HAL
open science

Beyond Volumetry in Longitudinal Deformation-Based Morphometry: Application to Sexual Dimorphism during Adolescence

Mehdi Hadj-Hamou

▶ **To cite this version:**

Mehdi Hadj-Hamou. Beyond Volumetry in Longitudinal Deformation-Based Morphometry: Application to Sexual Dimorphism during Adolescence. Medical Imaging. Universite Nice Cote d'Azur, 2016. English. NNT : . tel-01416569v1

HAL Id: tel-01416569

https://inria.hal.science/tel-01416569v1

Submitted on 14 Dec 2016 (v1), last revised 20 Mar 2017 (v2)

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.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

UNIVERSITÉ CÔTE D'AZUR

ECOLE DOCTORALE STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

THÈSE DE DOCTORAT

pour l'obtention du grade de

Docteur en Sciences

de l'Université Côte d'Azur

Domaine : Automatique, Traitement du Signal et des Images

Présentée et soutenue par

Mehdi Hadj-Hamou

Au delà de la volumétrie en morphométrie basée sur les déformations : application au dimorphisme sexuel durant l'adolescence

Directeur : Nicholas Ayache

Co-Directeur : Xavier Pennec

preparée à INRIA Sophia Antipolis, équipe Asclepios

soutenue le 14 décembre 2016

Jury :

Rapporteurs :	Christian Barillot	-	CNRS-INSERM-INRIA (VisAGes)
	Olivier Colliot	-	CNRS-INSERM-INRIA-UPMC-ICM (ARAMIS)
Directeur :	Nicholas Ayache	-	INRIA (Asclepios)
Co-Directeur :	Xavier Pennec	-	INRIA (Asclepios)
Examinateurs :	Rachid Deriche	-	INRIA (Athena)
	Jean-Luc Martinot	-	INSERM-CEA
Invité :	Marco Lorenzi	-	UCL

UNIVERSITY CÔTE D'AZUR

DOCTORAL SCHOOL STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

PHD THESIS

to obtain the title of

PhD of Science

of the University Côte d'Azur

Specialty: Automation, Signal and Image Processing

Defended by

Mehdi Hadj-Hamou

Beyond Volumetry in Longitudinal Deformation-Based Morphometry: Application to Sexual Dimorphism during Adolescence

Thesis Advisor: Nicholas Ayache

Thesis Co-Advisor: Xavier Pennec

prepared at INRIA Sophia Antipolis, Asclepios Team

defended on December 14, 2016

Jury:

Reviewers:	Christian Barillot	-	CNRS-INSERM-INRIA (VisAGes)
	Olivier Colliot	-	CNRS-INSERM-INRIA-UPMC-ICM (ARAMIS)
Advisor:	Nicholas Ayache	-	INRIA (Asclepios)
Co-Advisor:	Xavier Pennec	-	INRIA (Asclepios)
Examinators:	Rachid Deriche	-	INRIA (Athena)
	Jean-Luc Martinot	-	INSERM-CEA
Invited:	Marco Lorenzi	-	UCL

Acknowledgments

This work was partially funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modelling and Analysis of Dynamic Medical Images).

Contents

1 Introduction **1**
1.1 Clinical and methodological context 1
1.1.1 Adolescence: the onset of the majority of psychological disorders 1
1.1.2 Neuroimaging to better understand sexual dimorphism during adolescence 1
1.1.3 Existing problems to study longitudinal brain images 2
1.2 Objectives and Contributions 3
1.2.1 Objectives 3
1.2.2 Organisation of the thesis 3
1.2.3 List of publications 4

2 Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: a Computational Framework based on Stationary Velocity Fields **6**
2.1 Introduction 7
2.2 Processing Pipeline for the Analysis of Longitudinal Images 8
2.2.1 Pre-Processing 9
2.2.2 Position Correction 12
2.2.3 Non-Linear Deformation Analysis 15
2.3 Application to the Analysis of the Longitudinal Changes in Alzheimer's Disease 24
2.3.1 OASIS database 24
2.3.2 Methods and Results 24
2.4 Conclusion and Discussion 31

3 Validation of Longitudinal Volumetric Measurement Methods **36**
3.1 Introduction 37
3.2 Methodology of validation 39
3.2.1 Choice of software packages to compare: Segmentation and Registration-based software 39
3.2.2 Using both real and simulated datasets 40
3.2.3 Comparing segmentation and registration-based techniques . 41
3.2.4 When is the log-Jacobian integration equivalent to computing the relative volume changes ? 44
3.3 Results 46
3.3.1 Comparison of segmentation and registration-based methods using segmentation propagation 46

3.3.2 Accuracy of the log-Jacobian with respect to the segmentation propagation, statistical power of the registration-based methods using the log-Jacobian, and influence of the numerical schemes . . . 51
3.4 Conclusion and Discussion . . . 54

4 Beyond Volumetry: Deformation based Morphometry for the differential analysis of longitudinal SVF trajectories of subgroups 64
4.1 Introduction . . . 65
4.2 Disentangling inter-group and longitudinal changes: a geometric interpretation . . . 67
4.3 Implementation in the SVF case . . . 69
4.3.1 Construction of the inter-group axes and group longitudinal axis . . . 70
4.3.2 Computation of the intelligible quantities of interest . . . 72
4.4 Application to Alzheimer's disease: OASIS database . . . 73
4.4.1 Convergence/divergence of the patients with Alzheimer's and controls evolutions . . . 73
4.4.2 Significant groups differences at 75 and 76 between controls and patients with Alzheimer's . . . 73
4.5 Conclusion and Discussion . . . 74

5 Effect of sexual dimorphism on the healthy morphological evolution of the brain during adolescence 78
5.1 Introduction . . . 79
5.2 Dataset and methods . . . 80
5.2.1 The IMAGEN database . . . 80
5.2.2 Image processing . . . 81
5.2.3 Definition of the regions of interest . . . 81
5.3 Longitudinal volumetric changes . . . 81
5.3.1 Regional volume changes: Male-Female Comparison . . . 81
5.3.2 Local volume changes: Male-Female Comparison . . . 85
5.4 Beyond volumetry to study the effects of sexual dimorphism during adolescence . . . 87
5.4.1 Multivariate comparison of female and male longitudinal SVF trajectories . . . 87
5.4.2 Sexual differences in the cross-sectional and longitudinal evolution between 14 and 16 years old . . . 88
5.5 Conclusion and Discussion . . . 91

6 Conclusion and Perspectives 94
6.1 Main achievements . . . 94
6.2 Perspectives . . . 96
6.2.1 Automatic quality control of the processing pipeline results . 96

6.2.2 Mental disorders during adolescence 96
6.2.3 Using continuous classification of subjects along the inter-group axis 97
6.2.4 Understanding longitudinal evolutions with a dictionary of transformations 98

A Towards a Validation of Interpolation and Extrapolation of Longitudinal Deformations Based on Stationary Velocity Fields 99
A.1 Introduction 100
A.2 Measures for the validation of SVF-based geodesic regression. 101
A.2.1 Metrics for similarity of the modeled image evolution. 102
A.2.2 Metrics for the similarity of the modeled volume changes. . . 102
A.3 Experimental data 102
A.3.1 OASIS database 102
A.3.2 Image processing. 103
A.3.3 Group-wise analysis. 103
A.4 Results 103
A.4.1 Interpolation of longitudinal trajectories 103
A.4.2 Extrapolation of longitudinal trajectories 105
A.5 Conclusion 106

Bibliography 107

Chapter 1

Introduction

Contents

1.1 Clinical and methodological context 1

1.1.1 Adolescence: the onset of the majority of psychological disorders 1

1.1.2 Neuroimaging to better understand sexual dimorphism during adolescence 1

1.1.3 Existing problems to study longitudinal brain images 2

1.2 Objectives and Contributions 3

1.2.1 Objectives 3

1.2.2 Organisation of the thesis 3

1.2.3 List of publications 4

1.1 Clinical and methodological context

1.1.1 Adolescence: the onset of the majority of psychological disorders

Mental disorders will become Europe's largest health challenge of the 21st century: 38.2% of the EU population is concerned, with an estimated cost of 277 billion Euros [Wittchen 2011]. In half of the cases, the first onset of mental disorders occurs during adolescence at around 14 years old. Early interventions could help reduce or prevent the development of mental disorders. Therefore, a detailed understanding of the disease effect on the adolescent brain development is needed. This first implies improving the knowledge on healthy brain development during that period. Moreover, it is important to focus on sex differences since the prevalence of mental disorders in females is twice as much important as in males. In this thesis, we define sex differences as what is related to biological differences by opposition to gender that relates to the characteristics attributed by the society to masculine or feminine.

1.1.2 Neuroimaging to better understand sexual dimorphism during adolescence

Adolescence is a period of important changes where the brain matures to its adult state. This implies relatively subtle structural changes with time for both sexes. In the literature, only volumes are generally investigated although brain evolution

might be more than volume changes. Moreover, the results of the studies sometimes disagree in some brain regions. This increases the difficulty to establish knowledge on healthy brain development and sexual dimorphism during adolescence. Thus, it would be desirable to have a method that gives accurate results even for subtle brain changes and to be able to compare male and female evolutions while giving interpretable results.

1.1.3 Existing problems to study longitudinal brain images

With the recent development of longitudinal databases replacing cross-sectional ones, the confounding effect of inter-individual morphological variability has been reduced by using each subject as his or her own control.

However, there still exist several potential biases that need to be avoided while processing the longitudinal data. The most important issue concerns the asymmetry biases [Ridgway 2015]. Two types of asymmetries can be distinguished. The first one is introduced by the resampling of all the follow-up images except the baseline. The second type of bias is related to the non-centrality of the time point where the subject longitudinal deformations are computed. It is now established that the sequencing of the processing steps is key to avoid adding bias to the data and thus obtaining a robust longitudinal processing pipeline. In addition to the order of the steps, robust similarity measures have been proposed instead of the Sum of Squared Differences metric to increase the robustness of registration methods. However, most registrations still remain performed on the whole head potentially biasing the results at the brain border (cortex region). Moreover, in the developing context of reproducible research that has gained interest over the last years, a good practice should be for the processing pipeline to be fully reproducible.

Even when controlling for the different potential sources of bias in the longitudinal processing pipeline, the inter-subject variability might be by far higher than the small intra-subject longitudinal changes (e.g for a healthy subjects dataset). This would result in the impossibility to capture the longitudinal changes that would be lost in this very noisy context. In this case, we might reach the statistical detection limit of the algorithms and different methods might lead to opposite results. Following [Fox 2011], it is thus of great importance to "improve and validate atrophy quantification" and to know if the results are due to a bias in the method. However, to our knowledge no study has investigated this topic and compared longitudinal volumetric methods on such small changes.

Finally, although volumetry has enabled the community to make great findings, we can question whether volume changes are sufficient to explain all the changes occurring in the brain. For example, in Alzheimer's disease, one can observe that the temporal lobes often have a significant rotation, which might mean that there are more than just volume changes. Although statistical methods exist for the comparison of multivariate fields, the interpretation of the results still remains more difficult than in the univariate case.

1.2 Objectives and Contributions

1.2.1 Objectives

The objective of this PhD work is to address the above problems in the context of longitudinal changes during adolescence. The problems are challenging and necessitate the development of a longitudinal processing pipeline for the robust estimation of longitudinal deformations. Since the adolescence period involves subtle brain changes, we need to start by ensuring that the algorithms we use are able to detect such small variations. Moreover, we aim at going beyond the volumetry in order to ease the interpretation of the comparison of the three-dimensional deformation fields. Finally, we need to keep in mind that all the proposed work and results in this thesis have to be reproducible.

1.2.2 Organisation of the thesis

In **chapter 2**, we propose a deformation-based morphometry computational framework to robustly estimate the longitudinal brain deformations from image data series. Our first contribution is in explicitly detailing all the processing steps required for the longitudinal analysis of neuroimages. The sequencing of the numerous steps has been designed to limit the potential biases. A second contribution is to modify the non-linear registration algorithm by masking the similarity term while keeping the symmetry of the formulation. This change increases the robustness of the results with respect to intensity artifacts located in the brain boundaries. Experimental results show that this contribution leads to increased sensitivity of the statistical study on the longitudinal deformations. The proposed processing pipeline is based on freely available software and tools [1]. It is thus fully reproducible.

In **chapter 3**, we evaluate the accuracy, reproducibility, detection limit, and statistical power of our non-linear registration method in terms of volumetry. We use simulated ground truth and real data for which changes are small. Since all the effects cannot be tested while evaluating only one algorithm, we compare our method with three other popular volumetric longitudinal methods (segmentation-based and registration-based).
We find several interesting results for the community. In the zone where the changes are large enough, registration based-methods are generally more accurate, reproducible, and have a greater statistical power than the segmentation-based method FreeSurfer. The former method is also generally more sensitive to local linear intensity bias than registration-based methods. We notice that registration-based methods highly under-estimate large changes (superior to 10%), especially SPM. Concerning the detection limit of the methods, we show that registration based-methods have a smaller detection limit than the segmentation-based method. Moreover, we show that for changes less than 10 % the log-Jacobian integration is equivalent to relative volume changes. Finally, we evaluate two different numerical schemes

[1] The tools are available at http://www-sop.inria.fr/teams/asclepios/software/

for the log-Jacobian computation and show that they are very similar and can be interchangeably used.

We propose, in **chapter 4**, to go beyond volumetry in the statistical analysis of two groups by studying the whole 3-dimensional deformation field. To ease the group comparison, we propose to disentangle the group differences from the longitudinal population evolution. In addition to this contribution, we ease the interpretation by presenting two indices. The first one concerns the convergence or divergence of the group longitudinal evolutions. The second index quantifies the group difference with respect to the population longitudinal evolution using an advance or delay in time concept. This enables a tangible comprehension of the group differences for clinicians.

In **chapter 5**, we use the methods developed in chapters 2 and 4, as well as the results from chapter 3 to study the effect of sexual dimorphism on the healthy morphological evolution of the brain during adolescence. We first analyse the volume changes: results are consistent with the literature in almost every region of the brain. For the few regions where the results are not consistent we perform an additional analysis with three popular volumetric methods. Contrarily to what is reported in the literature, white matter volume does not increase during 14 to 16 for females on our dataset. Moreover, we find that the pre-frontal cortex is the main region where statistically significant differences occur in evolution: the pre-frontal cortex volume decrease is more important for females than for males. We then go beyond volumetry to bring new insights on the evolution during that age period. We find that at 14 years of age, no difference exists between the males and females' pre-frontal cortex, and that an important differentiation occurs in this region during the two following years: at 16, females' pre-frontal cortex is in advance of around five months with respect to the population mean longitudinal evolution.

Finally, **chapter 6** concludes this thesis with a number of unsolved problems. While the volumetric methods generally give accurate results, going beyond volumetry still remains difficult. We discuss about the potential perspectives on interpreting multivariate analysis.

1.2.3 List of publications

The presented work led to a published journal publication and two more are in preparation of submission (as a first author):

Published paper

- *Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: a Computational Framework based on Stationary Velocity Fields.* Mehdi Hadj-Hamou, Marco Lorenzi, Nicholas Ayache and Xavier Pennec. Frontiers in Neuroscience, vol. 10, no. 236, 2016. [Hadj-Hamou 2016]

In preparation

- *Validation of Longitudinal Volumetric Measurement Methods* (Chapter 3) to be submitted in Transactions on Medical Imaging.
- *Beyond Volumetry: Deformation-based Morphometry for the differential analysis of sexual dimorphism during adolescence* (chapters 4 and 5) to be submitted in NeuroImage.

Chapter 2

Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: a Computational Framework based on Stationary Velocity Fields

Contents

2.1 Introduction **7**

2.2 Processing Pipeline for the Analysis of Longitudinal Images **8**

2.2.1 Pre-Processing 9

2.2.2 Position Correction 12

2.2.3 Non-Linear Deformation Analysis 15

2.3 Application to the Analysis of the Longitudinal Changes in Alzheimer's Disease **24**

2.3.1 OASIS database 24

2.3.2 Methods and Results 24

2.4 Conclusion and Discussion **31**

We propose and detail a deformation-based morphometry computational framework, called Longitudinal Log-Demons Framework (LLDF), to estimate the longitudinal brain deformations from image data series, transport them in a common space and perform statistical group-wise analyses. It is based on freely available software and tools, and consists of three main steps: ***i)*** *Pre-processing,* ***ii)*** *Position correction, and* ***iii)*** *Non-linear deformation analysis. It is based on the LCC log-Demons non-linear symmetric diffeomorphic registration algorithm with an additional modulation of the similarity term using a confidence mask to increase the robustness with respect to brain boundary intensity artifacts. The pipeline is exemplified on the longitudinal Open Access Series of Imaging Studies (OASIS) database and all the parameters values are given so that the study can be reproduced. We investigate the group-wise differences between the patients with Alzheimer's disease and the healthy control group, and show that the proposed pipeline increases the sensitivity*

with no decrease in the specificity of the statistical study done on the longitudinal deformations.

2.1 Introduction

An important topic in neuroimaging is to analyse the progression of morphological changes in the brain observed in time series of images, in order to model and quantify normal or pathological biological evolutions [Scahill 2002]. Deformation-Based Morphometry (DBM) [Ashburner 1998] characterises the morphological changes of the brain in terms of spatial transformations (here called deformations), estimated by means of non-linear registration. A sub-field of DBM, called Tensor-Based Morphometry (TBM) focuses on the first derivatives of the deformation. Depending on the cross-sectional or longitudinal nature of the dataset used, we can define on one hand cross-sectional DBM and on the other hand longitudinal DBM [Chung 2001] that we will focus on in this chapter. Longitudinal DBM main steps can be summarised as **i)** quantifying the evolution of the morphology of each subject by estimating the individual's longitudinal deformation from the time series of images, and **ii)** characterising how this evolution varies among a sample using a suitable normalisation for the individual biological variability.

A variety of DBM approaches can be found in the literature (e.g. [Davatzikos 2001, Cardenas 2007, Lorenzi 2011, Südmeyer 2012]), each of them associated to specific non-linear registration methods, and processing pipelines. The comparison between the different DBM methods is not straightforward: the efficiency of each DBM pipeline is generally demonstrated on different data sets (or different subsets of the same data set) and the tools the processing pipeline is composed of are generally not all available. In the existing DBM pipelines - e.g. `SPM` [Friston 2007], `FreeSurfer` [Reuter 2012], `PipeDream`[1], `Anima`[2] - the multivariate information coming from the three-dimensional deformation is generally not used for the statistical analysis. To do so, one would need to express the three-dimensional deformation of every subject in a common space to compare them. There exists few algorithms that compute this 3D transport (e.g. [Lorenzi 2013b]) and in the absence of this tool, the DBM analysis often becomes a TBM analysis only. Studies are thus generally performed on the Jacobian determinant of the deformation or on the segmented regions of interest - since it is easier to compute these scalar maps in a common space. Moreover, in the developing context of reproducible research that has gained interest over the last years [Nature 2013, McCormick 2014], a good practice should be for researchers to publish the full details of their methodology: source code, data and parameters.

This is the objective of this chapter: to gather all the details in the same paper and propose a pipeline for the community, following the examples of [Avants 2011] and [Ashburner 2013]. Our computational framework is a complement to the exist-

[1] http://sourceforge.net/projects/neuropipedream/
[2] https://github.com/Inria-Visages/Anima-Public/wiki

ing processing pipelines. It enables researchers to replicate and verify their findings with a third party reproducible pipeline, thus enhancing the convincing power of their results. Our pipeline is based on [Lorenzi 2011], who proposed a hierarchical framework for the group-wise analysis of time series of images using diffeomorphic deformations parametrised by Stationary Velocity Fields (SVF). We bring a complement to the already existing literature by explicitly detailing all the processing steps required for the longitudinal analysis of neuroimages by relying on freely available tools. In addition to this contribution, we integrate a modification to the non-linear registration algorithm by adding a masking to the similarity term as proposed by [Brett 2001] while keeping the symmetry of the formulation. This change increases the robustness of the results with respect to intensity artifacts located in the brain boundaries. The proposed processing pipeline is based on freely available software and tools (the complete list can be found in Appendix 2.4).

The chapter is structured as follows: in section A.2, we develop a comprehensive processing pipeline called Longitudinal Log-Demons Framework (LLDF); we present each elementary modules it is based on, and after introducing the mathematical formalism related to DBM, we modify the LCC log-Demons to incorporate a confidence mask. Experimental results show that this contribution leads to increased sensitivity of the statistical study on the longitudinal deformations. In section A.3, we show an illustration of the pipeline on the statistical analysis of longitudinal brain changes in Alzheimer's disease. Because it is freely and easily available for benchmarking, we use the data from the longitudinal Open Access Series of Imaging Studies (OASIS) database [Marcus 2010]. We finally conclude and present the perspectives of this work in section 3.4.

2.2 Processing Pipeline for the Analysis of Longitudinal Images

We consider longitudinal observations of MRI scans for a given subject S_i, at N_i time points t_0, t_1,..., t_{N_i-1} (all the subjects do not necessarily have the same number N_i of time points). The corresponding images are denoted as I_0^i, I_1^i,..., $I_{N_i-1}^i$ respectively. The aim of the processing pipeline is to estimate each subject's longitudinal deformation from the image time series, and then transport the deformations in a common space to perform statistical group-wise analyses.

The construction of the pipeline is based on elementary modules described in the following paragraphs and it can therefore be divided into three main parts (cf. Fig. 2.1): **1)** Pre-processing, **2)** Position correction, and **3)** Non-linear deformation analysis. The pipeline proposed in this work relies on a number of neuroimaging tools previously proposed and validated by different groups. Our choice was motivated by our personal experience and by the optimal performances obtained in the presented application. We however acknowledge that other tools could have been employed. For this reason, the modular nature of the pipeline allows the replacement of the proposed tools with specific ones, such as in the case of longitudinal analysis in

postnatal brain development (cf. 2.2.1.3).

2.2.1 Pre-Processing

In this initial part of the pipeline all the individuals' images are processed independently of the time points. The pre-processing consists of the following chain of elementary steps: **1)** Standard reorientation, **2)** Field of view reduction and, **3)** Intensity non-uniformity correction. Different criteria have been taken into account for choosing the tools and software used to perform these elementary steps. Firstly, we only selected freely available tools part of well-established software - so that the pipeline can be reproduced by anyone - relying on already validated tools. Secondly, to make the pipeline user-friendly, we chose tools that necessitate minimal fine tuning in terms of parameters.

2.2.1.1 Standard Reorientation

Images from the MRI scanner are not necessarily oriented following the standard orientation defined by the MNI152 [Fonov 2009] template (Figure 2.2). This misorientation would prevent us from properly processing the images.

We thus use `FSL - fslreorient2std` [Jenkinson 2012], to reorient each image to match the standard orientation. Starting with I, the image acquired by the scanner, this tool applies rotations of 0, 90, 180 or 270 degrees around the image axes to get I^{std}, the reoriented output image. Notice that this reorientation only changes the header and does not perform any interpolation.

2.2.1.2 Reduction of the Field of View

Brain scans can sometimes include the neck or the shoulders (cf. Figure 2.2), and analysing the whole image would increase the image processing time and lead to increased errors due to intensity artifacts. Therefore it is preferable to reduce the Field of View (FOV) of the image to include the head only.

For this purpose, we use `FSL - robustfov` [Jenkinson 2012]: given an image I, comprising the head and the neck, it automatically crops the neck and other regions outside the head by re-sizing the height of the image, starting at the top of the skull, to a default size of 170 mm so that we finally obtain I^{head}, the image containing the head only.

In some rare cases (in another study not reported here, one case out of 120), this automatic tool might provide a wrong result, leaving an important part of the neck in the image or cropping the head. In that case, one can still manually set the correct height of the head.

2.2.1.3 Intensity Inhomogeneity Correction

One of the most common artifact in MRI scans is the shading one: an intensity non-uniformity for voxels of the same tissue class (cf. Fig. 2.2). Therefore,

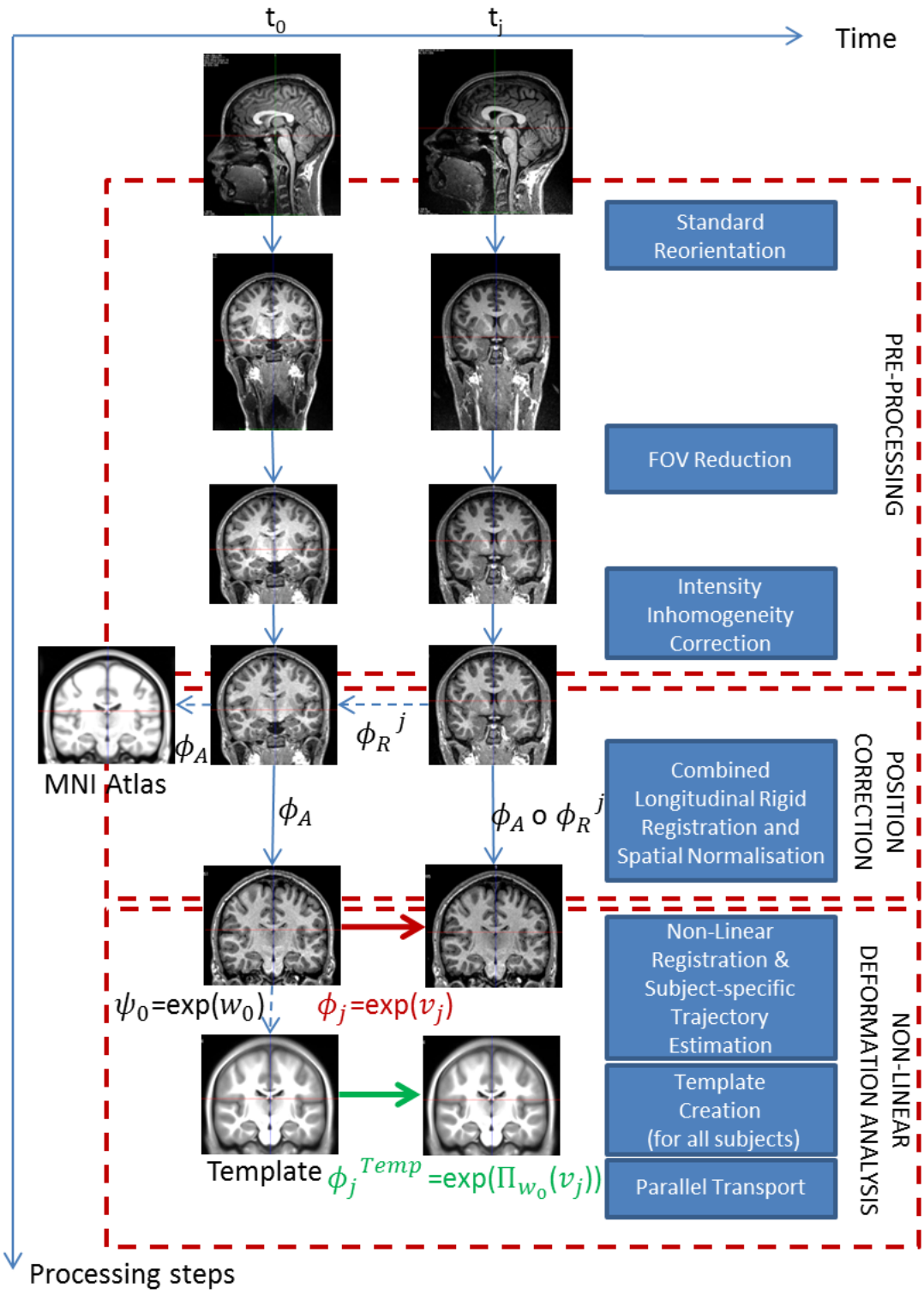

Figure 2.1: *Proposed processing pipeline for longitudinal analysis*: The pipeline is composed of three major steps. Starting with raw images, we first pre-process them, then correct the spatial position differences to end up with the longitudinal deformations for each subject in the template space. Dotted lines correspond to evaluated transformations whereas plain lines correspond to applied transformations.

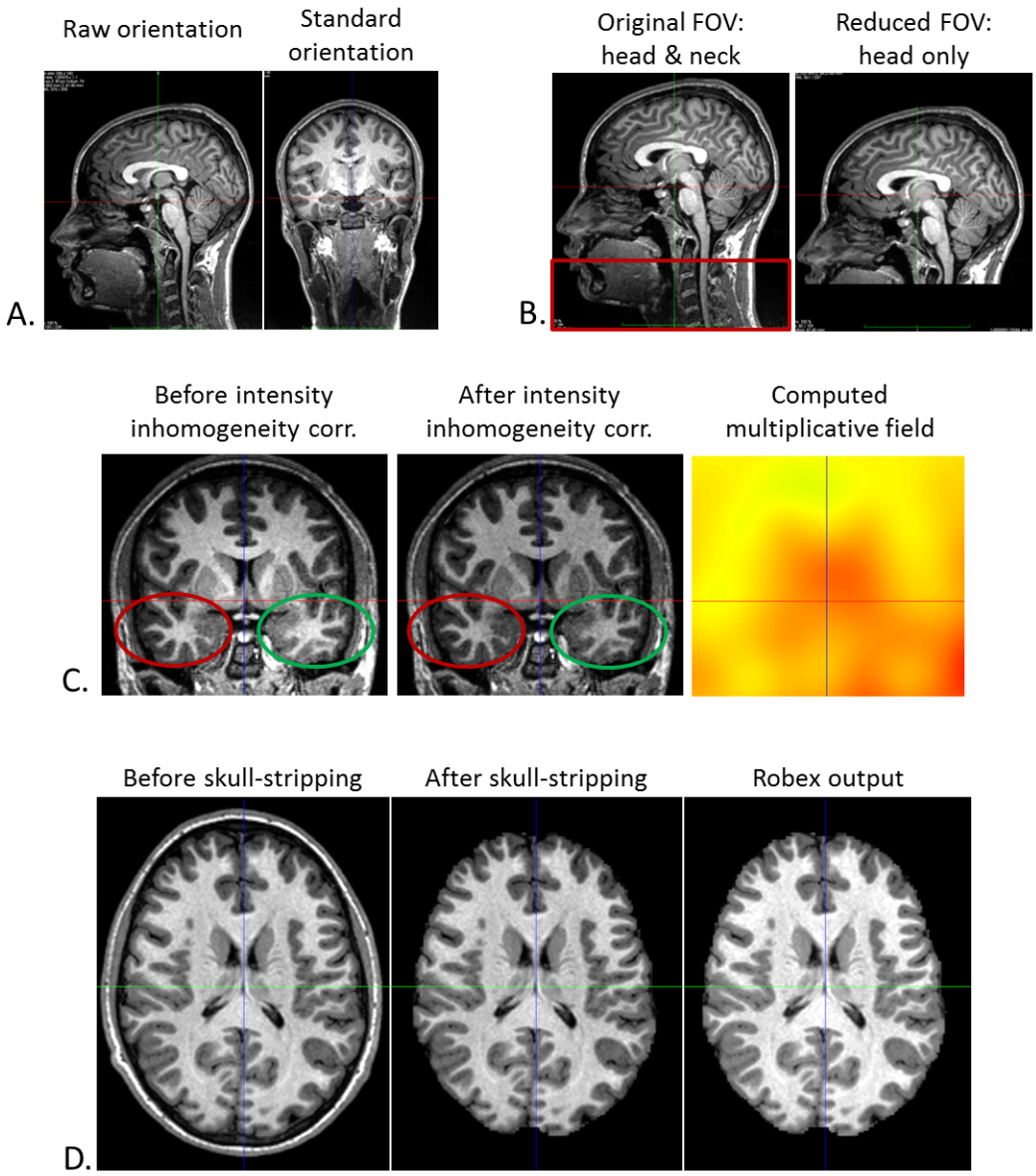

Figure 2.2: *Pre-processing steps*: **A. Reorientation of a subject coronal view**: Left: what is displayed initially as the coronal view is the sagittal one. Right: after reorientation it is truly the coronal view that is displayed. **B. Field of View Reduction**: Left: the original Field of View (FOV) including the head and neck (red rectangle). Right: after reduction, the cropped FOV does not contain the neck, but only the head. **C. Intensity Inhomogeneity Correction**: Left: the image has an intensity non-uniformity. The same tissue class has a lower intensity in the bottom left (red ellipse), and a higher intensity in the bottom right part of the image (green ellipse). Middle: after correction, the intensity appearance of the image is more homogeneous (cf. red and green ellipses). Right: estimated multiplicative field. **D. Skull-stripping**: Left: the head with its skull. Middle: the brain after the whole process of skull-stripping and image masking. We see that the resulting image has the same intensity as the original one; this is not the case of the image output by `Robex` (right image).

each MR image I undergoes an intensity non-uniformity correction using `ANTs - N4BiasFieldCorrection` [Avants 2011, Tustison 2010] to obtain the corrected image I^{Hom}. This algorithm improves the N3 Intensity Inhomogeneity correction [Sled 1998] and is based on the assumption that there exists a smooth, slowly varying multiplicative field F corrupting the image intensities: $I = I^{Hom} \times F$.

In the specific case of early brain development where heterogeneous myelination occurs, the default correction algorithm might be insufficient and a dedicated correction method could be used following [Prastawa 2004] example (or in the case of patients with multiple sclerosis lesions, [Karpate 2014] algorithm). The choice of the most appropriate algorithm is let to the user. In any case, the Local Correlation Criteria (similar to `ANTS` Cross-correlation [Avants 2011]) we use for the non-linear registration in 2.2.3.2 is robust to local intensity bias and is potentially able to cope with an incomplete inhomogeneity correction.

2.2.1.4 Skull-Stripping

It is often necessary (e.g in 2.2.2.1) to process the brain without its surrounding skull. For this reason, the pipeline includes a skull-stripping step (also called non-brain removal tool). We selected `Robex` [Iglesias 2011] for the robustness of its results with no parameter fine tuning: [Iglesias 2011] showed it generally performs better than six other popular algorithms (`BET` [Smith 2002], `BSE` [Shattuck 2001], `FreeSurfer`[3], `AFNI`[4], `BridgeBurner` [Mikheev 2008], and `GCUT` [Mahapatra 2012]). Our experiments were in agreement with this affirmation: when using Robex on our datasets, we no longer had large parts of the skull remaining which was sometimes the case when using `FSL - BET` with the default parameters.

Inputing I, the image with the brain and its surrounding skull, `Robex` outputs I^{robex} and I^{mask}, the skull stripped brain and the corresponding region mask respectively. In fact, `Robex` applies an additional intensity inhomogeneity correction and thus modifies the intensity of the output image I^{robex}. Therefore one has to use the output mask I^{mask} and mask the original image I to obtain I^{brain}, the image with the brain only (cf. Figure 2.2).

2.2.2 Position Correction

Contrary to the previous section, the images are now treated depending on the subject (and time point). This module consists of two combined steps: **1)** Longitudinal rigid registration, and **2)** Affine spatial normalisation. We first present these modules before explaining how we combine them.

2.2.2.1 Longitudinal Rigid Registration

For a single subject, the acquisition at different time points is usually not performed with the same position of the head in the scanner. This creates a global rigid (six

[3] http://surfer.nmr.mgh.harvard.edu
[4] http://afni.nimh.nih.gov/

degrees of freedom) misalignment of each subject data series. Since the aim of this work is to model the subtle local longitudinal brain changes, we need to account for this source of variability that generally exceeds the longitudinal variability. Taking the baseline I_0 as the reference position, we rigidly align the follow-up images I_1,..., I_{N-1} to the baseline I_0, using the rigid transformations ϕ_R^1,...,ϕ_R^{N-1}, to obtain the rigidly aligned image I_1^{al},..., I_{N-1}^{al} (cf. Fig. 2.3).

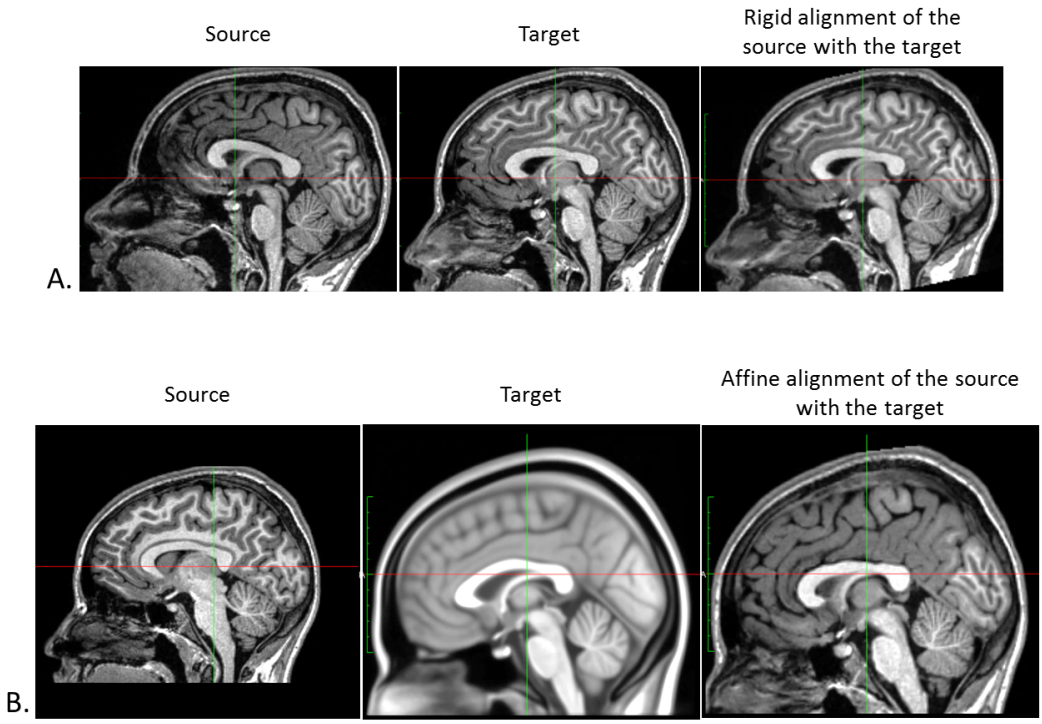

Figure 2.3: *Position correction steps*: **A. Rigid registration of subject images**: The image on the left is the follow-up image of a subject, the baseline (used as the reference) being the image in the middle. The image on the right is the subject image after rigid alignment. **B. Affine normalisation of a subject image**: Left: subject image. Middle: the MNI152 template. The subject image and the template differ in size and orientation. Right: result of the affine normalisation.

We choose to use `FSL - FLIRT` [Jenkinson 2001, Jenkinson 2002] for the linear registration as it is the benchmark linear registration framework used in the influential work of [Klein 2009] for the comparison of several state-of-the-art non-linear registration algorithms. The different steps of the rigid registration step are described in Algorithm 1. We note that despite the optimisation in two steps, only one single rigid transformation is applied. Composing the transformations from the - whole head and skull-stripped head - intra-subject rigid registrations minimises the potential resampling artifacts introduced by the repeated resampling of

Algorithm 1 Longitudinal Rigid Registration between 2 Images

Input: I_j with $j = 1, ..., N-1$, the image not necessarily aligned with the reference I_0.

Output: I_j^{al} with $j = 1, ..., N-1$, the image after rigid alignment with the reference I_0.

Find the rigid transformation ϕ_1^j that aligns I_j to I_0
$H_j = I_j \circ \phi_1^j$

Skull-strip (SS) I_0 and H_j
$J_j = SS(H_j)$ and $J_0 = SS(I_0)$

Find the rigid transformation ϕ_2^j that aligns J_j to J_0
$K_j = J_j \circ \phi_2^j$

Compose the 2 previously found transformations
$\phi_R^j = \phi_2^j \circ \phi_1^j$

Apply the composed transformation to the input image I_1
$I_j^{al} = I_j \circ \phi_R^j$

the data (during the different rigid registration steps). Lastly, we use B-splines as the interpolation method (more accurate than the standard tri-linear interpolation [Parker 1983]) and the normalised correlation as the cost function.

2.2.2.2 Affine Spatial Normalisation

Each brain differs in size and shape. In preparation for the group analysis and in order to align each subject anatomy in a common reference space, we normalise each subject head (shape and pose) to the MNI152 reference space using an affine (twelve degrees of freedom) transformation. Practically, the brain normalisation consists in resampling each subject baseline image I_0 in a common standard space S_{MNI} (MNI152 space) using an affine transformation ϕ_A computed with `FSL - FLIRT` to obtain the normalised image I_0^{MNI} (see Figure 2.3). We use B-splines as the interpolation method and the normalised correlation as the cost function.

2.2.2.3 Combined Longitudinal Rigid Registration and Spatial Normalisation

In the spirit of 2.2.2.1, we avoid as much as possible the potential resampling artifacts by composing the two spatial transformations ϕ_R and ϕ_A from the previous steps. The baseline I_0, is spatially normalised to the MNI152 space using ϕ_A (cf. 2.2.2.2). Concerning the follow-up images I_j, we apply the composition of ϕ_A and ϕ_R^j to I_j.

Since I_j^{al} and I_0 are already rigidly aligned the transformations that map both of them to the template S_{MNI} are the same.

2.2.3 Non-Linear Deformation Analysis

After the correction of the images in position and intensity, we can estimate the residual longitudinal morphological differences using non-linear registration. For this non-linear registration step, all the subjects are processed independently in order to compute each individual longitudinal deformation (expressed in every subject anatomy but with the same coordinate space). The final step is done in three stages: **1)** Estimation of the subject-specific longitudinal deformation trajectory using the previously computed longitudinal deformations, **2)** Study-specific template creation, and **3)** Transport of the subject-specific longitudinal deformation trajectory in the template (cf. Fig. 2.6). Before going further, we introduce the mathematical formalism related to Deformation Based Morphometry.

