Ecole Normale Supérieure de Cachan

Ecole doctorale Sciences Pratiques

Habilitation à diriger des recherches

Spécialité : **Mathématiques Appliquées**

Stéphanie ALLASSONNIÈRE

STATISTICAL ANALYSIS OF DEFORMABLE TEMPLATE MODELS FOR MEDICAL IMAGE UNDERSTANDING

Garant : **Alain** TROUVÉ

Rapporteurs :

Pr. **Ian** DRYDEN
Pr. **Marc** LAVIELLE
Dr. **Xavier** PENNEC

Soutenue publiquement le **13/11/2013** devant le jury composé de

Dr. Marc	**LAVIELLE**	INIRA PROJET POPIX ,	Rapporteur
Pr. Stéphane	**MALLAT**	ÉCOLE NORMALE SUPÉRIEURE, PARIS	Examinateur
Pr. Jean-Michel	**MARIN**	UNIVERSITÉ DE MONTPELLIER	Examinateur
Pr. Lionel	**MOISAN**	UNIVERSITÉ PARIS DESCARTES	Examinateur
Dr. Xavier	**PENNEC**	INRIA PROJET ASCLEPIOS	Rapporteur
Pr. Alain	**TROUVÉ**	ÉCOLE NORMALE SUPÉRIEURE, CACHAN	Garant

Ce manuscrit présente mon parcours dans le monde de la recherche et de l'enseignement et mes responsabilité scientifiques depuis ma soutenance de thèse en juillet 2007.

Le premier chapitre 1 récapitule mon CV de manière détaillé incluant mes enseignements. Le chapitre suivant 2 résume mes travaux dans le domaine des modèles statistiques en analyse d'images à partir de mon travail de thèse.

Les chapitres 3 à 5 présente de manière plus précise mon activité de recherche depuis 2007 autour de 3 axes :

1. Les généralisations du modèle BME (Bayesian Mixed Effect)- Template incluant les contraintes anatomiques et numériques liées à l'application aux données médicales,
2. L'analyse détaillée des corrélations des déformations,
3. Personnification de modèles pour la compréhension de pathologies.

Contents

CHAPTER 1

Curriculum Vitae détaillé

Contents

1. Curriculum Vitae

Stéphanie ALLASSONNIERE
née le 18 décembre 1979 à Lens (62), nationalité française, mariée.
Adresse : CMAP Ecole Polytechnique, Route de Saclay, 91128 Palaiseau cedex France.
Téléphone : 01-69-33-45-65
E-mail : Stephanie.Allassonniere@polytechnique.edu
Page web : www.cmapx.polytechnique.fr/ allassonniere/

Carrière et formation.

Sept 2008-: Professeur Chargé de Cours, Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique, France

2007-2008: Post-doctorat au Center for Imaging Science, Johns Hopkins University.

2003-2007: Doctorat de mathématiques appliquées soutenue le 4 juillet 2007, Université Paris 13

- Directeurs : Alain Trouvé et Laurent Younes
- Specialité: Mathématiques appliquées
- Titre: Representation and statistic estimation of deformable models for shape recognition and computational anatomy.

1999-2003: Elève de l'ENS de Cachan (DEA Mathmatiques / Vision / Apprentissage en 2003, Agrégation de Mathématiques en 2002, concours d'entrée en tant que Normalienne en 2001).

Prix.

Prix Excellencia 2010 catégorie Recherche, organisé par le comité Excellencia (Innoc.Europe avec Epita et Microsoft en partenariat avec la Comission Européenne).

2. Publications et communications

2.1. Liste de publications. Les publications listées ci-dessous sont pour la plupart disponibles sur ma page web ou sur les sites des journaux.

Articles de journaux acceptés.

A1: Toward a coherent statistical framework for dense deformable template estimation, S.ALLASSONNIÈRE, Y.AMIT, A.TROUVÉ. Journal of the Royal Statistical Society, Series B, 69(1), 2007, pp 3-29.

A2: Construction of Bayesian deformable models via stochastic approximation algorithm: A convergence study, S.ALLASSONNIÈRE, E.KUHN, A.TROUVÉ. Bernoulli Journal 16(3), 2010, pp 641-678.

A3: Stochastic Algorithm For Parameter Estimation For Dense Deformable Template Mixture Model, S.ALLASSONNIÈRE, E.KUHN. ESAIM-PS, 14, 2010, pp 382-408.

A4: Bayesian Consistent Estimation in Deformable Models using Stochastic Algorithms: Applications to Medical Images. S.ALLASSONNIÈRE, E.KUHN, A.TROUVÉ. Journal de la Société Française de Statistiques, 151(1), 2010, pp 1-16.

A5: A Stochastic Algorithm for Probabilistic Independent Component Analysis. S.ALLASSONNIÈRE, L.YOUNES. Annals of Applied Statistics, 6(1), 2012, pp 125-160.

A6: Tumor Growth Parameters Estimation and Source Localization From a Unique Time Point: Application to Low-grade Gliomas. I. REKIK, S. ALLASSONNIERE, H. DESLINGETTE, O. CLATZ, E. GEREMIA, N. AYACHE. Computer Vision and Image Understanding, 117(3), 2013, pp 238-249.

A7: Aircraft classification with a low resolution infrared sensor. S.LEFEBVRE, S.ALLASSONNIÈRE, J.JAKUBOWICZ, T.LASNE, AND E.MOULINES. . Machine Vision and Application Journal, 24(1), 2013, pp 175-186.

A8: Sparse Adaptive Parameterization of Variability in Image Ensembles. S. DURRLEMAN, S. ALLASSONNIERE, S.JOSHI. International Journal of Ccomputer Vision, 101(1), 2013, pp 161-183, .

A9: A Review of Medical Image Analysis Methods in MR/CT-imaged Acute-subacute Ischemic Stroke Lesion: Segmentation, Prediction and Insights into Dynamic Evolution Simulation Models. I. REKIK, S. ALLASSONNIERE, T. CARPENTER AND J. WARDLAW. NeuroImage Clinical, 1(1), 2012, pp 164-178.

A10: Deformable templates and statistical models for image and shape analysis. S.ALLASSONNIÈRE, J.BIGOT, J.A.GLAUNES, T.MAIRE AND F.RICHARD. Accepted for publication in the Annals Mathématiques Blaise Pascal.

A11: Spatio-temporal dynamic simulation of acute perfusion/diffusion ischemic stroke lesions evolution: a pilot study derived from longitudinal MR patient data. I. REKIK, S. ALLASSONNIERE, S. DURRLEMAN, T. CARPENTER AND J. WARDLAW.

Articles de journaux soumis.

A12: Convergent stochastic Expectation Maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation. S. ALLASSONNIERE AND E. KUHN.

A13: Probabilistic Atlas and Geometric Variability Estimation to Drive Tissue Segmentation. H.XU, B.THIRION AND S. ALLASSONNIÈRE

Articles en préparation.

A14: Bayesian Mixed Effect Atlas Estimation under Diffeomorphic constraint on the deformation model. S.ALLASSONNIÈRE, S.DURRLEMAN AND E.KUHN

A15: Multimodal probabilistic atlas estimation for functional IRM analysis. H.XU, B.THIRION AND S. ALLASSONNIÈRE (in preparation)

A16: A 4D Patient-specific metamorphosis-based method to model ischemic stroke lesion evolution. I. REKIK, S. ALLASSONNIERE, T. CARPENTER AND J. WARDLAW.

A17: Fréchet means on top and quotient space may not be consistent. S.ALLASSONNIÈRE AND A.TROUVÉ

A18: Computational anatomy applied to inverse problems in structural biology. O.ÖKTEM, P.RAVIKUMAR AND S.ALLASSONNIÈRE (in preparation)

A19: Dense Diffusion weight and anatomical brain atlas estimation. S.ALLASSONNIÈRE, M.GRASSEAU.

Manuscrit de thèse. **Representation and Statistical Estimation of Deformable Template Models for Shape Recognition and Computational Anatomy**.

Articles dans les actes de conférences à comité de lecture.

C1: Geodesic Shooting and Diffeomorphic Matching Via Textured Meshes, S. ALLASSONNIÈRE, A.TROUVÉ, L.YOUNES. Energy Minimization Methods in Computer Vision and Pattern Recognition, Lecture Notes in Computer Science, 3757, 2005, pp 365-381.

C2: Generative Model and consistent estimation algorithms for non-rigid deformable models, S.ALLASSONNIÈRE, E.KUHN, Y.AMIT, A.TROUVÉ. IEEE International Conference on Acoustics, Speech and Signal Processing, 5, 2006, pp V.

C3: MAP Estimation of Statistical Deformable Template Via Nonlinear Mixed Effect Models: Deterministic and Stochastic Approaches, S.ALLASSONNIÈRE, E.KUHN, A.TROUVÉ. 2nd Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop on Mathematical Foundations of Computational Anatomy (MFCA), 2008, pp 80-91.

C4: Consistent Atlas Estimation on BME Template Model : Applications to 3D Biomedical Images, S. ALLASSONNIÈRE, E. KUHN, J. T. RATNANATHER, A. TROUVÉ. Probabilistic Models for Medical Image Analysis (PMMIA) worshop of the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2009 conference.

C5: Detecting long distance conditional correlations between anatomical regions using Gaussian Graphical Models. S. ALLASSONNIÈRE, P.JOLIVET, C.GIRAUD. 3rd Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop on Mathematical Foundations of Computational Anatomy (MFCA), 2011, pp 111-122.

C6: Detection and classification of poorly known aircraft with a low-resolution infrared sensor. S. LEFEBVRE, S. ALLASSONNIÈRE, G. DURAND, J. JAKUBOWICZ, E. MOULINES, A. ROBLIN. Proc. of SPIE, 8050, Signal Processing, Sensor Fusion, and Target Recognition XX, 80501I, 2011.

C7: Mathematical Image Processing Algorithms in determining Stroke Tissue Status and Predicting its Fate: Systematic Review reveals Untapped Potentials. I. REKIK, S. ALLASSONNIÈRE, T. CARPENTER AND J. WARDLAW.. European Stroke Conference, Hamburg 2011.

C8: Dynamic Patient-specific Modeling of Ischemic Stroke Lesion Evolution: From Presentation to Final Damage using Diffusion, Perfusion and T2 MR imaging. I. REKIK, S. ALLASSONNIÈRE, S. DURRLEMAN, T.K. CARPENTER, J.M. WARDLAW. European Stroke Conference, Lisbon, 2012

C9: 4D spatio-temporal perfusion-diffusion Evolution scenarios estimation in Acute-subacute ischemic stroke. I. REKIK, S. ALLASSONNIÈRE, S. DURRLEMAN, T.K. CARPENTER, J.M. WARDLAW. International Society for Magnetic Resonance in Medicine Conference, Melbroune, 2012.

C10: A 4D Patient-specific metamorphosis-based method to model ischemic stroke lesion evolution from acute diffusion-weighted to final T2-defined outcome. I. REKIK, S. ALLASSONNIÈRE, T. CARPENTER AND J. WARDLAW. International Stroke Conference, Honolulu, 2013.

C11: Bayesian Estimation of Probabilistic Atlas for Anatomically-Informed Functional MRI Group Analyses. H. XU, B. THIRION, S. ALLASSONNIÈRE. Medical Image Computing and Computer Assisted Intervention (MICCAI) 2013.

Livre. **Zoom sur les suites numérique**. Manuel parascolaire pour les Terminales S, Edition Belin (1999)

2.2. Exposés.

Conférence invité.

- **Forum des jeunes mathématiciennes**, Mini-cours sur les enjeux statistiques en anatomie numérique (Novembre 2011)
- **Mathematics department, KTH, Sweeden**, Tutorial on Computational Anatomy, (October 2010)
- **Conférence Méthodes Mathématiques pour l'Image**, Exposé sur l'estimation d'altas en grande dimension, (Orléans, juin 2012)
- **Institute of Pure and Applied Mathematics (UCLA)**. Summer School: Mathematics in Brain Imaging, (July 2008).
- **Siemens Corporate Research**. Exposé sur l'estimation d'atlas pour une collaboration (04/02/08).
- **Workshop on Shape Spaces**. SAMSI (July 7-13 2007); talk at the New Researcher Session .

Exposés de séminaires ou conférences.

- **Séminaire BigMC**, (12/01/2012)
- **Institut de Mathématiques de Toulouse**, Journée thématique, (15/06/2011)
- **Imperial College London**, Shape meeting (17/05/2011)
- **SCI institut**, Séminaire d'Image, University of Utah (30/08/2010)
- **Probabilistic Models for Medical Image Analysis Worshop**, Poster, (Londres 2010).
- **Telecom ParisTech**, Image and statistics seminar (03/19/09).
- **Applied Mathematics and Statistics department**. Johns Hopkins University, (03/04/08).
- **Department of Mathematics and Statistics**. University of Maryland Baltimore County (UMBC). Seminar (16/11/07).
- **LATP University of Provense**. Marseille, seminar (12/01/06).
- **Colloquim of Centre de Mathematiques et leurs Applications**, Cachan, France, (05/18/06).
- **ICASSP conference**, Toulouse oral presentation (05/16/06).
- **Workshop on Shape Spaces**, IMA poster session (02-06/04/06).
- **Groupe de travail ISIS and MSPC**; oral presentation, ENST (09/06/05).
- **LAGA probability and statistics seminar**, (13/04/05).
- **Center for Imaging Science**, Johns Hopkins University (Baltimore) seminars (15/03/05).

3. Encadrement scientifique et enseignement

3.1. Encadrement scientifique.

Etudiants de Master :.

- Pierre Jolivet (M1) ; coencadrement avec Christophe Giraud
- Islem Rekik (M2) ; coencadrement avec Hervé Deslingette et Nicholas Ayache
- James Newling (M1) ; coencadrement avec Christophe Giraud et Emmanuel Bacry

Etudiants en thèse :.

- Islem Rekik (09/2010-) ; co-direction (50%) avec Joanna Wardlaw (Professeur en Neuro-radiologie à l'university d'Edinbourg), thèse préparée au sein du Department of Clinical Neurosciences de l'université d'Edinbourg.
- Hao XU (03/2011-) ; co-direction (80%) avec Bertrand Thirion (Directeur de recherche INRIA, Equipe Parietal), thèse préparée au sein du CMAP, Ecole Polytechnique.

3.2. Enseignement. Mes notes de cours manuscrites ainsi que les feuilles d'exercices (avec corrigés) sont mises en ligne sur mon site au fur et à mesure de l'avancement des cours.

2012-present Optimisation et simulations, cours du Master MathSV de l'université Paris 11 Orsay, en collaboration Alexandre d'Aspremont et Sylvain Faure.

2009-present Méthodes MCMC pour l'analyse d'images, cours du Master MVA de l'ENS Cachan en collaboration avec Gersende Fort et Eric Moulines.

2008-present
- Petites classes du cours d'Introduction à la théorie des probabilités de Sylvie Méléard, élèves de 1^{ere} année de l'Ecole Polytechnque.
- Petites classes du cours de Modélisation mathématique, élèves de 2^{eme} année de l'Ecole Polytechnque.
- Petites classes du cours de Traitement du Signal, élèves de 3^{eme} année de l'Ecole Polytechnque.
- Encadrement des projets de première (tronc commun), deuxième (PSC, Cours de Modélisation mathématique) et troisième année (Enseignements d'approfondissement pour le cours de Traitement du signal).
- Directrice d'option des stages de recherche mention Image et Signal (élèves de 3^{eme} année de l'Ecole Polytechnque).

2006-2007 ATER au département de mathematiques de l'Université Paris Dauphine
TD du cours de Statistiques Mathematiques, L3
TD / TP du cours d'analyse numerique, L3.

2003-2006 Monitrice au département de mathématiques de l'Université Paris 13
TD du cours de Statistiques pour l'economie, L3
TD / TP du cours d'analyse numerique, L3

2002-2003 Oraux hebdomadaires des classes préparatoires du Lycée Descartes d'Anthony.

3.3. Jury de thèse.

- Examinatrice de la thèse de Joan Bruna, thèse encadrée par Stéphane Mallat au sein du CMAP, Ecole Polytechnique.

4. Participation à la vie scientifique, responsabilité collectives

- Co-organisatrice du séminaire du CMAP depuis janvier 2009.
- Membre élue du conseil de département du département de Mathématiques appliquées à l'école Polytechnique depuis 2012.
- Responsable du master Mathématiques - Vision - Apprentissage à l'école Polytechnique depuis 2012.
- Participation à la commission de spécialiste pour le recrutement d'un maître de conférence au laboratoire de mathématiques de Toulouse section 26-60 en mai 2012.
- Organisation d'une session de la conférence du CANUM en juin 2012.
- Organisation d'une session de la conférence de la SIAM on Imaging Science en avril 2010.

5. Séjours de recherche à l'étranger

Sept 2007-Août 2008: Post-doctorat au Center for Imaging Science, Johns hopkins University.

2010-2011: Professeur invité pour 6 mois au Scientific Computing and Imaging institute, University of Utah.

6. Financements

- **MMoVNI**, Digiteo 2010. **Mathematical Modelling of between-subject Variability in NeuroImaging**. Role: Principal Investigator.
- **IRMgroup**, ANR programme Blanc. **Invariant Representations with Multiresolution Grouping**. (PI: Stéphane Mallat)
- **HM-TC**, ANR programme phare. **Hippocampus, Memory and Temporal Conciousness** (PI: Olivier Colliot)
- **Plan Alzheimer** Financements Fondation Plan Alzheimer. (PI. Jean-François Mangin)

CHAPTER 2

Introduction to my research activity

Contents

1. Introduction to the general context

In order to better understand the positioning of my research, I will first introduce the general framework I have considered.

1.1. Models and Shape Analysis. Most questions considered in my research are related to the representation and the analysis of geometrical structures in images upon which *deformations* can act. One central point is to provide a model of a population of images describing the observed objects as geometrical variations (characterised by the deformations) of one or several reference images and quantifying this variability inside this population. These references are called *templates* and this framework is known as "Deformable Templates".

On the one hand, one can see this issue as the definition of a sub-manifold generated by a family of studied images in an adequate space. This requires to exhibit a metric so that this space can be identified to a Riemanian manifold and provides a way to quantify variations as distances between shapes. In this setting, the analysis is simplified as there exists an exponential map which locally linearises the sub-manifold. The template is the element on the manifold where the exponential map is calculated. This is known as the pure geometrical approach which has been studied and improved for this past decade.

On the other hand, this can also be viewed with a statistical point of view. The problem is to construct probabilistic models of the variable objects represented in the image in order to quantify *in a statistical way* this variability. This aims at both describing the population of images (thanks to the distribution it follows) and at recognising a new shape which will be or not be a "normal" variation of this observed population. This aspect of the analysis is very important, leading to a new branch called "Statistical Shape Analysis". It has concentrated a lot of work since a few years, following the geometric advances which are more and more sophisticated.

One efficient probabilistic model was proposed by Grenander [**35, 36**]. This consists in proposing to model the observations as i.i.d realisations of a parametric generative statistical model. The observations are assumed to be noisy discretisation of a randomly deformed template image. Using statistical learning, one can expect to estimate the model parameters, in particular the template, as the one which best fit a set of observed data. This last point of view is the one I have adopted in my researches.

1.2. Computational Anatomy. The main motivation of my work is guided by medical applications. This is now possible thanks to the increasing number of scans of increasing resolution which are acquired from both clinical and research protocols.

Computational Anatomy [**37**] aims at building some probability models of the anatomical variability, in particular, human anatomy. A particular subject of study is neuro-anatomy, whose interest is focused on the human brain variability, all its different cortical and subcortical structures, function and connectivity. The goal is to build numerical **atlases** of the anatomy. We define an atlas as the coupling of a *representative numerical image* of the brain (the template) and a *statistical quantification* of their normal variability taking into account its diversity and complexity in tissues, shapes and functions. The template image is expected to exhibit common features which appear in the observed population. It should be an image representing a characteristic and ideal structure or object. The geometrical variability is analysed through the deformations which are applied to the template so that it looks like the observations. This technic is called registration and is the core

of the pure geometrical approach. Registering the template on an observation is the computation of an optimal (in a sense to be defined) deformation - in a specified a more or less complex class - so that the application of the deformation to the template produces an image equal (resp. close) to the observation (leading to exact - resp. inexact - matching). The notion of similarity between the deformed template and the observation is based on the amount of deformation required to register the template onto the target image. The statistical quantification of the normal variability summarises the information given by these deformations. This, together with the template, forms the atlas.

Important issues in atlas estimation are to :

(1) enable to register any new data in the coordinates of the template in order to transport the available information from the template image onto these new images : segmentation, functional activation, etc; This registration should be penalised as a function of its "normality";
(2) enable, when there is a probability law on the atlas, to quantify anatomical structure variability in different sub-groups in a given population as distance between distributions;
(3) and most challenging, enable to help clinician to give an early diagnostic of some pathologies which can be detected using some complex anatomic shape criteria.

In order to perform such an estimation, three objects are required. First, to quantify the distance between images, we have to set a metric on the space they belong to. This distance has to take into account the differences between images up to the normal variability of the population. Therefore, the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework (introduced in the following), however complex, is particularly well adapted as it enables to describe the data with a generative mechanism through a group action. Then, one needs to provide a statistical model to explain the generation of the observed population from the atlas. In particular, the atlas, as the parameters of this model, will be estimated thanks to the observations. Targeting medical imaging applications, the Bayesian framework appears appropriate as it enables to face the high-dimension-low-sample-size paradigm. Last, to perform the statistical estimation of the atlas, one requires performant algorithms.

As the deformations are central tool in this analysis, I will briefly recall the big picture of the registration issue in the particular settings I have worked with.

1.3. Registration Problem. Registration of images, i.e. matching data onto the reference image and the inverse problem of matching the template onto the data, using if possible diffeomorphic transformations - play an important role in this context [**65, 69, 37**]. Grenander in [**35**] proposed a theoretical framework to handle this and proposed to represent the possible variations of a given anatomical object as the orbit Gm_0 of a template m_0 under the action of the deformation group G. If one puts a metric on the group G, then the group action enables to transport this metric on the object space and defines a geodesic distance between images [**54**]. This provides a way of quantifying the differences between observed images. The way this is formulated is explained in the following paragraph.

1.3.1. *From "Small" to Diffeomorphic Deformation framework.* Rigid body transformations (mainly rotations and translations) or similitudes do not have enough degree of freedom to capture

the large heterogeneity of anatomical shapes. This requires to consider higher dimensional groups of deformations which would also be constrained to relevant anatomical variations.

The first stage is to consider that the deformation corresponds to the application of a dense vector field v on the domain Ω. This writes :

$$\phi = Id_\Omega + v \; . \tag{2.1}$$

This framework is usually called "small deformations" and has been widely used because of its simplicity. The matching problem is reduced to finding the "optimal" (in s sense to be defined) vector field v -w.r.t. the metric set on its ambient space- such that the deformed template is equal (resp. close) to the observation. Dimension reduction can be done by considering that the deformation field is only driven by control points and the dense field is given by spline interpolation ([**9**], [**13**]).

This really simple but important framework for applications has to face equally important problems. The first one is invertibility, as there are no explicit constraints on the norm of the derivative of the velocity vector field which would ensure the deformation to be one-to-one. Topological changes are therefore likely to appear. The second problem is that this formulation does not provide a group action (in the mathematical sense) of the set of transformations on the set of images. This prevents the previously mentioned "metrisation" of the set of objects through the "metrisation" of the transformation "group".

A solution has been proposed by A. Trouvé in [**69**]. The idea is to suppose that the deformation is a composition of infinitesimal small deformations close to identity leading at the limit :

$$\left\{ \begin{array}{lcl} \frac{\partial \phi_t}{\partial t} & = & v_t \circ \phi_t \\ \phi_0 & = & Id \; . \end{array} \right. \tag{2.2}$$

The final deformation is the solution at time $t = 1$ of this ordinary differential equation (under some assumptions that will ensure existence and uniqueness of the solution). Its action on an image, denoted by $\phi \; . \; I_0$, determines the template evolution toward a final state which we want to be close to the target. The group G of such deformations can be equipped with a right-invariant metric which also provides geodesic paths in the group. This property is transported to the set of objects through the group action which allows to define a distance between objects [**55**].

The only drawback of this framework is that the matching problem is apparently more complex since the deformation parametrisation $(v_t)_{0 \leq \leq 1}$ does not appear explicitly in the deformed template and the time dependency adds a new dimension to the variable of interest.

1.3.2. *Variational Formulation of Matching Problems.* In the two previously described framework, as well as in many others (e.g. [**74**]), the optimal mapping is formulated as the solution of a variational problem. Given a fixed "group" of deformations, one writes an energy which consists

of two terms, the deformation cost (its energy as a quantification of the amount of deformation required to map the template onto the object) given by d_V and the discrepancy between the target object y_{obs} (not necessarily in the orbit of the template) and the deformed template $\phi \cdot I_0$ given by g. The selected deformation is the one minimising this energy :

$$\hat{\phi} = \underset{\phi \in G}{\operatorname{argmin}} E_{I_0,V,\sigma^2}(\phi) = \underset{\phi \in G}{\operatorname{argmin}} \left\{ d_V(\mathrm{Id}, \phi)^2 + \frac{1}{2\sigma^2} g(y_{\text{obs}}, \phi \cdot I_0) \right\}. \tag{2.3}$$

Under weak assumptions on this data attachment term (cf [**34**]), this energy has a minimum in the group G of diffeomorphisms for all given positive constants σ^2. Note that this constant enables to balance the weight of the two terms depending on the user's confidence on the data and deformations.

This provides a theoretical framework to computes mappings not only between images but also many different data types going from labeled or unlabelled landmarks, curves, surfaces. The numerical solution can be obtained by a gradient descent on the previous energy.

2. Motivations and guiding line of my research

Now that the general context has been introduced, I will detail in this section my motivations and my guiding line which both have driven me for these years.

My interest quickly turned towards a major question: how can we estimate a complete atlas of the brain from a given population of images with theoretical guaranties about the output estimates? This question reveals three aspects: (1) modelling the wide range of observations we are provided with, (2) estimating this atlas numerically from this model and (3) proving the statistical relevance (in a way to be defined) of the estimate. These are the three directions I have explored.

First is the modelling issue. This started during my PhD by the proposition of a statistical generative parametric model based on Grenander's modelling in the context of "small deformations" and where the atlas is the model parameter **[A1]**, we called this model Bayesian Mixed Effect (BME) Template model. This model was performant to estimate atlases of handwritten digit images. However, targeting medical applications, one has to pay particular attention to the large variability even inside a healthy group. Therefore, the extension towards mixture of the previous model was natural **[A1]**. Moreover, dealing with real medical data implies to consider the anatomical constraints such as topological stability in the population. Thanks to the collaboration with Stanley Durrleman and Sarang Joshi, we proposed a finite dimension LDDMM variational formulation of the atlas construction **[A8]**. This comforted me in the interest of this deformation framework and motivated me to include it into the previous probabilistic model **[A14]**.

Another expectation regarding medical images is to be able to produce a complete atlas of the brain which shows all information we are able to capture with all the modalities. This challenge was guided by discussions with Bertrand Thirion and ended up in the creation of the MMoVNI (Mathematical Modelling of between-subject Variability in Neuro-Imaging) research project. The goal is of two fold : using available modalities such as MR, fMR, DW images and coupling their respective informations. This led to a three step generalisation of the BME Template model **[A13, C11, A15]** and **[A19]** which is still in progress to complete the DWI inclusion.

Medical images also come now along time to analyse the temporal evolution of a disease. To perform this analysis, we model the spatio-temporal scenario of two pathologies, brain glioma and strokes with two different approaches. First, we model the tumour propagation using Partial Differential equations (PDEs) **[A6]**. The stroke evolution was modelled using the spatio-temporal LDDMM setting **[A11]** and to take into account the scattered lesions, we proposed to extend this to metamorphosis **[A16]**. This interest gave me the chance to work in particular with Professor Joanna Wardlaw, Professor of Applied Neuroimaging in the Division of Clinical Neurosciences, Western General Hospital, Edinburgh.

These modelling shows the importance of the deformations. Therefore I was interested to analyse them in more details. I proposed to generalise two existing statistical models which capture and highlight complementary aspects of the deformations. On the one hand, Probabilistic Independent Component Analysis (ICA) shows that a small number of patterns were discriminative for the Alzheimer's disease diagnosis **[A5]**. On the other hand, the non local Gaussian Graphical Model (GGM) showed long distance dependencies between different areas of the hippocampus for healthy controls **[C5]**.

All these models enable to grasp different aspects of populations of brain images. Combining there analysis may lead to a better understanding of the brain structure, function and diseases.

Now that we have statistical models, the second aspect of my leading question is the numerical estimation process. It has to provide accurate and relevant estimates and in a reasonable computation time. My collaboration with Estelle Kuhn had a strong impact on the way I decided to solve this problem. After a test with deterministic algorithms **[A1]**, which were showing failures on some training samples, we proposed to use a stochastic algorithm to perform the numerical estimation of the atlas for the BME Template model **[A2]**. The natural extension to the mixture model required a particular attention for the numerics and we used a variant of the previous algorithm **[A3]**. However, both numerical methods suffered from their very high computational cost. This motivated us to propose a new Monte Carlo Markov Chain (MCMC) sampler - which appeared to be the numerical bottleneck -**[A12]** to boost the estimation while keeping accurate estimated.

