$\begin{aligned}\log(q_B(\theta|y_1, ...y_n)) \leqslant & \log \sum_{k=1}^{K} \left(\frac{\sum_i \sum_j (y_i^j - \mu_k)^2}{n|\Lambda|}\right)^{-\frac{n|\Lambda|}{2}} + \frac{a_g}{2}\log|R_g| \\ & - \frac{a_g}{2}\langle R_g, \Gamma_g^0\rangle - \sum_{k=1}^{K}\left(\frac{a_p\sigma_0^2}{2\sigma_k^2} + \frac{a_p}{2}\log\sigma_k^2\right) - \frac{(\mu_k - m_u)^2}{2\sigma_u^2} + C\end{aligned}$$

where $R_g = \Gamma_g^{-1}$, and C is a constant which does not depend on the parameters. If we denote η_g^0 the smallest eigenvalue of Γ_g^0 and $\|R_g\|$ the operator norm of R_g (which is also its largest eigenvalue), we get

$$\langle R_g, \Gamma_g^0\rangle \geqslant \eta_g^0\|R_g\| \text{and } \log(|R_g|) \leqslant (3k_g - 1)\log\|R_g\| - \log\|\Gamma_g\|$$

so that

$$\lim_{\|R_g\|+\|\Gamma_g\|\to\infty} -\frac{a_g}{2}\langle R_g, \Gamma_g^0\rangle + \frac{a_g}{2}\log|R_g| = -\infty.$$

Similarly, we can show that

$$\forall k \in [\![1, K]\!], \lim_{\sigma_k^2+\sigma_k^{-2}\to\infty} \frac{a_p\sigma_0^2}{2\sigma_k^2} + \frac{a_p}{2}\log\sigma_k^2 = -\infty.$$

Moreover, for all $k \in [\![1, K]\!]$, there exists at least one voxel j_k in one image i_k such that $y_{i_k}^{j_k} \neq \mu_k$, otherwise all μ_k would be equal and all the images would be constant. Thus

$$\log \sum_{k=1}^{K} \left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{|\Lambda|}(y_i^j - \mu_k)^2}{n|\Lambda|}\right)^{-\frac{n|\Lambda|}{2}} \leqslant \log\left(\sum_{k=1}^{K}\left(\min_{\substack{(i_k,j_k),\\ y_{i_k}^{j_k}\neq\mu_k}}\left(\frac{(y_{i_k}^{j_k} - \mu_k)^2}{n|\Lambda|}\right)\right)^{-\frac{n|\Lambda|}{2}}\right).$$

Furthermore,

$$\forall k \in [\![1, K]\!], \lim_{|\mu_k|\to\infty}\left(\frac{(y_{i_k}^{j_k} - \mu_k)^2}{n|\Lambda|}\right)^{-\frac{n|\Lambda|}{2}} = 0.$$

which implies that

$$\lim_{|\mu_k|\to\infty} \log\left(\sum_{k=1}^{K}\left(\min_{\substack{(i_k,j_k),\\ y_{i_k}^{j_k}\neq\mu_k}}\left(\frac{(y_{i_k}^{j_k}-\mu_k)^2}{n|\Lambda|}\right)\right)^{-\frac{n|\Lambda|}{2}}\right) = -\infty.$$

So that

$$\forall k \in [\![1,K]\!], \lim_{|\mu_k|\to\infty} \log\sum_{k=1}^{K}\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{|\Lambda|}(y_i^j-\mu_k)^2}{n|\Lambda|}\right)^{-\frac{n|\Lambda|}{2}} - \frac{(\mu_k-m_u)^2}{2\sigma_u^2} = -\infty.$$

Now considering the Alexandrov one-point compactification $\Theta \cup \{\infty\}$ of Θ, we have

$$\lim_{\theta\to\infty} \log(q_B(\theta|y_1,...,y_n)) \to -\infty.$$

Since $\theta \to \log(q_B(\theta|y_1,...y_n))$ is smooth on Θ, we get the result.

2.8 Proof of Theorem 2.3

In this Section, we prove Theorem 2.3. To this purpose, we will follow the path of proof in [15], i.e. prove that the stochastic approximation sequence satisfies assumptions (**A1'**)(**ii**), (**iii**), (**iv**), (**A2**) and (**A3'**). The fact that the critical points remain in a level set of the Lyapunov function remains an assumption because of the complexity of our model. We detail only the crucial steps and arguments of the proof which differ from the previously mentioned one and refer to [15] when it is identical.

The sufficient statistic vector S, the set $\mathcal{S}$ as well as the explicit expression of $\hat{\theta}(s)$ have been given in Subsection 3.2. As noted, $\hat{\theta}$ is a smooth function of $\mathcal{S}$.

2.8.1 Proof of assumption (A1').

We recall that the functions H, h and w are defined in Subsection 2.4.3. Thanks to these particular forms, we satisfy (**A1'(iii)**) and (**A1'(iv)**) as proved in [47].

Moreover, since the interpolation kernel K_p is bounded, there exist $A > 0$, $B > 0$, $C > 0$, $D > 0$ such that for any $(c, \boldsymbol{\beta}) \in \mathcal{Z}$, we have

$$0 < S_{0,k}(c,\boldsymbol{\beta}) \leqslant A, \|S_{1,k}(c,\boldsymbol{\beta})\| \leqslant B, 0 \leqslant S_{2,k}(c,\boldsymbol{\beta}) \leqslant C, 0 \leqslant S_3(c,\boldsymbol{\beta}) \text{ and } 0 \leqslant S_{4,k,l}(c,\boldsymbol{\beta}) \leqslant D.$$

We define the set $\mathcal{S}_a$ by

$$\mathcal{S}_a \triangleq \{S \in \mathcal{S} | 0 \leqslant S_{0,k} \leqslant A, \|S_{1,k}\| \leqslant B, 0 \leqslant S_{2,k} \leqslant C, 0 \leqslant S_3 \text{ and } 0 \leqslant S_{4,k,l} \leqslant D\}.$$

Since the constraints are obviously convex and closed, we get that $\mathcal{S}_a$ is a closed convex subset of $\mathbb{R}^s$ such that

$$\mathcal{S}_a \subset \mathcal{S} \subset \mathbb{R}^s$$

and satisfying

$$s + \rho H_s(c, \boldsymbol{\beta}) \in \mathcal{S}_a \text{ for any } \rho \in [0,1] \text{ any } s \in \mathcal{S}_a \text{ any } (c, \boldsymbol{\beta}) \in \mathcal{Z}.$$

We now focus on the first two points. As `l` and $\hat{\theta}$ are continuous functions, we only need to prove that $\mathcal{W}_M \cap \mathcal{S}_a$ is a bounded set for a constant $M \in \mathbb{R}_*^+$ with:

$$\mathcal{W}_M = s \in \mathcal{S}, w(s) \leqslant M,$$

where $w(s)$ is defined in Definition 1.

On $\mathcal{S}_a$, s_0, s_1, s_2 and s_4 are bounded; writing $\hat{\theta}(s) = (\alpha_k(s))_{1\leqslant k\leqslant K}, (\mu_k(s))_{1\leqslant k\leqslant K}, (\sigma_k^2(s))_{1\leqslant k\leqslant K}, \Gamma_g(s))$, we deduce from Equation (2.15) that $(\alpha_k(s))_{1\leqslant k\leqslant K}, (\mu_k(s))_{1\leqslant k\leqslant K}, (\sigma_k^2(s))_{1\leqslant k\leqslant K}$ are bounded on $\mathcal{S}_a$. Considering the sufficient statistic s_3, thus

$$w(s) \geqslant -\log\left(\int\int q(y, c, \boldsymbol{\beta}, \hat{\theta}(s)) dc d\boldsymbol{\beta}\right) \geqslant -\log(q_B(\hat{\theta}(s))) + C \geqslant -\log(q_{B|_\Gamma}(\Gamma(s))) + C,$$

where C is a constant independent of $s \in \mathcal{S}_a$. Since

$$-\log(q_{B|_\Gamma}(\Gamma(s))) = \frac{a_g}{2}(\langle \Gamma_g^{-1}, \Gamma_g^0 \rangle_F + \log|\Gamma_g|) \geqslant \frac{a_g}{2}\log|\Gamma_g|$$

and

$$\lim_{\|s\|\to+\infty, s\in\mathcal{S}_a} \log(|\Gamma_g(s)|) = \lim_{\|s\|\to+\infty, s\in\mathcal{S}_a} \log(|(s_3 + a_g\Gamma_g^0)/(n * |\Lambda| + a_g)|) = +\infty,$$

we deduce that

$$\lim_{\|s\|\to+\infty, s\in\mathcal{S}_a} w(s) = +\infty.$$

Since w is continuous and $\mathcal{S}_a$ is closed, this proves **(A1'(ii))**.

2.8.2 Proof of assumption (A2)

We prove the same condition (**DRI1**) defined in [15] which will imply (**A2**) under the condition that H and V are related. We, in fact, have the following property : $\exists\, C > 0$ such that :

$$\sup_{s\in\mathcal{K}} |H_s(c,\boldsymbol{\beta})| \leqslant C\; V(c,\boldsymbol{\beta})\,,$$

where we set $V : [\![1,K]\!]^{|\Lambda|} \times \mathbb{R}^{3k_g} \to [1,+\infty[$ as the following function

$$V(c,\boldsymbol{\beta}) = 1 + \|\boldsymbol{\beta}\|^2. \tag{2.16}$$

We now prove the following lemma which gives the existence of the small set $\mathtt{C}$ required by condition **(DRI1)**:

Lemma 2.4. *Any compact set of $\mathcal{Z} = [\![1,K]\!]^{|\Lambda|} \times \mathbb{R}^{3k_g}$ is a small set for $(\Pi_{\hat\theta(s)})_{s\in\mathcal{K}}$*

Proof. Let A be a Borel set of $\mathcal{Z}$ and $x \in \mathtt{C}$ a compact subset of $\mathcal{Z}$, then we have

$$\begin{aligned}
\Pi_{\hat\theta(s),t}(\boldsymbol{x},A) &\geqslant \int_{A^t} \left(1 \wedge \frac{\pi_{\hat\theta(s),t}(z^t)}{q(z^t,\hat\theta(s))}\frac{q(x^t,\hat\theta(s))}{\pi_{\hat\theta(s),t}(x^t)}\right) q(z^t,\hat\theta(s))dz^t \\
&\geqslant \int_{A^t} \left(\frac{q(z^t,\hat\theta(s))}{\pi_{\hat\theta(s),t}(z^t)} \wedge \frac{q(x^t,\hat\theta(s))}{\pi_{\hat\theta(s),t}(x^t)}\right) \pi_{\hat\theta(s),t}(z^t)dz^t\,, \\
&\geqslant \int_{A^t} \left(\frac{q(z^t,\hat\theta(s))}{\pi_{\hat\theta(s),t}(z^t)} \wedge \frac{q(x^t,\hat\theta(s))}{\pi_{\hat\theta(s),t}(x^t)}\right) \pi_{\hat\theta(s),t}(z^t)\mathbb{1}_{\mathtt{C}}(\boldsymbol{z})dz^t\,.
\end{aligned}$$

If we can prove that for any compact set $\mathcal{K} \in \mathcal{S}$, there exists a constant $C_{\mathcal{K},\mathtt{C}}$ such that

$$\frac{\pi_{\hat\theta(s),t}(z^t)}{q(z^t,\hat\theta(s))} \leqslant C_{\mathcal{K},\mathtt{C}}\,, \tag{2.17}$$

then:

$$\begin{aligned}
\Pi_{\hat\theta(s),t}(\boldsymbol{x},A) &\geqslant \int_{A^t} \frac{1}{C_{\mathcal{K},\mathtt{C}}}\pi_{\hat\theta(s),t}(z^t)\mathbb{1}_{\mathtt{C}}(\boldsymbol{z})dz^t &(2.18)\\
&\geqslant \int_{A^t} \frac{1}{C_{\mathcal{K},\mathtt{C}}}\pi_{\mathcal{K},t}(z^t)\mathbb{1}_{\mathtt{C}}(\boldsymbol{z})dz^t\,, &(2.19)
\end{aligned}$$

where $\forall \boldsymbol{z} \in \mathtt{C},\ \pi_{\mathcal{K},t}(z^t) = \min_{s\in\mathcal{K}} \pi_{\hat\theta(s),t}(z^t)$ is a positive measure thanks to the smoothness of the probability measure $q(z^t|\boldsymbol{z}^{-t},\boldsymbol{y},\hat\theta(s))$ in its parameter s for all $\boldsymbol{z} \in \mathtt{C}$.

Let us now prove (2.17).

$$\begin{aligned}
\pi_{\hat{\theta}(s),t}(\boldsymbol{z}) &= q(z^t|\boldsymbol{z}^{-t},\boldsymbol{y},\hat{\theta}(s)) \\
&= \frac{q(\boldsymbol{y}|\boldsymbol{z},\hat{\theta}(s))q(z^t,\hat{\theta}(s))}{q(\boldsymbol{y}|\boldsymbol{z}^{-t},\hat{\theta}(s))} \\
\frac{\pi_{\hat{\theta}(s),t}(\boldsymbol{z})}{q(z^t,\hat{\theta}(s))} &\leqslant \frac{1}{\int \prod_{j=1}^{|\Lambda|} \exp\left(-\frac{1}{2\hat{\sigma}^2_{c_j}(s)}(y_j-\mu_{c_j})^2\right) q(z^t,\hat{\theta}(s))\mathbb{1}_{\mathtt{C}}(\boldsymbol{z})dz^t}
\end{aligned}$$

Since there exists $a>0$ such that $\forall j, (y_j-\mu_{c_j})^2 \leqslant a$ and since $\sigma_c^2 \geqslant \sigma^2_{min}$, then we have

$$\exp\left(-\frac{1}{2\sigma^2_{c_j}}(y_j-\mu_{c_j})^2\right) \geqslant \exp\left(-\frac{a}{2\sigma^2_{min}}\right)$$

So that

$$\begin{aligned}
\frac{\pi_{\hat{\theta}(s),t}(\boldsymbol{z})}{q(z^t,\hat{\theta}(s))} &\leqslant \frac{1}{\exp\left(-\frac{a}{2\sigma^2_{min}}\right)^{|\Lambda|}\int q(z^t,\hat{\theta}(s))\mathbb{1}_{\mathtt{C}}(\boldsymbol{z})dz^t} \\
&\leqslant \exp\left(\frac{a}{2\sigma^2_{min}}\right)^{|\Lambda|} \quad \text{car} \int q(z^t,\hat{\theta}(s))\mathbb{1}_{\mathtt{C}}(\boldsymbol{z})dz^t = 1.
\end{aligned}$$

It is bounded by $C_{\mathcal{K},\mathtt{C}} = \exp\left(\frac{a}{2\sigma^2_{min}}\right)^{|\Lambda|}$ on $\mathcal{K}$ for any $z \in \mathtt{C}$. The complete transition kernel is a composition of the previous kernel for t from 1 to $3k_g+|\Lambda|$. Since the coordinate of z are independent we get:

$$\Pi_{\hat{\theta}(s)}(\boldsymbol{x},A) \geqslant \int_A \left(\frac{1}{C_{\mathcal{K},\mathtt{C}}}\right)^{3k_g+|\Lambda|} \left(\prod_{t=1}^{3k_g+|\Lambda|} \pi_{\mathcal{K},t}(z^t)\right) \mathbb{1}_{\mathtt{C}}(\boldsymbol{z})d\mathbf{z}$$

This yields the existence of the small set and the third condition of (**DRI1**) with

$$\delta\nu(A) = \int_A \left(\frac{1}{C_{\mathcal{K},\mathtt{C}}}\right)^{3k_g+|\Lambda|} \left(\prod_{t=1}^{3k_g+|\Lambda|} \pi_{\mathcal{K},t}(z^t)\right) \mathbb{1}_{\mathtt{C}}(\boldsymbol{z})d\mathbf{z}$$

and ends the proof. □

For proving the first conditions of (**DRI1**), we need to ensure that our acceptance rates are always strictly positive. We notice that $\tilde{r}_p(\boldsymbol{\beta}^p,b;\boldsymbol{\beta}^{-p},c,\theta) = \frac{q(c|\boldsymbol{\beta}_{b\to p})}{q(c|\boldsymbol{\beta})} \geqslant q(c|\boldsymbol{\beta}_{b\to p}) > 0$, because for $\forall\ (c,\boldsymbol{\beta}) \in \mathcal{Z},\ q(c|\boldsymbol{\beta}) \in]0,1[$ which is justified in Remark 4. Therefore, for any compact set $\mathcal{K} \subset \mathcal{S}$, $\exists\ a_{\mathcal{K}} > 0 :\ \forall\ r_p(\boldsymbol{\beta}^p,b;\boldsymbol{\beta}^{-p},c,\theta) \geqslant a_{\mathcal{K}}$. The other proof of the first and second conditions of (**DRI1**) is similar with the proof in [15] with the function V defined in Equation (2.16).

2.8.3 Proof of assumption (A3')

For proving the Hölder condition **(A3'(ii))**. We will use the lemma 6.4 and lemma 6.5 in [15] which state Lipschitz conditions on the transition kernel and its iterates. If we can prove that the derivative of the acceptance rates in our model are Lipschitz functions, we get the result of lemma 6.4.

Proof. Concerning the derivative of $r_p(\boldsymbol{\beta}^p, b; \boldsymbol{\beta}^{-p}, c, \theta)$, since

$$\log(\tilde{r}_p(\boldsymbol{\beta}^p, b; \boldsymbol{\beta}^{-p}, c, \theta)) = \sum_{j=1}^{|\Lambda|} \log\alpha_k(c_j = k|\boldsymbol{\beta}_{b\to p}) - \sum_{j=1}^{|\Lambda|} \log\alpha_k(c_j \neq k|\boldsymbol{\beta}),$$

we have $|\frac{d}{d\epsilon}\tilde{r}_p(\boldsymbol{\beta}^p, b; \boldsymbol{\beta}^{-p}, c, \theta))| \leqslant C_{\mathcal{K}}\|s' - s\|$.

Concerning the derivative of $r_j(c^j, k; c^{-j}, \theta)$,

$$\begin{aligned}\left|\frac{d}{d\epsilon} r_j(c^j, k; c^{-j}, \theta)\right| &= \left|-\frac{1}{2}\frac{d}{d\epsilon}\log\sigma_k^2 - \frac{d}{d\epsilon}\frac{(y_j - \mu_k)^2}{2\sigma_k^2}\right| \\ &= \left|-\frac{1}{2\sigma_k^2}\frac{d}{d\epsilon}\sigma_k^2 + \frac{(y_i - \mu_k)^2}{2\sigma_k^4}\frac{d}{d\epsilon}\sigma_k^2 - \frac{y_i - \mu_k}{2\sigma_k^2}\frac{d}{d\epsilon}\mu_k\right| \\ &\leqslant C_1\left|\frac{d}{d\epsilon}\sigma_k^2\right| + C_2\left|\frac{d}{d\epsilon}\mu_k\right|\end{aligned}$$

where $C_1, C_2 > 0$ and

$$\left|\frac{d}{d\epsilon}\mu_k\right| = \left|\frac{s_{1,k}}{s_{0,k}} - \frac{s'_{1,k}}{s'_{0,k}}\right| = \left|\frac{s_{1,k}(s'_{0,k} - s_{0,k})}{s_{0,k}s'_{0,k}} - \frac{(s'_{1,k} - s_{1,k})s_{0,k}}{s_{0,k}s'_{0,k}}\right| \leqslant C_{\mathcal{K}}\|s - s'\|.$$

Similarly,

$$\left|\frac{d}{d\epsilon}\sigma_k^2\right| = \left|\frac{n}{n + a_p}\left(\frac{s_{2,k}}{s_{0,k}} - \frac{s'_{2,k}}{s'_{0,k}} - \frac{s_{1,k}^2}{s_{0,k}^2} + \frac{s_{1,k}'^2}{s_{0,k}'^2}\right)\right| \leqslant C_{\mathcal{K}}\|s - s'\|.$$

Thus, we have

$$\left|\frac{d}{d\epsilon} r_j(c^j, k; c^{-j}, \theta)\right| \leqslant C_{\mathcal{K}}\|s' - s\|.$$

□

The next proof for (**A3'(ii)**) and the proofs for **(A3')(i)(iii)** are the same as the proofs in [15].