2.2.3.1 Mathematical Formalism for Deformation-based Morphometry

The longitudinal evolution of a point x of the brain between the initial biological time point $t_0 = 0$ and the biological time t_1 is defined by the *deformation* ϕ that maps the initial position $x(t_0)$ to the position $x(t_1)$:

$$\phi : \mathbb{R}^n \times \mathbb{R} \longrightarrow \mathbb{R}^n$$

$$(x, t) \mapsto x(t) = \phi(x, t)$$

In neuroimaging, the preservation of the brain topology is important; it can be obtained under the large deformation diffeomorphic setting [Joshi 2000, Beg 2005]. In this framework, we define the transformations φ that belong to the group $\mathcal{G}$ of diffeomorphisms: differentiable bijections with differentiable inverse. The transformations are parametrised by the flow of time-dependent velocity vector fields $v(x, s)$ (with the parametrisation time $s \in [0, 1]$) specified by the following ordinary differential equation:

$$\frac{\partial \varphi(x, s)}{\partial s} = v(\varphi(x, s), s),$$

with $\varphi(x, 0) = Id(x)$ (identity transformation). The resulting deformation ϕ, mapping $x(t_0)$ to $x(t_1)$ is given by the flow at $s = 1$: $\phi(x, t_1) = \varphi(x, 1)$. In the spirit of the log-Euclidean framework, [Arsigny 2006] proposed to restrict to the one-parameter subgroup of diffeomorphisms where the velocity vectors are stationary (i.e. constant over the parametrisation time s): $v(x, s) = v(x)$. In this case, the transformation $\phi(x, t_1)$ is encoded by the *stationary velocity field* (SVF) $v(x)$ via the Lie group exponential map: $\phi(x, t_1) = \exp(v(x))$; the exponential map is defined as the flow of the stationary ordinary differential equation:

$$\frac{\partial \varphi(x, s)}{\partial s} = v(\varphi(x, s)),$$

with $\varphi(x, 0) = Id(x)$ and $s \in [0, 1]$.

2.2.3.2 Non-Linear Symmetric Diffeomorphic Registration with Confidence Mask

We estimate the subtle longitudinal changes using symmetric non-linear diffeomorphic registration. The diffeomorphic deformations are parametrised using Stationary Velocity Fields (SVF), providing us with a rich mathematical and computational setting (see [Arsigny 2006, Vercauteren 2008, Lorenzi 2011]).

To non-linearly register I_i to I_j, we estimate the Stationary Velocity Field $v_{i\text{-}j}$ (cf. Figure 2.4) via an alternate minimisation of the following log-Demons energy with respect to $v_{i\text{-}j}$ and the auxiliary SVF v_c [Cachier 2003]. Instead of minimising a global energy, a correspondence field v_c is introduced, so that two simple, fast, and more efficient minimisation steps are performed, respectively for E_{Sim} and E_{Reg}. In the first step, E_{Sim} is minimised using a gradient descent method, whereas in the second step E_{Reg} can be solved explicitly as the Gaussian convolution of v_c when the regularisation term is chosen adequately:

$$E(v_{i\text{-}j}, v_c, I_i, I_j) = \overbrace{\frac{1}{\sigma_i^2}\mathrm{Sim}(v_c, I_i, I_j) + \underbrace{\frac{1}{\sigma_x^2}\mathrm{Corr}(v_{i\text{-}j}, v_c) + \frac{1}{\sigma_T^2}\mathrm{Reg}(v_{i\text{-}j})}_{E_{Reg}(v_{i\text{-}j}, v_c)}}^{E_{Sim}(v_{i\text{-}j}, v_c, I_i, I_j)} . \tag{2.1}$$

In this formula, σ_i is the parameter linked to the noise in the image, σ_x is linked to the uncertainty of the matching in the correspondence term, σ_T is the regularisation weight, **Sim** is the similarity criterion, **Reg** the regularisation term, and **Corr** is the correspondence term that links $v_{i\text{-}j}$ to v_c. The LCC log-Demons [Lorenzi 2013a] uses ρ the Local Correlation Coefficient (LCC) similarity metric [Cachier 2003] since it is robust to local intensity artifacts:

$$\rho(I_i, I_j) = \int_\Omega \frac{\overline{I_i I_j}}{\sqrt{\bar{I}_i^2 \bar{I}_j^2}} \text{ with } \bar{I} = G_\sigma * I(x),$$

where G_σ is the Gaussian smoothing operator with a kernel size of σ and Ω is the image domain.

Therefore, by considering the symmetric resampling $I_i' = I_i \circ \exp(\frac{v_c}{2})$ and $I_j' = I_j \circ \exp(-\frac{v_c}{2})$, the first term of equation (2.1) can be written as:

$$\mathrm{Sim}(v_c, I_i, I_j) = \rho^2(I_i', I_j') = \rho'^2(v_c, I_i, I_j) = \frac{\left[\overline{I_i \circ \exp(\frac{v_c}{2}).I_j \circ \exp(-\frac{v_c}{2})}\right]^2}{\overline{\left[I_i \circ \exp(\frac{v_c}{2})\right]^2}.\overline{\left[I_j \circ \exp(-\frac{v_c}{2})\right]^2}}.$$

If we define the update field $\delta v_{i\text{-}j}$ through the zeroth order term of the Baker-Campbell-Hausdorff (BCH) formula [Bossa 2007]:

$$\delta v_{i\text{-}j} = \log(\exp(-v_{i\text{-}j}) \circ \exp(v_c)) \approx -v_{i\text{-}j} + v_c,$$

then in the first part of the alternate optimisation of equation (2.1), E_{Sim} has to be minimised with respect to $\delta v_{i\text{-}j}$:

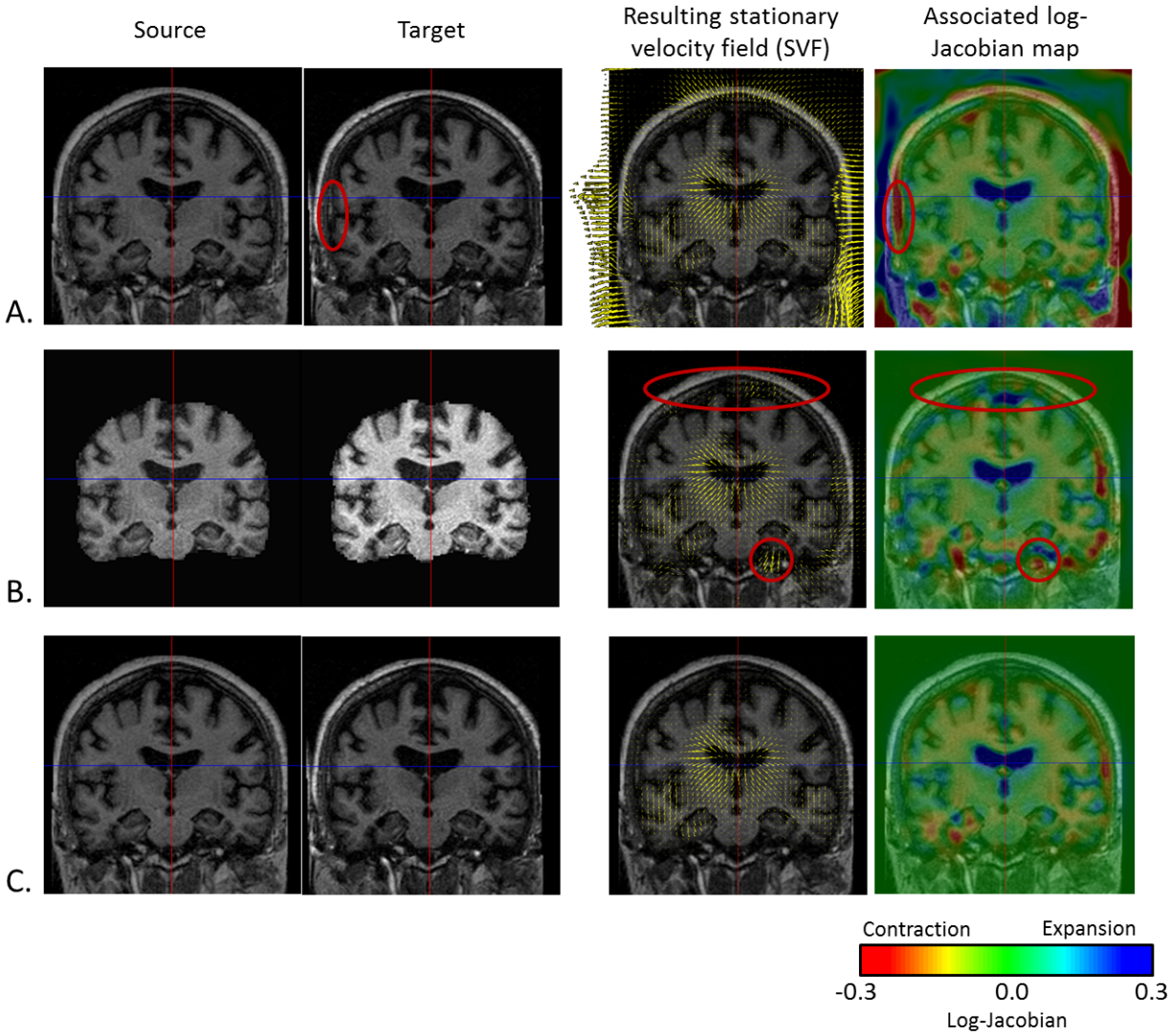

Figure 2.4: *Comparison of three non-linear diffeomorphic registration methods*: **First and second column:** we see the intensity bias affecting the source and target images. **A. Registration of the head with no confidence mask:** strong deformation fields are estimated in the skull and meninges that diffuse to the outer cortex region and bias the results (cf. red circle where a non-realistic expansion of 38% is found). **B. Registration of the skull-stripped images (no confidence mask):** the use of the skull-stripped images biases the result at the level of the outer cortex (cf. red circles) where non-existing high value deformations are found due to the high intensity gradient). In fact, skull-stripping imposes the outside brain intensity to be zero creating a high intensity gradient that biases the registration results (the update $\delta v_{i\text{-}j}$ is directly proportional to the image gradient). **C. Registration of the head with confidence mask:** the registration using the confidence mask enables us to estimate realistic transformation in the outer cortex.

$$E_{Sim}(\delta v_{i\text{-}j}, I_i, I_j) = -\frac{1}{\sigma_i^2}\rho'^2(\delta v_{i\text{-}j}, I_i, I_j) + \frac{1}{\sigma_x^2}||\delta v_{i\text{-}j}||^2,$$

with $\mathrm{Corr}(v_{i\text{-}j}, v_c) = ||\log(\exp(-v_{i\text{-}j}) \circ \exp(v_c))||^2 = ||\delta v_{i\text{-}j}||^2$. In the second part of the optimisation, E_{Reg} should be minimised with respect to $v_{i\text{-}j}$:

$$E_{Reg}(v_{i\text{-}j}, v_c) = \frac{1}{\sigma_x^2}||\log(\exp(-v_{i\text{-}j}) \circ \exp(v_c))||^2 + \frac{1}{\sigma_T^2}Reg(v_{i\text{-}j}).$$

The registered images generally comprise the brain and its surrounding skull which can lead to corrupted results. In fact, the resulting deformation field generally exhibits high values in the region of the meninges and the skull that diffuse through regularisation in the outer cortex (see Figure 2.4), potentially yielding to misleading discoveries.
One solution is to only register the brain tissues and the cerebrospinal fluid (CSF) obtained through skull-stripping. However, this solution may be prone to errors (small parts of the outer cortex could be cropped) and puts the outside brain intensity to zero creating a high intensity gradient that biases the registration results (as shown on Figure 2.4), since the update $\delta v_{i\text{-}j}$ is directly proportional to the image gradient.

Therefore, we modified the LCC log-Demons algorithm to incorporate the use of a confidence mask as proposed by [Brett 2001], and first introduced in the Demons algorithm by [Stefanescu 2004]. We consider that we do not want to align the structures outside the brain (skull, meninges,...). Therefore, the voxels outside the brain should have no influence in the similarity minimisation step. We define a probabilistic mask $\omega(x)$ such that its value is $\omega(x) = 1$ for a voxel inside the brain, $\omega(x) = 0$ outside, and in-between depending on the confidence we have for the voxel. The new log-Demons energy to minimise is:

$$E(v_{i\text{-}j}, v_c, I_i, I_j) = \boldsymbol{\omega}\frac{1}{\sigma_i^2}\mathrm{Sim}(v_c, I_i, I_j) + \frac{1}{\sigma_x^2}\mathrm{Corr}(v_{i\text{-}j}, v_c) + \frac{1}{\sigma_T^2}\mathrm{Reg}(v_{i\text{-}j}).$$

Thus, only the first part of the minimisation (E_{Sim}) is modified and we still get a closed-form solution leading to an effective computational scheme for the optimisation of E_{Sim} (cf. demonstration in Appendix 2.4):

$$\delta v_{i\text{-}j} = \begin{cases} -\frac{2\Lambda}{||\Lambda||^2 + \frac{1}{\omega}\frac{4}{\rho^2}\frac{\sigma_i^2}{\sigma_x^2}}, \text{ if } \omega > 0 \\ 0, \text{ if } \omega = 0 \end{cases}$$

with

$$\Lambda = \frac{G_\sigma * (I_i \nabla I_j^T)}{G_\sigma * (I_i I_j)} - \frac{G_\sigma * (I_j \nabla I_i^T)}{G_\sigma * (I_i I_j)} + \frac{G_\sigma * (I_i \nabla I_i^T)}{G_\sigma * (I_i^2)} - \frac{G_\sigma * (I_j \nabla I_j^T)}{G_\sigma * (I_j^2)}. \tag{2.2}$$

In order to keep a symmetric formulation of the registration, the probabilistic mask ω is defined using two masks. The first one is the brain mask M of the moving

image and the second one is the brain mask F of the fixed image. The mask ω is then defined as the average of the symmetric resampling of the two brain masks in the halfway space:

$$\omega = \frac{1}{2}\left[M \circ \exp(\frac{v_c}{2}) + F \circ \exp(-\frac{v_c}{2})\right]$$

Hence, the registration problem is still defined on the whole image domain but the update is weighted differently depending on the confidence on the brain areas. In our experiments, we defined the initial brain masks (for both fixed and moving images) as binary masks.

2.2.3.3 Estimation of the Subject-Specific Longitudinal Trajectory via Fully Symmetric SVF Regression

Given the previously estimated series of longitudinal deformations $\phi_{i\text{-}j} = \exp(v_{i\text{-}j})$ with $0 \leq i < j \leq N-1$ for a subject, we then model the *subject-specific longitudinal deformation trajectory* $\hat{\phi}$ as :

$$\hat{\phi}(x,t) = \exp(t \cdot \hat{v}(x)) \text{ with } t \in \mathbb{R},$$

where $\hat{v}$ is the best fit of a fully symmetric linear model in time - through the origin - of the series of SVFs $v_{i\text{-}j}$:

$$\hat{v} = \arg\min_{v} \sum_{0 \leq i < j \leq N-1} \|(t_j - t_i)v - v_{i\text{-}j}\|^2 = \frac{\sum_{0 \leq i < j \leq N-1}(t_j - t_i)v_{i\text{-}j}}{\sum_{0 \leq i < j \leq N-1}(t_j - t_i)^2}.$$

This model uses all the possible combinations of SVFs $v_{i\text{-}j}$ between the different time points while using the symmetry of the pairwise registration ($v_{i\text{-}j} = -v_{j\text{-}i}$) to simplify the problem. $\hat{v}$ and $\hat{\phi}(t=1) = \exp(\hat{v})$ represent the subject-specific evolution trajectory over a year. One should note that a linear model of the longitudinal SVFs does not lead to a linear model of the deformations. For up to three time points, our experience showed that a linear model in time is sufficient to explain the data. A higher-order model could be used for a higher number of time points at the cost of increasing the statistical complexity.

2.2.3.4 Unbiased Study-Specific Template Construction

In order to compare all the subject-specific longitudinal deformation trajectories , we need to have these deformations normalised in the same common reference anatomy called study-specific template A. Although each subject brain is normalised to the standard space (cf. Section 2.2.2.2), the affine alignment is not sufficient to compensate for the local anatomical differences (there is no voxel-to-voxel correspondence yet between the different anatomies). Among the available methods for the template construction, we chose to use the method from [Guimond 2000] consisting in the iterative averaging of intensities and deformations.

This iterative process is described in Algorithm 2 and illustrated on 136 subjects (Figure 2.5). In the following experiments, the iterative algorithm was stopped at the seventh iteration. At a given iteration there are two successive image resamplings due to the application of two deformations; this can bias the centering of the template. To ensure it is centered, we minimise the number of image resamplings at a given iteration by using a zeroth order term of the BCH: $\log(\exp(v_k^i) \circ \exp(-\bar{v}_k)) \approx v_k^i - \bar{v}_k$. Moreover, a good practice for the selection of the initialisation image for A_0 is to manually choose a subject image that is roughly centered with respect to the considered sample in order to avoid being blocked in a local minimum. In practice, we checked that changing the reference image for A_0 changed the final template A by only a negligible amount as shown on Figure 2.5.

Algorithm 2 Creation of an Unbiased Template A

Input: Set of study images I^i
Output: A: Study-specific template image

Initialisation: Select a reference image I^j among the M subjects images
$A_0 = I^j$

repeat

Non-linearly register the images to A_k
$A_k \approx I^i \circ \exp(v_k^i)$

Mean stationary velocity field
$\bar{v}_k = \frac{1}{M}(\sum_{i=1}^{M} v_k^i)$

Resample subjects' image
$L_k^i = I^i \circ \exp(v_k^i - \bar{v}_k)$

Template iteration k+1: Mean intensity image
$A_{k+1} = \frac{1}{M}(\sum_{i=1}^{M} L_k^i)$

until Variations of A_k and $\bar{v}_k$ are very small:
$\frac{1}{V}\sum_{i=1}^{V}(A_{k+1}(i) - A_k(i))^2$ and $||\bar{v}_{k+1} - \bar{v}_k|| < \varepsilon$

$A = A_{k+1}$

Here again the non-linear registrations are performed using our modified LCC log-Demons algorithm with confidence mask (we used the subjects images masks), in order to estimate the study-specific template while being robust to the artifacts on the brain boundaries.
Another point concerns the choice of the time point at which the template is created. There is no golden rule and the choice of the time point is usually let to the user.

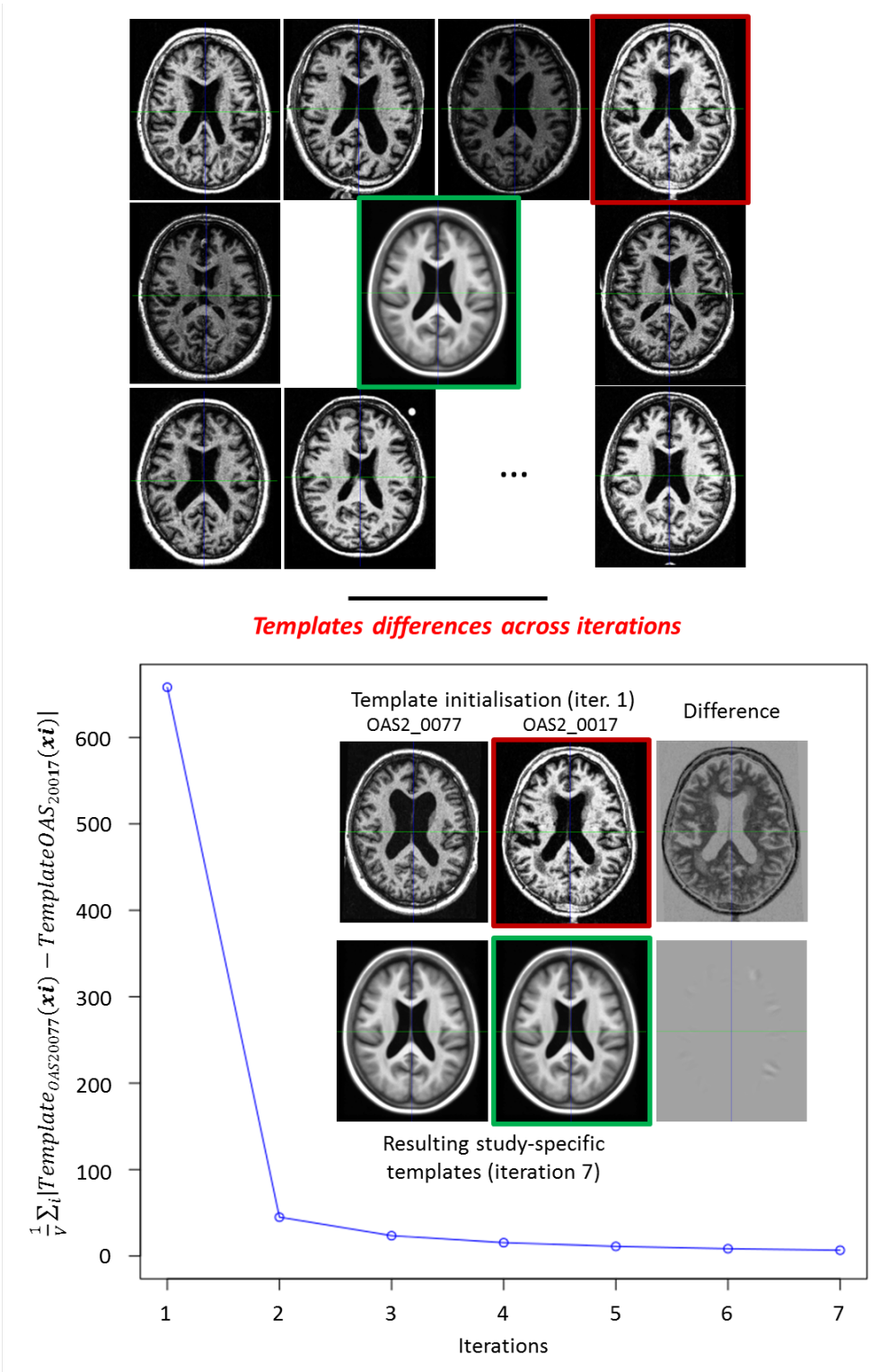

Figure 2.5: *Top: Iterative template Construction*: Example of the construction of the template (green frame) of a study of 136 subjects, 9 subjects are displayed. Red frame: the reference subject (OAS2_0017) used for the initialisation.*Bottom: Influence of the reference subject used to initialise the study-specific template*: We built a second study-specific template by initialising it with a different subject (OAS2_0077). We computed $\frac{1}{V}\sum_i |Template_{OAS2_0077}(x_i) - Template_{OAS2_0017}(x_i)|$ over the brain mask at each iteration. Although the initial reference images are dissimilar, we obtain two very similar templates at the end.

As for us, we use the images I_0 at the first time point t_0 to create the template.

2.2.3.5 Parallel Transport of the Subject-Specific Longitudinal Stationary Velocity Field

Now that a common brain anatomical image is defined, we need to express each subject-specific longitudinal deformation trajectory $\hat{\phi}$ in the template anatomy to be able to compare them. To do so, we use the parallel transport computed with the Pole ladder [Lorenzi 2013b] of the subject-specific longitudinal SVF trajectory $\hat{v}$ along the inter-subject SVF w_0 parametrising the cross-sectional transformation $\psi_0 = \exp(w_0)$ that maps I_0 to A (cf. Figure 2.6).

The result is $\hat{v}_T = \Pi_{w_0}(\hat{v})$, the subject-specific longitudinal SVF trajectory normalised in the template space. We can then compute the subject-specific longitudinal deformation trajectory in the template space $\hat{\phi}_T = exp(\hat{v}_T)$. The different steps necessary for the parallel transport are described in algorithm 3. It is then

Algorithm 3 Pole Ladder for the Parallel Transport of the Longitudinal Stationary Velocity Field

Input: $\hat{v}$: subject-specific SVF, I_0: subject image where the SVF is normalised and the template A

Output: $\hat{v}_T$: subject transported SVF

Non-linearly register the subject image I_0 into A

$A \approx I_0 \circ \exp(w_0)$

Parallel transport of $\hat{v}$ along w_0

Scaling step: find n such that w_0/n is smaller than 0.5 voxel in all dimensions

$n = ceiling(\frac{\max_{x\in\Omega} ||w_0(x)||}{0.5 \cdot voxelsize})$

repeat

Ladder step:

$v_k = \hat{v} + [\,\frac{w_0}{n}, \hat{v}] + \frac{1}{2}[\,\frac{w_0}{n}, [\,\frac{w_0}{n}, \hat{v}]\,]$ with $[\,,]$ the Lie brackets:

$[v, w](x) = D_w v(x) - D_v w(x) = \sum_i (w_i(x)\frac{\partial v_i(x)}{\partial x_i} - v_i(x)\frac{\partial w_i(x)}{\partial x_i})$

where $v_i(x)$ and $w_i(x)$ are respectively the components of the vector fields $v(x)$ and $w(x)$ in a Cartesian coordinates system of the point x with coordinates x_i. The numerical computation of the derivatives is performed using a centred difference scheme.

Let $\hat{v} = v_k$

until $k = n$

$\hat{v}_T = \Pi_{w_0}(\hat{v}) = v_n$

possible to perform a statistical analysis on these transported subject-specific longitudinal stationary velocity fields $\hat{v}_T$ as shown in section A.3.

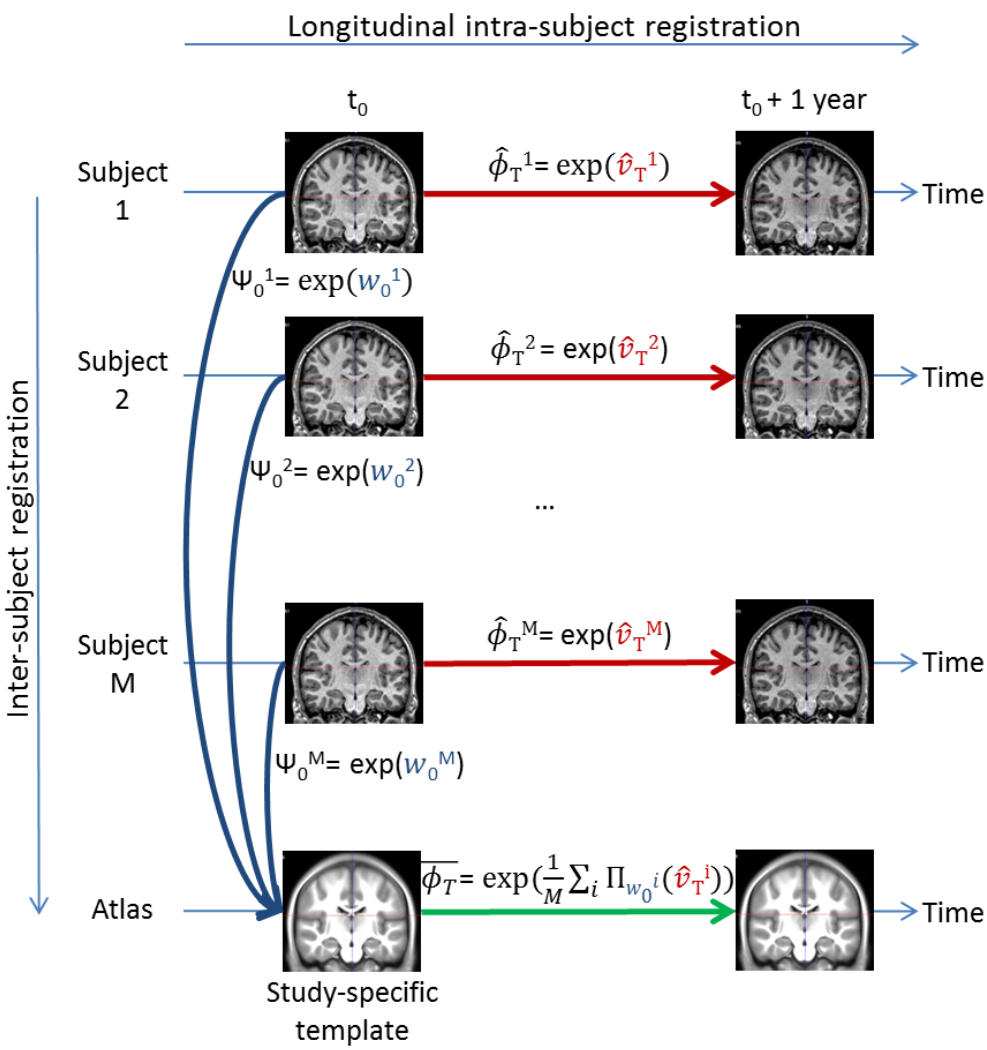

Figure 2.6: *Illustration of the parallel transport of the study transformations to the study-specific template*: After each subject longitudinal SVF transport, the mean transformation $\bar{\phi}_T$ is computed by taking the exponential of the average of all the transported subject-specific longitudinal SVFs $\Pi_{w_0^i}(\hat{v}_T^i)$.

2.3 Application to the Analysis of the Longitudinal Changes in Alzheimer's Disease

The aim of this section is to show an application of the proposed processing pipeline. We focused our illustration on Alzheimer's disease, a neuro-degenerative disease that causes dramatic changes in the brain anatomy over time. We use the Open Access Series of Imaging Studies (OASIS) database [Marcus 2010].

2.3.1 OASIS database

The clinical cohort considered in this study is composed of 64 patients diagnosed with very mild to moderate Alzheimer's disease, and 72 healthy individuals. For these subjects, 2 to 5 longitudinal brain acquisitions (T1 Magnetic Resonance Imaging) were available, corresponding to a follow-up time $t_{0\text{-}j} = t_j - t_0$ of 0.5 to 6.9 years. Further information can be found in Appendix 2.4.

2.3.2 Methods and Results

After applying the processing pipeline to the database (the parameters used for the different steps are summarised Table 3.1), we obtain the transported subject-specific longitudinal deformation trajectories $\hat{\phi}^i_T(t) = \exp(t \cdot \hat{v}^i_T)$ for each subject i in the study-specific template: we thus get 72 subject-specific longitudinal SVFs $\hat{v}^i_T$ for the healthy controls and 64 for the patients with Alzheimer's disease.

Concerning the non-linear registration parameters for the LCC log-Demons, the optimal parameters we propose here would of course be different for another study, and we recommend to fine-tune in priority the amount of regularisation (-b) and the number of iterations (-a). As for the SVF exponentiation (and log-Jacobian), all the computations were performed using an Euler forward integration scheme (option -z 1 in `SVFLogJacobian` tool).

Pipeline Step	Parameters Values
Standard reorientation	*Default*
Field of view reduction	*Default*
Intensity inhomogeneity corr.	*Default*
Skull-stripping	*Default*
Longitudinal rigid registration	-cost normcorr -interp spline -dof 6
Affine registration	-cost normcorr -interp spline
Non-linear reg.: Intra and Inter-subject	-r 2 -R 1 -C 3 -a 30x20x10 -x 0 -b 2.0 -S 0.15 -u 3.0 -V
Transport	*Default*

Table 2.1: Parameters used for each module of the longitudinal study.

Before discussing the results of the group-wise comparisons of the longitudinal evolutions, let us focus on an illustrative result concerning a single subject (OAS2_0002).

We computed the log-Jacobian map - which quantifies the relative volume changes associated to the longitudinal deformation - for the SVF $v_{0\text{-}2}$ of the longitudinal evolution between t_0 and t_2; the result can be seen on Figure 2.7. We can observe the expansion of the ventricles and more particularly in the temporal horn of the lateral ventricles, as well as the contraction in the hippocampi. Moreover, there exists an artifact outside the brain (left hand edge of the follow-up image on Figure 2.7). The use of the non-linear registration with confidence mask enables us to avoid any artifactual volume change in our log-Jacobian map and therefore provides more stable results. This is illustrated in Figure 2.4, where we compare the deformation found with and without the use of the confidence mask; we see on the left hand of the image (red circle on image A.) that this kind of artifact can locally bias the estimation of longitudinal deformations when the mask is not explicitly accounted for [Ashburner 2013].

Concerning the groups study, we consider the subject-specific deformations over a year ($t = 1$) so that we study the SVFs $\hat{v}_T^i$. It is then possible to visualise the mean volume changes during one year for each group of patients with Alzheimer's disease and healthy controls. After computing the average SVF for the non-demented group and the Alzheimer's one, we compute the associated log-Jacobian maps[5] (cf. Figure 2.8), and compare the modelled group-wise evolutions. We can see that the main expansion region is located in the lateral ventricles with higher values for the Alzheimer's patients group than for the healthy control one. Moreover, for the patients with Alzheimer's disease we can see an expansion in the temporal horn of the lateral ventricles that does not exist in the control group. Finally, the atrophy is higher for the Alzheimer's patients and mainly located in several parts of the white matter, in the thalamus and in the hippocampi whereas there is no visible contraction in the hippocampi or in the thalamus for the control group. These results are coherent with the findings reported in the literature [Braak 1991, Fox 1996, Jack 2004, Schott 2005, de Jong 2008].

2.3.2.1 Two-Sample t-Test: Alzheimer's patients versus Healthy Controls

We now statistically investigate the group-wise differences between the modelled longitudinal evolutions of the Alzheimer's patients group and the healthy control group by using a voxel-wise two-sample t-test on the log-Jacobian maps. For illustrative purposes, we show here a standard univariate analysis on a scalar map, but the use of the parallel transport in our pipeline enables us to do statistics directly on the subject-specific SVFs as shown in 2.3.2.3. The null hypothesis is that there exists no difference between the mean of the two groups.

We used SPM8 (see [Friston 2007]) for this test and corrected for multiple testing using the Family-Wise Error rate (FWE) with a corrected p-value of 0.05 in order to control for the same level of specificity. The t-test was limited to the brain mask. The

[5] The log-Jacobian maps (for OAS2_0002 and the different groups) are available on NeuroVault [Gorgolewski 2015] at http://neurovault.org/collections/YBADDEIH/

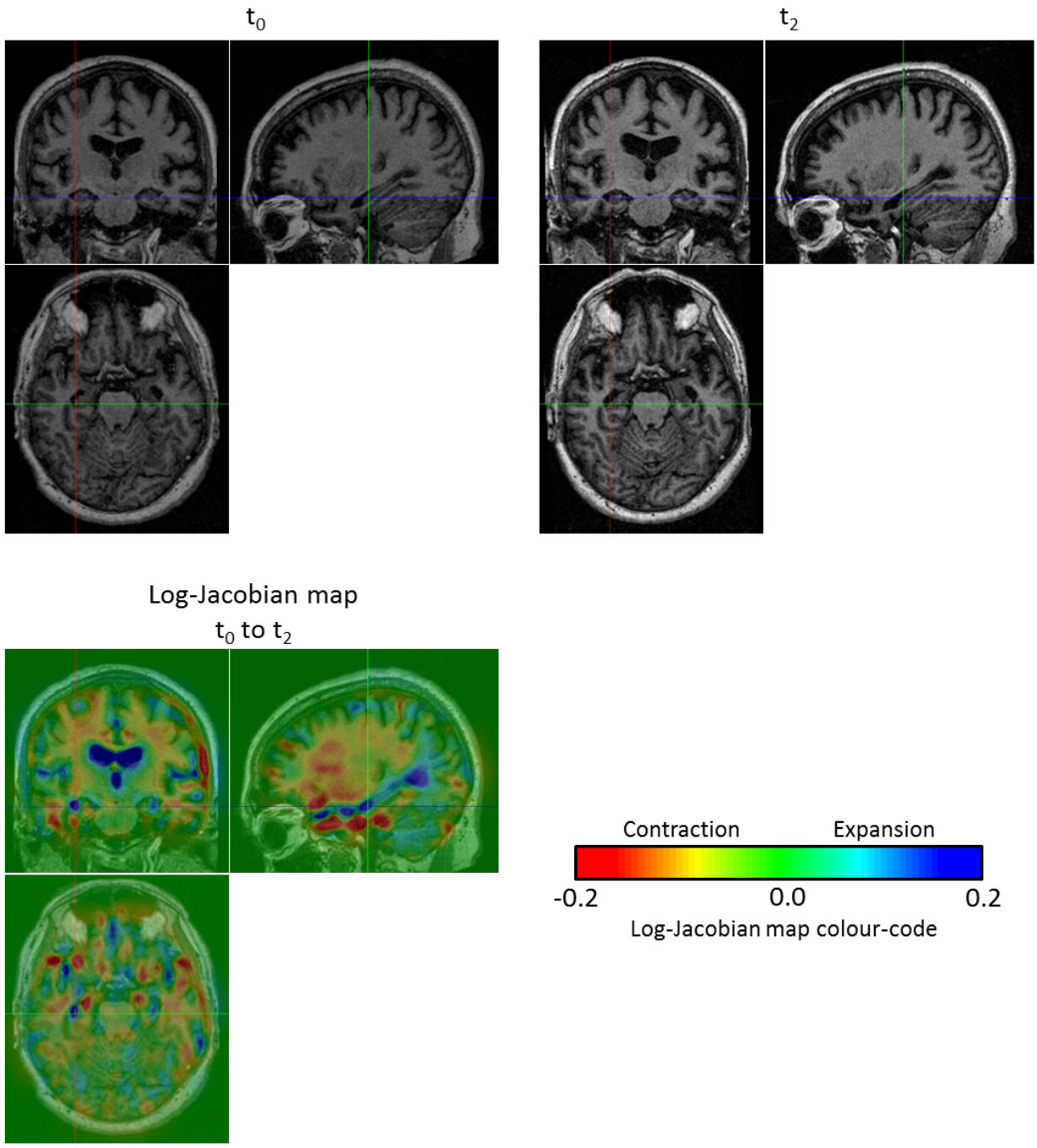

Figure 2.7: *Log-Jacobian map for the subject OAS2_0002*: We computed the log-Jacobian map - which represents the relative change of volume - for the SVF of the longitudinal evolution between t_0 and t_2. We can observe an expansion in the ventricles and more particularly in the temporal horn of the lateral ventricles and a contraction in the hippocampi. Moreover, although there is an artifact outside the brain (left hand edge of the follow-up image at t_2), the use of the non-linear registration with confidence mask enables us to avoid any artifactual volume change in our log-Jacobian map.

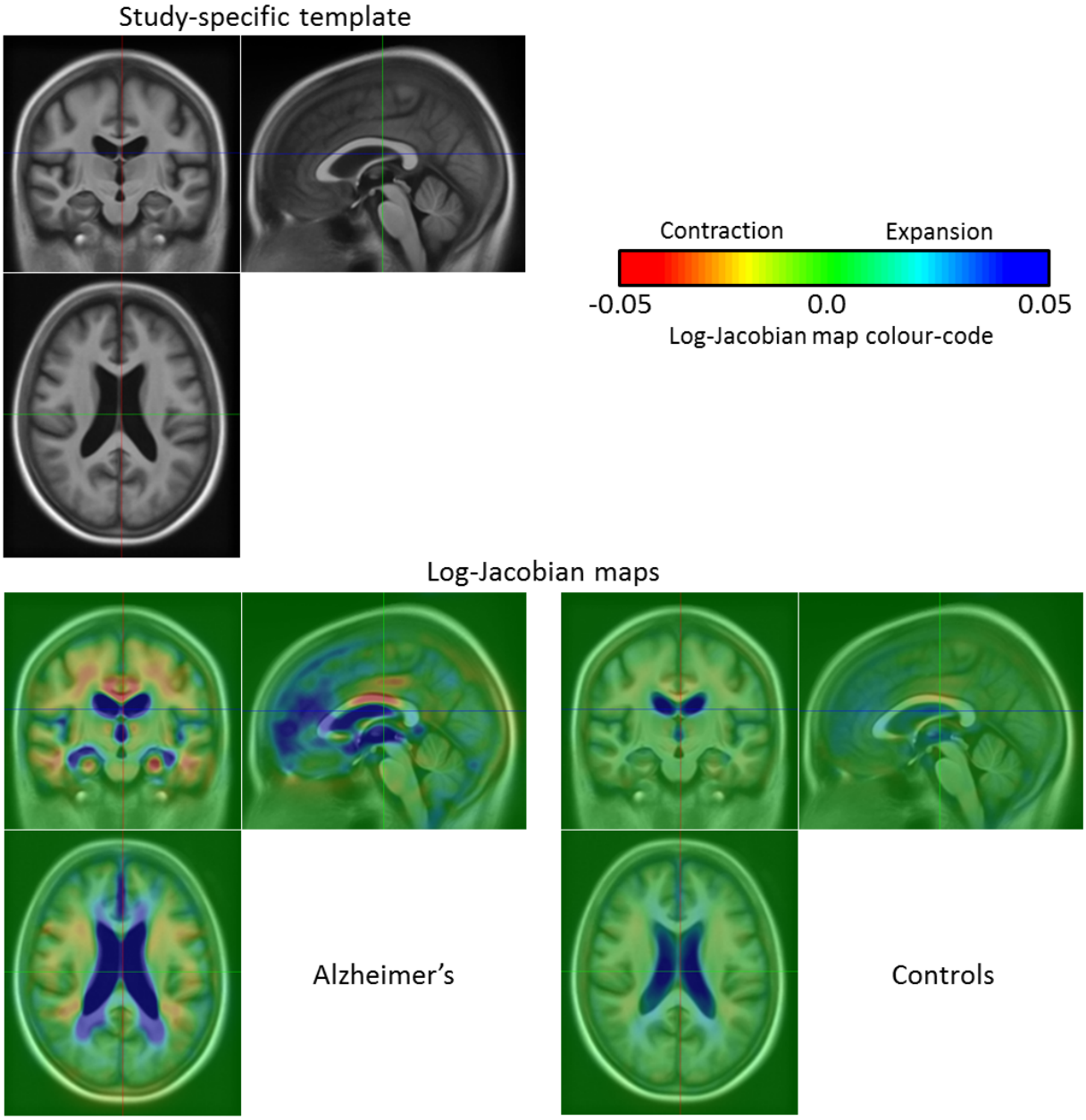

Figure 2.8: *Template for the 136 OASIS subjects at t_0 and log-Jacobian maps (one year evolution) of the patients with Alzheimer's disease and the healthy control group*: The main expansion region concerns the lateral ventricles where the Alzheimer's patients exhibit higher values when compared to the healthy subjects. Moreover, for the patients with Alzheimer's disease we can see an expansion in the temporal horn of the lateral ventricles that does not exist in the non-demented control group. Finally, the atrophy is higher for the Alzheimer's patients and mainly located in several parts of the white matter, in the thalamus and in the hippocampi whereas there is no visible contraction in the thalamus or the hippocampi for the healthy group.

result map with the thresholded t-values can be seen on Figure 2.9. The statistically different volume changes occur in the lateral ventricles, more particularly in the temporal horn, and also in the thalamus.