All these models and algorithms are producing very interesting results, however are they really statistically relevant? This is the last side of my leading question: can we rely on them and more precisely what are the (asymptotic) properties of the estimators and the algorithms? To provide an answer to this question, we proved the existence of the estimators we have chosen for different models given a population **[A1, 13]**, its consistency for the BME Template model **[A1]** and the convergence of all our stochastic algorithms **[A2, A3, A12, A13]** in particular the almost sure convergence of the estimated sequence towards the theoretical estimator and its asymptotic normal behaviour. Along this we also proved the geometrical ergodicity of our new MCMC sampler **[A12]** which provides important information to the user for any other use of it.

To conclude, I would like to highlight the strengths of this statistical approach. First of all, it enables to propose a unique solution to the three aspects of my leading question: modelling of the brain, numerical estimation and statistical guaranties. Moreover, it has several other advantages.

To begin with, this approach enables to reinterpret the variational formulation of the atlas estimation. Indeed, the estimation we consider is based on the assumption that the deformations from the template to the observations are not observed random variables. Our estimator does not depend directly on them but rather on their probability distribution. An approximation of this estimator can be computed using the best matchings. This involves to compute the best registration for each image and minimise iteratively with respect to the atlas, given the deformations and with respect to these mappings, given the atlas. This clearly shows the link between the variational approach which appears as a simplification of this loop, fixing the deformation probability distribution and the tradeoff term (related to d_V and σ^2 in Eq. (2.3)).

A second advantage lives in the wide range of generalisation this model allows for. In particular, the mixture model is important as the population of interest are most often not homogeneous and require to perform a unsupervised clustering. Moreover, it enables to include and couple all modalities to increase the knowledge of the brain.

Finally, the model we propose belongs to a bigger class which are known as Mixed Effect models. The numerical estimation algorithms we have developed can therefore be used for many other models of this class, specially for estimations in high dimension with high dimensional missing variables. As an example, the non local GG model is now tested on bird migration data.

Now that the motivations are introduced, let us move to the details of the different works.

3. Atlas estimation

3.1. What is a good atlas? The main goal of estimating an atlas is to transport (thanks to a deformation) information from the template onto a new acquisition giving insight about the second one. This requires a condition on the template. This image should be an image of an ideal shape capable of characterising a whole sample of images from its class but only its class. This implies a noise free and sharp image defined on the whole domain of interest.

Another fundamental quantity is the "metric" between images which has to take into account the geometrical structures the image is showing. In particular, this metric is related to the choice of the regularity of what will be defined as normal variations. On the one hand, a very smooth mappings would prevent from matching thin structures. On the other hand, too few constrains allow to take into account small details but generate over-fitting and may prevent from discriminating the images. A tradeoff has to be found to focus only the informative elements.

The choice of these two elements has to be made jointly as the constrains on the mappings will force the template to be more or less blurry and the other way around, the level of details of the template induces more or less regularity for the mapping.

A last choice is how to set the weight σ^2 which balances the two parts of the energy defined in Eq: (2.3). As we will see later, this parameter "measures" the noise rate in the data and possibly model mispecifications.

In the well-known variational approach of the registration problem, the previous quantities have been left to the user's expertise. However, it is natural to think that they are population dependent and should be learnt together. This is what the Bayesian Mixed Effect Template (BME)

model estimations and their generalisations aims at producing.

My PhD contributions will be summarised in Section: 4. The improvements and generalisations of that work will be detailed in the three axis of my following up work, Sections: 5, 6 and 7.

4. My PhD contributions

4.1. Dense registration method based on control points [C1]. For this first contribution, let us come back to the deformation group construction and the properties of the geodesic paths arising from landmark matching problems. We consider the framework introduced and developed in [**55, 56, 73**] where the deformation is assumed to be driven by a finite number of landmark displacements $q_t = (q_t^k)_{1\leq k\leq n}$ with known initial positions $q_0 = (q_0^k)_{1\leq k\leq n}$.

The velocity vector field at each time point is assumed to belong to a Reproducing Kernel Hilbert Space (RKHS) V given by its kernel K. In this setting, one can show that the optimal (as the minimiser of the trajectory kinetic energy $\int_0^1 \|v_t\|_V^2 dt$) dense velocity field v_* along the trajectory has a finite dimensional representation $v_{*,t}(.) = \sum_{k=1}^n K(., q_t^k)p_t^k$ where $p_t = (p_t^k)_{1\leq k\leq n}$ appear as momenta variables. Going a step further, and introducing the Hamiltonian $H(q,p) = \frac{1}{2}p^t K(q)p$, the finite dimensional Hamiltonian variables (q,p) evolve according to the usual Hamitonian system:

$$\left\{\begin{array}{rcl} \dfrac{dq_t}{dt} & = & \dfrac{\partial H}{\partial p} = K(x,.)p \\ \\ \dfrac{dp_t}{dt} & = & -\dfrac{\partial H}{\partial q}\,. \end{array}\right. \tag{2.4}$$

Given the initial momentum p_0 at time $t = 0$ such that $v_0(x) = \sum_{k=1}^n K(x, q_0^k)p_0^k$ and template landmarks q_0, the geodesic path is reconstructed for all $t \in [0,1]$ using the following interpolation equation which provides the time dependent velocity vector field v:

$$v_t(x) = \sum_{i=1}^n K(x, q_t^i)p_t^i\,, \tag{2.5}$$

and the flow equation (2.2). This yields a finite dimensional parametrisation by the initial momentum p_0 of the space of geodesics denoted here $(\phi_t^{p_0})_{t\in[0,1]}$.

The solutions of inexact landmark matching problems satisfy, in Hamiltonian variables (q,p), the so called transversality equation, on the final momentum p_1:

$$p_1 + \frac{1}{2\sigma^2}\nabla_{q_1} g = 0\,, \tag{2.6}$$

where g is the data attachment term depending only on the final position q_1. My contribution was to solve this equation in the dense image setting using a second order numerical algorithm (Newton's method) which accelerates the convergence of the estimation. To this purpose, we had to adapt the dense image data attachment term so that it only depends on the final control point positions q_1.

Note that the theoretical framework of this Hamiltonian point of view has been more and more developed [**53, 55, 38**] and will be the base of one part of my following development **[A8, A14]**.

In this work, the choices of the template and the metric on the vector field space were arbitrary. This made us move to the atlas estimation issue described in the following.

4.2. Bayesian Mixed Effect (BME) Template model for atlas estimation [A1, A2].

4.2.1. *BME Template Model and MAP Estimator **[A1]**.* The Model we proposed during my PhD takes place in the small deformations framework for deformable templates. We proposed a generative statistical model to learn the parameters given by the template, the metric of the set of deformations which induce the metric on the image space and the noise variance. The observed images are assumed to be Gaussian distributed conditional to the deformations We define the template image and the velocity vector field v (cf Eq :(2.1)) through parametric spline models which enable to reduce the dimension to a finite dimensional problem using control points (in the same spirit of the use of control points in Section: 4.1). The template image is encoded by a vector $\alpha \in \mathbb{R}^{k_p}$ as well as the deformation field by a another vector $\beta \in \mathbb{R}^{2k_g}$, together through two interpolation kernels.

Since we were looking for a global geometrical behaviour and a generative model, we assumed that the geometrical variables β are random variables drawn through a Gaussian distribution with zero mean and a covariance matrix that has to be estimated. The deformations are thus considered as random non observed variables. The covariance matrix is part of the model parameters and does not have any particular pattern of zeros. This enables to estimate both local and global correlations between control point movements. Indeed, the deformations are likely to be a global transformation of the shape drawn on the image. Therefore, to allow for any such correlated map, the covariance matrix has to be as few constrained as possible.

Note that this matrix can also be related to the metric used on the vector space V as it parametrises the kernel which defines the inner product of this Hilbert space of deformations.

We consider a Bayesian framework since our medical image problem falls typically in the High-Dimension-Low-Sample-Size (HDLSS) setting. The advantage of introducing priors on the parameters turns out to be of particular interest as it regularises the estimates and enables to introduce the knowledge people acquired by experiments on the matching problem into the model.

The estimation of the parameters -template parameter, covariance matrix and noise variance- called θ is done by maximising the posterior likelihood of the parameters given the observations: let θ be our parameters to estimate, then:

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmax}}\, q(\theta|y_1, ..., y_n)\,,$$

where we use q to denote likelihoods.

We have proved that the Maximum A Posteriori (MAP) estimator exists given a n-sample and that it is consistent in the sense that it tends to an optimal value of the parameter in the Kulback-Leibler sense under the "true" probabilistic law of the observations when the sample size tends to infinity.

Let $\mathbb{P}$ be the distribution governing the observations and Θ the set of all possible parameters. Let then Θ_* be the set of distributions coming from a model "close" to P:

$$\Theta_* = \{\ \theta_* \in \Theta \mid E_{\mathbb{P}}(\log q(y|\theta_*)) = \sup_{\theta\in\Theta} E_{\mathbb{P}}(\log q(y|\theta))\}.$$

Theorem 1 (Consistency while the sample size tends to infinity, **[A1]**). *Assume that Θ_* is not empty. Then, for every compact set $K \subset \Theta$,*

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

where δ is the metric compatible with the usual topology on Θ.

We proved that under some other weak assumptions the set of solutions is not empty and the solutions themselves are almost surely in a compact :

Theorem 2 (Consistency on bounded prototypes, **[A1]**). *Assume that $P(dy) = p(y)dy$ where the density p is bounded with exponentially decaying tails and that the observations y_1^n are i.i.d under P. Under weak conditions on the template and deformation variable dimension, the space $\Theta_*^R\ (= \Theta_* \cap \{bounded\ templates\}) \neq \emptyset$ and for any $\epsilon > 0$*

$$\lim_{n\to\infty} P(\delta(\hat{\theta}_n^R, \Theta_*^R) \geq \epsilon) = 0\,,$$

where δ is any metric compatible with the topology on Θ^R.

4.3. Statistical Learning Procedure: [A1]. To compute the MAP, we wanted to use the usual Expectation Maximisation (EM) algorithm [**23**] which was designed specially to handle observed likelihood maximisation in such missing data models. However, like in many cases where the EM algorithm should be used, we could not run it directly as the expectation with respect to the missing data posterior law (E-step) does not have a tractable form.

In **[A1]**, we have proposed an approximation of this law by a Dirac distribution on its mode. This approximation enables to carry out this "EM-like" algorithm easily and gives a fast procedure. The final process is reduced to two maximisations done alternatively. First, we compute the best match between each training sample image and the current template with the current metric and noise variance. The second step maximises with respect to the parameters θ the complete log-likelihood evaluated on these particular values of the deformation fields. In addition to this easy implementation, as already mentioned, this approximation links this approach to the variational approach previously described: this uses the same loops with an additional update: the geometrical covariance matrix which was fixed by the user in the variational approach and sometimes analysed a posteriori of the template estimation to describe modes of variability using a PCA decomposition for example.

4.4. Multicomponent Model: [A1]. The model was then generalised to the case where the data are heterogeneous and would benefit from being clustered. The context is the one of unsupervised learning as the label of each image is not known and has to be automatically assign by the algorithm. This led to a model where both the deformations and the image labels were considered as missing variables. The parameters were therefore extended to the component dependent templates, metrics and noises as well as the weight of each sub-group. The estimation of these parameters was done in the same way by maximising the posterior likelihood in a Bayesian framework, using the same approximation of the posterior distribution of the deformations as above.

An approximation of the posterior weights of each class given the observations were also approximated using the best deformations.

Experiments were performed on the USPS handwritten digit database. The results were relevant for both the photometry (accurate and sharp templates, good approximation of the noise variances of such images) and the geometry (relevant deformations can be drawn through the estimated laws). Although not a new classifier, this statistical model enables to perform classification tasks as a quantification of the validity of the results. Classification performances were competitive with the state of the art and even better when considering only very small training sample size. All these results are detailed in **[A1]**.

4.5. Stochastic Algorithm and Convergence study [A2]. The previous algorithm is based on the guess that given an observation, the probability distribution of the deformations of the template to match this image is well-approximated by a Dirac measure on its mode. This assumption seems realistic as long as this distribution is smooth, unimodal and symmetric. Unfortunately, although this sounds reasonable for ideal images, this becomes questionable when dealing with real ones (and even more as soon as the data are noisy). In **[A1]**, a toy example was presented to show that this expected behaviour appears even in the very simple context of translation of a step function. This can be explained by the fact that the posterior distribution is in such case likely to be irregular, spiky or multimodal. Selecting a mode of the distribution (subject to the fact that we never know if it is the global maximum or a local one) leads to choosing a single non realistic deformation as the best match and therefore to construct a non relevant atlas.

In **[A2]**, we have therefore proposed a second algorithmic approach based on a statistical variants of the EM algorithm treating the case of the single component model. We used the Stochastic Approximation EM (SAEM) algorithm introduced by [**21**] coupled with Monte Carlo Markov Chain (MCMC) methods to face the problem of the complex expression of the posterior distribution. This coupling was introduced in [**45**] and proved convergent for compact support hidden data. In **[A2]**, we extended this convergence results to unbounded hidden variables introducing weaker conditions on the transition kernel of the Markov chain. We also proved that our MCMC method, a hybrid Gibbs sampler, in the particular case of our model, was also convergent.

Theorem 3 (Convergence of the coupled procedure **[A2]**)**.**
Under some assumptions, let $(\theta_k)_{k\geq 0}$ be the sequence of estimated parameters. Then, for every initial point of the Markov chain β_0 and initial value of the parameters θ_0 set in two given compacts, we have $\lim_{k\to\infty} d(\theta_k, \mathcal{L}) = 0$ $\bar{\mathbb{P}}_{x_0,s_0}$-a.s, where $\mathcal{L}$ is the set of critical points of the observed log-likelihood, and $\bar{\mathbb{P}}_{x_0,s_0}$ a probability measure associated to the Markov chain.

4.6. Stochastic algorithm for the multi-component model [A3 part1]. Concerning the multicomponent case, extending the one-class-model MCMC method was not possible for numerical failures as the sampling was facing numerical trapping states. We proposed to go back to the basic SAEM algorithm and tried to simulate the missing variables through a distribution as close as possible to the posterior distribution.

The experimental results are similar to those found with the mode approximation method when the two of them numerically converge (non noisy data). For noisy data, the results of the stochastic algorithms are much better and this difference increases with the noise variance of the training images.

In the same way as we have used a stochastic method for the learning procedure, we also proposed several ways to face the classification problem which involves the same integration with respect to the missing variables. This was previously done by an approximation of expectation by its highest value (mode approximation). The solutions we suggested were based on different sampling schemes to approximate by Monte Carlo sums the expectation. This led to the same behaviour as the estimation part where noisy observations have to be handled with stochastic algorithms to get better results.

In the following sections, I describe the research projects I have developed since my PhD defence which can be articulated around three axes. More precisely, (1) I have extended the BME-Template model to increase its use in medical image analysis where the high dimension of the data forces to optimise the stochastic algorithm. Moreover, medical data are now acquired with different modality and the segmentation of brain tissue is an important challenge. (2) I also extend the analysis of the deformations in populations both globally and locally and (3) looked for personalisations some mathematical models to understand patient specific pathology evolution.

5. Axis 1: Generalisation of the BME Template model and of its stochastic estimation

Along my PhD thesis, I have developed a statistical generative model in order to explain a population of shape images though a mean (in the space of shapes) and what we can relate to a covariance (geometric covariance through the deformation covariance matrix and photometric covariance through the residual variance). This model was based on the deformable template model and generalised to heterogeneous data.

However, as mentioned in Section 2, coming back to the medical application, several aspects of this analysis had to be improved both in terms of modelling and numerical performances. Indeed, dealing with medical images implies to consider the anatomical constrained such as topological stability in the population. The “small deformation” setting did not ensure that the simulated deformations were preserving the topology of the shapes captured by the images. Holes or overlaps are not prohibited. This constraint should be included into the model. Moreover, brain images carry an important information about tissues in the grey levels. Hence, the statistical model should be able to consider this difference in grey level to produce a probabilistic template which provides the probability of the tissues at each voxel. Furthermore, more and more acquisitions are based on several modalities (T1-MRI, diffusion, fMRI, etc). These available information should be included in the model to be able to increase the accuracy of the estimation of each modality by taking into account their correlations. On the numerical side of the estimation, the high dimension of the data obliges us to optimise the stochastic algorithm in order to achieve computational attractive performances keeping theoretical convergence properties. This is the core of the following works.

5.1. Convergence of the multi-component stochastic algorithm [A3 part2]. First, the stochastic algorithm used for the multicomponent model did not fit into the condition of convergence of Theorem 3. Therefore, we prove in **[A3 part2]** that the sequence of parameters generated thanks to the proposed MCMC scheme converges almost surely towards a critical point of the observed likelihood.

This ensures the convergence of our estimated sequence as well as relaxes the SAEM algorithm constraints while keeping its convergence property.

5.2. New MCMC method to boost numerical performances [A12]. The bottleneck of the estimation algorithm in terms of computational cost is the MCMC method. Indeed, because of the large dimension of the random hidden variable, the hybrid Gibbs sampler (also called Metropolis Hastings within Gibbs) shows the best efficiency. It enables to explore the large dimension support of the random variable keeping a reasonable acceptance rate. The other samplers, as symmetric random walk, propose candidates which were very likely to be rejected (as a non interesting deformation). For this reason, the Gibbs sampler which samples each coordinate of the high dimensional vector one at a time, is visiting the support while still proposing relevant deformations (w.r.t. the target distribution).

Unfortunately, the power of this method, looping over the random vector dimension, is also its weakness. The computational time required by a Gibbs sampler increases linearly with the dimension of the vector, and becomes unrealistic when dealing with 3D images as the length of the deformation vector also increases. This is even more critical as soon as the acceptance ratio requires at each iteration of the loop has a heavy computational cost.

For this reason, we develop in **[A12]** a new MCMC sampler based on the Metropolis Adjusted Langevin Algorithm (MALA) where the proposal follows the anisotropy of the target distribution. This sampler is optimised in order to be used inside the SAEM estimation algorithm. To this purpose, it is not desirable to have an adaptive sampler as the adaptation is already produced by the estimation iterations leading to a non homogeneous Markov chain.

The expectation of the proposal of our Anisotropic MALA (AMALA) is obtained as the sum of the current iterate plus a drift which is proportional to the gradient of the logarithm of the target distribution. We construct the covariance matrix as a regularisation of the Gram matrix of this drift which replaces the isotropic diagonal covariance matrix of the MALA proposal. This new sampler can be seen as a stochastic version of the deterministic algorithm we used in **[A1]**. Indeed, the AMALA sampler proposes a candidate in the direction of the (increasing) gradient of the posterior likelihood. This gradient is in fact the (opposite) gradient of the energy that we minimised to get the optimal deformation at each iteration of the deterministic algorithm used in **[A1]**. The interest in the AMALA is that it explores the neighbourhood of this gradient direction (the drift) using a random perturbation given by the covariance matrix of the Gaussian proposal which is even though carrying the local anisotropy and correlations of the deformations. Allowing for a random search around an interesting value enables to stride the support of the deformation and take into account all possible deformation with their probabilities.

We prove the geometric ergodicity of the AMALA sampler uniformly on any compact set assuming some usual regularity conditions on the target distribution. We also prove the almost

sure convergence of the parameter estimated sequence generated by the coupling of AMALA and SAEM algorithms (AMALA-SAEM) toward the maximum likelihood estimate under some regularity assumptions on the model. Moreover, we prove a Central Limit Theorem for this sequence under weak conditions on the model. Note that the central limit theorem also applies for the previous MCMC sampler.

5.3. Diffeomorphic deformation BME Template [A8, A14]. The generalised model allowing for multi-component population was not accurate enough to deal with medical anatomical images. Although the cluster found in the observed populations were convincing and most of the time separating images with different topologies, no specific requirement in the model was ensuring that this topological constraint was in fact respected. Therefore, we specify the BME Template model to take into account diffeomorphic deformations. We use the LDDMM framework which was promising specially when considering the "landmark" (more precisely control point) driven deformation. However, we do not only introduce the diffeomorphic constraint into the model, which would certainly have been interesting. But we also take the benefit of the rich LDDMM theoretical framework.

As a first step towards introducing the diffeomorphic model into the BME Template stochastic estimation, we try a variational approach to see the benefit of such framework. In **[A8]**, we use an interesting property of the LDDMM framework when the deformation is driven by a set of control points. This property comes from the evolution equations, specially the Hamiltonian system (2.4). Indeed, the deformation is not only characterised by the initial momenta p_0 but also the initial positions of the control points q_0 carrying these momenta. This enables to consider that the optimisation of the matching between two images can be more precise and informative if we optimise not only the initial momenta but also the initial positions of the control points. This can be performed easily by assuming that our matching energy is a function of both variables. To perform the atlas estimation, we minimise the sum of the contribution of each observation through a gradient descent in both the atlas and the state of the deformation system (q_0, p_0). In **[A8]**, we show that the computational cost of this refinement was negligible as the computation of the energy gradient with respect to the momenta includes -because of their coupling evolution (2.4)- the computation of the one with respect to the control point initial positions.

This analysis shows another advantage. The energy we consider is a function of two very high dimensional (denoted k_{qp}) vectors, the control points q_0 and their momenta $(p_0^{(i)})_{1\leq i\leq n}$ where n is the number of observations, most often much smaller that k_{qp}. Moreover, the LDDMM deformation shows a lot of redundancy as soon as two control points are closer than its intrinsic scale (given by the scale of the kernel K in (2.4)) or if they point towards the same direction. For these reasons, it is natural to think that in real cases, the dimensionality of the observed variability has a small number of degree of freedom. Therefore we assume that the observed variability can be approximated by an "optimal" number of control points which should depend on the observations. In **[A8]**, we use the well-known l^1 penalty on initial momenta to force some of the momenta to vanish when matching the template to the observations. Since the control points are also allowed to move, the optimum is achieved for a reduced number of these points localised at strategic positions. In the experiments, the expected behaviour of the moving control points is observed: they

tend to move where the deformation is important and mostly near the boundaries of the shapes.

Now that the variational approach shows the interest of the LDDMM framework, the optimisation of the position and number of control points, the goal is to include them into the BME-Template model and use a stochastic estimation so that we would benefit from its theoretical and numerical convergence properties. In **[A14]**, we complete the statistical model to reach this goal and perform the atlas estimation. The atlas here contains in addition the reference control point positions. The gain is also noticeable as we use the previously described AMALA algorithm. The computation of the gradient with respect to the initial momenta (missing variable) provides us the gradient with respect to the control point initial positions. Therefore, the sampling of the couple of vectors (q_0, p_0) can be performed in a single run leading to a very efficient algorithm. Because of the very complex dependency of the diffeomorphic deformation (cf Eq. (2.4), (2.5) and (2.2)) with respect to the state of the system (q_0, p_0), some assumptions on the model required by the convergence theorems proved in **[A12]** cannot be checked. Therefore, no convergence of the estimated sequence can be guarantied with this estimation algorithm although reasonable to expect.

5.4. Probabilistic tissue atlas estimation [A13]. In parallel to the introduction of a diffeomorphic constraint, we also generalise the model so that it deals with the medical imaging requirements. The first important issue is that in brain images, the grey levels correspond to different tissue types. Therefore, the segmentation of the brain plays an important role in brain image analysis. Moreover, even with very recent scanners, the resolution of the acquired images is still very low compared to the tissue boundary scale. This produces voxels at tissue boundaries with a grey value which may be mixtures between the grey values of the different tissues it represents. This is called partial volume effect (PVE). For this reason, estimating a probabilistic atlas of each tissue, taking into account the geometric variabilities of the brain in the population brains and the PVE is an important challenge. In **[A13]**, we propose to extend our BME-Template model to handle these constraints.

Given a set of grey level MR images from different subjects, we estimate a probabilistic atlas which includes probability maps of each tissue of the brain, the covariance metric of the deformations and the distributions of the grey levels of each tissue. We introduce the voxel tissue type as a second missing variable for each patient. The estimation is performed by a MCMC-SAEM as presented above (using a hybrid Gibbs sampler because of the non differentiability of the model with respect to the deformation parameter) and proved to converge. An important aspect of this model is that the segmentation and registration are considered dependent; this enables to take advantages of both to improve the accuracy of each of them.

This method enables to get as output the individual segmentations of the observed brains along with the probabilistic atlas. Moreover, it enables to segment new MR images as we are able to map our probabilistic atlas onto the new observations and transport the tissue information.

5.5. Multimodal brain atlas [C11, A15]. The second step towards a full atlas of the brain is to take into account the multi-modal information we have for several patients. For example, functional MRI (fMRI) becomes a key modality in order to localise functional areas in the cortex and deep nuclei by measuring MRI signal changes associated with the hemodynamic response which is assumed to be correlated with neural activity. To deal with this couple of images for each

subject, we generalise in **[C11, A15]** the probabilistic model of Section 5.4, in order to create a probabilistic atlas combining the probabilistic templates of each tissue with the degree of activation on grey matter (GM) voxels- as well as the geometric variability of the population under study. The input is now a vector which encodes multimodal patient observations (grey level T1 and f -MR images). The interest interest of this model is that we estimated active areas only conditional to GM segmentation where the hemodynamic response actually occurs. We also performe the estimation by coupling the segmentation and registration steps. In this analysis, the output of the algorithm is of great interest as it includes the probabilistic atlas, the individual active areas and the means and variances of each tissue type in each modality.

5.6. On going works. In the same spirit of the MRI-and-fMRI atlas estimation, we also investigate the introduction of Diffusion Weighted images to incorporate information both in grey and white matters (**[A19]**). DW images correspond to the response of the water molecules to an additional magnetic field. The idea is to detect the brain fibres thanks to the measure of the water molecule movements while applying several different magnetic fields. Therefore, DW images are at least acquired for 7 directions and nowadays, we reach about 50 directions of acquisition. These directions of the extra magnetic field are called gradients. The gradients are vectors which discretise the unit sphere. Therefore, this task is more complex as it requires to process a stack of dozen of images at once.

In order to estimate a DWI atlas, it is important to model the set of images for each patient as a discretisation of an element of $\mathcal{S}^2 \times \Omega$, where $\mathcal{S}^2$ is the unit sphere in $\mathbb{R}^3$ and Ω the image domain in $\mathbb{R}^3$. This enables to get a dense representation on the continuous domain. Indeed, each patient may have been scanned with different gradient directions. This requires to be able to deal with any direction on the sphere. The goal is to be able to reconstruct a template of Diffusion Tensor Images (DTIs) [**20**] or representation of this water motion variability with higher order descriptors such as Q-balls [**72**]. These descriptors are estimated using the DWIs with more or less gradient directions in order to get a visualisation of the fibres where the water molecules are constrained. This requires to be able to sample any number of gradient response and force to have a dense estimation on the domain $\mathcal{S}^2 \times \Omega$.

This is still work in progress, specially the implementation as it requires to be able to deal with a large number of images (more or less $n \times 50$ where n is the number of subjects in the population and each has been submitted to 50 gradient directions).

Another question which remains is the quantification of the bias the deterministic approach induced and its dependence with respect to the noise variance of the data. We are investigating this direction in **[A17]** where the question is does the orbit of the Frechet mean of the data under the deformation group action coincides with the Frechet mean of the orbits?

At the end of this investigation, we hope we will be able to estimate for different populations their full brain atlas including all available modalities. This sounds promising for several applications. First, the different populations may show differences from each other on either the anatomical or diffusion or functional templates. Their geometrical variability will also be more accurate as it will take into account not only one modality but the whole information forcing the

deformations to fit in all brain regions. Then, when a new patient is scanned for example only with MRI, then the information available on the atlas could be mapped onto its scan and reflects interesting characteristics. This paves the way to both a better understanding of the brain and its pathologies and the early diagnosis transporting the diagnosis from a known template onto the new patient.

6. Axe 2: Geometrical variability statistical analysis

In the BME-Template model, the deformations are analysed through the covariance matrix of their distribution. Although this provides an interesting knowledge about them, one may want to interpret them in other way for example exhibiting a small number of discriminative patterns or long-distance correlations between subgroups of control points. This is what have motivated the following works.

6.1. Probabilistic Independent Component Analysis [A5]. Independent Component Analysis (ICA) is a statistical technique that aims at representing a data set of random vectors as linear combinations of a fixed family of vectors with statistically independent coefficients. It was initially designed to solve source separation problems in acoustic signals [**14**] and rapidly found a large range of applications in particular in medical image analysis [**15, 16**] where ICA has become one of the standard approaches. In the medical imaging analysis, the data are high dimensional but have small to moderate sample size, which complicates statistical analysis. A common way to address this issue is to apply dimension-reduction techniques to reduce the information to a smaller number of highly informative statistics. Probabilistic ICA can be used for this purpose, and in many cases, the representations it provides are qualitatively very different from those obtained using decorrelation methods such as principal components analysis (PCA) [**78**].

The numerical method described in **[A5]** estimates the maximum likelihood estimator associated to the probabilistic ICA model, where the likelihood is for the observations, therefore averaging over the unobserved independent components. This differs from the solution which is often adopted in the literature, which solves the parametric estimation and hidden variable (independent components of the model) reconstruction problems at the same time. This is the same context of missing data model as the BME-Template was exhibiting. The numerical algorithm to solve this estimation problem is also based on the MCMC-SAEM algorithm.