Chapter 3

Bayesian Estimation of Probabilistic Atlas for Tissue Segmentation

Automatic anatomical brain image segmentation is still a challenge. In particular algorithms have to address the partial volume effect (PVE) as well as the variability of the gray level of internal brain structures which may appear closer to gray matter (GM) than white matter (WM). Atlas based segmentation is one solution as it brings prior information. For such tasks, probabilistic atlases are very useful as they take into account PVE. In this chapter, we provide a detailed analysis of a generative statistical model based on dense deformable templates that represents several tissue types observed in medical images. The inputs are gray level data whereas our atlas is composed of both an estimation of the deformation metric and probability maps of each tissue (called class). This atlas is used to guide the tissue segmentation of new images. Experiments are shown on brain T1 MRI datasets. This method only requires approximate pre-registration, as the latter is done jointly with the segmentation. Note however that an approximate registration is a reasonable pre-requisite given the application.

3.1 Introduction

Medical image analysis often requires the accurate delineation of different tissue types, based on the contrasts observed in different image modalities and a spatial model of the tissues. The use of automated segmentation is important both for the sake of efficiency and for reproducibility. Many different methods have already been proposed for segmentation, such as level set methods [141], model-based segmentation [72], atlas-based approaches [155] among a huge literature. Here we consider atlas-based segmentation, the success of which depends crucially on the choice of the atlas. Probabilistic atlases are attractive because they make it possible to take into account the uncertainty on the underlying tissue type, which is related to partial volume effect or to perfectible registration. Joint registration-segmentation [106] is generally more effective than sequential registration and segmentation [70, 71]: an accurate segmentation is needed to finesse the registration, and is made easier in turn by an accurate registration to a tissue template. In this chapter, we address atlas estimation, i.e. the joint estimation of a probabilistic template together with the geometric variability of a population, and, once the atlas has been learned, we use it to segment new observations, while taking into account the estimated anatomical variability.

Several solutions have been proposed previously to deal with one or the other part of this problem. In [112], Riklin-Raviv et al. proposed an automated method for brain segmentation via latent templates; however it requires a pre-registration. In [83], the registration is used for the templates estimation; however it requires pre-segmentation. In [117], a probabilistic model was proposed for segmentation, however it does not do the registration jointly with the segmentation. In [20], Ashburner et al. proposed to segment a single image providing tissue probability maps. However it deals with a single image, and thus cannot be used to understand the geometric variability among individuals. In [15, 153], a method was proposed to create an atlas containing the template and the geometric variability. However it does not perform segmentation jointly with atlas estimation and the learned atlases are deterministic. In [25], a method was proposed to do the segmentation and registration jointly, while creating an average brain atlas, however it does not learn the geometric variability. Therefore the deformations follow a fixed distribution which is not representative of the population. The posterior segmentation is coarse as it is not conditioned to relevant deformations.

In this paper, we propose a generative approach for the estimation of a probabilistic atlas including geometric variability. During this process, registration and segmentation are coupled with atlas estimation. We use the same observation model as in [20], generalize

it and use the algorithm in [15] for the estimation. We also learn the geometry as the metric on the space of deformation, which reduces the possible deformations to those that are common in the population. To estimate model parameters, we use a stochastic algorithm that has demonstrated good performance on real data in [14]. The output of the algorithm is the probabilistic atlas, the individual tissue segmentation and the means and variances of each tissue type. We use this atlas as an anatomical prior for segmentation of new individuals. Two tests were performed on the different databases.

3.2 Material

Objective: From the T1 MR images, our objective is to provide an atlas construction which includes both the probabilistic template and the geometric variability. During the process, the registration and segmentation are done jointly. Then we use the estimated atlas to segment new MR images by constraining the template to subject deformation with the estimated metric.

Dataset:
a- Internet Brain Segmentation Repository (IBSR) [3]: 18 T1 MRI with their manual segmentations. Eight with resolution $0.9735 \times 0.9735 \times 1.5mm^3$, six with resolution $1 \times 1 \times 1.5mm^3$, and four with resolution $0.837 \times 0.837 \times 1.5mm^3$. These data are provided by the Center for Morphometric Analysis at Massachusetts General Hospital.

b- Open Access Series of Imaging Studies (OASIS) [5]: 416 subjects aged 18 to 96 with resolution $1 \times 1 \times 1mm^3$.

c- MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling [4]: 15 (10 females) T1 MRI from OASIS project with their manual segmentation provided by Neuromorphometrics Inc., aged from 19 to 34 with resolution $1 \times 1 \times 1mm^3$.

Each image in these three datasets is the size of $256 \times 256 \times 128$. We reduce the background in our experiment, therefore the image size is $160 \times 160 \times 128$.

3.3 Methods

3.3.1 Statistical Model.

We consider n gray level MR images from n patients. This set, $(y_i)_{1 \leqslant i \leqslant n}$, of images are observed on a grid of voxels Λ embedded in a continuous domain $D \subset \mathbb{R}^3$. We denote

$x_j \in D$ the location of voxel $j \in \Lambda$. We consider that each image is composed of voxels belonging to one class among K with an associated probability P_k. We assume that the signal in the K tissue classes is normally distributed with class dependent means $(\mu_k)_{1\leqslant k\leqslant K}$ and variances $(\sigma_k^2)_{1\leqslant k\leqslant K}$.
As mentioned previously, the gray level images have not been pre-segmented. Therefore, the unobserved class of each voxel is assumed to follow a distribution which is the discretization on Λ of a random deformation of tissue probability maps $(P_k)_{1\leqslant k\leqslant K}$. These probability maps correspond to the probability of each voxel to belong to each class in the template domain. They form the probabilistic template of the population. The random deformations from these template maps to each subject are also unobserved. We define them through a random field $z : \mathbb{R}^3 \to \mathbb{R}^3$ such that for $j \in \Lambda$ the prior probability of a voxel j from subject i to be in the k^{th} class ($c_i^j = k$) is given by:

$$\mathbb{P}(c_i^j = k) = P_k(x_j - z(x_j)) \tag{3.1}$$

In order to reduce the problem to finite dimension, we define the deformation field as a finite linear combinations of a given kernel K_g centered at some fixed equi-distributed control points in the domain $D : (x_g)_{1\leqslant g\leqslant k_g}$ with parameter $\boldsymbol{\beta} \in (\mathbb{R}^3)^{k_g}$

$$\forall x \in D, z_\beta(x) = (\mathbf{K_g}\boldsymbol{\beta})(x) = \sum_{k=1}^{k_g} K_g(x, x_g)\boldsymbol{\beta}(k) \tag{3.2}$$

where K_g is chosen as a radial Gaussian Kernel. Note that we can also choose the control points non uniformly, we can fix more control points in regions of interest to get a more accurate result in areas of high variability.
As for the deformation model, the templates $P_k : \mathbb{R}^3 \to [0,1], \forall k \in [\![1, K]\!]$ which are the tissue probability maps are parameterized by the coefficients $\alpha_k \in [0,1]^{k_p}$, which satisfy $\forall l \in [\![1, k_p]\!], \sum_{k=1}^{K} \alpha_k^l = 1$. We pick a fixed set of landmarks $(p_l)_{1\leqslant l\leqslant k_p}$ which may be different from the geometric ones:

$$\forall x \in D, P_k(x) = \mathbf{K_p}\alpha_k(x) = \sum_{l=1}^{k_p} K_p(x, p_l)\alpha_k^l, \tag{3.3}$$

where $K_p(x, p_l) = 1$ if p_l is the nearest neighbor of x among the set of points $(p_j)_j$ and 0 otherwise.
The previous hypothesis provides a generative statistical model for a sample of gray level images (Fig. 3.1). The random variables are the deformation vector $\boldsymbol{\beta}$, the class of each voxel c and the observed gray levels of the images y. We assume that the deformation vector follows a normal distribution with mean zero and non-diagonal covariance matrix Γ_g. The covariance matrix Γ_g is not assumed to have any particular pattern of zeros.

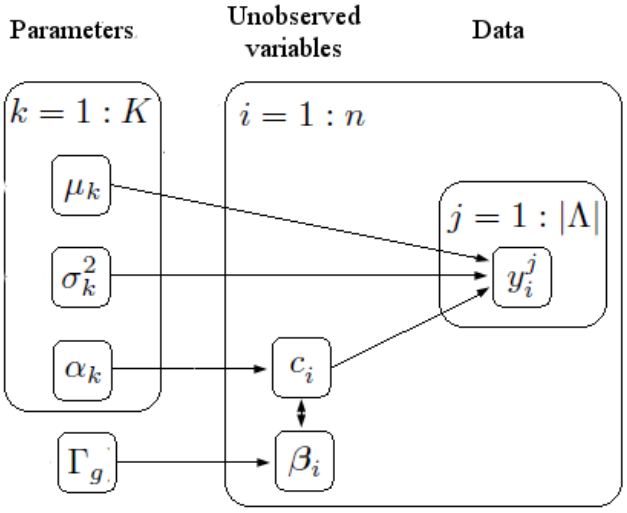


FIGURE 3.1: The generative statistical model.

This makes it possible to model local and global correlations between control point moves, in particular, very correlated displacements can be captured such as translation of a large area of the images.

The parameters to estimate are the covariance matrix Γ_g of the deformation vector β (Equation (3.2)), the coefficients that define the templates $(\alpha_k)_{1\leqslant k\leqslant K}$ (Equation (3.3)), the class dependent means $(\mu_k)_{1\leqslant k\leqslant K}$ and variances $(\sigma_k^2)_{1\leqslant k\leqslant K}$.
Medical images are typically high-dimensional, but usually come in small samples. Therefore to regularize the statistical model, we propose to work in a Bayesian framework. We use standard conjugate priors for the covariance matrix, the class dependent means and variances with fixed hyper-parameters.

3.3.2 Estimation Algorithm.

We use the maximum a posterior (MAP) estimator: $\hat{\theta}_n = \arg\max_{\theta\in\Theta} q_B(\theta|y_1, ..., y_n)$ where q_B denotes the posterior distribution of the parameters given the n observations $y_1, ..., y_n$.
As we are in an incomplete-data setting, we choose to use the Stochastic Approximation Expectation-Maximization (SAEM) [15, 47] to maximize the likelihood. This algorithm coupled with a Markov Chain Monte Carlo method has demonstrated good theoretical [15] and numerical [14] performances in such settings. The SAEM algorithm is an iterative procedure that consists of three steps. First, we simulate the missing data using a Metropolis-Hastings within Gibbs sampler. In our model, the missing data are the deformation vector and the tissue classes. Thanks to our choice of sampler, we get an online registration and segmentation. Then, a stochastic approximation is done on

the sufficient statistics using the simulated value of the missing data. Last, we maximize the log-likelihood with respect to the model parameters. See [151] for more details.

3.3.3 Segmentation of new individuals.

Once the atlas has been estimated, one would like to perform some posterior segmentation of new observations. This can easily be done using atlas-based segmentation methods as in [123, 155]. Our model can be used directly but this typically requires heavy computations. This complexity is not a problem when creating an atlas since this step has to be only performed once. However, the atlas based segmentation procedure has to be numerically efficient. To that purpose, we propose to use a different tool keeping all the specific aspects of the model, i.e. the parameters μ_k, σ_k^2, α_k and Γ_g.

More precisely, thanks to our estimated probabilistic template $(\hat{P}_k)_{1\leqslant k\leqslant K}$ given by Equation (3.3) with the estimated weights $(\hat{\mu}_k)_{1\leqslant k\leqslant K}$, we define the estimated gray level template image as

$$\hat{I} = \sum_{k=1}^{K} \hat{\mu}_k \hat{P}_k.$$

This template is defined on the whole space D. Note that this formulation of the template accounts for PVE in voxels.

Our atlas also provides the geometric variability of the population through the covariance matrix $\hat{\Gamma}_g$. We use this matrix as a metric for the space of deformations to constrain the registrations according to the learned distribution.

Given a target image y, the template $\hat{I}$ is deformed non-rigidly and registered to the target image by minimizing the classical energy:

$$E(\varphi) = \frac{1}{2}\|\varphi\|^2_{\hat{\Gamma}_g} + \int \frac{1}{2\hat{\sigma}^2_{c_{\varphi^{-1}(x)}}}(y(x) - \hat{I}\circ\varphi^{-1}(x))dx$$

The first term on the right hand side yields the cost of the deformation using the metric given by $\hat{\Gamma}_g$ whereas the second term quantifies the similarity between the observed image and the deformed template. The trade off between these two terms is given by the noise variances which have also been estimated to best fit the noise in the training dataset. Note that this noise also accounts for the fact that the images are not drawn by this simple approximating model.

Remark 5. We notice that in the numerical experiment, the variances of the tissue gray levels are very close to each other so that we assume in this posterior segmentation that they are all equal to σ^2.

Our model assumes that the deformations are linearized deformations given by control point movements. We keep this assumption and therefore the energy only depends on the vector of control point displacement $\boldsymbol{\beta}$. Using also Remark 5, we can approximate the integral by

$$E(\varphi_{\boldsymbol{\beta}}) \simeq \frac{1}{2}\boldsymbol{\beta}^T \hat{\Gamma}_g^{-1}\boldsymbol{\beta} + \frac{1}{2\sigma^2}\sum_{x\in\Lambda}\left(y(x) - \hat{I}(x - z_{\boldsymbol{\beta}}(x))\right)^2. \tag{3.4}$$

We use a gradient descent algorithm to minimize the criterion in Equation (3.4), which yields

$$\boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}\in\mathbb{R}^{3k_g}}{\operatorname{argmin}} \frac{1}{2}\boldsymbol{\beta}^T \hat{\Gamma}_g^{-1}\boldsymbol{\beta} + \frac{1}{2\sigma^2}\sum_{x\in\Lambda}\left(y(x) - \hat{I}(x - z_{\boldsymbol{\beta}}(x))\right)^2.$$

Then the tissue c_j^* for each voxel j of the new observation is chosen to be the class that maximizes the posterior probability of that voxel to belong to each class, given this deformation field $\boldsymbol{\beta}^*$,

$$c_j^* = \underset{c_j\in[\![1,K]\!]}{\operatorname{argmax}} \left[\log\left(q(y_j|c_j,\hat{\theta})q(c_j|\boldsymbol{\beta}^*,\hat{\theta})\right)\right].$$

The segmentation is therefore constrained by both the estimated template and the learned geometric variability.

3.4 Experiments and Results

We tested our algorithm on the previously described data and compared with SPM8 [7], package FAST [158] in the software FSL [1] and DARTEL algorithms [19].

As a first experiment, the proposed method is tested on data derived from manual annotations that are publicly available at the IBSR. The images were considered to have 2 tissue classes: GM and WM. We take 800 control points to drive the deformation model given in (Equation (3.2)) and 4066 points to drive the template model given in (Equation (3.3)).

At first, we run a five fold cross validation of our algorithm with 10 patient images. We visualize one group of our online segmentation of these training images and compare it to their manual segmentation (Fig. 3.2). Each column corresponds to one slice (160×128 voxels) of three training data, the manual segmentation, the segmentation obtained by first our method, SPM8 and lastly FAST. SPM8 and FAST outputs are probabilistic, the deterministic tissue class of each voxel is defined by the tissue that has the maximum probability in this voxel (as for our method cf. subsection 3.3.3). These two methods assume that there exists three classes and therefore CSF is considered as

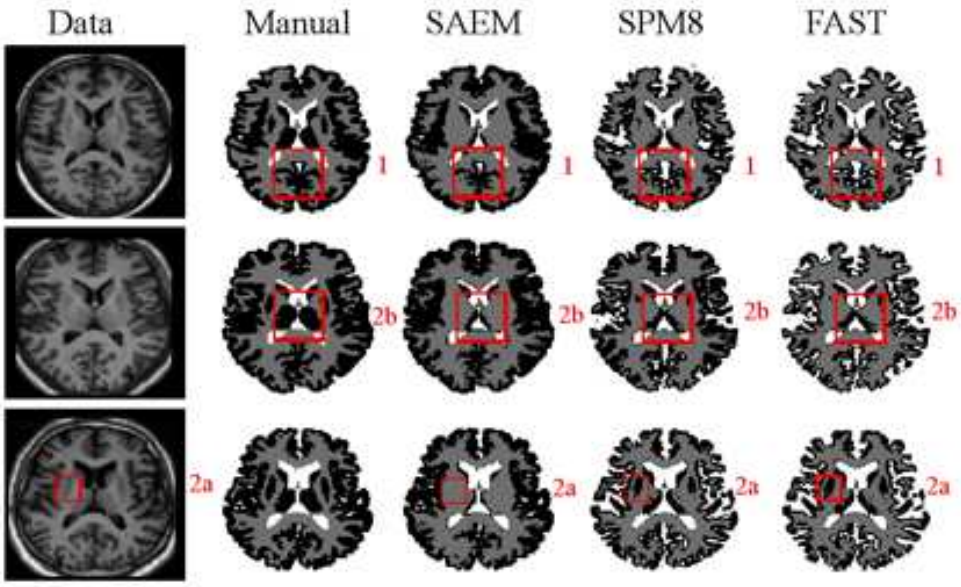


FIGURE 3.2: Each column corresponds to the one slice of 3 data images, the manual segmentation and the segmentation obtained by our method, SPM8 and FAST. Our method shows each fold of the GM (zone 1). Our method does not manage to segment the subcortical structures (zone 2a), others segment successfully with the strong prior (however not entirely, see zone 2b).

the background. The segmentation given by our algorithm looks accurate, for instance it shows the correct folds of the cortex in zone 1. Moreover, as it takes into account that the gray level of the tissues have a variance, which is estimated along the algorithm iterations, there are few misclassification of voxels contrary to SPM8 and FAST that overestimate the CSF, thus creating hole patterns in the cortex. This can be seen in the same brain areas in all patients as well as on all sides of the cortex folds. Furthermore, the thickness of the cortex segmented by both SPM8 and FAST is much smaller than the one given by the manual segmentation. This may come from two different aspects of these algorithms. First they rely on an atlas that may create a bias on the thickness of the cortex whereas our segmentation are performed at the same time as the atlas segmentation, therefore learned on the population. Moreover, they do not compute the registration simultaneously with the segmentation. This may also create this bias as the deformation is crucial as already noticed in [47]. The SPM8 tissue probability maps are sharp, leading to non accurate segmentation in particular due to PVE. The FAST tissue probability maps are fuzzy, as neither registration nor segmentation is sufficiently accurate; the uncertainty associated with this poor model results in a fuzzy template. As we can see in box 2a-b, our method fails to segment the subcortical structures. The voxels belonging to these structures have gray level values between the GM and WM means in the training set. Therefore, they are either classified by GM or by WM. A solution of this problem is to use an anatomical prior as in [119]. Zone 2a highlights the putamen that is misclassified by our method. As we can see its gray level is closer to WM

	FAST	SPM	SAEM
GM	59.3%	64.9%	**78.9%**
WM	**76.4%**	75.5%	69.9%

TABLE 3.1: The mean Jaccard Index for different methods in the five fold cross validation. Our method gives a much higher value of Jaccard index for GM. However a little worse for WM, it is because our method does not manage to segment the subcortical structures as GM which is even difficult to segment manually. FAST and SPM8 use an anatomical prior, therefore they segment successfully these structures.

than GM mean. On the other hand, in zone 2b, as the gray level of the thalamus reaches a value closer to GM, our algorithm manages to perform better. We can notice that both SPM8 and FAST capture these structures (however not entirely, see zone 2b). This is made possible thanks to their prior templates used for segmentation which contain these structures and thus force the segmentation around these positions. In our model, there is no informative anatomical prior set on the template and on the segmentations. Hence, our algorithm fails to fully classify these parts as GM, however it is not forced to construct a template with specific patterns.

As a quantitative test, we calculate the mean Jaccard index for each class for different methods (Table 3.1). Our method gets a better jaccard index for the gray matter than SPM8 and FAST with a significant improvement ($p \leqslant 0.05$ on the paired t-test comparing 5 pairs of jaccard index), which results from the fact that our method segments the cortex with the right thickness whereas SPM8 and FAST shrink it. As expected, we perform a little worse for the white matter, this phenomenon is a consequence of the misclassified subcortical structures.

Now we compare our estimated template with DARTEL template that uses the SPM's segmentation as input. The first two columns in Fig. 3.3 show one slice of DARTEL template (first column) and our probabilistic maps with 8 subjects (second column), the first and second rows correspond to GM and WM respectively. DARTEL atlas is smoother than our SAEM template. This comes from a smoothing step in the algorithm that creates these regular contours. Moreover, it also relies on an anatomical prior template so that the output looks very contrasted (almost binary). Although our atlas is less smooth, it only takes into account the training data and thus avoids creating a bias such as the large line of CSF in zone 3 that appears much thinner in the data and by consequence in our estimated template. The non smooth template may also be an advantage as it explains the uncertainty on voxels coming from both the PVE and the registration level of details. Another bias is shown in zone 4 where our probabilistic maps capture the presence of WM.

As a last experiment, we create our probabilistic atlas with 20 images from OASIS [5], and then segment 15 new images from MICCAI Challenge [4] using the estimated atlas.

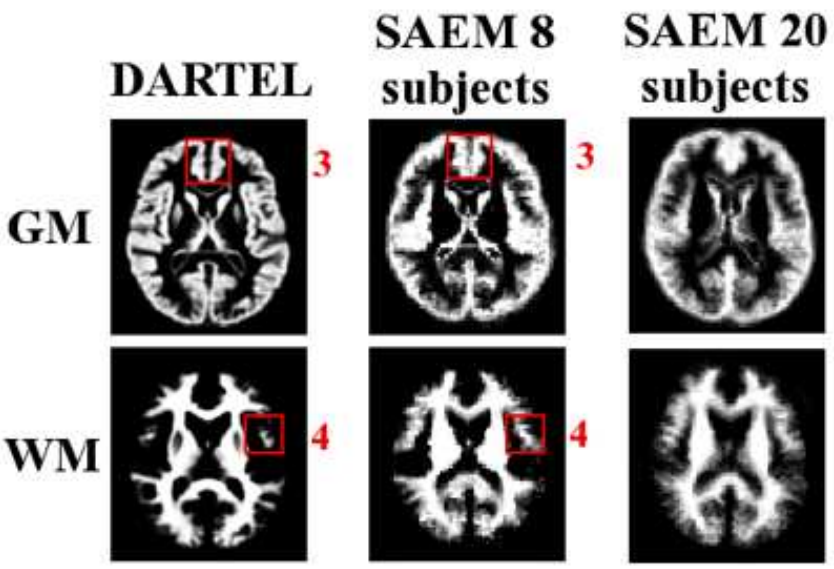


FIGURE 3.3: The atlas obtained by DARTEL with 8 subjects(first column), our method with 8 subjects (second column) and 20 subjects (third column). GM for the first row and WM for the second. DARTEL atlas shows a wide line of CSF in zone 3 which is thin in the data images. Our probabilistic maps capture the presence of WM in zone 4, however DARTEL only detects a small region. Our estimated atlas with 8 subjects is good. With 20 subjects, we get more details on the boundary of two types.

	FAST	SPM	SAEM
GM	66.4%	66.2%	**73.2%**
WM	72.8%	**76.3%**	72.8%

TABLE 3.2: The Jaccard Index for different methods (our segmentation method for new individual, FAST and SPM8) average across 15 subjects.

For the pre-processing, we use BET [73] to remove the non-brain tissue from the images of the database OASIS. The posterior segmentation for the 15 new images (Fig. 3.4) looks quite accurate (however, with the same misclassification of the subcortical structures). We calculate the Jaccard index for each class (Table 3.2). We get a Jaccard index around 73% for each tissue type. These rates outperform SPM8 and FAST for GM as we already noticed for the previous experiment. Concerning the WM, our classification rate is similar to those of other algorithms, which highlights the classification power of our model. In MICCAI Challenge, two classes (cortical and non-cortical) are considered for evaluating the result. The subcortical structures are considered as non-cortical: this confirms that these structures are difficult to segment with the gray level images. This particular detail will lead to low jaccard indexes for algorithms based on wrong priors (as SPM8 and FAST). The third column in Fig. 3.3 shows one slice of our probabilistic maps with 20 subjects. Compared with our template obtained with 8 subjects, the template obtained with 20 subjects captures more details on the boundary of two classes. This appears in particular on both right and left cerebral cortex areas

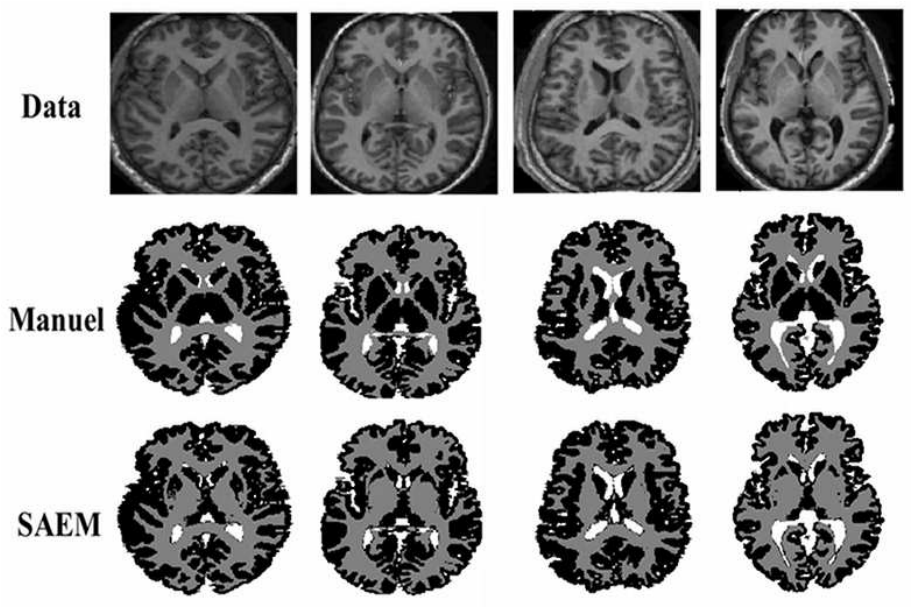


FIGURE 3.4: Each row corresponds to the one slice of 4 data images, the manual segmentation and the segmentation obtained by our method.

where the 8 subjects template classifies voxels only belonging to GM whereas the 20 subjects template captures the presence of WM voxels. Not unexpectedly, the procedure readily benefits from larger learning sets; at which level it asymptotes is an interesting question for the future. Since we only use the gray level images to create our atlas without any other information, the segmentation does not work well if the new brain is not similar to the population used to estimate the atlas. The constraint to use our method is that the new brain should belong to the same population in the sense that estimating an atlas from healthy brains will fail to segment a test brain with a tumor. Concerning the computational cost of our method, the atlas construction takes about 10 days for 8 images and almost a month for 20 images, whereas it takes about 5 minutes to segment the new image. The high dimensional covariance matrix makes the template construction long, however, it is interesting to spend some time to learn the geometric variability since the template construction should be done only once. Moreover, our algorithm can be parallelized for the simulation step by distributing one image per processor, which makes it possible to use more training data for the atlas construction within fixed computational time.

3.5 Conclusion and Discussion

In this study, we proposed a statistical model and used a stochastic algorithm to do registration and segmentation simultaneously in order to create a probabilistic atlas. Our algorithm has several advantages. First, we do not need a precise pre-registration which

is done jointly with the segmentation during the atlas estimation. Second, the atlas contains both the probabilistic templates and the geometric variability of the population. Third, we can use this anatomical prior for segmenting new individuals. Our experiments also show that our method fails to segment the subcortical structures as GM. This is easily explained by the fact that the voxels belonging to these structures are between the GM and WM means in the datasets. This gives a problem for segmentation without the prior information. One possible solution is to use the anatomical prior as SPM8 and FAST. Another solution is to use multimodal images that take advantage of the different information given by different imaging modalities. In particular, current work extends this model to anatomo-functional data which manage to segment the subcortical structures as GM thanks to some activations that only appear in GM in these structures [150].

Chapter 4

Bayesian Estimation of Probabilistic Atlas for Anatomically-Informed Functional MRI Group Analyses.

Traditional analyses of Functional Magnetic Resonance Imaging (fMRI) use little anatomical information. The registration of the images to a template is based on the individual anatomy and ignores functional information; subsequently detected activations are not confined to gray matter (GM). In this chapter, we propose a statistical model to estimate a probabilistic atlas from functional and T1 MRIs that summarizes both anatomical and functional information and the geometric variability of the population. Registration and Segmentation are performed jointly along the atlas estimation and the functional activity is constrained to the GM, increasing the accuracy of the atlas.

4.1 Introduction

Brain atlases are a useful tool in medical image analysis for both segmentation and registration tasks. Probabilistic atlases yield a useful summary of a given dataset [63, 84], as they take into account the uncertainty on the underlying tissue type, which is related to partial volume effect (PVE) or to perfectible registration. In [20], a probabilistic framework was proposed for joint nonlinear registration, intensity normalization and segmentation of a single image, from which it infers tissue probability maps. In [110], a probabilistic model was proposed to segment a heterogeneous data set of brain MRIs simultaneously while constructing the probabilistic atlases. In spite of its convincing results, this model is not consistent, as the deformations are considered as parameters (whereas segmentation is an unobserved random variable). In [151], the model proposed in [20] was generalized in order to provide estimates of individual segmentations as well as the probabilistic atlas from a set of anatomical images. This approach handles both the segmentation and registration as hidden variables, leading to a coherent convergent statistical estimator. However, this model is limited to scalar images. Here, we generalize it to create a probabilistic atlas that provides the probabilistic templates of each tissue as well as the degree of activation on GM voxels and the geometric variability.

Functional Magnetic Resonance Imaging of the brain is used to localize functional areas in the cortex and deep nuclei by measuring MRI signal changes associated with neural activity. It is a tool of choice for cognitive studies that aim at identifying specific regions of the brain that are activated in perceptual, cognitive or motor tasks. The most popular type of analysis is Statistical Parametric Mapping (SPM) [59], an approach that estimates the probability that some activation can be due to chance alone and provides p-value maps. Group analysis is then used to detect regions that show a positive mean activation across subjects [58, 132]. Accurate realignment of individual scans is most often obtained by normalizing individual anatomical images to a T1 MRI template. These processing steps are done without considering the complementarity of the anatomical and functional information available in each subject. Therefore, detected activations are not confined to gray matter. Few fMRI segmentation methods have been proposed to take into account multi-modal data, such as T1 and functional MRI. An implementation of cortical-based analysis of fMRI data was proposed in [16]. The fMRI data are mapped to the cortical surface, then activations are detected on the surface. It has been shown to achieve anatomically accurate activation detection. In [98], Markov Random Fields (MRF) were used as a spatial regularization in fMRI detection and anatomical information was incorporated into the MRF-based detection framework. In [118], both anatomical and functional data are used to improve the group-wise registrations. Anatomical information appears helpful in fMRI detection; however, the approaches

so far do not incorporate a group model into the analysis. In this paper, we process multi-modal data jointly to ensure that the detected active areas are conditioned to gray matter while registration is informed by functional information. More specifically, group analysis first performs the alignment of individual images to a T1 MRI template and then segments active regions by thresholding. However, performing registration and segmentation jointly is generally more effective than performing them sequentially [151, 153]. In this paper, we take advantage of such coupling.

To deal with all the issues described above, we propose an atlas estimation procedure that can improve the template image estimation and the detection of the active areas. We generalize the model proposed in [151]. The input is now multivariate, as it encodes multi-modal patient observations (gray level T1 and functional MRIs). The estimated active areas are conditioned to GM segmentation. We perform the estimation by coupling the segmentation and registration steps. We estimate a probabilistic atlas that accounts for the variability of active areas in the population. We also learn the geometry as the metric on the space of deformations that drive the coupled segmentation. We use a stochastic algorithm with known guarantees on the convergence of the estimation procedure. The output of the algorithm is the probabilistic atlas, the individual active areas and the means and variances of each tissue type in each modality.

The rest of this paper is organized as follows. In Section 2, we present the model, the estimation procedure, the algorithm. Section 3 yields experimental results on simulated and real data.

4.2 Methods

4.2.1 Statistical Model.

Let us consider n groups of T1- and f- MRIs $(y_{1,i}, y_{2,i}, ..., y_{m,i})_{1 \leqslant i \leqslant n}$ from n patients. Each image is observed on a grid of voxels Λ embedded in a continuous domain $D \subset \mathbb{R}^3$. We denote $x_j \in D$ the location of voxel j. We consider that each T1 MRI is composed of voxels belonging to one of the four classes, corresponding to four tissue types: gray matter, white matter (WM), CSF and background (BG). Each fMRI is composed of voxels belonging to one class among $3+K$, corresponding to WM, CSF and BG, where no activation is expected to occur, and K different levels of activation in gray matter. We assume that the signal in the $(3+K)^{(m-1)}$ classes is normally distributed with class dependent means $(\mu_{1,f(k_2)}, \mu_{2,k_2}, ..., \mu_{m,k_m})_{k_s \in \{WM,CSF,BG,GM_1,...GM_K\}}$ and variances $\left(\sigma^2_{1,f(k_2)}, \sigma^2_{2,k_2}, ..., \sigma^2_{m,k_m}\right)_{k_s \in \{WM,CSF,BG,GM_1,...,GM_K\}}$, where $2 \leqslant s \leqslant m$ and $f(k) = k$ if $k \in \{WM, CSF,$

$BG\}$, GM otherwise. We assume that $\forall(s,s') \in [2,m]^2,\ f(k_s) = f(k_{s'})$. The whole set of parameters is denoted by Θ.

As mentioned previously, we are working with gray level images which have not been pre-segmented. The unknown class of each voxel is supposed to be the discretization on Λ of a random deformation of probability maps $(P_{s,k})_{1\leqslant k\leqslant K+3, 2\leqslant s\leqslant m}$. These probability maps correspond to the probability of each voxel to belong to each class in the template domain. They form the *probabilistic template* of the population. The random deformations from this template to each subject are also unobserved as the images are not pre-registered. We define them through a random field $z : \mathbb{R}^3 \to \mathbb{R}^3$ such that for $j \in \Lambda$ the prior probability of a voxel j from subject i in the s^{th} image to be in the k_s^{th} class is given by:

$$\mathbb{P}(c_{s,i}^j = k_s) = P_{s,k}(x_j - z(x_j))\,. \tag{4.1}$$

We define the deformation field as a finite linear combination of a given kernel K_g centered at some fixed equi-distributed control points in the domain D, $(x_g)_{1\leqslant g\leqslant k_g}$, with parameter $\beta \in (\mathbb{R}^3)^{k_g}$

$$\forall x \in D, z_\beta(x) = \sum_{g=1}^{k_g} K_g(x, x_g)\beta(g)\,, \tag{4.2}$$

where K_g is chosen as a radial Gaussian Kernel. Note that we expect the tissue-specific information to be found in all the brain volume, hence the whole volume has to be covered with control points. As for the deformation model, the probability template maps $P_{s,k} : \mathbb{R}^3 \to]0,1[, \forall k_s \in [\![1, K+3]\!]$ are parametrized by the coefficients $\alpha_{s,k} \in]0,1[^{k_p}$ that satisfy $\forall l \in [\![1, k_p]\!], \sum_{k=1}^{K+3} \alpha_{s,k}^l = 1$. Let $(p_l)_{1\leqslant l\leqslant k_p}$ be some control points :

$$\forall x \in D, P_{s,k}(x) = \sum_{l=1}^{k_p} K_p(x, p_l)\alpha_{s,k}^l, \tag{4.3}$$

where $K_p(x, p_l) = 1$ if p_l is the nearest neighbor of x among $(p_j)_j, 0$ otherwise.

The previous hypothesis provides a generative statistical model for a sample of pairs of gray level images. The random variables are the deformation vector β, the class of each voxel c and the observed gray levels of the images. We assume that the deformation vector follows a normal distribution with mean zero and non-diagonal covariance matrix

Γ_g. The hierarchical model is given by:

$$\left\{\begin{array}{l} \beta_i \sim \mathcal{N}(0,\Gamma_g)|\Gamma_g, \\ c_{i,s}^j \sim \sum_{k=1}^{K+3} \delta_k P_{s,k}(x_j - z_{\beta_i}(x_j))|\beta_i, \\ \begin{pmatrix} y_{1,i}^j \\ y_{2,i}^j \\ \vdots \\ y_{m,i}^j \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{1,f(k_2)} \\ \mu_{2,k_2} \\ \vdots \\ \mu_{m,k_m} \end{pmatrix}, \begin{pmatrix} \sigma^2_{1,f(k_2)} & 0 & \cdots & 0 \\ 0 & \sigma^2_{2,k_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma^2_{m,k_m} \end{pmatrix}\right) \end{array}\right| c_{i,s}^j = k_s, \begin{pmatrix} \mu_{1,f(k_2)} \\ \mu_{2,k_2} \\ \vdots \\ \mu_{m,k_m} \end{pmatrix}, \begin{pmatrix} \sigma^2_{1,f(2)} \\ \sigma^2_{2,k_2} \\ \vdots \\ \sigma^2_{m,k_m} \end{pmatrix} \tag{4.4}$$

where $\mathcal{N}(\cdot;\mu,\sigma^2)$ is the normal density with mean μ and variance σ^2 and δ_k is a Dirac function. The covariance matrix Γ_g is not assumed to have any particular pattern of zeros. This makes it possible to model local and global correlations between control point moves, in particular, very correlated displacements can be captured such as translation of a large area of the images.

The parameters to estimate are the covariance matrix Γ_g of the deformation distribution (Eq. (4.2)), $(\alpha_{s,k})_{1\leqslant s\leqslant m, 1\leqslant k\leqslant K+3}$ the coefficients that define the template maps (Eq. (4.3)), the class dependent means and variances. As medical images are high-dimensional but usually come in small samples, we work in a Bayesian framework. We use the standard conjugate priors for the covariance matrix, the class dependent means and variances with fixed hyper-parameters. All priors are assumed independent.

4.2.2 Estimation Algorithm.

A maximum a posteriori (MAP) approach yields estimates of the model parameters:

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} q_B(\theta|(y_{1,1},y_{2,1}),\cdots,(y_{1,n},y_{2,n})),$$

where q_B denotes the posterior distribution of the parameters given the n observations $(y_{1,1},y_{2,1},...,y_{m,1}),...,(y_{1,n},y_{2,n},...,ym,n)$. As we are in an incomplete-data setting, we choose the Stochastic Approximation Expectation-Maximization (SAEM) algorithm coupled with a Markov Chain Monte Carlo method to take advantage of its theoretical and numerical properties [15, 151]. The SAEM algorithm is an iterative procedure that consists of three steps. First, we simulate the missing data using a Metropolis-Hastings algorithm within Gibbs sampler. Then a stochastic approximation is done on the sufficient statistics using the simulated value of the missing data. Last, we maximize the expected log-likelihood with respect to the model parameters. The whole algorithm is detailed in Algorithm 2, in particular the steps of the hybrid Gibbs sampler.

Algorithm 2 SAEM-MCMC Algorithm (with no reprojection)

Require: $c = (c_2, ..., c_m)_0$, $\boldsymbol{\beta} = (\boldsymbol{\beta})_0$, θ_0, s_0, Δ
  `Stochastic Approximation Expectation-Maximization`
  for $t = 0$ to iters **do**
    `Simulation step using Gibbs sampler:`
    for $i = 1$ to n **do**
      for $p = 1$ to $3k_g$ **do**
        `Metropolis-Hastings procedure`
        $b \sim \mathcal{N}\left(\frac{\sum_{q \neq p} R_{p,q}\boldsymbol{\beta}_i^q}{R_{p,p}}, \frac{1}{R_{p,p}}\right)$
        `Compute` $r_p(\boldsymbol{\beta}_i^p, b; \boldsymbol{\beta}_i^{-p}, c_2, ..., c_m, \theta_t) = \left[\prod_{s=2}^{m} \frac{q(c_s|\boldsymbol{\beta}_{i,b\to p})}{q(c_s|\boldsymbol{\beta}_i)} \wedge 1\right]$
        `With probability` $r_p(\boldsymbol{\beta}_i^p, b; \boldsymbol{\beta}_i^{-p}, c_2, ..., c_m, \theta_t)$, `update` $\boldsymbol{\beta}_i^p : \boldsymbol{\beta}_i^p \leftarrow b$
      end for
      `Update` $\boldsymbol{\beta}_{i,m+1} \leftarrow \boldsymbol{\beta}_i$
      for $j = 1$ to $|\Lambda|$ **do**
        for $s = 2$ to m **do**
          $C_s \sim \sum_{k=1}^{K} \delta_k P_{s,k}(x_j - z_{\boldsymbol{\beta}_{i,k+1}}(x_j))$
        end for
        `Compute` $r_j(c_i^j, C_s; c_{s,i}^{-j}, \theta_t) = \left[\frac{q(y_{1,i}|f(c_{2,i}), f(C_2)\to j, \theta_t)}{q(y_{1,i}|f(c_{2,i}), \theta_t)} \prod_{s=2}^{m} \frac{q(y_{s,i}|c_{s,i}, C_s\to j, \theta_t)}{q(y_{s,i}|c_{s,i}, \theta_t)} \wedge 1\right]$
        `With probability` $r_j(c_i^j, C; c_i^{-j}, \theta_t)$, `update` $c_i^j : c_i^j \leftarrow C$
      end for
      `Update` $c_{i,t+1} \leftarrow c_i$
    end for
    `Stochastic approximation step:`
    $s_{t+1} = s_t + \Delta_t[\mathcal{S}((c)_{t+1}, (\boldsymbol{\beta})_{t+1}) - s_t]$
    `Maximization step:`
    $\theta_{t+1} = \arg\min_{\theta\in\Theta} \hat{\theta}(s_{t+1})$.
  end for

4.3 Experiments and Results

We test our algorithm on both simulated data and real data. As the SAEM algorithm is an iterative procedure, we run 250 iterations which was checked to reach convergence. We initialize $\beta_0 = 0$ and a random classification c_0.

4.3.1 Simulated data.

We use a pair of $64 \times 64 \times 8$ images as the reference images ($s = 2$). We consider here $K = 3$, i.e. three different levels of activation in GM and 6 classes in total. We define the means and the standard deviations as follows (taking values that are observed in

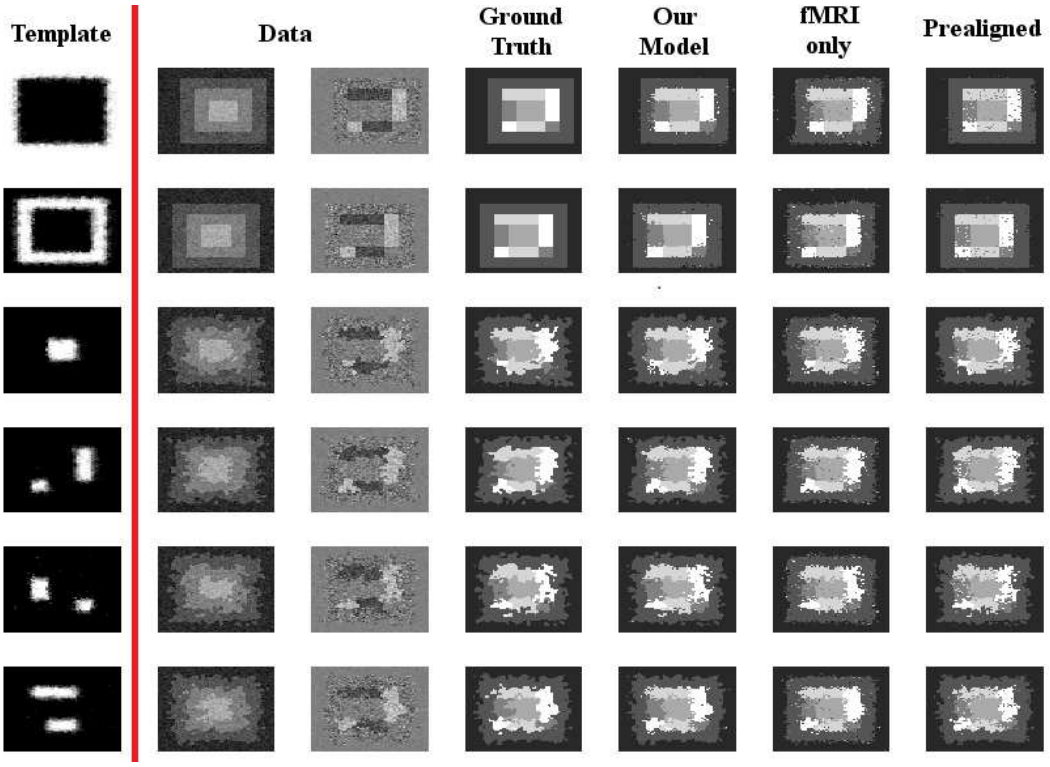


FIGURE 4.1: Experiments on simulated data. The first column displays the first slice of the probabilistic template, each row corresponding to a class and white/black colors to high/low probability. The second and third columns show one slice of six pairs of data images. The fourth to seventh columns correspond to the ground truth and the estimated segmentation for different models.

real fMRI for the standard deviations):

$$\begin{pmatrix} \mu_{1,1:6} \\ \mu_{2,1:6} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 4 & 3 & 3 & 3 \\ 0 & 0 & 0 & 2.5 & 0 & -2.5 \end{pmatrix}, \quad \begin{pmatrix} \sigma_{1,1:6} \\ \sigma_{2,1:6} \end{pmatrix} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.24 & 1.22 & 0.91 & 0.78 & 0.71 & 0.83 \end{pmatrix}$$

The training data is composed of 20 pairs of images with random deformations of our template following Eq. (4.4) with previous parameters. We take 64 fixed control points for the deformation model given in Eq. (4.2), i.e. one control point in each $4 \times 4 \times 4$ cube. We take all the points in the image as landmarks for the template model given in Eq. (4.3).

The most important output of our estimation procedure is the probabilistic template. The estimated probabilistic maps are shown in the first column in Fig. 4.1, each row corresponding to one class. The white/dark colors represent high/low probability of the tissues. Our probabilistic maps are sharp, as most voxels in each class have a probability larger than 0.9. Only voxels at the boundary between two classes are fuzzy, which account for both the accuracy of deformation and the level of noise.

	BG	CSF	WM	GM_1	GM_2	GM_3
Our model	**98.5%**	**92.3%**	**88.4%**	**88.0%**	**73.1%**	**90.4%**
fMRI only	89.0%	75.0%	69.7%	78.3%	48.2%	81.4%
Pre-Aligned	96.6%	91.1%	87.0%	86.6%	67.8%	86.6%

TABLE 4.1: Experiments on synthetic data. Jaccard Index for the different methods averaged across all data.

As mentioned previously, our model uses both the T1- and f- MRIs because we want the active areas to be conditioned to GM. We compare our model with the segmentation model in [151] using fMRI only. The result is shown in Fig. 4.1. The second and third columns correspond to one slice of six pairs of data images. The ground truth and the final estimated segmentation of different methods (our model, the model using fMRI only) are shown from the fourth to sixth columns. From the fourth and fifth columns, we can see that the segmentation obtained with our atlas estimation is accurate. From the fifth and sixth columns, we see the improvement using the information provided from the T1 MRI. We calculate the Jaccard index for each class as a quantitative validation (Table 4.1) for each method. Our model yields an accurate segmentation, as only few voxels are misclassified. For the model using fMRI only, we are only interested in GM_1 and GM_3 which correspond to the active areas. As the other classes are non-active, the means of these classes are close to zero, therefore they are difficult to segment without the MRI tissue type information, leading to lower values for these classes. Moreover, the Jaccard indexes for the active area are lower than those obtained with our procedure, which shows that the coupling of information from both images increases the accuracy of detection.

In our model, the registration and segmentation are done jointly, which avoids any pre-registration. In the preprocessing, each fMRI is pre-aligned to its corresponding MR image. However, the inter-subject non-rigid registration is not done, as it would require a template and would not take into account the fMRI observation to drive this preprocessing step. We compare our model with the pre-aligned model which does the registration and the segmentation sequentially. Fisrt, we use the segmentation model in [151] using the T1 MRIs, we get the deformation vector and individual tissue segmentation as our output. Then we apply the same deformation to the fMRI and detect the activation only in GM. The estimated segmentation of the pre-aligned model is shown in the seventh columns in Fig. 4.1. Comparing the fifth and seventh columns, the segmentations look similar which makes it difficult to say which method gives the better result. However our model gives less isolated points. Moreover, looking at the Jaccard indexes (Table 4.1), we see that our model outperforms the pre-aligned model. This shows the improvement of doing registration and segmentation jointly.

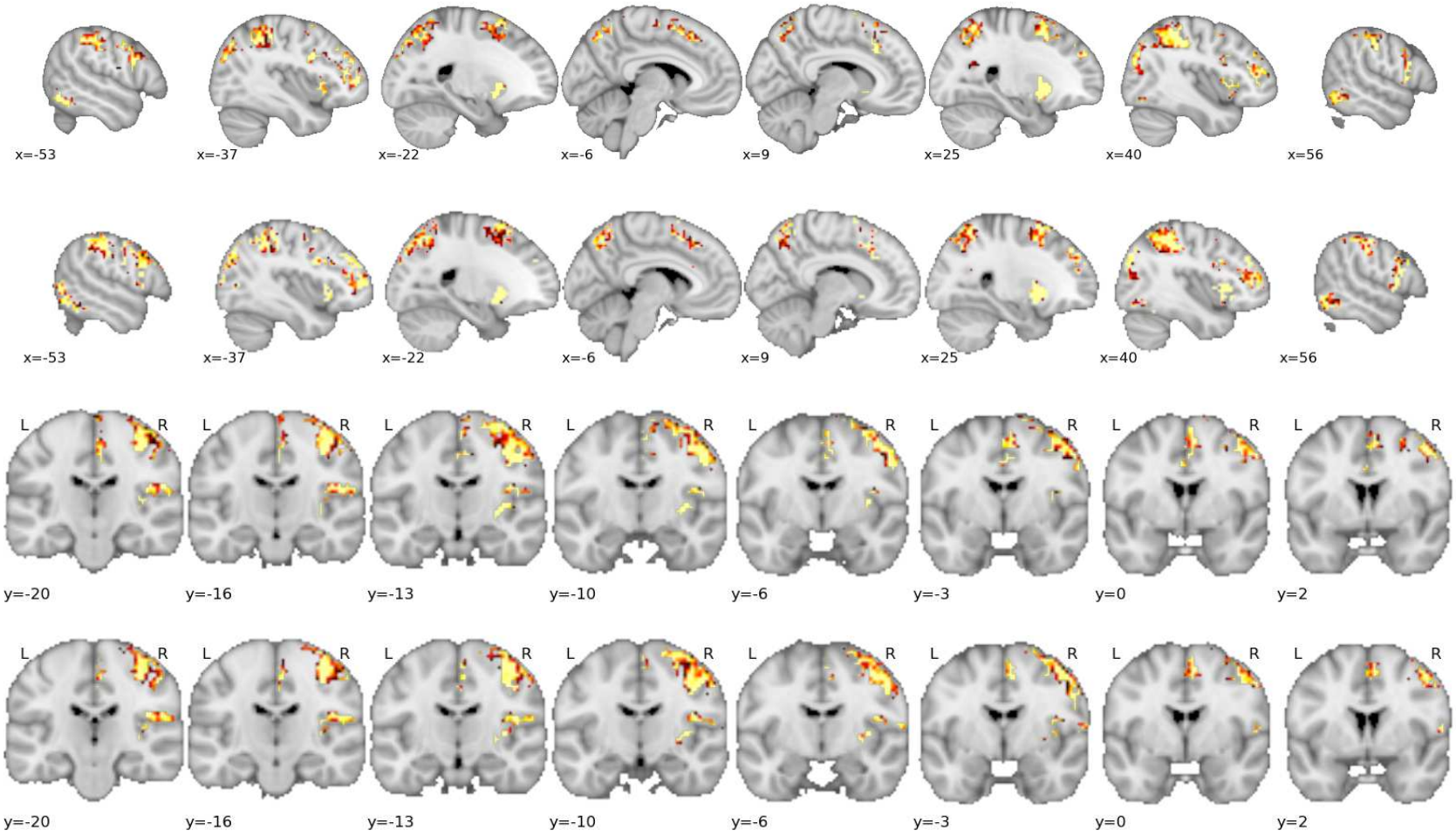


FIGURE 4.2: Estimated functional templates on real data for the experiments that use only one fMRI (the first and third rows) and the experiment that uses two fMRIs (the second and fourth rows). The yellow/red colors correspond to high/low probability of the activation for the computation task (the first two rows) and the left motor task (the third and fourth rows).

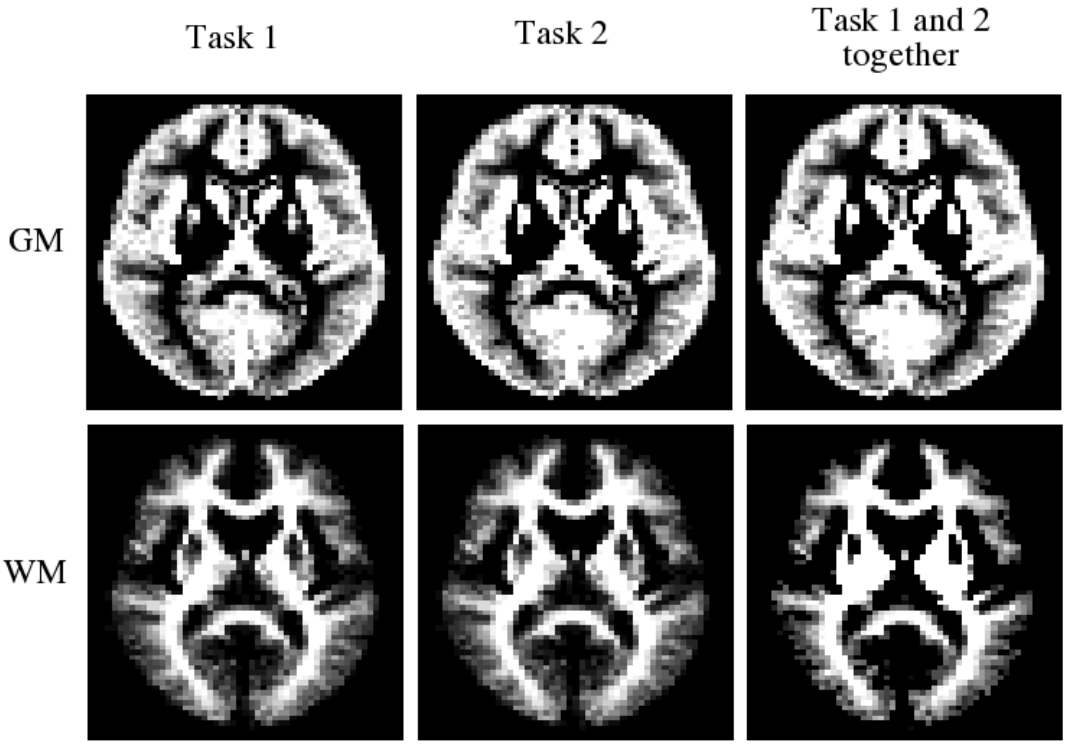


FIGURE 4.3: The estimated anatomy template for the experiments use only one fMRI (the first two columns) and the experiment uses two fMRIs (the third column).

4.3.2 In-vivo data.

The proposed method was also tested on a real MRI and fMRI dataset described in [103]. Both anatomical and functional data were subject to standard preprocessing using SPM8, including spatial normalization and General Linear Model analysis. Images are sampled at 3mm resolution, yielding volumes of shape $46 \times 53 \times 63$. We select a contrast from the fMRI that yields differential effect of a computation task versus a simple instruction reading/listening. We have $K = 3$ levels of activation in the GM.

We take 792 fixed control points for the deformation model given in Eq. (4.2), corresponding to one control points in each $6 \times 6 \times 6$ cube and $23 \times 27 \times 32$ points in the image as the landmarks for the template model given in Eq. (4.3), corresponding to one landmark in each $2 \times 2 \times 2$ cube.

At first, we use $s = 2$ that means we use the T1 image and one fMRI. The estimated functional probabilistic maps, thresholded at the $p > .95$ level, are shown in the first two rows of Fig. 4.2. The yellow/red colors correspond to high/low probability for the computation task activation (the first row) and the motor left task activation (the third row). Our probabilistic maps are sharp. The detected areas are well conditioned to GM and fits the known active areas for the computation task. For example, in the slice $x = 25mm$ in the first row, one can clearly see the Putamen. The anatomical template is

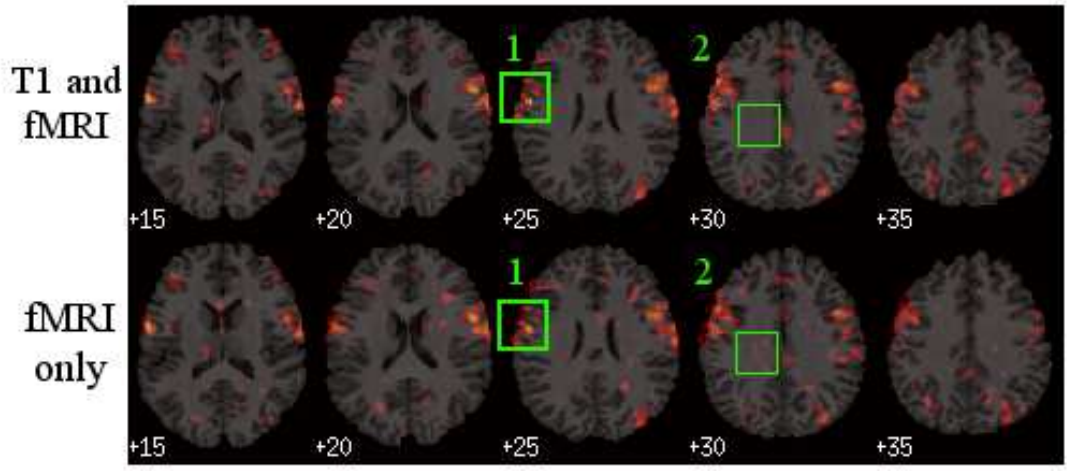


FIGURE 4.4: Experiments on real data showing the detected active areas $p > .95$ for the computation task. The first row for our method using both T1- and f- MRI and the second row for the standard method using fMRI only. Each column corresponds to one slice of the same patient.

shown in the first two columns of Fig. 4.3. The templates for the two tasks are similar. In Chapter 2, we have noticed that our model fails to segment the subcoritcal structures. However, we mange to segment these structures as GM with the information given by fMRI.

We compared our model with the standard method that thresholds the group-level mean activation. We represent the active areas in the computation task overlaid on T1 images. The results of one patient are shown in Fig. 4.4. The first row for our method uses both T1- and f- MRI and the second row for the method uses fMRI only. Each column corresponds to one slice of the same patient. In zone 1, we see that the areas detected as active by our method are limited to the GM. However, a part of the detected active areas by the non-anatomically aware method are outside of the brain. In zone 2, the standard method detects some active areas in WM, while our method does not. These show that we reach our goal, i.e. the detected active areas are well conditioned to GM. The detected active areas by our method are similar to those by the standard method in GM, this shows that our segmentation is accurate.

As the second experiment, we use $s = 3$ that means we use the T1 image and two fMRIs. The estimated functional probabilistic maps are shown in the second and fourth rows of Fig. 4.2. The results are smoother than the one use only one fMRI. The anatomical templates are shown in the third column of Fig. 4.3. The templates are similar to these obtained with one fMRI, however a little smoother, especially for the GM near the ventricles.

4.4 Conclusion

In this study, we proposed a statistical model to detect the active areas in the brain using both T1 and functional MRI. We used a stochastic algorithm to perform registration, segmentation and to create a probabilistic atlas simultaneously. Our model has several advantages. First, the probabilistic atlas contains both the templates and the geometric variability of the population. Second, we do not need any pre-registration to perform the segmentation which is automatically obtained as an output. Third, the detected active areas are confined to GM with the information provided from the MRI data. Our experiments show that we get better results with our algorithm than the standard method. The detected active areas are well conditioned to GM and the atlas is sharp.

Chapter 5

Including Shared Peptides for Estimating Protein Abundances: A Significant Improvement for Quantitative Proteomics.

Inferring protein abundances from peptide intensities is the key step in quantitative proteomics. The inference is necessarily more accurate when many peptides are taken into account for a given protein. Yet, the information brought by the peptides shared by different proteins is commonly discarded. We propose a statistical framework based on a hierarchical modeling to include that information. Our methodology, based on a simultaneous analysis of all the quantified peptides, handles the biological and technical errors as well as the peptide effect. In addition, we propose a practical implementation suitable for analyzing large datasets. Compared to a method based on the analysis of one protein at a time (that does not include shared peptides), our methodology proved to be far more reliable for estimating protein abundances and testing abundance changes. The source codes are available at http://pappso.inra.fr/bioinfo/all_P/.

5.1 Introduction

One of the main application of proteomics is the quantification of large and complex sets of proteins extracted from biological samples [49]. Among the available quantitative methods [120], the MS based technologies have become increasingly popular. In these approaches, proteins are digested and the resulting mixture of peptides is separated and analyzed by LC-MS/MS. Data are recorded either as a number of tandem mass spectra per protein or as peptide intensities derived from peak areas [120]. Generally, these data are processed to estimate protein abundances and identify proteins responsive to treatments of interest. The estimation of protein abundances and the detection of abundance changes are necessarily more reliable when many peptides are considered for a given protein. Shared peptides constitute a relevant source of information, especially when proteins are represented by few proteotypic peptides (see Fig. 5.1). Shared peptides are common, particularly when genes are duplicated. In some databases, they represent over 50% of the peptides [105]. In most proteomic studies, these peptides are discarded because of the difficulty to deconvolve the information they carry. Several approaches were developed to address this issue in the case of spectral counting [57, 157] or peptide intensity data [32, 74, 85]. Among the latter, Bukhman et al. [32] proposed a statistical model to infer protein abundances from both shared and proteotypic peptide intensities. However, their model suffers from two drawbacks. First, the peptide intensities are supposed to be normally distributed whereas they are widely assumed to be log-normally distributed [105]. Second, it does not take into account experimental variability.

In this paper, we propose an improved statistical framework to estimate protein abundances from peptide intensities that (i) includes shared peptide information and (ii) takes into account experimental variability, allowing normalization of signal intensities across multiple samples. In addition, we propose a practical implementation suitable for handling large datasets. Our methodology considers data derived from label-free experiments but it can be applied to other LC-MS approaches. The source codes are available at http://pappso.inra.fr/bioinfo/all_P/. They can be easily reproduced and adapted to other experimental designs.

5.2 Method

Our statistical model, referred to as the all-proteins (all-P) model, is based on the following assumptions.

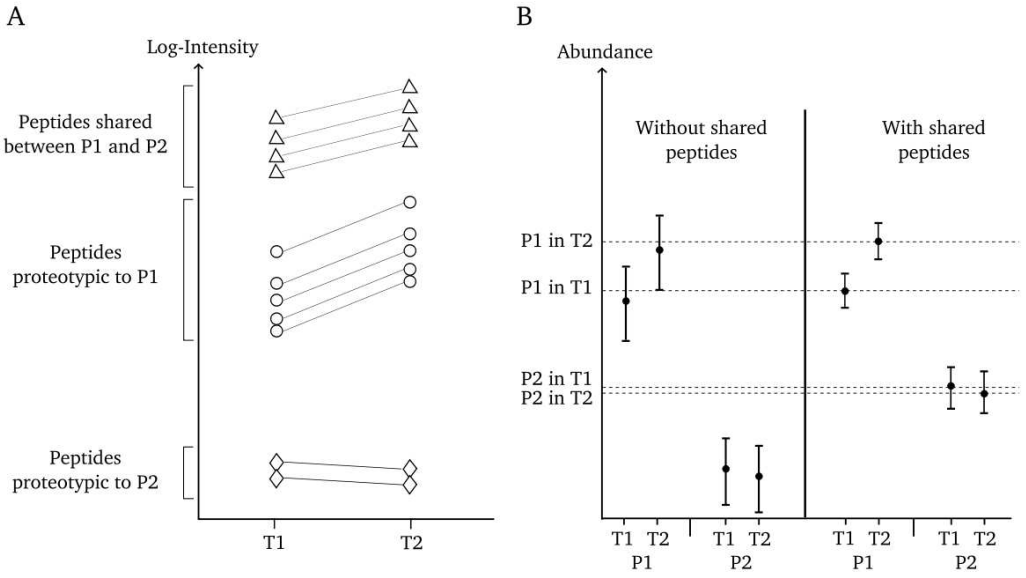


FIGURE 5.1: Relevance of shared peptides for quantitative proteomics illustrated by two proteins P1 and P2, quantified in two treatments, T1 and T2. (A) Intensities measured for shared and proteotypic peptides derived from P1 and P2. (B) Estimated (black dots) and real (dashed lines) protein abundances. Including shared peptides in protein quantification increased the power of detecting P1 abundance change between the treatments and allowed to better estimate P2 abundance.

(i) The abundance of a peptide is the sum of the abundances of the different proteins this peptide belongs to [32]:

$$a_i = \sum_k \delta_{ik} P_k \tag{5.1}$$

where a_i is the abundance of peptide i, P_k is the abundance of protein k, $\delta_{ik} = 1$ if peptide i belongs to protein k and $\delta_{ik} = 0$ otherwise.

(ii) The measured intensity of a peptide is proportional to its abundance in a sample, with a coefficient of proportionality depending on the peptide [32]. The measured intensity of a peptide is proportional to its abundance in a sample, with a coefficient of proportionality depending on the peptide [32].

$$I_i = \alpha_i a_i \tag{5.2}$$

where I_i = intensity of peptide i and α_i is the coefficient of proportionality for peptide i.

(iii) The intensity measures are assumed to be approximately log-normally distributed [105].

$$log(I_{ip}) = log(I_i) + E_{ip} \tag{5.3}$$

where I_{ip} is the pth intensity measured for peptide i (for instance different treatments and/or replicates), $E_{ip} \sim \mathcal{N}(0, \sigma^2)$ is the error. Combining Eqs.(5.1), (5.2) and (5.3), we get

$$log(I_{ip}) = log\left(\sum_k \delta_{ik} P_k\right) + D_i + E_{ip} \tag{5.4}$$

where the peptide effect $D_i = log(\alpha_i)$ reflects the LC-MS response of peptide i.

(iv) Experimental variability can be properly estimated by considering experimental parameters globally, for all the proteins [97]. To account for biological variability, a classical proteomic experiment containing the different treatments of interest is repeated several times independently. We thus decomposed the error term of Eq.(5.4) in an error B_r due to the biological variation of replicate r, an error C_{tr} due to the technical variation in treatment t and replicate r and a residual error ϵ_{itr}

$$E_{ip} = E_{itr} = B_r + C_{tr} + \epsilon_{itr} \tag{5.5}$$

where

$$B_r \sim \mathcal{N}(0, \sigma_B^2), C_{tr} \sim \mathcal{N}(0, \sigma_C^2), \epsilon_{itr} \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

(v) The goal of our model was to accurately estimate the protein abundance P_{kt} that, for this reason, was considered as a fixed effect. Conversely, the peptide effect D_i, of less interest, was regarded as random:

$$log(I_{ip}) = log\left(\sum_k \delta_{ik} P_{kt}\right) + D_i + B_r + C_{tr} + \epsilon_{itr} \tag{5.6}$$

where $D_i \sim \mathcal{N}(0, \sigma_D^2)$.

To recover a classical log-linear model for proteotypic peptides, we set $\theta_{kt} = log(P_{kt})$, leading to the all-P model

$$log(I_{itr}) = log\left(\sum_k \delta_{ik} \exp(\theta_{kt})\right) + D_i + B_r + C_{tr} + \epsilon_{itr} \tag{5.7}$$

Because of shared peptides, the all-P model is nonlinear. So the estimation of its parameters is not straightforward. To solve the problem of estimating the parameters θ_{kt}, σ_B^2, σ_C^2, σ_D^2, σ_ϵ^2, we used a Bayesian hierarchical framework that presented the advantage to be easier to implement than the other existing methods. It consists in characterizing the posterior distributions of the parameters given experimental data and non-informative prior information. The model is described in Fig. 5.2. Since the parameters θ_{kt} reflected biological values, we chose as prior distribution a centered Gaussian distribution with a variance large enough to be non-informative. The variances σ_B^2, σ_C^2, σ_D^2, σ_ϵ^2 being

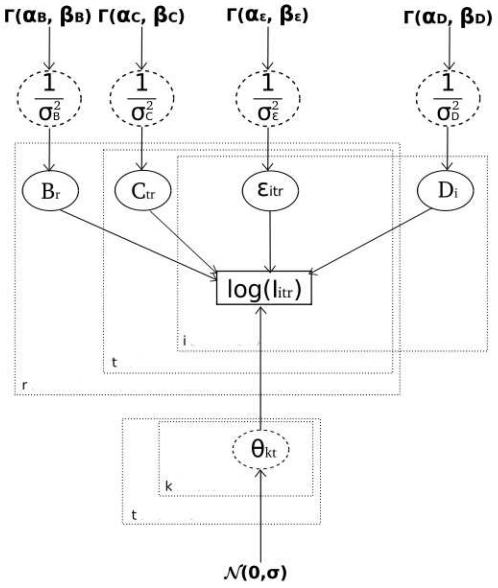


FIGURE 5.2: Directed Acyclic Graph for the Bayesian hierarchical model used in the all-P model. The indices k, i, t and r refer to proteins, peptides, treatments and replicates, respectively. Non-informative prior distributions (shown in bold) were assigned to the parameters θ_{kt}, σ_B^2, σ_C^2, σ_D^2 and σ_ϵ^2. An ergodic sample of the posterior distribution was generated by a Monte Carlo Markov Chain algorithm called Gibbs sampler.

positive values, we chose gamma laws as priors. To avoid unnecessary Central Unit Processing (CPU) time waste, the expectations of these gamma laws were chosen as the variances of D_i, B_r, C_{tr} and ϵ_{itr} estimated in the one-protein-at-a-time (one-P) model previously described [39]. Each parameter θ_{kt} was estimated by the empirical mean of the sample of the posterior distribution generated by JAGS [104]. From the classical asymptotic Bayesian theory, it is known that the distribution of this estimator is close to the distribution of the maximum likelihood.

The performances of the all-P model were compared to the performances of the one-P model. For each protein k, we have

$$log(I_{iktr}) = \theta_{kt} + \Delta_i + B_r + C_{tr} + \epsilon_{itr} \tag{5.8}$$

where I_{iktr} denotes the intensity of peptide i, proteotypic to protein k, observed in treatment t and replicate r and Δ_i is the peptide effect.

The parameters of this model were estimated for each protein k, only on the basis of proteotypic peptide intensities. Therefore, for some proteins, the number of observations might be too small to infer accurately the variances of the random effects. To overcome this drawback, the peptide effect Δ_i was considered as fixed. This model was

implemented in R [108] by using the *lme* function from the *nlme* package.

The procedure used to detect the proteins with differential abundance across the treatments was similar in the all-P and one-P models. Let T be the number of treatments. For each protein k, we tested the null hypothesis $H_k : \theta_{kt} = \theta_{kt'}$ for every $1 \leqslant t \leqslant t' \leqslant T$ against the alternative hypothesis A_k: there exists (t, t'), $t \neq t'$, such that $\theta_{kt} \neq \theta_{kt'}$. The Wald test based on a χ^2 distribution with (T-1) degrees of freedom is classically used to perform this test. However, the computation of the Wald statistic in the all-P model is rather complicated and may be instable when T is large. For these reasons, for testing H_k against A_k, we used a multiple testing procedure based on the following statistic:

$$s_k = \max_{t,t'} \{s_{k(t,t')}\} \tag{5.9}$$

where

$$s_{k(t,t')} = (\hat{\theta}_{kt} - \hat{\theta}_{kt'})^2 / \hat{\nu}_{k(t,t')},$$

$\hat{\nu}_{k(t,t')}$ is an estimator for the variance of $(\hat{\theta}_{kt} - \hat{\theta}_{kt'})$ calculated from the sample of the posterior distributions in the all-P model and given in the *lme* outputs in the one-P model. Under H_k, the distribution of $s_{k(t,t')}$ is approximated by a χ^2_1 distribution. The probability π_k to wrongly decide that the abundance of protein k changed between treatments t and t' is estimated by

$$\pi_k = P\left[\chi^2_1 > \max_{t,t'} \{s_{k(t,t')}\}\right]. \tag{5.10}$$

Since a large number of proteins had to be tested simultaneously, we applied the BenjaminiHochberg procedure [24] to decide which hypotheses H_k we rejected. To have a fair control of the false discovery rate (FDR), we had to adjust each π_k to $\tilde{\pi}_k = T(T-1)\pi_k/2$ so that $\tilde{\pi}_k$ satisfied $P(\tilde{\pi}_k \leqslant x) \leqslant x$.

5.3 Material

Original yeast proteome dataset

Four monosporic derivates obtained from two *S. cerevisiae* strains (VL1 supplied by LAFFORT Œnologie, Bordeaux, France and NRRL-Y-7327 supplied by ARS/NRRL culture collection, Peoria, Illinois, USA) and two *S. uvarum* strains (BR20.1 supplied by ADRIA NORMANDIE, Villers-Bocage, France and LC3 supplied by ISVV, Faculté d'Œnologie, Villenave d'Ornon, France) were inoculated in the Sauvignon must at 10^6

cells *per* mL and grown in anaerobic culture at 18°C. This experiment was repeated three times independently. Five mL of fermentative media were harvested when 30% of the fermetation was completed. Proteins were extracted in TCA-β-mercaptoethanol in acetone, denatured in urea, reduced, alkylated and digested with trypsin. LC-MS/MS analyses were performed using an Ultimate 3000 LC system (Dionex) connected to an LTQ Orbitrap mass spectrometer (Thermo Electron). Ionization was performed with a 1.3-kV spray voltage applied to an uncoated capillary probe. Peptide ions were analyzed using Xcalibur 2.0.7 (Thermo Electron). Dynamic exclusion was set to 90s. A custom FASTA format database of 5885 sequences of *S. cerivisiae* and 4966 sequences of *S. uvarum* downloaded from the Saccharomyces Genome Database website (http://downloads.yeastgenome.org) was searched by using X!Tandem (version 2010.01.01.4) (http://www.thegpm.org/TANDEM). The decoy database comprised he reverse protein sequences of the custom database. False discovery rate was less than 1% for both peptide and protein identification. Peptide intensities were quantified by integration of their peak area by using MassChroQ software as described by Valot et al. Protein abundances were estimated by the all-P model described in this paper.

Synthetic yeast proteome dataset

The parameters estimated from the original data using the all-P model were used together with Eq.(5.7), the design matrix and the peptide-protein relationships to generate synthetic datasets. To reduce the CPU time, we arbitrarily chose 100 proteins among those quantified in the original dataset. For 50 of them that exhibited significant abundance change between yeast strains, estimated abundances $\hat{\theta}_{kt}$ were kept unchanged. For the 50 remaining proteins, $\hat{\theta}_{kt}$ was replaced by $\tilde{\theta}_{kt} = 1/T(\sum_{t=1}^{t=T} \hat{\theta}_{kt})$. Hence, we expected to find 50 proteins exhibiting significant abundance changes between the treatments. In the 100 proteins, 58 of them were only represented by proteotypic peptides (type P proteins) and 42 by shared and proteotypic peptides (type S proteins).

Human yeast proteome dataset

The human yeast proteome dataset was obtained from the Clinical Proteomic Technology Assessment for Cancer (CPTAC) study 6. Forty eight human proteins (Sigma UPS1) were spiked in five different amounts (0.25, 0.74, 2.2, 6.7 and 20 fmol/μl) in a yeast reference proteome (60ng/μl). Samples were all prepared at the National Institute for Standards and Technology (NIST) and then distributed in five laboratories for MS analyses on seven different mass spectrometers. Each sample was analyzed in triplicates on each instrument. Material and methods are detailed in. In the present study,

we used the datasets obtained on one LTQ-XL-Orbitrap (Thermo), one LTQ-Orbitrap (Thermo) and one LTQ-Orbitrap(Jamie Hill Intruments) in two different laboratories (site 65 and 86) for two different amounts of human proteins (6.7 and 20 fmol/μl). Raw datafiles were transformed to mzXML open source format using ReadW software (v 4.3.1, http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW). During transformation profile MS data were centroided. The FASTA file containing the human, yeast and contaminant protein sequences available on the CPTAC website was searched with X!Tandem (version 2010.01.01.4; http://www.thegpm.org/TANDEM/) with the following settings. Enzymatic cleavage was declared as a trypsin digestion with one possible misscleavage. Carboxyamidomethylation of cysteine residuals and oxidation of methionine residuals were set to static and possible modifications, respectively. Precursor mass precision was set to 20 ppm. Fragment mass tolerance was 0.5 Th. A refinement search was added with the same settings, except that semi-trypsic peptides and protein N-ter acetylations were also searched. Only peptides with an E-value smaller than 0.1 were reported.

Identified proteins were filtered and sorted by using the X!Tandem pipeline (http://pappso.inra.fr/bioinfo/xtandempipeline/). Criteria used for protein identification were: *i.* at least two different peptides identified with an E-value smaller than 0.05. *ii.* a protein E-value (product of proteotypic peptide E-values) smaller than 10^4. To take into account that the same peptide sequence can be found in several proteins, proteins sharing at least one peptide were gathered in groups generally corresponding to proteins of similar functions. Within each group, proteins with at least one proteotypic peptide were reported as sub-groups. Peptides were quantified based on extracted ion chromatograms using MassChroQ software.

5.4 Results

We compared the performances of the estimation procedures used in the all-P and one-P models with simulation experiments. A total of 122 synthetic data sets were generated from the same original data. For 22 of them, the algorithm used in the one-P model did not converge for some proteins, probably owing to the small number of observations and to the model complexity (with fixed and random effects). This is in itself a limitation for the use of the one-P model. The all-P model did not present this drawback since it relies on all proteins to estimate the parameters. However, it was more time consuming, taking three days to complete one simulation on a computer with an Intel Core Xeon W3520 processor running at 2.60 GHz with 4 GB of RAM.

	All proteins (×1000)	Proteins of type S (×1000)	Proteins of type P (×1000)
One-P	2.6 (0.5)	2.5 (0.8)	2.7 (0.6)
All-P	2.1 (0.3)	1.4 (0.3)	2.6(0.5)

TABLE 5.1: Means and standard-errors (in parentheses) of mean squared error (MSE) calculated over 100 simulations, for all proteins, for proteins quantified from both shared and proteotypic peptides (type S), and for proteins quantified from proteotypic peptides only (type P). The means and standard-errors are averaged over all pairs (protein k, treatment t)

The performances of the one-P and all-P models for estimating protein abundances in the 100 remaining synthetic data sets were assessed by the mean squared error (MSE). MSE was defined, for each estimator $\hat{\theta}_{kt}$, as the expectation of $[(\hat{\theta}_{kt} - \theta_{kt})/\theta_{kt}]^2$ (Table 5.1). The all-P model outperformed the one-P model, since the mean of the MSE calculated over 100 simulations was 19.2% lower in the all-P model. This result illustrates the instability of the one-P model. The estimation of the variance of the random effects B_r and C_{tr} for each protein in the data set, with sometimes few observations per protein contributed to its poor performance (see Fig.5.3). The difference between the two models was more important for the type S proteins (Table 5.1), showing that including the shared peptides in the model improved by 44% the accuracy of the estimators. Because the testing procedure relied on the assumption that the statistics $(\hat{\theta}_{kt} - \hat{\theta}_{kt'})/\sqrt{\hat{\nu}_{k(tt')}}$ were distributed as centered standardized Gaussian variables, we checked the normality of these quantities (see Fig. 5.3). In the one-P model, their distribution was dramatically more dispersed than standardized Gaussian variables, while in the all-P model, their distribution satisfied the assumption.

The performances of the one-P and all-P models for detecting the proteins of variable abundance are shown in Fig. 5.4 A as curves of power (proportion of proteins actually variable that were declared significantly variable) versus FDR (proportion of proteins that were wrongly declared significantly variable). These quantities were estimated by averaging on the 100 simulated experiments. Again, the all-P model clearly outperformed the one-P model. For example, for a FDR below 5%, the all-P model allowed to increase the power by more than 10% compared with the one-P model. When considering the proteins of type S and type P, the gain in power obtained using the all-P model was superior by 14% and 6%, respectively. For α (level of threshold of the adjusted p-values) fixed at 0.05, the all-P model resulted in a gain of power (92.5% against 89.6%) and FDR (1.8% against 10.2%) for proteins of type S. Conversely, the one-P model showed a higher power for proteins of type P (82.6% against 77%) but this was compensated by a higher FDR (6.8% against 2.3%). The high FDR values obtained using the one-P model are easily explained by the discrepancy between the distribution of the test statistics and the Gaussian distribution previously shown (see Fig. 5.2). Indeed,

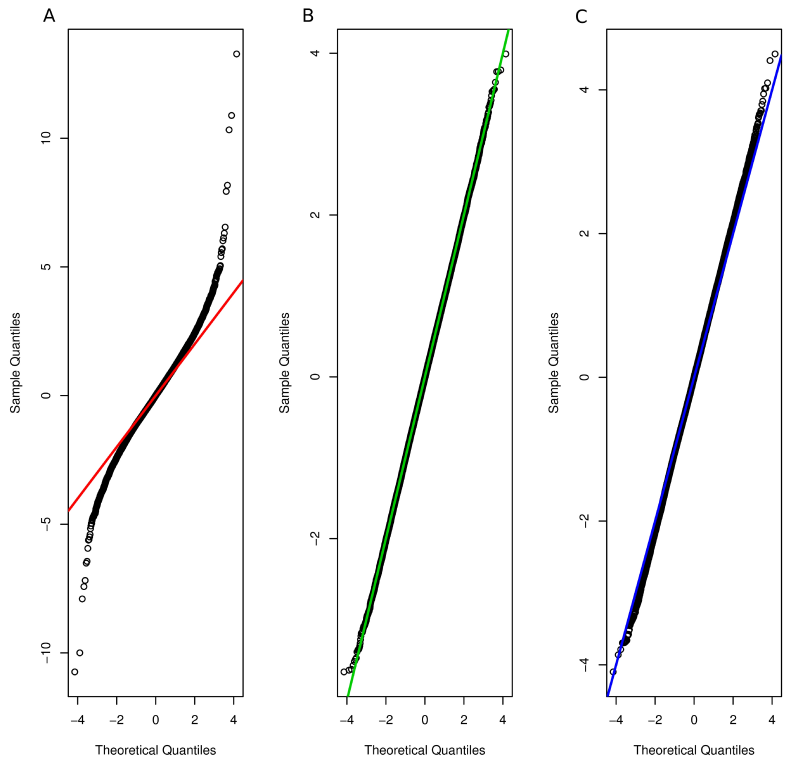


FIGURE 5.3: Synthetic dataset. Normal QQ plots of the statistics $(\hat{\theta}_{kt} - \hat{\theta}_{kt'})/\sqrt{\hat{\nu}_{k(t,t')}}$ for the one-P (A) and all-P models (B). (C) same as (A), but the difference $(\hat{\theta}_{kt} - \hat{\theta}_{kt'})$ resulting from the one-P model was normalized by the standard deviation $\sqrt{\hat{\nu}_{k(t,t')}}$ estimated by the all-P model for the proteins showing no differential abundance. This graph indicates that the bad fit observed in (A) is due to a bad estimation of the variance of the random effects by the one-P model.

the quantiles of the distribution of the statistics were larger than those of a Gaussian distribution, leading to π_k values smaller than expected. Finally, these results show, as expected, that the advantage of the all-P model over the one-P model was greater for proteins of type S.

To validate the performances of the all-P model, we analyzed a publicly available real data set [100], in which 48 human proteins were spiked in a yeast reference proteome in two amounts (i.e. treatments) differing by a one-to-three ratio (Supporting Information Material and Methods). Each humanyeast proteome mix was analyzed on three different mass spectrometers (i.e. replicates). A total of 763 proteins, of which 41 human, were reproducibly quantified by using the MassChroQ software [138]. Among them, 9.8% of the human proteins and 12.9% of the yeast proteins were of type S (Table 5.2). To validate our choice to consider peptide intensities as log-normal, we analyzed the residuals obtained in all-P model. The results confirmed the absence of any particular structure, except a very slight increase of absolute residuals versus fitted values (see Fig. 5.5). The performances of the one-P and all-P models for detecting the human proteins are shown in Fig. 5.4 B. For α=0.05, compared to the one-P model, the all-P model

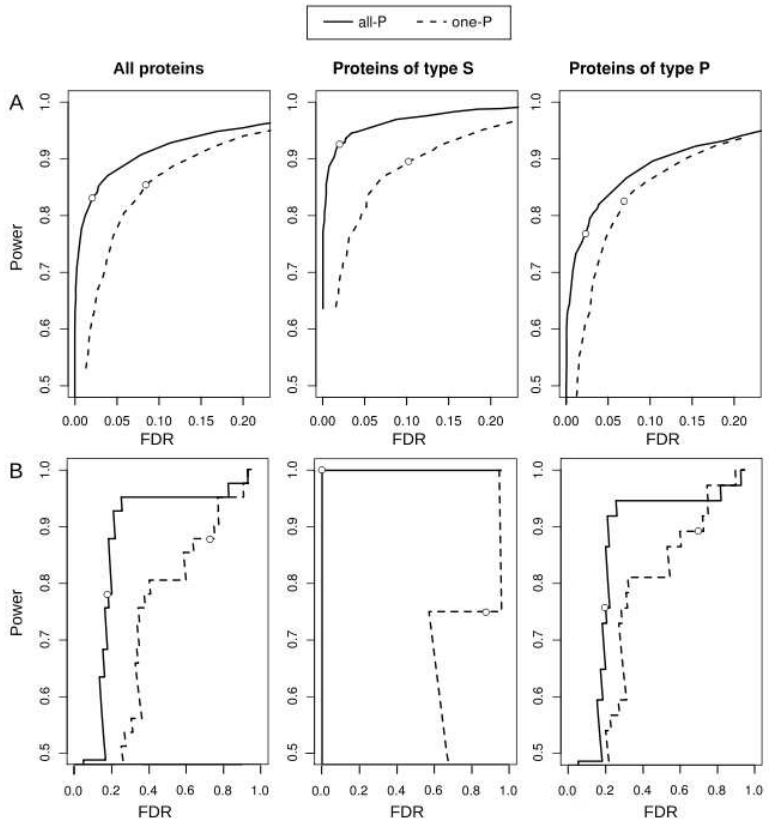


FIGURE 5.4: Power *versus* FDR estimated for all proteins, for proteins with both shared and proteotypic peptides (type S) and for proteins with proteotypic peptides only (type P) with the synthetic dataset (A) and the human-yeast dataset (B). The curves were generated by varying α. Circles indicate the power and FDR values obtained for α=0.05.

showed a much lower FDR (0% against 87% for proteins of type S; 20% against 70% for proteins of type P), a better power for proteins of type S (100% against 75%) and a lower power for proteins of type P (75.7% against 89.2%). These results thus confirm those obtained with the synthetic data sets.

5.5 Conclusion

In conclusion, the all-P model presented in this paper proved to be far more reliable and more powerful than the one-P model. We particularly highlighted the very poor performances of the one-P model regarding the FDR. Our results also point out that the information carried by shared peptides is highly valuable and has to be considered in the data analyses. The all-P model is time consuming, but a computer automatically does it, and in any case, this time is negligible compared to the time required to design a project and perform the experiments. Moreover, with the development of high throughput technologies that generate spectacular amounts of data, this type of statistical analysis is expected to become increasingly useful in biology.

	Number of proteins	Number of proteins with shared peptides	Number of shared peptides	Number of proteotypic peptides
Synthetic dataset				
Whole dataset	100	42	122	712
Set of variable proteins	50	19	72	545
Set of invariable proteins	50	23	91	167
Human-yeast dataset				
Synthetic dataset				
Whole dataset	763	97	321	3571
Set of variable proteins	41	4	7	113
Set of invariable proteins	722	3	321	3460

TABLE 5.2: Composition of the synthetic and human-yeast datasets. Note that a peptide can be shared between variable and invariable proteins.

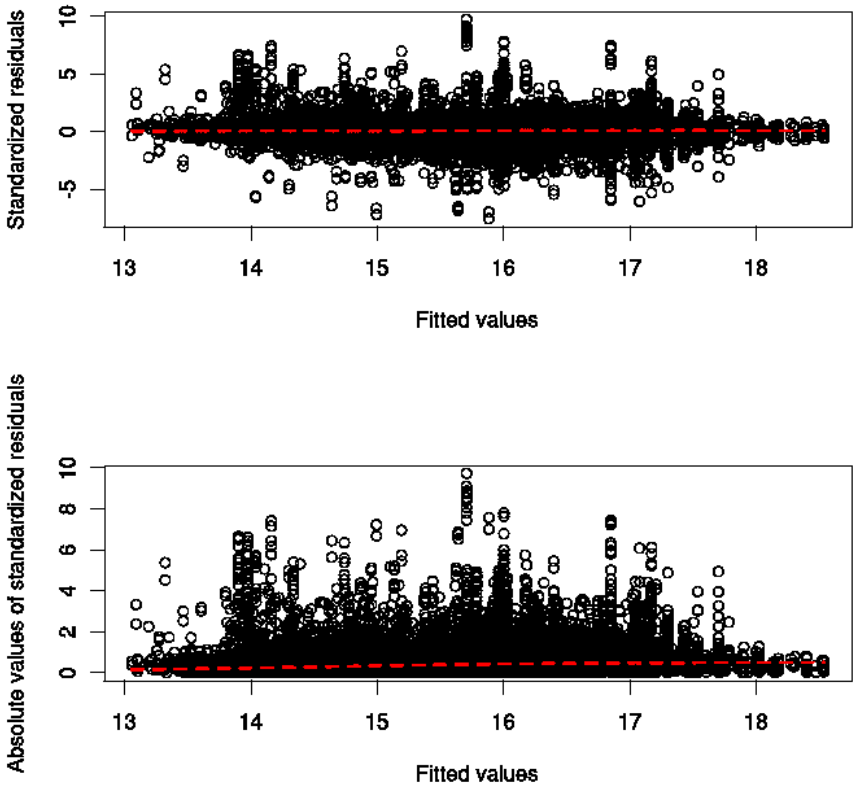


FIGURE 5.5: Human-yeast data set. Graphic of standardized residuals $\tilde{\epsilon}_{itr}$ corrected from the peptide random effect versus fitted values $\hat{\mu}_{it}$: $\hat{\mu}_{it} = \log(\sum_k \delta_{it}\exp(\hat{\theta}_{kt}))$, $\hat{\epsilon}_{itr} = \log(I_{itr} - \hat{\mu}_{it} - \hat{D}_i$ and $\tilde{\epsilon}_{itr} = [\log(I_{itr} - \hat{\mu}_{it} - \hat{D}_i]/sd(\hat{\epsilon}_{itr})$. The red dotted line represents a local polynomial smoother of the scatter plot. This graphic shows that the distribution of the residuals is heavy tailed, but does not show any particular structure, except a very slight increase of absolute residuals versus fitted values.

Chapter 6

Conclusion and Discussion.

We summarize here our contributions and propose some open questions for future work.

6.1 Summary

In this work, we have first proposed a generative statistical model for estimating a probabilistic atlas that includes both the photometric and geometric characterization of the observed population. The photometric information is given by a probabilistic template that contains each tissue probability map. This model makes it possible to take into account the uncertainty on the underlying tissue type, which is known as the partial volume effect (PVE). The geometrical characterization of the population is given by the covariance matrix of the template-to-subject deformation distribution. This highlights both what would be the normal variability in this population and the correlations underlying the deformations across different parts of the image.

The atlas estimation is based on non-registered and non-segmented data. In our estimation process, the registration and segmentation are performed jointly along the atlas estimation. Compared to the method that performs them sequentially, it increases the relevance between these two important and correlated steps and also increase the accuracy of our atlas.

The estimation algorithm is based on a Stochastic Approximation Expectation Maximization (SAEM) algorithm which is coupled with a Metropolis within Gibbs sampler to provide an efficient sampling procedure.

This has been developed in Chapter 2.

The estimated atlas does not only provide a characterization of the observed population: it can also be used to perform an atlas-based segmentation. The new subject image is a gray level observation drawn from the same population (for example healthy controls) as our training sample. This image has neither been pre-segmented nor pre-registered to any template image. We first construct a gray level template thanks to our estimated tissue probability maps and mean gray levels of each tissue. This provides us with a gray level template image. Thanks to the estimated geometric variability, we can constrain the gray level template-to-new-subject deformations to those that are common in the population. To this purpose, we introduce the estimated covariance matrix as a penalty term for the deformation and add this term to the usual data attachment term. This results in a classic matching energy to minimize. The registration is therefore obtained by minimizing this energy. The segmentation is then obtained by transporting the template segmentation to the observation frame by the estimated deformation. Our posterior segmentation method of new observations is fast and gives an accurate result (see Chapter 3).

We have then generated our single modality model to a multi-modal one (Chapter 4). The inputs are now multivariate, as they encode multi-modal patient observations (gray level T1 and functional MRIs). In this case, we have managed to impose meaningful biological constraints, such as the fact that the activation appears only in the gray matter. We got both the activation probability maps and the tissue probability maps as our probabilistic template. Thanks to this model, we were able to improve the accuracy of the estimated template as well as estimated geometrical variability; in particular, we better localized the activation patterns. We have tested our model on simulated data, which has shown the good performance with a quantitative test. We also managed to do experiments on real data. The results appear to better approach to the ground truth by coupling registration-segmentation and anatomo-functional images.

Concerning the theoretical study of these estimators on the proposed models and the estimation algorithms, we have proved the existence and the consistency of the estimator on our first model. We have also proved that our algorithm converges to this estimator. These two asymptotic properties enabled us to increase the confidence we have on our estimates which are also validated by the experiments.

In Chapter 5, we have proposed a model to analysis all the protein at same time (include shared peptides). We have proposed to use Gibbs samplers for handling large datasets. Compared to a method based on the analysis of one protein at a time (that does not include shared peptides), our methodology proved to be far more reliable for estimating protein abundances and testing abundance changes.

6.2 Large deformations for deformable template estimation

In this work, we use the small deformations that is to say, deformations that are driven by the displacement of each point. This writes: for all x in the domain of interest D:

$$\varphi(x) = x + \nu(x)\,, \tag{6.1}$$

where ν is a velocity vector field with fixed regularity.

As mentioned in the Introduction, these deformations, although very meaningful at first sight, are not optimal, as they are not constrained to be diffeomorphic. In the

medical applications that we target, this may be critical as it may delete some tissue parts by creating some overlaps. An improvement would be to consider diffeomorphic deformations as in [134] or [140]. This control on the deformations would help to respect anatomical constraints.

In particular, our model would become the following in the Large Diffeomorphic Deformation Metric Mapping (LDDMM) setting introduced in [134]. We will only consider here the generalization of the single modality model in Chapter 2, the multi-modal model is straightforward.

The probability of observing a data with intensity y_i^j for the ith image in the jth voxel given that it belongs to the kth class ($c_i^j = k$) remains defined as follows:

$$\mathbb{P}(y_i^j | c_i^j = k, \mu_k, \sigma_k^2) \sim \mathcal{N}(y_i^j; \mu_k, \sigma_k^2), \tag{6.2}$$

For $j \in \Lambda$ the prior probability of a voxel j from subject i to be in the k^{th} class would now be given by:

$$\mathbb{P}(c_i^j = k) = \varphi_i \cdot P_k(x_j) = P_k((\varphi_1^v)^{-1}(x_j))\,. \tag{6.3}$$

where φ_1^v is the solution of the diffeomorphic flow equation for a velocity vector field $(\nu_t)_{t\in[0,1]}$:

$$\begin{cases} \dfrac{\partial \varphi_t}{\partial t} = \nu_t \circ \varphi_t; \\ \\ \varphi_0 = Id. \end{cases} \tag{6.4}$$

The velocity vector field v_t should belong for all time $t \in [0,1]$ to a smooth deformation vector space V in order to ensure that there exists a unique solution to the flow equation and that the solution at $t = 1$ is a diffeomorphism. The space V is defined as a RKHS for which stating the kernel enables to easily deal with the smoothness of its elements.

This LDDMM formulation has the drawback that it requires to parametrize the deformation with the dense time dependent vector field $(v_t)_{t\in[0,1]}$. It has been shown in [52] that an approximation of the deformation can be done by considering that the initial velocity vector field is driven by the displacement of some control points. Thanks to this formulation and due to the theoretical properties of the generated deformation along the geodesic path [95], this parametrization with a finite number of control points is preserved along time.

Let us consider some given control points $(x_0^k)_{1\leqslant k\leqslant k_g}$ and initial momentum vectors $\boldsymbol{\beta}_0(k)$, then the initial velocity vector field writes:

$$\forall x \in D, v_0(x) = (\mathbf{K_g}(\boldsymbol{x_0})\boldsymbol{\beta}_0)(x) = \sum_{k=1}^{k_g} K_g(x, x_0^k)\boldsymbol{\beta}_0(k)\,, \tag{6.5}$$

where K_g is the RKHS kernel. Along the geodesic we get:

$$\forall x \in D, v_t(x) = (\mathbf{K_g}(\boldsymbol{x_t})\boldsymbol{\beta}_t)(x) = \sum_{k=1}^{k_g} K_g(x, x_t^k)\boldsymbol{\beta}_t(k)\,, \tag{6.6}$$

where the couple of vectors $(\boldsymbol{x_t}, \boldsymbol{\beta}_t)$ satisfies the Hamiltonian system:

$$\begin{cases} \dfrac{dx_t(k)}{dt} = \displaystyle\sum_{l=1}^{k_g} K_g(x_t(k), x_t(l))\beta_t(l) \\ \dfrac{d\beta_t(k)}{dt} = -\left(\displaystyle\sum_{l=1}^{k_g} d_{x_t(k)}(K_g(x_t(k), x_t(l))\beta_t(l))\right)^t \beta_t(k) \end{cases} \tag{6.7}$$

Using this formulation, the diffeomorphic deformation can be parametrized as in the small deformation setting by a finite vector which contains initial momenta. This formulation makes it possible to also take into account in our estimation process the initial positions of the control points $\boldsymbol{x_0}$. This has been recently studied in [11] for the gray level template estimation and could be applied in our models. The statistical model would look therefore very similar to the one presented in a previous chapter. Let $\boldsymbol{\gamma}_0 = (\boldsymbol{x_0}, \boldsymbol{\beta}_0))$ then:

$$\begin{cases} (\Gamma_g, \mu_k, \sigma_k^2) \sim \nu_g \otimes \nu_m \otimes \nu_p; \\ (\boldsymbol{\gamma}_0)_i \sim \mathcal{N}(0, \Gamma_g)|\Gamma_g; \\ c_i^j \sim \displaystyle\sum_{k=1}^{K} \delta_k \varphi_i^{(\boldsymbol{\gamma}_0)_i} \cdot P_k(x_j)|(\boldsymbol{\gamma}_0)_i; \\ y_i^j \sim \mathcal{N}(\mu_k, \sigma_k^2)|c_i^j = k, \mu_k, \sigma_k^2, \end{cases} \tag{6.8}$$

where δ_k is a Dirac measure on k and ϕ is .

After calculating the log-likelihood, we get the same sufficient statistics S_0, S_1, S_2, S_3 and S_4 (with an additional one if we optimize the position of initial control point as well) given in the section 2.4.1. The same SAEM-MCMC algorithm may be used with an optimization of the MCMC sampler as the dimension increases (see [12] for one example).

6.3 Multicomponent generalization of the models

Each of our models is designed for an homogeneous population. However, many of the acquired data bases may contain several clusters that are still labeled control for example, yet carry different but normal patterns. This may be easily treated in our context by considering mixtures of our models. The observed images of the data base are not pre-clustered so that their component label inside the population is not known. This same framework has been introduced in [13] for the gray level template estimation problem in the small deformation setting. This adds a new unobserved variable for each image which is the component label.

Although quite natural and easy to write and study theoretically, this mixture model requires to change the MCMC method. As this sample step is the bottleneck of our estimation algorithm, considering this generalization would require to optimize (in a way different from the diffeomorphic case) the MCMC sampler.

6.4 Other remarks

6.4.1 Extension of the multi-modal atlas

In our multi-modal generative model, we have introduced the functional MR images in order to get an anatomo-functional atlas that satisfies some biological constraints. In the same way, one can include complementary information given by the fiber directions of the Diffusion Weighted Images (DWIs). These images provides the local directions of the fibers in the brain which form the white matter. The white matter appears with the same gray level in the MRIs therefore the registration mappings are not constrained at all in these regions. This is the same issue when considering the gray matter without the functional information.

To overcome this issue, following the same lines of our anatomo-functional brain atlas, we can add a third input image - as the Fractional anisotropy obtained from a set of DWIs - in our input vector. The registration will be constrained now by all the input together forcing some particular deformations in all regions of the brain depending on the gray or white matter tissue. This extension can be inspired by the ideas in [35, 88, 114, 125] among others.

This generalization will have to be carefully done because of the large memory this may require to deal with so many images per subjects.

6.4.2 Kernel choice

One remark is the choice of the interpolation kernels. For the template model in section 2.5, the templates are probability maps, which yields the following constraint

$$\forall x \in D, \sum_{k=1}^{K} P_k(x) = 1\,. \tag{6.9}$$

Using a step function (described by the nearest neighbor interpolation) as the interpolation kernel enables to rewrite this as $\forall l \in [\![1, k_p]\!], \sum_{k=1}^{K} \alpha_k^l = 1$. This reduces the constraints to a single finite dimensional constraint on our parameters $(\alpha_k^l)_{1\leqslant k\leqslant K,\ 1\leqslant l\leqslant|\Lambda|}$. However, we may prefer to use others in order for the template probability maps to be smoother and to reduce the dimension of the parameter vector.

Other choices can be done using for example regular kernels as Gaussian kernels. However, the dense constraint 6.9 has to be carefully taken into account so that this does not increase the computational time or the theoretical properties of the model and estimator.

We have made several attempts to keep an easy formulation. For example, we have tried to use all the neighbors (first in a deterministic way). Since we consider all the neighbors and calculate the mean, we get a smoother result and there are less isolated points than the model with the nearest neighbor. Although this enables to remove the isolated missegmented voxels, the structures with small volume are misclassified. This would require to better estimate the smoothness that we expect on the template maps. Moreover, this kernel needs more computation time.

Another attempt was to consider the spatial dependencies between class labels in the prior law $\mathbb{P}(c_i^j = k)$. One solution is defining a Markov random field on class labels. We have defined the grid with a neighborhood system and an Ising model on class label. Unfortunately, this drastically increased the computation time, whereas not improving the results very much. This may be due to the bad choice of the system and model, and we could investigate other Markov random field that would better fit our expectations.

6.4.3 Algorithm implementation optimization

A common problem for our two previous algorithms is the computation time. Our model has high dimensional parameters and high dimensional observed and unobserved

random variables. This makes the computation time long because of the need of memory and the need to sample from the posterior distribution. However our algorithm can be parallelized for this particular simulation step. We are currently working on a parallel C++ version of our code to make it possible to increase the training set size as well as the size of the training image data and decrease the computation time.

6.4.4 Bias field correction

In all this work, we have been considering MR images as our observations. Although never pre-registered nor pre-segmented, the images may sometimes not be used directly as they come out the scanner. Indeed, they suffer from what is know as the bias field. This effect is related to the high magnitude of the magnetic field that is applied to the patient for the imaging process. Its effect on the images is to make the contrast smoothly evolve from one corner to the opposite one. This affects significantly the gray levels of each tissues which, if not corrected, may lead to mis-segmentation. One direction of interest would be to include this bias field in the generative model in order to account for this effect in our atlas estimation.

Appendix A

Definition of the most used similarity measures.

We defined the most used similarity measures that quantify the similarity between two images A and B here.

The correlation coefficient :

Press [107] proposes the correlation coefficient of two images as the similarity measure. This metric calculates the correlation between intensity values divided by the square root autocorrelation of two images:

$$r_{cc} = \frac{\sum_{i=0}^{N-1}(A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i=0}^{N-1}(A_i - \bar{A})^2}\sqrt{\sum_{i=0}^{N-1}(B_i - \bar{B})^2}}, \tag{A.1}$$

where A_i, B_i are the values of the voxel i in the two images, $\bar{A}$, $\bar{B}$ are the mean values of the voxel in the two images, N is the number of the voxels.
If the registration is well done, the two images should be strongly correlated. This means that the registration is better if we have a high correlation of two images, i.e. the registration is the best when r_{cc} is 1, and the worst when r_{cc} is 0 or negative.

Minimum Mean Square Error (MMSE) :

Umbaugh [135] proposes the minimum mean square error of two images as the similarity measure. This metric is used as the default metric in most applications.

It calculates the sum of squared differences between intensity values:

$$MSE = \sqrt{\sum_{i=0}^{N-1} [A_i - B_i]^2}, \tag{A.2}$$

where A_i, B_i are value of the voxel i in the two images.
If the registration is well done, the gray level value of the voxel in the image A should be close to the one in the image B. This means that the smaller the minimum mean square error is, the better the registration is.

Mutual Information (MI) and Normalized Mutual Information (NMI) :

Collignon et al. [42] propose the Mutual Information as the similarity measure for registration. It is defined as

$$MI(A, B) = H(A) - H(B|A), \tag{A.3}$$

where $H(A)$ is the marginal entropy of the histogram of image A and $H(B|A)$ is the conditional entropy that gives the uncertainty about the gray level in B when the gray level in A is given.
The entropy gives the uncertainty of image voxel intensities. An image consisting of almost a single intensity will have a low entropy value; it contains very little information. A high entropy value will be yielded by an image with more or less equal quantities of many different intensities, which is an image containing a lot of information. MI is the amount of information A contains about B, i.e. their dependency. It is a symmetric measure: A and B can be interchanged:

$$MI(A, B) = H(A) + H(B) - H(A, B), \tag{A.4}$$

where $H(A)$, $H(B)$ are the marginal entropy of the two images and $H(A, B)$ is the joint entropy that calculated from the joint histogram of A and B.
This formula shows that maximizing mutual information is related to minimizing joint entropy. Registration aims at maximizing mutual information: the images have to be aligned in such a manner that the amount of information they contain about each other is maximal. The joint entropy is calculated for the joint histogram of the images and it is therefore sensitive to the size and the contents of overlap. An issue that can occur when using joint entropy on its own, is that low values can be found for complete misregistrations.
Studholme et al. [130] proposes to use the normalized mutual information to eliminate the effect of the overlap area. The normalization makes the alignment

measure sesnitive to overlap area. It is defined as

$$NMI(A,B) = \frac{H(A) + H(B)}{H(A,B)}, \tag{A.5}$$

where $H(A)$, $H(B)$ are the marginal entropy of the two images, and $H(A,B)$ is the joint entropy that calculated from the joint histogram of A and B.

Bibliography

[1] FMRIB Software Library (FSL), http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/

[2] Freesurfer, surfer.nmr.mgh.harvard.edu/

[3] Internet Brain Segmentation Repository (IBSR), http://www.cma.mgh.harvard.edu/ibsr

[4] MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labelling, http://masi.vuse.vanderbilt.edu/workshop2012/index.php/MainPage

[5] Open Access Series of Imaging Studies (OASIS), http://www.oasis-brains.org/

[6] Segmentation and Registration Toolkit (ITK) , www.itk.org

[7] Statistical Parametric Mapping (SPM8), http://www.fil.ion.ucl.ac.uk/spm/software/spm8/

[8] Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag, A.A., Member, S., Moriarty, T.: A modified fuzzy c-means algorithm for bias field estimation and segmentation of mri data. IEEE Transactions on Medical Imaging 21(3), 193–199 (2002)

[9] Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 46(3), 726 – 738 (2009)

[10] Allassonnière, S., Amit, Y., Trouvé, A.: Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(1), 3–29 (2007)

[11] Allassonnière, S., Durrleman, S., Kuhn, E.: Bayesian mixed effect atlas estimation under diffeomorphic constraint on the deformation model. (submitted)

[12] Allassonnière, S., Kuhn, E.: Convergent stochastic expectation maximization algorithm with efficient sampling in high dimension. application to deformable template model estimation. (submitted)

[13] Allassonnière, S., Kuhn, E.: Stochastic algorithm for bayesian mixture effect template estimation. ESAIM: Probability and Statistics 14, 382–408 (2010)

[14] Allassonnière, S., Kuhn, E., Ratnanather, T., Trouvé, A.: Consistent atlas estimation on bme template model: Applications to 3D biomedical images. In: PMMIA (2009)

[15] Allassonnière, S., Kuhn, E., Trouvé, A.: Construction of bayesian deformable models via stochastic approximation algorithm: A convergence study. Bernoulli Journal 16(3), 641–678 (2010)

[16] Andrade, A., Kherif, F., Mangin, J.F., Worsley, K.J., Paradis, A.L., Simon, O., Dehaene, S., Bihan, D.L., Poline, J.B.: Detection of fMRI activation using cortical surface mapping. Human Brain Mapping 12, 79–93 (2001)

[17] Andrieu, C., Moulines, E., Priouret, P.: Stability of stochastic approximation under verifiable conditions. SIAM Journal on Control and Optimization 44(1), 283–312 (2005)

[18] Ardon, R., Cohen, L.D.: Fast constrained surface extraction by minimal paths. International Journal of Computer Vision 69(1), 127–136 (2006)

[19] Ashburner, J.: A fast diffeomorphic image registration algorithm. NeuroImage 38(1), 95 – 113 (2007)

[20] Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage 26(3), 839–851 (2005)

[21] Bardinet, E., Bhattacharjee, M., Dormont, D., Pidoux, B., Malandain, G., Schpbach, M., Ayache, N., Cornu, P., Agid, Y., Yelnik, J.: A three-dimensional histological atlas of the human basal ganglia. ii. atlas deformation strategy and evaluation in deep brain stimulation for parkinson disease. Journal of Neurosurgery 110(2), 208–19 (2009)

[22] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal of Imaging Sciences 2(1), 183–202 (2009)

[23] Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (2005)

[24] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B 57, 289–300 (1995)

[25] Bhatia, K.K., Aljabar, P., Boardman, J.P., Srinivasan, L., Murgasova, M., Counsell, S.J., Rutherford, M.A., Hajnal, J., Edwards, A.D., Rueckert, D.: Groupwise combined segmentation and registration for atlas construction. In: MICCAI. pp. 532–540 (2007)

[26] Bhatia, K.K., Hajnal, J.V., Puri, B.K., Edwards, A.D., Rueckert, D.: Consistent groupwise non-rigid registration for atlas construction. In: ISBI. pp. 908–911 (2004)

[27] Blezek, D.J., Miller, J.V.: Atlas stratification. Medical Image Analysis 11(5), 443 – 457 (2007)

[28] Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9), 1124–1137 (2004)

[29] Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary amp; region segmentation of objects in N-D images. In: Proc. of Eighth IEEE International Conference on Computer Vision. vol. 1, pp. 105–112 (2001)

[30] Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.P., Osher, S.: Fast global minimization of the active contour/snake model. Journal of Mathematical Imaging and Vision 28(2), 151–167 (2007)

[31] Bro-Nielsen, M., Gramkow, C.: Fast fluid registration of medical images. In: Proceedings of the 4th International Conference on Visualization in Biomedical Computing. pp. 267–276 (1996)

[32] Bukhman, Y.V., Dharsee, M., Ewing, R., Chu, P., Topaloglou, T., et al.: Design and analysis of quantitative differential proteomics investigations using lc-ms technology. Journal of Bioinformatics and Computational Biology 6(1), 107–123 (2008)

[33] Campbell, N.A., Wu, X.: Gradient cross correlation for sub-voxel matching. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 1065–1070 (2008)

[34] Cauchy, A.L.: Méthode générale pour la résolution des systèmes d'équations simultanées. Compte Rendu des Séances de L'Académie des Sciences XXV Série A(25), 536–538 (1847)

[35] Ceritoglu, C.: Multichannel Large Deformation Diffeomorphic Metric Mapping and Registration of Diffusion Tensor Images. Ph.D. thesis (2008)

[36] Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing 10(2), 266–277 (2001)

[37] Charpiat, G., Faugeras, O., Keriven, R.: Shape metrics, warping and statistics. In: Proc. International Conference on Image Processing. pp. 627–630 (2003)

[38] Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable templates using large deformation kinematics. IEEE Transactions on Image Processing 5(10), 1435–1447 (1996)

[39] Clough, T., Key, M., Ott, I., Ragg, S., Schadow, G., Vitek, O.: Protein quantification in label-free LC-MS experiments. Journal of Proteome Research 8, 5275–5284 (2009)

[40] Cocosco, C.A., Kollokian, V., Kwan, R.K.S., Pike, G.B., Evans, A.C.: Brainweb: Online interface to a 3D MRI simulated brain database. NeuroImage 5, 425 (1997)

[41] Cohen, L.D., Cohen, I.: Finite-element methods for active contour models and balloons for 2-D and 3-D images. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), 1131–1147 (1993)

[42] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., G., M.: Automated multimodality image registration using information theory. In: IPMI. pp. 263–264 (1995)

[43] Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C.: Design and construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging 17(3), 463–468 (1998)

[44] Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance model. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 681–685 (2001)

[45] Cour, T., Gogin, N., Shi, J.: Learning spectral graph segmentation

[46] Dambreville, S., Niethammer, M., Yezzi, A., Tannenbaum, A.: A variational framework combining level-sets and thresholding. In: British Machine Vision Conference. p. 1–10 (2007)

[47] Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the EM algorithm. The Annals of Statistics 27(1), 94–128 (1999)

[48] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Serie B 39(1), 1–38 (1977)

[49] Domon, B., Aebersold, R.: Options and considerations when selecting a quantitative proteomics strategy. Nature Biotechnology 28(7), 710–721 (2010)

[50] Drapaca, C.S., Cardenas, V., Studholme, C.: Segmentation of tissue boundary evolution from brain MR image sequences using multi-phase level sets. Computer Vision and Image Understanding 100(3), 312–329 (2005)

[51] Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics 3(3), 32–57 (1973)

[52] Durrleman, S.: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Ph.D. thesis (2010)

[53] Durrleman, S., Allassonnière, S., Joshi, S.: Sparse adaptive parameterization of variability in image ensembles. International Journal of Computer Vision 101(1), 161–183 (2013)

[54] Fan, J., Zeng, G., Body, M., Hacid, M.S.: Seeded region growing: An extensive and comparative study. Pattern Recognition Letters 26(8), 1139–1156 (2005)

[55] Farneback, G.: Polynomial expansion for orientation and motion estimation. Ph.D. thesis (2002)

[56] Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International Journal of Computer Vision 59(2), 167–181 (2004)

[57] Fermin, D., Basrur, V., Yocum, A.K., Nesvizhskii, A.I.: Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. Proteomics 11, 1340–1345 (2011)

[58] Friston, K.J., Holmes, A.P., Price, C.J., Büchel, C., Worsley, K.J.: Multisubject fMRI studies and conjunction analyses. NeuroImage 10(4), 385–396 (1999)

[59] Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Firith, C.D., Frackowiak, R.S.J.: Statistical prametric maps in functional imaging: a general linear approach. Human Brain Mapping 2(4), 189–210 (1995)

[60] Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-6(6), 721–741 (1984)

[61] Glasbey, C.A., Mardia, K.V.: A penalised likelihood approach to image warping. Journal of the Royal Statistical Society, Serie B 63, 465–492 (2001)

[62] Gonzalez, R.C., Woods, R.E.: Digital Image Processing (3rd Edition). Addison-Wesley Longman Publishing Co., Inc. (2006)

[63] Gouttard, S., Styner, M., Joshi, S., Smith, R.G., Cody, H., Gerig, G.: Subcortical structure segmentation using probabilistic atlas priors. SPIE (2007)

[64] Grady, L.: Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1768–1783 (2006)

[65] Grady, L., Schwartz, E.L.: Isoperimetric graph partitioning for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(3), 469–475 (2006)

[66] Grenander, U., Miller, M.I.: Computational anatomy: an emerging discipline. Quarterly of Applied Mathematics LVI(4), 617–694 (1998)

[67] Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: A convergence study. Computer Vision and Image Understanding 77(2), 192 – 210 (2000)

[68] Hachama, M., Desolneux, A., Richard, F.J.P.: Bayesian technique for image classifying registration. IEEE Transactions on Image Processing 21(9), 4080–4091 (2012)

[69] Haxhimusa, Y., Ion, A., Kropatsch, W.G.: Evaluating hierarchical graph-based segmentation. In: ICPR (2). pp. 195–198 (2006)

[70] Heckemann, R.A., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33, 115–126 (2006)

[71] Heckemann, R.A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J., Hammers, A.: Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. NeuroImage 51, 221–227 (2010)

[72] Hu, S., Collins, D.L.: Joint level-set shape modeling and appearance modeling for brain structure segmentation. NeuroImage 36(3), 672 – 683 (2007)

[73] Jenkinson, M., Pechaud, M., Smith, S.: BET2: MR-based estimation of brain, skull and scalp surfaces. In: Eleventh Annual Meeting of the Organization for Human Brain Mapping (2005)

[74] Jin, S., Daly, D.S., Springer, D.L., Miller, J.H.: The effects of shared peptides on protein quantitation in label-free proteomics by LC/MS/MS. Journal of Proteome Research 7, 164–169 (2008)

[75] Johnson, K.A., Becker, J.A.: The Whole Brain Atlas (1997)

[76] Joshi, S., Davis, B., Jomier, B.M., B, G.G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)

[77] Joshi, S., Miller, M.I.: Landmark matching via large deformation diffeomorphisms. IEEE Transactions on Image Processing 9(8), 1357–1370 (2000)

[78] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1(4), 321–331 (1988)

[79] Kikinis, R., Shenton, M.E., Iosifescu, D.V., McCarley, R.W., Saiviroonporn, P., Hokama, H.H., Robatino, A., Metcalf, D., Wible, C.G., Portas, C.M., Donnino, R.M., Jolesz, F.A.: A digital brain atlas for surgical planning, model-driven segmentation, and teaching. IEEE Transactions on Visualization and Computer Graphics 2(3), 232–241 (1996)

[80] Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46(3), 786 – 802 (2009)

[81] Klein, A., Mensh, B., Ghosh, S., Tourville, J., Hirsch, J.: Mindboggle: Automated brain labeling with multiple atlases. BMC Medical Imaging 5(1), 7 (2005)

[82] Kovacevic, N., Chen, J., Sled, J., Henderson, J., Henkelman, M.: Deformation based representation of groupwise average and variability. In: MICCAI. vol. 3216, pp. 615–622 (2004)

[83] Leemput, K.V.: Encoding probabilistic brain atlases using Bayesian inference. IEEE Transactions on Medical Imaging 28(6), 822–837 (2009)

[84] Leemput, K.V., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Transactions on Medical Imaging 20(8), 677–688 (2001)

[85] Liu, Q., Tan, G., Levenkova, N., Li, T., Pugh Jr., E.N., Rux, J.J., Speicher, D.W., Pierce, E.A.: The proteome of themouse photoreceptor sensory cilium complex. Molecular & Cellular Proteomics 6, 1299–1317 (2007)

[86] Ma, J., Miller, M.I., Trouvé, A., Younes, L.: Bayesian template estimation in computational anatomy. NeuroImage 42(1), 252 – 261 (2008)

[87] MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability. vol. 1, pp. 281–297 (1967)

[88] Malinsky, M., Peter, R., Hodneland, E., Lundervold, A.J., Lundervold, A., Jan, J.: Registration of FA and T1-weighted MRI data of healthy human brain based on template matching and normalized cross-correlation. Journal of Digital Imaging 26(4), 774–785 (2013)

[89] Malladi, R., Sethian, J.A., Vemuri, B.C.: Shape modeling with front propagation: A level set approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(2), 158–175 (1995)

[90] Marsland, S., Twining, C.J., Taylor, C.J.: A minimum description length objective function for groupwise non-rigid image registration. Image and Vision Computing 26(3), 333 – 346 (2008)

[91] Marsland, S., Twining, C.J., Taylor, C.J.: Groupwise non-rigid registration using polyharmonic clamped-plate splines. In: MICCAI (2). vol. 2879, pp. 771–779 (2003)

[92] McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: a survey. Medical Image Analysis 1(2), 91 – 108 (1996)

[93] Meyer, C.R., Boes, J.L., Kim, B., Bland, P.H., Zasadny, K.R., Kison, P.V., Koral, K., Frey, K.A., Wahl, R.L.: Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. Medical Image Analysis 1(3), 195–206 (1997)

[94] Miller, M., Priebe, C.E., Qiu, A., Fischl, B., Kolasny, A., Brown, T., Park, Y., Ratnanather, J.T., Busa, E., Jovicich, J., Yu, P., Dickerson, B.C., Buckner, R.L.: Morphometry BIRN. collaborative computational anatomy: An MRI morphometry study of the human brain via diffeomorphic metric mapping. Human Brain Mapping 30(7), 2132–2141 (2009)

[95] Miller, M.I., Trouvé, A., Younes, L.: On the metrics and Euler-Lagrange equations of computational anatomy. Annual Review of biomedical Engineering 4 (2002)

[96] Niethammer, M., Shi, Y., Benzaid, S., Sanchez, M., Styner, M.: Robust model-based transformation and averaging of diffusion weighted images – applied to diffusion weighted atlas construction. In: MICCAI, International Workshop on Computational Diffusion MRI (CDMRI' 10) (2010)

[97] Oberg, A.L., Mahoney, D.W., Eckel-Passow, J.E., Malone, C.J., Wolfinger, R.D., et al.: Statistical analysis of relative labeled mass spectrometry data from complex samples using ANOVA. Journal of Proteome Research 7, 225–233 (2008)

[98] Ou, W., Wells III, W.M., Golland, P.: Combining spatial priors and anatomical information for fMRI detection. Medical Image Analysis 14(3), 318–331 (2010)

[99] Park, H., Bland, P.H., Hero, A.O., Meyer, C.R.: Least biased target selection in probabilistic atlas construction. In: MICCAI (2). vol. 3750, pp. 419–426 (2005)

[100] Paulovich, A.G., Billheimer, D., J., H.A., Vega-Montoto, L., Rudnick, P.A., et al.: Interlaboratory study characterizing a yeast performance standard for benchmarking lc-ms platform performance. Molecular & Cellular Proteomics 9, 242–254 (2010)

[101] Peyre, G., Cohen, L.D.: Geodesic remeshing using front propagation. International Journal of Computer Vision 69(1), 145–156 (2006)

[102] Pham, D.L., Pham, D.L., Prince, J.L., Prince, J.L.: Adaptive fuzzy segmentation of magnetic resonance images. IEEE Transactions on Medical Imaging 18(9), 737–752 (1999)

[103] Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Bihan, D.L., Poline, J.B., Dehaene, S.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neuroscience 8 (2007)

[104] Plummer, M.: JAGS: A program for analysis of bayesian graphical models using Gibbs sampling. Proceedings of DSC, Vienna, Austria (2003)

[105] Podwojski, K., Eisenacher, M., Kohl, M., Turewicz, M., Meyer, H., Rahnenfuhrer, J., Stephan, C.: Peek a peak: a glance at statistics for quantitative label-free proteomics. Expert Review of Proteomics 7, 249–261 (2010)

[106] Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells III, W.M.: A bayesian model for joint segmentation and registration. NeuroImage 31, 228–239 (2006)

[107] Press, W.H.: Numerical Recipes in Fortran 77: The Art of Scientific Computing. Fortran numerical recipes, Cambridge University Press (1996)

[108] R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2010)

[109] Régis, J., Mangin, J.F., Frouin, V., Sastre, F., Peragut, J.C., Samson, Y.: Generic model for the localization of the cerebral cortex and preoperative multimodal integration in epilepsy surgery. Stereotactic and Functional Neurosurgery 65, 72–80 (1995)

[110] Ribbens, A., Hermans, J., Maes, F., Vandermeulen, D., Suetens, P.: Sparc: Unified framework for automatic segmentation, probabilistic atlas construction, registration and clustering of brain MR images. IEEE International Symposium on Biomedical Imaging pp. 856–859 (2010)

[111] Richard, F., Samson, A., Cuénod, Charles, A.: A SAEM algorithm for the estimation of template and deformation parameters in medical image sequences. Statistics and Computing 19(4), 465–478 (2009)

[112] Riklin Raviv, T., Leemput, K.V., Wells III, W.M., Golland, P.: Joint segmentation of image ensembles via latent atlases. Medical Image Analysis 14, 654–665 (2010)

[113] Robles-Kelly, A.: Segmentation via graph-spectral methods and Riemannian geometry. In: Proc. of the 11th International Conference on Computer Analysis of Images and Patterns. pp. 661–668 (2005)

[114] Roche, A., Malandain, G., Pennec, X., Ayache, N.: The correlation ratio as a new similarity measure for multimodal image registration. In Proc. 1st MICCAI 1496 of Lecture Notes in Computer Science, Cambridge, MA, 1115–1124 (1998)

[115] Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3d statistical deformation models using non-rigid registration. IEEE Transactions on Medical Imaging 22, 77–84 (2003)

[116] Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic Registration Using B-Splines. In: MICCAI. pp. 702–709 (2006)

[117] Sabuncu, M.R., Yeo, B.T.T., Leemput, K.V., Fischl, B., P., G.: A generative model for image segmentation based on label fusion. IEEE Transactions on Medical Imaging 29(10), 1714–1729 (2010)

[118] Sabuncu, M.R., Singer, B.D., Conroy, B., Bryan, R.E., Ramadge, P.J., Haxby, J.V.: Function-based intersubject alignment of human cortical anatomy. Cerebral Cortex 20, 130–140 (2010)

[119] Scherrer, B., Forbes, F., Garbay, C., Dojat, M.: Distributed loca lMRF models for tissue and structure brain segmentation. NeuroImage 44(3), 1152–1162 (2009)

[120] Schulze, W.X., Usadel, B.: Quantitation in mass-spectrometry-based proteomics. Annual Review of Plant Biology 61, 491–516 (2010)

[121] Seghers, D., D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: Construction of a brain template from MR images using state-of-the-art registration and segmentation techniques. In: MICCAI(1). vol. 3216, pp. 696–703 (2004)

[122] Sezgin, M., Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging 13(1), 146–168 (2004)

[123] Shan, L., Charles, C., Niethammer, M.: Automatic atlas-based three-label cartilage segmentation from MR knee images. In: 2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. pp. 241–246 (2012)

[124] Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39(3), 1064 – 1080 (2008)

[125] Siless, V., Guevara, P., Pennec, X., Fillard, P.: Joint T1 and brain fiber diffeomorphic registration using the Demons. Multimodal Brain Image Analysis 7012, 10–18 (2011)

[126] Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE Transactions on Medical Imaging 32(7), 1153–1190 (2013)

[127] Stringer, J.R.: Software review. Teaching and Learning in Medicine: An International Journal 7(3), 194–195 (1995)

[128] Studholme, C., Hill, D.L.G., Hawkes, D.J.: Incorporating connected region labelling into automated image registration using mutual information. In: Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis. pp. 23–31 (1996)

[129] Studholme, C.: Simultaneous population based image alignment for template free spatial normalisation of brain anatomy. In: WBIR. vol. 2717, pp. 81–90 (2003)

[130] Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3d medical image alignment. Pattern Recognition 32(1), 71–86 (1999)

[131] Talairach, J., Tournoux, P.: Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging (Thieme Classics). Thieme Medical Pub (1988)

[132] Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., Poline, J.B.: Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. NeuroImage 35, 105–120 (2007)

[133] Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis 2(3), 243–260 (1998)

[134] Trouvé, A.: Diffeomorphisms groups and pattern matching in image analysis. International Journal of Computer Vision 28, 213–221 (1998)

[135] Umbaugh, S.E.: Digital Image Processing and Analysis: Human and Computer Vision Applications with CVIPtools, Second Edition. Taylor & Francis (2010)

[136] van der Vaart, A.W.: Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press (2000)

[137] Vaillant, M., Miller, M, I., Trouvé, A., Younes, L.: Statistics on diffeomorphisms via tangent space representations. NeuroImage 23(S1), S161–S169 (2004)

[138] Valot, B., Langella, O., Nano, E., Zivy, M.: MassChroQ: a versatile tool for mass spectrometry quantification. Proteomics 11, 3572–3577 (2011)

[139] Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45, 61–72 (2009)

[140] Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1,Supp.1), S61–72 (2009)

[141] Vese, L.A., Chan, T.F., Tony, Chan, F.: A multiphase level set framework for image segmentation using the mumford and shah model. International Journal of Computer Vision 50, 271–293 (2002)

[142] Wang, L., He, L., Mishra, A., Li, C.: Active contours driven by local gaussian distribution fitting energy. Signal Processing 89(12), 2435–2447 (2009)

[143] Wang, Y.J., Farneback, G., Westin, C.F.: Multi-affine registration using local polynomial expansion. Journal of Zhejiang University Science 11(7), 495–503 (2010)

[144] Wang, Y., Teoh, E.K.: Dynamic B-snake model for complex objects segmentation. Image and Vision Computing 23(12), 1029–1040 (2005)

[145] Wells III, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of mri data. IEEE Transactions on Medical Imaging 15(4), 429–442 (1996)

[146] Wertheim, S.L.: The brain database: A multimedia neuroscience database for research and teaching. Symposium on Computer Application in Medical Care 8(1), 399–404 (1989)

[147] Williams, D.J., Shah, M.: A fast algorithm for active contours and curvature estimation. CVGIP: Image Underst. 55(1), 14–26 (1992)

[148] Wu, M., Rosano, C., Lopez-Garcia, P., Carter, C.S., Aizenstein, H.J.: Optimum template selection for atlas-based segmentation. NeuroImage 34(4), 1612 – 1618 (2007)

[149] Wyatt, P.P., Noble, J.A.: MAP MRF joint segmentation and registration of medical images. Medical Image Analysis 7(4), 539 – 552 (2003)

[150] Xu, H., Thirion, B., Allassonnière, S.: Bayesian estimation of probabilistic atlas for anatomically-informed functional MRI group analyses. In: MICCAI (3). vol. 8151, pp. 592–599 (2013)

[151] Xu, H., Thirion, B., Allassonnière, S.: Probabilistic atlas and geometric variability estimation to drive tissue segmentation. Statistics in Medicine (accept)

[152] Yelnik, J., Damier, P., Demeret, S., Gervais, D., Bardinet, E., Bejjani, B., Fran¸cois, C., Houeto, J., Arnule, I., Dormont, D., Galanaud, D., Pidoux, B., Cornu, P., Agid, Y.: Localization of stimulating electrodes in patients with parkinson disease by using a three-dimensional atlas-magnetic resonance imaging coregistration method. Journal of Neurosurgery 99(1), 89–99 (2003)

[153] Yeo, B.T.T., Sabuncu, M.R., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical demons: Fast diffeomorphic landmark-free surface registration. IEEE Transactions on Medical Imaging 29(3), 650–668 (2010)

[154] Zahn, C.T.: Graph-theoretical methods for detecting and describing gestalt clusters. IEEE Transactions on Computers 20(1), 68–86 (1971)

[155] Zhang, D., Wu, G., Jia, H., Shen, D.: Confidence-guided sequential label fusion for multi-atlas based segmentation. In: MICCAI (3). vol. 6893, pp. 643–650 (2011)

[156] Zhang, K., Song, H., Zhang, L.: Active contours driven by local image fitting energy. Pattern Recognition. 43(4), 1199–1206 (2010)

[157] Zhang, Y., Wen, Z., Washburn, M.P., Florens, L.: Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Analytical Chemistry 82, 2272–2281 (2010)

[158] Zhang, Y., Brady, M., Smith, S.: 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 (2001)

Résumé

Cette thèse s'inscrit dans le domaine de l'analyse statistique d'images du cerveau. Dans la première partie, nous avons proposé un nouveau m odèle statistique génératif basé sur des templates déformables qui prend en compte les différents types de tissus observés dans les images cérébrales. Un prototype (appelé template) est appris conjointement avec la variabilité géométrique des formes représentées dans les observations. Ces deux éléments forment un atlas qui contient à la fois une estimation des cartes de probabilité de chaque tissu (appelée classes) et la métrique de déformation. Le recalage et la segmentation, qui sont directement intégrés dans le processus d'estimation, sont effectué conjointement. Nous utilisons un algorithme EM stochastique (SAEM) couplée avec des méthodes MCMC pour l'estimation de cet atlas probabilistes. Nous avons prouvé l'existence et consistance de l'estimateur ainsi que la convergence de l'algorithme d'estimation. Cet atlas est ensuite utilisé pour segmenter les nouvelles images par une méthode de segmentation contrainte par notre atlas en particulier la métrique de l'espace des déformations admissibles. Le modèle est ensuite généralisé à des données multidimensionnelles permettant la prise en comte de diverses modalités comme l'anatomie couplée au fonctionnel. Ce modèle prend en compte les contraintes biologiques inhérentes au cerveau en particulier le fait que l'activation fonctionnelle n'apparat que dans la matière grise. Inférer l'abondance des protéines de l'intensité de peptides est l'étape clé dans la protéomique quantitative. La conclusion est nécessairement plus précis quand de nombreux peptides sont pris en compte pour une protéine donnée. Pourtant, l'information apportée par les peptides partagées par différentes protéines est souvent jeté. Dans la troisième partie du travail, nous proposons un système statistique basée sur une modèle hiérarchique à inclure cette information. Notre méthodologie, basée sur une analyse simultanée de tous les peptides quantifiés, gère les erreurs biologiques et techniques ainsi que l'effet des peptides. En outre, nous proposons une mise en uvre pratique adapté à l'analyse de grandes bases de données en utilisant des méthodes MCMC pour l'estimation d'abondance.