2.3.2.2 Reliability of the LCC log-Demons with a Confidence Mask

We tested the reliability of the implemented LCC log-Demons registration with a confidence mask. We compared it with the original LCC log-Demons applied to full head images or skull-stripped images. We therefore ran three similar processing pipelines where the only difference was the non-linear registration method used; the processing pipeline using the LCC log-Demons with a confidence mask is denoted as **LLDF**, the one using the registration of the whole head is called **Pipeline Head**, and the pipeline registering skull-stripped images is denoted as **Pipeline Skull-stripped**. Similarly to 2.3.2.1, we investigated the differences between the Alzheimer's patients group and the healthy control group in each case and compared the obtained results to see which method has the highest statistical sensitivity to find volume changes between the two groups. We notice that on average the LCC log-Demons registration with confidence mask is faster (around 1.1 times) than the LCC log-Demons registration of the full head.

The three corrected t-maps are presented Figure 2.9[6]. The three results present similar patterns with most of the statistical differences in the ventricular region and more particularly in the temporal horn of the lateral ventricles. Other statistical differences can be found in the thalamus. The volume of the regions of statistical significant differences are 10.4, 16.5 and 17.5 cm^3 for respectively 'Pipeline Skull-stripped', 'Pipeline Head', and 'LLDF'. Moreover, the t-values are higher with the 'LLDF' than with the two other methods. In average on the same statistical region (the smallest region, obtained by computing the intersection of the three statistically significant regions), we obtain an absolute t-value of 6.13 with 'LLDF' against 5.98 with 'Pipeline Head', and 5.69 with 'Pipeline Skull-stripped'. This increase of the t-values can be explained by the increased group difference for 'LLDF' compared to the group differences of the other two methods and not by a reduction of the variance. On the same statistical region, we observe a relative increase of 23.4% with respect to 'Pipeline Head' and of 23.7% with respect to 'Pipeline Skull-stripped'. Therefore, the LLDF pipeline enables us to have an increased statistical sensitivity with no decrease of the specificity.

2.3.2.3 Illustration of a DBM analysis: Hotelling's two-sample T^2-test

Finally, we illustrate the main advantage of the LLDF: by using the parallel transport in our pipeline it is then possible to perform statistics directly on the subject-specific longitudinal trajectories. We therefore perform a multivariate Hotelling's two-sample T^2-test to show the group-wise differences between the modelled subject-

[6]The t-maps as well as the group diference and estimated variance maps for the three methods are available at http://neurovault.org/collections/YBADDEIH/

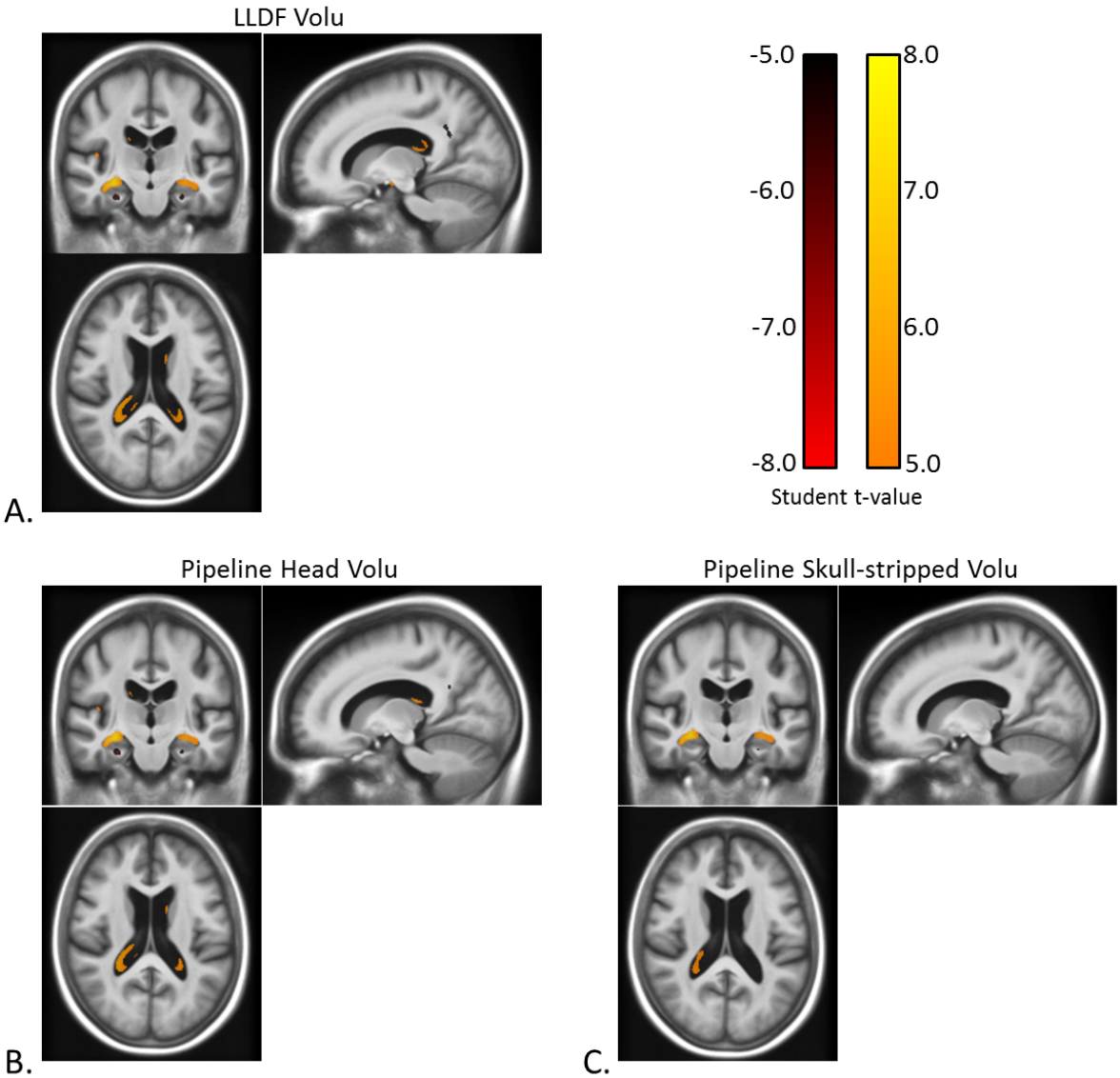

Figure 2.9: *Corrected t-statistic map for the volume changes differences between the patients with Alzheimer's disease and the healthy control group (for the 3 registration methods) on one slice*: The three results present similar patterns with statistical differences in the ventricular region, more particularly in the temporal horn of the lateral ventricles, and also in the thalamus. The volume of the regions of statistical significant differences are 10.4, 16.5 and 17.5 cm^3 for respectively 'Pipeline Skull-stripped', 'Pipeline Head', and 'LLDF'. Moreover, the t-values are higher with the 'LLDF' than with the two other methods. (Correction for multiple testing using the Family-Wise Error rate with a corrected p-value of 0.05).

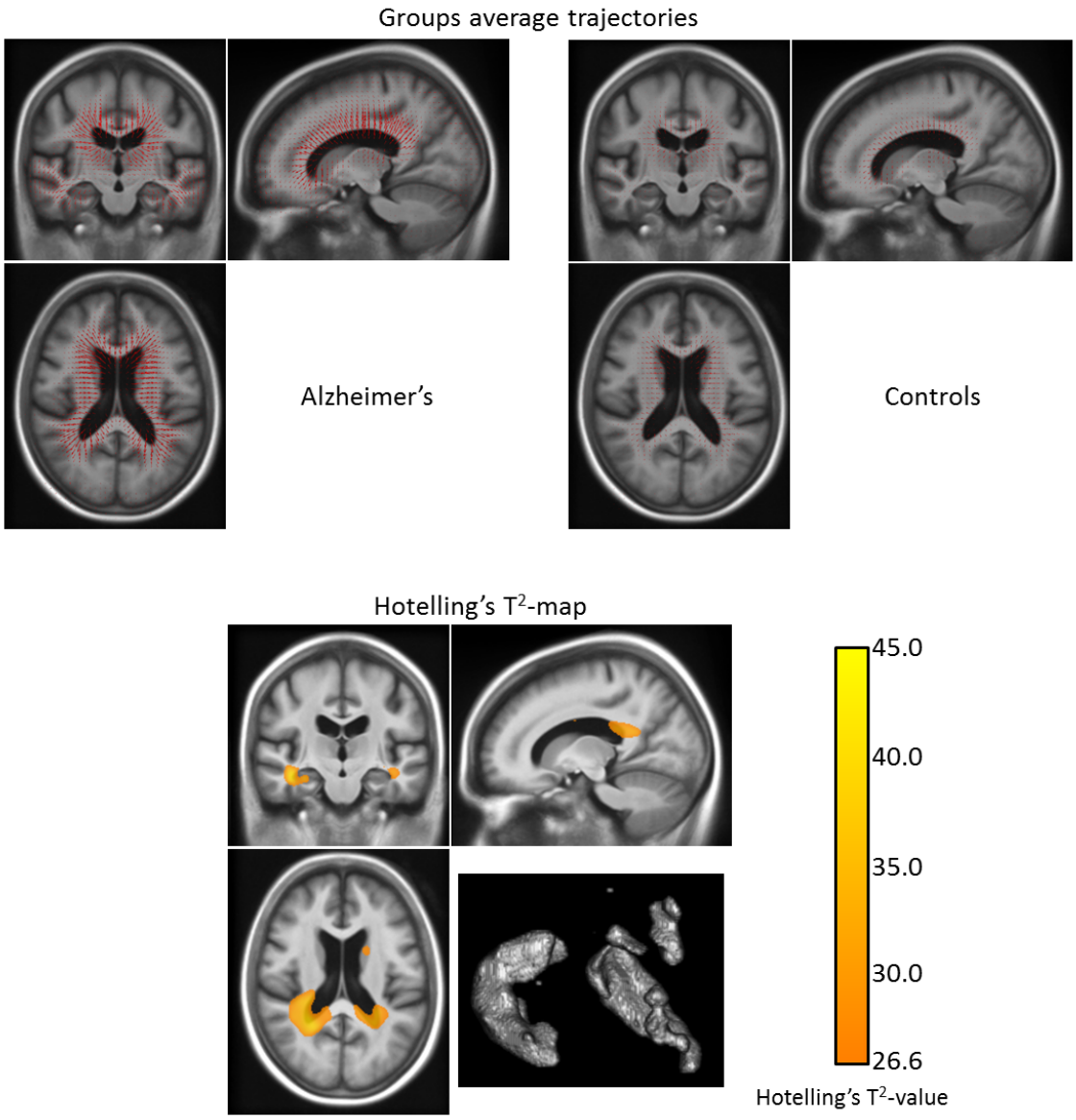

Figure 2.10: *Top: Group longitudinal trajectories for the patients with Alzheimer's disease and the healthy control group (obtained with the LLDF method)*: We can see that the mean trajectory for the demented group has a higher magnitude than the control one. *Bottom: Corrected T^2-map for the longitudinal trajectories differences between the patients with Alzheimer's disease and the healthy control group (for the LLDF method) on one slice*: The statistical differences between the demented and the control groups are located in the lateral ventricles, in the temporal horn of the ventricles, in the hippocampi, and in the caudate nuclei. The volume of the regions of statistical significant differences is 41.0 cm^3. (The Hotelling's T^2-test was corrected for multiple testing using 5000 permutations and the map is thresholded for a corrected p-value of 0.05).

specific longitudinal trajectories of the Alzheimer's patients group and the healthy control group - obtained using the confidence mask. The null hypothesis is that there exists no difference between the mean of the two groups. We corrected for multiple testing using 5000 permutations and we limited the test to the brain mask. The resulting T^2-map thresholded for a corrected p-value of 0.05 can be seen on Figure 2.10 [7].

We can see that the statistical differences between the demented and the control groups are located in the lateral ventricles, in the temporal horn of the ventricles, in the hippocampi and, in the caudate nuclei. The volume of the regions of statistical significant differences is larger than the one found using the univariate test: 41.0 cm^3. The observed differences in the statistically significant regions between the univariate t-test (cf. 2.3.2.1) and the multivariate Hotelling's T^2-test can be explained by the fact that in the first case the study is restricted to the volumetry only whereas in the second case it focuses on the displacement field - which in addition to the volumetry also includes translations and rotations. With this difference in mind, we can say that the patterns found in the two tests are coherent. For example concerning the caudate nuclei, although there is no statistically significant differences in the volume changes between the patients with Alzheimer's and the healthy subjects, there exist statistically significant differences in the displacements of the caudate nuclei between the two groups.

2.4 Conclusion and Discussion

We proposed and detailed a new processing pipeline [8] for the longitudinal analysis of image data series. It is based on freely available software and tools so that anyone can reproduce our study, use this pipeline to replicate and verify findings conducted with other pipelines or use it to perform new studies. Moreover, we also implemented a masking of the similarity term in the non-linear registration (with a formulation that ensures symmetry). It enhances the robustness of the registration results with respect to intensity artifacts in the boundary of the brain, thereby increasing the sensitivity of the statistical studies done on the longitudinal deformations. We finally showed on an open-access database that the results obtained with this pipeline are consistent with the findings from the literature.

The use of the parallel transport in our pipeline enables us to perform both standard univariate analysis on a scalar map and also statistics directly on the SVFs as illustrated by the multivariate Hotelling's T^2-test. Therefore, changes other than the ones linked to volumetry (like rotations or translations of the brain structures) could be studied. Concerning the confidence mask, initialising it with probabilistic masks of the fixed and moving image (instead of binary ones) could be used to take into account the uncertainty linked to the skull-stripping at the brain boundaries.

[7] The T^2-map is available at http://neurovault.org/collections/YBADDEIH/

[8] The whole pipeline will be released as a complement of the already available LCC log-Demons software.

However, in our experiment the use of binary masks was sufficient to increase the sensitivity of the statistical group-wise analysis while not decreasing the specificity. Intensities artifacts inside the brain such as prominent blood vessels could also be incorporated in the confidence mask if a blood vessels segmentation was available (a method similar to [Samaille 2012] could be applied).

The most important issue for the longitudinal processing pipeline is related to the asymmetry biases [Ridgway 2015] that need to be avoided in the processing. Two types of asymmetries can be distinguished. The first one, described in [Reuter 2011] and [Yushkevich 2010], is introduced by the resampling of all the follow-up images except the baseline. In our case, all the images (including the baseline image I_0 at $t = 0$) are resampled once and only once in the common reference space. In the case of the follow up images, the transformation used to resample the image is the combination of a rigid and an affine transformation (cf. 2.2.2.3), whereas in the case of the baseline image, we use the subject to reference space affine transformation only. This aspect of the pipeline has some similarity to that of [Rohrer 2013] - where again, some repeated interpolations are avoided, while other interpolations are symmetric by virtue of being in MNI space rather than in the native baseline space. It could be possible to go one step further and to avoid any explicit interpolation by initialising the non-linear registration (in the LCC log-Demons software) with the combined affine/rigid transformation using the software parameter: –initial-linear-transform. However, this would still imply an implicit internal resampling and in this case we would no longer follow the assumption made in LDDMM and the SVF framework that all the field tends towards zero when we get away from the center of the image (i.e. beyond the borders of the image). In practice, we observe edge-effects and a proper way to deal with the problem should be to revise the LCC log-Demons algorithm in order to explicitly handle the two transformation separately and make sure that the criterion (and the discretisation) would be affine invariant.

The second type of bias is related to the non-centrality of the time point where the subject longitudinal deformations are computed (also referenced as favoring a particular time point). Several non-stationary velocity fields-based methods (LDDMM) have taken great care of that [Avants 2011, Niethammer 2011b, Ashburner 2013]. In these methods, the initial velocity (or equivalently the momentum map) is different at different time points along a geodesic. In that case, for more than two time points, it is necessary to choose a time point for the subject-specific template, and this time point is generally the average (or median) of the observed time points. The momentum maps (from the template to all the time points) can then be compared in the template reference space only. In the stationary velocity field framework, the velocity field is - by definition - stationary. Thus, the SVF resulting from the registration is the same all along the trajectory: it is not expressed in material coordinates at a specific time point but in Eulerian coordinates which are not attached to a given time point. Therefore, in the symmetric LCC log-Demons any subject time point can be chosen to perform the pairwise registrations without needing a subject-specific template. Moreover, the annualised log-Jacobian map is valid for all time points even if its value for a material point

changes with time along its trajectory. Finally, even if each registration is fundamentally pairwise , the effect of the multiple time points is taken care of using the fully symmetric linear model in time described in section 2.2.3.3. This model uses all the possible combinations of SVFs in order to avoid favouring any specific time point. Notice that this approach is sub-optimal with unbalanced data where large variations exist in the number of time points N_i between the subjects. This can be corrected using methods like the one described in [Guillaume 2014]. However, in the study presented here, only 13 subjects out of 136 had more than three time points. The majority had two or three time points which did not unbalance the data too much.

Apart from the bias, one can wonder what would be the best method between LDDMM and the SVF framework. At first sight, LDDMM might appear as a better theoretical model for an elastic mechanical deformation since it is based on the conservation of the Hamiltonian. However, it is not completely clear that the longitudinal evolution of a brain (intra-subject) is an elastic deformation that conserves the energy. Moreover, in practice [Lorenzi 2013b] showed that for the longitudinal registration the differences between the two methods are very subtle and the stationary velocity field framework can be used.

Acknowledgments

This work was partially funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modelling and Analysis of Dynamic Medical Images).
OASIS grant numbers: P50 AG05681, P01 AG03991, R01 AG021910, P20 MH071616, U24 RR021382.

Appendix

Optimisation of the LCC log-Demons energy with a confidence mask

We detail here the optimisation of the LCC log-Demons energy function (defined section 2.2.3.2) with a confidence mask ω and the closed-form solution of the update δv. We consider the first part of the optimisation where E_{Sim} is minimised with respect to v_c, the auxiliary SVF:

$$E_{Sim}(\delta v, I_i, I_j) = -\frac{\omega}{\sigma_i^2}\rho'^2(\delta v, I_i, I_j) + \frac{1}{\sigma_x^2}||\delta v||^2. \tag{2.3}$$

We refer here to the optimisation of the LCC-correspondence of [Lorenzi 2013a], using Λ from equation (2.2) we know that the squared LCC $(\rho^{\delta v})^2$ can be approximated by

$$(\rho^{\delta v})^2 \approx (\rho + \frac{\rho}{2}\Lambda\delta v)^2 = \rho^2(1 + \frac{1}{2}\Lambda\delta v + \frac{1}{4}\delta v^T\Lambda^T\Lambda\delta v).$$

Its gradient is $D((\rho^{\delta v})^2) = \frac{\rho^2}{2}\Lambda$, and its Hessian is $H((\rho^{\delta v})^2) = \frac{\rho^2}{4}\Lambda^T\Lambda$.

Therefore the optimal of the energy (equation (2.3)) is given by :

$$\left(\omega H((\rho^{\delta v})^2) + \frac{\sigma_i^2}{\sigma_x^2} Id\right) \delta v = -\omega D((\rho^{\delta v})^2),$$

and we can deduce that the solution is :

$$\delta v = \begin{cases} -\frac{2\Lambda}{||\Lambda||^2 + \frac{1}{\omega}\frac{4}{\rho^2}\frac{\sigma_i^2}{\sigma_x^2}}, \text{ if } \omega > 0 \\ 0, \text{ if } \omega = 0 \end{cases}$$

List and information about the OASIS Subjects Used

For the reader to be able to replicate the presented results, the list of OASIS subjects we used can be downloaded here: http://www-sop.inria.fr/teams/asclepios/data/Pipeline/lists/OASIS_List_Subjects.csv. We kept the complete set of demented and non-demented subjects and only excluded the subjects who converted - from non-demented to demented - during the study. The images can be found here (we used the mpr-1 acquisitions only): http://www.oasis-brains.org/app/template/Tools.vm;jsessionid=9A56E751939B1DAE41049A46BBAAFB3E#services. Socio-demographic and clinical information on the sample can be found Table 4.1.

Group	Non-Demented	Demented
Age at baseline (years)	75 ± 8	75 ± 7
Female/Male	50/22	28/36
Education (years)	15 ± 3	14 ± 3
Mini Mental State Examination (MMSE)	29 ± 1	25 ± 3

Table 2.2: Socio-demographic and clinical information of the study cohort.

We notice that there is a statistically significant (by Fisher exact test) gender imbalance across the two groups (50/72 vs. 28/64). These gender effects could bias the study and were not accounted for. However, since we benchmark (cf. 2.3.2.2) the three methods - LLDF, Pipeline Head, and Pipeline Skull-strip - using the same sample, this gender imbalance should not be an issue.

Versions of the Software Used and Links to Download Them

In the following list, the reader can find the version of the software used and the links to download them so that he can replicate the processing pipeline (all the mentioned software were installed on Linux):

- FSL version 5.0.2.1: http://fsl.fmrib.ox.ac.uk/fsldownloads/
- ANTS version 1.9.v4: https://github.com/stnava/ANTs/tarball/master
- Robex version 1.0: https://www.nitrc.org/projects/robex

- LCC log-Demons with confidence mask: http://www-sop.inria.fr/teams/asclepios/software/LCClogDemons/LCClogDemonsV1.2.tar.gz
- Parallel transport (Pole ladder): http://www-sop.inria.fr/teams/asclepios/software/LCClogDemons/Ladder.tar.gz
- Stationary Velocity Field Tools: http://www-sop.inria.fr/teams/asclepios/software/LCClogDemons/SVF_Tools.tar.gz
- SPM8: http://www.fil.ion.ucl.ac.uk/spm/software/spm8/

Chapter 3

Validation of Longitudinal Volumetric Measurement Methods

Contents

3.1 Introduction **37**

3.2 Methodology of validation **39**

3.2.1 Choice of software packages to compare: Segmentation and Registration-based software 39

3.2.2 Using both real and simulated datasets 40

3.2.3 Comparing segmentation and registration-based techniques . 41

3.2.4 When is the log-Jacobian integration equivalent to computing the relative volume changes ? 44

3.3 Results **46**

3.3.1 Comparison of segmentation and registration-based methods using segmentation propagation 46

3.3.2 Accuracy of the log-Jacobian with respect to the segmentation propagation, statistical power of the registration-based methods using the log-Jacobian, and influence of the numerical schemes 51

3.4 Conclusion and Discussion **54**

Volumetric change is a biomarker to quantify brain evolution that can be obtained either by segmentation-based techniques or registration-based techniques. We believe it is greatly important to investigate the performance of the different existing methods. However there exist few evaluations or comparisons of the different techniques, especially for small deformations of the brain. We thus develop a framework to evaluate in a consistent way the accuracy, the reproducibility, the detection limit, and the statistical power of volumetric longitudinal methods. We apply it to four popular software packages for brain volumetry using both simulated ground truth and real data. Our study shows that registration based-methods are generally more accurate, reproducible, and have a greater statistical power than the segmentation-based method `FreeSurfer`*. In addition to these results we show that registration based-methods have a smaller detection limit than the segmentation-based method. Although all the methods are sensitive to local linear intensity bias, segmentation ones are generally more sensitive. As for the registration-based methods, we notice that for high volume changes (superior to 10%) they highly under-estimate changes,*

especially `SPM12`*. Moreover, we show that for changes less than 10 % the log-Jacobian integration is equivalent to relative volume changes. Finally, the two different evaluated numerical schemes for the log-Jacobian computation are very similar and can be interchangeably used.*

3.1 Introduction

In longitudinal brain studies, volumetric changes are one of the most used indices to quantify brain structural changes - either for the whole brain or for specific brain structures. In addition to being easily interpretable - due to its physical meaning - volumetry has an important clinical value being a biomarker [Giorgio 2013] that can for example support disease diagnosis (e.g. hippocampal atrophy in patients with Alzheimer's disease).

Given a subject's pair of longitudinal images, current automated approaches quantify brain volume changes, either by segmentation-based techniques (`FreeSurfer`, `FSL-FAST`...) or by using registration-based methods (`ANTs`, `FSL-FNIRT`...). With the latter type of techniques, the volumetry is not derived directly from parcellations but indirectly by the integration of a Jacobian determinant map. This map is computed from the gradient of the deformation field estimated by non-linear registration of the pair of images (Tensor-Based Morphometry).

In comparison to the great number of published studies that use fully-automated volumetric methods, only few articles have tried to evaluate these techniques [Cash 2015]. This might be explained in part by the absence of ground truth but also by the lack of framework to **consistently** compare the different methods. In our work consistently refers to the capacity to evaluate methods in the same exact way as to be fair and objective. Thus, people generally focus on the plausibility of the result whereas the precision of the methods - defined as the closeness of the two results obtained using two different methods - is less evaluated. In the absence of validation, one cannot be certain that the effect detected is not due to a confounding factor introducing variance such as the data acquisition or the data processing (e.g. software used...).

We believe that validation is of great importance to compare objectively the performance of the different methods and that this might be encouraged by the recent development of brain atrophy simulators [Camara 2008, Khanal 2016b]. This simulators are able to provide us with a ground truth against which independent volumetric methodologies can be validated. In [Camara 2008] the authors use simulated ground truth to test the **accuracy** of different methods. In our paper, accuracy is used to define how close the methods are from the ground truth. The authors use a dataset of patients with Alzheimer's disease to evaluate the accuracy of two registration-based methods without taking into account segmentation-based techniques. In [Cash 2015], no ground truth is available so the authors focus on a real dataset of patients with Alzheimer's disease that exhibit large brain deformations. They use the sample size effect as a quantitative measure to compare both

registration and segmentation-based methods. The authors also study the **reproducibility** of both types of methods; which consists in evaluating the variations in measurements made on a dataset with different measurement methods. In their work, the definition of the regions of interest (e.g. hippocampus) is specific to each method which makes it difficult to compare the methods consistently. No distinction is made between the variability of the volumetric method and the variability of the tool used to define the regions of interest. To fine tune the comparison between the methods, it could be useful to limit the study to the variability of the method only. Finally, the authors show that if the interval between the two longitudinal scans is lower than six months, it is not possible to detect a brain evolution. This highlights the existence of a detection limit for the algorithms which prevents them from quantifying the brain change really occurring: this can be detecting a change when no change exists or the opposite.

General guidelines for a thorough validation of the quality of the longitudinal volumetric methods have been proposed by [Fox 2011]. The authors recommend to compare the techniques on the same dataset using simulated images, and note that sample size should not be the only measure. Following their advice, our objective is to "improve and validate atrophy quantification" with a brain volumetry methodological comparison. We find this type of volumetric comparison poorly developed especially in the case of small deformations. Let us consider one of this case for which the brain evolution is very subtle for example during adolescence or if the interval during the two scans is short (e.g one month). In these cases we need to know if the results given by the methods are reliable.

Therefore, we need to answer four complementary questions in the case of small deformations: **i)** Are the results **accurate** i.e. close to the ground truth? **ii)** Are they **reproducible**? **iii)** What is the method **detection limit**? Since the output of the methods are generally used in statistical studies to detect a clinical effect, **iv)** what is the **statistical power** of the method i.e. is it able to detect an existing effect?

We thus detail a comparison framework to consistently evaluate the accuracy of volumetric longitudinal methods. However, no matter the validation some effects cannot be tested. So, it is interesting to check that we obtain the same results with other independent methods. Thus, we apply the validation to four popular software packages for brain volumetry using both simulated ground truth and real data. We also challenge the assumption that the log-Jacobian determinant is not equal to the relative changes. To our knowledge no such comparison of longitudinal volumetric methods was proposed before for small volume changes. We think that such a validation could be useful not only for the software developers but also for the community in order to know the range of use of the method. We focus only on automated methods, which represent the vast majority of the methods in large scale brain studies.

In section 3.2, we present a method to consistently evaluate all the longitudinal volumetric techniques using both simulated and real images. We then present the results of the comparison of the four methods in section 3.3. Finally, in section 3.4,

we discuss about the previous results.

3.2 Methodology of validation

We briefly present the evaluated software packages and the datasets. We then develop a method to consistently compare the results from both segmentation and registration-based techniques. Finally, we focus on the log-Jacobian integration, another technique used for computing the relative volume changes, only available for registration-based methods. We detail a methodology to answer the question: "when is it valid to approximate the log-Jacobian determinant integration by the relative change of volume?"

3.2.1 Choice of software packages to compare: Segmentation and Registration-based software

In addition to the evaluation of our non-linear algorithm (`LCC log-Demons`), we propose to compare it to other registration-based methods. This enables us to check whether other volumetric algorithms provide the same results. This is particularly interesting since we are dealing with small deformations and some of the methods might reach their limit of detection before other methods. Instead of using an unlimited number of methods we choose a concise number of methods guided by our experience and the practices of the clinicians we are working with. We choose to take into account two more popular registration methods: `ANTs`, `SPM12`, and add a wide-spread segmentation based method `FreeSurfer`. We notice that the evaluation framework proposed below is generic enough to be applied to all the other software packages that are not evaluated here. We kept the default or the recommended parameters for each software and no fine tuning was made with respect to the datasets. The detailed parameters used for each tool can be found in Appendix 3.4.

Segmentation-based software: `FreeSurfer`

We selected `FreeSurfer` [Reuter 2012], one of the references in the segmentation-based category. It outputs a full parcelation of the brain. We here use the longitudinal version as well as the default parameters. One should note that contrary to classical segmentation methods that process each time point independently, the longitudinal version of `FreeSurfer` integrates information from both time points in order to be more longitudinally consistent.

Registration-based software: `ANTs`, `SPM12`, and `LCC log-Demons`

`ANTsRegistration`: This pairwise registration algorithm ranked among the best registration software packages in [Klein 2009] experiments based on overlap and similarity criteria. It is based on the Large Deformation Diffeomorphic Metric Mapping and provides the inverse consistent displacement field. In addition to the recommended parameters, we used the brain binary mask option in order to limit the

registration to the brain area. It is important to notice that the similarity criterion used is the cross-correlation [Avants 2011].

`SPM12` longitudinal registration: We consider here the new serial longitudinal registration module [Ashburner 2013] released with `SPM12`. This registration module is based on the Large Deformation Diffeomorphic Metric Mapping: it creates a halfway subject-specific template and two deformation fields that map the template to either the subject baseline or the follow-up image. The similarity criterion is a Sum of Squared Difference (SSD). Here, the whole head is registered since it is not possible to limit the registration to the brain only. In this chapter, for a matter of space, we will refer to the serial longitudinal registration module of `SPM12` using only the name `SPM12`.

`LCC log-Demons` with confidence mask: We evaluate the `LCC log-Demons` with confidence mask [Hadj-Hamou 2016] for which the Large Deformation Diffeomorphic setting is restricted to the one-parameter subgroup of diffeomorphisms. The algorithm outputs the inverse consistent deformation field parameterised by a Stationary Velocity Field. The Local Correlation Criterion (LCC) similarity metric is similar to the one used in `ANTs`. The confidence masks are defined using the brain binary masks from the two time points.

3.2.2 Using both real and simulated datasets

The choice of the dataset is key for our study. The dataset should focus on small deformations with the objective of evaluating different apects: reproducibility, accuracy, statistical power, and detection limit. Using either real data or synthetic one would only enable us to answer a part of the questions. Indeed, the two types of data are complementary: the reproducibility and the statistical power of a method can only be performed on real datasets whereas accuracy and detection limit can only be evaluated when a ground truth exists i.e. for synthetic data. In this study, we consider longitudinal observations of T1-weighted MRI scans for a given subject, at 2 time points t_0 and t_1. The corresponding images will be described as I_0 and I_1 respectively. We present in the following sections the real dataset, IMAGEN, as well as the creation of the synthetic data.

Real data: the IMAGEN dataset

We selected the IMAGEN database since it consists of healthy subjects for which small brain structural changes are occurring. More precisely, we use the French subset of the European longitudinal IMAGEN database [Schumann 2010] that consists of 120 healthy adolescents aged 14 at the beginning of the study t_0, and 16 at t_1. The images are 3 Tesla, T1-weighted MRI scans. The sequence used is the Magnetisation Prepared Rapid Acquisition Gradient Echo (MPRAGE) based on the ADNI protocol [Jack 2008].

The data first went through a manual quality control. We then applied the preprocessing and position correction parts of the Longitudinal Log-Demons framework [Hadj-Hamou 2016] (and chapter 2). All the resulting pairs of subject's images were then processed with each of the four evaluated software packages.

Synthetic Subjects

Among the existing realistic brain MRI simulators that enable the user to create synthetic images with controlled amount of atrophy we use `SimulAtrophy` [Khanal 2016b], since it is the only freely available software. Starting with a user-defined atrophy map (with regions based on the baseline image I_0 segmentation), the algorithm outputs the simulated deformation field $\phi^S_{0\text{-}1}$. Then two simulated follow-up images I'_1 and I''_1 can be created (cf. Figure 3.1).

1. The image I'_1 results from the warping of I_0 with $\phi^S_{0\text{-}1}$. In this case, I'_1 exhibits the same intensities as I_0.

2. In practice, if the acquisition of I_0 and I_1 is not performed on the same scanner, the two images may not exhibit the same intensities and local intensity bias may exist. This is the case with the IMAGEN dataset we use. The simulator enables us to recreate this realistic type of bias using the deformation field $\phi_{0\text{-}1}$ that maps I_0 to I_1 in which the bias is visible. The simulated image I''_1 is then obtained by warping of I_1 with $\phi_{0\text{-}1}$ and $\phi^S_{0\text{-}1}$. For the sake of not privileging our in-house software (`LCC log-Demons`), we used `ANTs` to compute $\phi_{0\text{-}1}$. We could also have used a software not evaluated in the study. However we chose `ANTs` since it ranked among the best software package for registration in [Klein 2009].

Thus two working sets are created, one with no intensity local bias between I_0 and I'_1, and the second closer to reality with an intensity local bias between I_0 and I''_1. We use two subjects randomly chosen from the IMAGEN database, and for each of them we create 14 simulations by prescribing different atrophy maps (cf. Figure 3.2).

3.2.3 Comparing segmentation and registration-based techniques

Due to the different nature of the registration and segmentation-based methods' outputs - respectively segmented binary regions of interest R^j_0 and R^j_1 at t_0 and t_1, and a deformation field $\phi_{0\text{-}1}$ from which a Jacobian map is derived - it is not possible to directly compare the results consistently [Nakamura 2014]. Depending on the type of method used, the relative volume changes are computed in a differently. For segmentation-based methods, the computation is direct as the ratio of the segmented volumes. Concerning for the registration based-methods, the computation is less straightforward. In general, the change of volume is computed using the integration of the Jacobian determinant over the region segmentation. As the

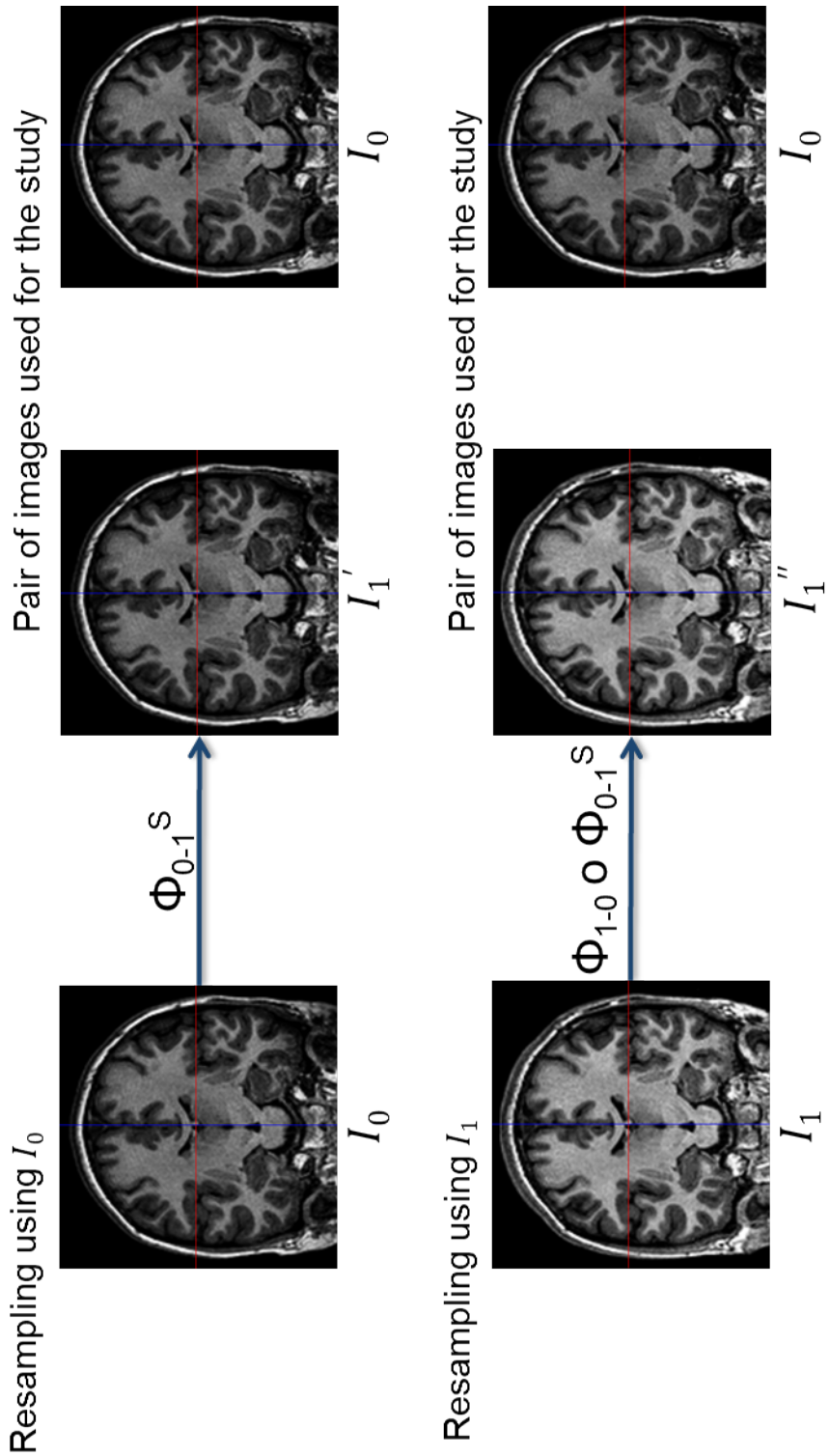

Figure 3.1: **Creation of the two types of synthetic images: I_1' and I_1''.** *Top:* I_1' results from the warping of I_0 with $\phi_{0\text{-}1}^{S}$. The images I_0 and I_1' look very similar. *Bottom:* I_1'' is then obtained by warping of I_1 with $\phi_{0\text{-}1}$ and $\phi_{0\text{-}1}^{S}$. In this last case, we can clearly see an intensity local shift between I_0 and I_1''.

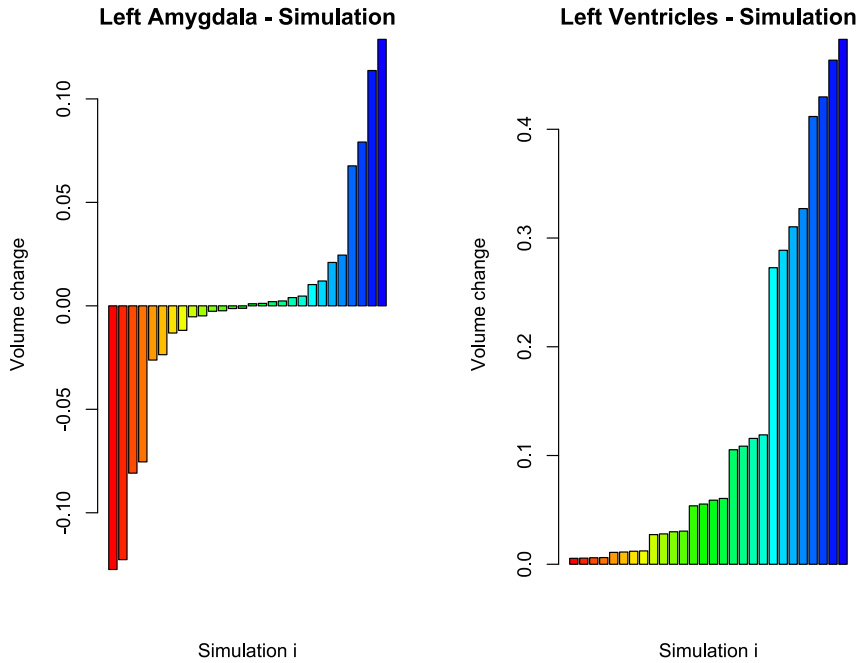

Figure 3.2: **Simulated volume changes for two subjects with 14 different atrophy maps for each one.** The simulator enables us to create a wide range of volume changes as shown for the left amygdala and the left lateral ventricle.