An advantage of our learning algorithm is that it applies to many different probabilistic distributions. There are almost no restrictions to the range of statistical models that can be used for the unobserved independent variables. As examples, we present in **[A5]** different models that all fit into this same framework, but which correspond to different statistical contexts. We introduce both continuous probability distributions with different tail behaviour and distributions with one or several atoms so that the independent coefficients may be switched off or describe an activation / inhibition process. This introduces the sparsity constrain in the model to be able to treat not only medical images but also real images.

We apply this analysis to the deformations of a hippocampus template onto 101 observed and segmented hippocampi. We exhibited 5 characteristic patterns of deformations which were

significant in terms of discriminating Alzheimer's disease (AD patients) at two different stages from control patients.

6.2. Long distance conditional correlations [C5]. It is quite natural to think that any region of the brain depends on all the other regions at least via some detours. This is equivalent to say that these regions are all correlated to each other. Several studies [**41, 47, 12**] have tried to design correlation patterns by computing the correlation matrix given by correlations between any two regions (using a PCA decomposition for example). Then, they highlight the most significant or stable ones by thresholding. However, correlations describe the global statistical dependencies between variables, corresponding to both direct and indirect interactions. The *direct* relations between two of these regions are less numerous and harder to capture but they carry some interesting information as well. Focusing on these direct dependencies, called *conditional correlations*, enables to see which areas directly affect the behaviour of a given region and avoids the burden given by the indirect interactions.

In **[C5]**, we study the geometric conditional dependencies of the amount of deformation of a template hippocampus to subjects carried by a discrete grid. A natural approach to estimate these conditional correlations is given by Gaussian Graphical Modelling (GGM) [**46, 52, 31**]. The response is modelled as a Gaussian random vector. The conditional correlations of the response variables are depicted by a graph. Each node represents a variable (grid point intensity) and an edge in the graph is set between two variables if they are conditionally dependent given all the remaining ones. This corresponds to a non-zero entry of the inverse covariance matrix called the precision matrix.

In the context of medical data, we face two problems with Gaussian Graphical Modelling. The first one is the high dimension of the data compared to the low sample size. The second problem is that in this framework, the neighbour points are very likely to be conditionally correlated since the medical phenomenon is usually continuous and these local dependencies can be predominant in the estimated graph possibly hiding any other links. We propose an algorithm and a selection criterion which address both questions.

To face the high-dimension-low-sample-size paradigm, sparse representation is well known to be a very powerful tool. Although we know that the underlying real graph is not sparse, it is interesting to perform a *sparse estimation of its structure*. Such an estimation provides both a stable estimate (with respect to the noise in the training set) of the conditional dependencies and a selection of the edges with the highest conditional correlations. This sparsity constraint is crucial as it prunes the graph keeping a small number of edges.

As noticed above, local dependencies of medical images are predominant while estimating the conditional correlations. We go beyond that and focus on long-distance dependencies introducing this knowledge as a prior to the estimation. The prior knowledge is modelled as a neighbouring graph provided by the user. Because of this, it can take into account geographical closeness or some anatomical proximity given for example by fibres. The estimation of the remaining graph is done as for a classical sparse GGM.

To avoid subjective choices of free parameters, we generalise a previously designed criterion to the case of neighbourhood prior. This criterion enables to select the most representative graph in terms of its prediction power.

This method was applied on the template-to-subject mapping we introduced previously to complete their analysis. It appears, on the preliminary results, that the AD patients "lose" long-distance conditional correlations which would suggest that their brain has a more random shape compared to the template.

7. Axe 3 : Personalisation of mathematical models to understand pathology evolutions

Previous to any statistical group analysis, one has to study the validity and the power of the mathematical models we propose. To that purpose, one has to find what are the parameters of interest in the models and whether they provide a informative understanding of the pathology on separated patients. This is known as personalisation of mathematical models. In the following, I focus on two dynamical models for brain tumours and stroke to see their impact on explaining and predicting the disease evolution.

7.1. Reaction diffusion model for brain tumour analysis [A6]. Brain gliomas represent about 50% of all primary brain tumours and can be classified according to their grade of malignancy. Low grade gliomas (LGG) are slow invaders of brain tissue as they keep growing for many years, presenting one of the most controversial decision treatment areas. High grade gliomas (HGG) remain unfortunately incurable with an average life expectancy of one year after its discovery. The diagnosis of brain gliomas includes the analysis of various MRI sequences of the brain which partially reveal the tumour invasion. Based on those images and other clinical information, neurologists try to determine the grade of the gliomas and to estimate their current and further spatial extent and if possible their source location.

For more than a decade, mathematical models of brain tumours have been devised to help clinicians answer these questions. Microscopic models study the cellular mechanisms and macroscopic models describe the evolution of tumour cell density. More realistic reaction-diffusion models have been proposed based on the fact that tumour cells migrate faster on white matter fibre tracts myelin sheaths.

A key issue for those models to answer clinical questions is their personalisation, *i.e.* the estimation of some patient-specific parameters from medical images. The main parameters to be identified based on reaction-diffusion models are a combination of tumour diffusion tensors in white and grey matter, its proliferation rate, its initial point and its initial time.

In **[A6]**, we aim at characterising the nature of the glioma, more precisely LGG, from a single MR image. Indeed, we hypothesised that the tumour shape is dependent on the proliferating or infiltrating nature of the tumour. Given a segmented brain glioma from an MR image, we solve an inverse problem in order to estimate the diffusivity ratio and the tumour source position. By localising the tumour source and estimating the invaded tissue characteristic using this ratio, our objective is to provide clinicians with new indices that can be used for diagnosis from the first acquired MR images, combined with a subsequent prediction of tumour invasive margins as it grows from the initially observed boundary. This additional information may help in surgical and/or radio-therapeutic treatment planning especially when it comes to determining the margins of the brain matter to take out following craniotomy or to radiate.

Additionally, after solving the inverse problem, we analyse the location of tumour sources, and their stability over time. Also, knowing the diffusivity ratio and the tumour source from a

single image, we evaluate whether this information gives insights into predicting further tumour shape evolution.

7.2. Stoke analysis: from acute to subacute stages [A9, A11, A16]. The spatio-temporal evolution of stroke lesions, from acute injury to final tissue damage, is complex. Although time sequences of multimodal images are acquired, there was no 4D dynamic model simulating the continuous evolution of ischemic stroke using MR observations **[A9]**.

Acutely, the diffusion-weighted (DWI) and perfusion- weighted imaging (PWI) detection of early ischemic changes led to the concept of the "perfusion- diffusion mismatch" hypothetically reflecting salvageable tissue, also known as the "ischemic penumbr". The mismatch concept assumes that the lesion core tends to grow into the observed mismatch region without treatment till reaching the final damaged tissue extent depicted by T2-w imaging at > 1 month. To better investigate the DWI-PWI mismatch concept, its key assumptions and prognostic potentials, we determine in **[A11]** patient-specific 4D evolution scenarios for both PWI (measured as Mean Transit Time, MTT) and DWI lesions from acute to subacute time-points by estimating a current-based diffeomorphic deformation for each scenario.

Knowing that the mismatch concept holds two main hypotheses: first, DWI areas with high expansion correlate with corresponding PWI areas with no dynamic change and second, DWI areas that does not expand correlate with corresponding PWI areas with high contraction, we introduce in this paper a mathematical function to spatially define the key term "'corresponding" and check the validity of these mismatch-driven hypotheses.

On pursuit of a better understanding of the mismatch evolution concept, we determine the spatial coherence between the estimated DWI/MTT time-evolving lesion boundary and the final T2-w and estimate the time-point (in hrs from stroke) when the DWI/MTT lesion boundaries are spatially closest to the final T2 lesion.

The evaluation of the spatio-temporal patient-specific perfusion-diffusion evolution scenarios showes very promising results as they fit into the manually-delineated lesion boundaries. Meanwhile, the depicted dynamic spatio-temporal contraction/expansion behaviours do not show any common kinetic evolution pattern thus not confirming the perfusion-diffusion mismatch evolution hypotheses commonly used in planning stroke treatment. Looking at the final tissue outcome, 4D modelling of acute ischemic lesions reveales large inter-patient spatial-temporal variation in optimal agreement and its timing of DWI-PWI vs final T2 lesions, questioning the validity of the assumptions underlying the perfusion-diffusion mismatch hypothesis.

The use of diffeomorphic model in that work cannot handle lesions with topology change such as fragmented lesions with a time-varying number of components. This motivates us to develop a dynamic model to fill in this gap in stroke research field as scattered patterns of perfusion and diffusion progression are very common **[A16]**. We explore a two-image based registration model including both diffeomorphic deformation and also photometric change in image evolution (the part that handles topology change) presented introduced by A. Trouvé and L. Younes [**71**] called metamorphosis.

Our goal is to extend the theory of metamorphosis to handle longitudinal data and to estimate a spatiotemporal evolution scenario of fragmented DWI lesion, given a time sequence of stroke-bearing DWI images from the same individual. This allows us to retrieve the dynamics of

DWI lesion evolution from acute to subacute stages taking into account in between observations and to look at a fundamental question in understanding stroke evolution process: “Does acute perfusion abnormality influence the way DWI lesion evolves?”. In this work, we investigate the predictive value of MTT perfusion map in determining DWI infarction evolution and to improve the definition of oligaemic (hypoperfused tissue that is not ischemic yet), at risk of infarction and dead tissue using the DWI dynamic measures that we derive from the estimated spatiotemporal metamorphosis.

CHAPTER 3

Around the BME-Template model

*This chapter summarises the works presented in **[A1, A2, A3, A4, A7, A8, A10** and work in progress and future directions contained in **A12, A13, A14, A15, A17, A18, A19]**.*

The model we proposed in **[A1]** during my PhD was based on simple assumptions and was a first attempt to estimate an atlas of a given population of images. Although providing interesting results, we investigated the ways to improve it so that it would take into account both the geometrical and photometrical issues arosen by the medical application, such as topology preserving, online segmentation of brain tissue, multi-modal data, etc. Moreover, the estimation process had to be improved as well in order to deal with high dimensional 3D data.

Contents

1. BME-Template model [A1,A2]

For sake of clarity, let us recall the model introduced during my PhD and presented in [**2**] and on which the following works are based. This model is a statistical generative model based on the deformable template theory developed by Grenander in [**35**] which tries to approximate the generation of observed images as smooth deformations of an ideal characteristic image called template. The Bayesian Mixed Effect (BME) Template model formulates this idea in the context of missing data variables. The atlas, composed of both the template image and a quantification of the geometrical variability, is common to all obervations and refered as the fixed effect. On the other hand, the deformations are related to each observation in a different way and are mathematical non observed tools which are considered as the random effects. As the targeted application is medical imaging, we work in the Bayesian framework to face the high-dimension-low-sample paradigm.

1.1. The observation model. We are given gray level images $(y_i)_{1\leq i\leq n}$ observed on a grid of pixels $\{r_s \in D \subset \mathbb{R}^2, s \in \Lambda\}$ where D is a continuous domain and Λ the pixel grid. Although the images are observed only at the pixels $(r_s)_s$ we are looking for a template image $I_0 : \mathbb{R}^2 \to \mathbb{R}$ defined on the whole plan. For each observation y, we assume the existence of an unobserved deformation field $z : \mathbb{R}^2 \to \mathbb{R}^2$ such that for $s \in \Lambda$

$$y(s) = I_0(r_s - z(r_s)) + \sigma\epsilon(s)$$

where $\sigma\epsilon$ denotes an additive noise.

Our model takes into account two complementary sides: photometry -indexed by p, and geometry -indexed by g. To work with finite dimensional elements, we parametrise both the template I_0 and the deformation z. They are assumed to belong to reproducing kernels Hilbert spaces V_p and V_g determined by their respective kernels K_p and K_g restricted to the subset of linear combinations of the kernels centred at some fixed control points in the domain D: $(r_{p,k})_{1\leq k\leq k_p}$ respectively $(c_k^g)_{1\leq k\leq k_g}$. This yields the following parametrisation by the finite dimensional coefficients $\alpha \in \mathbb{R}^{k_p}$ and $\beta = (\beta^{(1)}, \beta^{(2)}) \in \mathbb{R}^{k_g} \times \mathbb{R}^{k_g}$: $\forall r \in D$,

$$I_\alpha(r) = (\mathbf{K_p}\alpha)(r) = \sum_{k=1}^{k_p} K_p(r, r_{p,k})\alpha(k) \tag{3.1}$$

and

$$z_\beta(r) = (\mathbf{K_g}\beta)(r) = \sum_{k=1}^{k_g} K_g(r, c_k^g)(\beta^{(1)}(k), \beta^{(2)}(k)). \tag{3.2}$$

1.2. Statistical model. We suppose that all the data can be explained through the following statistical model: $\forall 1 \leq i \leq n$,

$$\begin{cases} \beta_i \sim \mathcal{N}_{2k_g}(0, \Gamma_g) \mid \Gamma_g \\ \\ y_i \sim \mathcal{N}_{|\Lambda|}(z_{\beta_i} I_\alpha, \sigma^2 \mathrm{Id}) \mid \beta_i, \alpha, \sigma^2 \end{cases} \tag{3.3}$$

where $zI_\alpha(s) = I_\alpha(r_s - z(r_s))$, for s in Λ. The parameters of interest are α, σ^2 - the variance of the additive noise - and the covariance matrix Γ_g of the variables β. This covariance matrix reflects the geometrical variability of the shape drawn in the images around the "mean" image which is the template. It captures the characteristic patterns of deformations. Moreover, as it does not have a

particular structure of zeros (as opposed to [**63, 49**]), it captures both local and global deformations through the correlations of the control point displacement. We assume that $\theta = (\alpha, \sigma^2, \Gamma)$ belongs to the parameter space Θ defined as the open set $\Theta = \{\ \theta = (\alpha, \sigma^2, \Gamma) \mid \alpha \in \mathbb{R}^{k_p},\ |,\ \sigma > 0,\ \Gamma \in \Sigma^+_{2k_g,*}(\mathbb{R})\ \}$, here $\Sigma^+_{2k_g,*}(\mathbb{R})$ is the set of strictly positive symmetric matrices. In the following, we call atlas the couple of the template image and the covariance matrix.

Even though the parameters are finite dimensional, the maximum-likelihood estimator can yield degenerate estimates when the training sample is small. Therefore, we introduce independent prior distributions on the parameters and work in the Bayesian framework.

This is the first model we dealt with and which showed a lot of advantages in spite of a complex interweaving of the parameters and missing variables. In particular, we were enabled to link this model to the well-known and very popular variational method. Moreover, this model enables to use the Maximum a Posteriori estimator which we proved to be consistent. The numerical aspect will be detailed below, taking advantage of the mixed effect setting.

Unfortunately, this model also shows some limitations in particular when targeting medical image applications. These are the aspects we investigated since 2007 and which are presented in the next sections.

2. Multicoponent case [A3]

In many situations, object classes can not be described as smooth deformations of one template. It is therefore natural to extend the model to include a clustering of the data. This is expressed by using a mixture of deformable template model. The parameters to estimate are now all cluster dependent atlases as well as the weights of each cluster in the population[1]. The observations are now grey level images from a non homogeneous population which has not been pre-clustered.

Let τ_m be the maximum number of components we expect in the training set. For each observation y_i, we consider the pair $\xi_i = (\beta_i, \tau_i)$ of unobserved variables which correspond respectively to the deformation field and the label of image i.

We introduce the following notation:

$$\eta = (\theta, \rho) \text{ with } \theta = (\theta_t)_{1 \leq t \leq \tau_m} \text{ and } \rho = (\rho_t)_{1 \leq t \leq \tau_m}\ ,$$

where $\rho = (\rho_t)_{1 \leq t \leq \tau_m}$ are the mixture coefficients. As before each parameter of the model θ_t is composed of a photometric part (α_t, σ_t^2) and a geometric part $\Gamma_{g,t}$.

[1] PhD contributions

We canonically extend the statistical model define in Section 1, adding a prior distribution on the mixture coefficients:

$$\begin{cases} \rho \sim \nu_\rho \\ \\ \theta = (\alpha_t, \sigma_t^2, \Gamma_{g,t})_{1\leq t\leq \tau_m} \sim \otimes_{t=1}^{\tau_m}(\nu_g \otimes \nu_p) \\ \\ \tau_1^n \sim \otimes_{i=1}^n \sum_{t=1}^{\tau_m} \rho_t \delta_t \mid \rho \\ \\ \beta_1^n \sim \otimes_{i=1}^n \mathcal{N}(0, \Gamma_{g,\tau_i}) \mid \ \tau_1^n,\ \theta \\ \\ y_1^n \sim \otimes_{i=1}^n \mathcal{N}(z_{\beta_i} I_{\alpha_i}, (\sigma_{\tau_i})^2 Id_\Lambda) \mid \beta_1^n,\ \tau_1^n,\ \theta \end{cases}$$

with $I_{\alpha_i} = \mathbf{K_p}\alpha_{\tau_i}$ and $z_{\beta_i} = \mathbf{K_g}\beta_i$ for all $1 \leq i \leq n$.

The goal is now to estimate the different cluster atlases given the non clustered training population of images. Note that the cluster atlases may differ from each other through either the template, the covariance matrix or both. This enables to treat data where the photometric mean may be equivalent but with different variability of the geometrical structures drawn in the images.

The following models are based on the previously described ones, however they include several requirements that arise when dealing with medical images.

3. Probabilistic atlas [A13]

3.1. Motivations and challenges. Medical image analysis often requires the accurate delineation of different tissue types, based on the contrasts observed in one or several images. The use of automated segmentation is important both for the sake of efficiency and of reproducibility. Moreover, tissue segmentation is an important information which enables to refine the atlas -template and geometrical covariance matrix again- estimation both concerning the location of each tissue and the accuracy of the deformations the brains are subject to. Therefore, we focus here on integrating an online segmentation in the atlas estimation to improve the accuracy of our estimates. Therefore, the observations remain grey level MRIs which have not been pre-segmented nor pre-registered. As these images are not binary, they show what is called partial volume effect (PVE): voxels -due to their resolution- contain several tissue types specially at the transition from one tissue to another. This leads us to choose to modify our model and to consider a probabilistic template model as it makes it possible to take into account this uncertainty on the underlying tissue type at each voxel of the template image.

Another fundamental issue is the coupling of registration and segmentation. Indeed, performing registration and segmentation jointly is generally more effective than performing them sequentially: the information conveyed by segmentation can improve the registration of the template to an individual image and reciprocally, individual segmentations are improved when an accurate registration from the template provides a model of the tissue types. As noticed in **[A2]**, estimating the deformation probability distribution together with the template (grey level one in **[A2]**) increases the population classification accuracy, as the model better fits the observations. In this work, we

use the same framework in which we consider a probabilistic template (instead of a grey level one) and an online coupling of the segmentation and registration of the observations.

We get as final output of the estimation process an estimation of both the probabilistic template (which contains the probability maps of each tissue in the template domain) and the geometrical variability. Although the individual deformation and segmentation are not parameters of the model, the algorithm also returns individual deformations segmentations. These are formally defined as an approximation of the best registration of the estimated template to the observation and of its best segmentation given the model. Additional parameters are also learned by this procedure, such as the means and variances of each tissue grey level distributions.

Some of the experimental results are presented in Fig. 8 to 11 in Section 7. This shows the probabilistic templates (one image per tissue) compared to the results of other atlas estimation algorithms (Fig. 8). We also present the online segmentation of the training data that we get as a complementary output of our estimation algorithm (Fig. 9). The generative model enables to simulate synthetic images, five of them are presented in Fig 10 highlighting characteristic deformations captured in the training sample.

Based on the resulting atlas, an atlas based segmentation of a new observation can be performed by maximising the posterior probability of each of its voxel to belong to each tissue given the observation, the atlas and the computed deformation. The accuracy of the deformation is of high importance, as it impacts the quality of the final segmentation (Fig. 11). This deformation is constrained by the normal distribution which has been learned on the training set, allowing only for meaningful (with regards to the data) deformations.

3.2. Statistical generative model. We now describe the statistical model we propose to integrate all these requirements. We consider here *n* individual MR images from *n* patients. This set $(y_i)_{1\leqslant i\leqslant n}$ of images are observed on a grid of voxels Λ embedded in a continuous domain $D \subset \mathbb{R}^3$. We denote $x_j \in D$ the location of voxel $j \in \Lambda$. We consider that each image is composed of voxels belonging to one class among K, corresponding to K tissues types. We assume that the signals in the K tissue classes are normally distributed with class dependent means $(\mu_k)_{1\leqslant k\leqslant K}$ and variances $(\sigma_k^2)_{1\leqslant k\leqslant K}$ as proposed in [**6**]. Therefore the probability of observing a data with intensity y_i^j for the ith image in the jth voxel given that it belongs to the kth class $(c_i^j = k)$ is defined as follows:

$$\mathbb{P}(y_i^j | c_i^j = k, \mu_k, \sigma_k^2) = \mathcal{N}(y_i^j; \mu_k, \sigma_k^2), \tag{3.4}$$

where $\mathcal{N}(.; \mu, \sigma^2)$ is the normal density with mean μ and variance σ^2. This expression results from the assumption that given the class, the voxels are assumed to have independent grey level.

In order to take the geometrical variability in shape of the brain along a population, we consider that there exists a random deformation from the template to the subject that acts as follows: the *unobserved* classes of the voxels of the data y are assumed to follow the probability distribution given by the discretisation on Λ of the warped probabilistic template. This template is defined by the probability maps $(P_k)_{1\leqslant k\leqslant K}$ which provides the probability of each voxel to belong to each class in the template domain. In other words, the probability maps are deformed to match the observation y ; then they are discretised on Λ to provide, at each voxel, a voxel-dependent discrete

probability measure for this point which gives the probability of each voxel to belong to each class. Let $z : \mathbb{R}^3 \to \mathbb{R}^3$, for $j \in \Lambda$ the prior probability of a voxel j from subject i to be in the k^{th} class is given by:

$$\mathbb{P}(c_i^j = k) = P_k(x_j - z(x_j))\,. \tag{3.5}$$

As previously, their exists a parameter $\boldsymbol{\beta} \in (\mathbb{R}^3)^{k_g}$ such that

$$\forall x \in D, z_\beta(x) = (\mathbf{K_g}\boldsymbol{\beta})(x) = \sum_{k=1}^{k_g} K_g(x, c_k^g)\boldsymbol{\beta}(k)\,, \tag{3.6}$$

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

As for the deformation model, the templates $P_k : \mathbb{R}^3 \to [0,1], \forall k \in \{1, K\}$, which are the tissue probability maps, are defined on the whole domain D. We parametrize the templates by the coefficients $\alpha_k \in [0,1]^{k_p}$, which satisfy $\forall l \in \{1, k_p\}$, $\sum_{k=1}^{K} \alpha_k^l = 1$. Then, we write for fixed control points $(r_{p,l})_{1 \le l \le k_p}$

$$\forall x \in D, P_k(x) = \mathbf{K_p}\alpha_k(x) = \sum_{l=1}^{k_p} K_p(x, r_{p,l})\alpha_k^l, \tag{3.7}$$

where $K_p(x, r_{p,l}) = 1$ if $r_{p,l}$ is the nearest neighbour of x among the set of points $(r_{p,j})_j$ and 0 otherwise.

The previous hypothesis provides a generative statistical model for a sample of grey level images. The random variables are the deformation vector $\boldsymbol{\beta}$, the class c of each voxel and the parameters that characterise the grey levels of the tissues. The probability distributions of the former two elements are given by Equation (3.4) and (3.5). We assume that the deformation vector follows a normal distribution with mean zero and full covariance matrix. The hierarchical model is given by: $\forall i \in \{1, n\},\ \forall j \in \Lambda$

$$\begin{cases} \boldsymbol{\beta}_i \sim \mathcal{N}(0, \Gamma_g)|\Gamma_g; \\ c_i^j \sim \sum_{k=1}^{K} \delta_k P_k(x_j - z_{\boldsymbol{\beta}_i}(x_j))|\boldsymbol{\beta}_i; \\ y_i^j \sim \mathcal{N}(\mu_k, \sigma_k^2)|c_i^j = k, \mu_k, \sigma_k^2, \end{cases} \tag{3.8}$$

where δ_k is a Dirac measure on k. As for the previous cases, a Bayesian approach is used to face the high-dimension-low-sample-size paradigm.

In this model, we estimate $(\mu_k, \sigma_k^2, \alpha_k)_{1 \le k \le K}$ and Γ_g which are the fixed effects whereas the deformations and voxel classes are the random unobserved variables. This model falls again in the mixed effect setting.

4. Multimodal images [A15, C11, A19]

With their increasing number and quality, most of the acquisition protocols combine different modalities in order to refine the analysis or diagnosis (for example, the acquisition done for the HM-TC ANR project combine 3 and 7T-MRIs, DWIs and MEG recordings). This provides us with at least pairs of images per subject. These images are obviously complementary as they show different aspects of the same brains. T1-MRIs highlight the anatomy (basically grey / white matter

and CSF) whereas functional MRIs show the activated areas when the subject is performing a cognitive task. These two aspects of the brain are dependent as activity only appears in grey matter regions. One can wonder how the anatomy is distributed in a population but also how the activity changes within this anatomy. Moreover, the activity shows a region of grey matter (up to the PVE) and then may be used to increase the accuracy of the brain segmentation. To combine all these aspects and take advantages of available modalities, we propose a model which includes this couples of images.

4.1. T1-and-f-MRI [A15, C11]. Let us now consider n pairs of T1- and f- MRIs $(y_{1,i}, y_{2,i})_{1\leqslant i\leqslant n}$ from n patients. We keep the previous section notations. We consider that each MR image is composed of voxels belonging to one of the four classes, corresponding to four tissue types: grey matter (GM), white matter (WM), CSF and background (BG). Each fMR image is composed of voxels belonging to one class among $3+K$, corresponding to WM, CSF and BG, where no activation is expected to occur, and K different levels of activation in grey matter. We assume that the signal in the $3+K$ classes is normally distributed with class dependent means $(\mu_{1,f(k)}, \mu_{2,k})_{k\in\{WM,CSF,BG,GM_1,\dots GM_K\}}$ and variances $(\sigma^2_{1,f(k)}, \sigma^2_{2,k})_{k\in\{WM,CSF,BG,GM_1,\dots GM_K\}}$, where $f(k)=k$ if $k\in\{WM,CSF,BG\}$, GM otherwise. On the other hand, this also avoids to segment active areas which belong for example to the background, that is to say, outside of the brain, which appears in some acquisitions (See Section 7 and **[C11]**).This description forces the activation to appear only in the grey matter voxels. This constrains the segmentation to classify active voxels as belonging to GM and also provides more information to increase the accuracy of the registration.

As mentioned previously, we are working with grey level images which have not been neither pre-segmented nor pre-registered. Let us now describe the generative model we chose to approximate the generation of our observed images. It is based on the probabilistic atlas model described in Section 3. The unknown class of each voxel is supposed to be the discretisation on Λ of a random deformation of probability maps $(P_k)_{1\leqslant k\leqslant K+3}$. These probability maps correspond to the probability of each voxel to belong to each class in the template domain. We use the same assumptions on the probability maps and deformation field as in Section 3 namely Eq: (3.5), (3.6) and (3.7).

The previous hypothesis provides a generative statistical model for a sample of pairs of grey level images (again in a Bayesian framework) :

$$
\left\{
\begin{array}{l}
\beta_i \sim \mathcal{N}(0,\Gamma_g)|\Gamma_g;\\
c_i^j \sim \sum_{k=1}^{K} \delta_k P_k(x_j - z_{\beta_i}(x_j))|\beta_i,\\
\\
\begin{pmatrix} y^j_{1,i}\\ y^j_{2,i}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mu_{1,f(k)}\\ \mu_{2,k}\end{pmatrix}, \begin{pmatrix}\sigma^2_{1,f(k)} & 0\\ 0 & \sigma^2_{2,k}\end{pmatrix}\right)\left| c_i^j = k, \begin{pmatrix}\mu_{1,f(k)}\\ \mu_{2,k}\end{pmatrix}, \begin{pmatrix}\sigma^2_{1,f(k)}\\ \sigma^2_{2,k}\end{pmatrix}\right.
\end{array}
\right. \tag{3.9}
$$

where $\mathcal{N}(\cdot;\mu,\sigma^2)$ is the normal density with mean μ and variance σ^2 and δ_k is a Dirac function.

The parameters to estimate are the covariance matrix Γ_g of the deformation distribution (Eq. (3.6)), $(\alpha_k)_{1\leqslant k\leqslant K+3}$ the coefficients that define the template maps extended to $3+K$ classes

(Eq. (3.7)), $(\mu_{1,f(k)}, \mu_{2,k})_{1\leqslant k\leqslant K+3}$ and $(\sigma^2_{1,f(k)}, \sigma^2_{2,k})_{1\leqslant k\leqslant K+3}$ the class dependent means and variances.

At the end of the estimation process, we get a probabilistic template which shows both the tissue probability maps as well as for the grey matter the probability of different level of activation for the given cognitive task. One of this image is presented in Fig. 12 where the activation areas have been projected back to the segmented cortical surface (segmentation based on our anatomical probability maps). As well as for the probabilistic anatomical atlas, along the algorithm, we also get the individual segmentations of the training images. Some of these are presented in Fig. 13 where we compare the accuracy of the localisation of the active voxels with the same algorithm which only deals with fMR images.

At the end of this estimation, we get detailed information about the grey matter. One can also wonder if the information in the white matter can also be included into the model. This is what we try the following paragraph.

4.2. TI-MRI and DWI (work in progress [A19]). We are working on proposing a second model in order to estimate a Diffusion Weight Image (DWI) atlas. DWIs are acquired in order to provide the directions of the fibre bundles in the white matter. In order to get such directions, the patient is scanned several times with different magnetic field directions. The relaxation times are recorded for each direction (call gradient) and form the diffusion weigh images. The number of gradient direction changes according to the precision of the fibre bundles the user expects. The simplest model is know as Diffusion Tensor Image (DTI) [**20**]. It models at each voxel the water diffusion as a Gaussian distribution. The covariance matrices are tensors (3×3 matrices) which show the directions of the underlying fibres thanks to the matrix eigen vectors. This model requires 7 different gradients to get an estimate. However, many more can be acquired increasing the precision of the estimate and allowing for higher degree tensor models (called HARDI).

If we consider the DTIs, the registration is an important challenge as well as the construction of a template image. This becomes even more complicated as the model of diffusivity gets complicated as well.

Our idea is to construct a DWI template rather than choosing a tensor representation. In this framework, the data images do not need to be co-registered in a preprocessing step. We use as input of the model the DWI images of each patient and construct a template of continuous DWI in space and in gradient directions which live on the sphere S^2. This provides an access to what would be the raw data of the ideal person that is to say the 3D grey-level volumes for any gradient direction on the sphere. If you resample the sphere afterwards with the number of directions you want, you can estimate the template diffusion tensors as well as any other HARDI image like q-balls [**72**].

We are currently testing an estimation algorithm based on this model. The implementation has been done in `C` language by Michael Grasseau, Research Engineer from CMAP Ecole Polytechnique. This will enable to insert this program in any platform like MedInria or Deformetrica to provide a new toolbox to their users. This platform can be used for anatomical atlas estimation

as well as for the DWI case depending on the number of gradient images that are provided per patient. Indeed, the generalised model above includes the single T1 or T2 atlas estimation. The input in this particular case is a database with a unique image per patient. The output is therefore of the same nature: a continuous anatomical image (T1 or T2) in space. As soon as the database has more that one image per patient, the DWI estimation will be performed

At the end of this step, it will be coupled with the probabilistic fMRI atlas estimation to provide a complete atlas of the brain. In this atlas, the activation is constrained to the grey matter, the segmentation and registration of the data images are performed along the atlas estimation and involve the informations provided by both the activation areas and the fibre bundles.

5. Diffeomorphic deformations [A8, A12]

The previous models were all considering small deformations. Although easy to use, this is not entirely satisfactory when analysing medical images as there is no guaranty that the topology is preserved. In specific cases such as stroke study (see Chapter 5), this is not an interesting nor even required condition. However, in the analysis we perform here (mainly atlas estimation of normal brains) this constraint becomes meaningful. Moreover, the framework we use offers a great advantage which is that it provides a natural group action of the group of constructed diffeomorphisms on the set of images. This allows to provide a distance on this set thanks to a metric on the diffeomorphisms. For these reasons, we specified our BME-Template model so that it takes into account this diffeomorphic condition.

We focus here on large diffeomorphic deformations in the LDDMM setting producing deformations ϕ_i [**70, 55**] which we recall very briefly here. In this framework, a large group of diffeomorphisms is seen as "Riemannian manifold" of infinite dimension. The equivalent of the logarithm of a diffeomorphism is a continuous squared integrable velocity field. The conjugate variable of the velocity field is the momentum, which is used to define intrinsic tangent-space statistics on deformations [**73, 26, 64, 25**]. For image matching, the momentum is encoded by an image of infinite dimension, or numerically of the size of the input images. However, it has been shown in [**26**] that such continuous momentum maps can be efficiently approximated by a finite set of well-chosen Dirac delta momenta, where momenta here stand for vectors attached to control points. Therefore, we use in this modelling a finite-dimensional parameterisation of the momenta, in the spirit of [**4**] and [**39**] as described in the following.

5.1. Finite dimensional representation of diffeomorphic deformations. The set of geometric control points, which may be located anywhere in the image domain, defines a potentially infinite-dimensional basis of the parameterisation of the deformations. The vectors attached to them define the weights of the decomposition of a given deformations onto this basis. Defining an adapted basis for the description of the variability means finding the optimal positions of a finite number of control points: both the position and the number of the geometric control points should be optimised altogether, given an image ensemble. Indeed, an optimal set of geometric control points are unlikely to be equally distributed in the image domain; instead, they should be located at the most variable parts of the image. The optimal positions of the points are characteristic of the image ensemble and therefore shared by all the template-to-subjects deformations. The momentum vectors attached to these control points parameterise each of these deformations

and are therefore specific to each observation. We use the property of the LDDMM setting that the optimisation of the position of the control points could be done along with the estimation of the momentum vectors without introducing any additional cost in the computation of the gradient. We propose here to take advantage of the geodesic shooting equations in order to guarantee that at each step of the optimisation process, the computed deformations are geodesic. Here are the details.

In our approach, we assume a discrete parameterisation of the driving velocity field v_t via a convolution operator:

$$v_t(x) = \sum_{k=1}^{k_g} K_g(x, c_k^{\mathrm{g}}(t))\alpha_k(t), \tag{3.10}$$

where for each time t, $c_i^{\mathrm{g}}(t)$ denotes a set of k_g geometric control points, $\alpha_i(t)$ a set of k_g momentum vectors attached to them. K_g is a fixed positive definite kernel that defines a RKHS.

We denote $z(t) = \{c_k^{\mathrm{g}}(t), \alpha_k(t)\}_{k=1,\dots,k_g}$ (a $2dk_g$ vector, where $d = 2$ in 2D and 3 in 3D) the state of the system at time t. Knowing the state of the system at any time $t \in [0, 1]$ defines a flow of diffeomorphisms. Indeed, any point x_0 in the ambient space follows the path $x(t) = \phi_t(x_0)$ which satisfies the ODE:

$$\begin{cases} \dot{x}(t) = v_t(x(t)) = \sum_{k=1}^{k_g} K_g(x(t), c_k^{\mathrm{g}}(t))\alpha_k(t) \\ x(0) = x_0 \end{cases} . \tag{3.11}$$

The path $x(t)$ depends therefore only on the initial condition x_0 and the state of the driving system $z(t)$. The final position $x(1)$ is by definition $\phi_1(x_0)$.

We take advantage of the fact that among all paths $t \to z(t)$ connecting ϕ_0 to ϕ_1 there is one which satisfies a minimum energy principle: the 'geodesic paths.'

Its kinetic energy depends only on the time-varying state of the system $z(t)$. Following mechanical principles, it has been shown in [**56**] that the extremal path connecting ϕ_0 and ϕ_1 is such that the state of the system $z(t)$ satisfies the following set of ODEs:

$$\begin{cases} \dfrac{dc_k^{\mathrm{g}}(t)}{dt} = \displaystyle\sum_{l=1}^{k_g} K_g(c_l^{\mathrm{g}}(t), c_k^{\mathrm{g}}(t))\alpha_k(t) \\ \dfrac{d\alpha_k(t)}{dt} = -\displaystyle\sum_{l=1}^{k_g} \alpha_k(t)^t \alpha_l(t) \nabla_1 K_g(c_l^{\mathrm{g}}(t), c_k^{\mathrm{g}}(t)) \end{cases} , \tag{3.12}$$

given initial conditions $\alpha_k(0) = \alpha_{0,k}$ and $c_k^{\mathrm{g}}(0) = c_{0,k}^{\mathrm{g}}$. Denoting $z_0 = \{\alpha_{0,k}, c_{0,k}^{\mathrm{g}}\}_k$ the initial state of the system, (3.12) can be re-written in short as:

$$\begin{cases} \dot{z}(t) = F(z(t)) \\ z(0) = z_0 \,. \end{cases} . \tag{3.13}$$

These differential equations can be interpreted as the motion of k_g self-interacting particles without external forces. The interaction between particles is given by the kernel K_g. The first equation in (3.12) gives the speed of the control points; the second one, its acceleration. Note that the first equation is consistent with the definition of the velocity field in (3.10), since it reads $\frac{dc_k^{\mathrm{g}}(t)}{dt} = v_t(c_k^{\mathrm{g}}(t))$.

These equations show that the whole flow of diffeomorphisms is entirely determined by the initial state of the system z_0. Indeed, given z_0, the integration of (3.12) gives the state of the system at any later time t: $z(t)$ (the motion of the control points and the momentum vector over time). Then, the integration of (3.11) gives the motion of any point x_0 in the ambient space according to the flow of diffeomorphisms ϕ_t. The generation of diffeomorphisms ϕ can be fully controlled by the finite-dimensional vector z_0.

Accordingly, the inverse map ϕ_1 is also fully determined by z_0. Given a point position y in the image domain Ω, the position given by the inverse flow $\phi_1^{-1}(y)$ can be computed by integrating the following ODE backward in time (where the velocity field has been reversed):

$$\frac{dy(t)}{dt} = -v_t(y(t)), \qquad y(1) = y. \tag{3.14}$$

Then, the final value at time $t = 0$ gives the mapped position: $y(0) = \phi_1^{-1}(y)$.

Let Y be an image of vectors, which gives the position of every voxel in the image domain. In the continuous setting, we have $Y(y) = y$ for any $y \in \Omega$, where Y is seen as a squared integrable map in $L^2(\Omega, \mathbb{R}^d)$. The domain Ω is deformed by the inverse diffeomorphism ϕ_1^{-1}: the inverse flow can be computed by integrating the following ODE:

$$\begin{cases} \dfrac{dY(t)}{dt} = G(Y(t), z(t)) \\ Y(1) = \mathrm{Id}_{L^2} \end{cases}, \tag{3.15}$$

where

$$\begin{aligned} G(Y(t), z(t)) &= -v_t(Y(t)) \\ &= -\sum_{k=1}^{k_g} K_g(Y(t)(.), c_k^{\mathrm{g}}(t))\alpha_k(t) \end{aligned} \tag{3.16}$$

maps an image of vectors in $L^2(\Omega, \mathbb{R}^3)$ and a $2dk_g$-dimensional vector to an image of vectors in $L^2(\Omega, \mathbb{R}^3)$.

Once integrated backward from time $t = 1$ to $t = 0$, the final image of vectors $Y(0)$ maps the domain Ω to $\phi_1^{-1}(\Omega)$. As a consequence, the deformation of the template image I_0 can be written as:

$$I_0(\phi_1^{-1}(y)) = I_0(Y(0)(y)). \tag{3.17}$$

Eventually, one can easily verify that the geodesic paths of the state of the system are energy conservative: for any time t, $\|v_t\|_V^2 = \|v_0\|_V^2$. Therefore, the total kinetic energy of a given path is given as:

$$L(z_0) = \sum_{k=1}^{k_g} \sum_{l=1}^{k_g} \alpha_{0,k}^t K_g(c_{0,k}^{\mathrm{g}}, c_{0,l}^{\mathrm{g}})\alpha_{0,l}. \tag{3.18}$$

This is a function of only the initial state of the system, which will be used as a measure of regularity of the deformations in the objective function. This yields a matching energy -and therefore a template estimation energy- which only depends on this initial state of the particle system.

In this particular setting, the template estimation is performed by a gradient descent considering both the unobserved deformation momenta and the template parametrisation as parameters to be optimised. The total energy to minimise sums the individual contributions of each observation, the energy of image I_i being the tradeoff between $L(z_0)$ and $\|I_0(\phi_1^{-1}(y)) - I_i\|^2_{L^2}$ (we consider that images are L^2 functions).

5.2. Sparcity of the deformation representation. Now that we are able to optimise the position of the control points, one step further is: can we also optimise their number? The dimension of the parameterisation of the deformations is determined by the number of geometric control points. We adjust the dimension of this parameterisation, so that it better reflects the true number of degrees of freedom that is needed to describe the variability of the image set. An optimal set of geometric control points would be concentrated near the contours of the template image, where the need of deformation is the most important.

The kinetic energy is used as a L^2 regularity term used in the previous energy. The effect of this term is to spread the "total amount of momentum" that is needed over the whole set of control points. This is in contradiction with our goal to select a small amount of relevant control points to describe the variability of the image set. To enforce the distribution of momenta to be concentrated on a small set of control points, we introduce an additional L^1 penalty term in the spirit of elastic nets [**77**]. Note that this is not exactly the elastic net paradigm, since we do not use the usual Euclidean norm on the momentum vectors for the L^2 penalty ($\boldsymbol{\alpha}_{0,i}{}^t\boldsymbol{\alpha}_{0,i}$) but the L^2 metric induced by the metric K_g instead $L_i(\mathbf{c}_0^{\mathrm{g}}, \alpha_{0,i}) =(\boldsymbol{\alpha}_{0,i}{}^t K_g \boldsymbol{\alpha}_{0,i})$:

$$E(\mathbf{w}, \mathbf{c}_0^{\mathrm{g}}, \{\boldsymbol{\alpha}_{0,i}\}_{i=1,\dots,N}) = \sum_{i=1}^{N} \left\{ A_i(\mathbf{w}, \mathbf{y}_i(0)) + L_i(\mathbf{c}_0^{\mathrm{g}}, \alpha_{0,i}) + \gamma_{\mathrm{g}} \sum_{p=1}^{k_g} \|\alpha_{0,i,p}\| \right\}, \tag{3.19}$$

where $\|.\|$ denotes the Euclidean norm in the ambient 2D or 3D space and A_i is the data attachment term related to the i^{th} observation.

The effect of this prior is to enforce momentum vectors of small magnitude to vanish. This enforces the deformations to be encoded in a small number of non-zero parameters.

As before, this particular optimisation is performed by a gradient descent. The FISTA [**10**] algorithm can be easily implemented to deal with the non-differentiable penalty.

5.3. LDDMM BME-template model. As we have already pointed in Section 2, the variational approach is subject to convergence failure in particular when considering noisy data but not only (see Section 8). Therefore, now that this diffeomorphic constrain has demonstrated its potential, the idea is to introduce it into the BME-Template model and to perform the atlas (not only template) estimation in a stochastic way which will be described in Section 6. First of all, one has to construct the model and then to look for a way to introduce the sparsity condition. This is what is presented here.

Let $\phi_1^{\mathbf{v}}$ be the solution of the flow equation (5.6) for a regular velocity vector field $\mathbf{v} = (v_t)_{t\in[0,1]}$. Then, for all $s \in \Lambda$, we assume that the observations satisfy :

$$y(s) = I_0((\phi_1^{\mathbf{v}})^{-1}(r_s)) + \sigma\epsilon(s),$$

where $\sigma\epsilon$ denotes an additive random noise on the grid Λ and r_s is the pose of voxel s in the continuous domain D.

As in Section 5.1, we assume that the velocity vector field $\mathbf{v}$ is only driven by the position of the geometric control points $\mathbf{c}_0 = (c_{k,0})_{1\leq k\leq k_g}$ and the associated initial momenta $\boldsymbol{\alpha}_0$ on the domain D. This provides a simple finite dimensional characterisation of the diffeomorphism $\phi_1^{\mathbf{v}}$ as a function of the initial state of the system $z = (\mathbf{c}_0, \boldsymbol{\alpha}_0)$. This yields the notation in terms of z: $\phi_1^{\mathbf{z}}$. Note that the geometrical control points leave in the template domain D. Therefore, these control points are shared by all the observations and drive a different deformation for each image thanks to different initial momentum. However, for numerical efficiency, we have to consider these control points as random non observed variables. Nonetheless, they are still shared by all observations.

According to these elements, we propose the following probabilistic modelling: the initial momenta $\boldsymbol{\alpha}_0 = (\boldsymbol{\alpha}_0^i)_{1\leq i\leq n}$ are independently identically distributed following a Gaussian density with mean 0 and covariance matrix Γ_g. The geometrical initial control points $\mathbf{c}_0 = (c_k^g)_{1\leq k\leq k_g}$ follow a Gaussian distribution with mean $\bar{\boldsymbol{c}}_{\mathbf{0}}$ and covariance $a_c Id$ where Id is the identity matrix of dimension $d \times k_g$. The momenta $\boldsymbol{\alpha}_0$ are assumed to be independent of the control points $\mathbf{c}_0$ given Γ_g.

Following the same lines as above, we parametrize the template function. The action of a diffeomorphism onto this templates summarises as the linear combination of the deformed kernel with the same weights: $\forall r \in D$,

$$I_0 \circ (\phi_1^{\mathbf{z}})^{-1}(r) = K_p^{(\mathbf{c}_0,\boldsymbol{\alpha}_0)}\mathrm{w}(r) = \sum_{k=1}^{k_p} \mathbf{K_p}\left((\phi_1^z)^{-1}(r), r_{p,k}\right)\mathrm{w}(k)\,. \tag{3.20}$$

The complete model writes:

$$\begin{cases} \theta = (\mathrm{w}, \sigma^2, \Gamma_g, \bar{\boldsymbol{c}}_{\mathbf{0}}) \sim (\nu_p \otimes \nu_g) \\ \\ \mathbf{c}_0 \sim \mathcal{N}(\bar{\boldsymbol{c}}_{\mathbf{0}}, a_c Id) |\, \theta, \\ \\ \boldsymbol{\alpha}_0^i \sim \mathcal{N}_{dk_g}(0, \Gamma_g) |\, \theta,\ \forall 1 \leq i \leq n\,, \\ \\ y_i \sim \mathcal{N}_{|\Lambda|}(\mathbf{K_p}^{(\mathbf{c_0},\boldsymbol{\alpha}_\mathbf{0}^\mathbf{i})}\mathrm{w}, \sigma^2 Id_\Lambda) \mid (\mathbf{c}_0, \boldsymbol{\alpha}_0^i),\ \theta,\ \forall 1 \leq i \leq n\,. \end{cases} \tag{3.21}$$

We summarise here the different quantities involved in the model. The parameters of interest are $\theta = (\mathrm{w}, \sigma^2, \Gamma_g, \bar{\boldsymbol{c}}_{\mathbf{0}})$ and the variables $(\boldsymbol{\alpha}_0, \mathbf{c}_0)$ are considered as hidden random variables. The atlas estimation has therefore to estimate both the template image, the covariance matrix of the momenta and the initial position of the geometrical control points. The way it is performed is described in the following Section. Before that, now that we optimise the position of the control points, the same question of how many are really required to capture the geometrical variability of

the population of images arises again. We will introduce in the next paragraph the control point selection in a different spirit of what we have done in the deterministic model.

Remark 1. *This model is very similar to the small deformation one. One can wonder if optimising the position of the geometrical control points would have been also possible in that framework. Although possible, this would not have provide a good estimation. Indeed, the interest of the LDDMM framework is that the time evolution of the control points are linked to their momenta through the Hamiltonian system (Eq.* (3.12)*). This constrains the positions of the points to be at relevant locations in accordance to the shooting momenta. In the small deformation setting, these coupling does not exists and would most probably lead to non identifiability of the model and arguable estimates.*

5.4. Heuristic of proposed sparse model. In the deterministic setting, the control point selection is driven by a L^1 penalty on the momenta in the energy minimisation process. This means that small momenta will be vanished and all other will be slightly reduced. A control point which does not contribute to at least one of the template-to-observation deformations is called non active. So that, the set of active control points is given by the union of all active individual points.

In the following, the selection of control points is performed using a population based criterion. The idea is to inactivate a control point whose contribution quantified by its momenta in the whole population is below a given threshold and does not have a strong correlation with the other control point momenta. Therefore our procedure selects control point locations as well as their number which are relevant with regards to the whole population.

This mathematical constraint on the momenta is taken into account in the model by assuming that the geometrical covariance matrix Γ_g is of the form $\Gamma_g = A_g + \varepsilon_g Id$ where ε_g is a small positive value and A_g is a sparse symmetric positive matrix.

To construct A_g, we threshold the contribution of each momentum in the empirical covariance matrix. Let c_k^g be one of the control points. We compute the sum of the Frobenius norms of the sub-matrix of the sufficient statistic s_3 given by the stochastic approximation of the empirical covariance of this control point with all others:

$$t_k = \sum_{j=1}^{k_g} \|s_3(c_k^g, c_j^g)\|_{Fro}\,, \tag{3.22}$$

where $s_3(c_k^g, c_j^g)$ is the $d \times d$ sub-matrix of s_3 corresponding to the control points c_k^g and c_j^g. If $t_k \geq \lambda$, this control point c_k^g is said active.

Let us denote $\mathcal{A}$ the set of active points. Then we define A_g as follows:

$$\forall (k,j) \in \{1, ...k_g\} \quad A_g(c_k^g, c_j^g) = s_3(c_k^g, c_j^g) \mathbb{1}_{c_k \in \mathcal{A}} \mathbb{1}_{c_j \in \mathcal{A}}\,. \tag{3.23}$$

This forms the sparse matrix A_g and the matrix Γ_g is update as follows:

$$\Gamma_g = \frac{1}{n + a_g}(A_g + a_g Id_{dk_g})\,. \tag{3.24}$$

6. Stochastic Algorithm for parameter estimation

All the previous statistical models fall in the same category namely parametric Bayesian missing data or mixed effect models. We want to estimate our atlases using statistical learning methods which enable the parameter to fit a given population and therefore to adapt to different groups. The natural approach is therefore the Expectation-Maximisation (EM) [**24**] algorithm which was introduced to face exactly this issue. Unfortunately, even in the simplest context we have consider, the direct implementation of this algorithm is not possible since the posterior distribution of the missing data given the observations has a very complex close form (in particular, regarding the normalising constant). Therefore, a first approximation of this algorithm was proposed in **[A1]** which was subject to convergence failure. We proposed in **[A2]** to use a stochastic version of the EM algorithm which we proved convergent. This algorithm is a coupling of the Stochastic Approximation EM (SAEM) introduce in [**22**] with Monte Carlo Markov Chain (MCMC) methods for the simulation step. The algorithm is describes in the following paragraph.

6.1. MCMC-SAEM [A3]. We present here the algorithm for the mixture model, the one-component one being a special case. All the generalisations of the simple BME-Template model including different deformations or input images can be extended to their mixture cases. Although complex, our models all belong to the curved exponential family which simplifies the implementation of the algorithm as well as is a condition for convergence.

We will present in this section algorithms which can be used for any other mixed effect model satisfying mild assumptions.

We introduce the complete log likelihood : $L : \mathcal{S}\times\Theta\times\varrho \to \mathbb{R}$ as $L(s;\eta) = -\psi(\eta)+\langle s, \phi(\eta)\rangle$, where $\mathcal{S}$ is the set of sufficient statistics s, η is the parameter vector and ϕ and ψ are two measurable functions. Moreover, one assumes the existence of a critical function $\hat{\eta} : \mathcal{S} \to \Theta \times \varrho$ which satisfies: $\forall \eta \in \Theta \times \varrho, \forall s \in \mathcal{S}, L(s; \hat{\eta}(s)) \geq L(s;\eta)$ (note that this condition is satisfied by our models). Then, iteration l of this algorithm consists of the following four steps.

Simulation step: The missing data, denoted here (β, τ) to refer to our models (where β parametrises the deformation and τ is the cluster), are drawn using a transition probability of a convergent Markov chain having the posterior distribution as stationary distribution:

$$(\beta_{l+1}, \tau_{l+1}) \sim \Pi_{\eta_l}((\beta_l, \tau_l), \cdot) ,$$

where η_l is the current estimate of the vector of parameters.

Stochastic approximation step: Since the model is exponential, the stochastic approximation is done on the sufficient statistics using the simulated values of the missing data:

$$s_{l+1} = s_l + \Delta_{l+1}(S(\beta_{l+1}, \tau_{l+1}) - s_l) ,$$

where $(\Delta_l)_l$ is a decreasing sequence of positive step-sizes.

Truncation step: A truncation is done on the stochastic approximation. Let $(\mathcal{K}_q)_{q\geq 0}$ be an increasing sequence of compact subsets of $\mathcal{S}$ such as $\cup_{q\geq 0}\mathcal{K}_q = \mathcal{S}$ and $\mathcal{K}_q \subset \text{int}(\mathcal{K}_{q+1}), \forall q \geq 0$. If $\bar{s}_{l+1}$ wanders out of $\mathcal{K}_{l+1}$ then the algorithm is reinitialised in a given compact set. Otherwise,

set $s_{l+1} = \bar{s}_{l+1}$.

Maximisation step: The parameters are updated:

$$\eta_{l+1} = \hat{\eta}(s_{l+1}) \, .$$

6.2. Transition kernels. We now explain the first choice of the transition kernel of the Markov chain Π_η used in the simulation step. In this mixture case, one has to pay attention to the trapping state problem: mainly, images are assigned to a cluster at the first iteration and due to numerical problems (low probability) never change along the estimation.

As we aim to simulate (β_i, τ_i) through a transition kernel whose stationary distribution is $q(\beta, \tau | y_i, \eta)$, we first simulate τ_i with a kernel whose stationary distribution is $q(\tau | y_i, \eta)$ and then β_i through a transition kernel that has $q(\beta | \tau, y_i, \eta)$ as stationary distribution. Given any initial deformation field $\xi_0 \in \mathbb{R}^{dk_g}$, we run, for each component t, J_l iterations of a hybrid Gibbs sampler (for each coordinate of the vector, a Hasting-Metropolis sampling is done given the other coordinates) $\Pi_{\eta,t}$ using the conditional prior distribution $\beta^j | \beta^{-j}$ as the proposal for the j^{th} coordinate, β^{-j} referring to β without its j^{th} coordinate. So that we get J_l elements $\xi_{t,i} = (\xi_{t,i}^{(k)})_{1 \leq k \leq J_l}$ of an ergodic homogeneous Markov chain whose stationary distribution is $q(\cdot | y_i, t, \eta)$. Denoting $\xi_i = (\xi_{t,i})_{1 \leq t \leq \tau_m}$, we simulate τ_i through the discrete density with weights given by:

$$\hat{q}_{\xi_i}(t | y_i, \eta) \propto \left(\frac{1}{J_l} \sum_{k=1}^{J_l} \left[\frac{f_t(\xi_{t,i}^{(k)})}{q(y_i, \xi_{t,i}^{(k)}, t | \eta)} \right] \right)^{-1} ,$$

where f_t is the density of the Gaussian distribution $\mathcal{N}(0, \Gamma_{g,t})$. Then, we update β_i by re-running J_l times the hybrid Gibbs sampler Π_{η,τ_i} starting from a random initial point β_0.

We prove in **[A3]** that the sequence $(\eta_l)_l$ generated through this algorithm converges a.s. towards a critical point of the penalised likelihood of the observations.

The combination of three different statistical tools -EM algorithm, stochastic approximation and MCMC methods- in a single algorithm led us to assume three usual types of conditions. The convergence assumptions due to the EM algorithm require some regularity of the model. The hypothesis concerning the stochastic approximation focuses on the step-size sequence and on the control of the random perturbation and the residual term. To ensure the convergence of the MCMC method, assumptions similar to the usual Drift conditions are sufficient. We refer to **[A3]** for further details.

Remark 2. *The proof of convergence of the one component algorithm has been addressed in **[A2]** as the algorithm is simpler in the sense that there is no auxiliary chains and $J_l = 1$ for all l (a single hybrid Gibbs sampler step). The two proofs of convergence are different in the sense that one is based on a generalisation of a Theorem from* [**22**]*, the other comes back to the general convergence of the stochastic approximations* [**5**].

However convergent, this algorithm is unfortunately not up and running. Indeed, the hybrid Gibbs sampler becomes computationally very expensive when sampling very high dimensional variables. Although it reduces the dimension of the sampling to one which enables to stride easier the target density support, it loops over the sampling variable coordinates, which becomes computationally unusable as soon as the dimension is very large or as the acceptation ratio involves heavy

computations. In the following, we propose a new MCMC sampler to deal with high dimensional random variables. This sampler is designed to be optimal within the estimation algorithm.

6.3. Anisotropic MALA. We propose an anisotropic version of the well-known Metropolis Adjusted Langevin Algorithm (MALA). So let us first recall the steps of the Metropolis Adjusted Langevin Algorithm (MALA). Let $\mathcal{X}$ be an open subset of $\mathbb{R}^l$, the $l-$dimensional Euclidean space equipped with its Borel $\sigma-$algebra $\mathcal{B}$. Let us denote π the probability density function (pdf) with respect to the Lebesgue measure on $\mathcal{X}$ of the target distribution. We assume that π is positive continuously differentiable. At each iteration k of this algorithm, a candidate X_c is simulated with respect to the Gaussian distribution with expectation $X_k + \frac{\sigma^2}{2} D(X_k)$ and covariance $\sigma^2 Id_l$ where X_k is the current value,

$$D(x) = \frac{b}{\max(b, |\nabla \log \pi(x)|)} \nabla \log \pi(x)\,, \tag{3.25}$$

Id_l is the identity matrix in $\mathbb{R}^l$ and $b > 0$ a fixed truncated threshold. Given this candidate, the next value of the Markov chain is updated using an acceptance ratio $\alpha_{MALA}(X_k, X_c)$ as follows: $X_{k+1} = X_c$ with probability $\alpha_{MALA}(X_k, X_c) = \min\left(1, \frac{\pi(X_c) q_{MALA}(X_c, X_k)}{q_{MALA}(X_k, X_c)\pi(X_k)}\right)$ and $X_{k+1} = X_k$ with probability $1 - \alpha_{MALA}(X_k, X_c)$.