Jacobian determinant is centered on one (no change), there exists another variant symmetric with respect to zero: the log-Jacobian determinant. The integration of the log-Jacobian determinant over the region segmentation (also called flux) computes the relative changes of volume. This is only true for infinitesimal changes though it is used in some studies to quantify large volume changes even if this is not theoretically correct.

Instead of mixing all types of results, we here propose a method to compare the two types of methods in a consistent manner on the volumes using Segmentation Propagation [Calmon 2000]. It consists in deforming[1] the segmented region R_0^j using the computed $\phi_{0\text{-}1}$ to obtain $R_1^j = R_0^j \circ \phi_{01}$. The volume of R_1^j at t_1, $V_1^j = Volume(R_1^j)$ is computed as a weighted sum of the pixels over the deformed ROI. Therefore, for all the methods, we define a consistent index, the relative volume (RV) changes for the studied ROI j:

$$RV_j = \frac{V_1^j - V_0^j}{V_0^j}$$

[1]We use tri-linear interpolation to avoid negative values for the warped segmentation (that can occur with B-Spline interpolation).

$$\text{with } V_0^j\text{: FS vol. and } V_1^j = \begin{cases} Volume(R_1^j), \text{ for FreeSurfer (FS)} \\ Volume(R_0^j \circ \phi_{01}^{reg}), \text{ for registration meth.} \\ Volume(R_0^j \circ \phi_{01}^{simu}), \text{ for gold std.} \end{cases}$$

In this study, we define $j = 20$ regions of interest (ROI) which represent 10 structures per hemisphere: **1)** the cortex, **2)** the white matter, **3)** the lateral ventricles, and 7 subcortical grey matter structures: **4)** the thalamus, **5)** the caudate nucleus, **6)** the putamen, **7)** the globus pallidus, **8)** the hippocampus, **9)** the amygdala, and **10)** the accumbens. The details for the application of the method with the different registration-based algorithms can be found in Appendix 3.4.

3.2.4 When is the log-Jacobian integration equivalent to computing the relative volume changes ?

In Tensor-Based Morphometry, the log-Jacobian map integrated over a region represents the relative change in volume in this region only for infinitesimal changes. However, some studies do use the log-Jacobian integration with no restriction on the size of the changes and the results seem plausible. Therefore, it would be interesting to know if in practice the log-Jacobian integration is equivalent to the relative volume changes and what is the range of validity of this assumption. If the assumption was true, the values obtained from the log-Jacobian determinant integration should directly be comparable to the previous index (obtained by segmentation propagation). The aim here is to quantify the difference between the log-Jacobian integration and the relative volume changes as defined in the previous part. For the sake of comparison, the study uses the same regions of interest (as shown on Figure 3.3): the log-Jacobian determinant is integrated on the ROI R_0^j previously computed with `FreeSurfer`.

Moreover, the practical computation of the log-Jacobian depends on the numerical scheme used. Thus, we would also like to quantify the agreement between two existing numerical schemes that we briefly present here:

Centered Finite Difference (FD): We use `ANTs` implementation `CreateJacobianDeterminantImage` for the displacement fields (the parameters can be found in Appendix) except for `SPM` deformation fields for which we used `SPM deformation toolbox` that is more direct on `SPM` results. We previously verified that the results given by the two implementations of the same method were the same.

Euler Forward (EF): Contrary to the previous scheme, this method is only available for algorithms who parameterise the deformation with a Stationary Velocity Field. Thus, it is here only used with the `LCC log-Demons` results. The implementation is based on [Lorenzi 2013a]. During the analysis of the results, we discovered an error in the version of the log-Jacobian computation tool. It lied in the finite difference implementation that did not take into account the orientation

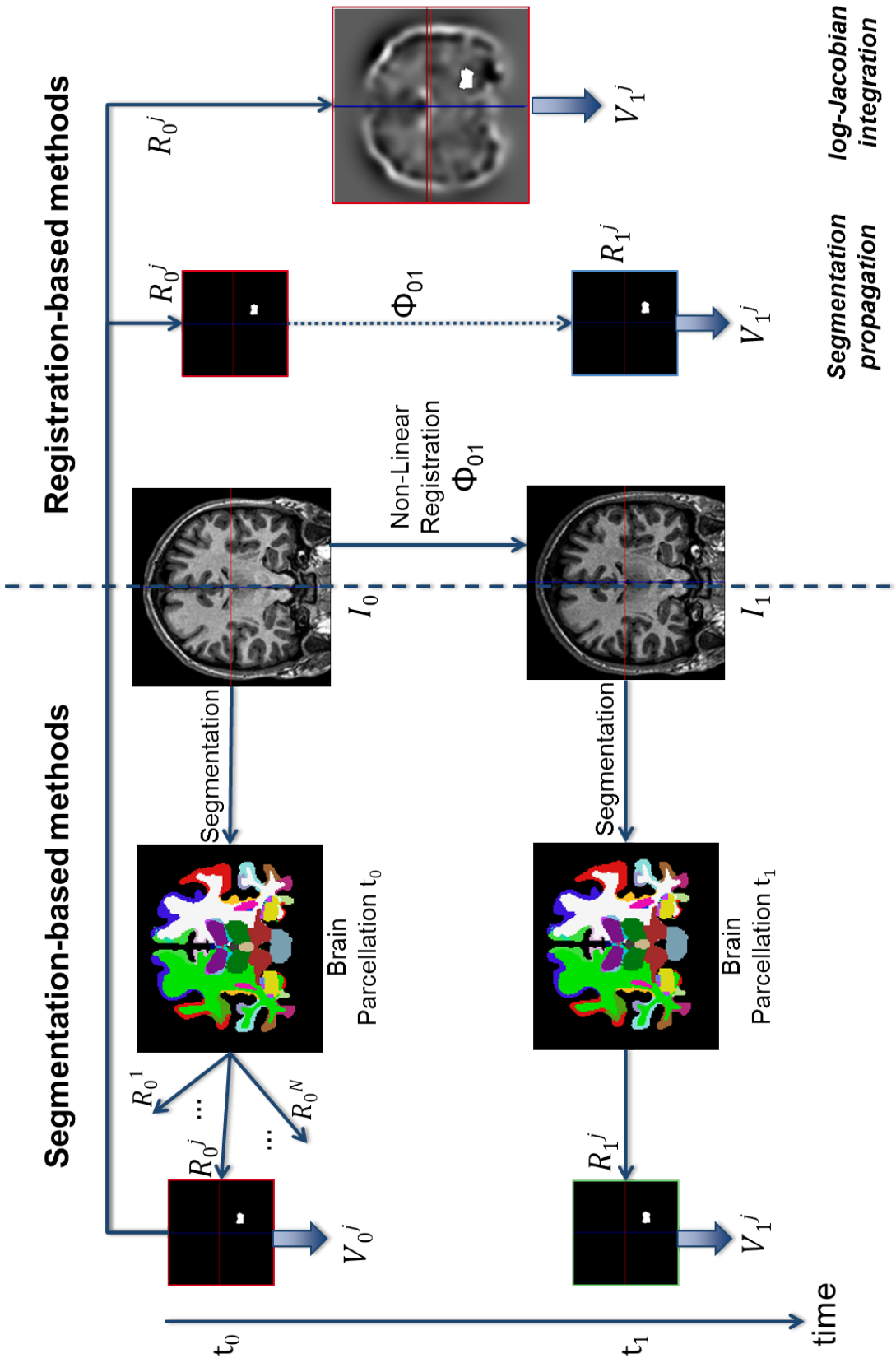

Figure 3.3: **Overview of the consistent comparison method for the different segmentation/registration techniques.** The pipeline can be applied to any registration or segmentation-based method. *Left:* Method for the computation and comparison of volume changes using segmentation-based methods. *Right:* Overview of the comparison method for registration-based methods (segmentation propagation and log-Jacobian integration).

of the image. This caused errors for images that do not have a direct orientation (orientation matrix different from the identity). We thus corrected this problem and released a new version of the tool[2] which we use for the study.

3.3 Results

We first evaluate consistently segmentation and registration-based methods on different aspects. We use both synthetic and real images to evaluate the accuracy, the detection limit, the reproducibilty and the statistical power of four popular volumetric methods. We then focus on the registration-based methods to evaluate the accuracy of the log-Jacobian integration with respect to the simulated volume changes, and to analyse the influence of the numerical scheme used to compute the log-Jacobian.

3.3.1 Comparison of segmentation and registration-based methods using segmentation propagation

Using synthetic data we evaluate the accuracy of the volumetric methods. This corresponds to quantifying the agreement between the evaluated methods and the simulated ground truth. We use Bland-Altman plots [Bland 1999] instead of correlation studies since the latter measure the strength of a relation between two variables and not the agreement between them. A Bland-Altman plot (cf. Figure 3.4) consists in a scatter plot XY. The Y-axis represents the difference between the volume measured by a method and the simulated ground truth volume (called thereafter the two measurements). This difference is plotted against the two measurements (X-axis). If the measurements obtained by a method were to give exactly the same results as the simulated ground truth then all the differences should be equal to zero. In practice this is not possible since variations exists (e.g. in the computation). Therefore, the agreement between the methods is assessed by studying the standard deviation (also known as critical difference and represented by the turquoise dashed lines) of the differences between two measurements. An accurate method will have a small critical difference.
We also use Bland-Altman plots to evaluate the detection limit of a method by studying the mean of the differences (represented by a dark blue dashed line) to see if any bias exists (if so the mean will be different from zero).

Finally, when no ground truth exists (real dataset), we do not use Bland-Altman plots, but study study the reproducibility of the methods using box plots. Moreover the statistical power is assessed using statistical tests.

Although the two different types of methods (segmentation and registration-based) give results that are in relative agreement with each other, registration-based methods perform better that segmentation-based methods on different aspects: ac-

[2]The new version is available at http://www-sop.inria.fr/teams/asclepios/software/

curacy, detection limit, reproducibility, and statistical power. The results are detailed below.

Accuracy of segmentation and registration-based methods

We first consider the regions with non-zero volume changes to study the accuracy of the methods. In the simplest case where the simulated image I_1' is resampled from I_0, the segmentation-based method - `FreeSurfer` - is less accurate than the registration-based method (2.5 times less accurate) for the different brain structures. Figure 3.4 shows the results for the amygdala region. The results in this region summarise the general behavior of most of the structures where changes occur.
In the more realistic cases, the simulated image includes an intensity local bias. In this case, both types of methods are less accurate than in the previous case. However, the segmentation-based method is even less accurate than the registration-based methods (3.5 times less accurate).

We notice that `FreeSurfer` behaves differently in the ventricles (cf. Figure 3.5). In the case of same intensity, `FreeSurfer` is the most accurate method. However we notice an important shift (5%) in the mean of the error when local intensity bias is introduced. This shows that the segmentation-based method is more sensitive to the intensity change than registration based methods (for the registration based method the non-zero mean is due to the error made for the high changes).

Concerning the registration-based methods it is harder to rank them. However, using the results in the ventricular region, we notice that for important changes (superior to 10%) the registration methods tend to under-estimate the changes. In this case, the `LCC log-Demons` results are closer to the simulated truth than the other algorithms. This is particularly true with respect to `SPM` that can only capture around 60% of the changes (for simulated changes of 10% and more).

Detection limit of segmentation and registration-based methods

The simulated truth also enables us to quantify the limit of detection of the different methods using regions where no changes (or very small changes) were simulated. This is the case for the cortex region where the changes vary from 0% to -0.15%. In this region, if the methods had no detection limit they should be able to detect zero change as well as the most subtle changes. On Figure 3.6, we see that when there is no intensity change, the methods exhibit no systematic bias. However, we notice that the segmentation-based method's variability on the error remains very high.
In the more realistic case where there exists a change in intensity between the pair of images, we see a systematic bias for all the methods. We notice that this bias is on average nearly four times larger for segmentation-based method than for registration ones. In conlusion, `ANTs` ranks first with the lowest detection limit (less than 1%) followed by `SPM` (around 1.5%), the `LCC log-Demons` (around 2%) and lastly `FreeSurfer` (more than 6%).

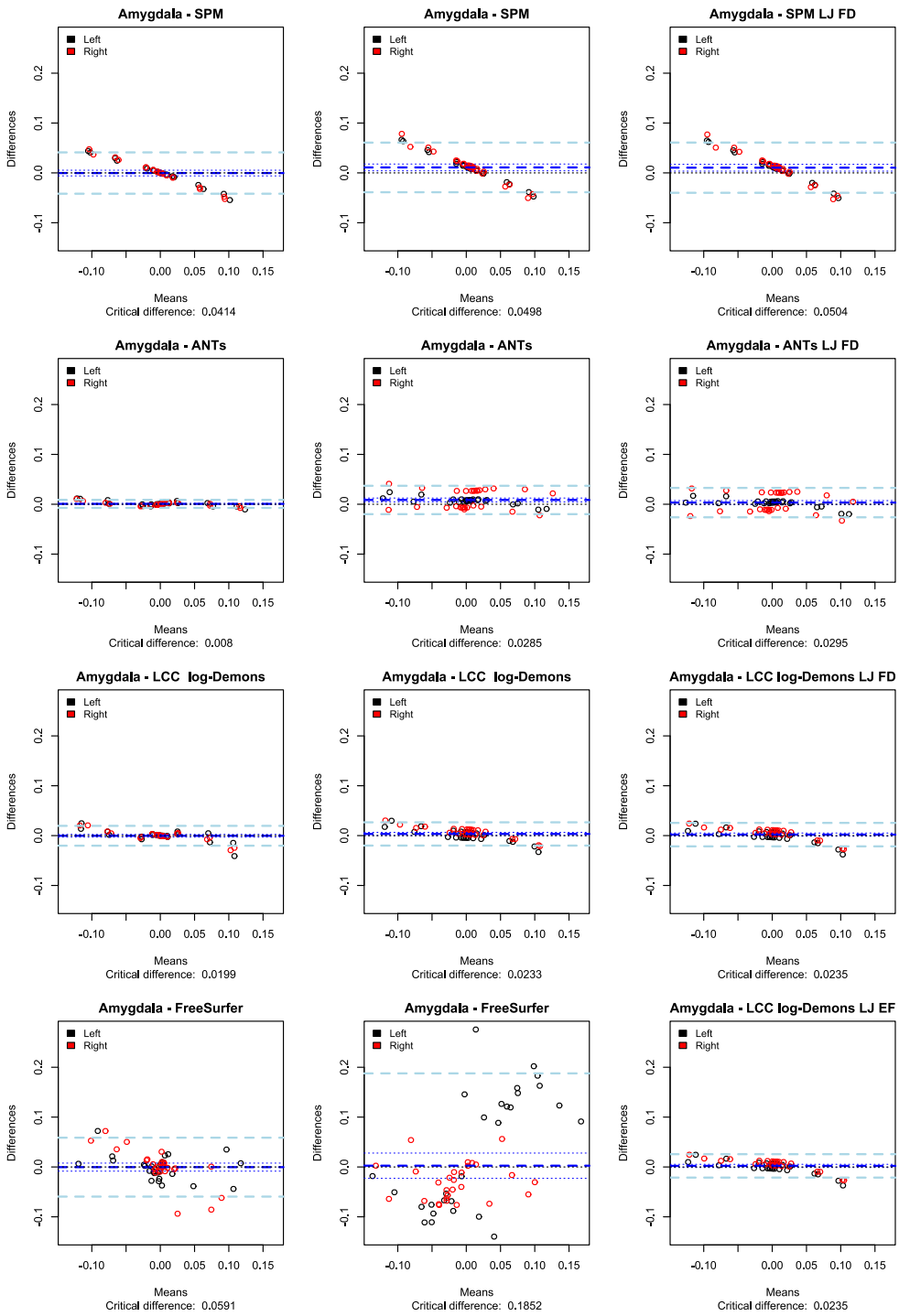

Figure 3.4: **Bland-Altman plot for the amygdala region for the four studied methods.** *Left column:* No intensity change between the 2 time points. *Middle column:* Intensity change between the baseline and follow-up images. In both cases registration-based methods provide results that are more accurate than the segmentation-based method. *Right column:* Volume changes (with intensity change between the 2 time points) obtained by log-Jacobian integration over the ROI (available for registration-based methods only). For the range of simulated changes, we see that using the log-Jacobian is equivalent to computing the volume changes.

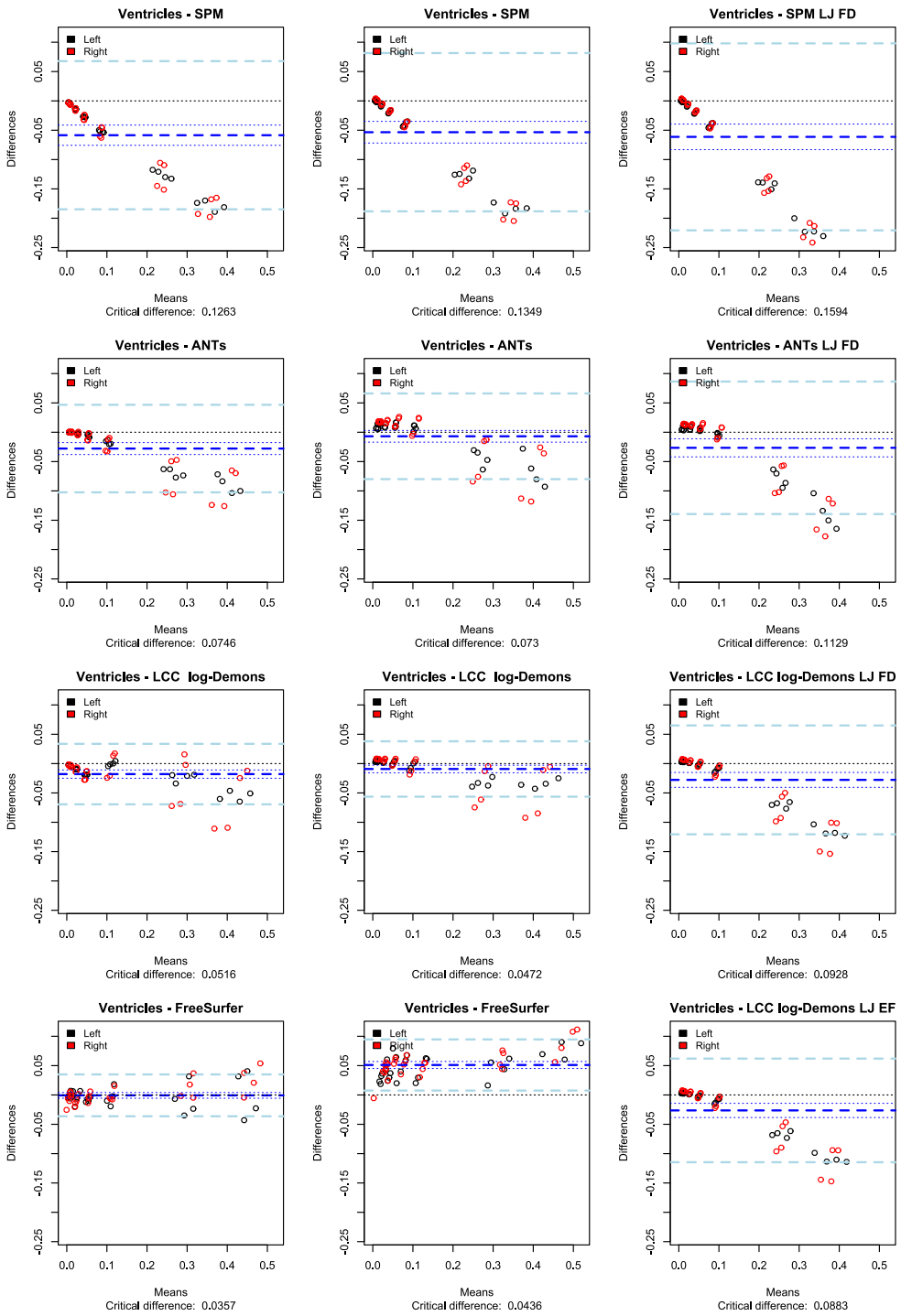

Figure 3.5: **Bland-Altman plot for the ventricles region for the four studied methods.** *Left column:* No intensity change between the 2 time points. *Middle column:* Intensity change between the baseline and follow-up images. The ventricles are the only region where the segmentation-based method is more accurate than registration-based methods. However, the segmentation-based method is the most sensitive method to intensity change. *Right column:* Volume changes (with intensity change between the 2 time points) obtained by log-Jacobian integration over the ROI (available for registration-based methods only). We see that for changes less than 10%, it is possible to approximate the volume changes by the log-Jacobian.

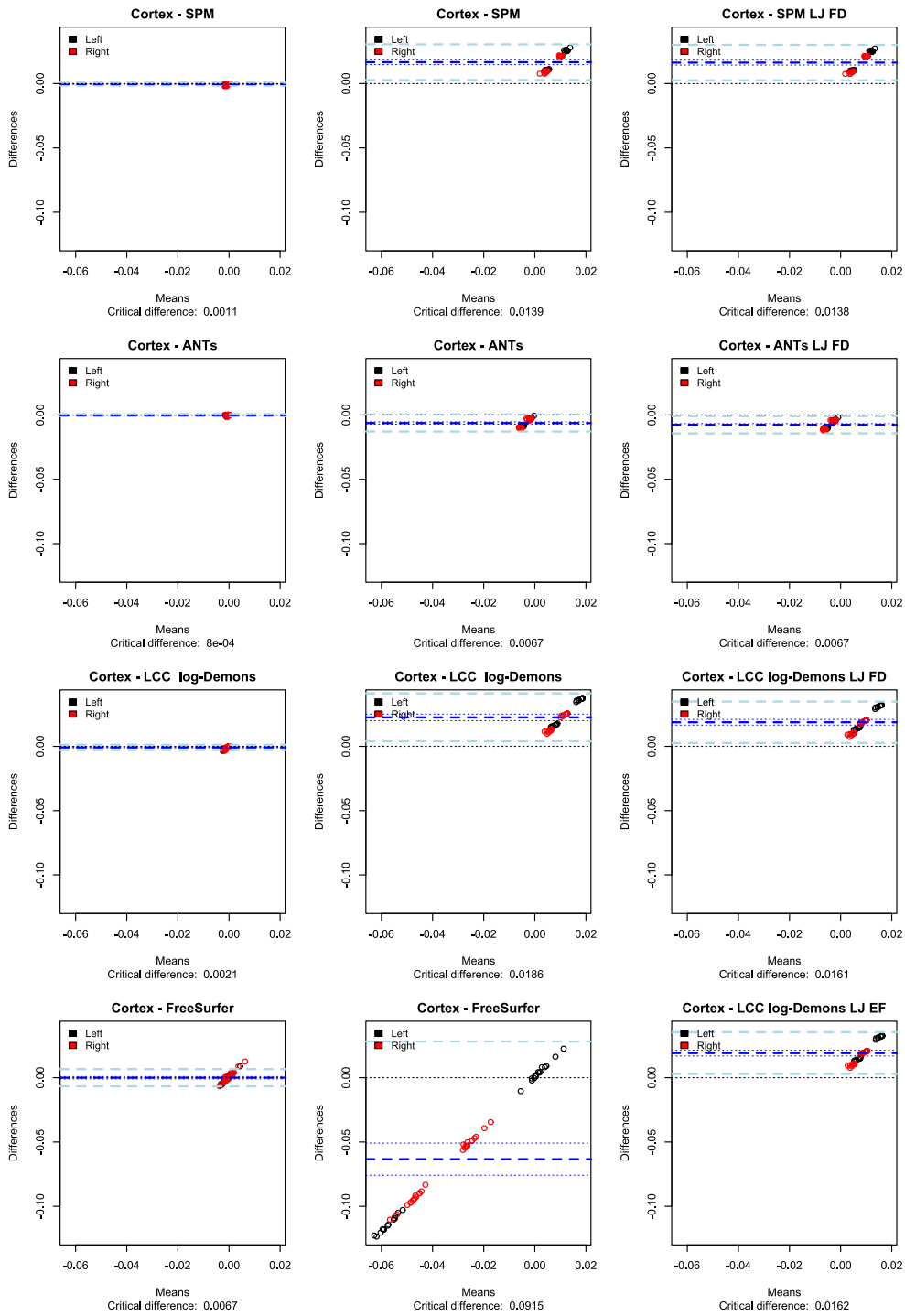

Figure 3.6: **Bland-Altman plot for the cortex region for the four studied methods.** *Left column:* No intensity change between the 2 time points. There exists no bias. *Middle column:* Intensity changes between baseline and follow-up images. A bias (mean of the error not equal to zero) is present for all methods, more important for the segmentation-based than for registration-based methods. This shows the existence of a detection limit for the different algorithms. The algorithms are not able to detect that no volume change (or very subtle ones) occurred. *Right column:* Volume changes (with intensity change between the 2 time points) obtained by log-Jacobian integration over the ROI (available for registration-based methods only). For the range of simulated changes, using the log-Jacobian is equivalent to computing the volume changes.

Reproducibility and statistical power of segmentation and registration-based methods

Let us now focus on the real dataset case where an intensity change occurs between the two time points.The volumetry results for the four different methods are plotted on Figure 3.7. We can see that the segmentation-based method has a higher variability than the registration ones (on average more that 2 times). Moreover, we notice a very good agreement of the registration-based methods between themselves and less agreement between segmentation and registration-based methods.

We also assess the statistical power of the four different methods using the IMAGEN dataset. During adolescence females and males have different longitudinal trajectories [Giedd 2006], a statistical test should therefore detect such differences between the two groups. For each of the four methods, we perform two-sample t-tests on the different regions for the volume changes between females and males. The null hypothesis being that there is no difference in the mean of the two groups in the considered region for a given method.

The resulting p-values for the left hemisphere regions are summarised in Table 3.1: the method with the lowest p-value being the methods with the greatest statistical power. We apply a Bonferroni correction for multiple comparisons (the traditional thresholded p-value of 0.05 corresponds to a corrected thresholded p-values of 0.005 for the 10 regions). The segmentation-based method `FreeSurfer` has a far lower statistical power than the registration-based methods. In particular it fails to detect statistical differences in 7 regions (out of 10) whereas the registration-based methods do not. Concerning the latest, we notice that `SPM12` and the `LCC log-Demons` are the methods with the greatest statistical power, with an statistically significant effect detected in 9 regions out of 10 (even if the `LCC log-Demons` has a slightly lower statistical power than `SPM12`). `ANTs` is the registration-based methods with the lowest statistical power: a statistically significant effect is detected in 8 regions out of 10. Similar results are found in the right hemisphere regions.

3.3.2 Accuracy of the log-Jacobian with respect to the segmentation propagation, statistical power of the registration-based methods using the log-Jacobian, and influence of the numerical schemes

We now only consider the registration-based methods and focus on the log-Jacobian integration. We first analyse the validity of approximating the relative volume changes by the log-Jacobian integration. We then compare the influence of different numerical schemes for the log-Jacobian integration.

Accuracy and statistical power of the log-Jacobian

In this section, for the sake of comparison, the log-Jacobian integration for all the methods is performed using a centered finite difference scheme. For medium and

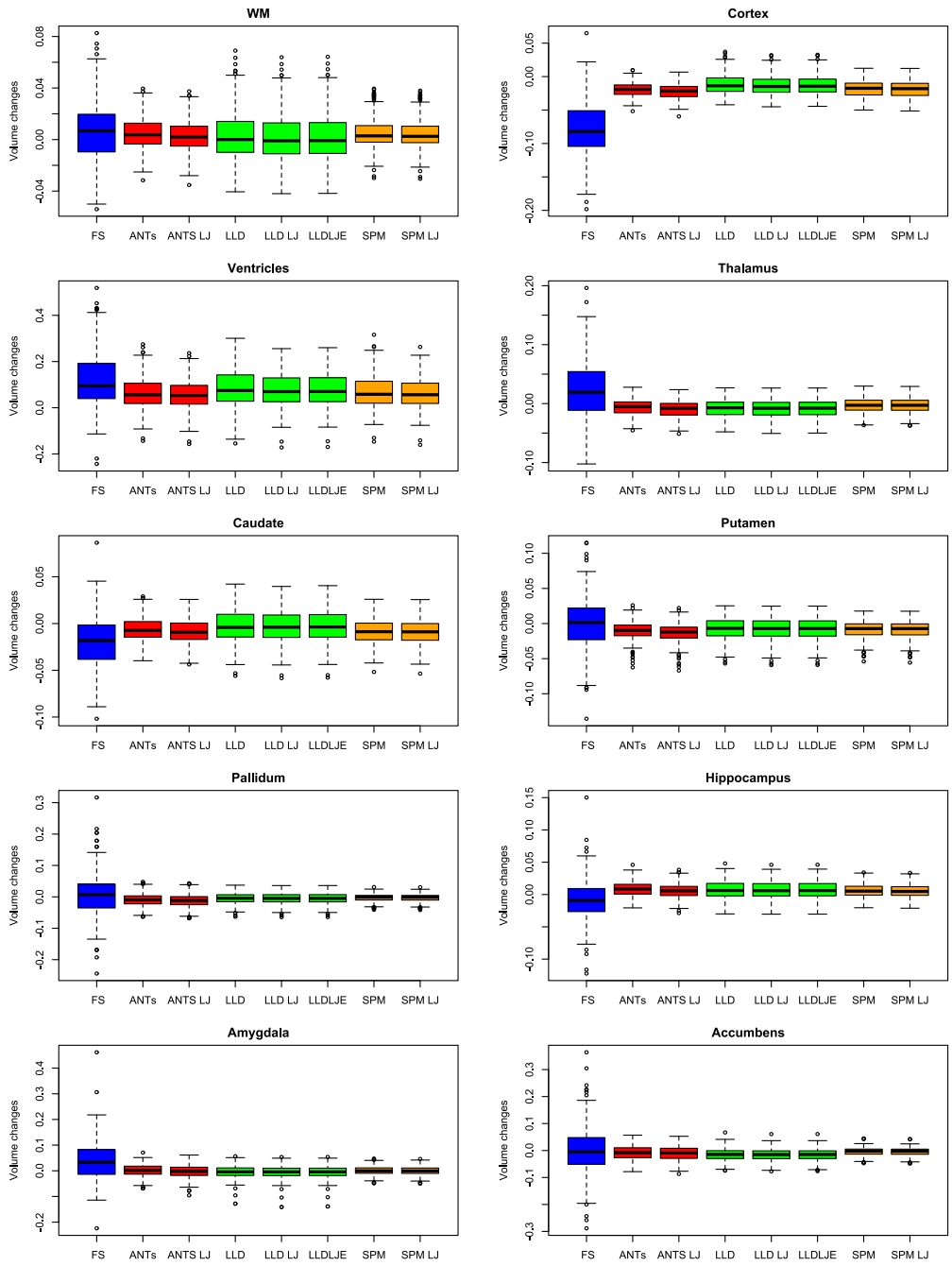

Figure 3.7: **Volumetry results (obtained by segmentation propagation) for the 120 IMAGEN subjects on 10 ROIs with the four different algorithms.** We see that the segmentation-based method results are more variable than for the registration-based ones. (Acronyms used: FS=FreeSurfer, LLD=LCC log-Demons, LJ=log-Jacobian, WM=White matter)

ROI	FreeSurfer	SPM12	ANTs	LCC log-Demons
WM	**$8.6\ 10^{-09}$**	**$3.5\ 10^{-13}$**	**$2.2\ 10^{-10}$**	**$1.4\ 10^{-16}$**
Cortex	**$1.1\ 10^{-3}$**	**$1.1\ 10^{-16}$**	**$1.3\ 10^{-06}$**	**$1.8\ 10^{-18}$**
Ventricles	$9.0\ 10^{-1}$	$6.3\ 10^{-1}$	$2.0\ 10^{-1}$	$6.7\ 10^{-1}$
Thalamus	$3.3\ 10^{-2}$	**$4.8\ 10^{-13}$**	**$5.3\ 10^{-6}$**	**$1.8\ 10^{-6}$**
Caudate	**$1.3\ 10^{-3}$**	**$2.7\ 10^{-11}$**	**$1.9\ 10^{-4}$**	**$5.3\ 10^{-4}$**
Putamen	$7.6\ 10^{-1}$	**$9.3\ 10^{-9}$**	**$1.0\ 10^{-4}$**	**$1.6\ 10^{-10}$**
Pallidum	3.410^{-1}	**$1.0\ 10^{-7}$**	**$3.4\ 10^{-5}$**	**$8.6\ 10^{-9}$**
Hippo.	$5.6\ 10^{-2}$	**$2.5\ 10^{-10}$**	**$1.5\ 10^{-5}$**	**$3.1\ 10^{-9}$**
Amygdala	$6.3\ 10^{-1}$	**$3.2\ 10^{-12}$**	$2.4\ 10^{-2}$	**$1.6\ 10^{-8}$**
Accumbens	$7.3\ 10^{-1}$	**$1.9\ 10^{-9}$**	**$3.2\ 10^{-3}$**	**$1.1\ 10^{-6}$**

Table 3.1: **p-values for the two-sample t-tests on regional volume changes (left hemisphere) between females and males.** p-values inferior to the threshold p-value of 0.005 (correction for multiple comparisons) are in bold. We notice that the segmentation-based method `FreeSurfer` has a far lower statistical power than registration-based methods.

large volume changes (from 10% and more - cf. Figure 3.5), the log-Jacobian underestimates more the changes than the segmentation propagation (the error generally increases as the changes increase). For example in the ventricles, the log-Jacobian integration adds an average 5% bias which decreases the accuracy of the log-Jacobian. If the changes are small (cf. Figure 3.4), there is no difference between the relative volume changes and the log-jacobian values.

Therefore, the log-Jacobian integration is equivalent to computing relative volume changes for volume changes inferior to 10%. In practice the only region that generally exhibits changes superior to 10% is the lateral ventricles. For the regions with large deformations, a corrected log-Jacobian should be computed in order to take into account the underestimation (in addition to the underestimation linked to the algorithm).

We now study the statistical power of the methods using the log-Jacobian integration. For each method, we perform a two-sample t-test on the real dataset between females and males. The null hypothesis is that there exists no difference between the mean of the two groups. In table 3.2, we see on the first four columns that the ranking from section 3.3.1 does not change. SPM12 and the LCC log-Demons have the greatest statistical power (a statistically significant difference between the two groups is detected in 8 regions out of 10). FreeSurfer has the lowest statistical power with an significant effect detected in only 3 regions out of 10.

Influence of the numerical scheme for the log-Jacobian integration

We now compare the finite difference (FD) scheme with respect to the Euler forward (EF) scheme. A consistent comparison of the two schemes is possible using the `LCC`

ROI	SPM12	ANTs	LLD FD	LLD EF
WM	**$2.6\ 10^{-13}$**	**$2.5\ 10^{-10}$**	**$5.6\ 10^{-17}$**	**$5.8\ 10^{-17}$**
Cortex	**$1.3\ 10^{-16}$**	**$1.4\ 10^{-6}$**	**$1.3\ 10^{-18}$**	**$1.2\ 10^{-18}$**
Ventricles	$6.0\ 10^{-1}$	$1.9\ 10^{-1}$	$6.2\ 10^{-1}$	$6.3\ 10^{-1}$
Thalamus	**$4.5\ 10^{-13}$**	**$5.2\ 10^{-6}$**	**$2.1\ 10^{-6}$**	**$1.9\ 10^{-6}$**
Caudate	**$3.3\ 10^{-11}$**	**$2.4\ 10^{-4}$**	**$3.3\ 10^{-4}$**	**$4.0\ 10^{-4}$**
Putamen	**$9.5\ 10^{-9}$**	**$9.3\ 10^{-5}$**	**$1.6\ 10^{-10}$**	**$1.6\ 10^{-10}$**
Pallidum	**$8.8\ 10^{-8}$**	**$2.1\ 10^{-5}$**	**$8.3\ 10^{-9}$**	**$8.1\ 10^{-9}$**
Hippo.	**$2.5\ 10^{-10}$**	**$3.2\ 10^{-5}$**	**$3.1\ 10^{-9}$**	**$2.9\ 10^{-9}$**
Amygdala	**$2.6\ 10^{-12}$**	$2.9\ 10^{-2}$	**$2.0\ 10^{-8}$**	**$1.8\ 10^{-8}$**
Accumbens	**$1.5\ 10^{-9}$**	$5.3\ 10^{-3}$	**$1.3\ 10^{-6}$**	**$1.3\ 10^{-6}$**

Table 3.2: **p-values for the two-sample t-test between regional log-Jacobian integrations (left hemisphere) of girls versus boys.** p-values inferior to 0.005 (corrected threshold for multiple comparisons) are in bold. Using the log-Jacobian does not change the statistical power of the methods. Moreover, no difference in the statistical power can be found between the two different numerical schemes for the computation of the log-Jacobian.

`log-Demons`. This is the only volumetric method that enables the use of both schemes. Using the simulated cases, on figure 3.4, no difference can be seen between the results of the two different schemes. They both exhibit the same accuracy. The same trend is visible for the real dataset on Figure 3.7 where no difference can be seen on the reproducibility. Finally we also compare the statistical power of the two schemes. We perform two-sample t-tests for each numerical scheme between females and males. The null hypothesis being that there is no difference in the mean between the two groups in a region for a given method. The results are visible in the two last columns on the right in Table 3.2. We see that the two numerical schemes have the same statistical power. Therefore we can conclude that using one or the other numerical scheme does not modify the results and that the schemes can be interchangeably used.

3.4 Conclusion and Discussion

In conclusion, we developped a framework to consistently compare both segmentation and registration-based methods and applied it to 4 software packages. The results of our study show that registration based-methods are generally more accurate, reproducible and have a higher statistical power than the segmentation-based method `FreeSurfer`. Moreover, the segmentation-based method has a higher limit of detection. Although all the methods are sensitive to local linear intensity bias, the segmentation one is generally more sensitive. We notice that the registration-based methods highly under-estimate changes for high changes (superior to 10%)

especially SPM12. Moreover, we showed that for changes inferior to 10 % the log-Jacobian integration is equivalent to the relative volume changes. Finally, the two different evaluated numerical schemes (finite differences and Euler Forward) for the computation of the log-Jacobian are very similar and can be interchangeably used.

Therefore the results from this study tend to privilege registration-based methods over segmentation-based methods. However one should keep in mind that the registration-based methods are not perfect and all possess different limitations. Among the different evaluated registration-based methods, it is difficult to privilege one of them. They all perform relatively well on the four aspects and each of the methods outperforms the others on one aspect: for large volume changes, the LCC log-Demons should be privileged whereas for small changes ANTs is the most accurate method. Concerning the discrimination between two groups, SPM or the LCC log-Demons should be preferred. For datasets where there exists a local intensity bias between the baseline image and the follow-up, SPM is more sensitive to change in intensity than ANTs or the LCC log-Demons. This can be explained by the SSD metric used by SPM whereas the other registration methods use a local correlation criteria proven to be more robust.

The aim of this study was to propose methods, criteria, and guidelines for users and software designers. The study is of course not exhaustive (number of evaluated software, number of simulations) but we hope it will encourage people to evaluate and compare their software package. In order to do so we make the simulated images available as well as the results of our study. We identify two immediate next steps. The first one concerns the correction of the log-Jacobian determinant when changes are superior to 10%. This would enable people to use a corrected log-Jacobian determinant for the volumetric studies. In practice a first correction would consist in quantifying the log-Jacobian bias on a "simple and controlled" example (e.g. rectangle for which the amount of change is known) and then correct the log-Jacobian accordingly. The study from section 3.3.2 could be then done again to check that the correction enhances the results.
The second concerns the Jacobian determinant that we did not studied here. It would be interesting to quantify in practice the agreement between the Jacobian and the volume changes computed by segmentation propagation.
The other next steps should be to launch more simulations on more subjects. As for the detection limit, further simulations would be needed to exactly quantify it for each method as an important part of the changes in the brain might in the range of the detection limit. The use of scan-rescan - that were not available for our database - should be used if available.

An important limitation in our methodology concerns the use of FreeSurfer as a reference for the ROI definition although we know it is not as accurate. We limit the influence of this effect by remaining consistent in the study and using the same region definition for all the methods. However, for a future work it would be necessary to fine tune the results using structure-dedicated segmentation tools such as the hippocampus automatic segmentation tool from [Chupin 2009].

Acknowledgments

This work was partially funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modelling and Analysis of Dynamic Medical Images).

Appendix

Parameters for the packages

We here summarise the parameters used for the different methods:

- `LCC log-Demons` with confidence masks: The parameters are the one used in chapter 2: -r 2 -R 1 -C 3 -a 30x20x10 -x 0 -b 2.0 -S 0.15 -u 3.0 -V.
- `SPM 12`: we use the default parameters (Times=[0 1] for simulations and the Times=[t_0 t_1] with t_0, t_1 the real age at the time of the scans for the IMAGEN dataset) and warping regularisation: [0 0 100 25 100]
- `ANTsRegistration` with 2 binary masks: -d 3 -m CC[I_1, I_0, 1 , 4] -t SyN[0.25, 3, 0] -c [30x20x10] -s 0x0x0vox -f 4x2x1 -u 1 -z 0 -x [Binary mask of I_1, Binary mask of I_0].
- `FreeSurfer`: We use the default parameters and the segmentation is done in 3 steps (the first and third steps are performed for each time point):
  - recon-all -i -all,
  - recon-all -base -tp -tp -all,
  - recon-all -long -all .
- log-Jacobian determinant: `CreateJacobianDeterminantImage` with options 1 0 and `SVFLogJacobian` tool with -z 1.

SPM12 deformation composition: The method for resampling the volumes (segmentation propagation) described in section 3.2.3 can be applied as is for ANTs and LCC log-Demons but needs to be adapted for SPM12 since the computed deformations are halfway from the subject-specific template to an image. Starting with the two deformation fields expressed in the halfway template to I_0 and I_1 we use SPM12 deformation toolbox to compose the inverse of ϕ_{h_0} with ϕ_{h_1} in order to get the full deformation from I_0 to I_1. The process is then the one described in section 3.2.3 to warp R_0^j with the computed deformation field. This warping is done using SPM deformation toolbox.

Results for the other regions

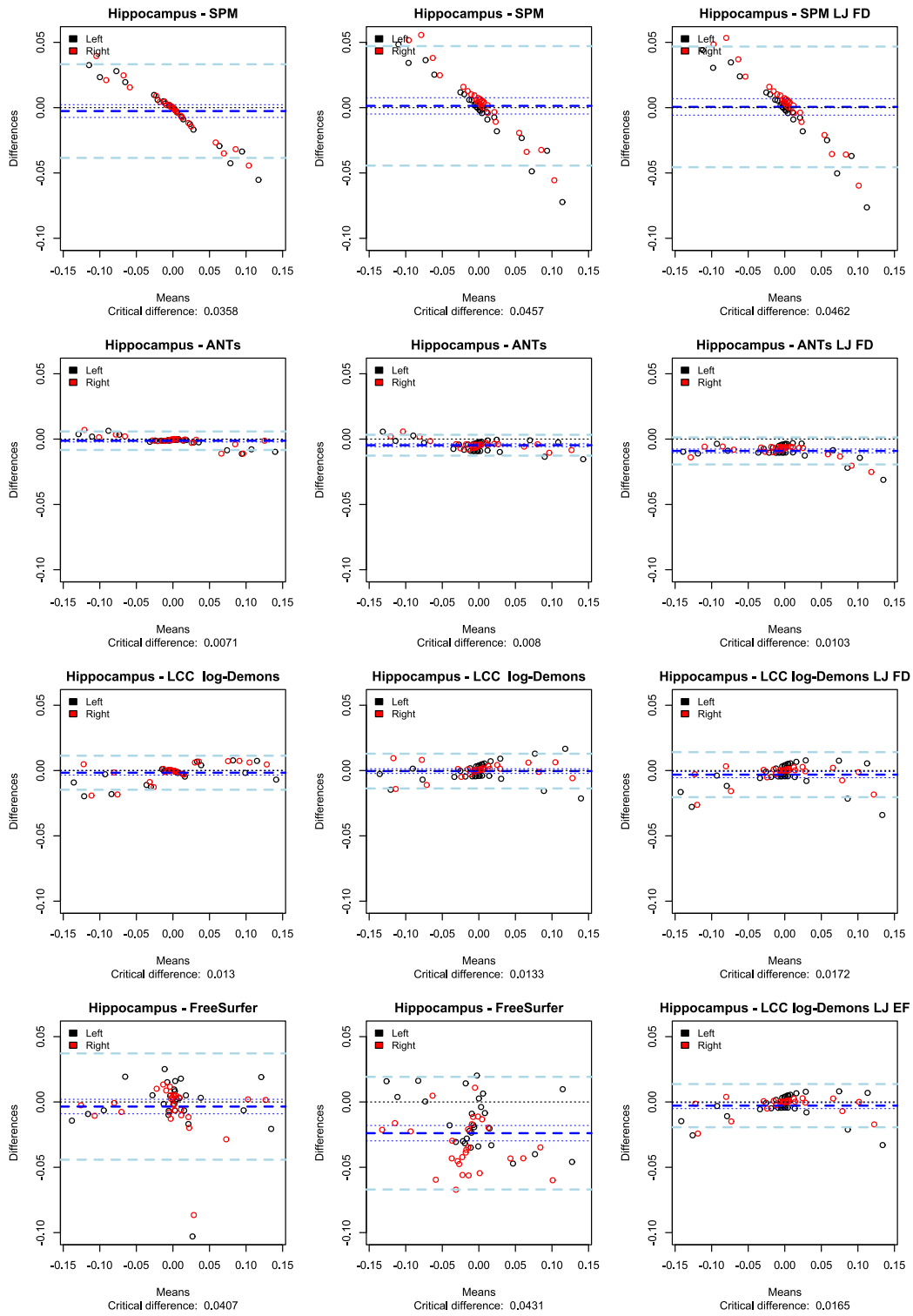

Figure 3.8: Bland-Altman plot for the hippocampus egion for the four studied methods.

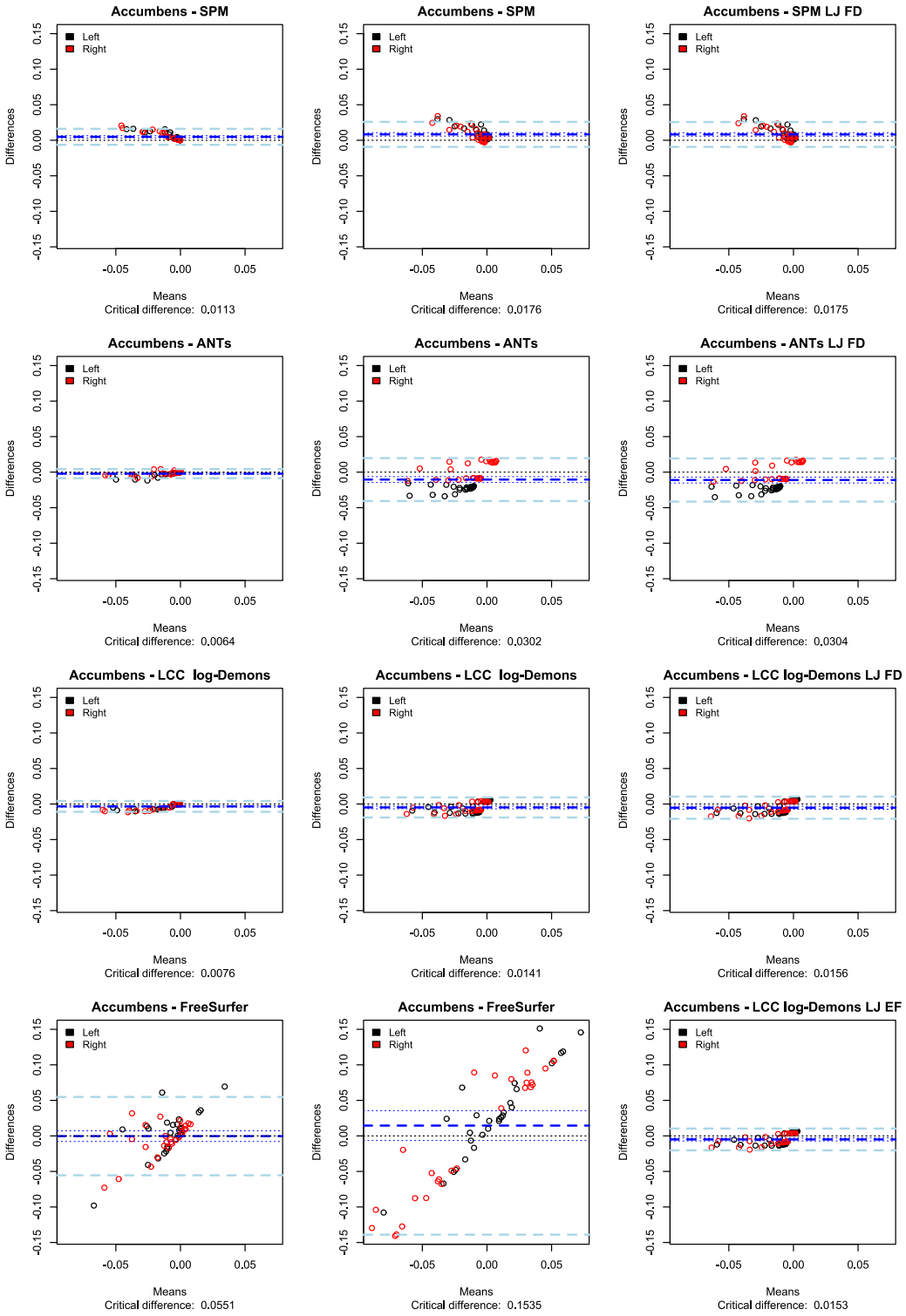

Figure 3.9: Bland-Altman plot for the accumbens region for the four studied methods.

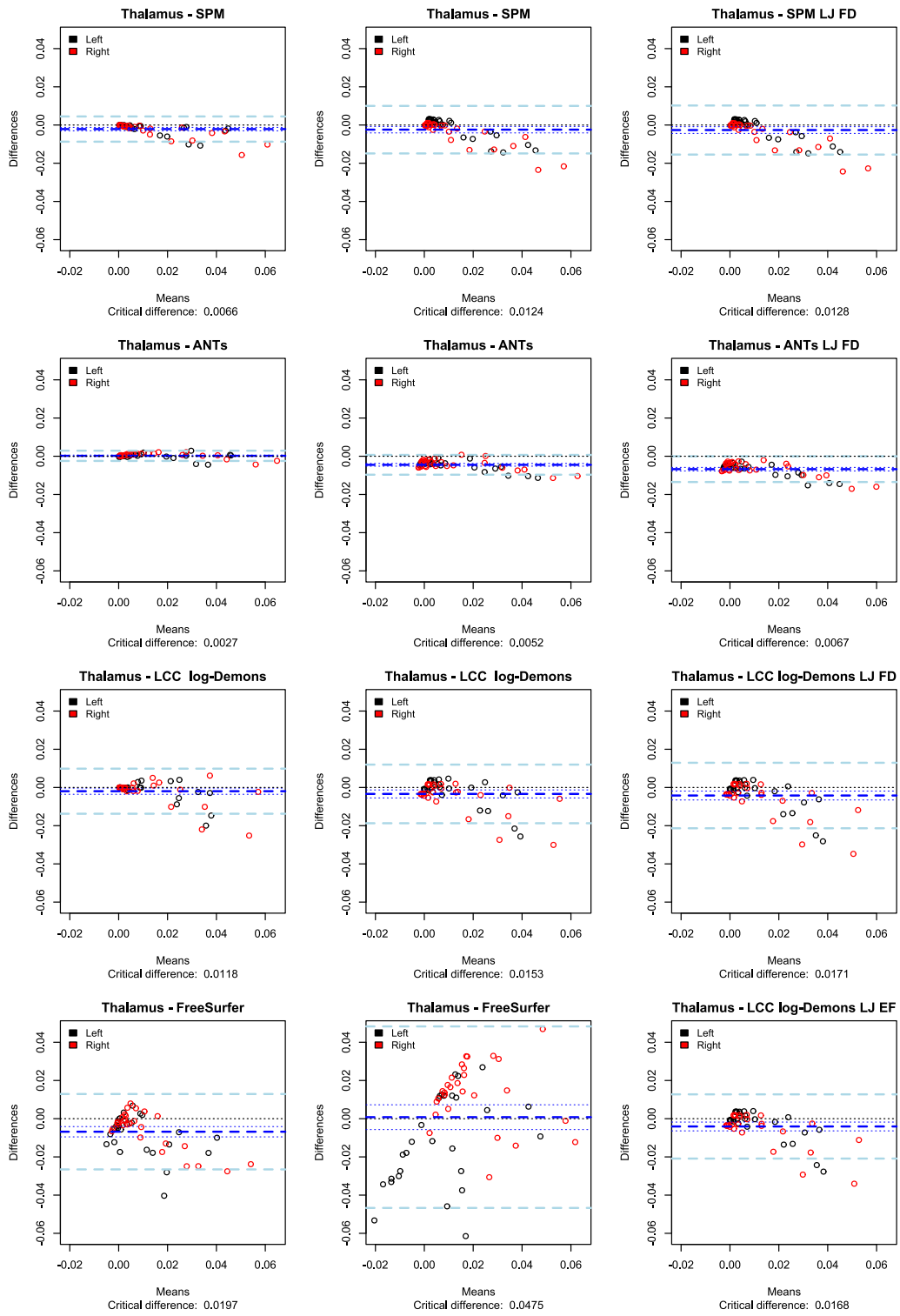

Figure 3.10: Bland-Altman plot for the thalamus region for the four studied methods.

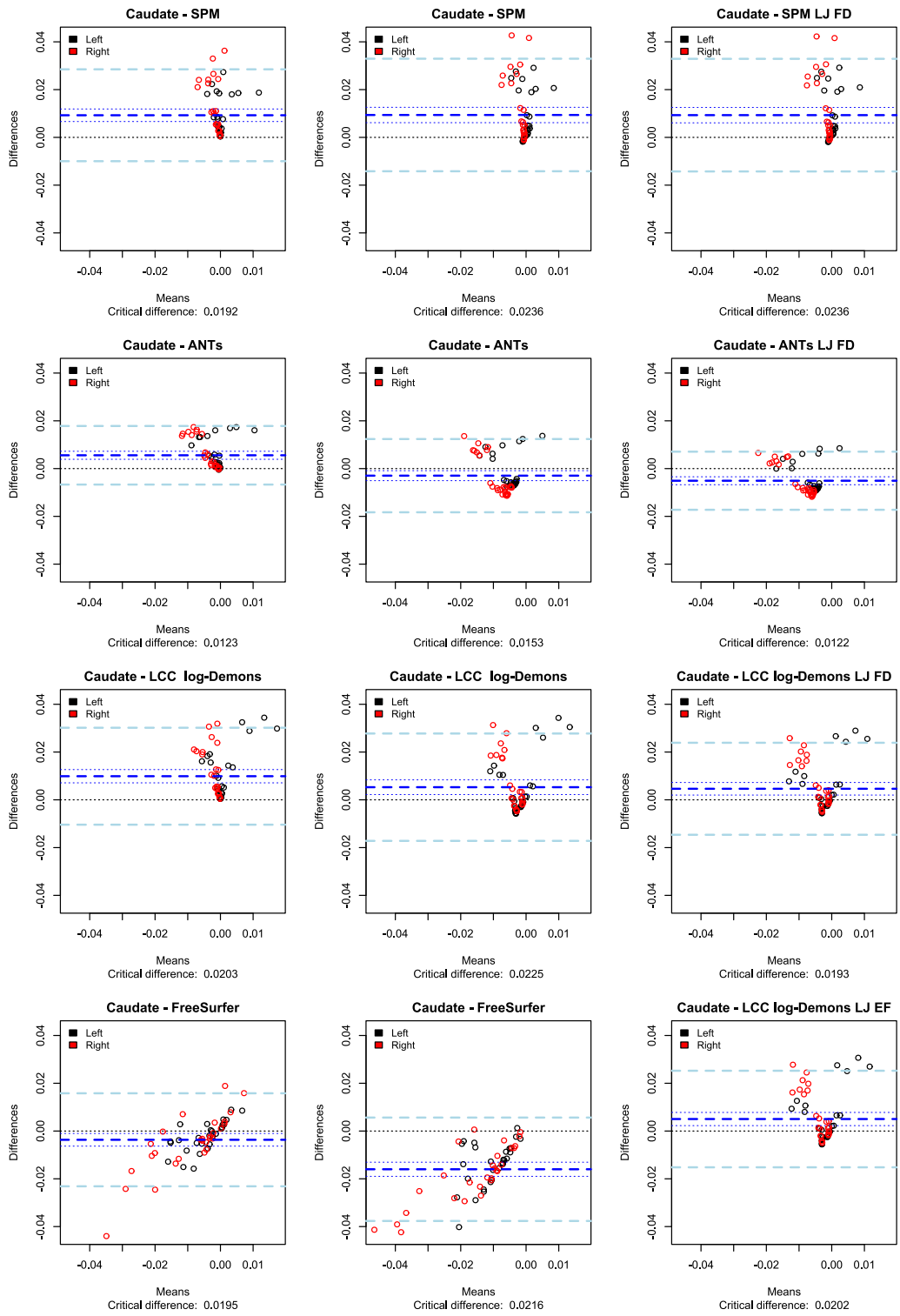

Figure 3.11: Bland-Altman plot for the caudate region for the four studied methods.

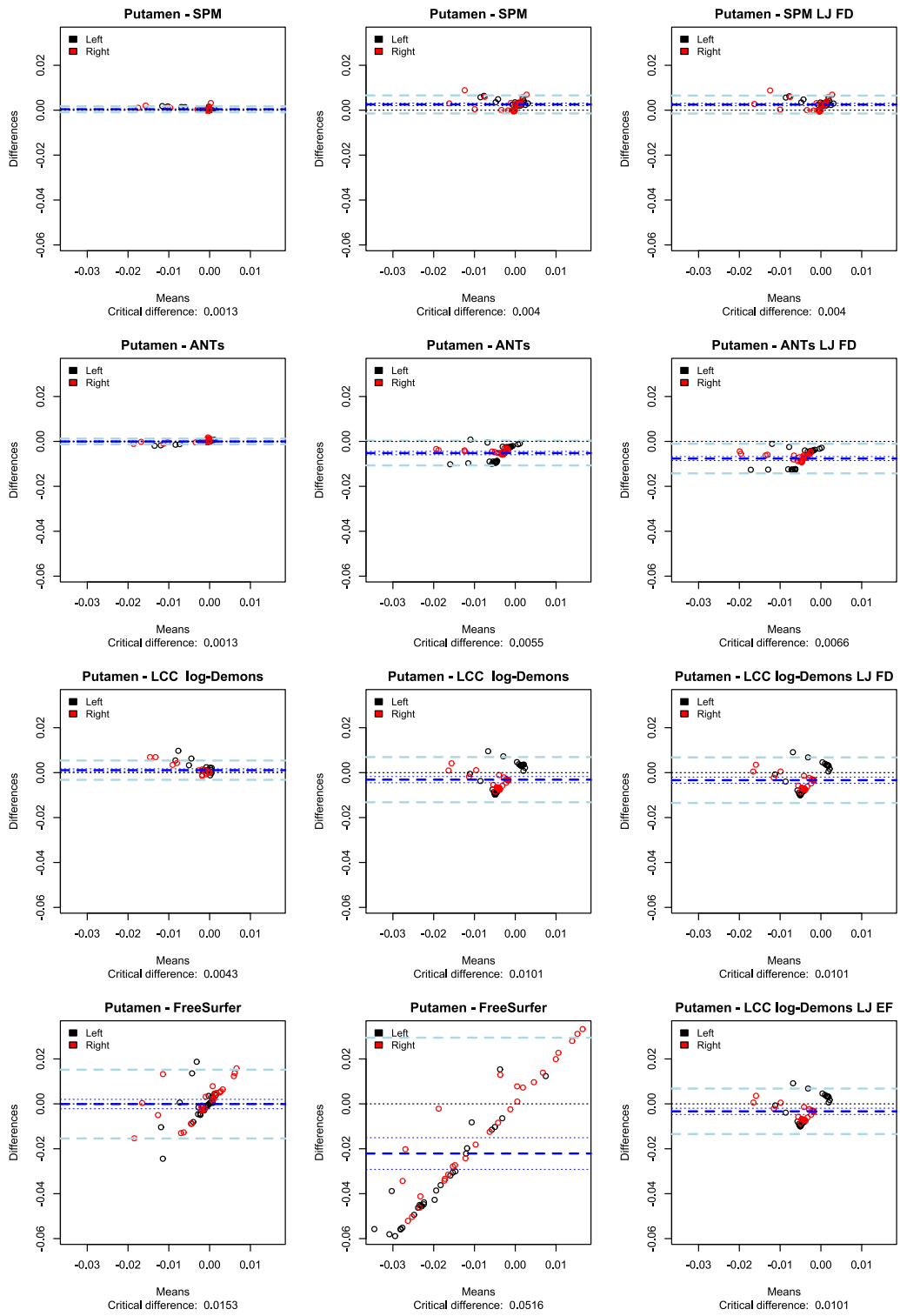

Figure 3.12: Bland-Altman plot for the putamen region for the four studied methods.

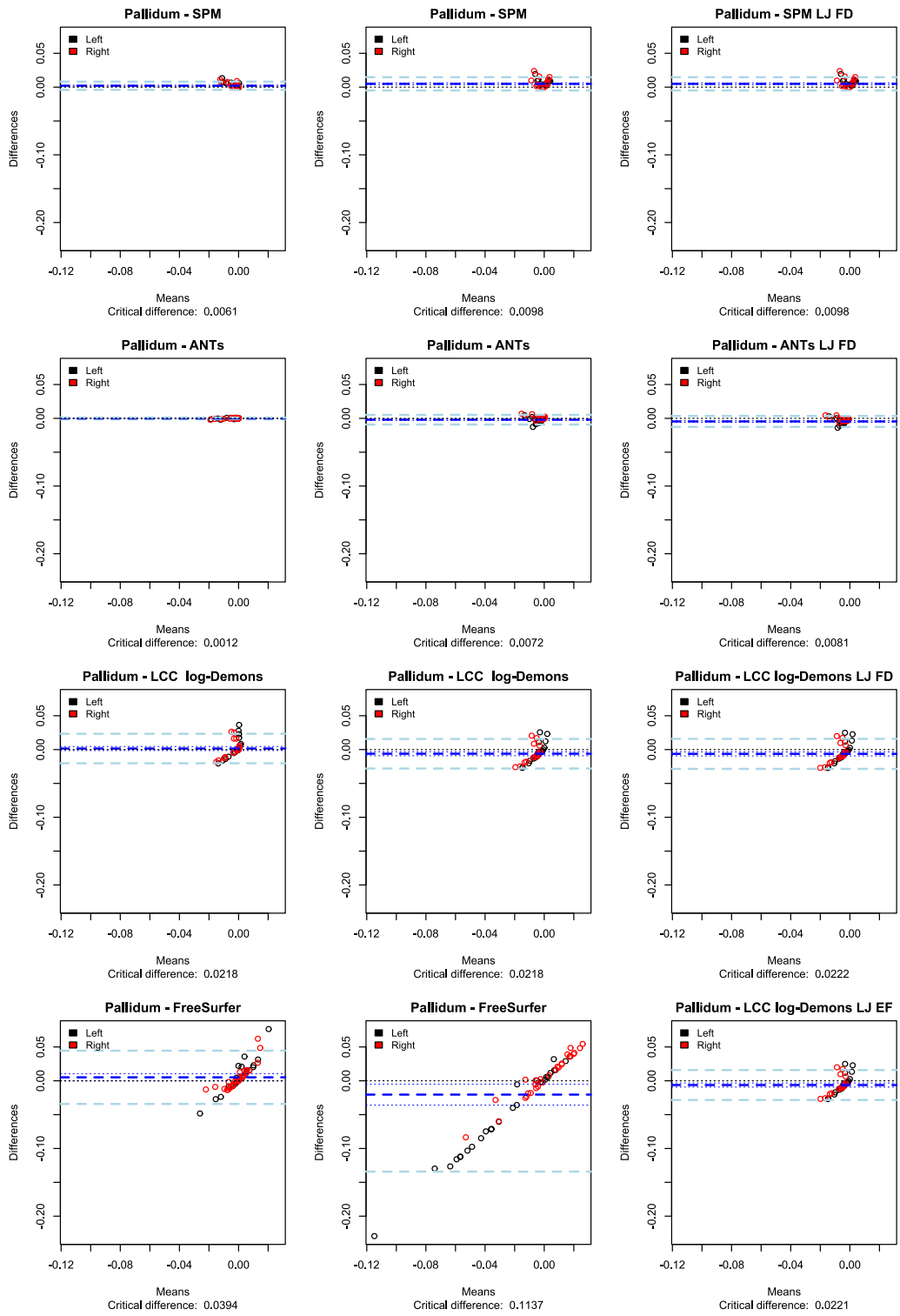

Figure 3.13: Bland-Altman plot for the pallidum region for the four studied methods.

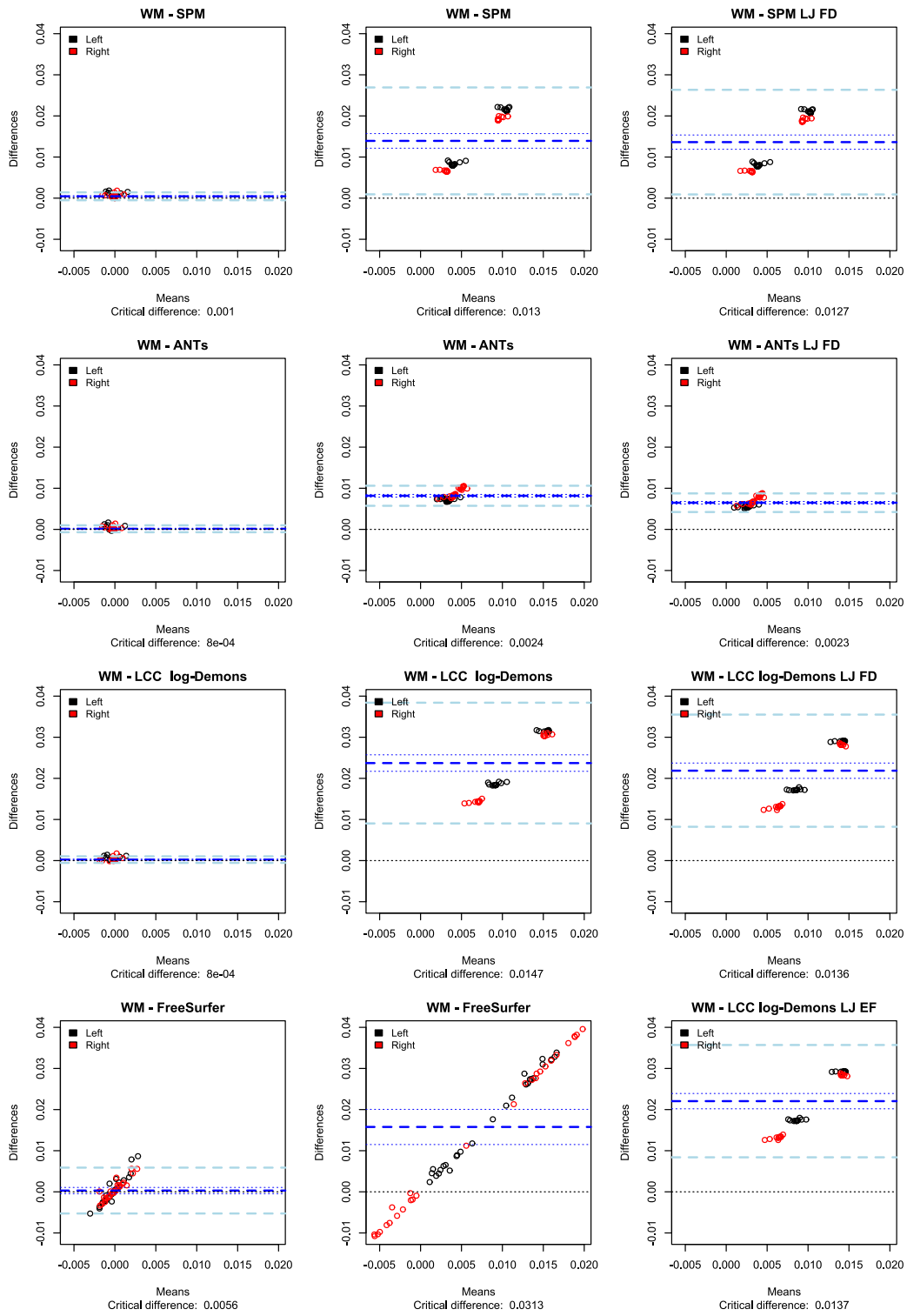

Figure 3.14: Bland-Altman plot for the white matter egion for the four studied methods.

Chapter 4

Beyond Volumetry: Deformation based Morphometry for the differential analysis of longitudinal SVF trajectories of subgroups

Contents

4.1 Introduction **65**

4.2 Disentangling inter-group and longitudinal changes: a geometric interpretation **67**

4.3 Implementation in the SVF case **69**

4.3.1 Construction of the inter-group axes and group longitudinal axis 70

4.3.2 Computation of the intelligible quantities of interest 72

4.4 Application to Alzheimer's disease: OASIS database **73**

4.4.1 Convergence/divergence of the patients with Alzheimer's and controls evolutions 73

4.4.2 Significant groups differences at 75 and 76 between controls and patients with Alzheimer's 73

4.5 Conclusion and Discussion **74**

In Deformation-based Morphometry (DBM), the comparison between the groups of deformation fields is generally difficult to interpret. We propose a method that brings an intelligible way to quantify both group differences and longitudinal evolutions by disentangling the group differences from the longitudinal evolutions. To do so, we first build the geometric frame components in a consistent way with respect to each other. Second, we quantify the relative evolutions in terms of convergence/divergence and advance/delay with respect to the population mean trajectory. We illustrate the method on the OASIS database and show that the proposed indices are intelligible thus easing the interpretation of the results. For example we show that at 75 and 76 years, the lateral ventricles are in advance of around 5 months with respect to the population mean trajectory. We also show that the differences between patients and controls do not evolve during this time period.

4.1 Introduction

In neuroimaging, identifying longitudinal structural differences between two groups is generally done by using one of the three principal morphometric methods [Frackowiak 2003]: Voxel-based Morphometry, Tensor-based Morphometry, and Deformation-based Morphometry, that we describe below. These methods are classified based on their variable of interest and each of them has its advantages and inconvenients.

The first method, Voxel-based Morphometry (VBM) [Ashburner 2000], relies on intensities or tissue probability maps to discriminate between two groups and thus makes no assumption on the transformation. However, this method only indicates a difference of tissue concentration between groups and is generally not able to detect very local changes since it involves a smoothing filtering (e.g. a Gaussian filtering of 8mm).

The second method, Tensor-based Morphometry (TBM), focuses on the first derivative of the deformation field (Jacobian matrix) resulting from the registration of a pair of subject images. In the simplest case of TBM, the Jacobian determinant is used, making the group differences easily interpretable in terms of volume changes (scalar). Although many findings have been made thanks to this index, reducing the brain evolution to only a change in volume is often thought to be too simplistic and some information is lost. For example, it does not qualify the potential rotation of a region.

The third and last method is the Deformation-based Morphometry (DBM) [Ashburner 1998] in which the longitudinal evolution is defined by the deformation field resulting from the non-linear registration of a pair of subject images (or by the parameters of this deformation field). DBM goes beyond volumetry, since rotations and translations are included in addition to volume changes. This results in a much more complete description. However, the comparison between the group three-dimensional deformation fields is generally difficult to interpret. This lack of interpretability can partly be explained by the DBM configurations used to perform the group comparisons. Understanding their limitations, we could then propose a new way of comparison in DBM.

In order to describe the two main configurations that can generally be found, let us consider a population composed of two groups: A and B. We assume in this chapter that we are in a generic case where the deformation is parametric and can either be parameterised by an initial momentum, spline coefficients, or a Stationary Velocity Field (SVF). Thus the comparison focuses on the longitudinal deformation parameters' trajectories v_A and v_B.

The first configuration (Figure 4.2 A.) - used for example in the Hotelling's T^2 test - consists in comparing the two longitudinal trajectories in the common population-specific template space. The major drawback of this setting is that it only focuses on the longitudinal evolutions but cannot quantify the cross-sectional group differences.

The second configuration addresses this problem by using the two group tem-

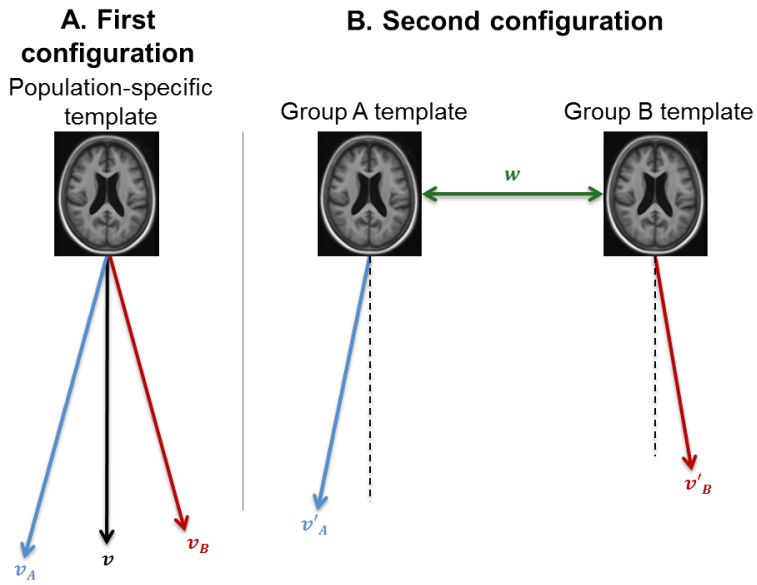

Figure 4.1: *Overview of the framework*: **A. First DBM configuration**: Only the population-specific template is used. The major drawback of this setting is that it only focuses on the longitudinal evolutions but cannot quantify the cross-sectional group differences. **B. Second DBM configuration** The two group templates are used. In addition to the groups' longitudinal trajectories, this setting enables the use of the cross-sectional differences w. However, the two group templates are defined independently - $T_A \neq (T_B \circ \phi_w)$ - which would bias the study towards one group anatomy.

plates instead of the population-specific template (Figure 4.2 B.). In addition to the groups' longitudinal trajectories, this setting enables the use of the cross-sectional differences w (also called inter-group differences) at the baseline t_0 and at the follow-up time point t_1 (not represented on the figure). Using the inter-group axis it is possible to define the measurable concept of convergence/divergence. If the inter-group differences decrease between t_0 and t_1 then the longitudinal trajectories are converging. Conversely, if w increases with time then the trajectories are diverging. Although useful for the group comparison, this second configuration still has a major drawback. The two templates are not directly comparable since they are defined independently. Indeed, $T_A \neq (T_B \circ \phi_w)$. This biases the group comparison towards one group anatomy.

Using the pros and cons of the two previous configurations, we can define the features of the ideal DBM comparison method. Ideally we would like to compare the trajectories with respect to the common longitudinal trajectory v as it is done in the first configuration (so that no group anatomy is privileged), while keeping the possibility of studying the cross-sectional differences. The aim of the statistical group comparison DBM method would be to analyse and interpret the differences

and similarities between two different groups using both longitudinal and cross-sectional information in a consistent way. However to define such a method, two important questions are to be answered. First, instead of defining the templates independently as it is usually done, "how can the different templates and evolutions be defined in a consistent way with respect to each other?". Second, "how can we disentangle the inter-group differences from the longitudinal evolutions?".

We therefore propose a method that brings an intelligible way to quantify both group differences and longitudinal evolution in the DBM context. It facilitates the interpretation of the comparison results by using two measurable indices that quantify the relative evolutions in terms of convergence/divergence and advance/delay with respect to the population mean trajectory. To our knowledge, no similar method has been proposed yet in the longitudinal context.

In section 4.2, we develop a new methodology for the interpretable group comparison of both longitudinal and cross-sectional deformations of the brain. We detail its construction step-by-step in section 4.3. In section 4.4, we then illustrate the method on the open-access OASIS database comparing patients with Alzheimer's and controls trajectories. This shows how the method eases the interpretation of the results. Finally, in section 4.5, we conclude about the proposed method and present the potential perspectives.

4.2 Disentangling inter-group and longitudinal changes: a geometric interpretation

In this section, we present the principles of the method and we then detail its construction step-by-step in the next section. The method relies on a simple geometric structure - also called frame - (cf. Figure 4.2) defined by five main components: the population mean longitudinal trajectory v, the two inter-group axes w_0 at t_0 and w_1 at t_1, and the two group longitudinal trajectories v_A for the group A and v_B for the group B. An important point of the methods concerns the definition of all these components dependently as well as which are the meaningful indices.

Let us consider the following initial configuration: the population-specific template at t_0 has already been computed by an iterative averaging of intensities and deformations as described in chapter 2. Here we chose the initial time point t_0 as the reference: it is a convenient configuration for describing the method. However, different possibilities exist for the choice of the reference time point. This topic is discussed in the section 4.5. The population longitudinal trajectory v has also been computed. One must note that all the deformations will be linearised around the population specific template at the reference time point t_0. The first important step of the method consists in defining consistently the inter-group axis w_0 at t_0. The inter-group axis can be seen as the trajectory that best discriminates between the two groups. Instead of creating the two group templates at t_0 independently, we propose to derive them from the population-specific template at t_0. We use the inter-subject deformations resulting from the population-template creation. They

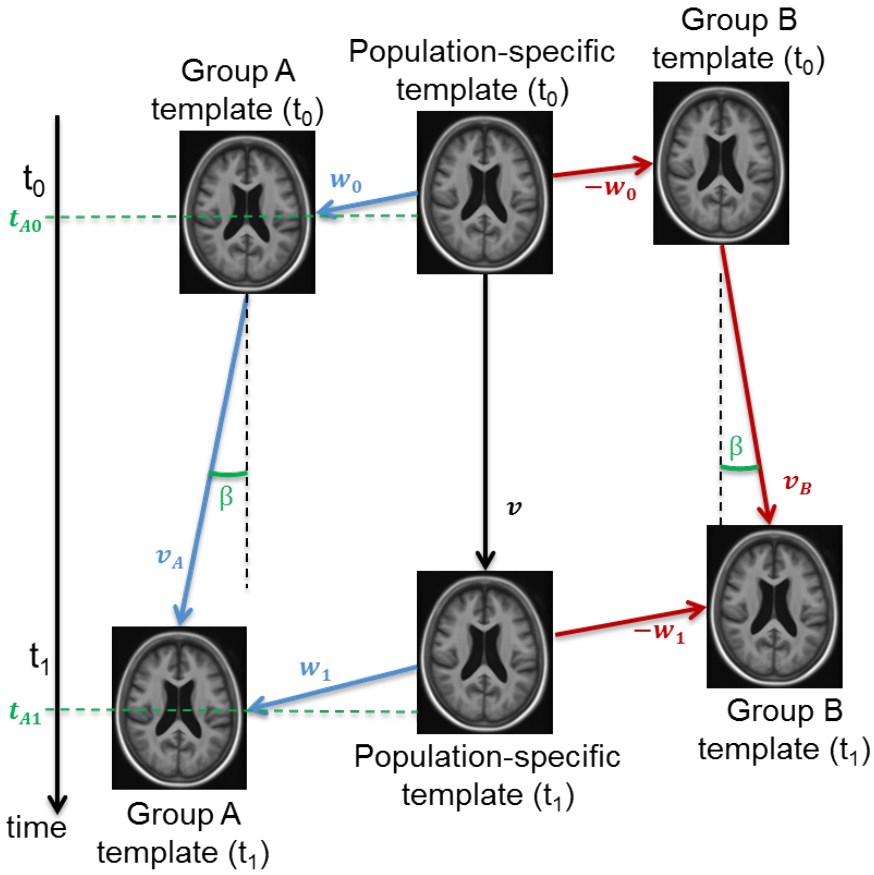

Figure 4.2: *Overview of the geometrical frame*: Five main components define this geometric frame: the population mean longitudinal trajectory v, the two inter-group axes w_0 at t_0 and w_1 at t_1, and the two group longitudinal trajectories v_A for the group A and v_B for the group B. We define two intelligible indices for the quantification of the group differences: t_{A_0} and t_{A_1} the advance/delay time, and β the convergence/divergence angle

map the population-specific template to the subjects baseline images. The inter-group difference w_0 is then equal to the average, for the group A only, of these inter-subject deformations. Since the population-specific template is by construction centered (see Section 2.2.3.4 and Algorithm 2) with respect to the population, one can interchangeably choose to average the group A or the group B inter-subject deformations. Using this construction technique we can ensure that the population-specific template is centered with respect to the two group templates. The group atlases are then built by warping the population-specific template with the inter group axis w_0. We notice that due to the symmetric construction of the group templates (with respect to the population-specific template) it is then possible to focus on the construction of one group only.

The population-specific template at t_1 is also defined in a consistent way, since it is derived from the population-specific template at t_0 using v; if v is the annualised longitudinal trajectory then $t_1 = t_0 + 1$ year. Then, the inter-group axis w_1 can be built in a similar way as w_0 averaging the inter-subject deformations at t_1. In the case of w_1, the inter-subject deformations at t_1 are not computed initially. Using the linearity of the geometric frame, each deformation is defined by summing the inter-subject deformation at t_0, its longitudinal subject specific model and the longitudinal mean population. In the case of SVFs, this results from the approximation of the Baker-Campbell-Hausdorff (BCH) formula at the zeroth order. For other parameterisations, it has to be checked if this assumption remains valid. However, since we are dealing with relatively small longitudinal deformations, a Taylor expansion of the composition should be achievable. The two group longitudinal trajectories are then derived by linearity from the two inter-group axes and the population longitudinal axis. We use this group longitudinal trajectories to define the divergence/convergence index.

Once all the components of the geometric frame are defined consistently from the reference population-specific template, it is possible to define an intelligible index that will be used to quantify the group differences. This is done by disentangling the group differences from the longitudinal evolution, at t_0 and t_1. Let us consider the case of t_0. This consists in projecting w_0 on the longitudinal population trajectory v. This enables us to quantify the advance or the delay of one group at t_0 with respect to the mean longitudinal trajectory (in months). We notice that due to the symmetric construction of the group templates, if the group A is in advance with respect to the mean, then the group B is delayed of the same amount.

In conclusion, the key part of the method lies in the consistent definition of all the five components that are derived from the reference population-specific template instead of being computed independently. In addition to the convergence/divergence index, this enables us to define an intelligible index to quantify the group differences with respect to time (more precisely with respect to the population mean longitudinal trajectory). In the next section, we detail the way to implement our two-group comparison method in the specific case where the deformations are parameterised by SVFs.

4.3 Implementation in the SVF case

We now consider the case where the deformation is parametrised by a SVF and there are N_A and N_B subjects for respectively group A and group B with longitudinal observations for each of them. We apply the Longitudinal Log-Demons Framework (cf. chapter 2) [Hadj-Hamou 2016] to each subject: a pre-processing step and a position correction step are first performed. Then a non-linear registration is done followed by the creation of the subject-specific longitudinal SVF trajectory $\hat{v}_i$. The population-specific template T_0 is then created and the $\hat{v}_i$ are transported in T_0 along the inter-subject (subject-to-T_0) SVF w_i^0. The population mean longitudi-

nal trajectory v is then obtained by averaging the transported SVFs. Thanks to the symmetry of the method, from now on we only consider the group A for the construction of the frame and for the computation of the indices.