The Gaussian proposal of the MALA algorithm is optimised with respect to its expectation guided by the Langevin diffusion. One step further is to optimise also its covariance matrix. A first work in this direction was proposed by [**7**]. The covariance matrix of the proposal is given by a projection of a stochastic approximation of the empirical covariance matrix. It produces an adaptive Markov chain. This process involves some additional tuning parameters which have to be calibrated. Since our goal is to use this sampler in an estimation algorithm, the sampler has at each iteration a different target distribution (depending on the current estimate of the parameter). Therefore, the optimal tuning parameter may be different along the iterations of the estimation process. We rather take advantage of the dynamic of the estimation algorithm as an adaptation of the chain. On the other hand, an intrinsic solution has been proposed by [**33**] where the covariance matrix is given by the metric of the Riemann manifold of the variable to sample. Unfortunately, this metric may not be accessible and its empirical approximation not easy to compute.

For these reasons, we propose a sampler in the spirit of [**7**], [**33**] or [**51**]. The proposal remains a Gaussian distribution but both the drift and the covariance matrix depend on the gradient of the target distribution. At the k^{th} iteration, we are provided with X_k. The candidate is sampled from the Gaussian distribution with expectation $X_k + \delta D(X_k)$ and covariance matrix $\delta\Sigma(X_k)$ denoted in the sequel $\mathcal{N}\left(X_k + \delta D(X_k), \delta\Sigma(X_k)\right)$ where $\Sigma(x)$ is given as :

$$\Sigma(x) = \varepsilon Id_l + D(x)D(x)^T\,, \tag{3.26}$$

with D defined in Equation (3.25) and $\varepsilon > 0$ is a small regularisation parameter. We introduced the gradient of $\log \pi$ into the covariance matrix to provide an anisotropic covariance matrix depending on the amplitude of the drift at the current value. When the drift is large, the candidate is likely to be far from the current value. This large step may not be of the right amplitude and a large variance will enable more flexibility. Moreover, this enables to explore a larger area around these

candidates which would not be possible with a fixed variance. On the other hand, when the drift is small in a particular direction, it means that the current value is within a region of high probability for the next value of the Markov chain. Therefore, the candidate should not move too far neither with a large drift nor with a large variance. This enables to sample a lot around large modes which is of particular interest. This covariance also enables to treat the directions of interest with different amplitude of variances as the drift already does as well as providing dependencies between coordinates since the directions of large variances are likely to be different from the Euclidean axis. This is taken into account here by introducing the drift into the covariance matrix.

We denote by q_c the pdf of this proposal distribution. The transition kernel becomes:

$$\Pi(x, A) = \int_A \alpha(x, z) q_c(x, z) dz + \mathbb{1}_A(x) \int_{\mathcal{X}} (1 - \alpha(x, z)) q_c(x, z) dz \,, \tag{3.27}$$

where $\alpha(X_k, X_c) = \min\left(1, \frac{\pi(X_c) q_c(X_c, X_k)}{q_c(X_k, X_c)\pi(X_k)}\right)$.

We have proved that under usual regularity conditions on the target distribution, for all $s \in \mathcal{S}$ (the set of sufficient statistics), this kernel is V_s-uniformly ergodic. This property implies that the kernels are geometrically ergodic.

To evaluate numerically the efficiency of this new sampler, we compare its mixing properties with the MALA ones. We used both algorithms to sample from a 10 dimensional normal distribution with zero mean and non diagonal covariance matrix. Its eigenvalues range from 1 to 10. The eigen-directions are chosen randomly. The autocorrelations of both chains are plotted in Fig. 1 where we can see that there is a benefit of using the anisotropic sampler. To evaluate the weight of the anisotropic term $D(x)D(x)^T$ in the covariance matrix, we compute its amplitude (computed as its non zero eigenvalue since it is a rank one matrix). We see that it is of the same order as the diagonal part in mean and jumps up to 15 times bigger. This shows the importance taken by the anisotropic term. The last check is the Mean Square Euclidean Jump Distance (MSEJD) which computes the expected squared distance between successive draws of the Markov chain. The two methods provide MSEJD of the same order showing a very slight advantage in term of visiting the space for the AMALA sampler (1.29 versus 1.25 for the MALA).

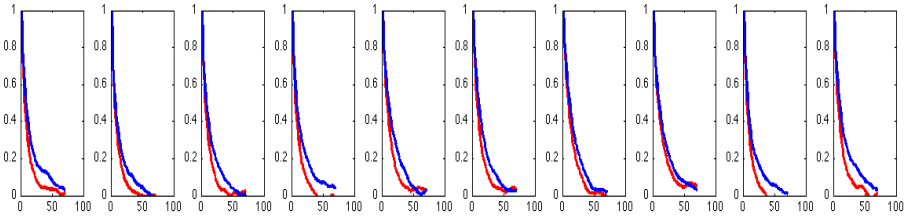


FIGURE 1. Autocorrelations of the MALA (blue) and AMALA (red) samplers to target the 10 dimensional normal distribution with anisotropic covariance matrix.

6.4. AMALA within SAEM. Now that we have an efficient sampler in high dimension, our goal is to use it into the SAEM to increase its computational efficiency. The complete algorithm is given in Algorithm 1.

This algorithm keeps the convergence properties of the previous MCMC-SAEM where the hybrid Gibbs sampler was used. Indeed, we proved that under usual conditions on the model

Algorithm 1 AMALA within SAEM

for all $k = 1 : k_{end}$ **do**

`Sample` z_c with respect to $\mathcal{N}(z_{k-1}+\delta D(z_{k-1},\theta_{k-1}), \delta\Sigma(z_{k-1},\theta_{k-1}))$ whose pdf is denoted $q_{s_{k-1}}(z_{k-1},.)$ where

$$\begin{cases} D(z_{k-1},\theta_{k-1}) & = \frac{b}{\max(b,|\nabla \log p(z_{k-1}|y;\theta_{k-1})|)} \times \nabla \log p(z_{k-1}|y;\theta_{k-1}) \\ \Sigma(z_{k-1},\theta_{k-1}) & = \varepsilon Id_l + D(z_{k-1},\theta_{k-1})D(z_{k-1},\theta_{k-1})^T. \end{cases}$$

`Compute` the acceptance ratio

$$\alpha_{s_{k-1}}(z_{k-1},z_c) = \min\left(1, \frac{p(z_c|y;\theta_{k-1})q_{s_{k-1}}(z_c,z_{k-1})}{q_{s_{k-1}}(z_{k-1},z_c)p(z_{k-1}|y;\theta_{k-1})}\right)$$

`Sample` $\bar{z} = z_c$ with probability $\alpha_{s_{k-1}}(z_{k-1},z_c)$ and $\bar{z} = z_{k-1}$ with probability $1 - \alpha_{s_{k-1}}(z_{k-1},z_c)$

`Do the stochastic approximation`

$$\bar{s} = s_{k-1} + \gamma_k\left(S(\bar{z}) - s_{k-1}\right),$$

where $(\gamma_k)_k$ is a sequence of positive step sizes.

if $\bar{s} \in \mathcal{K}_{\kappa_{k-1}}$ and $\|\bar{s} - s_{k-1}\| \leq \varepsilon_{\zeta_{k-1}}$ **then**

Set $(z_k, s_k) = (\bar{z}, \bar{s})$ and $\kappa_k = \kappa_{k-1}$, $\nu_k = \nu_{k-1} + 1$, $\zeta_k = \zeta_{k-1} + 1$

else

set $(z_k, s_k) = (\widetilde{z}, \widetilde{s}) \in \mathrm{K} \times \mathcal{K}_0$ and $\kappa_k = \kappa_{k-1} + 1$, $\nu_k = 0$, $\zeta_k = \zeta_{k-1} + \Psi(\nu_{k-1})$

where $\Psi : \mathbb{N} \to \mathbb{Z}$ is a function such that $\Psi(k) > -k$ for any k

and $(\widetilde{z}, \widetilde{s})$ is chosen arbitrarily.

end if

`Update` the parameter

$$\theta_k = \hat{\theta}(s_k)$$

end for

(namely the ones required by the EM convergence) and a tail behaviours of the family of posterior distributions (implying uniformity of the geometric ergodicity of transition kernel family), the sequence of parameters converges almost surely towards a critical point of the observed likelihood and reaches an asymptotic normal distribution.

Theorem 4 (Convergence Result for the Estimated Sequence generated by Algorithm 1)**.** *Assume some regularity conditions on the model and the usual condition on the step size sequences. Let* K *and* $\mathcal{K}_0$ *be two compact subsets of* $\mathcal{X}$ *and* $Conv(\bar{S}(\mathbb{R}^l))$*. Then, for all* $z_0 \in \mathrm{K}$ *and* $s_0 \in \mathcal{K}_0$*, we have* $\lim_{k\to\infty} d(\theta_k, \mathcal{L}) = 0$ *a.s. where* $(\theta_k)_k$ *is the sequence generated by Algorithm 1 and* $\mathcal{L} \triangleq \{\theta \in \Theta, \partial_\theta l(\theta) = 0\}$ *where* l *is the observed log likelihood.*

Theorem 5. *Under the usual assumptions for Central Limit Theorems (including the previous ones), the sequence of sufficient statistics* $(s_k - s^*)/\sqrt{\gamma_k}$ *converges in distribution to a Gaussian random vector with zero mean and covariance matrix* Γ*. Moreover,*

$$\frac{1}{\sqrt{\gamma_k}}(\theta_k - \theta^*) \to_{\mathcal{L}} \mathcal{N}(0, \partial_s\hat{\theta}(s^*)\Gamma\partial_s\hat{\theta}(s^*)^T)$$

where $\theta^* = \hat{\theta}(s^*)$.

Remark 3. *Note that this asymptotic normality property also applies to the previous Metropolis-Hastings within Gibbs sampler in the mixture or non mixture cases. The assumptions are very common and usually satisfied by many other MCMC samplers. Therefore, these convergences will also hold for many other MCMC methods (however maybe not efficient in practice).*

Remark 4. *Note that for the probabilistic and multi-modal template estimations, the unobserved parameter $\boldsymbol{\beta}$ appears in the indicator function and makes it impossible to compute the gradient with respect to $\boldsymbol{\beta}$ and thus precludes any algorithm based on alternative gradient descents. This indicator functions enables to easily deal with the constrain on α which would appear much harder with other smoother kernels.*

Remark 5. *This algorithm is particularly well adapted to the LDDMM model. Indeed, the random variables to sample from are the vectors of deformation momenta and the control point positions. Computing the gradient of the complete log likelihood with respect to one of these quantity automatically provides the gradient with respect to the second one. Therefore, for a really limited additional computational cost (due to the calculation of new sufficient statistics), it is easy to optimise with respect to these point positions.*

Remark 6. *Last remark, the interesting aspect of the AMALA algorithm (shared by the MALA sampler) is that it relates to a gradient ascent on the posterior distribution which is one step of the deterministic version of the EM we proposed in **[A1]**. Indeed, given the current estimates of the parameters, we simulate a candidate as a Gaussian perturbation of a vector which is in the direction of increasing posterior probability. The mean of the Gaussian proposal is a gradient ascent candidate. The power of the random sampler is that it explores around this proposed value and in particular for the AMALA, taking into account the anisotropy of the target distribution. This enables to look into the posterior distribution and not only focus on its mode. However, this relation between the two algorithm highlights the interest of the AMALA sampler which takes advantage of both the gradient descent efficiency and the random exploration of the posterior probability.*

7. Some of the results got with the previous estimations

7.1. Handwritten digits. The first experiments we run are on a toy (but informative) example, handwritten digits. The estimated templates for different estimation algorithms are presented in Fig. 2. This particular data base has a great advantage as it also contains a test base which can be used to compare classification performances. Although not designed to be a classifier, our model can be used, once the parameters have been learnt, as a classifier. We refer to the different published papers to see the classification performances.

The influence of the dimension of the random variable to simulate (here the deformation parametric vector β) on the sampler is highlighted in Fig. 3. This shows the importance of the MCMC sampler used in the SAEM algorithm. This is even more glaring when looking at synthetic samples generated from the estimated model (using noisy training examples) where the geometry is badly estimated when the sampler does not visit properly the variable support (Fig. 4).

Remark 7. *Another application of this estimation method used as a classifier was done with a ONERA team on recognition and detection of plane radar images (see **[A7]**) which has motivated a PhD subject treated by Florian Maire (see **[A13]**,* [**49**]*).*

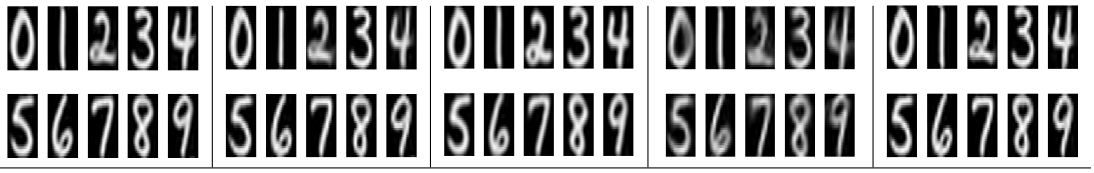
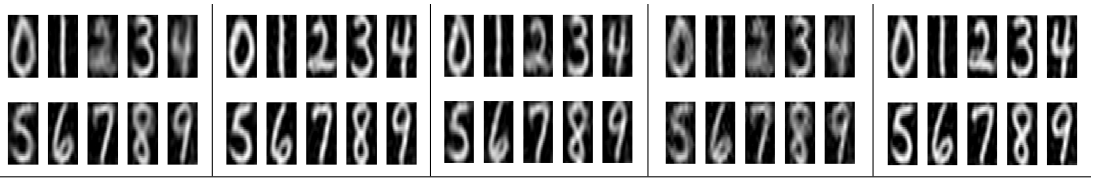
Algo./ Noise level	FAM	Hybrid Gibbs	MALA	Adaptive MALA	AMALA
No additive Noise					
Additive Noise of Variance 1					

FIGURE 2. Estimated templates using the five algorithms and noise free and noisy data. The training set includes 20 images per digit. The dimension of the hidden variable is 72.

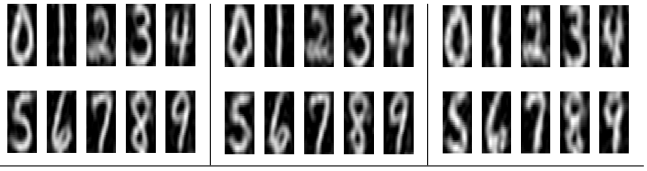
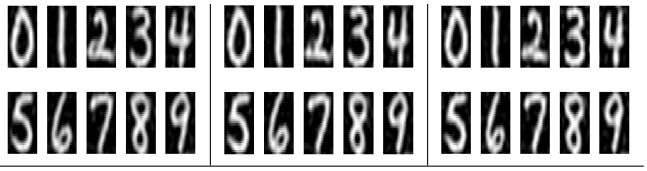
Dim. of deformation / Sampler	$2k_g = 72$	$2k_g = 128$	$2k_g = 200$
MALA			
AMALA			

FIGURE 3. Estimated templates using MALA and AMALA samplers in the stochastic EM algorithm on noisy training data. The training set includes 20 images per digit. The dimension of the hidden variable increases from 72 to 200.

7.2. Grey level Medical images. We also test our algorithm in much higher dimension using the dataset of murine dendrite spines (see [**1, 18, 19**]) already used by [**3**]. The dataset consists of 50 binary images of microscopic structures, tiny protuberances found on many types of neurones termed dendrite spines. The images are from control mice and knockout mice which have been genetically modified to mimic human neurological pathologies like Parkinson's disease.

The template estimated with either 30 or 50 observations are presented in Fig. 7. We obtain similar shapes which are coherent with what a mean shape could be regarding the training sample. To evaluate the estimated geometrical variability, we generate synthetic samples. Eight of these are shown in Fig. 6. We observe different twisting which are all coherent with the shapes observed

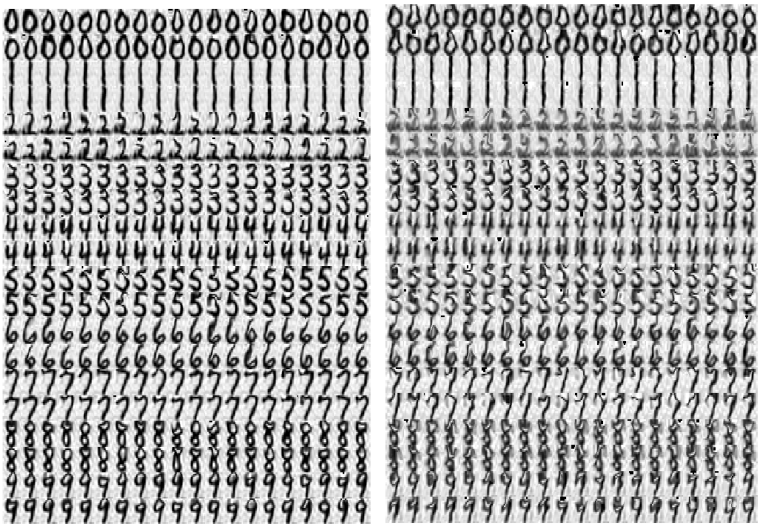


FIGURE 4. Synthetic samples generated with respect to the BME template model using the estimated parameters with AMALA-SAEM (left) and MALA-SAEM (right). For each digit, the two lines represent the deformation using $+$ and $-$ the simulated deformation z. The number of geometric control points is 64 leading to a hidden variable of dimension 128.

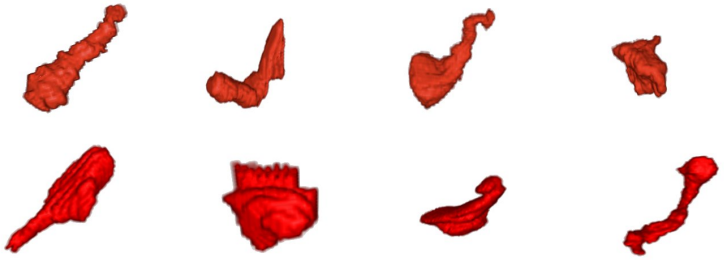


FIGURE 5. 3D views of eight samples of the data set of dendrite spines. Each image is a volume leading to a binary image.

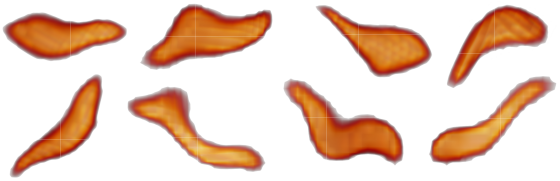


FIGURE 6. 3D views of eight synthetic data. The estimated template shown in Fig. 7 is randomly deformed with respect to the estimated covariance matrix.

in the dataset (some examples are presented in Fig. 5). Note that the training shapes have very irregular boundaries whereas the parametric model used for the template leads to a smoother image. Thus, the synthetic samples do not reflect the local ruggedness of the segmented murine dendrite spines. If the aim was to capture these local bumps, the number of photometrical control points has to be increased. However, the goal of our study was to detect global shape deformations.

7.3. Anatomical brain MRIs. Here are presented some experiments on brain anatomical images where we estimate the tissue probability maps shown in Fig. 8 (with comparisons with

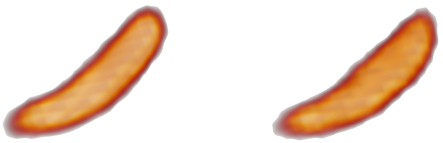


FIGURE 7. Estimated templates of murine dendrite spines. The training set is either composed of 30 (left) or 50 (right) images.

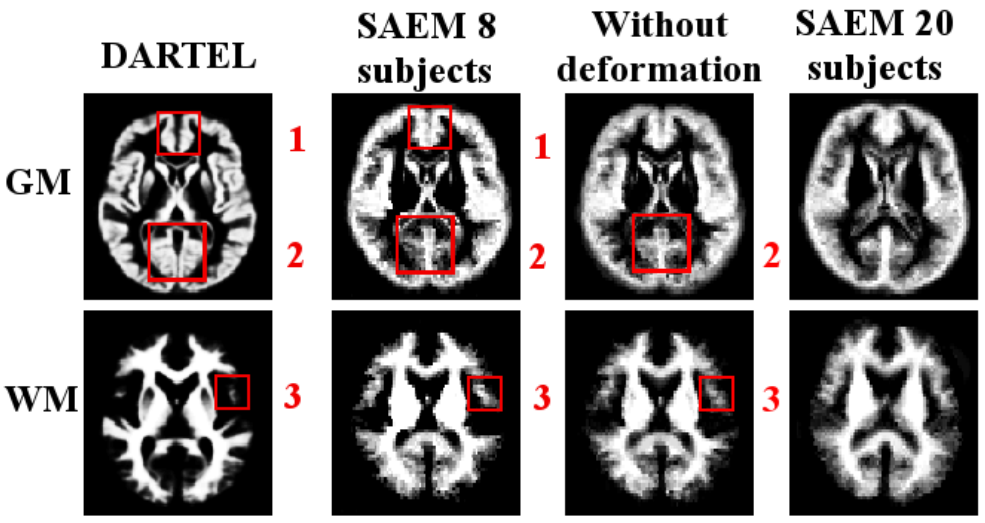


FIGURE 8. The atlas obtained by DARTEL (first column), our method with 8 subjects (second column) and 20 subjects (third column). GM for the first row and WM for the second. DARTEL atlas shows a wide line of CSF in zone 3 which is thin in the data images. Our probabilistic maps capture the presence of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

existing methods). The choice of the sampler (Metropolis-Hasting within Gibbs with the prior distribution used as proposal) enables to get along with the atlas the individual segmentations of the training images. Some of these segmentations are presented in Fig. 9 and synthetic generated images in Fig. 10. Posterior segmentation of new individuals are shown in Fig. 11. More details on the performances highlighted in red regions can be found in **[A13]**.

7.4. T1-andf-MRI atlas. We present here some results of the multi-modal atlas estimation when we consider anatomical and functional data together. The template is presented in Fig. 12 on the cortical surface segmented from our anatomical estimation. As well as for the previous anatomical case, the segmentation of active areas detected on one of the training subjects are part of the output of the algorithm. One of them is shown in Fig. 13 and compared to the result obtained when only using the functional images (without the anatomical image, the anatomical constrains disappear leading to activations in white mater voxels or even outside the brain). More details and comparisons on these experiments can be found in **[A15]**.

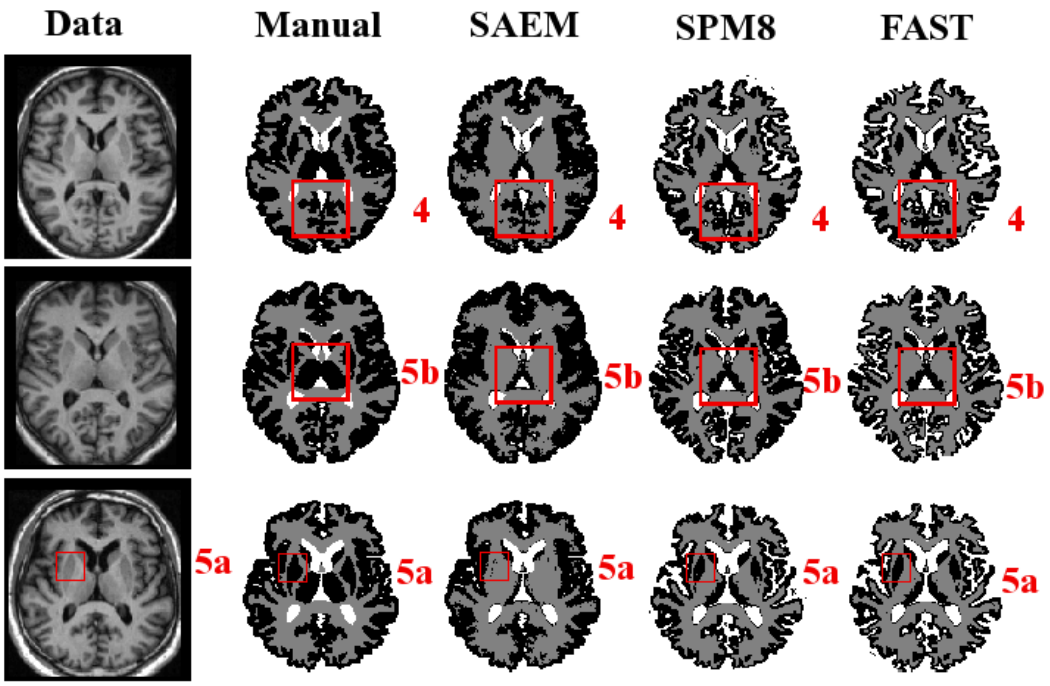


FIGURE 9. Experiments on real data. Each column corresponds to the one slice of 3 data images, the manual segmentation and the segmentation obtained by our method, SPM8 and FAST. Our methods shows each fold of the GM (zone 1). Our method does not manage to segment the subcortical structures (zone 2a), others segment successfully with the strong prior (however not entirely, see zone 2b).

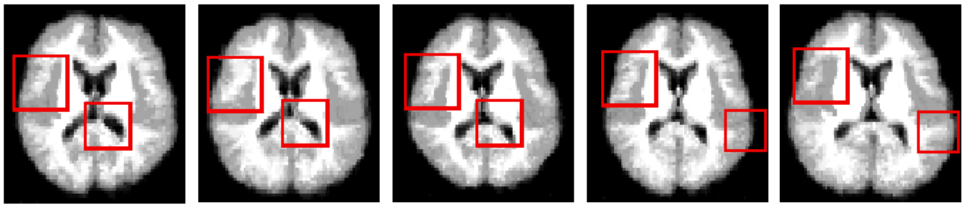


FIGURE 10. Five simulated images using the estimated template with 8 subjects. The deformations of the ventricles are realistic as well as the cortex foldings which look like some training ones. Moreover, the cortex thickness changes as we can see in the training set.

8. Open questions and future work

In addition to the DWI atlas estimation which has to be finished **[A19]**, several questions remain concerning the model and its parameter estimation.

8.1. Theoretical analysis [A17]. In **[A17]**, we are interested in the non convergence of the deterministic approach which has shown its limitation numerically. The question are of different aspects. First, can we make a parallel with the computation of Frêchet means in a certain space? Can we compute the bias the deterministic algorithm produces as a function of the noise variance of the images? Does there exist cases where this method will always be biased? Then, questions about what is the most important part of the stochastic algorithm which enables to overcome the

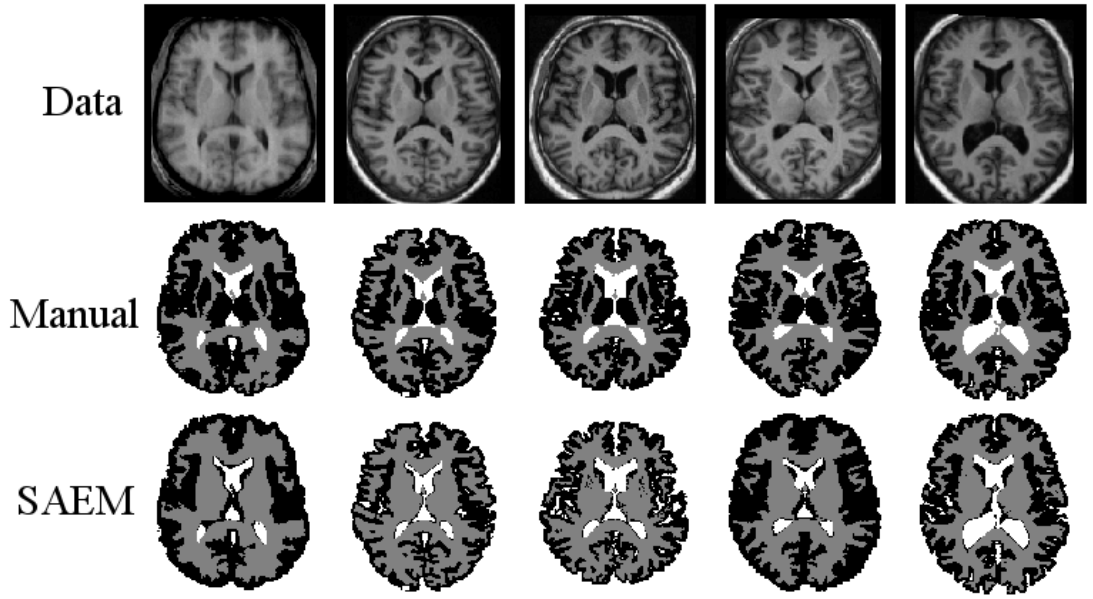


FIGURE 11. Segmentation for 5 new individuals using the atlas created by 20 individuals. Each row corresponds to the one slice of 5 data images, the manual segmentation and the segmentation obtained by our method.

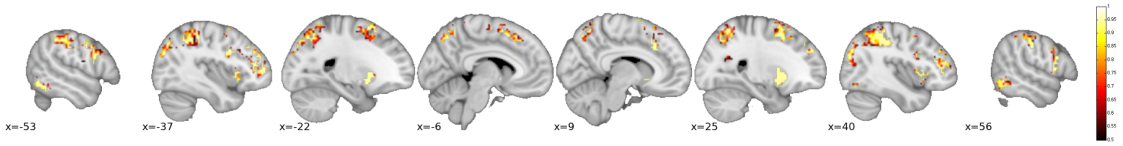


FIGURE 12. The estimated template of activation. The yellow/red colors correspond to high/low probability of the activation for the computation task.

previous limitations arise. These are all the directions we are exploring in **[A17]**.

8.2. Implementation of a package. Another side of this work which has to be considered is to propose a publicly available package so that other teams can test our methods on their data. This requires to code part of the existing implementation in another language (for example `C` as one part is already converted). This step is a join work with Michael Grasseau, research engineer at the CMAP and can be eventually included in the existing softwares MedInria or Deformetrica.

8.3. Shape based regularisation for 2D tomography [A18]. This work is a collaboration with Ozan Oktem from the Center for Applied and Industrial Mathematics in KTH, Sweeden. His issues are related to inverse problems in structural biology, and in particular the idea is to provide methods for image processing in Electron Microscopy that can account for 2D/3D shape information using shape analysis.

Image reconstruction and analysis is central in automating the imaging pipeline in Electron Microscopy. Due to the problem complexity, successful approaches need to properly accounted for: (a) the relation between image features and imaging data, (b) statistical properties of image noise and other uncertainties, and (c) a priori knowledge about the image features. The first two

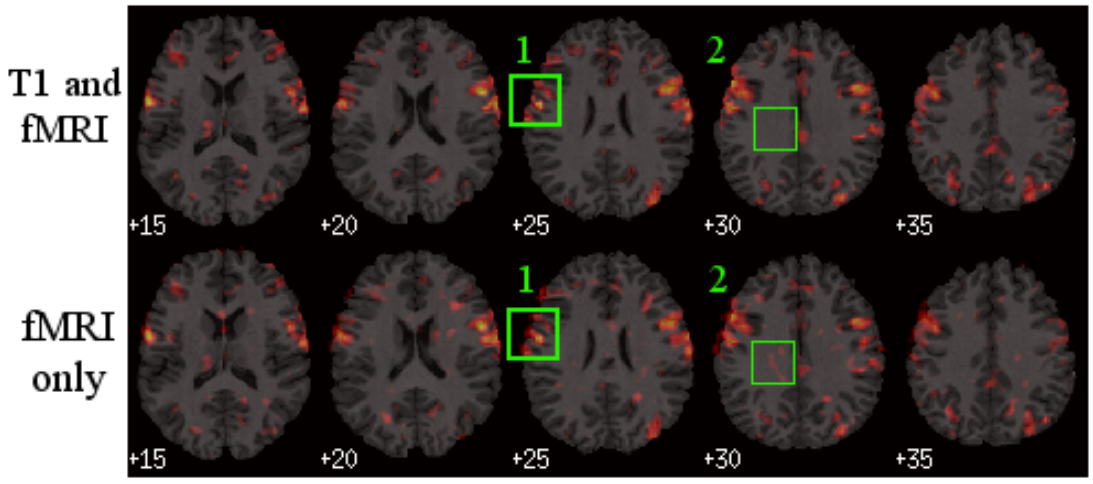


FIGURE 13. Experiments on real data showing the detected active areas $p > .95$ for the computation task. The first row for our method using both T1- and f-MRI and the second row for the standard method using only fMRI. Each column corresponds to one slice of the same patient.

are related to the physics of image formation, the last to the type of samples being investigated. Yet another issue is that most imaging problems in Electron Microscopy are ill-posed, so a priori knowledge about relevant image features must be used. Current approaches for image reconstruction and analysis makes use of a priori regularity assumptions (smoothness, edges, etc.) in the 2D/3D image features being processed. Knowledge about shapes are however not yet used in this context and it can have profound impact on the outcome.

The challenge lies in adopting the proper mathematical notion of shape and shape variability that is needed for analysing and statistically modelling the shape of structures in different samples being imaged. The mathematical framework has to be general enough to handle a large variety of structures, while still be computationally feasible. Most approaches are based on reducing the notion of shape to a set of scalar quantities, shape features. Hence, comparing shapes in this way requires some kind of feature extraction (typically involving segmentation and image registration) and shape similarity is then the Euclidian distance between the shape features. These approaches are plagued with several difficulties. Scalar shape features are often overly simplistic to describe the complex notion of shape, it is impossible to interpret the distance between shape features as a probability of two shapes being equal, and objects whose shapes are to be compared need to be described by the same shape features.

This project seeks to adapt Computational Anatomy to image reconstruction and processing problems in Electron Microscopy imaging. The issues are of two types (i) estimating some representative shapes of a given object (for example antibodies) and their generic geometric variations and (ii) detecting the same kind of objects in a new image of non isolated objects from different nature.

The first step will be based on simulations since once an antibody has been imaged, it is brunt so that no other image can be acquired of the same exemplar. Moreover, the acquired images are 2D images which do not reflect the 3D structure of the antibody. The second question is also

challenging as the 2D images are very noisy, containing many different objects of different nature. The goal is to be able, from the 3D representation of a given object (template and variability), detect some likely projections of elements from the same group.

This work is part of a grant proposal submitted by Ozan Oktem.

8.4. Coupling the atlas with genomic information. In the previous studies, the images we have dealt with were either control patients or identified as suffering from a disease (AD or Parkinson for example). This classification has been done by experts and are related to a pathology. However, several acquisition protocols are now including genomic descriptors as the European project Imagen (http://www.imagen-europe.com/) which includes 2000 teenager images together with their genetic analysis. The goal is to find the invariants of several subpopulations which share common image patterns as well as genetic features. As these two datasets are very different, it will require to change the statistical model so that it involves the genetic informations either as a population or as an individual effect or both. This work can be conducted in collaboration with Neurospin which is part of the Imagen project.

CHAPTER 4

Geometrical variability statistical analysis

*This chapter summarises the works presented in **[A5, C5]**.*

We have seen the large impact of the analysis of deformations in a population through the estimation of a population specific distribution. This normal distribution is a first statistical descriptor of the registrations inside a population. One can wonder in addition if there is some typical patterns of deformations and if the deformations show some specific dependencies between different part of the shape. These questions are developed in the following works.

Contents

1. Probabilistic Independent Component Analysis [A5]

Independent Component Analysis (ICA) is a statistical technique that aims at representing a data set of random vectors as linear combinations of a fixed family of vectors with statistically independent coefficients. This model is based on a decomposition of the $d-$dimensional data on d sources. Contrary to Principal Component Analysis where each direction is weighted by its variance, ICA does not provide a way to reduce the dimension to the "most significant" sources.

Probabilistic ICA has been designed in this way. The observations are decomposed on a small number of highly representative sources (leaving the rest of the information as noise). Interestingly, the sources are very different from PCA orthogonal directions as they are closer to characteristic patterns of the data base. We present here the probabilistic ICA model. We show its large flexibility, in particular allowing for distribution which encode the sparse representation of a database on something (that could be compared to a dictionary) although leading to the same and simple estimation algorithm.

1.1. Observation model of probabilistic ICA. We assume that the observation is a set of vectors which take values in $\mathbb{R}^d$. Let $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ be the training observations, which are assumed to be independent and identically distributed. We will denote by $\boldsymbol{X}$ a generic variable having the same distribution as the $\boldsymbol{X}_k$'s. The jth coordinate of $\boldsymbol{X}$ (resp. $\boldsymbol{X}_k$) will be denoted X^j (resp. X_k^j).

We assume that $\boldsymbol{X}$ can be generated in the form

$$\boldsymbol{X} = \boldsymbol{\mu}_0 + \sum_{j=1}^{p} \beta^j \boldsymbol{a}_j + \sigma \boldsymbol{\varepsilon}\,, \tag{4.1}$$

where $\boldsymbol{\mu}_0 \in \mathbb{R}^d$, $\boldsymbol{a}_j \in \mathbb{R}^d$ for all $j \in \{1, ..., p\}$, $\boldsymbol{\varepsilon}$ is a standard d dimensional Gaussian variable and $\beta^1, \ldots, \beta^p$ are p independent scalar variables, the distribution of which being specified later. Let $\boldsymbol{\beta}$ denote the p-dimensional variable $\boldsymbol{\beta} = (\beta^1, \ldots, \beta^p)$. To each observation $\boldsymbol{X}_k$ is therefore associated hidden realisations of $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$, which will be denoted $\boldsymbol{\beta}_k$ and $\boldsymbol{\varepsilon}_k$. Denote $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)$. It is a d by p matrix and one of the parameters of the model. Another parameter is σ, which will be a scalar in our case (a diagonal matrix being also possible).

The models that we describe are all identifiable, with the obvious restriction that $\boldsymbol{A}$ is identifiable up to a permutation and a sign change of its columns (the latter restriction being needed only when the distribution of $\boldsymbol{\beta}$ is symmetrical). This fact derives from identifiability theorems for factor analysis, like Theorem 10.3.1 in [**40**].

A large range of possible distributions falls into our general model and can be numerically treated the same way as described below. We give in the following paragraph some examples which include the well known cases (as the IFA) and also censored distributions allowing for some coefficients to vanish. The estimation algorithm will be presented later.

1.2. Continuous probability distributions for the ICA coefficients. We start with one of the most popular models, in which each β^j follows a logistic distribution with fixed parameter $1/2$. This is the model introduced in the original paper of Bell and Sejnowsky [**11**], and probably one of the most commonly used parametric model for ICA. One reason for this is that the logistic probability density function (p.d.f.) is easy to describe, smooth, with a shape similar to the

Gaussian, but with heavier, exponential, tails. Note that, for identifiability reasons, one cannot use Gaussian distributions for the components.

A simple variant is to take β^j to be Laplacian with density $e^{-|t|}/2$. The resulting model is very similar to the previous one with similar exponential tails, with the noticeable difference that the Laplacian p.d.f. it is not differentiable in 0. One consequence of this is that it leads to sparse maximum a posteriori reconstruction (not estimation) of the hidden variables (cf. Section 1.5).

Another model can be proposed to increase the tail density. Let $\beta^j = s^j Y^j$ where $\boldsymbol{Y}$ is a standard Gaussian vector, $s^1, \ldots, s^p$ are independent exponential random variables with parameter 1, also independent from Y and $\boldsymbol{\varepsilon}$. In this case, we can write

$$\boldsymbol{X} = \boldsymbol{\mu}_0 + \sum_{j=1}^{p} s^j Y^j \boldsymbol{a}_j + \boldsymbol{\sigma}\epsilon. \tag{4.2}$$

The p.d.f. of $\beta = sY$ tends to infinity at $\beta = 0$, and has sub-exponential tails. It therefore allows more frequent large values of the component coefficients. This may help to overcome the variability in intensity which appears in medical images for examples.

The Independent Factor Analysis (IFA) [**8, 58, 59**] model is a special case of probabilistic ICA in which the distribution of each coordinate β^j is assumed to be a mixture of Gaussians. We will here use a restricted definition of the IFA model which will be consistent with the other distributions that we are considering in this paper, ensuring that the β^j's are independent with identical distribution, and that this distribution is symmetrical.

1.3. Censored distributions for the ICA coefficients. In contrast with the continuous models for which coefficients vanish with probability zero, we now introduce a discrete switch which "turns them off" with positive probability. First, we model the hidden variables as a Gaussian-distributed scale factor multiplied by a Bernoulli random variable. We therefore define $\beta^j = b^j Y^j$, where $\boldsymbol{Y}$ is standard Gaussian and letting b^j have a Bernoulli distribution with parameter $\alpha = P(b^j = 1)$. We assume that all variables $b^1, \ldots, b^p, Y^1, \ldots, Y^p, \varepsilon$ are independent. The complete model for $\boldsymbol{X}$ has the same structure as before, namely

$$\boldsymbol{X} = \boldsymbol{\mu}_0 + \sum_{j=1}^{p} b^j Y^j \boldsymbol{a}_j + \sigma\boldsymbol{\varepsilon}. \tag{4.3}$$

Using a censoring distribution in the decomposition is a very simple way to enforce sparsity in the resulting model. The population is characterised by a set of p vectors, however, each subject is only described by a subset of these p vectors corresponding to the active ones. The probability of the activation of the vectors is given by α. As α increases, the sparsity in the subject decomposition increases as well whereas the dimension to explain the whole training set may remain equal to p. Censored models therefore arise naturally in situations where independent components are not expected to always contribute to the observed signals. This often occurs in spatial statistics, in situations for which observations combine basic components in space, not necessarily occurring all together. We will see an example of such a situation with hand-written digits where components can be interpreted as common parts of some of the digits, but not all, and therefore should not be selected every time. Functional magnetic resonance images (fMRIs), for which ICA methods have

been extensively used [**50, 15, 16**], are also important examples of similar situations. These three-dimensional images indicate active areas in the brain when a subject executes a specific cognitive task. People generally interpret components as basic processing units that interact in a complex task, but these units are not expected to be involved in every task for every subject. Similarly, genomic data, where a gene can activate a protein or not for particular patients, may fall into this context as well.

We now describe some possible variants within the class of censored models.

Combining EG- and BG-ICA, so that a scale factor and a censoring variable intervene together, we get a new complete model for $\boldsymbol{X}$ given by

$$\boldsymbol{X} = \boldsymbol{\mu}_0 + \sum_{j=1}^{p} s^j b^j Y^j \boldsymbol{a}_j + \sigma\boldsymbol{\varepsilon}. \tag{4.4}$$

The previous models include a switch which controls whether the component is present in the observation or not. One may want to further qualify this effect as "activating" or "inhibiting", which can be done by introducing a discrete model for $\boldsymbol{Y}$, each component taking values -1, 0 or 1. We define $\beta^j = s^j Y^j$, where $s^1, \ldots, s^p$ are i.i.d. exponential variables with parameter 1. We let $\gamma = P(Y^j = -1) = P(Y^j = 1)$, providing a symmetric distribution for the components of $\boldsymbol{Y}$. The interpretation of the decomposition is that each component has a fixed effect, up to scale, which can be positive, negative or null. The model can therefore be seen as a variation of the Bernoulli-Gaussian where the effect can be a weighted inhibitor as well as a weighted activator. This allows selective appearance of decomposition vectors and therefore refine the characterisation of the population.

This particular model makes all its sense when trying to model the generation of data with non-zero mean. Going back to our fMRI example, the mean image is more likely to be an active brain since all the patients are subject to the same cognitive task and the activation is always positive or zero. This will create some active areas in the mean brain ($\boldsymbol{\mu}_0$). However, as we already noticed, these areas can be active or not depending on the subject participating to the experiment. This can be modeled by a weighted activation or inhibition of its areas around the mean through the corresponding decomposition vectors. The decomposition vectors are still expected to correspond to the different active zones. This is what this model tries to capture. We will see in the experiments that it also applies to the hand written digits.

Remark 8. *The previous model can be simplified by assuming that the exponential scale factor is shared by all the components*

In these models, the sparsity of the representation will obviously depend on the number of selected components, p, that we suppose given here. When p is too small, it is likely that the model will find that censoring does not help and take $\gamma = 1/2$ (or $\alpha = 1$ in the Bernoulli-Gaussian model). Adding more components in the model generally results in α and γ decreasing, enabling some components to be switched off.

Again, these are only few examples of possible distributions that lead to the same estimation algorithm. Many others can be considered according to the data set the user is analysing. We

now describe the estimator and the estimation algorithm which we use to get the parameters which contain in particular the decomposition vectors.

1.4. Maximum likelihood estimation. Although very different from each other, the previous models are all built using simple generative relations: let Z be the missing variables (including $\boldsymbol{Y}, \boldsymbol{b}, \boldsymbol{s}$ depending on the model), then $\boldsymbol{Z} \to \boldsymbol{\beta}$ and $(\boldsymbol{\beta}, \boldsymbol{\varepsilon}) \to \boldsymbol{X}$. We estimate the model parameters by maximising the likelihood of the observation of n independent samples of $\boldsymbol{X}$ that we denote $\boldsymbol{x}^{*n} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, namely to find

$$\hat{\theta}_n = \operatorname*{argmax}_{\theta} q_{obs}^{*n}(\boldsymbol{x}^{*n}; \theta) \text{ with } q_{obs}^{*n}(\boldsymbol{x}^{*n}; \theta) = \prod_{k=1}^{n} q_{obs}(\boldsymbol{x}_k; \theta). \tag{4.5}$$

Numerically, we use the same MCMC-SAEM with hybrid Gibbs sampler to solve this optimisation problem as presented in Chapter 3, Section 6. Indeed, despite the large range of probability distribution we have presented here, the models all fit into the curved exponential family which allows for the easy use of this algorithm. Moreover, the small number p of sources enables the use of a Gibbs sampler without any numerical bottleneck.

1.5. Reconstruction. Assuming that the parameters in the model are known or have been estimated, the reconstruction problem consists in estimating the hidden coefficients of the independent components, $\hat{\boldsymbol{\beta}} \in \mathbb{R}^p$, based on a new observation of $\boldsymbol{x} \in \mathbb{R}^d$. Reconstruction with probabilistic ICA models is not as straightforward as with complete ICA, for which the operation reduces to solving a linear system. A natural approach is maximum likelihood.

This maximisation is not explicit, although simpler for our first two models. Both cases can be solved efficiently by convex programming. The Laplacian case is similar (up to the absence of normalisation of the columns of $\boldsymbol{A}$) to the Lasso regression algorithm [**66**], and can be minimised using an incremental procedure on the set of vanishing β^j's [**29**].

The other models also involve some form of quadratic integer programming, the general solution of which being NP-complete. When dealing with large numbers of components, one must use generally sub-optimal optimisation strategies (including local searches) that have been developed for this context (see [**48**], for example).

1.6. Experiments. Experiments on a synthetic database are presented and analysed in **[A5]** showing in particular the stability of our models and estimation algorithm with respect to noise. Moreover, it compares our performances to other numerical methods such as FastICA and other EM-based algorithms, demonstrating the advantages of our process.

Some experiments on handwritten digits and face patches showed the importance of the choice of the missing data distribution in particular highlighting the interest in censored ones (cf. **[A5]**).

Lastly, experiments on medical images were done in order to detect differences between patient suffering from Alzheimer's disease and control subjects. The 101 subjects in the data set are separated in 3 groups with 57, 32 and 12 patients, containing healthy patients in the first group and patients with Alzheimer's disease and semantic dementia (denoted the AD group later) at different stages in the last two groups.

Using our algorithm, we have computed $p = 5$ decomposition vectors based on the complete data set. Figures 1 to 3 present these decomposition vectors mapped on the meshed hippocampus for six selected models. The estimated mean is shown on the left side and the five corresponding

decomposition vectors are on the right side. Although results vary with the chosen model, we can see common features emerging.

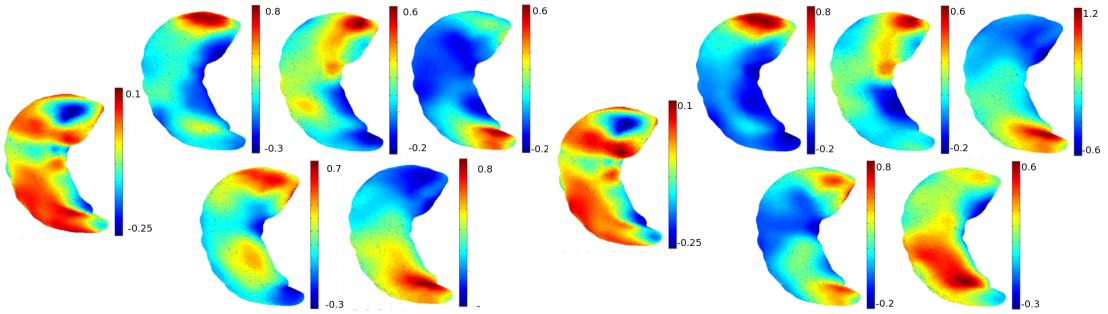


FIGURE 1. Left: Mean (left) and 5 decomposition vectors estimated with the Log-ICA model. Right: Mean (left) and 5 decomposition vectors estimated with the Lap-ICA model. Each image has its own color map to highlight the major patterns.

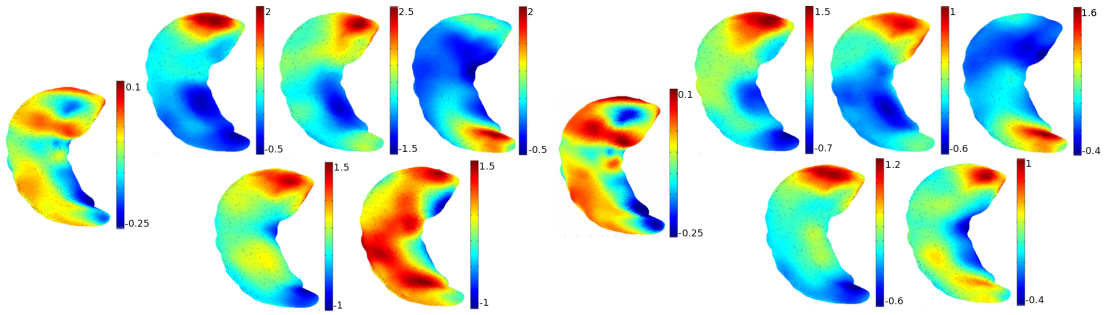


FIGURE 2. Left: Mean (left) and 5 decomposition vectors estimated with the EG-ICA model. Right: Mean (left) and 5 decomposition vectors estimated with the BG-ICA model. Each image has its own color map to highlight the major patterns.

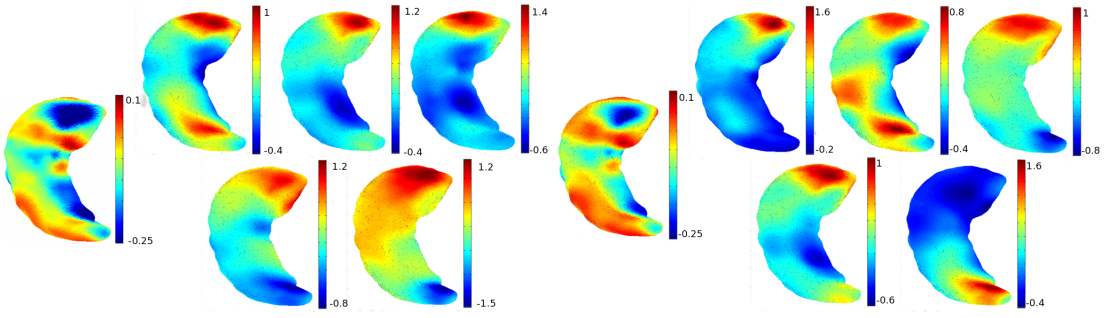


FIGURE 3. Left: Mean (left) and 5 decomposition vectors estimated with the ET-ICA model. Right: Mean (left) and 5 decomposition vectors estimated with the TE-ICA model. Each image has its own color map to highlight the major patterns.

In tables 1 and 2, we provide the p-values obtained from the comparison of the five ICA coefficients (β) among the three subgroups. The test is based on a Hoteling T-statistic evaluated on the coefficients, the p-value being computed using permutation sampling. The test is performed for two different comparisons: first we compare the healthy group with respect to the two pathological

groups together. This is what is shown in Table 1. The second test compares the healthy group with the group of 32 mild AD patients. The results are presented in Table 2. The results show that the estimated features are highly discriminant for the two stage of AD.

Model	Log-ICA	Lap-ICA	EG-ICA	BG-ICA	EBG-ICA
Mean on log $10^{-3}\times$	0.31	0.29	0.27	0.33	0.9
Std deviation on log $10^{-3}\times$	0.16	0.19	0.12	0.25	1.2

Model	ET-ICA	TE-ICA	TEoff-ICA
Mean on log $10^{-3}\times$	0.27	2.4	75.7
Std deviation on log $10^{-3}\times$	0.14	2.9	126.2

TABLE 1. Mean and standard deviation of the p-values for the eight models with the five decomposition vectors shown in Figures (1) to (3). The mean and the standard deviation are computed over 50 samples of the posterior distributions of the hidden variables to separate the first group (Control) with respect to the two others (AZ).

Model	Log-ICA	Lap-ICA	EG-ICA	BG-ICA	EBG-ICA
Mean on log $10^{-3}\times$	9.0	9.6	8.3	10.9	18.7
Std deviation on log $10^{-3}\times$	3.8	4.8	2.7	7.6	17.7

Model	ET-ICA	TE-ICA	TEoff-ICA
Mean on log $10^{-3}\times$	8.9	30.8	148.7
Std deviation on log$10^{-3}\times$	4.6	28.8	160.4

TABLE 2. Mean and standard deviation of the p-values for the eight models with the five decomposition vectors shown in Figures (1) to (3). The mean and the standard deviation are computed over 50 samples of the posterior distributions of the hidden variables to separate the first group (Control) with respect to the second one (mild AZ).

2. Long distance conditional correlations [C5]

The previous study enabled to exhibit some characteristic patterns which appear as a small size dictionnary on which each observation can be decomposed. However, these patters are quire difficult to interpret, in particular considering the hippocampus data, as they show a global deformations of the mean shape. One can wonder if we can extract other type of complementary information from these data, in particular whether some regions of the shape have dependent expansion or shrinkage. However, two major facts appears here: (i) since the deformation is global, all parts of the shape are likely to be dependent from each other and (ii) as the deformation is assumed smooth, when selecting the highest dependencies, the neighbouring voxels are the most likely to appear. Therefore, on has to look for *direct* dependencies (also called conditional dependencies) which are not connecting two neighbouring regions. These were the motivations of the following study.

2.1. Gaussian Graphical Models. Let us consider p points on a given shape that will compose the nodes of a graph. On these points, we observe n random discretisation (with p points) of a quantification of template-to-subject deformations (here the norm of the Jacobian). The p nodes of the graph are thus identified to p random variables denoted $(X_1, ..., X_p)$, this vector is assumed to be distributed as a multivariate Gaussian $\mathcal{N}_p(0, \Sigma)$. The graph G_Σ of conditional dependencies is defined as follows: there exists an edge between nodes a and b if and only if the variables X_a and X_b are dependent given all the remaining variables. This will be denoted $a \overset{G_\Sigma}{\sim} b$. The goal is to estimate this graph G_Σ given a population of observations.

2.2. Estimation processes. The estimation is done by a regression as presented in [**52**]. Let X_a be the a^{th} node of the graph. The goal is to estimate the matrix θ such that:

$$X_a = \sum_{b \neq a} \theta_{a,b} X_b + \varepsilon_a$$

where ε_a is assumed to follow a normal distribution with zero-mean and variance $1/(\Sigma^{-1})_{a,a}$. An important property of this modelling is that $\theta_{a,b} = -\frac{K_{a,b}}{K_{a,a}}$ where $K = \Sigma^{-1}$ is the precision matrix. Therefore $\theta_{a,b} = 0$ is equivalent to $a \overset{G_\Sigma}{\not\sim} b$. The estimation of θ is done by minimising an energy (which is actually the negative penalised log likelihood of the model) given by two terms. The first one is the l^2 distance between the observations and their estimation (coming from the Gaussian distribution of the noise). The second term is a prior on the parameters θ to estimate. In the case we consider, the prior is a Laplacian prior with parameter λ which corresponds to an l^1 penalty on the energy side. This regression technique is known as the *LASSO* algorithm [**67**] (or with slight relaxation, elastic net algorithm [**77**]) and forces the estimate to be a sparse matrix. Including this constraint on the candidate matrices enables to really focus on the most important conditional correlations. Therefore predicting one variable from the others is not much dependent on the training set. Minimising this energy node by node does not produce a symmetric estimated matrix. Therefore, we use different several symmetrisation techniques as proposed [**52**] to produce different graphs.

2.3. Penalised criterion to choose among a family of estimated graphs. In the previous energy, there are several parameters which have to be chosen. This may be very difficult and data depending. A classical choice is to use cross-validation. Giraud in [**31**] suggested instead to set a criterion so that after computing a family of possible graphs with different parameters, it enables to select the best one with respect to this criterion in terms of its accuracy in predicting one variable from the others.

2.4. Non local Gaussian Graphical Models.

2.4.1. *Introduction of G_0.* As noticed above, the neighbouring points - neighbour nodes of the graph- are very likely to be conditionally correlated. We want to put more attention onto the other correlations -which will be called long-distance ones in the sequel since they do not affect the neighbours. To this purpose, we have introduced in the estimation a neighbouring graph G_0 which carries the neighbour nodes of all the graph nodes. We assume that there exist correlations between these points but we are not estimating them rather looking for the other ones. There are two reasons for that. On the one hand, we are not interested in the local correlations since

they appear to be obvious. The long-distance ones however may reveal some important behaviour which may be characteristic from the population we are studying. These non obvious relations between regions may show some non random effect on the shape. This will be illustrated in the experiments.