The method then necessitates four main steps, which we present in details below: **1)** Construction of the inter-group axis w_0 at t_0, **2)** construction of the inter-group axis w_1 at t_1, **3)** construction of the longitudinal group trajectory v_A, **4)** computation of the quantities of interest.

4.3.1 Construction of the inter-group axes and group longitudinal axis

The three main steps for the construction of the geometrical frame are summarised in Algo 4, and illustrated for two subjects on Figure 4.3.

Algorithm 4 Construction of the inter-group axes at t_0 and t_1 and group A longitudinal trajectory

Input: v: population mean longitudinal SVF, w_0^i: inter-subject SVF (T_0 to subject i) at t_0, and $\hat{v}_i$: subject-specific SVF longitudinal trajectory.

Output: w_0: Inter-group SVF at t_0, w_1: Inter-group SVF at t_1, v_A: group A longitudinal trajectory .

Compute the inter-group SVF w_0 at t_0
$w_0 = \frac{1}{N_A} \sum_i w_0^i$

Compute the inter-group SVF w_1 at t_1 (Z is the BCH formula)
$w_1^i = Z(Z(w_0^i, -\hat{v}_i), v) \simeq w_0^i - \hat{v}_i + v$, for each subject of the group A
$w_1 = \frac{1}{N_A} \sum_i w_1^i$

Create the group A mean longitudinal trajectory SVF v_A
$v_A = Z(Z(w_1, -v), -w_0) \simeq w_1 - v - w_0$

Useful but additional step: Create the group-specific templates A_0 and A_1 respectively at t_0 and t_1 as well as the population-specific template T_1 at t_1
$A_0 = T_0 \circ \exp(w_0)$, $A_1 = T_0 \circ \exp(v + w_1)$, *and* $T_1 = T_0 \circ \exp(v)$

In practice there is no necessity to create the group-specific templates A_0 and A_1. However, it is easier to see the differences between the two images A_0 and T_0 than to visualise the three-dimensional inter-group SVF w_0. It is even possible to ease the visualisation by amplifying the effect of the deformation (also called caricature the evolution). We notice that the number of resamplings is minimised when creating T_1^A by using the zeroth order of the Baker-Campbell-Hausdorff formula: $\exp(v) \circ \exp(w_1) \approx \exp(v + w_1)$

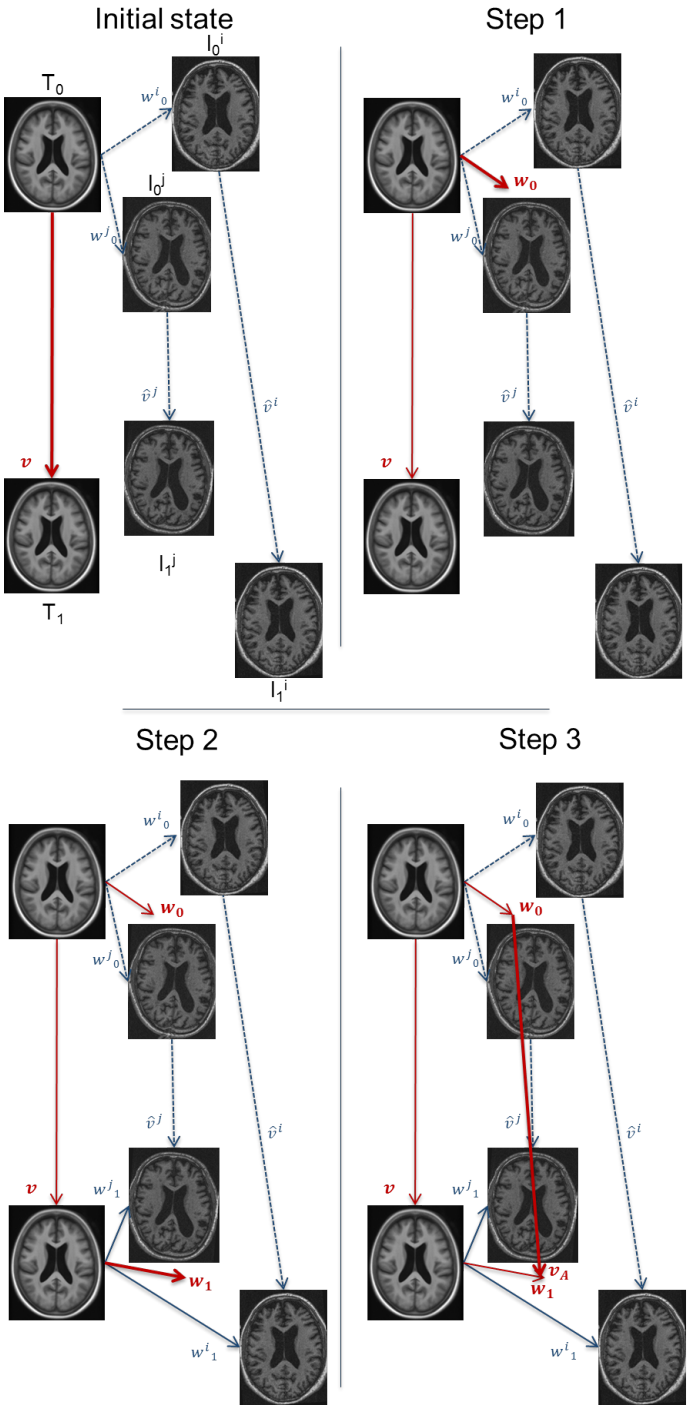

Figure 4.3: *Construction steps for the geometric frame*: The 3 steps are illustrated in the case of a group with two subjects. **A. Initial setting** We can see v, the population longitudinal SVF, w_0^i and w_0^j the inter-subject SVFs (T_0 to subject) at t_0, and $\hat{v}_i$, $\hat{v}_j$ the subject-specific SVF longitudinal trajectories. **B. First step**: Creation of the inter-group axis w_0 at t_0. **C. Second step**: Creation of the inter-group axis w_1 at t_1. **D. Third and final step**: Creation of the longitudinal group A trajectory SVF v_A.

4.3.2 Computation of the intelligible quantities of interest

We now detail the computation of the two quantitative indices in the SVF case. The first step in the quantification of the group differences at t_0 and t_1 is to determine if there actually are group differences, i.e if the inter-group SVF w_0 (or w_1) is significantly different from zero. We perform a Hotelling's T^2-test at each voxel:

$$T_A^2 = N_A {w_0}^T \Sigma_A^{-1} w_0, \ \Sigma_A\text{: covariance matrix for the subjects of group A.}$$

The null hypothesis is that w_0 is equal to zero. During this step, we apply no correction for multiple comparisons. If w_0 is null then it simply means that there is no group difference at the considered voxel for the given time point.

Quantifying the group differences with respect to the population longitudinal evolution

For the voxels where there exists a group difference, we quantify this difference using the orthogonal projection of w_0 on the population longitudinal trajectory v: t_{A_0} and t_{A_1} respectively at t_0 and t_1.

$$t_{A_0} = \frac{1}{\|v\|}(w_0 \cdot v), \text{ and } t_{A_1} = \frac{1}{\|v\|}(w_1 \cdot v).$$

If the index t_{A_0} is positive, the group A is in advance of t_{A_0} months at t_0 with respect to the population longitudinal trajectory v and by symmetry the group B has a corresponding delay with respect to v. Conversely, if t_{A_0} is negative, the group A is late of t_{A_0} months at t_0 with respect to the population longitudinal trajectory v and the group B would be symmetrically in advance with respect to v.

Quantifying the convergence/divergence of the group longitudinal evolutions

The other main advantage of our method is the possibility to quantify the two groups longitudinal evolution v_A and v_B in terms of convergence and divergence. We thus define the convergence/divergence ratio as β:

$$\beta = \arccos\left(\frac{v \cdot (n \times (v^A \times n))}{\|v\|\|(n \times (v^A \times n))\|}\right)\frac{180}{\pi}, \text{ with } n = \frac{w_0}{\|w_0\|} \times \frac{v}{\|v\|}.$$

Similarly to the t_{A_0} (or t_{A_1}) index (in months), the β index has an understandable unit. β is an angle (in degrees) which eases the comprehension of the convergence or divergence. Moreover, β is only defined if w_0 and v are different from zero. The convergence and divergence are defined according to the values of β, if:

- $\beta > 0$: divergence of the group evolutions,
- $\beta < 0$: convergence of the group evolutions,
- $\beta = 0$: the group differences remain the same.

In the particular case where w_0 is null, it is only possible to qualify the divergence. If there exist inter-group differences at t_1 - i.e w_1 is different from zero - then the group longitudinal trajectories are divergent. Conversely, if there is no inter-group difference at t_1, the group longitudinal trajectories remain the same.

4.4 Application to Alzheimer's disease: OASIS database

We now illustrate the proposed method on the OASIS database, following the study initiated in chapter 2 [Hadj-Hamou 2016]. The clinical cohort considered here remains the same and is composed of 64 patients diagnosed with very mild to moderate Alzheimer's disease, and 72 healthy individuals (further information can be found in Appendix 4.5). We use all the already processed results - using the Longitudinal log-Demons Framework - from chapter 2 and then start after the template construction. We compare the healthy subjects group versus the patients with Alzheimer's. Here, the inter-group axis corresponds to the Alzheimer's-healthy axis. Using the proposed method, we answer the two following questions. "Do patients with Alzheimer's and subjects have brain structures that converge ?" and "How much do structures differ between patients and controls at a given time point?".

4.4.1 Convergence/divergence of the patients with Alzheimer's and controls evolutions

Concerning the group longitudinal evolutions of the patients and the subjects, we find that the two groups longitudinal trajectories are not diverging nor converging. Thus, the group differences remain constant between 75 and 76 years old: the cross-sectional differences between patients with Alzheimer's and controls neither increase nor decrease. Let us now quantify these group differences using the concept of advance and delay.

4.4.2 Significant groups differences at 75 and 76 between controls and patients with Alzheimer's

We now consider the inter-group differences at $t_0 = 75$ and $t_1 = 76$ (from the point of view of the patients with Alzheimer's disease). We see on Figure 4.4 that the differences between the patients and controls are the same at 75 and 76 years of age. This is consistent with the convergence/divergence result. Instead of describing the group differences at both time points, we therefore focus on one time point $t_0 = 75$. We see that on average for the patients, the lateral ventricles are in advance of around 5 months with respect to the population mean trajectory. The other regions in advance are the corpus callosum, the occipital lobe and the cerebellum. Conversely, the temporal lobes have a delay with respect to the longitudinal mean population trajectory. More specifically, the hippocampi have a delay of around 1.8 months. Finally, we notice that there is no difference between the parietal cortex of the patients with Alzheimer's and the controls at t_0 or t_1.

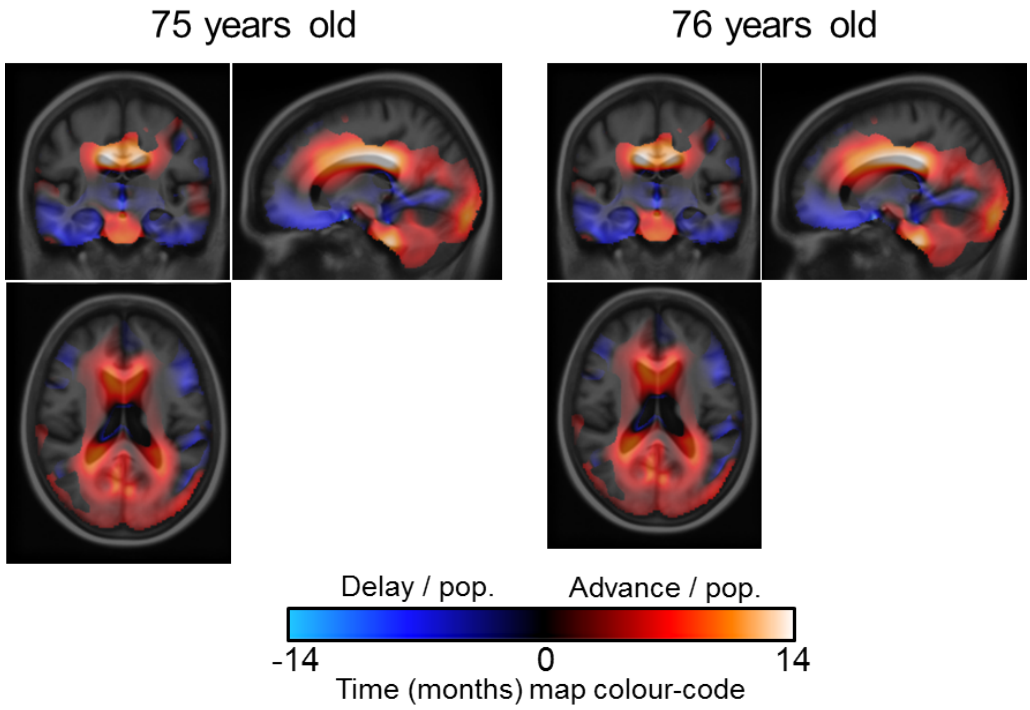

Figure 4.4: *Group differences*: **Quantification of the inter-group differences for the group of patients with Alzheimer's using t_{A_0} and t_{A_1}**: We see that the differences between the patients and controls are the same at 75 and 76 years of age. At $t_0 = 75$ years, on average for the patients with Alzheimer's, the lateral ventricles are in advance of around 5 months with respect to the population mean trajectory. The other regions in advance are the corpus callosum, the occipital lobe and the cerebellum. Conversely, the temporal lobes have a delay with respect to the longitudinal mean population trajectory. We notice that there is no difference between the parietal cortex of the patients with Alzheimer's and the controls at t_0 or t_1.

4.5 Conclusion and Discussion

We presented a novel method for the comparison of two groups based on DBM. This method enables to quantify the group differences in an intelligible way, which eases the interpretation of the results. More specifically, the method uses both cross-sectional and longitudinal information to quantify the cross-sectional group differences in terms of advance/delay with respect to the population longitudinal trajectory. Moreover it quantifies the group differences in the longitudinal evolution in terms of divergence/converge.

A potential advantage of the method not described in this chapter is that the comparison can be performed at different structure levels. In our illustration, we chose to use one vector per voxel (the most generic form of the method), but we

could have defined regions of interest (ROI) and averaged the deformation on each ROI so that we would only work with one vector per ROI. This would be equivalent to studying only the translations. Another possibility is to first summarise the deformation field using affine projections on each ROI (cf. Appendix 4.5) and work with only twelve affine parameters on each ROI.

Based on the brief illustration, we see that the t_A index (advance/delay) might have interesting applications for clinicians: the possibility to quantify the group differences in terms of months really makes sense and is more intelligible than some other existing indices.

Moreover, it could be interesting to study the similarities between this work and the work of [Schiratti 2015] where the authors propose a model of logistic trajectories with more degrees of freedom than our model (they define the acceleration in addition to the time shift). However, their focus is not the same since in our work the challenge is to distinguish between two groups.

Concerning the choice of the reference time point for the method, since we are following the disease's evolution, it seemed relevant to start at the initial time point, and we therefore used the baseline t_0 as the initial time point. However, the method could be used with t_1 as reference time point, meaning that the deformations would be linearised around the population-specific template at t_1 and potentially modifying the results. Hence, there is a potential asymmetry bias in the method we propose. We did not check if the results changed when modifying the reference time point. In order to limit the asymmetry to a minimum, it could be interesting to use the "middle" time point between t_0 and t_1 to perform the analysis. This was not the aim of this article as we wanted in first instance to propose an intelligible DBM comparison method and its related concepts. However, the way forward should be to study this possibility and ensure that the proposed indices are robust to the potential asymmetries. With this issue solved, a future direction for this work will be to work on the extension of the method to three groups or more. To do so, a preliminary step would be necessary, e.g. to perform a principal component analysis on the inter-group SVFs and then work on the first eigen-SVF.

Acknowledgments

This work was partially funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modelling and Analysis of Dynamic Medical Images).
OASIS grant numbers: P50 AG05681, P01 AG03991, R01 AG021910, P20 MH071616, U24 RR021382.

Appendix

Spatial Dimensionality Reduction of the Deformation using Affine Projection

This method consists in reducing the dimension of the deformation field, describing it with the use of the four elementary transformations: translation, rotation, scale and shear. This would provide us with a dictionary of words to describe a regional evolution. We propose to decrease the dimension of v_{01} by approximating this non-parametric transformation by a locally affine one. We thus divide the subject's brain into regions (anatomically defined for example). Therefore, each region is defined by a probabilistic mask with its probability weights w^r.
For each region R_r, we project the SVF v_{01} onto the linear space of log Euclidean affine transformation as described in [Seiler 2012].
We obtain the following affine approximation:

$$v_{01}^r(x_i) = M^r.\widetilde{x}_i$$

- $\widetilde{x}_i = \begin{bmatrix} x_i \\ 1 \end{bmatrix}$, are the homogeneous coordinates, x being the spatial coordinates,
- M^r is the matrix logarithm of affine transformation T^r in region R_r such that

$$\begin{bmatrix} M^r \\ 0 \end{bmatrix} = \log(T^r) = \log \begin{pmatrix} A^r & t^r \\ 0 & 1 \end{pmatrix}$$

  with :

  - A^r the linear part of the affine transformation,
  - t^r the translation part.

In the log-Euclidean Affine framework, the affine transformation M^r is the sum of 4 elementary transformations: a Translation vector t, a Rotation matrix R, a Scale matrix S and a shear matrix K : $M^r = t + R + S + K$.

List and information about the OASIS Subjects Used

The subjects are the same as the one used in [Hadj-Hamou 2016], the list of OASIS subjects we used can be downloaded here: `http://www-sop.inria.fr/teams/asclepios/data/Pipeline/lists/OASIS_List_Subjects.csv`. We kept the complete set of demented and non-demented subjects and only excluded the subjects who converted - from non-demented to demented - during the study. The images can be found here (we used the mpr-1 acquisitions only): `http://www.oasis-brains.org/app/template/Tools.vm;jsessionid=9A56E751939B1DAE41049A46BBAAFB3E#services`. Socio-demographic and clinical information on the sample can be found Table 4.1.

Group	Non-Demented	Demented
Age at baseline (years)	75 ± 8	75 ± 7
Female/Male	50/22	28/36
Education (years)	15 ± 3	14 ± 3
Mini Mental State Examination (MMSE)	29 ± 1	25 ± 3

Table 4.1: Socio-demographic and clinical information of the study cohort.

We notice that there is a statistically significant (by Fisher exact test) gender imbalance across the two groups (50/72 vs. 28/64).

Chapter 5

Effect of sexual dimorphism on the healthy morphological evolution of the brain during adolescence

Contents

5.1 Introduction **79**

5.2 Dataset and methods **80**

5.2.1 The IMAGEN database 80

5.2.2 Image processing 81

5.2.3 Definition of the regions of interest 81

5.3 Longitudinal volumetric changes **81**

5.3.1 Regional volume changes: Male-Female Comparison 81

5.3.2 Local volume changes: Male-Female Comparison 85

5.4 Beyond volumetry to study the effects of sexual dimorphism during adolescence **87**

5.4.1 Multivariate comparison of female and male longitudinal SVF trajectories 87

5.4.2 Sexual differences in the cross-sectional and longitudinal evolution between 14 and 16 years old 88

5.5 Conclusion and Discussion **91**

Adolescence is a period of major changes where the brain develops from an immature to an adult mature state. This maturation period is particularly marked by the divergence in males and females developmental trajectories.
In this article, we propose to apply a deformation-based morphometry pipeline to study the sex differences in the longitudinal evolution of the brain during adolescence using the IMAGEN longitudinal database.
The volumetric results are mainly consistent with the existing literature. We found an exception in the white matter. In this region, our study found a decreasing volume between 14 and 16 for the females whereas the literature reports an expansion of the white matter. Our finding was replicated using `SPM12` *longitudinal registration tool. Using the more local log-Jacobian map we found that only the pre-frontal cortex and parts of the white matter had significantly different volume evolutions between males and females. The pre-frontal cortex volume decrease is more important for females*

than for males.
We then go beyond volumetry using our geometric group comparison on the three-dimensional deformation fields. We show that the male-female differences increase with time in most of the brain. More specifically, we find that there is no difference between males and females cortex at 14. An important differentiation occurs during the following years, and at 16 years of age, females are in advance on the frontal cortex of around five months with respect to the population mean longitudinal trajectory.

5.1 Introduction

Adolescence is a period of major changes where the brain develops from an immature to an adult mature state. This maturation period is particularly marked by the divergence in males and females developmental trajectories, for both white and grey matter. For example, cortical and sub-cortical grey matter volumes are both reported to grow and then decrease, but they peak earlier in females than in males (around one to one and a half year) [Lenroot 2007]. Moreover in addition to being a period of physical changes, adolescence is often the onset of developmental disorders, which have a greater incidence in females than in males. It therefore appears necessary to understand the development of such disorders during adolescence. In order to do so we first need to understand healthy brain development and in particular the effects of sexual dimorphism during that period. Most of the major neuropsychiatric disorders are now thought to arise due to deviations from normal brain development during adolescence [Giedd 2008].

Studies on the topic mainly focus on volume comparisons and no consensus can generally be found between the studies results. For example, the hippocampus is reported to be larger in female than in male [Neufang 2009]. However, other studies report the opposite trend [Goddings 2014]. Different factors might explain these differences of results between studies. First, an important part of the studies focuses on cross-sectional data. In this setting, the confounding effect of inter-individual morphological variability is relatively high with respect to longitudinal datasets. Second, [Giedd 2006] has shown that the evolution of most of the grey matter regions consists in a volume increase at the beggining of adolescence followed by a volume decrease. Studying the adolescent brain over a large period of time (as it is done in most studies) might therefore capture both the increase and the decrease of the structures, resulting in an overall evolution with no changes. Third, the analysis methods used are generally segmentation-based methods, which have been shown in chapter 3 to be less accurate and reproducible than registration-based methods. To address this potential limitation, the use of Voxel-based Morphometry has increased. However there are so far very few analysis using non-linear registration to perform Tensor-based Morphometry or Deformation-based Morphometry (DBM).

In this article, we propose to apply the deformation-based morphometry pipeline presented in chapter 2 to study the sex differences in the longitudinal evolution

of the brain during adolescence. In order to find a consensus, we compare our results to the results of three other popular segmentation-based and registration-based methods. We use the IMAGEN dataset. It has the advantage of being longitudinal and limited to a small age range (14 to 16 years old). After validating the volumetric changes, we go beyond volume changes in the brain by analysing the three-dimensional deformation field. Using this longitudinal database, we wish to bring new insights on the the evolution during that age period.

In section 5.3, we compare our regional volumetric results to the literature and to other methods results. We then apply in section 5.4 the DBM method proposed in chapter 4. The main results show that at 14 years of age, no difference exists between males and females pre-frontal cortex and that an important differentiation occurs during the two following years in this region. In addition to the already known volumetric findings, our results give complementary relevant information for the understanding of sexual dimorphism during healthy brain development in adolescence.

5.2 Dataset and methods

5.2.1 The IMAGEN database

We use the IMAGEN database [Schumann 2010] as it is longitudinal. By using each subject as his or her own control, longitudinal datasets have the advantage to reduce as much as possible the high inter-subject variability that can cause important variations in the studies based on cross-sectional databases. Moreover, we try to limit the study to a small age range in order to focus on either the volume increase or the decrease. Otherwise, we might find no change (if the volume of the structure increases then decreases of the same amount). Among the different centers of the IMAGEN cohorts only the French and the German have scanned the adolescents at two years of interval (the other centers used a four-year interval). We thus use the French subset of the European longitudinal IMAGEN database. It consists (after quality control) of 120 healthy adolescents scanned at $t_0 = 14$ years of age (baseline) with a follow-up scan at $t_1 = 16$. The images are 3 Tesla, T1-weighted MRI scans. We did not include the German subset as it might be useful to replicate the findings of the French database. The French subset population has a female ratio of 50.8%, appropriate for studies on sexual dimorphism. More information can be found in Table 5.1.

Group	**Female**	**Male**
Age at baseline (years)	14.3 ± 0.5	14.3 ± 0.4
Age at follow-up (years)	16.8 ± 0.6	16.7 ± 0.6
Number of subjects	61	59

Table 5.1: Socio-demographic and clinical information of the study cohort.

5.2.2 Image processing

The data first went through a manual quality control. Then we applied the preprocessing (including intensity inhomogeneity correction) and position correction parts (intra-subject rigid registration and affine registration to the MNI152 atlas) of the Longitudinal Log-Demons framework [Hadj-Hamou 2016] (and chapter 2). Each pair of longitudinal subject images is then non-linearly registered using the LCC log-Demons with confidence mask. The construction of the population-specific template at $t_0 = 14$ years is followed by the transport of each subject longitudinal evolution into the template.

5.2.3 Definition of the regions of interest

During adolescence, the brain is supposed to go through a dual structural development [Mills 2014]. This development starts with subcortical structures and then continues with the cortex development and more particularly the pre-frontal cortex. We therefore decide to focus our analysis on both grey matter subcortical structures and cortex. In this study, we define 11 regions of interest (ROI): **1)** the cortex, **2)** the white matter, **3)** the lateral ventricles, **4)** the orbito-frontal cortex, and 7 subcortical grey matter structures: **5)** the thalamus, **6)** the caudate, **7)** the putamen, **8)** the globus pallidus, **9)** the hippocampus, **10)** the amygdala, and **11)** the nucleus accumbens. The regions of interest are segmented using Freesurfer [Reuter 2012].

5.3 Longitudinal volumetric changes

5.3.1 Regional volume changes: Male-Female Comparison

We first study the longitudinal evolution by brain region of interest, which enables us to directly compare our results with the literature. Here we study the average log-Jacobian values by ROI. For volume changes less than 10% we have seen in chapter 3 that this is equivalent to studying the relative changes of volume in the region. The results for males and females in the different ROIs can be found Figure 5.1. A red star under a box plot indicates that the mean changes are statistically different from zero (the null hypothesis being that there the mean of the group is not different from zero).

Three main evolution trends can be found. The first one concerns the lateral ventricles that increase with time for both females and males. The second type of evolution concerns a part of the grey matter structures (cortex, orbito-frontal cortex, thalamus, caudate nucleus, putamen, and accumbens). For this subset of structures the volume decrease is more important in females than in males (in some cases there is even no volume decrease for males). In the third case, for the remaining structures, the volume increase is more important for males than for females (in some cases there exists a volume decrease for females). We now detail the evolution region by region.

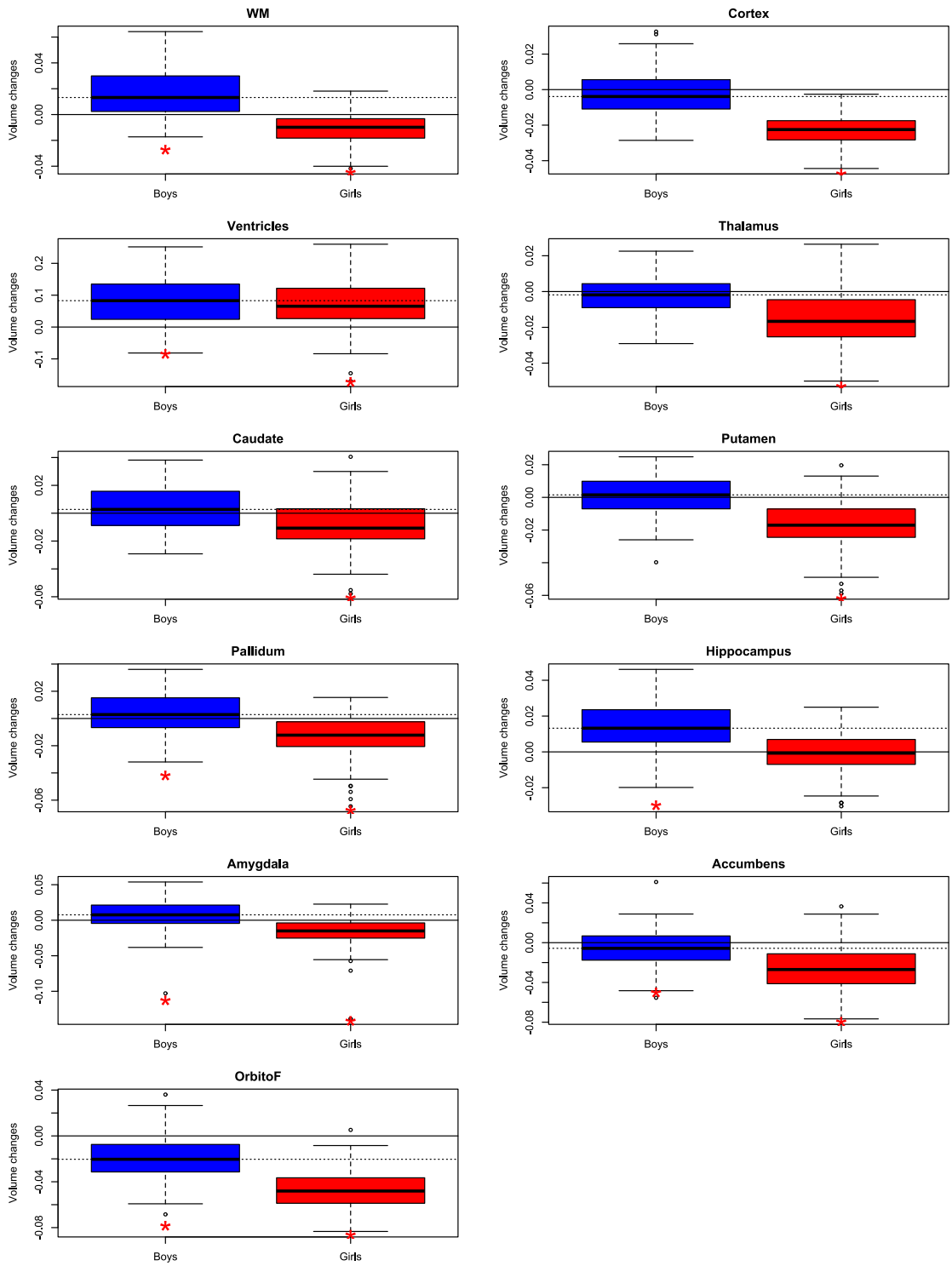

Figure 5.1: *Relative volume changes for the males (blue) and the females (red).* A red star indicates that the mean changes are statistically different from zero. We see that three main evolution trends can be found. The first one concerns the lateral ventricles that increase with time for both females and males. The second type of evolution concerns the cortex, the orbito-frontal cortex, the thalamus, the caudate nucleus, the putamen, and the accumbens. For this subset of structures the volume decrease is more important in females than in males. In the third case, for the remaining structures, the volume increase is more important for males than for females.

Lateral Ventricles

The lateral ventricles increase in volume for both females and males. The expansion is more important for males than for females. This is consistent with the literature [Giedd 2006] where the increase rate for males is more important than for the females.

White Matter

Although the white matter volume increases for males, our results show a decrease for females during that period of time. This is not reported in other studies that focus on a large age span (e.g. from childhood to adulthood). The validity of this finding is discussed in the following section.

Cortical Grey Matter

We find a contraction for females, while males have no statistical change in that region over that period of time. This is in agreement with [Giedd 2006] where the authors find that the females' cortical volume peaks earlier than in males.

Orbito frontal cortex: In this region, we find that there is an atrophy for both females and males. Females have a larger decrease rate than males. This is consistent with the fact that the cortical volume increases then decreases during adolescence (as described above).

Subcortical Grey Matter

Putamen, caudate nucleus, and globus pallidus: These structures are the principal components of the basal ganglia. For both the putamen and the caudate nucleus, we notice on one hand an atrophy for the females between 14 to 16 years old. On the other hand, no significant volume change occurs for males. These results are consistent with the literature [Giedd 2006], where these structures are shown to follow an increasing then decreasing trajectory similar to the cortical grey matter. Our findings mean that for females the putamen and the caudate nucleus have already reached their maximum volume peak. Conversely, for males the putamen and the caudate nucleus are reaching their maximum volume peak. This is consistent with [Lenroot 2007] where the authors find that the caudate nucleus reach their volume peak at around 14 for males. Concerning the globus pallidus, we find a volume atrophy for females and a volume increase for males.

Hippocampus and amygdala: These two structures are particularly of interest for studies of sexual dimorphism since they are rich in hormone receptors (respectively the amygdala for males and the hippocampus for females). We find that the volume of the hippocampus increases for males while no statistically significant change can be found for females. A previous study on a cross-sectional dataset found

ROI	Males vs. Females
WM	**$8.7\ 10^{-33}$**
Cortex	**$2.0\ 10^{-33}$**
Lateral ventricles	$3.2\ 10^{-1}$
Thalamus	**$8.3\ 10^{-12}$**
Caudate	**$1.4\ 10^{-08}$**
Putamen	**$1.1\ 10^{-21}$**
Pallidum	**$2.3\ 10^{-16}$**
Hippocampus	**$1.2\ 10^{-17}$**
Amygdala	**$4.4\ 10^{-16}$**
Accumbens	**$1.5\ 10^{-12}$**
Orbito-frontal cortex	**$1.2\ 10^{-27}$**

Table 5.2: **p-values for the two-sample t-test on the mean log-Jacobian integration between females and males.** p-values inferior to 0.0045 (the threshold of 0.05 was corrected for multiple comparisons using Bonferroni method) are in bold. We see that there exist statistical differences between the mean value for males versus females in all the ROIs except for the lateral ventricles.

the opposite, with the hippocampal volume increasing significantly only in females [Giedd 2006]. As for the amygdala, we find an expansion for males and an atrophy for females. This finding is consistent with [Giedd 2006].

Accumbens: This structure decreases in volume for both sexes. However, we notice a larger rate of decrease for females than for males.

Thalamus: We find that the volume of the thalamus decreases only for females while no significant volume change can be found for males.

To confirm our findings, we then tested if the regional trends were statistically different for the two sexes. For each region, we thus performed a two-sample t-test between females and males. The null hypothesis is that there is no difference between the mean volume change of the males and the mean volume change of the females. We correct the p-value threshold (initially 0.05) for multiple comparisons using Bonferroni method: the corrected p-value threshold is $0.05/11 = 0.0045$. The results of the t-tests can be found in Table 5.2. We see that apart from the lateral ventricles, all the regions of interest exhibit statistically significant differences between males and females on the average relative volume changes.

In the regional analysis, we saw that some of our volumetric findings (e.g. in the white matter) were not in agreement with the literature. In order to test the reliability of our results, we compared them with the results from three other popular volumetric methods. One segmentation-based method, `FreeSurfer`, and two

registration-based methods: `ANTs` and `SPM12`. The results can be found in Appendix 5.5. Our results are in agreement with the other methods in every region for both males and females (on average, two other methods agree with our method). The only exception concerns the cortex for the males where we find no statistical volume change, whereas the other methods agree on a decrease of the volume. However, we notice that the trend between males and females in the cortex is conserved: the contraction of the cortex is more important for females than for males. We also notice that in some regions (e.g. the thalamus for the females), although registration-based methods all agree (atrophy for the thalamus), the segmentation based method gives a different result (expansion for the thalamus). This might partly explain the lack of consensus in the literature. Depending on the type of method used (registration or segmentation method), the results can be opposite on the same dataset.

5.3.2 Local volume changes: Male-Female Comparison

One of the potential bias of the regional-based studies we performed previously is their dependence on the quality of the segmentation. Therefore, after working on a volumetric ROI-based analysis of sexual differences, we get rid of the parcellations and work on the local maps of volume changes. We use the log-Jacobian maps of the subject's longitudinal evolution to study the sexual dimorphism during adolescence. The average map for the population (females and males) can be found on the left-hand of Figure 5.2. We analysed the group differences in the longitudinal volume evolutions by performing a two-sample t-test on the log-Jacobian maps using `SPM12` [Friston 2007]. The null hypothesis is that there exists no difference in the mean of the two groups. The results are corrected for multiple comparison using Family-wise error (FWE) with a corrected p-value of 0.05. The result map can be found on the right-hand of Figure 5.2. We only represented the regions where the male and female volume changes are statistically different.

Concerning the population mean longitudinal evolution, we see on Figure 5.2 that the lateral ventricles are expanding between 14 and 16. We also see the expansion in the white matter and in the hippocampi. An atrophy can be found in the cortical gray matter, in the thalamus, as well as in the putamen. We also notice the expansion in the meninges (cerebrospinal fluid) around the brain, which is compatible with an atrophy of the cortical grey matter.

We now focus on the sexual differences in the longitudinal volumetric evolutions. On the right hand of Figure 5.2, the areas in blue are the regions where the volume atrophy is significantly more important for the females than for the males. These statistical differences are visible in the pre-frontal cortex as well as in the temporal lobes.

Conversely, the areas in red correspond to regions where the volume expansion is significantly more important for the males than for the females. We see that the volume expansion of the white matter in the frontal, occipital and parietal lobes is more important for the males than for the females. This is consistent with the reported higher volume changes in white matter for males than females [Giedd 2006].

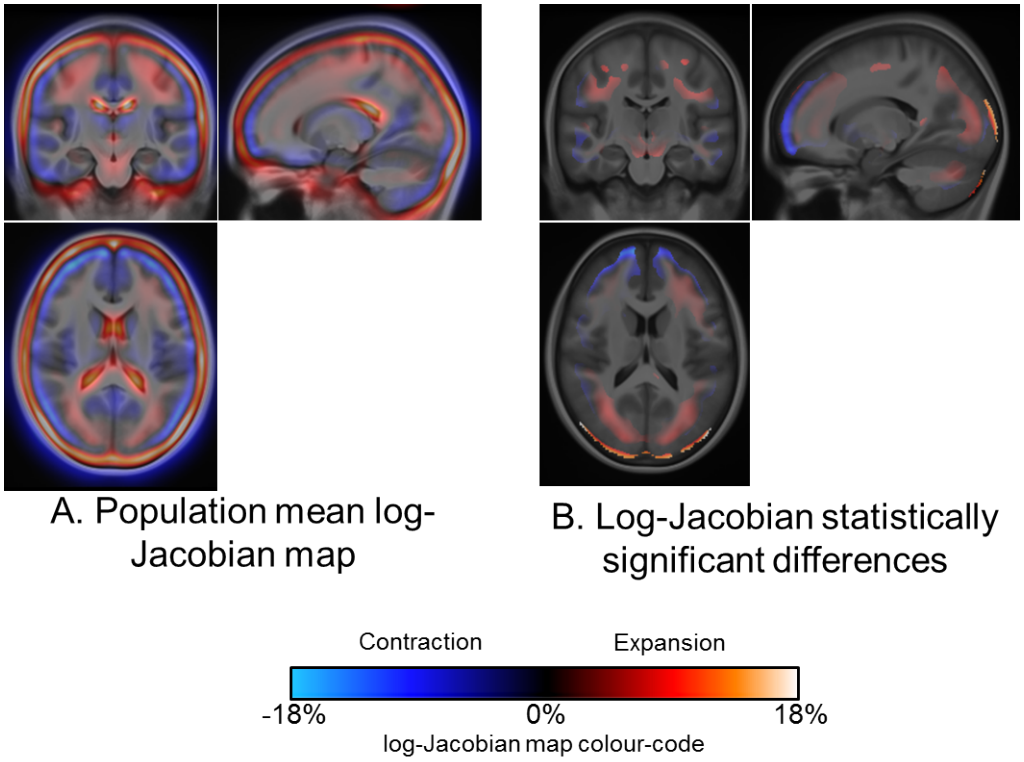

Figure 5.2: **Left:** *Log-Jacobian map for the mean longitudinal evolution.* We see that the lateral ventricles are expanding as well as the white matter and the hippocampus. An atrophy can be found in the cortical gray matter, in the thalamus, and in the putamen. **Right:** *Significant differences in volume changes between males and females.* We see that the volume atrophy is significantly more important for the females than for the males in the pre-frontal cortex and in the temporal lobes. The volume expansion is significantly more important for the males than for the females in the white matter (frontal, occipital and parietal lobes).

5.4 Beyond volumetry to study the effects of sexual dimorphism during adolescence

Now that we have compared our results to the existing literature on volumetric changes, we go beyond volumetry. We work on the three-dimensional longitudinal deformation field instead of focusing on the volume changes. We more specifically consider the Stationary Velocity Field (SVF) that parametrises the deformation field.

5.4.1 Multivariate comparison of female and male longitudinal SVF trajectories

One way to study the sexual dimorphism during adolescence is to perform a multivariate two-sample Hotelling's T^2 test on the three-dimensional subjects' longitudinal SVF trajectories. The null hypothesis is that there exists no difference in the mean SVF trajectory of the two groups. The test was corrected for multiple comparisons using 5000 permutations. The result can be found on Figure 5.3.

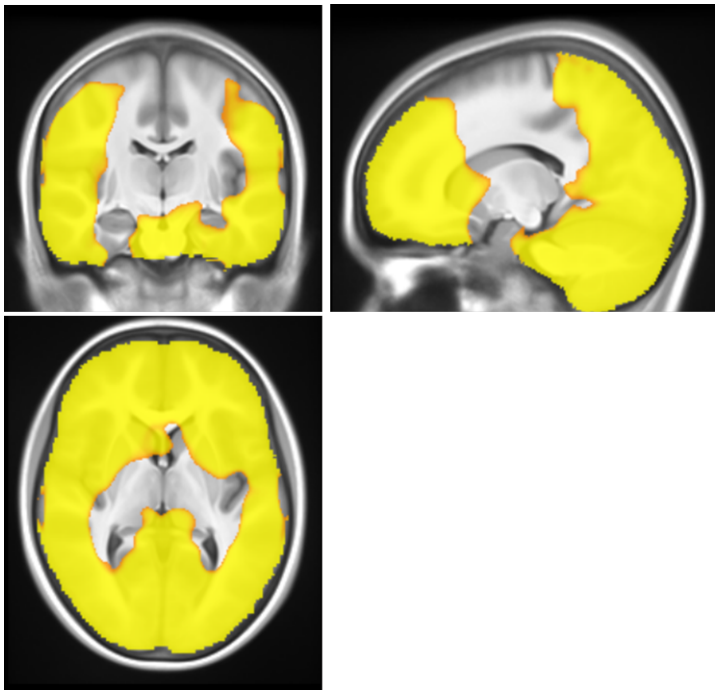

Figure 5.3: *Two-sample Hotelling's T^2 test on the transported 3 dimensional subject longitudinal trajectories*: The null hypothesis is that there exist no difference in the mean SVF trajectory of the two groups. The test was corrected for multiple comparison using 5000 permutations. The main differences are located on the cortex (except the parietal one).

As we can see, there exist many regions where the longitudinal SVF trajectories

of the two groups (males vs. females) are significantly different. The main differences are located on the cortex (except the parietal one). Although informative, the results are not easily understandable. Indeed, it is not possible to quantify the differences in an intelligible way. This might partly explain why studies generally prefer to focus on volume.

5.4.2 Sexual differences in the cross-sectional and longitudinal evolution between 14 and 16 years old

As we saw in section 5.4.1, interpreting the results of a multivariate test is generally not straightforward or easily understandable. In this section we propose to apply the method developed in chapter 4 in order to ease the understanding of sexual dimorphism during adolescence. The aim of the method is to disentangle the male/female cross-sectional differences from the population longitudinal evolution. It will then be possible to quantify these differences in terms of advance/delay with respect to the population longitudinal evolution at each time point, as well as the convergence/divergence of the 2 groups longitudinal evolutions.

Starting from the population-specific template at t_0 and the longitudinal population SVF trajectory, we compute the inter-group SVFs that best discriminate the females and the males at t_0 and t_1. We also compute the male longitudinal trajectory as well as the female longitudinal trajectory.

Convergence/Divergence of the Female and Male longitudinal SVF trajectories

We first study the convergence (or divergence) of the two groups longitudinal evolutions. If the group evolutions are diverging from each other this means that the male-female differences are increasing with time. Conversely, if the trajectories are converging, the difference between males and females are disappearing with time. This information is summarised by an angle value that quantifies the amount of convergence (negative values) or divergence (positive value). If the group difference have a steady evolution the angle is zero. The results can be seen on Figure 5.4.

We can see that the male-female differences are remaining constant in an important of the brain. Elsewhere, most of the regions are diverging between 14 and 16 years. This is consistent with the fact that the adolescence is a major period of differentiation between males and females. The only regions where convergence occurs is in the white matter (for the left hemisphere) and in a part of the cerebellum. To better understand the patterns of divergence and convergence, it is useful to focus on the male-female cross-sectional differences at t_0 and t_1. To do so, we propose to disentangle the cross-sectional differences from the longitudinal evolution.

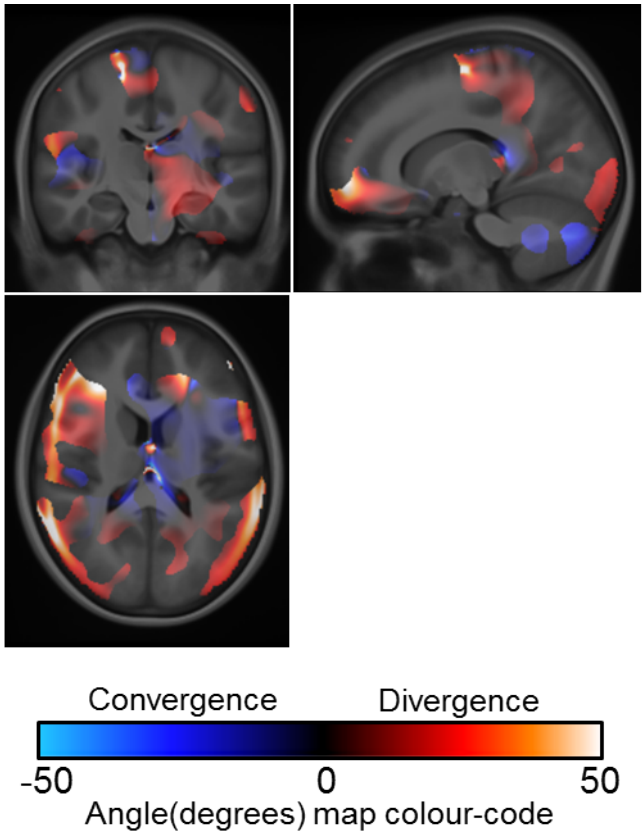

Figure 5.4: *Divergence/convergence of the group longitudinal evolutions*: The angles are displayed only for voxels where a statistically significant difference between females and males exists (this explains the artificial borders that can be seen). We can see that the male-female differences are remaining constant in an important of the brain between 14 and 16. Elsewhere, most of the regions are diverging and there is convergence only in a part of the white matter and in a part of the cerebellum.

Disentangling cross-sectional male-female differences from the longitudinal evolution

We now propose to study the cross-sectional group differences and quantify them with respect to the longitudinal evolution. The results for the females can be seen on Figure 5.5. One must note that the results for the males are not represented here since they are equal to the opposite of the females results. Therefore if the girls are in advance of two months with respect to the population longitudinal mean, then boys have a delay of two months.

At 14 years old, few structural differences exist between males and females; their location is mostly in the left hemisphere. After two years most of the regions are different. The region of statistically significant male-female difference nearly doubled (increased by 184%) over the two years.

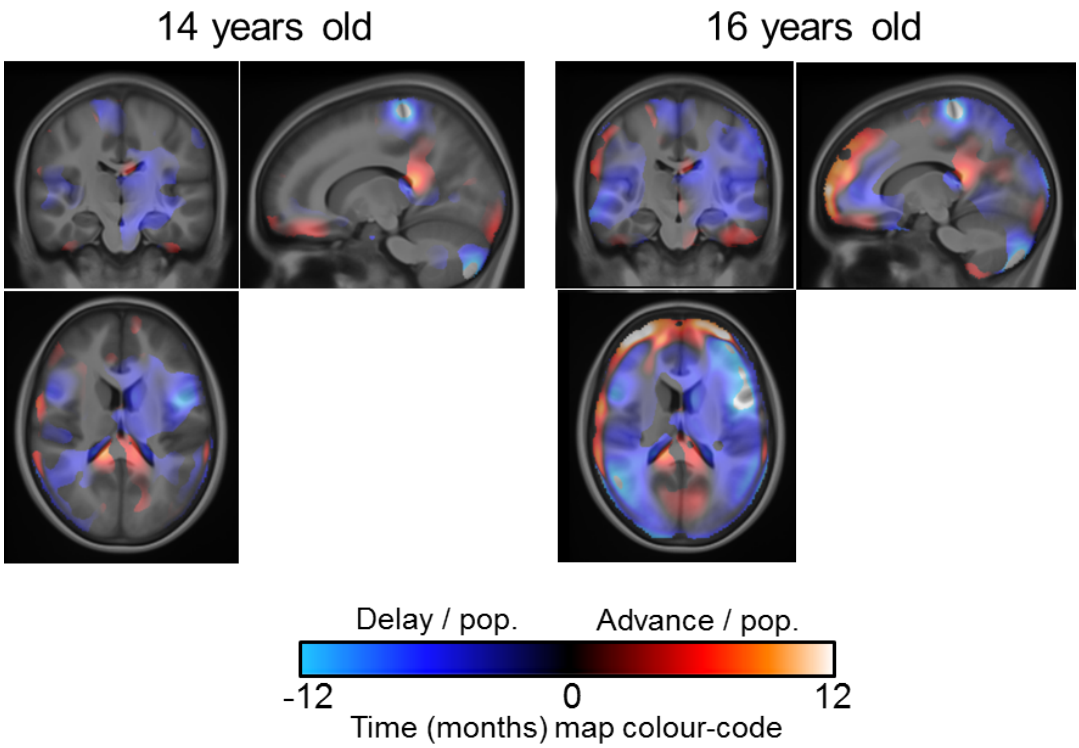

Figure 5.5: *Quantification (in months) of the group differences for females at 14 (left) and 16 (right) in terms of advance/delay with respect to the population longitudinal evolution*: We see that at 14 years old, few structural differences exist between males and females; mostly located in the left hemisphere. After two years most of the regions are different. At 14, females are in advance with respect to the longitudinal mean trajectory in very small parts of the cortex. On average, the female pre-frontal cortex has no advance. However, it develops with time and at 16 years of age, females are in advance on the frontal cortex (on average nearly five months). Conversely, females are in delay with respect to the mean in the white matter, in the ventricles, in the left thalamus, and in the left hippocampus.

More specifically, at 14 years of age, females are in advance with respect to the longitudinal mean trajectory in very small parts of the cortex. On average, the female pre-frontal cortex has no advance which means that there is no difference between males and females cortex at 14. However, the differentiation occurs during the following years, and at 16 years of age, females are in advance on the frontal cortex (on average nearly five months). Conversely, females are in delay with respect to the mean in the white matter (less than three months), in the ventricles (2.5 months), in the left thalamus (less than four months), and in the left hippocampus (4.5 months).

5.5 Conclusion and Discussion

In this chapter, we first analysed the effects of the sexual dimorphism during adolescence using volumetric indices. The volumetric results were mainly consistent with the existing literature. We found an exception in the white matter. In this region the volume was decreasing between 14 and 16 for the females whereas the literature reports an expansion of the white matter. Our finding was replicated using `SPM12` longitudinal registration tool. As adolescence involves very subtle brain changes, we notice that we are close to the detection limit of the algorithms. This might explain why the different volumetric methods might not agree. Moreover, we found a statistical difference between males and females in all the regions of interest except in the lateral ventricles. Using the more local log-Jacobian map we found that only the pre-frontal cortex and parts of the white matter had significantly different volume evolutions between males and females. The pre-frontal cortex volume decrease is more important for females than for males.

Then we went beyond volumetry by performing the group comparison on the three-dimensional deformation fields' parameters. We showed that most of the brain regions diverge during adolescence between males and females. This means that the male-female differences increase with time in most of the brain. We then quantified the cross-sectional group differences and show that the pre-frontal cortex is the region where the most important differentiation occurs during 14 and 16. We find that at 14 years of age, males and females have no difference at the level of the pre-frontal cortex. An important differentiation occurs during the two following years: at 16, females' pre-frontal cortex is in advance of around five months with respect to the mean population longitudinal evolution.

The next step would be to use the German database to check if we can replicate the findings. Moreover, as the beginning of adolescence is marked by puberty it could be interesting to incorporate the pubertal stage in the study. More specifically, we could control for the pubertal stage in the study, to verify if this has an impact on the changes we found among females and males. Finally, all the second follow-ups (at 18 years old) have been acquired for the IMAGEN database. Using the methods already performed in this chapter to analyse the third time point would enable us to see if the trajectory remains stable or changes.

Appendix: Comparison of the volumetric results given by four methods

Volumetric methods comparison

In this section, we show the results for the comparison between the `LCC log-Demons` and 3 popular algorithms (one segmentation-based method, `FreeSurfer`, and two registration-based methods: `ANTs` and `SPM12`). We can see that depending on the regions, the methods do not agree.

Volume changes for females with 4 different methods

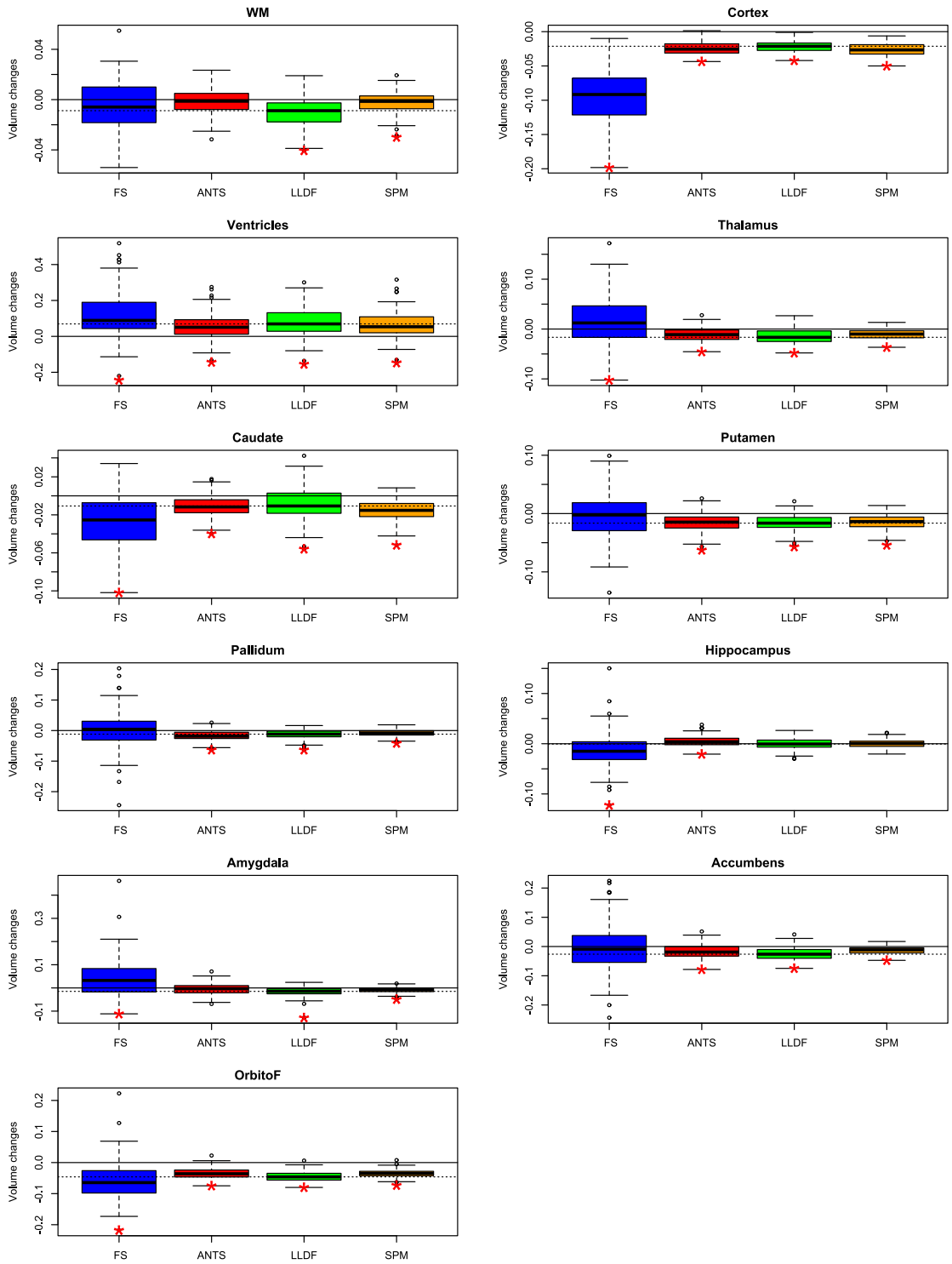

Figure 5.6: *Comparison of four volumetric methods for the females longitudinal ROI evolutions* Acronyms used: FS=FreeSurfer, LLDF=LCC log-Demons, WM=White matter, OrbitoF=Orbito-frontal cortex

Volume changes for males with 4 different methods

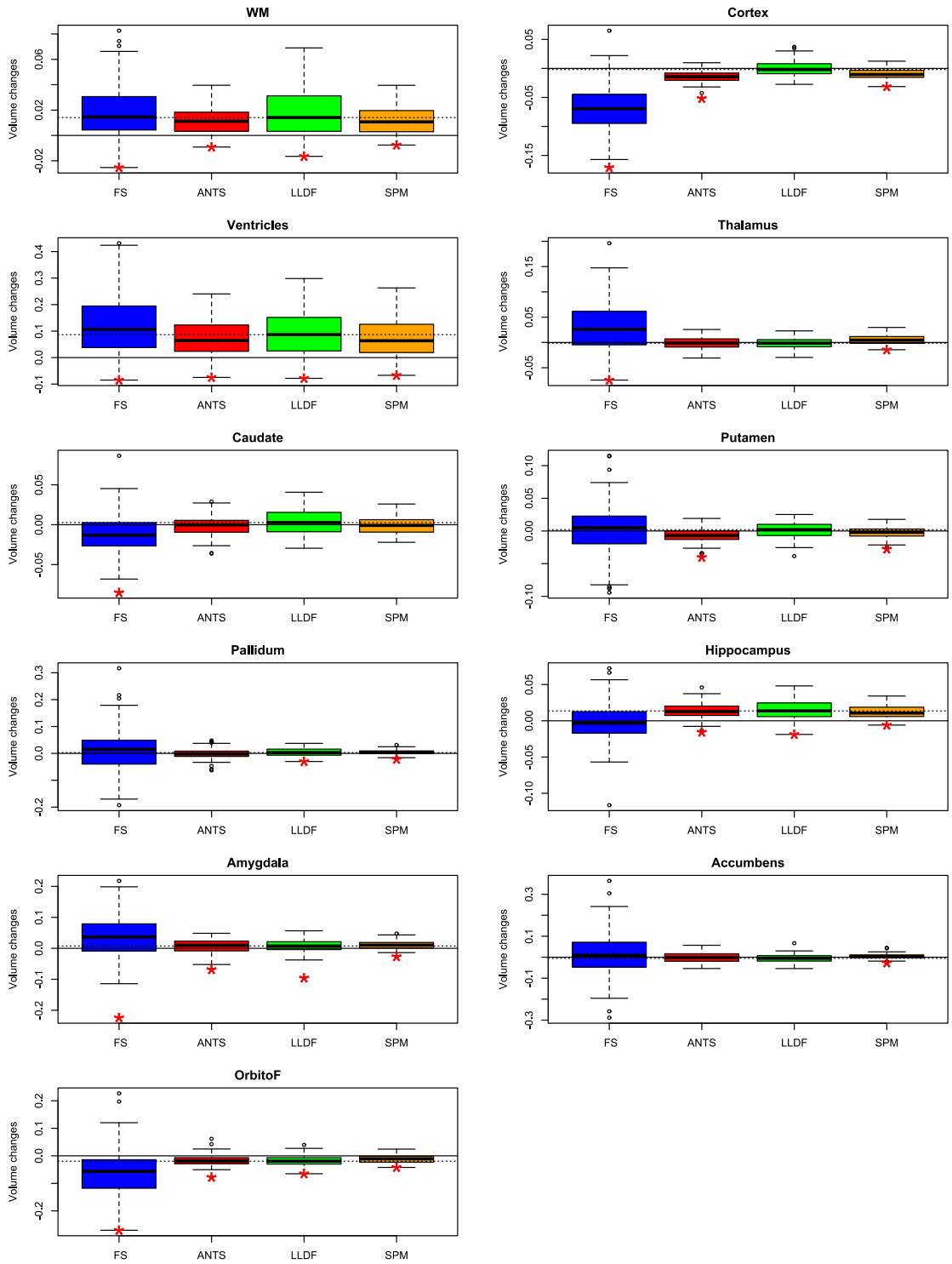

Figure 5.7: *Comparison of four volumetric methods for the males longitudinal ROI evolutions* Acronyms used: FS=FreeSurfer, LLDF=LCC log-Demons, WM=White matter, OrbitoF=Orbito-frontal cortex

Chapter 6

Conclusion and Perspectives

Contents

6.1 Main achievements **94**

6.2 Perspectives **96**

6.2.1 Automatic quality control of the processing pipeline results . 96

6.2.2 Mental disorders during adolescence 96

6.2.3 Using continuous classification of subjects along the inter-group axis 97

6.2.4 Understanding longitudinal evolutions with a dictionary of transformations 98

Each chapter of this manuscript contains conclusions on the performed work and perspectives. We here conclude on the main achievements of this Ph.D work. We then suggest perspectives and improvements to the work for the future.

6.1 Main achievements

Studying the brain structural evolution during adolescence poses several problems as numerous biases need to be avoided when capturing longitudinal evolutions. Moreover, when the intra-subject changes are very small, it is crucial to know if the available methods can capture the longitudinal evolutions with no bias. In most of the studies, these longitudinal changes are limited to scalar volumetric changes in order to ease their analysis. However, one can observe that the brain evolution is not limited to volumetry, and in this multivariate case, the interpretation is therefore more difficult. In this thesis we addressed the above problems in the context of the longitudinal study of the structural changes during adolescence.

In **chapter 2**, we proposed a deformation-based morphometry computational framework to robustly estimate the longitudinal brain deformations from image data series. In addition to limiting the potential processing biases, we showed that this processing pipeline leads to an increased sensitivity of the statistical study of the longitudinal deformations. This framework has thus a great potential for the analysis of large longitudinal datasets. The proposed processing pipeline is fully reproducible and is used in [Khanal 2016a]. We are preparing a visual implementation of the processing pipeline using the dtkComposer. The pipeline will be proposed as a package included in MedInria and the graphical form of the pipeline (cf. preview

on Figure 6.1) will facilitate its use by anyone. Indeed, this should prevents users from incurring in any mistakes in the sequencing of the steps, since everything will be readily available. This package should be freely available by the end of the year.

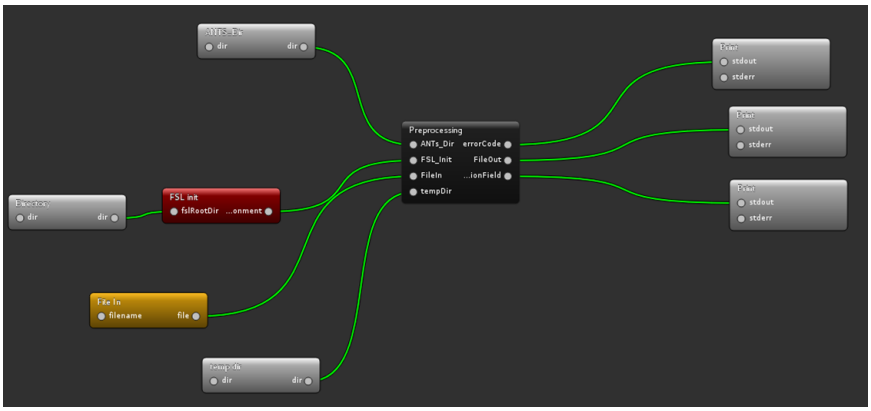

Figure 6.1: *Preview of the graphical Longitudinal Log-Demons Framework.* The processing pipeline will be be proposed as a package included in MedInria and will be based on dtkComposer.

In **chapter 3**, we evaluated the accuracy, reproducibility, detection limit, and statistical power of four popular segmentation-based and registration-based method in terms of volumetry. We found interesting results for the community: although registration-based methods generally performed better than segmentation-based methods on all the aspects cited above, the volumetric results from the two types of methods are in relative agreement. These findings should help raise awareness in the community about the range of use of the different algorithms. Moreover, we showed that the log-Jacobian integration is equivalent to relative volume changes for changes less than 10 %. This means that the log-Jacobian integration could be used as a volumetry index in most parts of the brain especially for healthy evolutions.

In **chapter 4**, we proposed to go beyond volumetry in the statistical analysis of two groups by studying the whole three-dimensional deformation field, while disentangling the group differences from the longitudinal population evolution. We thus developed two intelligible indices. The first one characterises the convergence or divergence of the group longitudinal evolutions, while the second index quantifies the group difference with respect to the population longitudinal evolution using an advance or delay (in units of time). We showed that the method eases the interpretation of the differences and therefore might be useful for clinicians.

In **chapter 5**, we used the methods developed in chapters 2 and 4, as well as the results from chapter 3, to study the effect of sexual dimorphism on the healthy morphological evolution of the brain during adolescence. By going beyond volumetry, we complemented the literature on sexual dimorphism. This work is a cross-collaboration with the INSERM-CEA U1000 team, specialised in psychiatry. During this thesis, we explained and shared with them our methodological knowledge on longitudinal image analysis. We are currently training a Ph.D student from the

U1000 team to use the Longitudinal Log-Demons Framework proposed in chapter 2.

6.2 Perspectives

6.2.1 Automatic quality control of the processing pipeline results

An important topic linked to the processing pipeline of chapter 2 and not addressed in this thesis is quality control. As databases are getting bigger, it is no longer feasible to manually check that every image has been processed properly. It is thus important to develop automated quality control methods. Such methods have been developed in the context of the Centre d'Acquisition et de Traitement des Images (CATI) [Mangin 2016] for study concerning Alzheimer's disease [Colliot 2008]. For our specific pipeline, instead of performing a quality control after each step, we believe that a good strategy is to identify a few key steps after which it is important to run the quality control. In order for the control to be relevant, it is important to define what is expected as a "good" output by opposition to an outlier. In the case of our Longitudinal Log-Demons Framework, a key step is after the non-linear registration. We could use an outlier detection algorithm on the computed SVFs. A simple and naive solution could be to perform a principal component analysis on the set of SVFs (for the whole set of subjects) to compute the three first eigen-SVFs. Then we could identify the potential outliers as the most distant SVFs from these eigen-SVFs.

6.2.2 Mental disorders during adolescence

As we deepen our knowledge of healthy brain development during adolescence, the next step is to develop a detailed understanding of the effects a disease could have on the healthy brain evolution. It would therefore be interesting to focus on psychological disorders. A part of the subjects in the IMAGEN study have developed a mental disorder, and a neuropathological score has been assessed for all subjects. We could therefore correlate this score to the subject longitudinal trajectory in order to model a trajectory for each psychological disorder and compare it to the estimated healthy trajectory. This would enable clinicians to detect earlier which subjects are potentially at risk. In practice, they would scan a subject at two timepoints and then compare his longitudinal trajectory to the reference (healthy) trajectory. This would necessitate a robust way to quantify the distance between the subject trajectory and the reference trajectory. This application has the potential to prevent teenagers from developing severe forms of depression that can lead to suicide, one of the leading cause of death among adolescents.

6.2.3 Using continuous classification of subjects along the inter-group axis

In chapter 4 we proposed a geometric configuration that consisted in a representation of both the inter-group differences and the longitudinal differences. An interesting application consists in focusing only on the inter-group differences and use it as a reference spectrum to classify subjects. Let us consider the case of the male-female axis from chapter 5. The idea would be to classify any subject using this reference axis: by projecting the subject brain on the axis, we would obtain a scalar λ (cf. Figure 6.2).

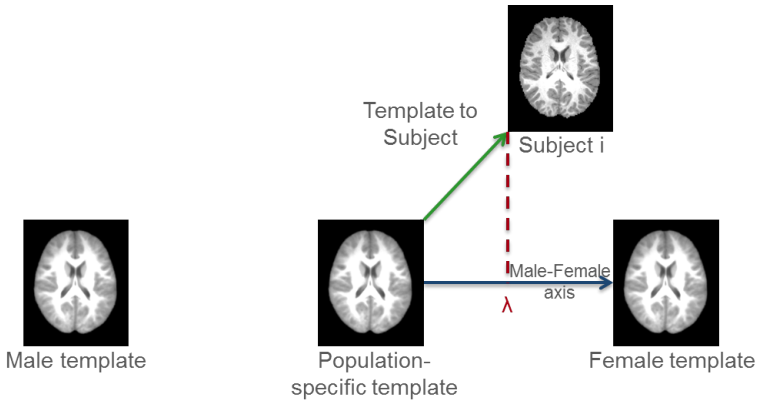

Figure 6.2: *Projection of a subject brain on the Female-Male axis* From a binary variable (sex) we obtain a continuous classification with λ.

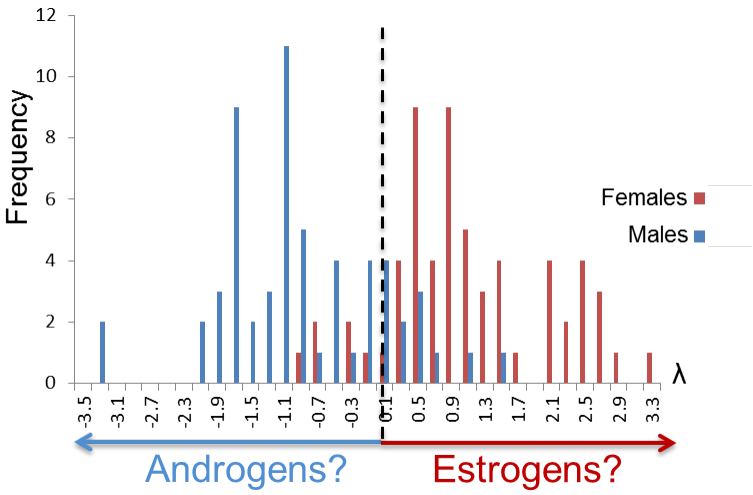

Figure 6.3: *Projection of the 120 IMAGEN subject brains (at 14) on the Female-Male axis.* We can see some of the male subjects (in blue) are classified on the female part of the axis. This could mean a higher rate of estrogens (female hormones).

This scalar would quantify the amount of "femaleness" or "maleness" of the brain. Indeed starting with a binary variable (here the sex), it is possible to obtain a continuous classification. In the case of a projection on the male-female axis, λ could be correlated with the rate of male or female hormones as shown on Figure 6.3 . Therefore, this method could be useful for clinicians to define the archetype of the male brain and female brain.

6.2.4 Understanding longitudinal evolutions with a dictionary of transformations

As we saw throughout this thesis, volumetric measurements can be relatively trusted and are easily understandable. This is still not the case for multivariate analysis of deformations, for which the interpretation still remains to be eased.

A dictionary of meaningful transformations could be defined and used to have a meaningful interpretation of the evolutions. For example, the deformation could be described as a combination of four elementary transformations: scale and shear for the volume changes, together with translation and rotation. The spatial dimensionality reduction of the deformation could for example be performed using affine projections such as the one performed with the polyaffine log-Demons [Seiler 2012] (the affine decomposition is defined in Appendix 4.5).

Appendix A

Towards a Validation of Interpolation and Extrapolation of Longitudinal Deformations Based on Stationary Velocity Fields

Contents

A.1 Introduction **100**

A.2 Measures for the validation of SVF-based geodesic regression. **101**

A.2.1 Metrics for similarity of the modeled image evolution. 102

A.2.2 Metrics for the similarity of the modeled volume changes. . . 102

A.3 Experimental data **102**

A.3.1 OASIS database 102

A.3.2 Image processing. 103

A.3.3 Group-wise analysis. 103

A.4 Results **103**

A.4.1 Interpolation of longitudinal trajectories 103

A.4.2 Extrapolation of longitudinal trajectories 105

A.5 Conclusion **106**

Analyzing the progression of morphological changes in the brain is an important topic in medical imaging. Diffeomorphic non-linear registration is a promising tool for modeling longitudinal changes, observed in T1 magnetic resonance (MR) images, as geodesic trajectories. In particular, diffeomorphic registration parametrized by Stationary Velocity Fields (SVF) has been applied to the modeling of longitudinal changes in Alzheimer's disease. However, the validity of these modeling assumptions to faithfully describe the observed anatomical evolution needs to be further investigated. In this work, we analyze the accuracy of linear geodesic regression of SVFs to describe anatomical deformations estimated from past and future observations of the MR images. The evaluation is performed by local and regional analysis of the longitudinal changes of the modeled images, and of the measured volume changes. The first experimental results show that trajectories generated by geodesic regression are compatible with those obtained by longitudinal registration of the follow-up images.

In particular, the volume changes encoded by geodesic regression are not statistically different from those measured by non-linear registration.
The results for this study are partial and the indices used are not sufficient to be able to conclude about the validation of interpolation and extrapolation based on SVFs. Hence, we discuss the necessary improvements and measures to be performed in order to conclude.

A.1 Introduction

An important topic in medical imaging is to analyze the progression of morphological changes in organs in order to model and quantify biological processes like development or disease [Goddings 2014] [Scahill 2003]. For example, longitudinal image analysis of brain changes in Alzheimer's disease [Lorenzi 2011] aims at understanding the pathological evolution in patients for clinical and diagnosis purposes.

Trajectories of longitudinal morphological changes can be measured by non-linear registration of follow-up T1 magnetic resonance (MR) images of a given subject. Among the most popular registration algorithms, we can distinguish the ones based on the large deformations paradigm where the deformation is a diffeomorphism parametrized by tangent velocity fields [Beg 2005, Ashburner 2007, Vercauteren 2008] and in particular the Large Deformation Diffeomorphic Metric Mapping (LDDMM)[Trouvé 1998] and the Stationary Velocity Field (SVF) framework [Arsigny 2006]. In LDDMM, geodesics are minimizing a suitable Riemannian distance, while in the SVF setting geodesics are the straight lines of the Cartan connection. In both settings, geodesics are parametrized by their initial tangent vector: initial momentum for LDDMM and the Stationary Velocity Field for SVF.

Based on this registration paradigm, geodesics regression on images has been performed either for predicting intermediate observations on sequences of images (*Interpolation*) [Davis 2010, Niethammer 2011a, Fletcher 2013], or for modeling trajectories beyond the current observation interval (*Extrapolation*) [Lorenzi 2011]. By performing geodesic regression as in [Hadj-Hamou 2016], longitudinal changes are identified by deformations parametrized by SVFs. The authors proposed to estimate a series of SVFs by non-linear registration of follow-up images to the baseline. They then proposed to model the "*deformation trajectory*" associated to the resulting series of deformations by a linear model in time of the resulting SVF. The methodological motivation relies on the tangent representation of SVFs which enables to describe sequences of diffeomorphisms by more tractable linear modeling of SVF in the tangent space.
The statistical analysis of the deformation trajectories provided meaningful description of the disease progression. However, the validity of such a linear assumption in reliably describing the observed anatomical evolution still needs to be evaluated. A comparison is thus needed between the trajectory generated by geodesic regression with the one described by longitudinal registration of follow-up images. The aim of this work is to study the accuracy of linear geodesic regression of SVF to describe

anatomical deformations estimated from past and future observations i.e. to quantify to error made by regression with respect to registration. In this study, we focus on the modeling of longitudinal trajectories in Alzheimer's disease since it is one of the prominent fields of application of registration.

In section A.2, we develop a new methodology and metrics to study the effectiveness of the linear geodesic regression. We then present the experimental setting in section A.3. Finally, in section A.4, we show that SVF-based models are able to accurately describe the trajectories estimated by non-linear registration, for both interpolation and extrapolation.

A.2 Measures for the validation of SVF-based geodesic regression.

We consider longitudinal observations of T1-weighted MRI scans for a given subject, at the 2 time points t_0 and t_1 (cf. Figure A.1). The corresponding images will be described as I_0 and I_1 respectively. By non-linearly registering I_0 to I_1, we estimate the SVF $\widetilde{v}$ that maximizes the similarity between I_1 and $I_0 \circ \exp(\widetilde{v})$.
Let $\alpha \in \mathbb{R}$ be the time factor and T the subject age at which we want to realize the prediction: $T = t_0 + \alpha(t_1 - t_0)$. According to the linear modeling assumption for the evolution of the SVFs, the predicted SVF $v(\alpha)$ is $v(\alpha) = \alpha \cdot \widetilde{v}$. We can thus define 2 predictive models depending on the values of α. For $0 < \alpha < 1$, we realize an interpolation, while for $\alpha < 0$ or $\alpha > 1$, we realize an extrapolation (in the past or in the future).

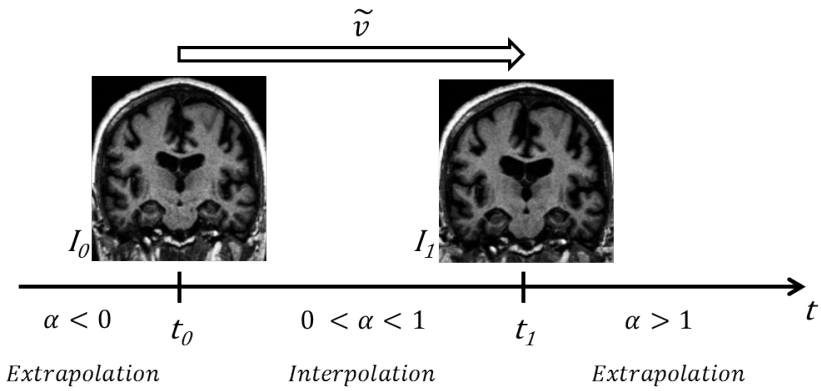

Figure A.1: Prediction of longitudinal trajectories in the case of 2 time points t_0 and t_1, based on the measured SVF $\widetilde{v}$. 2 different cases of prediction (interpolation or extrapolation) depending on the values of the time factor α.

In order to evaluate the effectiveness of the prediction, we use a third time point as the reference for what we are predicting. We will thus compare the predicted trajectory to the measured one, at the same time point. We propose 2 simple metrics for this evaluation.

A.2.1 Metrics for similarity of the modeled image evolution.

The first metric we propose concerns the matching between the predicted image $I^{Pred} = I_0 \circ \exp(v(\alpha))$ and the optimally registered image $I^{Meas} = I_0 \circ \exp(\widetilde{v})$. Among the numerous metrics that exist, we choose the Sum of Squared Differences (SSD). This make sense since we are comparing two deformations of the same image. However, we verified that we obtain the same conclusions with the Local Correlation Criteria (LCC).
Since the baseline image I_0 of different subjects may not have the same scale of intensities, in addition to an inhomogenity bias correction (pre-processing, ANTs - N4BiasFieldCorrection [Avants 2011, Tustison 2010]) we normalize the SSD by the maximum intensity of the baseline image I_0. We then build the SSD map as:

$$SSD = \frac{(I^{Pred} - I^{Meas})^2}{(max(I_0))^2}$$

Regional analysis was also performed by computing the average SSD in 3 different areas: the whole brain, the ventricles and the hippocampi. The 2 latter zones are known to be the regions where the most dramatic changes occur during aging and Alzheimer's disease.

A.2.2 Metrics for the similarity of the modeled volume changes.

The second metric concerns the transformation itself. One of the main goal of longitudinal studies on Alzheimer's disease is the estimation of brain atrophy. Therefore, we study the differences (lJD) in the log-Jacobian (lJ) associated to the predicted evolution $v(\alpha)$ and the measured one $\widetilde{v}$: $lJD = lJ^{Pred} - lJ^{Meas}$.
As for the SSD metric, we compute the average lJD in the 3 same anatomical regions previously defined.

A.3 Experimental data

A.3.1 OASIS database

For this study we use the longitudinal OASIS (Open Access Series of Imaging Studies) database [Marcus 2010] since it contains healthy evolutions but also demented evolutions for which the prediction would be helpful. Moreover, this database is Open Access which facilitates the reprodcution and compariosn of the rults. Our working set consists of N=40 subjects aged 60 to 92 for which 3 acquisitions were available: 24 subjects are non-demented, while 16 have dementia. For the time points t_0, t_1 and t_2, we tested the interpolation from $[t_0, t_2]$ to t_1, and the extrapolation from $[t_0, t_1]$ to t_2. Each subject has a different time factor α that varies from 0.2 to 0.83 for the interpolation, and from 1.2 to 5 for the extrapolation.

A.3.2 Image processing.

Every subject image time series underwent bias field correction [Tustison 2010], followed by linear alignment to the MNI reference space [Fonov 2009]. For each subject, the follow-up images were independently rigidly aligned to the baseline. Both affine and rigid transformations were performed with FSL Flirt [Jenkinson 2012].
Then, pairwise non-rigid registration between follow-up images and baseline was performed with the LCC-LogDemons algorithm [Lorenzi 2013a]. A preliminary study was realized on a set of 10 images - which were then discarded - to find the best parameters for the registration algorithm ($\sigma_{elastic} = 1.5$, $\sigma_{fluid} = 0.5$, $\sigma_{LCC} = 3$).

A.3.3 Group-wise analysis.

Longitudinal analysis was separately performed on healthy and demented subjects. We built group-wise maps of SSD and lJD (using the anatomical template of healthy elderly population defined in [Lorenzi 2011]). Statistical differences between regional predicted and measured intensities and log-Jacobian were assessed by paired t-tests.

A.4 Results

A.4.1 Interpolation of longitudinal trajectories

The results for the interpolation are visible on Figure A.2 and A.4. We see that for the healthy and the AD groups, the interpolated and the measured images are very similar, in terms of intensities, in most of the brain regions.
Differences are located in the ventricles (cf. Fig. A.3 A.). In particular for the healthy group, the images are statistically different ($p < 0.05$, paired t-test) in the ventricles. This is not the case for the Alzheimer's group.
For the 2 groups, we observe few differences between the estimated volume changes and the measured ones, reinforced by the fact that these differences are not statically significant ($p > 0.05$, paired t-test) in every region of the brain.

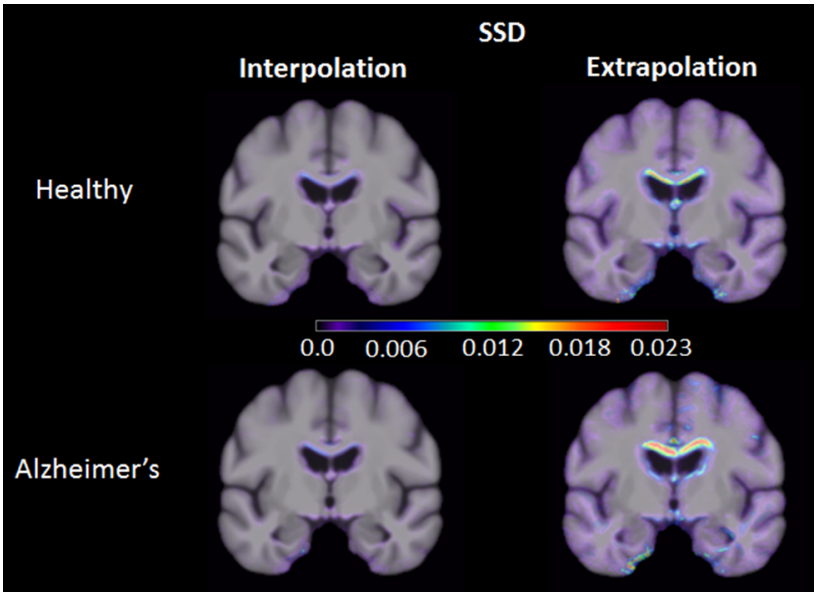

Figure A.2: *SSD*: Group-wise mean SSD map for the interpolation and the extrapolation. For both groups, the interpolated and extrapolated images are very similar to the measured ones in terms of intensity. The main dissimilarities can be seen in the ventricles with a higher dissimilarity for extrapolation.

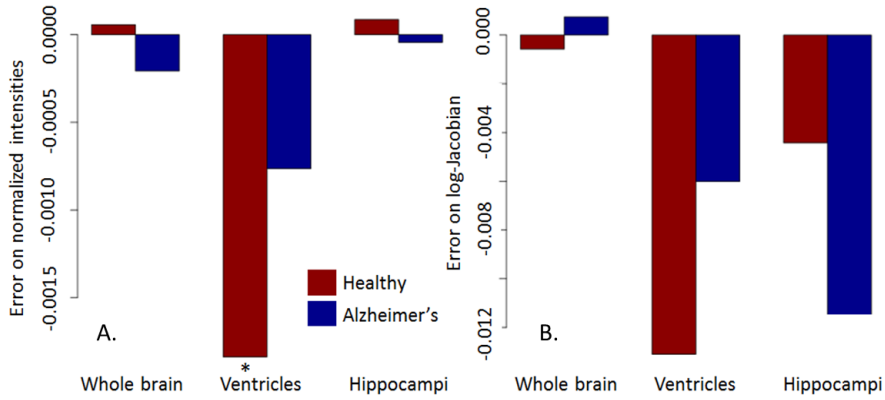

Figure A.3: *Interpolation*: Group-wise mean values of the error on the intensities (A.) and log-Jacobian (B.) for both interpolated and measured images, I^{Pred} and I^{Meas}, in 3 different areas: whole brain, ventricles and hippocampi. (*: indicates that there is a significant difference ($p < 0.05$, paired t-test)). I^{Pred} and I^{Meas} show no significant difference except for the intensity in the ventricles for the healthy group.

A.4.2 Extrapolation of longitudinal trajectories

Concerning the extrapolation, for the healthy and the Alzheimer's groups, the interpolated and the measured images are very similar, in terms of intensities (Fig. A.2), in most of the brain regions.
Statistical differences ($p < 0.05$, paired t-test) are located in the ventricles (cf. Fig. A.5 A.). This area is characterized by a shift of the cerebrospinal fluid (CSF) and white matter (WM) interface. The observed very high SSD is due to the high contrast between the low intensities of the CSF and the high intensities in the white matter. In order to quantify this shift we study the displacement field orthogonal to the ventricular border. The average error on the displacements in this zone is 0.257 mm which is less than 0.3 voxel. This might however be a negligible difference when compared to the usual large ventricular expansion reported in longitudinal studies, even for healthy subjects [Lehéricy 2007].
Concerning the map of the log-Jacobian differences between the extrapolated and the log-jacobian obtained from non-linear registration, we observe that large differences in the estimated volume changes are visible in the ventricles, temporal areas and the frontal cortex. Even though this might be indicative of an acceleration process of brain atrophy, we note that the regional quantification of atrophy did not show any significant difference ($p > 0.05$, paired t-test) between the average log-Jacobian values (cf. Fig. A.5 B.).

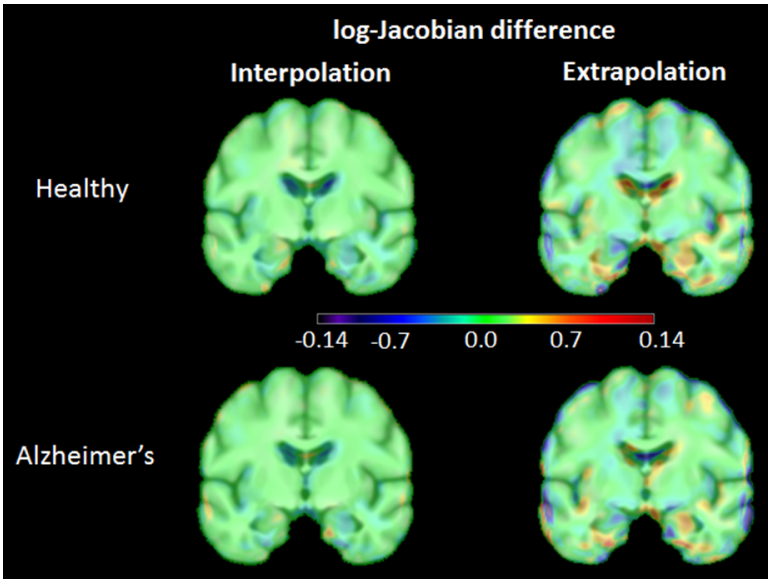

Figure A.4: *lJD*: Group-wise mean lJD map for the interpolation and the extrapolation. For both groups, the interpolated images are very similar to the measured ones in terms log-jacobian. The extrapolated images are different in the ventricles, the temporal and frontal areas.

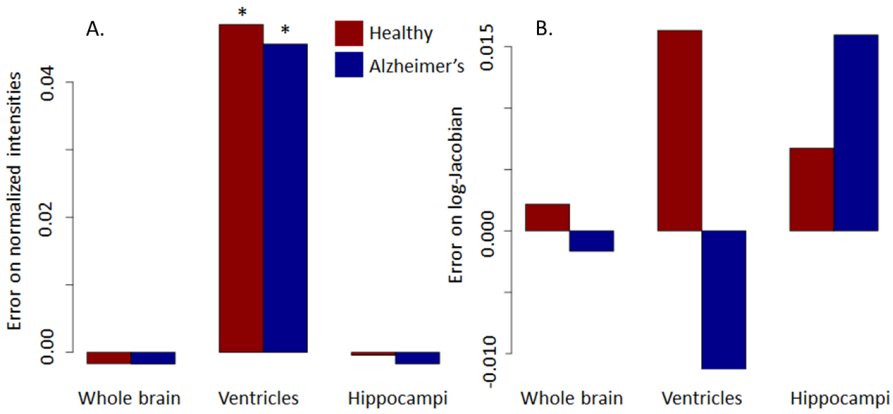

Figure A.5: *Extrapolation*: Group-wise mean values of the error on the intensities (A.) and log-Jacobian (B.) for both interpolated and measured images, I^{Pred} and I^{Meas}, in 3 different areas: whole brain, ventricles and hippocampi. (*: indicates that there is a significant difference ($p < 0.05$, paired t-test)). I^{Pred} and I^{Meas} show no significant difference except for the intensity in the ventricles for both groups.

A.5 Conclusion

We studied the accuracy of linear geodesic regression of SVF in describing anatomical deformations estimated from past and future observations of brain MR. We showed that trajectories generated by geodesic regression (interpolation or extrapolation) are compatible with those obtained by longitudinal registration of the follow-up images. In particular, we observed that the volume changes encoded by geodesic regression are not statistically different from those measured by non-linear registration. This result has important implications for the development and use of SVF-based models of brain atrophy. An exception was found for the ventricles, which showed to be characterized by an accelerated expansion not entirely captured by the model. This is indicative of the presence of more complex dynamics of brain evolution in those areas.

The results of this article are partial and further investigations should be performed to be able to give a definitive conclusion. In addition to the presented indices, segmentation-based measures could be used: the extrapolated brains and the reference brains should be segmented and overlaps indices (Dice, Jaccard...) should be used to quantify the similarity.

Moreover, to show the advantage of the SVF-based extrapolation, we could exaggerate the deformations by a factor ten and compare the results between the SVF based extrapolation and the direct multiplication of the deformation field. We could thus show that with the SVF-based extrapolation the brain still has the structure of a brain whereas with the other method the brain "explodes" (it no longer looks like a brain).

Bibliography

[Arsigny 2006] Vincent Arsigny, Olivier Commowick, Xavier Pennec and Nicholas Ayache. *A Log-Euclidean Framework for Statistics on Diffeomorphisms.* In Rasmus Larsen, Mads Nielsen and Jon Sporring, editeurs, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006, volume 4190 of *Lecture Notes in Computer Science*, pages 924–931. Springer Berlin Heidelberg, 2006. (Cited on pages 15, 16 and 100.)

[Ashburner 1998] John Ashburner, Chloe Hutton, Richard Frackowiak, Ingrid Johnsrude, Cathy Price and Karl Friston. *Identifying global anatomical differences: Deformation-based morphometry.* Human Brain Mapping, vol. 6, no. 5-6, pages 348–357, 1998. (Cited on pages 7 and 65.)

[Ashburner 2000] John Ashburner and Karl J. Friston. *Voxel-Based Morphometry-The Methods.* NeuroImage, vol. 11, no. 6, pages 805 – 821, 2000. (Cited on page 65.)

[Ashburner 2007] J. Ashburner. *A fast diffeomorphic image registration algorithm.* NeuroImage, vol. 8, no. 2, pages 135–160, 2007. (Cited on page 100.)

[Ashburner 2013] John Ashburner and Gerard R. Ridgway. *Symmetric diffeomorphic modelling of longitudinal structural MRI.* Frontiers in Neuroscience, vol. 6, no. 197, 2013. (Cited on pages 7, 25, 32 and 40.)

[Avants 2011] Brian B. Avants, Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein and James C. Gee. *A reproducible evaluation of ANTs similarity metric performance in brain image registration.* NeuroImage, vol. 54, no. 3, pages 2033 – 2044, 2011. (Cited on pages 7, 12, 32, 40 and 102.)

[Beg 2005] M.Faisal Beg, MichaelI. Miller, Alain Trouve and Laurent Younes. *Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms.* International Journal of Computer Vision, vol. 61, no. 2, pages 139–157, 2005. (Cited on pages 15 and 100.)

[Bland 1999] J Martin Bland and Douglas G Altman. *Measuring agreement in method comparison studies.* Statistical Methods in Medical Research, vol. 8, no. 2, pages 135–160, 1999. (Cited on page 46.)

[Bossa 2007] Matias Bossa, Monica Hernandez and Salvador Olmos. *Contributions to 3D diffeomorphic atlas estimation: Application to brain images.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, volume 10, pages 667–674, 2007. (Cited on page 16.)

[Braak 1991] H. Braak and E. Braak. *Alzheimer's disease affects limbic nuclei of the thalamus.* Acta Neuropathol., vol. 81, no. 3, pages 261–8, 1991. (Cited on page 25.)

[Brett 2001] Matthew Brett, Alexander P. Leff, Chris Rorden and John Ashburner. *Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking.* NeuroImage, vol. 14, no. 2, pages 486 – 500, 2001. (Cited on pages 8 and 18.)

[Cachier 2003] Pascal Cachier, Eric Bardinet, Didier Dormont, Xavier Pennec and Nicholas Ayache. *Iconic feature based nonrigid registration: the PASHA algorithm.* Computer Vision and Image Understanding, vol. 89, no. 2 - 3, pages 272 – 298, 2003. Nonrigid Image Registration. (Cited on page 16.)

[Calmon 2000] Guillaume Calmon and Neil Roberts. *Automatic measurement of changes in brain volume on consecutive 3D {MR} images by segmentation propagation.* Magnetic Resonance Imaging, vol. 18, no. 4, pages 439 – 453, 2000. (Cited on page 43.)

[Camara 2008] Oscar Camara, Julia A. Schnabel, Gerard R. Ridgway, William R. Crum, Abdel Douiri, Rachael I. Scahill, Derek L.G. Hill and Nick C. Fox. *Accuracy assessment of global and local atrophy measurement techniques with realistic simulated longitudinal Alzheimer's disease images.* NeuroImage, vol. 42, no. 2, pages 696 – 709, 2008. (Cited on page 37.)

[Cardenas 2007] Valerie A. Cardenas, Colin Studholme, Stefan Gazdzinski, Timothy C. Durazzo and Dieter J. Meyerhoff. *Deformation-based morphometry of brain changes in alcohol dependence and abstinence.* NeuroImage, vol. 34, no. 3, pages 879 – 887, 2007. (Cited on page 7.)

[Cash 2015] David M. Cash, Chris Frost, Leonardo O. Iheme and Devrim ˙*Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge.* NeuroImage, vol. 123, pages 149 – 164, 2015. (Cited on page 37.)

[Chung 2001] M.K. Chung, K.J. Worsley, T. Paus, C. Cherif, D.L. Collins, J.N. Giedd, J.L. Rapoport and A.C. Evans. *A Unified Statistical Approach to Deformation-Based Morphometry.* NeuroImage, vol. 14, no. 3, pages 595 – 606, 2001. (Cited on page 7.)

[Chupin 2009] Marie Chupin, Emilie Gérardin, Rémi Cuingnet, Claire Boutet, Louis Lemieux, Stéphane Lehéricy, Habib Benali, Line Garnero and Olivier Colliot. *Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI.* Hippocampus, vol. 19, no. 6, pages 579–587, 2009. (Cited on page 55.)

[Colliot 2008] Olivier Colliot, Gaël Chételat, Marie Chupin, Béatrice Desgranges, Benoît Magnin, Habib Benali, Bruno Dubois, Line Garnero, Francis Eustache and Stéphane Lehéricy. *Discrimination between Alzheimer Disease, Mild Cognitive Impairment, and Normal Aging by Using Automated Segmentation*

of the Hippocampus. Radiology, vol. 248, no. 1, pages 194–201, 2008. PMID: 18458242. (Cited on page 96.)

[Davatzikos 2001] Christos Davatzikos, Ahmet Genc, Dongrong Xu and Susan M. Resnick. *Voxel-Based Morphometry Using the RAVENS Maps: Methods and Validation Using Simulated Longitudinal Atrophy.* NeuroImage, vol. 14, no. 6, pages 1361 – 1369, 2001. (Cited on page 7.)

[Davis 2010] C. Davis B., T. Fletcher, E. Bullitt and S. Joshi. *Population shape regression from random design data.* NeuroImage, vol. 8, no. 2, pages 135–160, 2010. (Cited on page 100.)

[de Jong 2008] L. W. de Jong, K. van der Hiele, I. M. Veer, J. J. Houwing, R. G. Westendorp, E. L. Bollen, P. W. de Bruin, H. A. M. Middelkoop, M. A. van Buchem and J. van der Grond. *Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study.* Brain, vol. 131, no. 12, pages 3277–3285, 2008. (Cited on page 25.)

[Fletcher 2013] T. Fletcher. *Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds.* NeuroImage, vol. 8, no. 2, pages 135–160, 2013. (Cited on page 100.)

[Fonov 2009] VS Fonov, AC Evans, RC McKinstry, CR Almli and DL Collins. *Unbiased nonlinear average age-appropriate brain templates from birth to adulthood.* NeuroImage, vol. 47, Supplement 1, no. 0, page S102, 2009. Organization for Human Brain Mapping 2009 Annual Meeting. (Cited on pages 9 and 103.)

[Fox 1996] N. C. Fox, E. K. Warrington, P. A. Freeborough, P. Hartikainen, A. M. Kennedy, J. M. Stevens and M. N. Rossor. *Presymptomatic hippocampal atrophy in Alzheimer's disease.* Brain, vol. 119, no. 6, pages 2001–2007, 1996. (Cited on page 25.)

[Fox 2011] Nick C. Fox, Gerard R. Ridgway and Jonathan M. Schott. *Algorithms, atrophy and Alzheimer's disease: Cautionary tales for clinical trials.* NeuroImage, vol. 57, no. 1, pages 15 – 18, 2011. (Cited on pages 2 and 38.)

[Frackowiak 2003] R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, C.J. Price, S. Zeki, J. Ashburner and W.D. Penny. Human Brain Function. Academic Press, 2nd édition, 2003. (Cited on page 65.)

[Friston 2007] K. Friston. *Chapter 2 - Statistical parametric mapping.* In Karl Friston, John Ashburner, Stefan Kiebel, Thomas Nichols and William Penny, editeurs, Statistical Parametric Mapping, pages 10 – 31. Academic Press, London, 2007. (Cited on pages 7, 25 and 85.)

[Giedd 2006] Jay N. Giedd, Liv S. Clasen, Rhoshel Lenroot, Dede Greenstein, Gregory L. Wallace, Sarah Ordaz, Elizabeth A. Molloy, Jonathan D. Blumenthal, Julia W. Tossell, Catherine Stayer, Carole A. Samango-Sprouse, Dinggang Shen, Christos Davatzikos, Deborah Merke and George P. Chrousos. *Puberty-related influences on brain development.* Molecular and Cellular Endocrinology, vol. 254-255, pages 154 – 162, 2006. (Cited on pages 51, 79, 83, 84 and 85.)

[Giedd 2008] Jay N. Giedd, Matcheri Keshavan and Tomas Paus. *Why do many psychiatric disorders emerge during adolescence?* Nature Reviews Neuroscience, vol. 9, no. 12, pages 947 – 957, 2008. (Cited on page 79.)

[Giorgio 2013] Antonio Giorgio and Nicola De Stefano. *Clinical use of brain volumetry.* Journal of Magnetic Resonance Imaging, vol. 37, no. 1, pages 1–14, 2013. (Cited on page 37.)

[Goddings 2014] Anne-Lise Goddings, Kathryn L. Mills, Liv S. Clasen, Jay N. Giedd, Russell M. Viner and Sarah-Jayne Blakemore. *The influence of puberty on subcortical brain development.* NeuroImage, vol. 88, no. 0, pages 242 – 251, 2014. (Cited on pages 79 and 100.)

[Gorgolewski 2015] Krzysztof Jacek Gorgolewski, Gael Varoquaux, Gabriel Rivera, Yannick Schwartz, Satrajit S Ghosh, Camille Maumet, Vanessa V Sochat, Thomas E. Nichols, Russell A. Poldrack, Jean-Baptiste Poline, Tal Yarkoni and Daniel S. Margulies. *NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain.* Frontiers in Neuroinformatics, vol. 9, no. 8, 2015. (Cited on page 25.)

[Guillaume 2014] Bryan Guillaume, Xue Hua, Paul M. Thompson, Lourens Waldorp and Thomas E. Nichols. *Fast and accurate modelling of longitudinal and repeated measures neuroimaging data.* NeuroImage, vol. 94, pages 287 – 302, 2014. (Cited on page 33.)

[Guimond 2000] Alexandre Guimond, Jean Meunier and Jean-Philippe Thirion. *Average Brain Models: A Convergence Study.* Computer Vision and Image Understanding, vol. 77, no. 2, pages 192 – 210, 2000. (Cited on page 19.)

[Hadj-Hamou 2016] Mehdi Hadj-Hamou, Marco Lorenzi, Nicholas Ayache and Xavier Pennec. *Longitudinal Analysis of Image Time Series with Diffeomorphic Deformations: a Computational Framework based on Stationary Velocity Fields.* Frontiers in Neuroscience, vol. 10, no. 236, 2016. (Cited on pages 4, 40, 41, 69, 73, 76, 81 and 100.)

[Iglesias 2011] J.E. Iglesias, Cheng-Yi Liu, P.M. Thompson and Zhuowen Tu. *Robust Brain Extraction Across Datasets and Comparison With Publicly Available Methods.* Medical Imaging, IEEE Transactions on, vol. 30, no. 9, pages 1617–1634, Sept 2011. (Cited on page 12.)

[Jack 2004] C.R. Jack, M.M. Shiung, J.L. Gunter, P.C. O'Brien, S.D. Weigand, D.S. Knopman, B.F. Boeve, R.J. Ivnik, G.E. Smith, R.H. Cha, E.G. Tangalos and R.C. Petersen. *Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD.* Neurology, vol. 4, no. 62, 2004. (Cited on page 25.)

[Jack 2008] Clifford R. Jack, Matt A. Bernstein, Nick C. Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J. Britson, Jennifer L. Whitwell, Chadwick Ward, Anders M. Dale, Joel P. Felmlee, Jeffrey L. Gunter, Derek L.G. Hill, Ron Killiany, Norbert Schuff, Sabrina Fox-Bosetti, Chen Lin, Colin Studholme, Charles S. DeCarli, Gunnar Krueger, Heidi A. Ward, Gregory J. Metzger, Katherine T. Scott, Richard Mallozzi, Daniel Blezek, Joshua Levy, Josef P. Debbins, Adam S. Fleisher, Marilyn Albert, Robert Green, George Bartzokis, Gary Glover, John Mugler and Michael W. Weiner. *The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods.* Journal of Magnetic Resonance Imaging, vol. 27, no. 4, pages 685–691, 2008. (Cited on page 40.)

[Jenkinson 2001] Mark Jenkinson and Stephen Smith. *A global optimisation method for robust affine registration of brain images.* Medical Image Analysis, vol. 5, no. 2, pages 143 – 156, 2001. (Cited on page 13.)

[Jenkinson 2002] Mark Jenkinson, Peter Bannister, Michael Brady and Stephen Smith. *Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images.* NeuroImage, vol. 17, no. 2, pages 825 – 841, 2002. (Cited on page 13.)

[Jenkinson 2012] Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich and Stephen M. Smith. *FSL.* NeuroImage, vol. 62, no. 2, pages 782 – 790, 2012. (Cited on pages 9 and 103.)

[Joshi 2000] S.C. Joshi and M.I. Miller. *Landmark matching via large deformation diffeomorphisms.* Image Processing, IEEE Transactions on, vol. 9, no. 8, pages 1357–1370, Aug 2000. (Cited on page 15.)

[Karpate 2014] Yogesh Karpate, Olivier Commowick, Christian Barillot and Gilles Edan. Longitudinal intensity normalization in multiple sclerosis patients, pages 118–125. Springer International Publishing, Cham, 2014. (Cited on page 12.)

[Khanal 2016a] Bishesh Khanal, Nicholas Ayache and Xavier Pennec. Simulating Realistic Synthetic Longitudinal Brain MRIs with known Volume Changes. working paper or preprint, July 2016. (Cited on page 94.)

[Khanal 2016b] Bishesh Khanal, Marco Lorenzi, Nicholas Ayache and Xavier Pennec. *A biophysical model of brain deformation to simulate and analyze longitudinal MRIs of patients with Alzheimer's disease.* NeuroImage, vol. 134, pages 35–52, June 2016. (Cited on pages 37 and 41.)

[Klein 2009] Arno Klein, Jesper Andersson, Babak A. Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E. Christensen, D. Louis Collins, James Gee, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P. Woods, J. John Mann and Ramin V. Parsey. *Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.* NeuroImage, vol. 46, no. 3, pages 786 – 802, 2009. (Cited on pages 13, 39 and 41.)

[Lehéricy 2007] S. Lehéricy, M Marjanska, L Mesrob and etal. *Magnetic resonance imaging of Alzheimer's disease.* NeuroImage, vol. 8, no. 2, pages 135–160, 2007. (Cited on page 105.)

[Lenroot 2007] Rhoshel K. Lenroot, Nitin Gogtay, Deanna K. Greenstein, Elizabeth Molloy Wells, Gregory L. Wallace, Liv S. Clasen, Jonathan D. Blumenthal, Jason Lerch, Alex P. Zijdenbos, Alan C. Evans, Paul M. Thompson and Jay N. Giedd. *Sexual dimorphism of brain developmental trajectories during childhood and adolescence.* NeuroImage, vol. 36, no. 4, pages 1065 – 1073, 2007. (Cited on pages 79 and 83.)

[Lorenzi 2011] Marco Lorenzi, Nicholas Ayache, GiovanniB Frisoni and Xavier Pennec. *Mapping the Effects of $A\beta_{1-42}$ Levels on the Longitudinal Changes in Healthy Aging: Hierarchical Modeling Based on Stationary Velocity Fields.* In Gabor Fichtinger, Anne Martel and Terry Peters, editeurs, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2011, volume 6892 of *Lecture Notes in Computer Science*, pages 663–670. Springer Berlin Heidelberg, 2011. (Cited on pages 7, 8, 16, 100 and 103.)

[Lorenzi 2013a] M. Lorenzi, N. Ayache, G.B. Frisoni and X. Pennec. *LCC-Demons: A robust and accurate symmetric diffeomorphic registration algorithm.* NeuroImage, vol. 81, no. 0, pages 470 – 483, 2013. (Cited on pages 16, 33, 44 and 103.)

[Lorenzi 2013b] Marco Lorenzi and Xavier Pennec. *Geodesics, Parallel Transport & One-Parameter Subgroups for Diffeomorphic Image Registration.* International Journal of Computer Vision, vol. 105, no. 2, pages 111–127, 2013. (Cited on pages 7, 22 and 33.)

[Mahapatra 2012] Dwarikanath Mahapatra. *Skull Stripping of Neonatal Brain MRI: Using Prior Shape Information with Graph Cuts.* Journal of Digital Imaging, vol. 25, no. 6, pages 802–814, 2012. (Cited on page 12.)

[Mangin 2016] Jean-FranÃ§ois Mangin, Marie Chupin, Marie-Odile Habert, Yann Cointepas, Bénédicte Batrancourt, Cyril Poupon, Olivier Colliot, Habib Benali and Stéphane Lehéricy. *CATI: A large instrument for muticenter neuroimaging.* Journal of Neuroradiology, vol. 43, no. 2, pages 71 – 72, 2016. 43rd Congress of the {SFNR}. (Cited on page 96.)

[Marcus 2010] Daniel S. Marcus, Anthony F. Fotenos, John G. Csernansky, John C. Morris and Randy L. Buckner. *Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults.* Journal of Cognitive Neuroscience, vol. 22, no. 12, pages 2677–2684, 2010. (Cited on pages 8, 24 and 102.)

[McCormick 2014] Matthew Michael McCormick, Xiaoxiao Liu, Luis Ibanez, Julien Jomier and Charles Marion. *ITK: Enabling Reproducible Research and Open Science.* Frontiers in Neuroinformatics, vol. 8, no. 13, 2014. (Cited on page 7.)

[Mikheev 2008] Artem Mikheev, Gregory Nevsky, Siddharth Govindan, Robert Grossman and Henry Rusinek. *Fully automatic segmentation of the brain from T1-weighted MRI using Bridge Burner algorithm.* Journal of Magnetic Resonance Imaging, vol. 27, no. 6, pages 1235–1241, 2008. (Cited on page 12.)

[Mills 2014] K L Mills, A-L Goddings, L. S Clasen, J. N Giedd and S-J Blakemore. *The Developmental Mismatch in Structural Brain Maturation during Adolescence.* Developmental Neuroscience, vol. 36, no. 3-4, pages 147 – 160, 2014. (Cited on page 81.)

[Nakamura 2014] Kunio Nakamura, Nicolas Guizard, Vladimir S. Fonov, Sridar Narayanan, D. Louis Collins and Douglas L. Arnold. *Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis.* NeuroImage: Clinical, vol. 4, pages 10 – 17, 2014. (Cited on page 41.)

[Nature 2013] Nature. *Announcement: Reducing our irreproducibility.* Nature, vol. 496, no. 7446, 2013. (Cited on page 7.)

[Neufang 2009] Susanne Neufang, Karsten Specht, Markus Hausmann, Onur Gúntúrkún, Beate Herpertz-Dahlmann, Gereon R. Fink and Kerstin Konrad. *Sex Differences and the Impact of Steroid Hormones on the Developing Human Brain.* Cerebral Cortex, vol. 19, no. 2, pages 464–473, 2009. (Cited on page 79.)

[Niethammer 2011a] M. Niethammer, Y. Huang and F-X. Vialar. *Geodesic Regression for Image Time-Series.* NeuroImage, vol. 8, no. 2, pages 135–160, 2011. (Cited on page 100.)

[Niethammer 2011b] Marc Niethammer, Yang Huang and François-Xavier Vialard. Medical image computing and computer-assisted intervention – miccai 2011: 14th international conference, toronto, canada, september 18-22, 2011, proceedings, part ii, chapitre Geodesic Regression for Image Time-Series, pages 655–662. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. (Cited on page 32.)

[Parker 1983] J.Anthony Parker, Robert V. Kenyon and D. Troxel. *Comparison of Interpolating Methods for Image Resampling.* Medical Imaging, IEEE Transactions on, vol. 2, no. 1, pages 31–39, March 1983. (Cited on page 14.)

[Prastawa 2004] Marcel Prastawa, John Gilmore, Weili Lin and Guido Gerig. *Automatic Segmentation of Neonatal Brain MRI.* In Christian Barillot, DavidR. Haynor and Pierre Hellier, editeurs, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004, volume 3216 of *Lecture Notes in Computer Science*, pages 10–17. Springer Berlin Heidelberg, 2004. (Cited on page 12.)

[Reuter 2011] Martin Reuter and Bruce Fischl. *Avoiding asymmetry-induced bias in longitudinal image processing.* NeuroImage, vol. 57, no. 1, pages 19 – 21, 2011. (Cited on page 32.)

[Reuter 2012] Martin Reuter, Nicholas J. Schmansky, H. Diana Rosas and Bruce Fischl. *Within-subject template estimation for unbiased longitudinal image analysis.* NeuroImage, vol. 61, no. 4, pages 1402 – 1418, 2012. (Cited on pages 7, 39 and 81.)

[Ridgway 2015] G.R. Ridgway, K.K. Leung and J. Ashburner. *Computing Brain Change over Time.* Brain Mapping, vol. 1, pages 417 – 428, 2015. (Cited on pages 2 and 32.)

[Rohrer 2013] Jonathan D. Rohrer, Francesca Caso, Colin Mahoney, Maya Henry, Howard J. Rosen, Gil Rabinovici, Martin N. Rossor, Bruce Miller, Jason D. Warren, Nick C. Fox, Gerard R. Ridgway and Maria Luisa Gorno-Tempini. *Patterns of longitudinal brain atrophy in the logopenic variant of primary progressive aphasia.* Brain and Language, vol. 127, no. 2, pages 121 – 126, 2013. (Cited on page 32.)

[Samaille 2012] Thomas Samaille, Ludovic Fillon, Rémi Cuingnet, Eric Jouvent, Hugues Chabriat, Didier Dormont, Olivier Colliot and Marie Chupin. *Contrast-Based Fully Automatic Segmentation of White Matter Hyperintensities: Method and Validation.* PLoS ONE, vol. 7, no. 11, pages 1–14, 11 2012. (Cited on page 32.)

[Scahill 2002] RI Scahill, JM Schott, JM Stevens, MN Rossor and NC Fox. *Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI.* Proc Natl Acad Sci U S A, vol. 99, pages 4703–7, 2002. (Cited on page 7.)

[Scahill 2003] RI Scahill, C Frost, R Jenkins, JL Whitwell, MN Rossor and NC Fox. *A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging.* Archives of Neurology, vol. 60, no. 7, pages 989–994, 2003. (Cited on page 100.)

[Schiratti 2015] J.-B. Schiratti, S. Allassonnière, A. Routier, O. Colliot and S. Durrleman. A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data, pages 564–575. Springer International Publishing, Cham, 2015. (Cited on page 75.)

[Schott 2005] J. M. Schott, S. L. Price, C. Frost, J. L. Whitwell, M. N. Rossor and N. C. Fox. *Measuring atrophy in Alzheimer disease: A serial MRI study over 6 and 12 months.* Neurology, vol. 65, no. 1, pages 119–124, 2005. (Cited on page 25.)

[Schumann 2010] G Schumann, E Loth, T Banaschewski, A Barbot, G Barker, C Buchel, P J Conrod, J W Dalley, H Flor, J Gallinat, H Garavan, A Heinz, B Itterman, M Lathrop, C Mallik, K Mann, J-L Martinot, T Paus, J-B Poline, T W Robbins, M Rietschel, L Reed, M Smolka, R Spanagel, C Speiser, D N Stephens, A Strohle and M Struve. *The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology.* Mol Psychiatry, vol. 15, pages 1128–1139, December 2010. (Cited on pages 40 and 80.)

[Seiler 2012] C. Seiler, X. Pennec and M. Reyes. *Capturing the Multiscale Anatomicale Shape Variability with Polyaffine Transformation Trees.* NeuroImage, vol. 8, no. 2, pages 135–160, 2012. (Cited on pages 76 and 98.)

[Shattuck 2001] David W. Shattuck, Stephanie R. Sandor-Leahy, Kirt A. Schaper, David A. Rottenberg and Richard M. Leahy. *Magnetic Resonance Image Tissue Classification Using a Partial Volume Model.* NeuroImage, vol. 13, no. 5, pages 856 – 876, 2001. (Cited on page 12.)

[Sled 1998] J.G. Sled, A.P. Zijdenbos and A.C. Evans. *A nonparametric method for automatic correction of intensity nonuniformity in MRI data.* Medical Imaging, IEEE Transactions on, vol. 17, no. 1, pages 87–97, Feb 1998. (Cited on page 12.)

[Smith 2002] Stephen M. Smith. *Fast robust automated brain extraction.* Human Brain Mapping, vol. 17, no. 3, pages 143–155, 2002. (Cited on page 12.)

[Stefanescu 2004] Radu Stefanescu, Olivier Commowick, Grégoire Malandain, Pierre-Yves Bondiau, Nicholas Ayache and Xavier Pennec. *Non-rigid Atlas to Subject Registration with Pathologies for Conformal Brain Radiotherapy.* In Christian Barillot, DavidR. Haynor and Pierre Hellier, editeurs, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004, volume 3216 of *Lecture Notes in Computer Science*, pages 704–711. Springer Berlin Heidelberg, 2004. (Cited on page 18.)

[Südmeyer 2012] Martin Südmeyer, Peter Pieperhoff, Stefano Ferrea, Holger Krause, Stefan Jun Groiss, Saskia Elben, Lars Wojtecki, Karl Zilles, Katrin Amunts and Alfons Schnitzler. *Longitudinal Deformation-Based Morphometry Reveals Spatio-Temporal Dynamics of Brain Volume Changes in Patients*

with Corticobasal Syndrome. PLoS ONE, vol. 7, no. 7, page e41873, 07 2012. (Cited on page 7.)

[Trouvé 1998] A. Trouvé. *A fast diffeomorphic.* NeuroImage, vol. 8, no. 2, pages 135–160, 1998. (Cited on page 100.)

[Tustison 2010] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich and J. C. Gee. *N4ITK: Improved N3 Bias Correction.* IEEE Transactions on Medical Imaging, vol. 29, no. 6, pages 1310–1320, June 2010. (Cited on pages 12, 102 and 103.)

[Vercauteren 2008] Tom Vercauteren, Xavier Pennec, Aymeric Perchant and Nicholas Ayache. *Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach.* In Dimitris Metaxas, Leon Axel, Gabor Fichtinger and Gábor Székely, editeurs, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008, volume 5241 of *Lecture Notes in Computer Science*, pages 754–761. Springer Berlin Heidelberg, 2008. (Cited on pages 16 and 100.)

[Wittchen 2011] H.U. Wittchen, F. Jacobi, J. Rehm, A. Gustavsson, M. Svensson, B. Jónsson, J. Olesen, C. Allgulander, J. Alonso, C. Faravelli, L. Fratiglioni, P. Jennum, R. Lieb, A. Maercker, J. van Os, M. Preisig, L. Salvador-Carulla, R. Simon and H.-C. Steinhausen. *The size and burden of mental disorders and other disorders of the brain in Europe 2010.* European Neuropsychopharmacology, vol. 21, no. 9, pages 655 – 679, 2011. (Cited on page 1.)

[Yushkevich 2010] Paul A. Yushkevich, Brian B. Avants, Sandhitsu R. Das, John Pluta, Murat Altinay and Caryne Craige. *Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: An illustration in ADNI 3 T MRI data.* NeuroImage, vol. 50, no. 2, pages 434 – 445, 2010. (Cited on page 32.)

Beyond Volumetry in Longitudinal Deformation-Based Morphometry: Application to Sexual Dimorphism during Adolescence

Abstract: Analysing the progression of brain morphological changes in time series of images is an important topic in neuroimaging. Although the development of longitudinal databases has helped reducing the inter-individual variability, there still exist numerous biases that need to be avoided when capturing longitudinal evolutions. Moreover, when the intra-subject changes are very small with respect to the inter-subject variability it is crucial to know if the available methods can capture the longitudinal change with no bias. In most of the studies, these longitudinal changes are limited to scalar volumetric changes in order to ease their analysis. However, one can observe that brain changes are not limited to volumetry. In this multivariate case, the interpretation is more difficult. This thesis addresses these problems along three main axes.

First, we propose a longitudinal Deformation-based Morphometry processing pipeline to robustly estimate the longitudinal changes. We detail the whole sequencing of the processing steps as they are key to avoid adding bias. In addition to this contribution we integrate a modification to the non-linear registration algorithm by masking the similarity term while keeping the symmetry of the formulation. This change increases the robustness of the results with respect to intensity artifacts located in the brain boundaries and leads to increased sensitivity of the statistical study on the longitudinal deformations. The proposed processing pipeline is based on freely available software and tools so that it is fully reproducible.

The second axis is dedicated to the evaluation of the accuracy and reproducibility of our non-linear registration method in terms of volumetry. We compare our method to three other popular volumetric longitudinal methods (segmentation-based and registration-based). We use simulated ground truth and real data for which changes are small. We show that registration based-methods are generally more accurate, consistent and reproducible than the segmentation-based method FreeSurfer. We also notice that for high changes (superior to 10%) registration-based methods highly under-estimate changes. Finally, we show that log-Jacobian integration is equivalent to volumetry for changes less than 10 % and that the numerical schemes used for its computation have no significant influence.

Finally, we present a method to go beyond volumetry with the multivariate statistical analysis of two groups by studying the whole three-dimensional deformation field. We propose to disentangle the group differences from the longitudinal population evolution. The results are easily interpretable with respect to other existing methods. We apply the proposed method to the study of sexual dimorphism during adolescence. Results show that at 14 years of age, there is no difference between females and males pre-frontal cortex and that an important differentiation occurs during the two following years: at 16, girls' pre-frontal cortex is in advance of around five months.

Keywords: Longitudinal images, Evaluation, Volumetric methods, Non-rigid registration, Processing pipeline, Multivariate statistics, Group comparison, Sexual dimorphism

Au delà de la volumétrie en morphométrie basée sur les déformations : application au dimorphisme sexuel durant l'adolescence

Résumé : L'analyse des changements morphologiques du cerveau dans des séries temporelles d'images est un sujet important en neuroimagerie. Bien que le développement des bases de données longitudinales ait aidé à réduire la variabilité inter-individu, il reste encore de nombreux biais qui doivent être évités lors de l'estimation des évolutions longitudinales. De plus, lorsque les changements intra-sujets sont très faibles par rapport à la variabilité inter-sujet, il est crucial de savoir si les méthodes existantes peuvent capturer sans biais les changements longitudinaux. Dans la plupart des études, les changement longitudinaux sont limités à leur composante volumétrique scalaire afin d'en faciliter l'analyse. Cependant, les changements cérébraux ne sont généralement pas uniquement volumétriques et dans ce cas multivarié, l'interprétation est alors plus difficile. Cette thèse adresse ces problèmes en suivant trois axes principaux.

Premièrement, nous proposons une chaîne de traitement longitudinale reposant sur la morphométrie à partir de déformations et ayant pour but d'estimer de manière robuste les changements longitudinaux. Afin d'eviter de rajouter du biais, nous détaillons tout l'enchaînement des étapes de traitement. En plus de cette contribution, nous intégrons une modification de l'algorithme de recalage non-linéaire qui consiste à masquer le terme de similarité tout en conservant la symétrie de la formulation. Cette contribution augmente la robustesse des résultats vis-à-vis des artefacts d'intensité situés en bordure du cerveau et augmente ainsi la sensibilité de l'étude statistique realisée sur les déformations longitudinales. Pour que la chaîne de traitement proposée soit totalement reproductible, seuls des logiciels et outils disponibles librement sont utilisés.

Le deuxième axe est dédié à l'évaluation de la précision et de la reproductibilité de notre méthode de recalage non-linéaire en terme de volumétrie. Nous comparons notre méthode à trois autres méthodes populaires pour la volumétrie en longitudinal utilisant la segmentation et le recalage. Nous utilisons une vérité terrain simulée et des données réelles pour lesquelles les changements sont faibles. Nous montrons que les méthodes de recalage sont plus précises, cohérentes et reproductibles que la méthode de segmentation FreeSurfer. On note aussi que pour des changements importants (supérieurs à 10%) les méthodes de recalage sous-estiment très fortement les changements. Finalement, nous montrons que l'intégration du log-jacobien est équivalente aux changements de volume pour des changements inférieurs à 10% et que les schémas numériques utilisés pour son calcul n'ont pas d'influence significative.

Finalement, nous présentons une méthode pour aller au delà de la volumétrie avec l'analyse statistique multivariée de deux groupes en étudiant le champ de déformation tri-dimensionnel. Nous proposons de séparer les différences de groupe de

l'évolution longitudinale de la population. Les résultats sont facilement interprétables en comparaison d'autres méthodes existantes. Nous appliquons cette méthode à l'étude du dimorphisme sexuel pendant l'adolescence. Les résultats montrent qu'à 14 ans il n'existe pas de différence entre le cortex préfrontal des garçons et celui des filles. Une différentiation importante s'opère cependant durant les deux années suivantes : à 16 ans, le cortex préfrontal des filles est en avance d'environ cinq mois.
Mots-clés : Images longitudinales, Evaluation, Méthodes pour la volumétrie, Recalage non-linéaire, Chaîne de traitement, Statistiques multivariées, Comparaison de groupe, Dimorphisme sexuel

---