On the other hand, the value of the penalty is directly related to the complexity of the families of possible candidate graphs. Shrinking the families to those of graphs containing G_0 reduces their complexity and therefore the penalty. The effect it that more edges appear and therefore the long-distance ones.

The way we introduce G_0 in the estimation is as follows. The estimation of the conditional correlations is done in the orthogonal space of the neighbouring graph G_0. This leads to replace the random variable X_a by

$$X_a - X_{m_a}(X_{m_a}^T X_{m_a})^{-1} X_{m_a}^T X_a \,, \tag{4.6}$$

where X^T is the transposition and X_{m_a} is the matrix defined as follow: if we denote X the $n \times p$ matrix of all the data and m_a the list of neighbours of node a in G_0, then $X_{m_a} = X(., m_a)$ is of dimension $n \times card(m_a)$.

This orthogonality constraint may have to be relaxed since the projection may lead to ill-conditioned matrices. Moreover, it will also enable to capture some edges with small projection onto the orthogonal of G_0 which may appear stronger since we know that the local dependencies are more likely to be the strongest and summarise the main information. This can be express in terms of introducing a small ridge (driven by a new parameter γ_0 chosen via the criterion). Denoting Id the identity matrix, this yields

$$X_a - X_{m_a}(X_{m_a}^T X_{m_a} + \gamma_0 Id)^{-1} X_{m_a}^T X_a \,. \tag{4.7}$$

2.4.2. *Selection of graph: Change in the penalty.* When we introduce an a priori graph G_0, we shall change the selection criterion to take into account this a priori knowledge. To fit with this situation, the previous criterion is adapted by changing the penalty part which reflects the change of complexity induced by the graph G_0. It ensures a control of the prediction error similar to Theorem 1 in [**31**].

We have therefore modified the existing package GGMselect in two ways (which is included in the new update of the package). If a prior is given, we change the data so that the new random variable is given by eq. (4.6) or (4.7). Then we adapted the penalty criterion.

2.5. Experiments. The results of the estimation using four of the previously described methods are presented in Fig. 4 where the graph selected by our criterion is highlighted with a red star. It shows that there are long-distance correlations which appear in the deformation patterns in the population (see the pruned graph in Fig. 5 top image). These edges were not appearing when the estimation was not including the prior neighbouring graph. In order to see the different areas which are conditionally correlated, we use the spectral clustering method which are presented in Fig. 5 bottom images. Comparing the sub-populations, it is interesting to see whether they carry differences with respect to conditional correlations. However, since the late AD group is pretty small, we only cluster it into two groups: controls versus AD (both mild and late together). The goal is to see if the disease changes the kind of conditional correlated deformations that are present in a population. This is actually what the preliminary results tend to show.

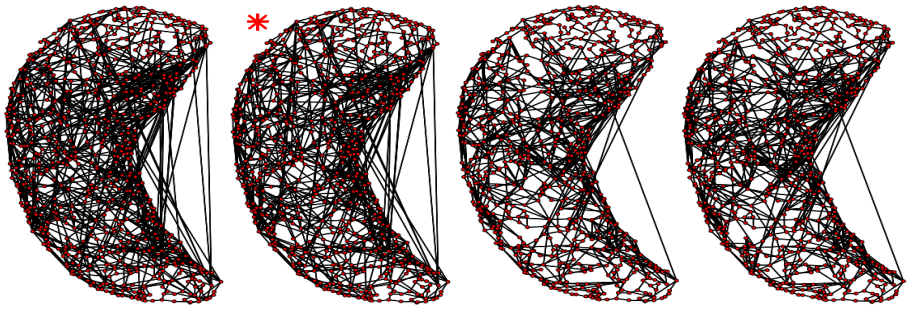


FIGURE 4. Estimated graphs with the eight methods using a neighbouring graph. Some results of the methods presented above. The red star shows the graph selected by our criterion.

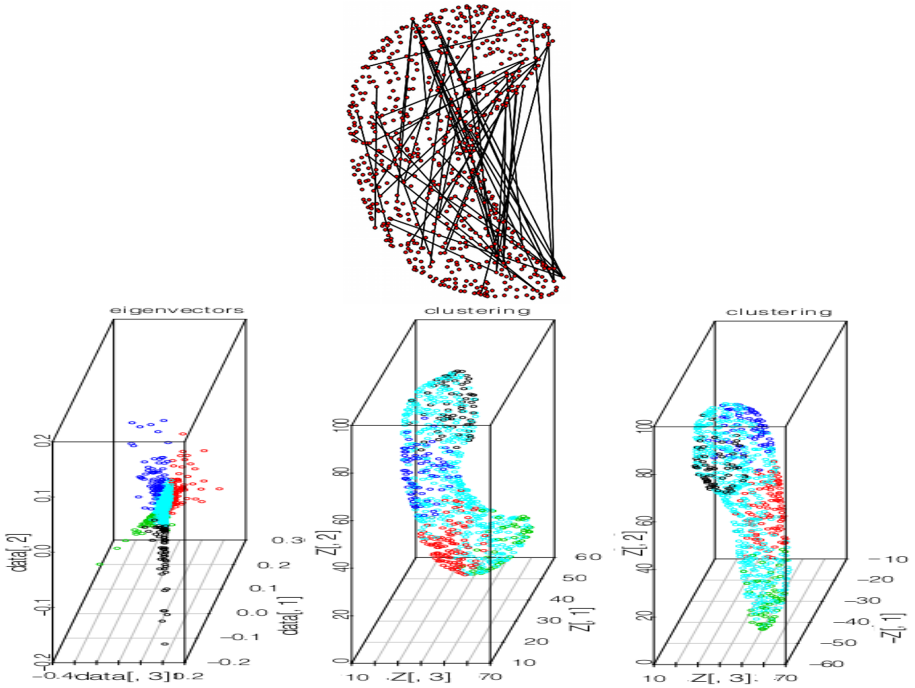


FIGURE 5. Best graph analysed. Left: the edges between close nodes have been removed for better visualisation. Right: k-mean clustering.

3. Future work

3.1. Mixture model with common correlation pattern. The previous Gaussian Graphical model can be naturally extended to a mixture model where the population will be clustered in a unsupervised way. This natural extension becomes really interesting when the correlation patterns can be constrained so that all subpopulations share a common correlation structure and differ from each other by sparse additional correlations. This falls naturally in our analysis of AD vs control hippocampi where the neighbouring correlations are shared by all individuals but the long distance ones may appear only for one subgroup. This generalised model will prevent from fixing a graph that we expect to be shared but which can be doubtful. This writes: $(X_1, ..., X_p)$ is assumed to be distributed as a mixture of multivariate Gaussian distributions $\sum_{k=1}^{K} w_k \mathcal{N}_p(0, \Sigma_k)$ where Σ_k is constrained to be of the form $\Sigma_k = \Sigma_0 + \Gamma_k$ with Γ_k sparse. Other constrains can be set for

example on Σ_0 which can be also sparse or with low rank making the link with recent works from Christophe Giraud [**32**] and many others.

3.2. GGM for vector value or time sequence data. In the GG model the data are real valued. This covers a large variety of samples however, one may want to generalise this approach to vector value inputs for example time sequences on a graph or time sequences of images (the voxels becoming the node of the underlying graphical structure). This comes into play when considering for example fMRI data which are (without any preprocessing) time evolution of the activation in the brain. Electro-encephalograms (EEGs) or magneto-encephalograms (MEGs) are also data which are time sequences, each electrode provides a time sequence of the evolution of activation on the cortex or in the subcortical structures. The interest is now to find correlation patterns between the time evolutions of the voxels.

A first attempt has been done in collaboration with Thomas Fletcher in the University of Utah where the time sequences of fMRIs have been projected on a orthonormal basis to produce almost independent coefficients. These coefficients were then treated as augmented data. This work is still in progress to find first if this method is promising (choosing the correct basis) or if other technics should be considered.

CHAPTER 5

Personalisation of mathematical models to understand pathology evolutions

*This chapter summarises the works presented in **[A6, A9, A11, A16]**.*

For many pathologies, the important effects are highlighted along time. This time evolution may be more or less long and subject to know or not known dynamic. In this chapter, we concentrate on two pathologies and the modelling of their evolution. First is the tumour growth for which the evolution is modelled as a reaction-diffusion phenomenon. The second one is stroke for which the dynamic is not known and which we therefore model as time evolving deformable templates.

Contents

1. Tumour growth estimation [A6]

1.1. Tumour Growth Modelling. Glial cells dynamics are essentially governed by two biological phenomena : proliferation and invasion. They can be jointly modelled by a reaction-diffusion equation which describes the change over time of the normalised tumour cell density u:

$$\begin{cases} \frac{\partial u}{\partial t} = \nabla.\,(\mathbf{D}(\mathbf{x})\nabla u) + \rho u(1-u) \\ \mathbf{D}\,\nabla u.\,n_{\partial\Omega} = 0 \end{cases} \tag{5.1}$$

where ρ is the proliferation rate, $\mathbf{D}$ the local diffusion tensor, and $n_{\partial\Omega}$ is the normal vector at the domain boundary surface $\partial\Omega$. In the first equation, the proliferation of tumour cells follows a logistic growth parameterised by ρ whereas the tumour infiltration into neighbouring neural fibres is captured by an anisotropic diffusion parameterised by $\mathbf{D}$. The second equation indicates that there is no flux of tumour cells outside the domain Ω.

The diffusion tensor characterises the mobility of tumour cells that is considered to be isotropic in grey matter but anisotropic in white matter. More precisely, the tumour diffusion tensor (TDT) may be written as $\mathbf{D}(\mathbf{x}) = d_g\mathbf{I}_3$ in grey matter, where d_g is the diffusivity coefficient. In white matter, we propose to use the following white matter tumour diffusion tensor model:

$$\mathbf{D}(\mathbf{x}) = \ \mathbf{V}(\mathbf{x})[diag(e_1(\mathbf{x})d_w, d_g, d_g)]\mathbf{V}(\mathbf{x})^T \tag{5.2}$$

where d_w is the white matter diffusivity coefficient, $\mathbf{V}(\mathbf{x})$ represents the matrix of sorted eigenvectors of $\mathbf{D}_{water}(\mathbf{x})$, the measured diffusion tensor of water molecules in the tissues, and $e_1(\mathbf{x})$, is the normalised largest eigenvalue (between 0 and 1) of $\mathbf{D}_{water}(\mathbf{x})$. With this choice, tensors have a non-homogeneous anisotropy ratio which is always less than d_w/d_g but is maximised at the centre of the white matter fibres and continuously decreases towards their boundaries.

The reaction-diffusion equations (5.1) are not practical when dealing with clinical images. Indeed, in MR images, tumour cell density u cannot be observed but the visible tumour boundary can. Hence, we use a front motion approximation for the reaction-diffusion equation which was introduced by Konukoglu *et al.* [**42**] assuming that the visible contour is associated with iso-density contour $u = 0.4$ [**68**]. They introduced an Anisotropic Eikonal Equation (AEE) describing the time $T(\mathbf{x})$ at which the evolving tumour front passes through the location $\mathbf{x}$ accounting for the fact that the tumour front speed increases over time to reach an asymptotic value equal to $2\sqrt{\rho\ \mathbf{n}^\star\mathbf{D}\mathbf{n}}$ where $\mathbf{n}$ is the normal direction of the front. Before reaching the asymptotic stage, we use a dynamic formula of the time-evolving speed of the propagating tumour front including the effect of the front curvature $\kappa_{eff}(\mathbf{x})$:

$$F = \frac{4\rho T - 3}{2\sqrt{\rho}T} - 0.3\sqrt{\rho}\left(1 - e^{-|\kappa_{eff}|/0.3\sqrt{\rho}}\right) \tag{5.3}$$

This is especially important at the early stage of the tumour growth when the front is similar to a small sphere.

This last formulation is no longer a Hamilton-Jacobi equation due to the second-order curvature term and therefore cannot be solved by fast sweeping methods such as the Anisotropic Fast Marching (AFM) [**43**]. However, a multi-pass approach was proposed [**42**] to solve efficiently

this equation by applying several times the AFM method while estimating the curvature front from previous iterations. This is the approach we followed.

1.2. Parameter estimation problem from a unique MR image. Based on the previously exposed model, we can simulate the growth of a glioma given its initial source S for which we assume that $T(S) = 0$. From this boundary condition and the knowledge of diffusivity d_w, d_g and proliferation rate ρ we can compute the time $T(\mathbf{x})$ at which the visible tumour front reaches a given point $\mathbf{x}$.

In this work, we are interested in solving the following inverse problem: given a visible tumour boundary S_{Seg} in an MR image, can we extract the growth parameters $S, d_w, d_g, \rho, T_{\mathrm{Obs}}$ that best explain the observed tumour boundary. The duration T_{Obs} between the onset of the tumour and the MR image acquisition is indeed also unknown.

Based on [**42**], it has been already established that several combinations of ρ, d_w, d_g lead to the same front speed and therefore the same tumour growth simulations. Therefore, it is sufficient in this inverse problem to consider a fixed value of the proliferation rate ρ corresponding to the tumour grade and to estimate the source and *diffusivity ratio* also called *spikiness index*:

$$r = \frac{d_w}{d_g}\,.$$

Here, the "spikiness index" represents a biology-driven estimated measure which quantifies the tortuousness of the boundary of the tumour. We consider that $\{S, r\}$ appropriately characterise well a tumour extent if its visible boundary in MRI, S_{Seg}, is an isocontour of the simulated tumour growth initiated at S with diffusivity ratios equal to r. Therefore, we propose to estimate the patient specific parameters by minimising the criterion which estimates the symmetric distance between the observed shape and the generated one with our proliferation model.

Our motivation to use this criterion to get good estimates of our unknowns (S, r) derives from the fact that a tumour boundary (propagating front) is simultaneously defined as an isotime and an isosurface. Thus, to quantify how good is the estimation of the parameters guiding the spatio-temporal evolution of the tumour shape, we need to quantify how closely the simulated isosurface matches the observed one (manually delineated boundary).

In order to efficiently minimise the previously outlined criterion, we use the multidimensional unconstrained minimisation algorithm without gradient introduced by Powell in [**62**]. This algorithm suits our case since our parameters are bounded in both biological and geometrical spaces. Moreover, the derivative of minimisation criterion is not easy to compute. To better study the convergence of this algorithm and evaluate its outcome, 15 tests were performed using synthetic tumours. Further evaluation of this method was then studied using real data: two patients with LGG.

1.3. Experiments. After the optimisation of our criterion, we extract the simulated isocontour which is closest from the visible tumour boundary in MRI. For patient studies, we take advantage of the fact that several FLAIR MR images are available for different time points. We first check the stability of source locations for the M time points per patient and correlate the spikiness index with the observed tumour evolution. To estimate the two target parameters $[S, r]$, the proliferation rate was fixed to $0.012/day$. Table 1 represents the outcome of the estimation method for the four time points (t_1, t_2, t_3, t_4) of one patient. The estimated source remains very stable

	t_1	t_2	t_3	t_4
S	(120, 111, 31)	(120, 111, 31)	(122, 115, 30)	(121, 116, 30)
$d(G, S)$ in mm	2.00	1.73	1.42	5.38
r	41.98	40.47	40.58	41.38

TABLE 1. Outcome of the minimisation algorithm of Patient A which was separately computed at four time points. The distance d denotes the Euclidean distance between the tumour barycenter G and its estimated source S.

which comforted us with our modelling. An important part of the community was assuming that the source of the tumour was necessary its barycenter which appears here subject to caution. The spikiness index is highly correlated with the tumour shape : low index for sphere-like tumours and higher value for star-shape tumour boundary.

Second, we study the possibility to predict further spatial tumour boundaries evolution based on parameter evaluation. More precisely, after estimating patient specific parameters at time t_i, we run a forward simulation using those estimated parameters trying to predict the tumour invasion process as time evolves. We notice a limited discrepancy between the observed contour at time point t_i and the predicted one at t_{i+1} since the symmetric distance between these contours ranges from 1.74 mm to 1.97 mm. Moreover the volume overlap $C_{overlap}$ between the real tumour volume and the extracted one ranges from 85.63% to 88.04%. It is remarkable that the estimated parameters at a specific time point t_i captures the tumour spatial evolution at the successive time point t_{i+1}. As highlighted in Fig. 1 and Fig. 2, the estimated-parameters based prediction shows promising results.

2. Stroke evolution [A11, A16]

Both of the works described in this chapter have been motivated by the application directly. In particular, this one which started from discussions with Joanna Wardlaw, Professor of Applied Neuroimaging in the Division of Clinical Neurosciences, Western General Hospital, Edinburgh. She is a specialist in stroke lesions and wanted mathematical tools to help her understand and discuss several conjectures related to this disease. Therefore, this section highlights the application results rather than the model itself as it was the primary goal.

Stroke is the third leading cause of death in industrialised countries, the second commonest cause of death worldwide, and a major social and financial burden. 80% of stroke is ischemic and so far the only effective treatment is with thrombolytic drugs to dissolve the occluding thrombus. These treatments are potentially hazardous. Many have suggested that their use could be focused on patients who are most likely to benefit by using techniques such as Magnetic Resonance (MR) imaging to identify the location and spatial extent of ischemic tissue and differentiate salvageable from non-salvageable tissue [**60, 75**].

To capture the individual differences and enable a model that is as patient-specific as possible, we chose a spatio-temporal approach that captures individual differences independent of clinical factors such as stroke severity, the effected vascular territory or the location of the occluded artery and that did not require any information except the DWI and PWI lesion outlines. We propose two models to estimate the spatio-temporal scenario of evolution of strokes. The first one is a piecewise

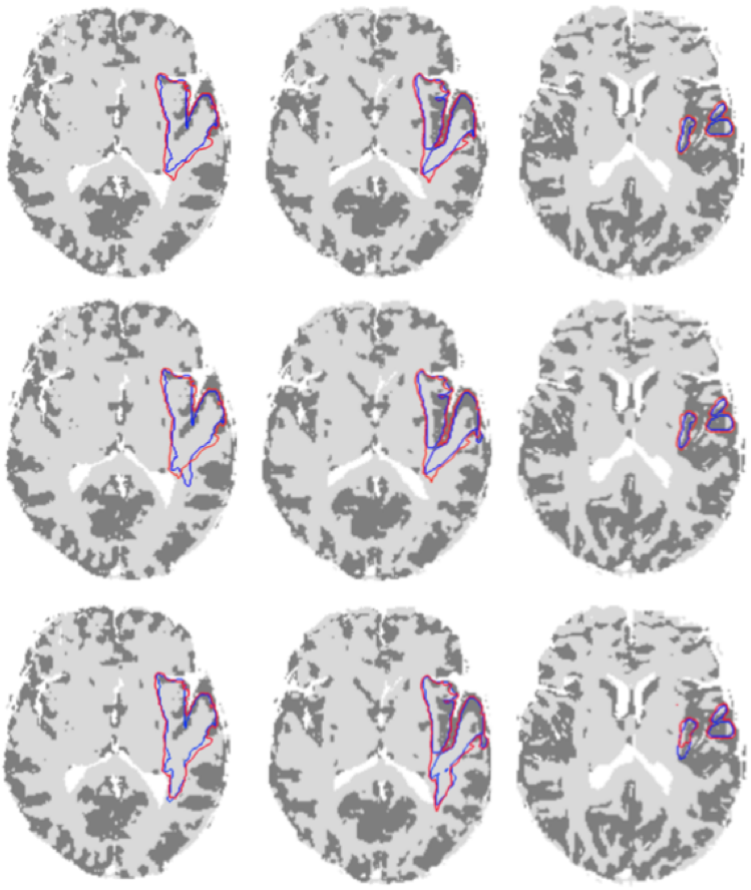


FIGURE 1. Different axial slices of Patient A showing the discrepancy between the real boundary and the estimated one at 3 different successive time points. First row: using the algorithm outcome at t_1, we spatially determine the tumour evolution at t_2. The blue contour represents the manually delineated tumour at the second time point and the red one is the extracted iso-time surface $\hat{S}_{isoTime}$. Second row: using the algorithm outcome at t_2, we predict the tumour evolution at t_3. Tumour boundary at t_3 is coloured in blue and the estimated one in red. Third row: using the algorithm outcome at t_3, we predict the tumour evolution at t_4. Same colours are used to show the similarity between the estimated and the ground-truth tumour contours.

geodesic curve in the LDDMM setting initiated in [**28**]. This provides a first insight about patient specific evolutions and enables to discuss several paradigms from the stroke community. The second model takes into account the fact that stoke lesions may change topology in particular the number of connected components of the lesion is subject to evolution along time. Therefore, we propose to adapt the previous model to metamorphosis [**71**] which is particular adapted to this issue.

2.1. Spatio-temporal model [A11]. In this first model, we used the diffeomorphic model to address three aspects of stroke lesion evolution where imaging could provide valuable insight. Firstly, we modelled individual DWI and PWI lesion dynamic changes separately to identify sub-regional differences in expansion and contraction of the respective DWI and PWI lesions and determined how well the model fitted the true lesion evolution. Secondly, we explored the pattern

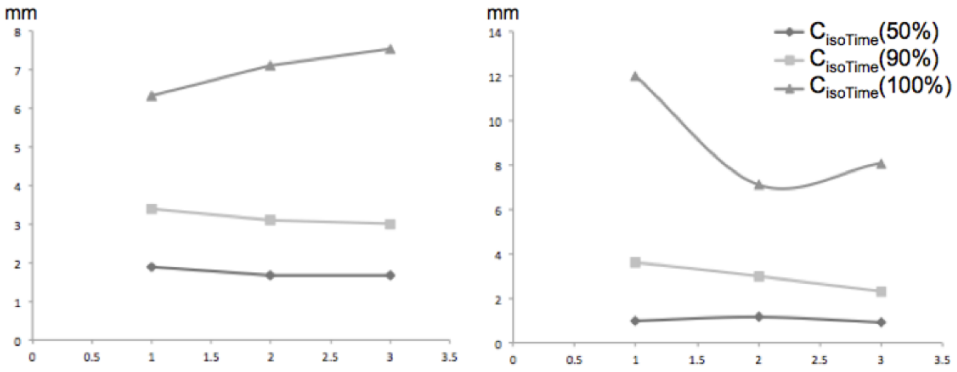


FIGURE 2. Evaluation curves plotted using a distance error measure. For each patient (A or B), we try the predict the evolution at a successive time point based on the estimated parameters at the previous one.

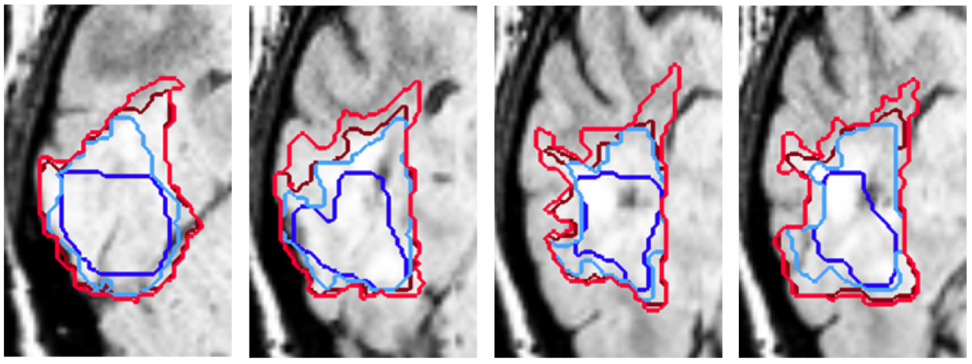


FIGURE 3. Four different axial slices of Patient B showing the remarkable tumour boundary spikiness evolution from the initial time point t_1 to the final one t_4. Tumour contours at successive time points are respectively coloured in: dark blue for t_1, light blue for t_2, dark red for t_3 and light red for t_4.

of change in the DWI and PWI lesions in relation to each other in individual patients, imputed to 3 hourly time intervals, from the three acquired imaging snapshots spaced many days apart. Thirdly, we investigated the model prognostic potential by identifying the time-point at which the PWI/DWI lesions best matched the final observed T2 lesion as an indicative estimate of the ”final” lesion damage.

We use the approach introduced by Durrleman et al. **[27, 28]** to investigating geometric variability of anatomical structures using the framework of currents to model the shapes of the stroke lesions, define the derived distance between shapes and the deformation metric. This sets out a realistic and efficient framework to represent stroke lesion surfaces as dynamic (rather than static) structures which contract, expand, swell, and shrink in different directions, with variation in the speed of these local changes, without imposing prior assumptions. To estimate the pattern of change in ischemia, we use a sequence of time-indexed surfaces $\{S_0, S_1, ..., S_n\}$ from time-points t_0 to t_n.

Considering a source baseline shape S_0 (in our case the 3D meshed lesion surface at the first acquisition timepoint) and $\{x_p\}_{\{1,..,N\}}$ set of points representing the centers of facets of the triangulation of S_0, the goal of the model is to estimate a continuous evolution deformation function $\chi(t)$ deforming the source shape S_0 successively onto the next shape until reaching the final shape S_n: $S(t) = \chi_t(S_0)$. The previously introduced diffeomorphic flow equation guides the deformation function χ^v_{DWI} (resp. χ^v_{MTT}) of the DWI (resp. MTT) lesion shapes, without assuming any prior anatomical spatial constraints or any interdependencies of spatio-temporal MTT/DWI shapes.

To ascertain these smooth trajectories morphing one shape into another passing by the observed timepoints, we introduce a temporal regression functional $J(v)$ that we minimize through running a gradient descent algorithm as described in [**27**]. The minimum of the functional $J(v)$ achieves simultaneously: a) the minimum of the regularity of the deformation that represents physically the kinetic energy of the estimated deformation; and b) a second term that represents the fidelity of the estimated data to the observed data.

The estimation of the "spatiotemporal" variability in DWI and MTT lesions provides us with a clear representation of the similarities and differences between MTT and DWI kinetic patterns, and thereafter to evaluate some biological assumptions.

To show the differences and similarities in the dynamic behaviour of the MTT and DWI lesions as they evolve over time, we need to define appropriate mathematical tools that would adequately set a connection between the changes in the DWI lesion and the MTT lesion as they both evolved with time. This connection is defined as a geodesic diffeomorphism $\phi_t(x)$, that automatically identified corresponding areas in MTT and DWI baseline surfaces. We used the same framework of currents as above to estimate this additional deformation function. Therefore, we can now spatially link both estimated MTT and DWI evolution scenarios, respectively noted as χ^v_{MTT} and χ^v_{DWI}, by using ϕ_{t_1} and its inverse $\phi^{-1}_{t_1}$ as a continuous junction to connect DWI and MTT lesions at the first time point (Fig. 4). We can also follow back in time the spatio-temporal change respectively within each MTT and DWI lesion evolution pattern by using the inverses of the estimated evolution functions $\chi^{-1}_{MTT}(t)$ and $\chi^{-1}_{DWI}(t)$. Now that we can navigate throughout each DWI and MTT 4D evolution scenario separately, and between anatomically corresponding DWI and MTT coordinates at each individual time point, we can extract the kinetic similarities and differences in DWI versus MTT lesion evolution as follows:

a) We firstly compute the norm of the speed at each time step t_i at each point of the continuously evolving lesion 3D surface,

(b) We then extract contracting (corresponding to inward speed direction to the surface) and expanding (corresponding to outward speed direction to the surface) local areas at t_1 of the DWI-estimated evolution scenario $\chi^v_{DWI}(t)$. We compute the mean evolution speed of the initially depicted contraction (*vs.* expansion) areas at each time step of $\chi^v_{DWI}(t)$. Using the mean evolution speed minus (*vs.* plus) its standard deviation as an automatic contraction (*vs.* expansion) threshold, we only mark areas with high contractions (*vs.* expansions) from the DWI lesion surface at t_1 then spatially map them into the MTT lesion at t_1, as deformed by ϕ_{t_1}.

(c) Finally we compute the mean speed for these contracting and expanding areas over their spatio-temporal evolution for the 8 representative patients, in both DWI and MTT abnormalities.

Experiments are presented in Fig. 5 and commented in Section 2.3.

Our study has limitations. Many of the original 48 patients were excluded at this proof-of-concept stage because their lesions consisted of a changing number of components. However scattered multi-focal lesions are common in stroke. For this reason, we proposed to handle this using metamorphosis.

2.2. Spatiotemporal metamorphosis for general stroke analysis [A16]. First, we briefly introduce the theory of metamorphosis, its application to two images (a source $I_0 = I_{temp}$ and a target $I_T = I_{target}$) and present our main contribution: extending a two-image based metamorphosis into handling time series of images. We will not review the extensive literature on its mathematical formulations or numerical solution when applied to data as excellent reviews on this topic can be found in [**57, 71, 76**].

The idea of metamorphosis differs from the large deformation diffeomorphic metric mapping (LDDMM) image deformation scheme in that it allows the template to vary by combining the geometric change, induced by the effect of the group action $\mathbf{g}$ on a manifold M- with photometric residual displacements $\mathbf{Z}$ -also known as the "deformation residual". This defines "metamorphosis" for two-image registration coupling photometric and geometric variability measures between I_{temp} and I_{target}.

Therefore, metamorphosis is given by the pair $\eta = (\mathbf{I_t}, \mathbf{v_t})$ on V and M where $\mathbf{I_t}$ denotes the intensity scalar field associated with image evolution and $\mathbf{v_t}$ belongs to V a Hilbert space of smooth, compactly supported vector fields on M, driving image space deformation. Metamorphosis is then formed on the basis of the LDDMM framework and inherits from several of its properties. In particular, the norm of a curve η on M is defined in [**30, 76**] as follows: $||\eta||^2_M = inf\{|\mathbf{v}|^2_V + \sigma^2|\mathbf{Z}|^2_{\mathbf{M}} : \ \mathbf{Z} = \eta + \nabla\mathbf{I}.\mathbf{v}\}$, where $\mathbf{Z}$ represents the photometric residual and $\mathbf{v}$ the estimated vector field. This new metric designed for metamorphosis is inspired from the advection equation (partial differential equation that governs the motion of a conserved scalar field $\mathbf{I_t}$ as it is advected by a velocity vector field $\mathbf{v_t}$) where the residual image variation $\mathbf{Z_t}$ of the deformation is injected into; therefore, we move from a pure advection equation:

$$d\mathbf{I_t}/d_t + \nabla\mathbf{I_t}.\mathbf{v_t} \ = \ 0 \tag{5.4}$$

to the metamorphosis equation:

$$d\mathbf{I_t}/d_t + \nabla\mathbf{I_t}.\mathbf{v_t} \ = \ \sigma^2\mathbf{Z_t}. \tag{5.5}$$

The term $\nabla\mathbf{I_t}.\mathbf{v_t}$ represents the spatial variation of the moving image I_t in the direction $\mathbf{v_t}$ and σ^2 is an error parameter. Furthermore, the moving intensity scalar field $\mathbf{I_t}$ is defined under the action of the diffeomorphism $\mathbf{g_t}$ on $\mathbf{I_0}$: $\mathbf{I_t} = \mathbf{g_t}.\mathbf{I_0}$. This nicely brings us back to the diffeomorphic flow equation (Eq (5.6)).

$$\begin{cases} \frac{d\mathbf{g_t}(x)}{dt} & = & \mathbf{v_t}(\mathbf{g_t}(x)), \quad t \in [0,T] \\ \mathbf{g_0} & = & \mathbf{Id}\,. \end{cases} \tag{5.6}$$

All the previous equations have the fundamental property to remain geodesic equations in the image manifold M and demonstrated to enable efficient and robust estimation of intensity variation and diffeomorphisms -in case of large deformations [**71**].

Within this well-defined mathematical framework, we expand the work presented in [**30**] from two-image based metamorphosis into time-series based spatiotemporal metamorphosis. Image boundary conditions ($at\ t = 0\ \mathbf{I_0} = I_{temp}\ and\ at\ t = T\ \mathbf{I_T} = I_{target}$) initialises on scalar field $\mathbf{I_t}$ leading to an exact metamorphic matching. Now, considering a set of N images $\{\mathbf{I_0}, ..., \mathbf{I_i}, ..\mathbf{I_T}\}$, we include the new images as we update the initialisation of the intensity field $\mathbf{I_t}$ using piecewise geodesic and continuous interpolation. These intensity constraints force the metamorphosis to spatially identify these new observations. We optimise this matching and find the piecewise geodesic path on the Riemannian manifold M connecting all these images. This can be defined as a variational problem over the intensity scalar field $\mathbf{I_t}$ and the deformation vector field $\mathbf{v_t}$. To estimate optimal metamorphic fields $(\mathbf{I_t}, \mathbf{v_t})$, we define the energy of metamorphosis as introduced in [**30**]:

$$\begin{aligned} J(\mathbf{I}, \mathbf{v}) &= \int_0^1 ||\mathbf{v_t}||_V^2 dt + \sigma^2 \int_0^1 ||\mathbf{Z_t}||_{L^2}^2 dt \\ &= \int_0^1 ||\mathbf{v_t}||_V^2 dt + \tfrac{1}{\sigma^2} \int_0^1 ||d\mathbf{I_t}/d_t + \nabla \mathbf{I_t}.\mathbf{v_t}||_{L^2}^2 dt\,. \end{aligned} \tag{5.7}$$

under the constraints that the geodesic path passes through the observations between t_0 and t_T.

Minimising this energy $J(\mathbf{I}, \mathbf{v})$ under these constraints using gradient descent algorithm estimates the spatiotemporal metamorphosis.

2.3. Experiments.

2.3.1. *Diffeomorphic scenario.* Despite the small sample of only eight representative patients, we see wide variance in the spatiotemporal interaction between PWI and DWI lesion surfaces in terms of correspondence between areas of high contraction and expansion. We also see significant dynamic changes in MTT lesions, including expansion as well as contraction including in 6 of 8 patients the DWI lesion expanding more rapidly in areas of MTT expansion than in areas where MTT is static. Thus, as well as dynamic contraction and expansion patterns that were in line with that expected from mismatch theory (Figure 2), we see the DWI lesion surface contracting faster than the corresponding MTT, and the hypoperfused MTT surface expanding faster than corresponding DWI. This means that this approach can be used to understand lesion evolution in much greater detail and in relation to anatomic, patient-specific, stroke-specific and treatment-specific factors than is possible through analysis of 2D or 3D image data. Through analysis of evolution of mean speed of lesion change at different time-points of DWI and MTT lesions, we identify an overall common pattern implying that diffusion lesions tend to change more slowly than MTT lesions. We also see that DWI hyperintense tissue can contract (DWI reversal phenomenon), consistent with recent data [**44**]. There are variable rates of expansion and contraction of the DWI lesion demonstrating that the DWI lesion surface change is heterogeneous [**61**].

The model also highlights wide inter-patient variability in the time at which the estimated MTT and DWI evolution scenarios match the final T2 lesion. Larger studies using this approach and an accurate analysis of the speed of DWI and PWI lesion evolution in a much more diverse range of patients are now justified to determine reasons for these variations in lesion behaviour.

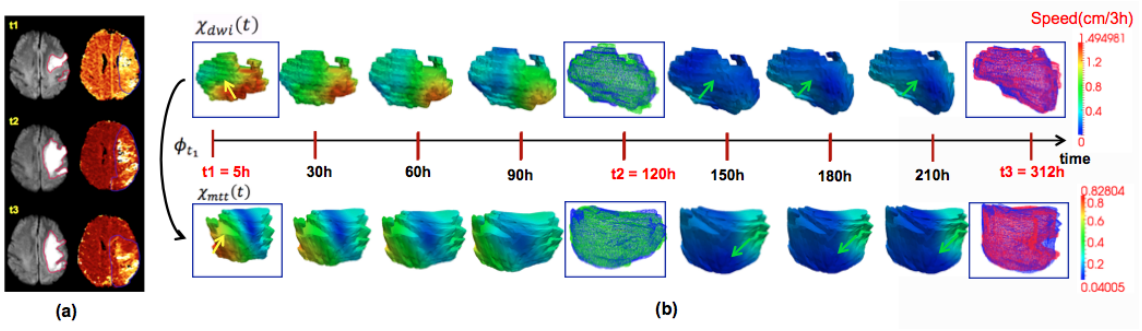


FIGURE 4. *Estimated DWI/MTT lesion evolution scenarios.*(a) DWI (vs. MTT) axial slice at 3 time-points in the first (vs. second) column. (b) Estimated DWI/MTT evolution scenarios with estimated velocity norm (speed) in cm/3h. At t2 (vs. t3), the green (vs. red) surface represents the observed lesion and the blue one represents the estimated lesion. Green arrows represent contraction areas and yellow ones expansion areas.

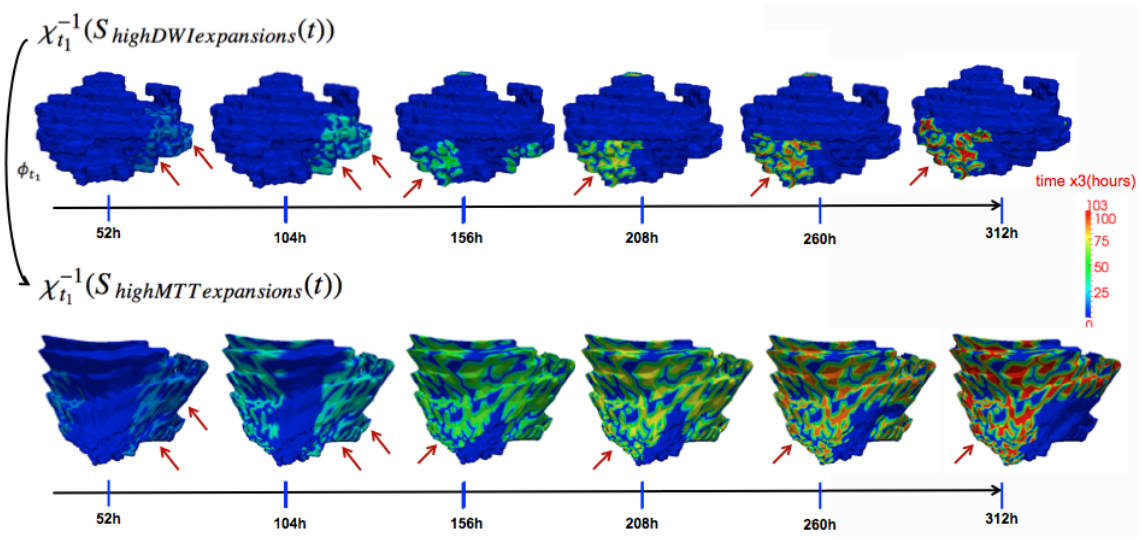


FIGURE 5. *Time-indexed penumbra and core lesion boundaries combined to final T2-w boundary imaged at* $>$ 1 *month.* On the right: MTT manually delineated penumbra boundaries consecutively at t_1 (in orange), t_2 (in yellow) and t_3 (in red). On the left: DWI lesion boundaries ranging from light to dark blue as time evolves from t_1, to t_2 then t_3. In both MR axial slices, the final T2-boundary is displayed in green shading.

2.3.2. *Metamorphosis scenario.* Figure 7 shows the different steps for our data analysis: first, we simulate the spatiotemporal evolution of scattered DWI lesion evolution to track both intensity and velocity changes in the lesion (step 1). Then, we generate the total DWI deformation magnitude map from t1 to t3 and used its absolute values, without considering the direction of the deformation. For each voxel, we plot its mean MTT perfusion value against its mean absolute DWI deformation magnitude (step 2). In each of the 14 patients, the generated data fit into Gaussian curves ($RMSE = 0.0206 + / - 0.0104$) although their means and standard deviations

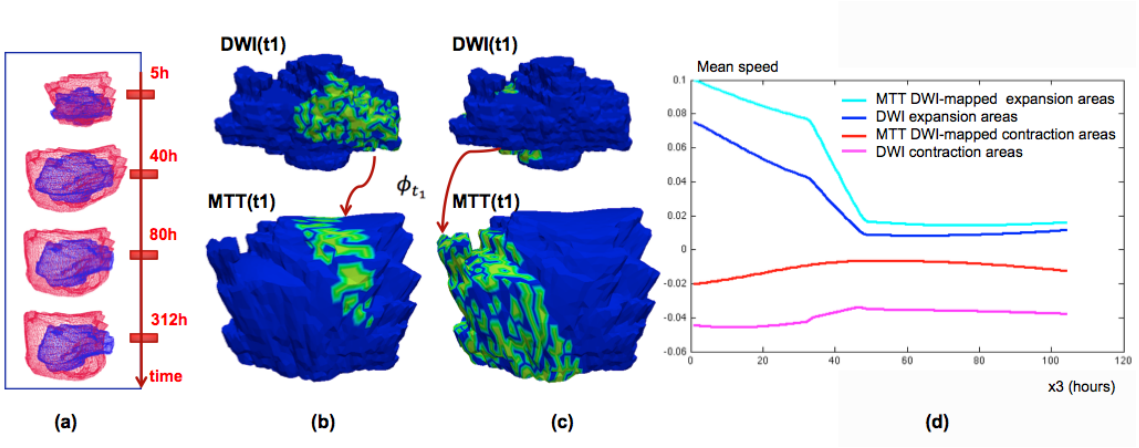


FIGURE 6. *Comparison between highly expanding DWI and MTT areas extracted across the spatiotemporal estimated scenarios.* As the color ranges from blue to green to red, the different successive areas with high expansion speed are 'lit up' at each time step, all mapped back at the baseline lesion shape acquired at t_1 using the inverse of the estimated evolution function $\chi_{t_1}^{-1}(S_{highDWIexpansions}(t))$ for DWI lesion and $\chi_{t_1}^{-1}(S_{highMTTexpansions}(t))$ for MTT lesion. The mapping-back enables us to foresee the upcoming dynamic changes –more precisely expansions– on the acute baseline lesion surface. The red arrows point to highly expanding areas in DWI lesion and their corresponding areas that also expanded in the MTT lesion. The dark blue areas –that didn't change colour as time evolved– did not expand *ie* they contacted.

differ between patients indicating differing perfusion thresholds for the point at which brain tissue becomes at risk or dies between patients. This key finding shows that MTT patient-specific perfusion values seem to reflect the magnitude of DWI lesion evolution from acute to subacute stages. Indeed, the Gaussian distribution defines two MTT perfusion thresholds ($p_1 = \mu_g + 2\sigma_g$ and $p_2 = \mu_g - 2\sigma_g$ with μ_g the mean of the fitted Gaussian curve and σ_g its standard deviation) that present a novel way for discriminating three statuses of abnormal MTT tissue: i) fast MTT perfusion values less than p_1, *ie.* good perfusion, combined with very low DWI deformation magnitude (quasi-static DWI lesion) indicate that this part of ischemia is oligaemic (a potential implication of this observation is that DWI lesion may also contain some viable tissue), ii) MTT perfusion values extending from p_1 to p_2 with high DWI deformation magnitude indicate that the corresponding tissue is rapidly dying, and iii) MTT perfusion values greater than p_2, *ie.* extremely low perfusion, that are associated with very low DWI deformation magnitude show that the undermined tissue is dead. This is in line with the finding in [**17**], where MTT lesion was demonstrated to include parts of ischemic tissue that is dead or oligaemic.

The patient variation in Gaussian means μ_g and their corresponding standard deviations σ_g was very important in our data which tends to indicates that we cannot define universal absolute thresholds for discriminating dead from penumbral or penumbral from oligaemic tissue.

3. Future work

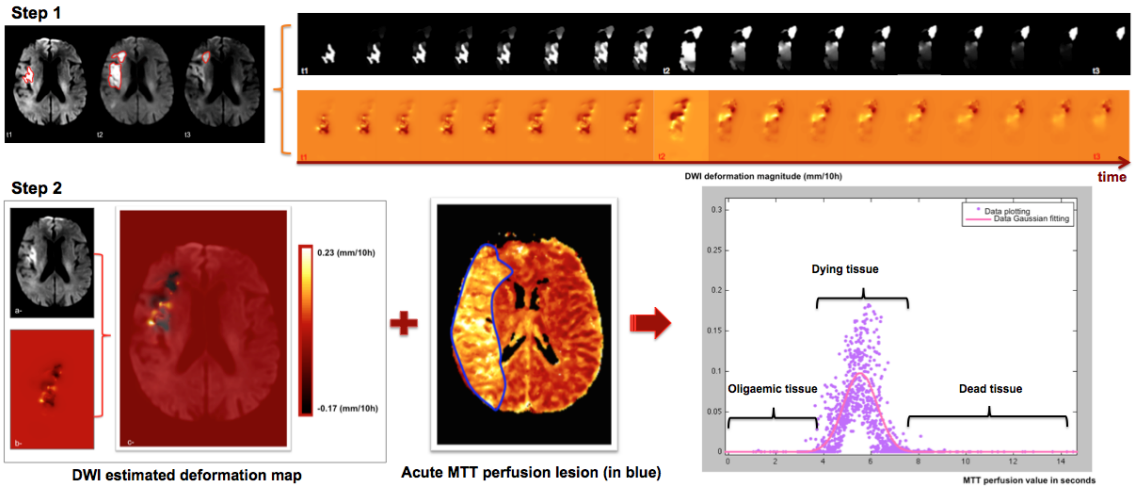


FIGURE 7. *Algorithm pipeline for each patient:* Step 1) Using three successive DWI acquisition time-points, we estimate the spatiotemporal metamorphosis for each patient through estimating the intensity scalar field (top row) and the velocity field (bottom row). Step 2) We consider the absolute values of the generated deformation magnitude map (right) to plot the mean deformation magnitude associate with each perfusion value in MTT lesion (left).

3.1. From personalisation to group analysis. The previous models try to mimic the evolution of a pathology for each patient individually. This is of course a first step to understand the disease itself and try to detect interesting behaviours. The second step is now to analyse these evolutions for a group of patients. These two step can be compared to the registration - template estimation questions. Indeed, in order to analyse a population of images through the estimation of a template, one had first to look for a relevant tool which enabled to highlight the differences between a pair of images. This tool appeared to be the registration mapping. Given this comparison technique, the group analysis became possible. The personalisation of mathematical model appears to be the similar step to what is the "registration step" in the sense that it enables to exhibit tools which could now be used to perform population analysis.

Thanks to the spikiness index and the source location, one can now build models which would enable to cluster the tumours, relate this to potential surgery, etc. Concerning strokes, the modelling is still challenging as it has now to be adapted to patients who were treated with thrombolytic drugs or surgery. In a following step, one have to extract form the evolution scenario some characteristic elements which would be used in a group model.

3.2. Stochastic PDEs. A last challenge would be to introduce random effects in the reaction-diffusion equations so that it better represent the evolution of the tumour (or equivalently in a model for stroke) in each patient. This part requires a close collaboration with medical doctor specialist of each disease to identify the sources of randomness which is relevant to include in the model.

Bibliography

[1] G. Aldridge, J. Ratnanather, M. Martone, M. Terada, M. Beg, L. Fong, E. Ceyhan, A. Kolasny, T. Brown, E. Cochran, S. Tang, D. Pisano, M. Vaillant, M. Hurdal, J. Churchill, W. Greenough, M. Miller, and M. Ellisman. Semi-automated shape analysis of dendrite spines from animal models of fragilex and parkinson's disease using large deformation diffeomorphic metric mapping. *Society for Neuroscience Annual Meeting, Washington DC*, 2005.

[2] S. Allassonnière, Y. Amit, and A. Trouvé. Toward a coherent statistical framework for dense deformable template estimation. *JRSS*, 69:3–29, 2007.

[3] S. Allassonnière, E. Kuhn, and A. Trouvé. Bayesian consistent estimation in deformable models using stochastic algorithms: Applications to medical images. *Journal de la Société Française de Statistique*, 151(1):1–16, 2010.

[4] S. Allassonnière, A. Trouvé, and L. Younes. Geodesic shotting and diffeomorphic matching via textured meshes. In A. Y. Anand Rangarajan, Baba Vemuri, editor, *Proc. of the Energy Minimization Methods for Computer Vision and Pattern Recognition (EMMCVPR 05)*, pages 365–381, November 9-11 2005.

[5] C. Andrieu, E. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim.*, 44(1):283–312 (electronic), 2005.

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

[7] Y. Atchadé. An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. *Methodol. Comput. Appl. Probab.*, 8:235–254, 2006.

[8] H. Attias. Independent factor analysis. *Neural Computation*, 11:803–851, 1999.

[9] R. Bajcsy and C. Broit. Matching of deformed images. In *The 6th international conference in pattern recognition*, pages 351–353, 1982.

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

[11] A. J. Bell and T. J. Sejnowski. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1004–1034, 1995.

[12] P. Bickel and E. Levina. Covariance regularization by thresholding. *Annals of Statistics*, 36(6):2577–2604, 2008.

[13] L. Bookstein, F. *Morphometric tools for landmark data; geometry and biology*. Cambridge University press, 1991.

[14] O. Bremond, E. Moulines, and J.-F. Cardoso. Séparation et déconvolution aveugle de signaux bruités: modélisatin par mélange de gaussiennes. *GRETSI*, 1997.

[15] V. Calhoun and T. Adali. A method for making group inferences from functional MRI data using independent component analysis. *Human Brain mapping*, 14(3):140–151, 2001.

[16] V. Calhoun, T. Adali, and V. McGinty. fMRI activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. *NeuroImage*, 14(5):1080–1088, 2001.

[17] E. Carrera, J. Jones, P.S.and Alawneh, I. Mikkelsen, T. Cho, S. Siemonsen, J. Guadagno, K. Mouridsen, L. Ribe, and N. Hjort.

[18] E. Ceyhan, L. Fong, T. Tasky, M. Hurdal, M. Beg, M.F.and Martone, and J. Ratnanather. Type-specific analysis of morphometry of dendrite spines of mice. *5th Int. Symp. Image Signal Proc. Analysis, ISPA*, pages 7–12, 2007.

[19] E. Ceyhan, R. Ölken, L. Fong, T. Tasky, M. Hurdal, M. Beg, M. Martone, and J. Ratnanather. Modeling metric distances of dendrite spines of mice based on morphometric measures. *Int. Symp on Health Informatics and Bioinformatics*, 2007.

[20] L. B. D and B. E. Imagerie de diffusion in-vivo par résonance magnétique nucléaire. *C R Acad Sci (Paris)*, 301,15:1109–1112, 1985.

[21] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27(1):94–128, 1999.

[22] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27(1):94–128, 1999.

[23] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1:1–22, 1977.

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1:1–22, 1977.

[25] S. Durrleman, P. Fillard, X. Pennec, A. Trouvé, and N. Ayache. Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NeuroImage*, 55(3):1073 – 1090, 2011.

[26] S. Durrleman, X. Pennec, A. Trouvé, and N. Ayache. Statistical models on sets of curves and surfaces based on currents. *MedIA*, 13(5):793–808, 2009.

[27] S. Durrleman, X. Pennec, A. Trouvé, N. Ayache, and J. Braga. Comparison of the endocranial ontogenies between chimpanzees and bonobos via temporal regression and spatiotemporal registration. *Journal of human evolution*, 2011.

[28] S. Durrleman, X. Pennec, A. Trouvé, J. Braga, G. Gerig, and N. Ayache. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *International Journal of Computer Vision*, 103(1):22–59, 2013.

[29] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 2004.

[30] L. Garcin and L. Younes. Geodesic image matching: A wavelet based energy minimization scheme. *EMMCVPR LNCS*, 3757:349–364, 2005.

[31] C. Giraud. Estimation of gaussian graphs by model selection. *Electron. J. Stat.*, 2:542–563, 2008.

[32] C. Giraud. Low rank multivariate regression. *Electronic Journal of Statistics*, 5:775–799, 2011.

[33] M. Girolami and B. Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B*, 73(2):1–37, 2011.

[34] J. A. Glaunes. *Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et lánatomie numérique*. PhD thesis, Université Paris 13, 2005.

[35] U. Grenander. *General Pattern Theory*. Oxford Science Publications, 1993.

[36] U. Grenander and M. I. Miller. Representation of knowledge in complex systems (with discussion section). *J. Royal Stat. Soc.*, 56(4):569–603, 1994.

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

[38] R. Holm, D, T. Ratnanather, J, A. Trouvé, and L. Younes. Soliton dynamics in computational anatomy. *Neuroimage*, 23:S170–S178, 2004.

[39] S. Joshi and M. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE transactions in image processing*, 9(8):1357–1370, 2000.

[40] A. Kagan, Y. Linnik, and C. Rao. Characterization problems in mathematical statistics. *Wiley*, 1973.

[41] S.-G. Kim and et al. Stuctural connectivity via the tensor-based morphometry. *Biannual Meeting of Korean Society of Human Brain Mapping*, 2010.

[42] E. Konukoglu, O. Clatz, B. H. Menze, M.-A. Weber, B. Stieltjes, E. Mandonnet, H. Delingette, , and N. Ayache. Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations. *IEEE Transactions on Medical Imaging*, 29(1):77–95, 2009.

[43] E. Konukoglu, M. Sermesant, O. Clatz, J. Peyrat, H. Delingette, and N. Ayache. A recursive anisotropic fast marching approach to reaction diffusion equation: application to tumor growth modeling. *In Proceedings of the 20th international conference on Information processing in medical imaging, Springer-Verlag*, pages 687–99, 2007.

[44] P. Kranz and J. Eastwood. Does diffusion-weighted imaging represent the ischemic core? an evidence-based systematic review. *American Journal of Neuroradiology*, 30(6):1206–1212, 2009.

[45] E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM Probab. Stat.*, 8:115–131 (electronic), 2004.

[46] S. Lauritzen. Graphical models. 17, 1996.

[47] H. Lee and et al. Discriminative persistent homology of brain networks, 2011.

[48] D. Li and X. Sun. *Nonlinear Integer Programming*. 2006.

[49] F. Maire, S. Lefebvre, E. Moulines, and R. Douc. Aircraft classification with low infrared sensor. *Statistical Signal Processing Workshop (SSP), IEEE*, 2011.

[50] S. Makeig and T. Jung. Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, 94(20):10979–10984, 1997.

[51] T. Marshall and G. Roberts. An adaptive approach to langevin MCMC. *Statistics and Computing*, 22 (5):1041–1057, 2012.

[52] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 2006.

[53] P. Michor and D. Mumford. Riemannian geometries on spaces of planar curves. *J. Eur. Math. Soc.*, 2005. To appear.

[54] I. Miller, M and L. Younes. Group action, diffeomorphism and matching: a general framework. *Int. J. Comp. Vis*, 41:61–84, 2001. (*Originally published in electronic form in: Proceeding of SCTV 99, http://www.cis.ohio-state.edu/ szhu/SCTV99.html*).

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

[56] M. I. Miller, A. Trouvé, and L. Younes. Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*, 24(2):209–228, 2006.

[57] M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41(1/2):61–84, 2001.

[58] J. W. Miskin and D. MacKay. Ensemble learning for blind source separation. *Independent Component Analysis: Principle and Practice*, S. Roberts and R. Everson (Eds), Cambridge University Press:209–233, 2001.

[59] E. Moulines, J.-F. cois Cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. volume 5.

[60] T. Neumann-Haefelin, H. Wittsack, F. Wenserski, M. Siebler, R. Seitz, U. Modder, and H. Freund. Diffusion - and perfusion-weighted MRI. the DWI/PWI mismatch region in acute stroke. *Stroke*, 30:1591–1597, 1999.

[61] T. Phan, G. Donnan, V. Srikanth, J. Chen, and D. Reutens. Heterogeneity in infarct patterns and clinical outcomes following internal carotid artery occlusion. *Archives of neurology*, 66(12):1523, 2009.

[62] M. Powell. The bobyqa algorithm for bound unconstrained optimization without derivatives. *Cambridge NA Report NA2009/06*, University of Cambridge, 2009.

[63] F. Richard, A. Samson, and C. A. Cuenod. A SAEM algorithm for the estimation of template and deformation parameters in medical image sequences. *Statistics and Computing*, 19:465–478, 2009.

[64] N. Singh, P. Fletcher, J. Preston, L. Ha, R. King, J. Marron, M. Wiener, and S. Joshi. Multivariate statistical analysis of deformation momenta relating anatomical shape to neuropsychological measures. In T. Jiang, N. Navab, J. Pluim, and M. Viergever, editors, *Proc. MICCAI'10*, volume 6363 of *Lecture Notes in Computer Science*, pages 529–537. Springer, 2010.

[65] D. W. Thompson. *On growth and form /*. Cambridge :University Press ;, 1917.

[66] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B*, 58, No. 1:267–288, 1996.

[67] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B*, 58 (1):267–288, 1996.

[68] P. Tracqui, G. Cruywagen, D. Woodward, G. Bartoo, J. Murray, and J. E. Alvord. A mathematical model of glioma growth: the effect of chemotherapy on spatio-temporal growth. *Cell Proliferation*, 28(1):17–31, 1995.

[69] A. Trouvé. An infinite dimensional group approach for physics based model. Technical report, 1995.

[70] A. Trouvé. Diffeomorphism groups and pattern matching in image analysis. *Int. J. of Comp. Vis.*, 28(3):213–221, 1998.

[71] A. Trouvé and L. Younes. Metamorphoses through lie group action. *Foundations of Computational Mathematics*, 5(2):173–198, April 2005.

[72] D. Tuch. Q-ball imaging. *Magn Reson Med*, 52:1358–1372, 2004.

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

[74] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *Neuroimage*, 45:61–72, 2009.

[75] J. Wardlaw. Neuroimaging in acute ischaemic stroke: insights into unanswered questions of pathophysiology. *Journal of internal medicine*, 267(2):172–190, 2010.

[76] L. Younes. Shapes and diffeomrophisms. *Springer*, 2011.

[77] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 2005.

[78] M. Uzumcu, A. F. Frangi, J. H. Reiber, and B. P. Lelieveldt. Independent component analysis in statistical shape models. *SPIE Medical Image Analysis*.

N° d'impression :

Représentation et Estimation Statistique de Modèles Déformables pour la Reconnaissance de Forme et l'Anatomie Numérique

Résumé :

Mots-clés : Modèles déformables, tir géodésic, modèles statistiques, estimation et apprentissage statistique, consistence.

Representation and Statistical Estimation of Deformable Template Models for Shape Recognition and Computational Anatomy

Abstract:

Keywords: Deformable template, geodesic shooting, statistical model, estimation and statistical learning, consistency.

AMS Classification: