



# Generative diffeomorphic modelling of large MRI data sets for probabilistic template construction



Claudia Blaiotta<sup>a,\*</sup>, Patrick Freund<sup>a,b</sup>, M. Jorge Cardoso<sup>c</sup>, John Ashburner<sup>a</sup>

<sup>a</sup> Wellcome Trust Centre for Neuroimaging, University College London, London, UK

<sup>b</sup> Spinal Cord Injury Center Balgrist, University Hospital Zurich, University of Zurich, Zurich, Switzerland

<sup>c</sup> Translational Imaging Group, CMIC, University College London, London, UK

## ARTICLE INFO

### Keywords:

Atlas construction  
Image segmentation  
Image registration  
Generative modelling  
MRI

## ABSTRACT

In this paper we present a hierarchical generative model of medical image data, which can capture simultaneously the variability of both signal intensity and anatomical shapes across large populations. Such a model has a direct application for learning average-shaped probabilistic tissue templates in a fully automated manner. While in principle the generality of the proposed Bayesian approach makes it suitable to address a wide range of medical image computing problems, our work focuses primarily on neuroimaging applications. In particular we validate the proposed method on both real and synthetic brain MR scans including the cervical cord and demonstrate that it yields accurate alignment of brain and spinal cord structures, as compared to state-of-the-art tools for medical image registration. At the same time we illustrate how the resulting tissue probability maps can readily be used to segment, bias correct and spatially normalise unseen data, which are all crucial pre-processing steps for MR imaging studies.

## Introduction

Over the past two decades neuroimaging techniques have been widely recognised as powerful non-invasive tools to answer neuroscientific questions regarding, for instance, the variability of neuroanatomy across different populations (Good et al., 2001; Ashburner and Friston, 2000), the role of structural changes in determining disease onset and progression (Fox et al., 2000; Chetelat et al., 2005; Brex et al., 2002), as well as the relationships between neural activations and mental functions (Friston et al., 1996) during common cognitive, somatosensory or motor tasks.

One of the main challenges, which is encountered in all neuroimaging studies, originates from the difficulty of mapping between different anatomical shapes. In particular, a fundamental problem arises from having to ensure that this mapping operation preserves topological properties and that it provides, not only anatomical, but also functional overlap between distinct instances of the same anatomical object (Brett et al., 2002).

This explains the rapid development of the discipline known as computational anatomy (Grenander and Miller, 1998), which aims to provide mathematically sound tools and algorithmic solutions to model high-dimensional anatomical shapes, with the ultimate goal of encoding,

or accounting for, their variability. In this framework, the problem of anatomical mapping is most commonly addressed by assuming that a population of subjects can be fully described by means of an average-shaped template, together with a set of high-dimensional transformations, which, once applied to the template, generate the observed anatomical configurations (Thompson et al., 2000). In this setting, an ideal template is one that can accurately be matched to the observed data while minimising the geodesic distance between the template itself and each observation (Avants and Gee, 2004).

One of the practical advantages of this approach is that, once a template has been constructed, it can be used to model unseen data drawn from the same population, thus providing an ideal reference coordinate system for neuroimaging studies and for statistical testing of neuroscientific hypotheses.

In this paper we propose a general modelling scheme and a training algorithm, which, given a large cross-sectional data set of MR scans, can learn a set of average-shaped tissue probability maps, either in an unsupervised or semi-supervised manner. This is achieved by building a hierarchical generative model of MR data, where image intensities are captured using multivariate Gaussian mixture models, after diffeomorphic warping (Ashburner and Friston, 2011; Joshi et al., 2004) of a set of unknown probabilistic templates, which act as anatomical priors. In

\* Corresponding author.

E-mail address: [claudia.blaiotta.13@ucl.ac.uk](mailto:claudia.blaiotta.13@ucl.ac.uk) (C. Blaiotta).

addition, intensity inhomogeneity artefacts are explicitly represented in our model, meaning that the input data does not need to be bias corrected prior to model fitting.

Our work builds on a number of state-of-the-art techniques, some of which were already explored and validated individually as part of our previous work (Ashburner and Friston, 2005, 2011; Blaiotta et al., 2016). In particular, we rely on the variational Bayes expectation-maximisation (VBEM) algorithm for image segmentation proposed in Blaiotta et al. (2016), which is a robust generalisation of the segmentation method implemented in the SPM12 software (Ashburner and Friston, 2005). Additionally, we exploit the diffeomorphic image registration framework described in Ashburner and Friston (2011), which is an accurate, fast converging and memory efficient strategy to align complex anatomical shapes using diffeomorphic transformations.

To the best of our knowledge, the particular mathematical formulation that we adopt to combine such modelling techniques has never been adopted before. The resulting approach enables processing simultaneously a large number of MR scans in a groupwise fashion and particularly it allows the tasks of image segmentation, image registration, bias correction and atlas construction to be solved by optimising a single objective function, with one iterative algorithm. This is in contrast to a commonly adopted approach to mathematical modelling, which involves a pipeline of multiple model fitting strategies that solve sub-problems sequentially, without taking into account their circular dependencies. Our strategy instead, addresses the fact that learning anatomical priors requires collecting accurate segmentations from a training population, while such segmentations can only be obtained if suitable priors are available. Thus, the two modelling problems are solved more effectively within an integrated framework, rather than independently.

Related work was done by Bhatia et al. (2007), however their method relies on classical point estimation techniques to perform image segmentation, as opposed to our variational approach, which instead allows estimating full posteriors on the intensity distribution parameters. Additionally we incorporate bias correction and also explore semi-supervised learning, as opposed to the fully unsupervised scheme adopted in their work. The work presented in Lorenzen et al. (2004) is also along a similar line but it only address the problem of Bayesian diffeomorphic template construction given a set of pre-computed segmentations, without embedding image segmentation and atlas construction in a single mathematical model of the data.

The main aim of this paper is to demonstrate the methodological validity of the proposed approach, which, for this purpose, has been thoroughly tested on both synthetic and real MR neuroimaging data sets. However, the method is a general one and, since it was not tuned or optimised for a particular type of imaging data, we believe that it holds a potential for application in a wider range of medical image computing problems, some of which we aim to explore as part of our future work.

## Methods

Let us consider a population of  $M$  subjects belonging to a homogeneous group, from an anatomical point of view, and let us assume that  $D$  image volumes of different contrast are available for each subject.

From a generative perspective, the image intensities  $\mathbf{X} = \{\mathbf{X}_i\}_{i=1,\dots,M}$ , which constitute the observed data, can be thought of as being generated by sampling from  $D$ -dimensional Gaussian mixture probability distributions, after non-linear warping of a probabilistic anatomical atlas (Evans et al., 1994). The use of Gaussian mixture models to capture the probability density function of MR data is a well established approach (Ashburner and Friston, 2005; Greenspan et al., 2006; Zhang et al., 2001). For an extensive review of such methods see Balafar (2014).

The probabilistic atlas carries prior anatomical knowledge, in the form of average shaped tissue probability maps. From a mathematical modelling point of view, the atlas encodes local mixing proportions  $\Theta_\pi = \{\pi_j\}_{j=1,\dots,N_x}$  of the mixture model, with  $j$  being an index set over the  $N_x$

**Table 1**

List of mathematical symbols used in this paper.

Symbol	Meaning
$x_{ij}$	Observed image intensity at voxel $j$ for subject $i$ .
$\mathbf{z}_{ij}$	Vector of latent class membership probabilities.
$\pi_j$	Tissue priors at voxel $j$ .
$\mu_{ik}$	Mean intensity of class $k$ for subject $i$ .
$\Sigma_{ik}$	Covariance of intensities for class $k$ and subject $i$ .
$\mathbf{W}_{0k}$	Scale matrix of Wishart prior distribution on $\Lambda_k = (\Sigma_k)^{-1}$ .
$\nu_{0k}$	Degrees of freedom of Wishart prior distribution on $\Lambda_k$ .
$\mathbf{m}_{0k}$	Mean of Gaussian prior distribution over $\mu_k$ .
$\beta_{0k}$	Scaling hyperparameter of Gaussian prior distribution over $\mu_k$ .
$\alpha_0$	Hyperparameter governing the Dirichlet prior on $\pi$ .
$\Theta_\beta$	Bias field parameters.
$\mu_\beta$	Prior mean of bias parameters.
$\Sigma_\beta$	Prior covariance matrix of bias parameters.
$\Theta_a$	Affine transformation parameters.
$\mu_a$	Prior mean of affine transformation parameters.
$\Sigma_a$	Prior covariance matrix of affine transformation parameters.
$\mathbf{w}_i$	Weights for rescaling the tissue priors.
$\mathbf{u}_{ij}$	Initial velocity at voxel $j$ for subject $i$ .
$L_u$	Differential operator to compute penalty on $\mathbf{u}_i$ .

template voxels (for a list of all the mathematical symbols used in this section see Table 1). Each vector  $\pi_j$  has  $K$  elements, which indicate the prior probability of voxel  $j$  belonging to one of  $K$  tissue classes. These spatially varying mixing proportions can also be thought of as the coefficients used to parametrise  $K$  continuous scalar functions of space  $\{\pi_k\}_{k=1,\dots,K}$ .

### Tissue priors

Each image voxel,  $j \in \{1, \dots, N_x\}$ <sup>1</sup> for each subject  $i \in \{1, \dots, M\}$  is considered as being drawn from  $K$  possible tissue classes. The following prior latent variable model defines the probability of finding tissue type  $k$ , at a specific location  $j$  (i.e. centre of voxel  $j$ ), in image  $i$ , prior to observing the corresponding image intensity signal

$$p(z_{ijk} = 1 | \Theta_\pi, \Theta_w, \Theta_u) = \frac{w_{ik} \pi_k(\xi_i(\mathbf{y}_j))}{\sum_{c=1}^K w_{ic} \pi_c(\xi_i(\mathbf{y}_j))}, \quad (1)$$

or equivalently

$$p(\mathbf{z}_{ij} | \Theta_\pi, \Theta_w, \Theta_u) = \prod_{k=1}^K \left( \frac{w_{ik} \pi_k(\xi_i(\mathbf{y}_j))}{\sum_{c=1}^K w_{ic} \pi_c(\xi_i(\mathbf{y}_j))} \right)^{z_{ijk}}. \quad (2)$$

Class memberships, for each subject and each voxel, are encoded in the latent variable  $\mathbf{z}_{ij}$ , using a one-of- $K$  scheme (i.e.  $\mathbf{z}_{ij}$  is  $K$ -dimensional vector with all elements equal to zero except for one, which is equal to one). In particular,  $z_{ijk} \in \{0, 1\}$  is equal to one if voxel  $j$  of image  $i$  belongs to tissue class  $k$  and zero otherwise, which also explains the equivalence of (1) and (2).  $\{\pi_k\}_{k=1,\dots,K}$  are continuous scalar functions  $\pi_k : \Omega_\pi \rightarrow \mathbb{R}$ , defined on the template domain  $\Omega_\pi$ . Such functions are common across the entire population, which satisfy the constraint

$$\sum_{k=1}^K \pi_k(\mathbf{y}) = 1, \quad \forall \mathbf{y} \in \Omega_\pi \subset \mathbb{R}^3, \quad (3)$$

with  $\mathbf{y}$  being a continuous coordinate vector field. Global weights  $\Theta_w = \{\mathbf{w}_i\}_{i=1,\dots,M}$ , with  $\mathbf{w}_i \in \mathbb{R}^K$ , are introduced to further compensate for individual differences in tissue composition.

In equation (1),  $\xi_i$  denotes a generic spatial transformation,

<sup>1</sup> Our formulation tries to be as general as possible, therefore we do not assume that different images will have the same size, as this often does not occur in practice. The size of the template is calculated so as to cover the field of view of all the images in the data set.

parametrised by  $\Theta_{i_s}$ , which allows projecting prior anatomical information onto individual data, with  $\xi_i : \Omega_i \rightarrow \Omega_\pi$  being a continuous mapping from the domain  $\Omega_i \subset \mathbb{R}^3$  of image  $i$ , into the space of the tissue priors  $\Omega_\pi \subset \mathbb{R}^3$ . In this work we adopt a large deformation diffeomorphic model (Ashburner and Friston, 2011; Beg et al., 2005; Trounev, 1998), where, as explained in the following subsection, the diffeomorphisms are parametrised by means of an initial velocity field, denoted by  $\mathbf{u}$ .

Since digital image data is a discrete signal, defined on a tridimensional voxel grid, each mapping  $\xi_i$  needs to be discretised as well, via sampling at the centre of every voxel  $j \in \{1, \dots, N_i\}$ , to give the discrete mapping  $\{\xi_i(\mathbf{y}_j)\}_{j=1, \dots, N_i}$  that appears in (1).

As opposed to the modelling approach described in Ashburner and Friston (2005) and Blaiotta et al. (2016), where the tissue priors were considered as fixed and known *a priori* quantities, here the tissue probability maps are treated as random variables, whose point estimates or full posteriors can be inferred via model fitting (Bhatia et al., 2007; Ribbens et al., 2014).

For this purpose, a finite dimensional parametrisation of the continuous functions  $\{\pi_k\}_{k=1, \dots, K}$  needs to be defined. Typically, whenever a continuous function needs to be reconstructed from a finite discrete sequence, it is possible to formulate the problem as an interpolation that makes use of a finite set of coefficients and continuous basis functions. Since the priors  $\{\pi_k\}_{k=1, \dots, K}$  are bounded to take values in the interval  $[0, 1]$  on the entire domain  $\Omega_\pi$  (see equation (3)), not all basis functions are well suited here. For this reason we use first degree B-splines, which ensure that the tissue priors satisfy the above mentioned constraint, while being quite a computationally efficient choice compared to higher order interpolation methods. The coefficients used to parametrise the tissue priors belong to the discrete set  $\Theta_\pi = \{\pi_j\}_{j=1, \dots, N_\pi}$  of  $K$ -dimensional vectors, with

$$\sum_{k=1}^K \pi_{jk} = 1, \quad \forall j \in \{1, \dots, N_\pi\}. \quad (4)$$

and  $N_\pi$  being the number of template voxels. Such coefficients can be learned directly from the data, as it will be shown in the following section.

Additionally, prior distributions on the parameters  $\{\pi_j\}_{j=1, \dots, N}$  can be introduced (Bishop, 2006). Dirichlet priors are the most convenient choice here, since they are conjugate to multinomial forms of the type in (2), and they can be expressed as

$$p(\boldsymbol{\pi}_j) = \text{Dir}(\boldsymbol{\pi}_j | \boldsymbol{\alpha}_0) = C(\boldsymbol{\alpha}_0) \prod_{k=1}^K \pi_{jk}^{\alpha_k - 1}, \quad (5)$$

where the normalising constant is given by

$$C(\boldsymbol{\alpha}_0) = \frac{\Gamma(\bar{\alpha})}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)}, \quad (6)$$

with  $\Gamma(\cdot)$  being the gamma function and

$$\bar{\alpha} = \sum_{k=1}^K \alpha_k. \quad (7)$$

### Diffeomorphic image registration

As anticipated in the previous sections, the generative interpretation of imaging data that this work relies on involves warping an unknown, average-shaped atlas to match a series of individual scans.

Such a problem, that is to say template matching via non-rigid registration, has been largely explored in medical imaging, mainly for solving image segmentation or structural labeling problems, in an automated fashion (Ashburner and Friston, 2005; Shen and Davatzikos, 2004; Christensen, 1999; Chui et al., 2001; Bajcsy et al., 1983; Iglesias et al.,

2012; Pluta et al., 2009; Warfield et al., 1999; Khan et al., 2008; Bowden et al., 1998).

Indeed, the modelling of spatial mappings between different anatomies can be approached in a variety of manners, depending on the adopted model of shape and on the objective function (i.e. similarity metric and regularisation) that the optimisation is based on, thus leading to a variety of algorithms with remarkably different properties (Penney et al., 1998; Denton et al., 1999; Klein et al., 2009).

The work presented here is formulated according to the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework (Younes, 2010), where the transformations mapping between the source images and the target image are assumed to belong to a Riemannian manifold<sup>2</sup> of diffeomorphisms (Ashburner, 2007). A diffeomorphism  $\phi : \Omega \rightarrow \Omega$  is a smooth differentiable map (with a smooth differentiable inverse  $\phi^{-1}$ ) defined on a compact, simply connected domain  $\Omega \subset \mathbb{R}^3$ .

One way of constructing transformations belonging to the diffeomorphic group  $\text{Diff}(\Omega)$  is to solve the following non-stationary transport equation (Joshi and Miller, 2000)

$$\frac{d}{dt} \phi(\mathbf{y}, t) = \mathbf{u}(\phi(\mathbf{y}, t), t), \quad \phi(\mathbf{y}, 0) = \mathbf{y}, \quad t \in [0, 1], \quad (8)$$

where  $\mathbf{u}(\phi(\mathbf{y}, t), t) \in \mathcal{H}$  is a time dependent, smooth velocity vector field, in the Hilbert space<sup>3</sup>  $\mathcal{H}$ .

The initial map, at  $t = 0$ , is equal to the identity transform  $\phi(\mathbf{y}, 0) = \mathbf{y}$ , while the final map, endpoint of the flow of the velocity field  $\mathbf{u}$ , can be computed by integration on the unitary time interval  $t \in [0, 1]$  (Beg et al., 2005).

$$\phi(\mathbf{y}, 1) = \int_0^1 \mathbf{u}(\phi(\mathbf{y}, t), t) dt + \phi(\mathbf{y}, 0). \quad (12)$$

A procedure known as geodesic shooting (Miller et al., 2006; Ashburner and Friston, 2011; Allasonnière et al., 2005; Vialard et al., 2012; Beg and Khan, 2006) is applied, within the work presented here, to compute the final diffeomorphism given the initial map (i.e. the identity transform) and the velocity field at  $t = 0$ . Such a procedure exploits the principle of conservation of momentum (Younes et al., 2009), which is given by  $\mathbf{L}_u^\dagger \mathbf{L}_u \mathbf{u}$ , with  $\mathbf{L}_u^\dagger$  being the adjoint of the differential operator  $\mathbf{L}_u$ , to integrate the dynamical system governed by (8) without having to compute and store an entire time series of velocity fields. The implementation adopted here relies on the work presented in Ashburner and Friston (2011).

A diffeomorphic path  $\phi$  is not only differentiable, but also guaranteed to be a one-to-one mapping. Such a quality is highly desirable for finding morphological and functional correspondences between different anatomies without introducing tears or foldings, which would violate the conditions for topology preservation (Christensen, 1999). Additionally, the diffeomorphic framework provides metrics to quantitatively evaluate distances between anatomies or shapes. It should also be noted that diffeomorphisms are locally analogous to affine transformations (Avants

<sup>2</sup> A Riemannian manifold, in differential geometry, is a smooth manifold  $M$  equipped with a Riemannian metric (inner product). In particular, the Riemannian metric  $G_p$  on the  $n$ -dimensional manifold  $M^n$  defines, for every point  $p \in M$ , the scalar product of vectors in the tangent space  $T_p M$ , in such a way that given two vectors  $\mathbf{x}, \mathbf{y} \in M$ , the inner product  $G_p(\mathbf{x}, \mathbf{y})$  depends smoothly on the point  $p$ . The tangent space represents the nearest approximation of the manifold by a vector space (Warner, 2013).

<sup>3</sup> A Hilbert space  $\mathcal{H}$  is a complete inner product space, where an inner product is a map  $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ , which associates each pair of vectors in the space with a scalar quantity. In particular given  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{H}$  and  $a, b \in \mathbb{C}$

$$\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a\langle \mathbf{x}, \mathbf{z} \rangle + b\langle \mathbf{y}, \mathbf{z} \rangle, \quad (9)$$

$$\langle \mathbf{x}, \mathbf{x} \rangle \geq 0, \text{ and } \langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}, \quad (10)$$

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle. \quad (11)$$

et al., 2006).

In practice, finding an optimal diffeomorphic transformation to equation a pair, or a group, of images involves optimising an objective function (e.g. minimising a cost function), in the space  $\mathcal{H}$  of smooth velocity vector fields defined on the domain  $\Omega$ . The required smoothness is enforced by constructing the norm on the space  $\mathcal{H}$  through a differential operator  $\mathbf{L}_u$  (Beg et al., 2005), such that a quantitative measure of smoothness can be obtained via

$$\mathcal{R}(\mathbf{u}) = \|\mathbf{L}_u \mathbf{u}\|_{L^2}^2, \quad (13)$$

where  $\mathbf{u}$  is a discretised version of  $u$ .

The form of the cost function will depend on how the observed data is modelled. For the work presented here, groupwise equationment is achieved via maximisation of the following variational objective function

$$\begin{aligned} \mathcal{E}(\Theta_u) &= \mathbb{E}_{\mathbf{Z}}[\log p(\bar{\mathbf{Z}}|\Theta_u, \Theta_w, \Theta_u)] + \log p(\Theta_u) + \text{const} \\ &= \sum_{i=1}^M \sum_{j=1}^{N_i} \sum_{k=1}^K \gamma_{ijk} \log \left( \frac{w_{ik} \pi_k(\phi_i(\mathbf{y}_j))}{\sum_{c=1}^K w_{ic} \pi_c(\phi_i(\mathbf{y}_j))} \right) \\ &\quad - \frac{1}{2} \sum_{i=1}^M \|\mathbf{L}_u \mathbf{u}_i\|_{L^2}^2 + \text{const}, \end{aligned} \quad (14)$$

where  $\mathbf{Z} = \{\mathbf{Z}_i\}_{i=1, \dots, M}$  is the set of latent variables across the entire population,  $\{\gamma_{ij}\}_{i,j} = \{\mathbb{E}[\mathbf{z}_{ij}]\}_{i,j}$  are  $K$ -dimensional vectors encoding the posterior probabilities of each voxel belonging to the  $K$  tissue types. The coordinate mappings  $\{\phi_i\}_{i=1, \dots, M}$  are encoded in the parameter set  $\Theta_u$ , which consists of  $M$  vectors of coefficients  $\{\mathbf{u}_i\}_{i=1, \dots, M}$ , containing  $3 \times N_i$  elements each. Such coefficients can be used to construct continuous initial velocity fields via trilinear, or higher order, interpolation.

The posterior membership probabilities  $\{\gamma_{ij}\}_{i,j}$  that appear in (14) can be computed by combining the prior latent variable model introduced in 2.1 with a class conditional likelihood model of image intensities, which will be described in subsection. In such a case, learning posterior label probabilities can be addressed as a standard mixture distribution inference problem, which can be conveniently solved using the expectation-maximisation algorithm or its variational extensions (Bishop, 2006; Blaiotta et al., 2016), thus leading to a fully unsupervised learning scheme.

Alternatively, when manual labels are available, binary posterior class probabilities can be derived directly from such categorical annotations, without performing inference from the observed image intensity data. In particular, if all input data has been manually labelled, then the resulting algorithm would implement a fully supervised learning strategy. Instead, if only some of the data has associated training labels, a hybrid approach can be adopted, which would fall into the category of semisupervised learning (Chapelle et al., 2006; Filipovych et al., 2011).

Essentially these three approaches, namely unsupervised, semi-supervised and fully supervised, differ in the relative ratio between the number of labelled and unlabelled voxels. Even so, they can all be framed within an expectation-maximisation setting, where depending on whether a voxel is labelled or not, the E-step is performed differently. Specifically, if  $\mathbf{x}_{ij}$  is an unlabelled observation

$$\begin{aligned} \gamma_{ijk} &= p(z_{ijk} = 1 | \mathbf{x}_{ij}, \Theta) \\ &= \frac{p(\mathbf{x}_{ij} | z_{ijk} = 1, \Theta) p(z_{ijk} = 1 | \Theta)}{\sum_{c=1}^K p(\mathbf{x}_{ij} | z_{ijc} = 1, \Theta) p(z_{ijc} = 1 | \Theta)}, \end{aligned} \quad (15)$$

with  $\Theta$  being the current estimate of model parameters at each iteration of the algorithm. Instead, if  $\mathbf{x}_{ij}$  has been manually labelled

$$\gamma_{ijk} = p(z_{ijk} = 1 | l_{ij}) = \begin{cases} 1, & \text{if } l_{ij} = k \\ 0, & \text{if } l_{ij} \neq k \end{cases} \quad (16)$$

where  $l_{ij}$  is a categorical manual label assigned to voxel  $j$  in image  $i$ .

Finally, it is also possible to take into account the uncertainty inherent in the process of manual rating. In such a case, the actual posterior probabilities, for labelled observations, can be computed by making use of the categorical output of manual labelling together with an estimate of the rater sensitivity and with a generative intensity model.

Making use of Bayes rule, this gives

$$\begin{aligned} \gamma_{ijk} &= p(z_{ijk} = 1 | \mathbf{x}_{ij}, \Theta, l_{ij}) \\ &= \frac{p(\mathbf{x}_{ij} | z_{ijk} = 1, \Theta) p(z_{ijk} = 1 | \Theta) p(z_{ijk} = 1 | l_{ij})}{\sum_{c=1}^K p(\mathbf{x}_{ij} | z_{ijc} = 1, \Theta) p(z_{ijc} = 1 | \Theta) p(z_{ijc} = 1 | l_{ij})}, \end{aligned} \quad (17)$$

where  $p(z_{ijk} = 1 | l_{ij})$  indicates the probability of voxel  $j$  in image  $i$  belonging to class  $k$ , given the manual label attributed to the same voxel.

A simple model for this, is

$$p(z_{ijk} = 1 | l_{ij}) = \begin{cases} \zeta_l, & \text{if } l_{ij} = k \\ \frac{1 - \zeta_l}{K - 1}, & \text{if } l_{ij} \neq k \end{cases} \quad (18)$$

where  $\zeta_l$  is the sensitivity of the rater that generated the set of labels  $\{l_{ij}\}_{j=1, \dots, N}$  for image  $i$ . The problem of how to evaluate the performance of a manual or automated rater is not addressed here. For instance, a probabilistic scheme, which has been widely used to assess segmentation performance in medical imaging, is presented in Warfield et al. (2004).

In the remainder of this paper we will be focusing on the semi-supervised approach and we will consider the general case where only some voxels in few images have been manually labelled by adopting the model in (17) with a fixed sensitivity parameter  $\zeta_l$  equal to 0.9.

#### Combining diffeomorphic with affine registration

Anatomical shapes are very high dimensional objects. The diffeomorphic model described in the previous section, which is encoded using  $3 \times N_i$  free parameters, can account for a significant amount of shape variability in the observed data. Nevertheless, it is still convenient, mainly for computational reasons, to combine such a local, high dimensional shape model with global, lower dimensional transformations, such as rigid body or affine transforms. In fact, by beginning to solve the registration problem from the coarsest deformation components (e.g. rigid body or affine), it is possible to ensure that the subsequent diffeomorphic registration starts from a good initial estimate of image alignment, that is to say closer to the desired global optimum (Lester and Arridge, 1999). This makes the optimisation problem faster to solve and at the same time it reduces significantly the chance of registration failure (Modersitzki, 2004). It is relatively common for non-linear registration algorithms to perform poorly in the presence of a large translational or size mismatch between the reference and the target images (Jenkinson and Smith, 2001).

A possible parametrisation that combines affine and diffeomorphic transformations is

$$\xi_i(\mathbf{y}) = \mathbf{T}_i \phi_i(\mathbf{y}) + \mathbf{t}_i, \quad \forall \mathbf{y} \in \Omega_i, \quad (19)$$

where  $\xi_i(\mathbf{y})$  is the resulting mapping from image of subject  $i$  into the template space. Such a mapping is obtained by affine transforming the diffeomorphic deformation field  $\phi_i$ . The transformation matrix  $\mathbf{T}_i$  encodes nine degrees of freedom (rotation, zooming and shearing) and is computed via an exponential map  $\mathbf{T}_i = \exp(\mathbf{Q}_i(\mathbf{a}_i))$  with  $\mathbf{Q}_i(\mathbf{a}_i) \in \mathfrak{ga}(3)$ , where  $\mathfrak{ga}(3)$  is the Lie algebra for the affine group in three dimension  $GA(3)$  and  $\mathbf{a}_i$  is a vector of nine parameters (Ashburner and Ridgway, 2013). Translations are modelled by the vector  $\mathbf{t}_i \in \mathbb{R}^3$ . The entire set of affine parameters is denoted as  $\Theta_a = \{\mathbf{a}_i, \mathbf{t}_i\}_{i=1, \dots, M}$ .



### Intensity model

From a general probabilistic perspective, classification of tissue types based on MR signal intensities requires a model of the observed data that is capable of capturing the probability of occurrence of each signal sample value  $\mathbf{x}_{ij}$ , provided that the true labels are known. In other words, the problem breaks down into defining suitable conditional probabilities  $p(\mathbf{x}_{ij}|z_{ijk} = 1)$ , for each  $k = \{1, \dots, K\}$  and then applying Bayes rule to infer the posterior class probabilities.

In the model adopted here, image intensity distributions are represented as Gaussian mixtures, with the unknown mean  $\mu_{ik}$  and covariance matrix  $\Sigma_{ik}$  of each Gaussian component  $k$ , for subject  $i$ , being governed by Gaussian-Wishart priors (Bishop, 2006; Blaiotta et al., 2016).

Correction of intensity inhomogeneities is also performed within the same modelling framework and it involves multiplying the uncorrected intensities of each image volume by a bias field, which is modelled as the exponential of a weighted sum of discrete cosine transform basis functions (Styner et al., 2000; Ashburner and Friston, 2005). Such an approach is conceptually equivalent to scaling the probability distributions of all Gaussian components by a local scale parameter, which is the bias itself, such that

$$p(\mathbf{x}_{ij}|z_{ijk}, \mu_{ik}, \Sigma_{ik}, \Theta_\beta) = \mathcal{N}(\mathbf{x}_{ij}|\hat{\mu}_{ik}, \hat{\Sigma}_{ik}), \quad (20)$$

with

$$\begin{aligned} \hat{\mu}_{ik} &= (\text{diag}(\mathbf{b}_{ij}))^{-1} \mu_{ik}, \\ \hat{\Sigma}_{ik} &= (\text{diag}(\mathbf{b}_{ij}))^{-1} \Sigma_{ik} (\text{diag}(\mathbf{b}_{ij}))^{-1}, \end{aligned} \quad (21)$$

where  $\mathcal{N}$  indicates a Normal distribution,  $\Theta_\beta$  denotes the set of bias field parameters and  $\mathbf{b}_{ij}$  is a  $D$ -dimensional vector representing the bias at voxel  $j$  of subject  $i$  for each of the  $D$  imaging modalities. The result in (20) is independent from the particular parametrisation of  $\mathbf{b}_{ij}$ , as long as the bias is multiplicative, and it can be easily proven by applying the change of variable  $\mathbf{x}_{ij} = \mathbf{c}_{ij}/\mathbf{b}_{ij}$ , with  $\mathbf{c}_{ij}$  being the corrected image intensity

$$\begin{aligned} p(\mathbf{x}_{ij}|z_{ijk}, \mu_{ik}, \Sigma_{ik}, \Theta_\beta) &= \\ \det(\text{diag}(\mathbf{b}_{ij})) p(\mathbf{c}_{ij}|z_{ijk}, \mu_{ik}, \Sigma_{ik}, \Theta_\beta) &= \\ \det(\text{diag}(\mathbf{b}_{ij})) \mathcal{N}(\mathbf{c}_{ij}|\mu_{ik}, \Sigma_{ik}) &= \mathcal{N}(\mathbf{x}_{ij}|\hat{\mu}_{ik}, \hat{\Sigma}_{ik}). \end{aligned} \quad (22)$$

### Graphical model

A graphical representation of the model adopted in this paper is depicted in Fig. 1, while a legend of the symbols used to indicate the different variables can be found in Table 1.

Given such a model, it is possible to define the following variational objective function  $\mathcal{L}$ , which constitutes a lower bound on the logarithm of the marginal joint probability  $p(\bar{\mathbf{X}}, \Theta_\beta, \Theta_\alpha, \Theta_u, \Theta_\pi | \Theta_w)$ , such that

$$\log p(\bar{\mathbf{X}}, \Theta_\beta, \Theta_\alpha, \Theta_u, \Theta_\pi | \Theta_w) \geq \mathcal{L} \quad (23)$$

and

$$\begin{aligned} \mathcal{L} &= \mathbb{E}_{\mathbf{Z}, \Theta_\mu, \Theta_\Sigma} [\log p(\mathbf{X}|\mathbf{Z}, \Theta_\mu, \Theta_\Sigma, \Theta_\beta)] + \mathbb{E}_{\mathbf{Z}} [\log p(\mathbf{Z}|\Theta_\pi, \Theta_w, \Theta_u, \Theta_\alpha)] \\ &+ \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\log p(\Theta_\mu, \Theta_\Sigma)] + \log p(\Theta_\pi) + \log p(\Theta_\beta) + \log p(\Theta_\alpha) \\ &+ \log p(\Theta_u) - \mathbb{E}_{\mathbf{Z}} [\log q(\mathbf{Z})] - \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\log q(\Theta_\mu, \Theta_\Sigma)], \end{aligned} \quad (24)$$

where the expectations indicated as  $\mathbb{E}_{\mathbf{Z}}$  and  $\mathbb{E}_{\Theta_\mu, \Theta_\Sigma}$  are computed with respect to variational posterior distributions (Blaiotta et al., 2016; Bishop, 2006), denoted by  $q(\cdot)$ , on the latent variables  $\mathbf{Z}$  and on the Gaussian means and covariances  $\{\Theta_\mu, \Theta_\Sigma\}$ , respectively. Optimisation of  $\mathcal{L}$ , which provides optimal parameter and hyperparameter estimates, will be discussed in the following section.

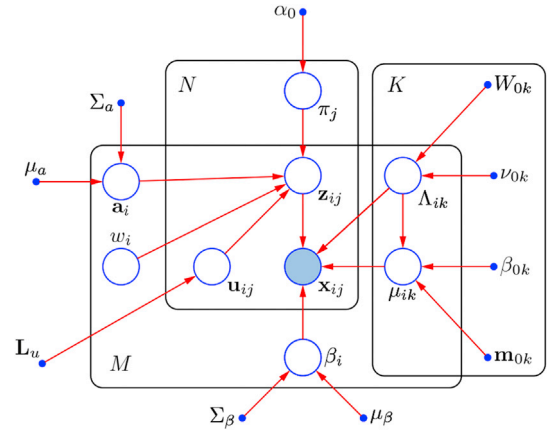


Fig. 1. Graphical representation of the model adopted in this paper. Observed variables  $\{\mathbf{x}_{ij}\}$  are represented by a filled circle. Latent variables  $\{z_{ij}\}$  as well as model parameters are depicted as unfilled circles. Blue solid dots correspond to fixed hyperparameters. The so called plate notation is adopted to indicated repeated variables. Symbols referring to all variables and parameters are listed in Table 1.

### Model fitting

The model described in the previous section can be fit to data sets of MR images by combining a variational expectation-maximisation (VBEM) algorithm with gradient based numerical optimisation techniques.

Indeed, the VBEM algorithm described in Blaiotta et al. (2016) is well-suited for addressing the model estimation problem discussed here, since it allows learning posterior distributions on the Gaussian mixture parameters, under the assumption that  $q(\bar{\mathbf{Z}}, \Theta_\mu, \Theta_\Sigma)$  factorises as  $q(\bar{\mathbf{Z}})q(\Theta_\mu, \Theta_\Sigma)$  (Bishop, 2006), and at the same time it is able to transfer the information encoded in such posteriors by estimating empirical intensity priors for each tissue type. As shown in our previous work (Blaiotta et al., 2016), this approach has several advantages over maximum likelihood estimation, including lower vulnerability to overfitting, faster convergence and higher robustness against misregistration, which inevitably occurs in the early iterations of any image registration algorithm.

Additionally, the algorithm proposed in this paper loops over all subjects in the population and, for each subject, it iterates over estimating the Gaussian-Wishart posteriors, the bias field parameters, the affine parameters and the initial velocities, which are all treated as conditional optimisations. Subsequently the tissue probability maps and the intensity priors are updated and the whole cycle is repeated until convergence.

Estimation of the bias field parameters  $\Theta_\beta$  can be conveniently performed via non-linear optimisation techniques. Here the problem is solved using the Gauss-Newton method (Bertsekas, 1999) with a backtracking line search, so as to maximise the objective function in (24) with respect to  $\Theta_\beta$ . The resulting implementation is very similar to the one described in Ashburner and Friston (2005), therefore further details are omitted here. Optimisation of the affine parameters  $\Theta_\alpha = \{\mathbf{a}_i, \mathbf{t}_i\}_{i=1, \dots, M}$  can also be carried out by means of a Gauss-Newton scheme and a brief description of the required computations can be found Appendix A. For the update of the weight parameters  $\Theta_w$  we adopt the same strategy outlined in Ashburner and Friston (2005); Blaiotta et al. (2016).

The following sections instead will present in detail the algorithmic scheme used to learn the average shaped tissue templates  $\Theta_\pi = \{\pi_j\}_{j=1, \dots, N_t}$  and to estimate the set of initial velocity fields  $\Theta_u = \{\mathbf{u}_i\}_{i=1, \dots, M}$ .

The pseudocode reported in alg:opt illustrates the different optimisation stages of the proposed algorithm. Essentially, the purpose of each

stage within the inner loop is to optimise a subset of parameters for a particular subject, while keeping all the other parameters fixed at their current estimates. This strategy leverages the conditional dependencies among different variables in the model (such conditional dependencies are illustrated in Fig. 1), with the aim of learning more accurate parameters compared to those estimated by solving the different optimisation problems independently (Mahapatra and Sun, 2012; Ashburner and Friston, 2005; Meng and Rubin, 1993). The outer loop serves to regenerate group priors both in the form of tissue probability maps and parametric intensity priors.

**Algorithm 1.** Optimisation algorithm for generating average-shaped tissue probability maps using the approach presented in this paper.

```

Input: A cross-sectional data set of MR images  $\bar{\mathbf{X}}$ 
1 begin
2   Set regularisation hyperparameters:
    $\{\mathbf{L}_u, \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\}$ ;
3   Initialise all model parameters:
    $\{\Theta_\pi, \Theta_w, \Theta_u, \Theta_a, \Theta_\mu, \Theta_\Sigma, \Theta_\beta\}$ ;
4   while objective function  $\mathcal{L}$  has not converged
   do
5     for iter = 1...,  $I_n$  do
6       for each subject  $i$  do
7         VE-step:
8         compute  $q(\{\mathbf{h}_j\}_{j=1}^N)$  (C.5);
9         compute  $q(\mathbf{Z}_i)$  (15),(16);
10        evaluate sufficient statistics
11        (C.8), (C.9);
12        VM-step:
13        compute  $q(\boldsymbol{\mu}_i | \boldsymbol{\Sigma}_i^{-1})$ ;
14        compute  $q(\boldsymbol{\Sigma}_i^{-1})$ ;
15        (Blaiotta et al., 2016);
16        M-step:
17        update  $\mathbf{w}_i$ ;
18        (Ashburner and Friston, 2005);
19        for iter $_\beta$  = 1...,  $I_\beta$  do
20          update  $\boldsymbol{\beta}_i$ 
21          (Ashburner and Friston, 2005);
22        end
23        for iter $_a$  = 1...,  $I_a$  do
24          update  $\mathbf{a}_i$  (Appendix A);
25        end
26        for iter $_u$  = 1...,  $I_u$  do
27          update  $\mathbf{u}_i$  (Appendix B);
28        end
29        end
30        update  $\Theta_\pi$  (37);
31        update  $\{\beta_{0k}, \boldsymbol{\mu}_{0k}, \nu_{0k}, \mathbf{W}_{0k}\}$ 
32        (Blaiotta et al., 2016);
33      end
34    end
35  end

```

#### Updating the tissue priors

At each outer iteration of the algorithm, the tissue priors  $\Theta_\pi = \{\pi_j\}_{j=1, \dots, N_\pi}$  need to be updated, given the current estimates of all the other parameters, which are kept fixed for each individual in the population.

Considering only the terms in (24) that depend on  $\Theta_\pi$  gives the following objective function, which has to be maximised with respect to  $\Theta_\pi$

$$\begin{aligned} \mathcal{L}_\pi &= \mathbb{E}_{\mathbf{Z}}[\log p(\mathbf{Z} | \Theta_\pi, \Theta_w, \Theta_u, \Theta_a)] + \log p(\Theta_\pi) + \text{const} \\ &= \sum_{i=1}^M \sum_{k=1}^K \int_{\Omega_i} \gamma_{ik}(\mathbf{y}) \log \left( \frac{w_{ik} \pi_k(\boldsymbol{\xi}_i(\mathbf{y}))}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_i(\mathbf{y}))} \right) d\mathbf{y} \\ &\quad + \log p(\Theta_\pi) + \text{const}. \end{aligned} \quad (25)$$

It should be noted that the parameters  $\Theta_\pi$  that need to be estimated are defined on the domain of the template  $\Omega_\pi$ , rather than on the individual spaces  $\{\Omega_i\}_{i=1, \dots, M}$ . For this reason equation (25), which is a sum of integrals on the native domains, needs to be mapped to  $\Omega_\pi$ , by inverting the warps  $\{\boldsymbol{\xi}_i\}_{i=1, \dots, M}$ , to give

$$\begin{aligned} \mathcal{L}'_\pi &= \sum_{i=1}^M \sum_{k=1}^K \int_{\Omega_\pi} \det \left( \frac{\partial \boldsymbol{\xi}_i^{-1}}{\partial \mathbf{y}} \right) \gamma_{ik}(\boldsymbol{\xi}_i^{-1}(\mathbf{y})) \log \left( \frac{w_{ik} \pi_k(\mathbf{y})}{\sum_{c=1}^K w_{ic} \pi_c(\mathbf{y})} \right) d\mathbf{y} \\ &\quad + \log p(\Theta_\pi) + \text{const}, \end{aligned} \quad (26)$$

where  $\boldsymbol{\xi}_i^{-1}$  is the transformation mapping from image  $i$  into the template space and the determinants of the Jacobian matrices of the deformations are included to preserve volumes after the change of variables.

Finally equation (26) is discretised on a regular voxel grid, whose centres have coordinates  $\{\mathbf{y}_j\}_{j=1, \dots, N_\pi}$ , to give

$$\begin{aligned} \mathcal{L}'_\pi &= \sum_{i=1}^M \sum_{j=1}^{N_\pi} \sum_{k=1}^K \det(\mathbf{J}_{ij}^{\boldsymbol{\xi}_i^{-1}}) \gamma_{ik}(\boldsymbol{\xi}_{ij}^{\boldsymbol{\xi}_i^{-1}}) \log \left( \frac{w_{ik} \pi_{jk}}{\sum_{c=1}^K w_{ic} \pi_{jc}} \right) \\ &\quad + \log p(\Theta_\pi) + \text{const}, \end{aligned} \quad (27)$$

where

$$\boldsymbol{\xi}_{ij}^{\boldsymbol{\xi}_i^{-1}} = \boldsymbol{\xi}_i^{-1}(\mathbf{y})|_{\mathbf{y}=\mathbf{y}_j}, \quad (28)$$

$$\det(\mathbf{J}_{ij}^{\boldsymbol{\xi}_i^{-1}}) = \det \left( \frac{\partial \boldsymbol{\xi}_i^{-1}(\mathbf{y})}{\partial \mathbf{y}} \right) |_{\mathbf{y}=\mathbf{y}_j}, \quad (29)$$

$$\pi_{jk} = \pi_k(\mathbf{y})|_{\mathbf{y}=\mathbf{y}_j} \quad (30)$$

The prior term  $p(\Theta_\pi)$  is given by the following Dirichlet distribution

$$p(\Theta_\pi) = \prod_{j=1}^{N_\pi} \text{Dir}(\pi_j | \boldsymbol{\alpha}_0) = C(\boldsymbol{\alpha}_0) \prod_{j=1}^{N_\pi} \prod_{k=1}^K \pi_{jk}^{\alpha_{0k}-1}. \quad (31)$$

Maximising equation (27) is a constrained optimisation problem, subject to

$$\sum_{k=1}^K \pi_{jk} = 1, \quad \forall j \in \{1, \dots, N_\pi\} \quad (32)$$

A closed form solution could be easily found if the rescaling weights  $\{\mathbf{w}_i\}_{i=1, \dots, M}$  were all equal to one. In such a case

$$\begin{aligned} \mathcal{L}'_\pi &= \sum_{i=1}^M \sum_{j=1}^{N_\pi} \sum_{k=1}^K \det(\mathbf{J}_{ij}) \gamma_{ik}(\boldsymbol{\xi}_{ij}^{\boldsymbol{\xi}_i^{-1}}) \log(\pi_{jk}) \\ &\quad + \sum_{j=1}^{N_\pi} \sum_{k=1}^K (\alpha_{0k} - 1) \log p(\pi_{jk}) + \text{const}, \end{aligned} \quad (33)$$

which could be maximised under the constraint (32), by making use of Lagrange multipliers (Falk, 1967), to give

$$\pi_{jk} = \frac{N_{jk} + \alpha_{0k} - 1}{\sum_{k=1}^K (N_{jk} + \alpha_{0k}) - K}, \quad (34)$$

with  $N_{jk} = \sum_{i=1}^M \det(\mathbf{J}_{ij}) \gamma_{ik}(\boldsymbol{\xi}_{ij}^{\boldsymbol{\xi}_i^{-1}})$ .

This solution would provide maximum a posteriori point estimates of  $\Theta_\pi = \{\pi_j\}_{j=1, \dots, N_\pi}$ . However for this problem, it would also be possible to

derive a full variational posterior distribution, which, like its prior, would take a Dirichlet form, with parameters  $\alpha_j = \alpha_0 + \mathbf{N}_j$ .

When rescaling of the tissue priors is allowed the optimisation problem becomes more complex. The strategy adopted here consists in finding an approximate solution to the unconstrained optimisation problem by setting the derivatives of the objective function in (26) to zero

$$\frac{\alpha_{0k} - 1}{\pi_{jk}} + \sum_{i=1}^M \det(\mathbf{J}_{ij}^{\xi^{-1}}) \gamma_{ik}(\xi_{ij}^{-1}) \left( \frac{1}{\pi_{jk}} - \frac{w_{ik}}{\sum_{c=1}^K w_{ic} \pi_{jc}} \right) = 0. \quad (35)$$

Solving with respect to  $\pi_{jk}$ , under the simplifying assumption that the term  $\sum_{c=1}^K w_{ic} \pi_{jc}$  can be treated as a constant, gives

$$\bar{\pi}_{jk} = \frac{N_{jk} + \alpha_{0k} - 1}{\sum_{i=1}^M \frac{\det(\mathbf{J}_{ij}^{\xi^{-1}}) \gamma_{ik}(\phi_{ij}^{-1}) w_{ik}}{\sum_{c=1}^K w_{ic} \pi_{jc}}}. \quad (36)$$

Such a solution is then projected onto the constraining hyperplane, by preserving tissue proportions at each voxel

$$\pi_{jk} = \frac{\bar{\pi}_{jk}}{\sum_{c=1}^K \bar{\pi}_{jc}}. \quad (37)$$

Experimental testing of this strategy indicated that it gave a constant improvement of the objective function at a relatively cheap computational cost. Alternatively, iterative constrained non-linear optimisation techniques (Powell, 1978) could have been exploited to solve the template update problem.

#### Computing the deformation fields

Groupwise image alignment is achieved by optimisation of the variational objective function defined in (24), with respect to the parameters used to compute the deformations. This is equivalent to adopting the following image matching or similarity term

$$\begin{aligned} \mathcal{S} &= \mathbb{E}_{\mathbf{Z}}[\log p(\mathbf{Z} | \Theta_{\pi}, \Theta_w, \Theta_u, \Theta_a)] \\ &= \sum_{i=1}^M \int_{\mathbf{y} \in \Omega_i} \sum_{k=1}^K \gamma_{ik}(\mathbf{y}) \log \left( \frac{w_{ik} \pi_k(\xi_i(\mathbf{y}))}{\sum_{c=1}^K w_{ic} \pi_c(\xi_i(\mathbf{y}))} \right) d\mathbf{y}. \end{aligned} \quad (38)$$

Additionally, working on discretised image grids, with associated voxel centres  $\{\mathbf{y}_{ij}\}_{j=1, \dots, N_i}$ , requires reformulating  $\mathcal{S}$  as

$$\mathcal{S} = \sum_{i=1}^M \sum_{j=1}^{N_i} \sum_{k=1}^K \gamma_{ik} \log \frac{w_{ik} \pi'_{jk}}{\sum_{c=1}^K w_{ic} \pi'_{jc}}, \quad (39)$$

with

$$\pi'_{jk} = \pi_k(\xi_i(\mathbf{y}))|_{\mathbf{y}=\mathbf{y}_{ij}}. \quad (40)$$

The penalty term for this groupwise image registration problem is given by

$$\begin{aligned} \mathcal{R} &= \mathcal{R}_{dif} + \mathcal{R}_{af} = \log p(\Theta_u) + \log p(\Theta_a) \\ &= -\frac{1}{2} \sum_{i=1}^M (\|\mathbf{L}_u \mathbf{u}_i\|_{L^2}^2 + \mathbf{a}_i^T \Sigma_a^{-1} \mathbf{a}_i) + \text{const}, \end{aligned} \quad (41)$$

with  $\mathbf{u}_i$  being a  $3 \times N_i$  dimensional vector of parameters used for representing the initial velocity field of image  $i$  and  $\mathbf{a}_i$  encoding affine deformation parameters used to compute the transformation in (19).

For each image  $i$  in the data set, updating the corresponding initial velocity field, given the current estimates of the templates and all the other model parameters, involves optimising the following objective function

$$\begin{aligned} \mathcal{E}_{dif}^{(i)} &= \mathcal{S}^{(i)} + \mathcal{R}_{dif}^{(i)} \\ &= \sum_{j=1}^{N_i} \sum_{k=1}^K \gamma_{ijk} \log \frac{w_{ik} \pi_k(\xi_{ij})}{\sum_{c=1}^K w_{ic} \pi_c(\xi_{ij})} - \frac{1}{2} \|\mathbf{L}_u \mathbf{u}_i\|_{L^2}^2, \end{aligned} \quad (42)$$

with respect to  $\mathbf{u}_i$ , under the following deformation model

$$\xi_{ij} = \xi_i(\mathbf{y}_{ij}) = \mathbf{T}_i \phi_i(\mathbf{y}_{ij}) + \mathbf{t}_i, \quad (43)$$

where  $\phi_i$  is a diffeomorphism computed via geodesic shooting (Ashburner and Friston, 2011) from the corresponding initial velocity field  $\mathbf{u}_i$ .

Here image registration is solved via Gauss-Newton optimisation, with a backtracking line search to ensure convergence. This requires computing both the first and second derivatives of the objective function (Hernandez and Olmos, 2008). Such derivatives can be found in Appendix B. This leads to a very high dimensional inverse problem, which unfortunately cannot be solved via numerical matrix inversion, since this would be prohibitively expensive from a computational point of view. The approach adopted in this work consists in treating this optimisation as a partial differential equation problem, which can efficiently be solved using multigrid methods (Modersitzki, 2004). In particular, we adopt the same full multigrid implementation as in Ashburner (2007).

#### Validation and discussion

In this section we present results obtained by applying the presented modelling framework to real MR scans acquired with different imaging systems and protocols, as well as to synthetic MR volumes. Both qualitative and quantitative measures will be provided to assess the behaviour of the proposed approach.

All validation analyses have been performed on the full model presented in this paper, since the validity of the individual components (i.e. bias correction, diffeomorphic registration, variational Gaussian mixture modelling) was already assessed in our previous work (Ashburner and Friston, 2005, 2011; Blaiotta et al., 2016), primarily at a single subject level. Additionally, numerous studies have demonstrated that solving image segmentation and registration tasks in a coupled and iterative manner, as opposed to sequential approaches, ensures more accurate results while reducing the chances of getting prematurely trapped in a local optimum (Ashburner and Friston, 2005; Pohl et al., 2006; Yezzi et al., 2001; Bhatia et al., 2007). This results from the conditional dependencies between the parameters controlling the deformation fields and the intensity distribution parameters, which are used to infer tissue labels. In other words, closely matching the template to each scan is a necessary condition in order to compute accurate segmentations while, in turn, accurate segmentations of the data help to improve alignment with the templates.

#### Template construction

We begin validating our approach by illustrating a set of templates obtained by fitting the proposed model to a large neuroimaging data set. Indeed, in remainder of this paper we will be using neuroimaging data for both training and testing, even if, given the generality of the presented approach, applications to many more types of data could be explored.

#### Data

The input data for training the model was obtained from three different databases, two of which are freely accessible for download, thus ensuring that the results presented here could readily be compared to those produced by competing algorithms for medical image registration or segmentation.

**OASIS data set.** The first data set consists of thirty five T1-weighted MR scans from the OASIS (Open Access Series of Imaging Studies)

database (Marcus et al., 2007). The data is freely available from the web site <http://www.oasis-brains.org>, where details on the population demographics and acquisition protocols are also reported. Additionally, the selected thirty five subjects are the same ones that were used within the 2012 MICCAI Multi-Atlas Labeling Challenge (Landman and Warfield, 2012).

**Balgrist data set.** The second data set consists of brain and cervical cord scans of twenty healthy adults, acquired at University Hospital Balgrist with a 3T scanner (Siemens Magnetom Verio). Magnetisation-prepared rapid acquisition gradient echo (MPRAGE) sequences, at 1 mm isotropic resolution, were used to obtain T1-weighted data, while PD-weighted images of the same subjects were acquired with a multi-echo 3D fast low-angle shot (FLASH) sequence, within a whole-brain multi-parameter mapping protocol (Weiskopf et al., 2013; Helms et al., 2008).

**IXI data set.** The third and last data set comprises twenty five T1-, T2- and PD-weighted scans of healthy adults from the freely available IXI brain database, which were acquired at Guy's Hospital, in London, on a 1.5T system (Philips Medical Systems Gyroscan Intera). Additional information regarding the demographics of the population, as well as the acquisition protocols, can be found at <http://brain-development.org/ixi-dataset>.

The complete data set therefore consists of eighty multispectral scans of healthy adults, obtained with fairly diverse acquisition protocols and using scanning systems produced by different vendors.

Unfortunately, not all the three modalities of interest (T1-, T2- and PD-weighted) are available for all of the subjects. To circumvent the difficulties arising from the presence of missing imaging modalities, without neglecting any of the available data (indeed deletion of entries with missing data is still, in spite of its crudity, a common statistical practice), the Gaussian mixture modelling approach discussed in Blaiotta et al. (2016) was generalised by introducing an additional variational posterior distribution over the missing data points.

In practice, the resulting variational EM scheme iterates over first estimating an approximated posterior distribution on the unknown image intensities, secondly updating the sufficient statistics of the complete (observed and missing) data and finally computing variational posteriors on the Gaussian mixture parameters. Additional computational details relative to this strategy are provided in Appendix C.

In synthesis, it was possible to fit the generative groupwise model described in this paper to the entire data set, in spite of having different imaging modalities available from the different acquisition sites. This is indeed a very common scenario in real life medical imaging problems, therefore it should be actively addressed by processing or modelling solutions that claim to be applicable to large population data (van Tulder and de Bruijne, 2015).

Manual brain labels are freely available for the selected subset of the OASIS data set. Such labels have been generated and made public by Neuromorphometrics, Inc. (<http://Neuromorphometrics.com>) under academic subscription and they provide a fine parcellation of cortical and non cortical structures, for a total of 139 labels across the brain.

Part of this label data was used for training of the model while the remainder was left out for testing and validation. In particular, brain labels of twenty out of the thirty five OASIS subjects were used to create gray and white matter ground truth segmentations, which were provided as training input for semisupervised model fitting. In spite of having defined only one gray matter training label, two distinct gray matter classes were introduced in the mixture model (top two rows in Fig. 2), to best capture the corresponding distribution of image intensities, which is poorly represented by a single Gaussian component, as opposed to the distribution of white matter intensities. In this case, posterior membership probabilities were computed making use of equation (17).

Spinal cord manual labels were created for forty subjects (twenty from Balgrist data set and twenty from the IXI data set), by manually delineating the contour of the cord in each transverse slice of the data beginning from the lower extremity of the medulla oblongata. Such labels were randomly split in half for training and half for subsequent test

analyses. Due to the limited resolution of the data it was not possible to manually delineate gray and white matter within the spinal cord. For this reason, each voxel classified as spinal cord in the training data was allowed to be assigned either to the gray or to the white matter tissue classes, based on the fit of its intensity value to the underlying Gaussian mixture model, as outlined in equation (17).

#### Tissue templates and intensity priors

The tissue probability maps obtained by applying the modelling framework presented in this paper to the data set described above are depicted in Fig. 2. The total number of tissue classes used for this experiment is equal to twelve but three classes, representing air in the background, are not shown.

The total number of classes was selected based on empirical evidence to obtain a reasonable trade-off between goodness of fit and computational cost. In principle the proposed algorithm would be able to automatically determine the optimal number of classes, as demonstrated in Blaiotta et al. (2016), however this would require setting an initial number of components higher than the unknown optimal one, which, for the size of the data set considered here, would have been computationally very expensive.

Apart from the number of Gaussians, the only hyperparameters that are not estimated from the data but need to be fixed *a priori* are those controlling the bias and registration regularisation. Indeed, one of the strengths of the proposed approach is that, in spite of its complexity most parameters are automatically inferred from the data, thus requiring minimal parameter tuning. For the experiments described in this section we used the settings described below.

The prior distribution on the bias field parameters was assumed to be Gaussian with zero mean and a precision matrix corresponding to a penalty on the Laplacian of the resulting non-uniformity field. Similarly, the affine model was regularised using a Gaussian prior with zero mean and a diagonal covariance matrix, where the magnitude of parameters modelling translations and rotations was less heavily regularised (the prior variance was set equal to 1), compared to that of parameters controlling scaling and shearing, whose prior variance was instead set equal to 0.001. Finally the regularisation settings for diffeomorphic registration were borrowed from the default settings of the Geodesic Shooting toolbox (Ashburner and Friston, 2011) provided with SPM12, which correspond to penalising a linear combination of absolute displacement, membrane energy, bending energy and linear elasticity (more details can be found in the documentation of the SPM12 software).

Fig. 2 shows how one of the two gray matter classes (first row) best fits the subcortical nuclei and also includes voxels affected by partial volume effects at the interface between gray and white matter, while the second one (second row) is more representative of cortical structures, with partial volume effects generated by the mixed presence of gray matter and CSF. The third row in Fig. 2 shows the white matter class, which also includes most of the brainstem and the spinal cord.

The remaining tissue classes were estimated in a purely unsupervised way. Therefore a non ambiguous anatomical interpretation is not straightforward.

Tissue class four (fourth row) mainly contains CSF, even if other tissues are also present, especially in the neck area. This should be attributed to the lack of CSF training labels as well as to a poor multivariate coverage of the cervical region in the available data. In fact, data from the OASIS set is truncated around the first cervical vertebra. The T1-weighted scans of the IXI data set cover up to the C2/C3 vertebral level, but the corresponding T2- and PD-weighted scans do not extend beyond the brainstem. Indeed, only the data from the second database (Balgrist hospital) provides more than one modality covering up to around the fourth cervical vertebra. In this case though, additional difficulties arose from poor inter-modality alignment of the data, a problem that turned out to be particularly severe in the cervical region and that, given its non-linearity, cannot be fully compensated for by affine inter-modality coregistration. This result confirms the importance and



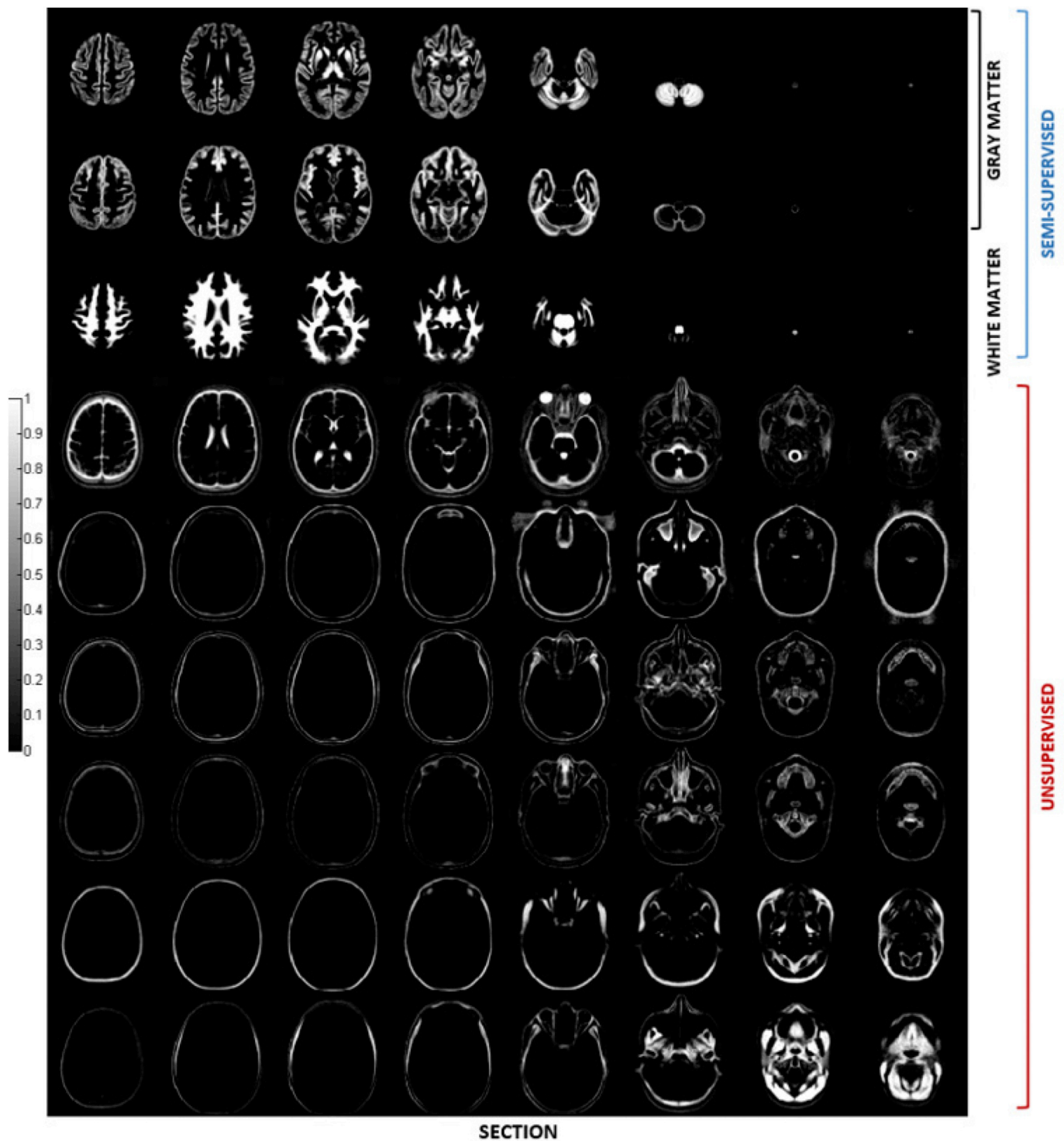


Fig. 2. Tissue probability maps obtained by applying the presented groupwise generative model to a multispectral data set comprising head and neck scans of eighty healthy adults, from three different databases.

usefulness of non-linear image coregistration tools, particularly when modelling highly deformable anatomical structures (Stroman et al., 2008; Fonov et al., 2014). Such a problem was not addressed in this paper, which limits the applicability of the proposed approach when scans of the same subject are not in good alignment. However it would in principle be possible to introduce inter-modality deformation fields within the same generative model adopted here and we aim to address this issue as part of our future work.

Bone tissue is also not easily identifiable from the data available for

this experiment, but it could have potentially been much better extracted by incorporating some CT scans into the training data.

Fat and soft tissues are mainly represented in the last two classes (bottom two rows in Fig. 2).

While quantitative evaluation analyses of the proposed modelling approach will be reported in the following sections, the results presented in Fig. 2 provide a qualitative insight into the performance of the algorithm discussed in this paper. In particular, the sharp appearance of the tissue probability maps suggests that the proposed model can capture a

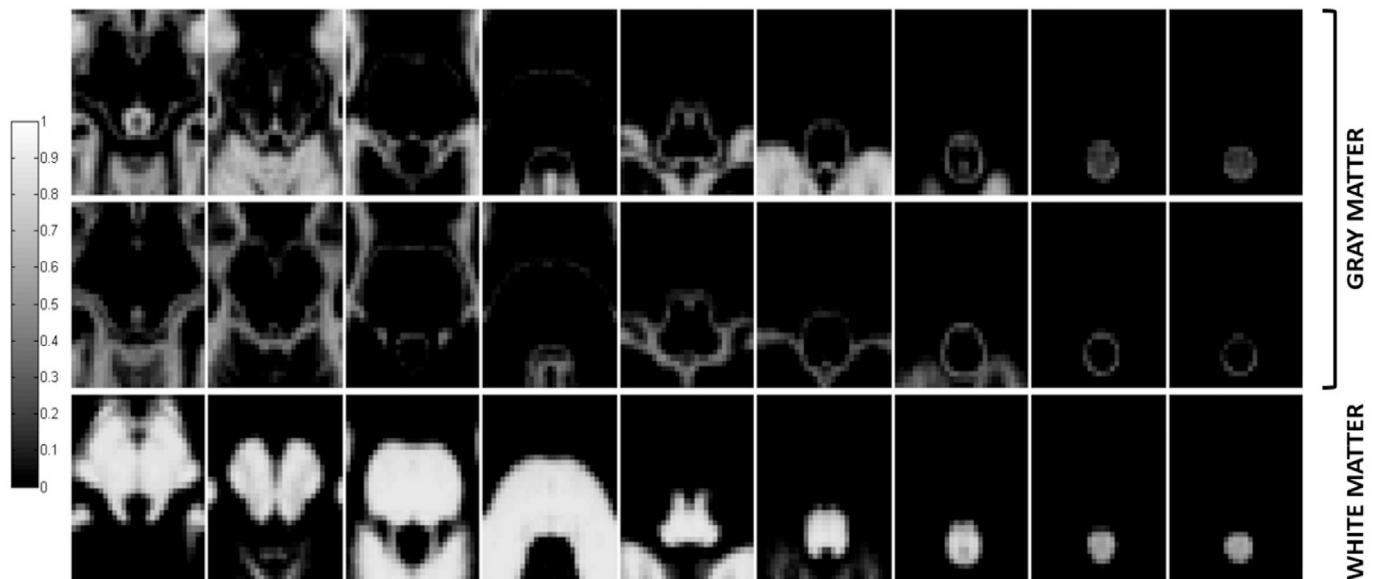


Fig. 3. Zoomed views of the gray (top two rows) and white matter (bottom row) tissue probability maps at the brainstem and spinal cord levels. Each transverse slice covers an area of  $45 \times 53 \text{ mm}^2$ , while the axial distance between adjacent slices along each row equals 7.5 mm. The order of tissue classes from top to bottom is the same as in the top three rows of Fig. 2, however more densely spaced transverse slices are depicted here.

significant amount of anatomical variability thus ensuring fine alignment of complex anatomical structures, whose shape varies significantly across individuals. In turn, this represents a very valuable property for the purpose of performing statistical group analyses on neuroimaging data, since most of these studies rely on the assumption that accurate anatomical and functional correspondences can be estimated through some form shape mapping procedure, which is most commonly implemented using non-linear image registration techniques.

Fig. 3 illustrates zoomed views of the gray and white matter tissue probability maps at the brainstem and spinal cord levels.

The empirical Bayes learning procedure, introduced in Blaiotta et al. (2016) to estimate suitable prior distributions for the parameters of the Gaussian mixture model, was applied here to the same data used to construct the templates. Some of the results are summarised in Fig. 4, where the estimated empirical prior distributions on the mean intensity of gray and white matter are depicted, with overlaid contour plots showing some of the individual posteriors (randomly selected across the entire population).

Such results indicate that the proposed empirical Bayes learning scheme can serve to capture, not only the variability of mean tissue intensity across subjects for each of the modalities of interest, but also the amount of covariance between such modalities. Information of this sort can potentially be used in a number of different frameworks, for solving problems such as tissue segmentation, pathology detection or image synthesis.

In Fig. 5 we report a plot of the lower bound during model fitting as a qualitative demonstration of convergence.

Run time for the data set presented here was around 50 h on an 8 Core PC at 3 GHz with 32 GB RAM.

#### Validity of groupwise registration

The performance of groupwise registration achieved by the presented algorithm was assessed by computing pairwise overlap measures for all possible couples of spatially normalised test label maps (i.e. label maps that were not used for training the model), which were obtained by applying the inverse of the estimated deformation fields to the maps in their native space. The Dice<sup>4</sup> score coefficient was chosen as a metric of

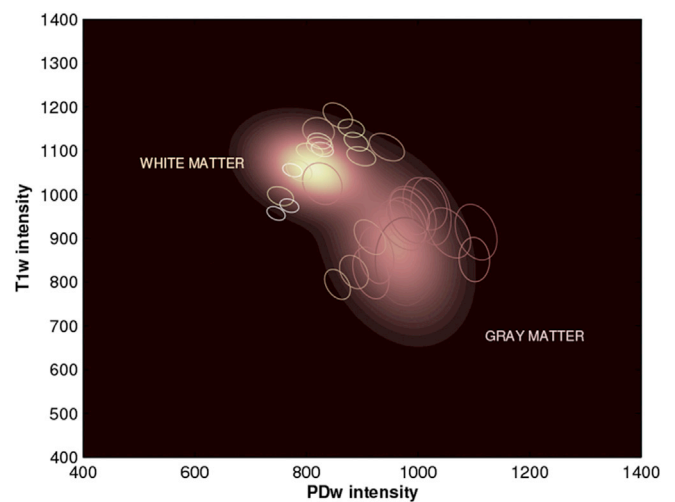


Fig. 4. Prior distribution over the mean intensity of gray and white matter, in T1- and PD-weighted data. Contour plots illustrate a number of randomly selected individual posteriors.

similarity.

Results are summarised in Fig. 6, where the accuracy of the algorithm presented here is compared to that achieved by the groupwise image registration method described in Avants et al. (2010), whose implementation is publicly available, as part of the Advanced normalisation Tools (ANTs) package, through the web site <http://stnava.github.io/ANTs/>. Indeed, the symmetric diffeomorphic registration framework implemented in ANTs has established itself as the state-of-the-art of medical image nonlinear spatial normalisation (Klein et al., 2009).

A number of options can be customised within the template construction framework distributed with ANTs. The experiments, whose results are reported here, were performed with the settings recommended in the package documentation for brain MR data, which are also reported in Table 2. The same strategy outlined above was applied in order to compute Dice score coefficients for all structural labels and all possible pairs of test images.

Results of this validation analyses indicate that the method presented

<sup>4</sup> The Dice score over two sets  $A$  and  $B$  is defined as  $DSC = 2 \frac{|A \cap B|}{|A| + |B|}$ .

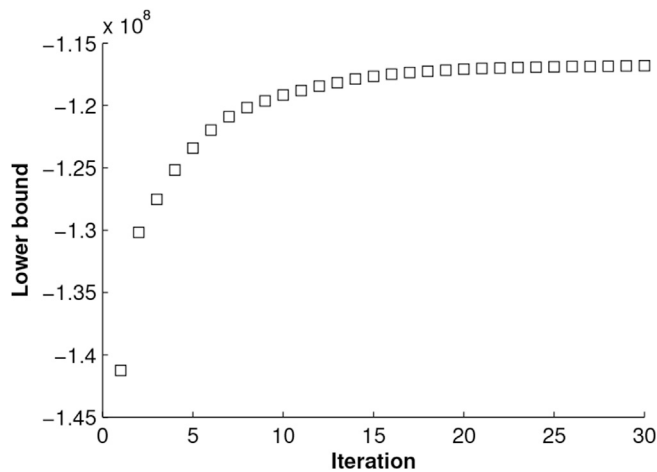


Fig. 5. Lower bound values as a function of iteration number during model training.

here, in spite of not being as accurate as ANTs for aligning some subcortical brain structures (e.g. thalamus, putamen, pallidum and brainstem), provided significantly better overlap when registering cortical regions, as assessed by means of paired t-tests with a significance threshold of 0.05 and without correcting for multiple comparisons.

#### Accuracy of tissue classification

The accuracy of tissue classification achieved by the method presented in this paper was first evaluated on test data that was used to create the templates but without providing manual labels for training the model. The aim in this case is to determine to which extent the proposed method can capture significant features of the training data, by learning from few annotated examples.

Dice scores were computed to compare the automated segmentations produced via semisupervised groupwise model fitting, with the ground truth, obtained by merging all the gray and white matter brain structures (labels) into two tissue classes respectively. The probabilistic gray and white matter segmentations produced by the proposed algorithm were thresholded at 0.5, in order to obtain binary label maps, directly comparable to the ground truth. Results are summarised in *seg*, which shows the distributions of Dice scores obtained for gray and white matter.

Such results were then compared to those produced by the brain segmentation algorithm implemented in SPM12, using the tissue probability maps illustrated in Fig. 2. Indeed the proposed approach extends and generalises the model underlying SPM12 segmentation method. This is achieved by replacing a small deformation approach to image registration with a diffeomorphic representation of shapes, as well as by introducing empirical intensity priors (within a variational Bayes framework) and by allowing learning of the tissue probability maps directly from the data.

Therefore, results of these analyses, which are also shown in Fig. 7, provide a direct insight into the performance gain achieved by incrementing the model available in SPM12 as indicated above. In particular, our experiments indicate that the proposed method enables higher tissue classification accuracy, for both gray and white matter (assessed by means of paired t-tests with a significance threshold of 0.05).

The experiments described in this section however did not test the generalisation capability of the proposed method, which would have required a k-fold cross-validation design in order to exploit as much as possible of the available data during training. Unfortunately this was highly impractical in this case due to the expensive computational cost of groupwise model fitting. The generalisation performance of the proposed model is instead assessed in the next section, using a hold-out approach and making use of both real and synthetic unseen brain MR data, for which tissue labels are known.

#### Modelling unseen data

Further validation experiments were performed to quantify the accuracy of the framework described in this paper to model unseen data, that is to say data that was not included in the atlas generation process.

In particular, we evaluated registration accuracy using data from the Internet Brain Segmentation Repository (IBSR), which is provided by the Centre for Morphometric Analysis at Massachusetts General Hospital (<http://www.cma.mgh.harvard.edu/ibsr/>). Experiments to assess bias correction and segmentation accuracy were instead performed on synthetic T1-weighted brain MR scans from the Brainweb database (<http://brainweb.bic.mni.mcgill.ca/>), which were simulated using a healthy anatomical model under different noise and bias conditions.

#### Accuracy of registration

To assess the performance of the presented method for spatially normalising unseen test data we made use of the ISBR data set, which consists of 18 T1-weighted brain images with manual labels of 84 anatomical structures. Such a data set was also used by Klein et al. (2009) for their evaluation of 14 brain image registration algorithms. Therefore the results presented in the remainder of this section allow comparing the proposed approach to a number of image registration methods available to the neuroimaging community.

For this experiment we registered each of the 18 unseen scans to the tissue probability maps presented in Fig. 2 by adopting the algorithmic framework presented in this paper. We then computed target overlap measures for all pairs of spatially normalised images. This was achieved by composing the inverse of each estimated deformation field with every other direct transformation and applying the resulting warps to the labels in their native space, so as to map between each couple of subjects. Our method, after averaging across different brain regions, achieved a median overlap score of 0.54 with the 25th and 75th percentiles equal to 0.51 and 0.57 respectively. Instead, the best performing method in the experiments performed by Klein et al. (2009) obtained a median score of 0.55 on the same data set.

Such results indicate that our approach, in spite of being intrinsically best suited for groupwise analyses, where the reference image is constructed iteratively from the same scans that are being spatially normalised (Avants et al., 2010), can provide accurate results also when mapping test data to a fixed reference that encodes the average shape of a different population. Indeed, this experiment was intended as a means to assess the generalisation capability of the templates constructed with the proposed approach, while ensuring a fairer comparison to the results of Klein et al. (2009), which were produced via pairwise non-linear registration rather than by estimating group averages.

#### Accuracy of bias correction

A healthy adult brain MR model was processed by means of the algorithm discussed here, using the brain and spinal cord templates previously constructed as tissue priors. Different levels of noise and bias field were added to the uncorrupted synthetic data, to test the behaviour of the proposed modelling scheme in different noise (1%, 3%, 7%) and bias conditions (20% and 40%).

The noise in these simulated images has Rayleigh statistics in the background and Rician statistics in the signal regions and its level is computed as a percent standard deviation ratio, relative to the MR signal, for a reference tissue (Cocosco et al., 1997).

Regarding the bias field instead, 20% bias is modelled as a smooth field in the range [0.9, 1.1] while 40% bias is obtained by rescaling of the 20% field, so as to range between 0.8 and 1.2.

Table 3 reports the Pearson product-moment correlation coefficients between the ground truth and the estimated bias fields, for the different bias ranges and noise levels. Results indicate that the similarity between the estimated and true bias decreases for more intense non-uniformity fields and higher noise levels.

Indeed this is not surprising, as the penalty term, which enforces



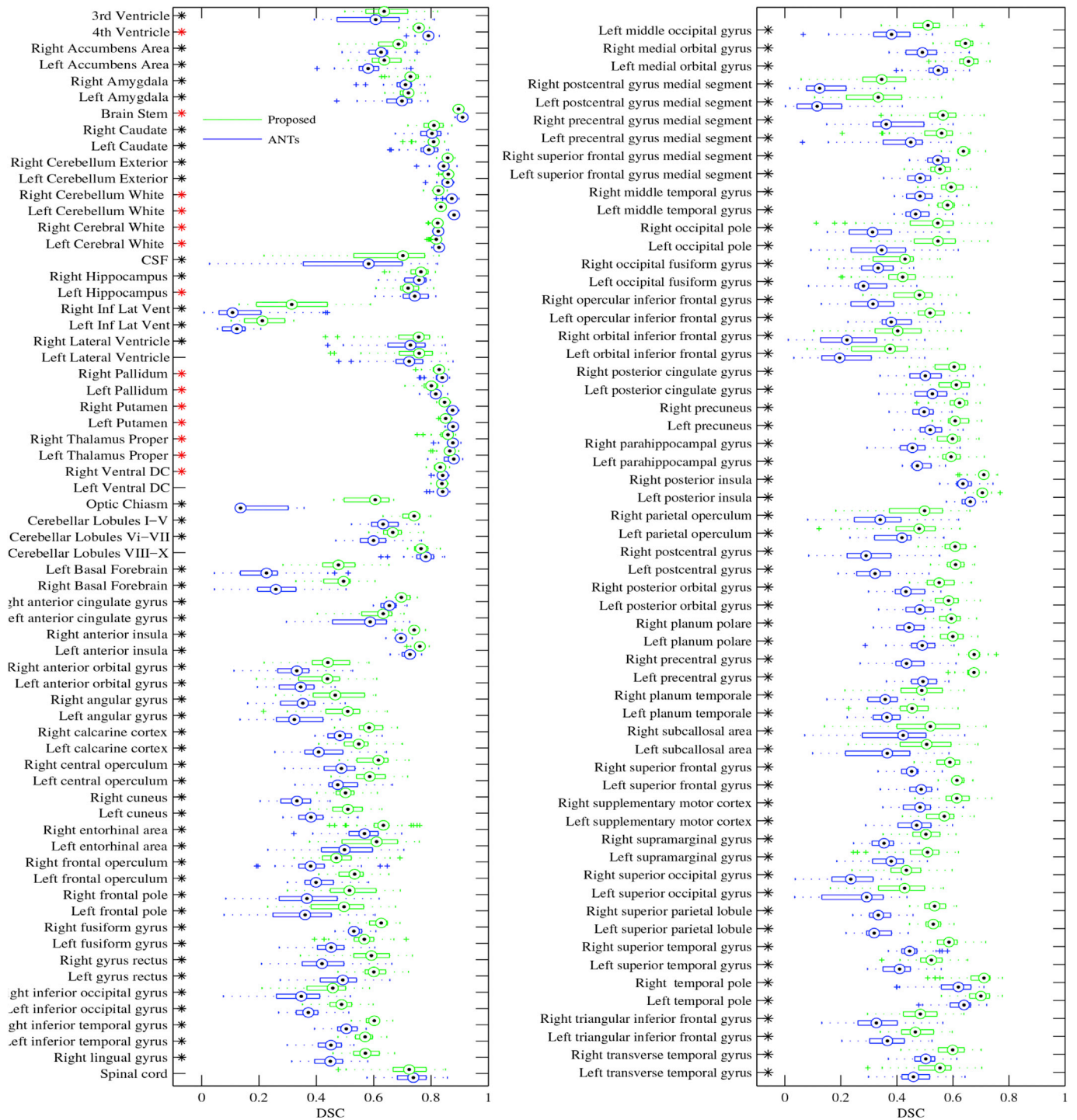


Fig. 6. Registration accuracy achieved by the presented method, compared to the performance of the diffeomorphic groupwise template construction method distributed with ANTs. Boxplots, in green for the proposed method and blue for ANTs, illustrate the distributions of overlap measures (Dice score coefficients) obtained by considering all possible pairs of spatially normalised test images, whose labels were not used for model fitting. Stars indicate statistically significant differences between the two methods, assessed by means of paired t-tests without correcting for multiple comparisons. Star colours encode the method that achieved higher performance (black for the proposed algorithm and red for ANTs).

smoothness of the bias field, has a greater impact in determining the shape of the estimated bias when the non-uniformity fields have a larger dynamic range. Nevertheless, results reported in the following section will show how this increased mismatch between the estimated and true bias, for higher non-uniformities, does not seem to affect the accuracy of tissue segmentation. On the other hand, the accuracy of bias correction is directly related to the amount of noise corrupting the data, mainly due to how this affects the precision associated with estimation of the Gaussian mixture parameters. For a comparison of these results with the

performance of SPM12 bias correction on simulated T1-weighted scans from the Brainweb database see Blaiotta et al. (2016).

*Accuracy of tissue classification*

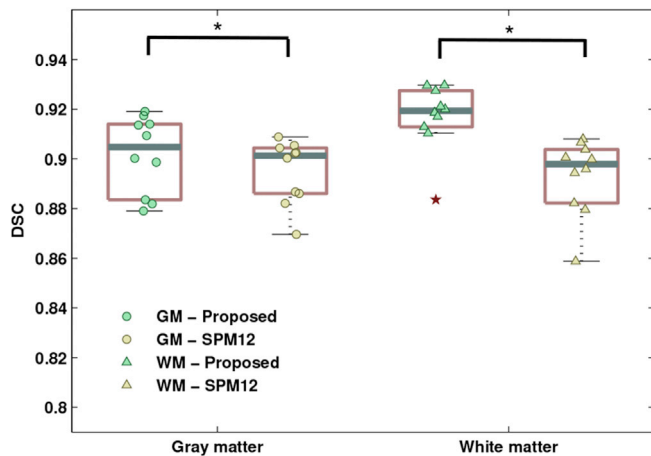
For the same synthetic data the accuracy of tissue classification was also evaluated, by comparing the similarity between the estimated gray and white matter segmentations and the underlying anatomical model.

Results are reported in Fig. 8, which shows the Dice score coefficients obtained under different bias and noise conditions.



**Table 2**  
Options selected to perform groupwise registration with ANTs, using the `antsMultivariateTemplateConstruction` script provided with the ANTs package.

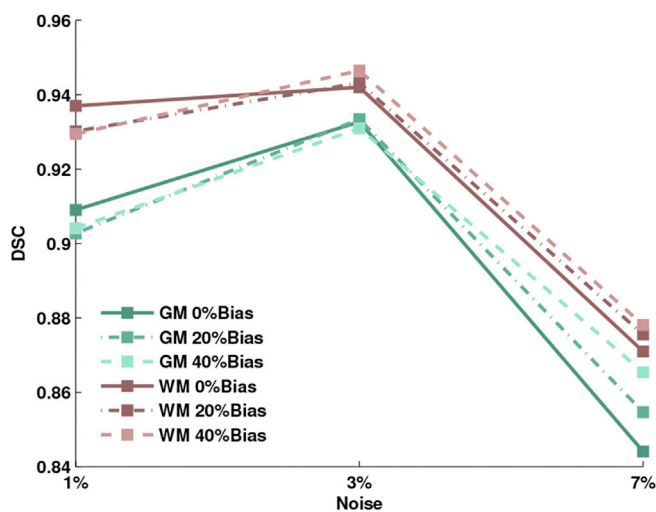
Option	Value
Similarity Metric	Cross-correlation (CC)
Transformation model	Greedy SyN (GR)
Initial rigid body	yes
N4 Bias Correction	yes
Number of resolution levels	4
Number of iterations	100 × 70 × 50 × 10
Gradient step	0.2
Number of template updates	4



**Fig. 7.** Brain segmentation accuracy of the presented method in comparison to SPM12 image segmentation algorithm. Boxplots indicate the distributions of Dice score coefficients, with overlaid scatter plots of the estimated scores. Red stars denote outliers.

**Table 3**  
Pearson's correlation coefficients between the ground truth bias fields and those estimated by the presented algorithm, for simulated T1-weighted data.

Bias	Noise		
	1%	3%	7%
20%	0.86	0.86	0.70
40%	0.72	0.72	0.51



**Fig. 8.** Dice scores between the estimated and ground truth segmentations for brain white matter and brain gray matter, under different noise and bias conditions, for synthetic T1-weighted data.

The Brainweb database has been extensively used in the neuroimaging community to validate MR image processing algorithms. Therefore the results reported here should be directly comparable to the performance of many brain segmentation techniques present in the literature.

**Conclusions**

This paper presented a comprehensive generative modelling framework, suitable to capture anatomical variability from large cross-sectional MR data sets. From a theoretical perspective, such a framework relies on variational probability density estimation techniques to model the observed data (i.e. MR image intensities). Additionally, a hierarchical modelling perspective is proposed, where observations from a population of subjects are used to construct empirical intensity priors, which can then serve to inform models of new data. Shape modelling is performed via groupwise diffeomorphic registration, thus ensuring bijective (i.e. one-to-one) differentiable mappings between anatomical configurations (Miller, 2004). Such an approach enables a rigorous mathematical encoding of anatomical shapes via deformable template matching (Christensen et al., 1996), therefore providing a quantitative framework for the analysis of shape variation and covariation.

Data for validating the method was collected from many different brain databases, most of which are publicly accessible to the research community. Results of our experiments, performed both on training and unseen test data, indicate that the proposed approach can accurately align complex anatomical shapes, such as the brain and spinal cord, and segment data into tissue types, while being robust to inter-scanner signal variations. Therefore, the proposed algorithm defines a convenient framework to extract volumetric and morphometric information from large structural neuroimaging data sets, in a fully automated manner. At the same time it provides outputs that can be readily interpreted, for instance via statistical hypothesis testing, with the ultimate goal of comparing different populations, treatment effects etc. (Ashburner and Friston, 2000).

Additionally, our results suggest that the proposed approach could be useful to construct templates capable of capturing the peculiar anatomical features of populations poorly represented by standard anatomical atlases (such as young or elderly populations, diseased populations, or individuals belonging to different ethnic groups (Tang et al., 2010; Fillmore et al., 2015)). This would not only lead to more accurate segmentation results, but as a direct consequence, also increase the reliability of subsequent data analyses, which build models of the segmented data to infer or predict clinically meaningful information.

It should be noted that, in spite of having been tested on neuroimaging data sets, our method was intentionally formulated as a general approach, which makes it potentially suitable to solve a wide range of imaging problems, for instance in the context of animal imaging studies or for the analysis of different human organs using multispectral MR data sets. We aim to explore some of this applications as part of our future work.

**Acknowledgments**

Claudia Blaiotta is co-funded by UCL and Zurich Balgrist Hospital, as part of the ‘UCL Impact’ award scheme. This research was supported by Wings for Life - Spinal Cord Research Foundation [WFL-CH-007/14]. The Wellcome Trust Centre for Neuroimaging is supported by core funding from the Wellcome Trust (091593/Z/10/Z). The OASIS project was funded by the National Institutes of Health grants P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584. The IXI project was supported by the EPSRC grant GR/S21533/02.

## Appendix A. Derivatives of the lower bound with respect to the affine parameters

The affine parameters, for each subject  $i$ , can be estimated (i.e. optimised) in a Gauss-Newton fashion, so as to maximise of the following objective function

$$\begin{aligned} \mathcal{E}_{af}^{(i)} &= \mathcal{D}^{(i)} + \mathcal{R}_{af}^{(i)} \\ &= \sum_{j=1}^{N_i} \sum_{k=1}^K \gamma_{ijk} \log \frac{w_{ik} \pi_k(\boldsymbol{\xi}_{ij})}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_{ij})} - \frac{1}{2} \mathbf{a}_i^T \boldsymbol{\Sigma}_a^{-1} \mathbf{a}_i, \end{aligned} \quad (\text{A.1})$$

with respect to  $\mathbf{a}_i$ .

The gradients and Hessians, which are useful to solve this problem are reported below. In particular, for the matching term, the following derivatives need to be computed

$$\frac{\partial \mathcal{D}^{(i)}}{\partial \mathbf{a}_i} = \sum_{j=1}^{N_i} \sum_{k=1}^K \left( \gamma_{ijk} - \frac{w_{ik} \pi_k(\boldsymbol{\xi}_{ij})}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_{ij})} \right) \mathbf{g}_{jk}^\pi, \quad (\text{A.2})$$

where  $\mathbf{g}_{jk}^\pi$  is defined as

$$\mathbf{g}_{jk}^\pi = \mathbf{B}_i^T ([\phi_{ij}, 1] \otimes \nabla [\log(\pi_k(\boldsymbol{\xi}_{ij}))]), \quad (\text{A.3})$$

with

$$\mathbf{B}_i^T = \frac{\partial \mathbf{S}_i}{\partial \mathbf{a}_i}, \quad (\text{A.4})$$

and

$$\mathbf{S}_i = \begin{bmatrix} \mathbf{T}_i & \mathbf{t}_i \\ 0 & 1 \end{bmatrix}. \quad (\text{A.5})$$

$$\begin{aligned} \frac{\partial^2 \mathcal{D}^{(i)}}{\partial \mathbf{a}_i^2} &= \sum_{j=1}^{N_i} \left( \sum_{k=1}^K \frac{w_{ik} \pi_k(\boldsymbol{\xi}_{ij})}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_{ij})} \mathbf{g}_{jk}^\pi \right) \\ &\times \left( \sum_{k=1}^K \frac{w_{ik} (\pi_k(\boldsymbol{\xi}_{ij}))}{\sum_{c=1}^K w_{ic} (\pi_c(\boldsymbol{\xi}_{ij}))} \mathbf{g}_{jk}^\pi \right)^T \\ &- \sum_{j=1}^{N_i} \sum_{k=1}^K \frac{w_{ik} \pi_k(\boldsymbol{\xi}_{ij})}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_{ij})} \mathbf{g}_{jk}^\pi (\mathbf{g}_{jk}^\pi)^T. \end{aligned} \quad (\text{A.6})$$

Gradients and Hessians of the penalty term are instead given by

$$\frac{\partial \mathcal{R}_{af}^{(i)}}{\partial \mathbf{a}_i} = -\boldsymbol{\Sigma}_a^{-1} \mathbf{a}_i, \quad (\text{A.7})$$

$$\frac{\partial^2 \mathcal{R}_{af}^{(i)}}{\partial \mathbf{a}_i^2} = -\boldsymbol{\Sigma}_a^{-1}. \quad (\text{A.8})$$

## Appendix B. Derivatives of the lower bound with respect to the initial velocities

Optimisation of the initial velocities, for each image  $i$ , requires maximising the following objective function

$$\begin{aligned} \mathcal{E}_{dif}^{(i)} &= \mathcal{D}^{(i)} + \mathcal{R}_{dif}^{(i)} \\ &= \sum_{j=1}^{N_i} \sum_{k=1}^K \gamma_{ijk} \log \frac{w_{ik} \pi_k(\boldsymbol{\xi}_{ij})}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_{ij})} - \frac{1}{2} \sum_{i=1}^M \|\mathbf{L}_u \mathbf{u}_i\|_{L^2}^2, \end{aligned} \quad (\text{B.1})$$

with respect to  $\mathbf{u}_i$ .

Here, we report the first and second derivatives of this objective function, which are useful to solve the registration problem using gradient-based techniques, such as the Gauss-Newton algorithm.

The gradient of the matching term  $\mathcal{D}^{(i)}$  with respect to  $\mathbf{u}_i$  is given by

$$\begin{aligned} \frac{\partial \mathcal{D}^{(i)}}{\partial \mathbf{u}_i} &= \sum_{k=1}^K \gamma_{ijk} \frac{\partial}{\partial \mathbf{u}_i} \left( \log \frac{w_{ik} \pi_k(\boldsymbol{\xi}_i)}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_i)} \right) \\ &= \sum_{k=1}^K \gamma_{ijk} \left( \mathbf{g}_k^\pi - \sum_{c=1}^K \frac{w_{ic} \pi_c(\boldsymbol{\xi}_i)}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_i)} \mathbf{g}_c^\pi \right), \end{aligned} \quad (\text{B.2})$$

which, making use of  $\sum_{k=1}^K \gamma_{ijk} = 1$ , can be rewritten as

$$\frac{\partial \mathcal{D}^{(i)}}{\partial \mathbf{u}_i} = \sum_{k=1}^K \left( \gamma_{ik} - \frac{w_{ik} \pi_k(\boldsymbol{\xi}_i)}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_i)} \right) \mathbf{g}_k^\pi, \quad (\text{B.3})$$

where  $\mathbf{g}_k^\pi$  is computed, at each voxel  $j$ , by

$$\mathbf{g}_{jk}^\pi = (\mathbf{T}_i, \mathbf{J}_{ij}^{\boldsymbol{\xi}})^T \nabla [\log(\pi_k(\boldsymbol{\xi}_{ij}))], \quad (\text{B.4})$$

and  $\mathbf{J}_i^{\boldsymbol{\xi}}$  indicates the Jacobian matrix of  $\boldsymbol{\xi}_{ij}$ .

An approximated positive semidefinite Hessian of  $\mathcal{D}$  can instead be computed by discarding the second derivatives of the logarithm of tissue priors

$$\frac{\partial^2}{\partial \mathbf{y}^2} \log \left( \frac{w_{ik}(\pi_k(\boldsymbol{\xi}_i(\mathbf{y})))}{\sum_{c=1}^K w_{ic}(\pi_c(\boldsymbol{\xi}_i(\mathbf{y})))} \right) = 0, \forall \mathbf{y} \in \Omega_i, \quad (\text{B.5})$$

to give

$$\begin{aligned} \frac{\partial^2 \mathcal{D}^{(i)}}{\partial \mathbf{u}_i^2} &= \left( \sum_{k=1}^K \frac{w_{ik} \pi_k(\boldsymbol{\xi}_i)}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_i)} \mathbf{g}_k^\pi \right) \\ &\times \left( \sum_{k=1}^K \frac{w_{ik}(\pi_k(\boldsymbol{\xi}_i))}{\sum_{c=1}^K w_{ic}(\pi_c(\boldsymbol{\xi}_i))} \mathbf{g}_k^\pi \right)^T \\ &- \sum_{k=1}^K \frac{w_{ik} \pi_k(\boldsymbol{\xi}_i)}{\sum_{c=1}^K w_{ic} \pi_c(\boldsymbol{\xi}_i)} \mathbf{g}_k^\pi (\mathbf{g}_k^\pi)^T. \end{aligned} \quad (\text{B.6})$$

Finally, the first and second derivatives of the penalty term  $\mathcal{R}$ , which are also required to optimise (B.1), can be computed by

$$\frac{\partial \mathcal{R}_{diff}^{(i)}}{\partial \mathbf{u}_i} = -\mathbf{L}_u^\dagger \mathbf{L}_u \mathbf{u}_i, \quad (\text{B.7})$$

$$\frac{\partial^2 \mathcal{R}_{diff}^{(i)}}{\partial \mathbf{u}_i^2} = -\mathbf{L}_u^\dagger \mathbf{L}_u. \quad (\text{B.8})$$

### Appendix C. Variational Gaussian mixtures: inference of missing data

The variational Bayes EM algorithm for fitting Gaussian mixture models, described in Blaiotta et al. (2016), can be generalised to handle the case where some components of the  $D$ -dimensional observation  $\mathbf{x}_j$  are missing.

Having denoted

$$\mathbf{x}_j = \begin{bmatrix} \mathbf{o}_j \\ \mathbf{h}_j \end{bmatrix}, \quad (\text{C.1})$$

with  $\mathbf{o}_j$  being the observed data and  $\mathbf{h}_j$  the missing (hidden) data, the Gaussian likelihood  $p(\mathbf{x}_j | z_{jk} = 1, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  can be expressed as

$$p(\mathbf{x}_j | z_{jk} = 1, \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = \mathcal{N} \left( \begin{bmatrix} \mathbf{o}_j \\ \mathbf{h}_j \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{\mu}_k^o \\ \boldsymbol{\mu}_k^h \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}_k^{o,o} & \boldsymbol{\Lambda}_k^{o,h} \\ \boldsymbol{\Lambda}_k^{h,o} & \boldsymbol{\Lambda}_k^{h,h} \end{bmatrix}^{-1} \right), \quad (\text{C.2})$$

by making use of block matrix notation to partition the mean vector  $\boldsymbol{\mu}_k$  and the precision matrix  $\boldsymbol{\Lambda}_k$ .

In this case  $\mathbf{h}_j$  is treated as an unobserved random variable. Thus, in a variational Bayes setting, an additional posterior factor can be introduced for each missing data point  $\mathbf{h}_j$  to give

$$\begin{aligned}
 q(\mathbf{H}, \mathbf{Z}, \Theta_\mu, \Theta_\Sigma) &= q(\mathbf{H})q(\mathbf{Z})q(\Theta_\mu, \Theta_\Sigma) \\
 &= q(\mathbf{Z})q(\Theta_\mu, \Theta_\Sigma) \prod_{j=1}^N q(\mathbf{h}_j) .
 \end{aligned}
 \tag{C.3}$$

Making use of the general result  $q_s(\Theta_s) \propto \exp(\mathbb{E}_{s \neq s}[\log p(\mathbf{X}, \Theta)])$  (Bishop, 2006), an approximated posterior on the missing data point  $\mathbf{h}_j$  can be computed by

$$\begin{aligned}
 \log q(\mathbf{h}_j) &= \mathbb{E}_{\mathbf{Z}, \Theta_\mu, \Theta_\Sigma} [\log p(\mathbf{x}_j, \mathbf{z}_j, \Theta_\mu, \Theta_\Sigma | \Theta_\pi)] + \text{const} \\
 &= \mathbb{E}_{\mathbf{Z}, \Theta_\mu, \Theta_\Sigma} [\log p(\mathbf{x}_j | \mathbf{z}_j, \Theta_\mu, \Theta_\Sigma)] + \mathbb{E}_{\mathbf{Z}} [\log p(\mathbf{z}_j | \Theta_\pi)] \\
 &+ \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\log p(\Theta_\mu, \Theta_\Sigma)] + \text{const} ,
 \end{aligned}
 \tag{C.4}$$

where  $\Theta_\pi$  denotes the mixing proportion parameter set, treated here via maximum likelihood, and  $p(\Theta_\mu, \Theta_\Sigma)$  is a conjugate Gaussian-Wishart prior on the means and covariances of the model.

Ignoring the terms independent from  $\mathbf{h}_j$ , equation (C.4) can be rewritten as

$$\begin{aligned}
 \log q(\mathbf{h}_j) &= \sum_{k=1}^K \gamma_{jk} \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\log \mathcal{N}(\mathbf{x}_j | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)] + \text{const} \\
 &= \frac{1}{2} \sum_{k=1}^K \gamma_{jk} \mathbf{h}_j^T \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\Lambda}_k^{h,h}] \mathbf{h}_j \\
 &+ \sum_{k=1}^K \gamma_{jk} \mathbf{h}_j^T \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\Lambda}_k^{o,h}] (\mathbf{o}_j - \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\mu}_k^o]) \\
 &- \sum_{k=1}^K \gamma_{jk} \mathbf{h}_j^T \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\Lambda}_k^{h,h}] \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\mu}_k^h] + \text{const} .
 \end{aligned}
 \tag{C.5}$$

The previous equation indicates that the unobserved value  $\mathbf{h}_j$  is drawn from a Gaussian mixture distribution with mixing proportions equal to the posterior (after having observed  $\mathbf{o}_j$ ) membership probabilities  $\{\gamma_{jk}\}_{k=1, \dots, K}$ , while the Gaussian means  $\{\mathbf{n}_{jk}\}_{k=1, \dots, K}$  and covariances  $\{\mathbf{P}_{jk}\}_{k=1, \dots, K}$  are given by

$$\begin{aligned}
 \mathbf{n}_{jk} &= \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\mu}_k^h] + (\mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\Lambda}_k^{h,h}])^{-1} \\
 &\times \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\Lambda}_k^{o,h}] (\mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\mu}_k^o] - \mathbf{o}_j) ,
 \end{aligned}
 \tag{C.6}$$

$$\mathbf{P}_k = \mathbb{E}_{\Theta_\mu, \Theta_\Sigma} [\boldsymbol{\Lambda}_k^{h,h}] .
 \tag{C.7}$$

Given the posteriors  $q(\mathbf{Z})$  and  $q(\mathbf{H})$ , the following sufficient statistics of  $\mathbf{X}$  can be computed

$$\mathbf{s}_{1k} = \begin{bmatrix} \sum_{j=1}^N \gamma_{jk} \mathbf{o}_j \\ \sum_{j=1}^N \gamma_{jk} \mathbf{n}_{jk} \end{bmatrix} ,
 \tag{C.8}$$

$$\mathbf{s}_{2k} = \begin{bmatrix} \sum_{j=1}^N \gamma_{jk} \mathbf{o}_j \mathbf{o}_j^T & \sum_{j=1}^N \gamma_{jk} \mathbf{o}_j \mathbf{n}_{jk}^T \\ \sum_{j=1}^N \gamma_{jk} \mathbf{n}_{jk} \mathbf{o}_j^T & \sum_{j=1}^N \gamma_{jk} (\mathbf{n}_k \mathbf{n}_k^T + (\mathbf{P}_k)^{-1}) \end{bmatrix} .
 \tag{C.9}$$

Once such sufficient statistics have been evaluated, they can be used to update the Gaussian-Wishart posteriors  $q(\Theta_\mu, \Theta_\Sigma)$  in the exact same way as in Blaiotta et al. (2016). Such posteriors are in turn used to compute the expectations that appear in equation (C.6) and (C.7), in an iterative EM fashion.

An inner product naturally induces a norm by  $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$ , therefore every inner product space is also a normed vector space (Dieudonné, 2013).

## References

Allasonnière, S., Trouvé, A., Younes, L., 2005. Geodesic shooting and diffeomorphic matching via textured meshes. In: International Workshop on Energy Minimisation Methods in Computer Vision and Pattern Recognition. Springer, pp. 365–381.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. *Neuroimage* 38 (1), 95–113.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry: the methods. *Neuroimage* 11 (6), 805–821.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. *Neuroimage* 26 (3), 839–851.

Ashburner, J., Friston, K.J., 2011. Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *Neuroimage* 55 (3), 954–967.

Ashburner, J., Ridgway, G.R., 2013. Symmetric diffeomorphic modeling of longitudinal structural MRI. *Front. Neurosci.* 6, 197.

Avants, B., Gee, J.C., 2004. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *Neuroimage* 23, S139–S150.

Avants, B.B., Schoenemann, P.T., Gee, J.C., 2006. Lagrangian frame diffeomorphic image registration: morphometric comparison of human and chimpanzee cortex. *Med. Image Anal.* 10 (3), 397–412.

Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J.C., 2010. The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49 (3), 2457–2466.

Bajcsy, R., Lieberman, R., Reivich, M., 1983. A computerised system for the elastic matching of deformed radiographic images to idealized atlas images. *J. Comput. Assisted Tomogr.* 7 (4), 618–625.



- Balafar, M., 2014. Gaussian mixture model based segmentation methods for brain MRI images. *Artif. Intell. Rev.* 41 (3), 429–439.
- Beg, M.F., Khan, A., 2006. Computing an average anatomical atlas using LDDMM and geodesic shooting. In: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006. IEEE, pp. 1116–1119.
- Beg, M.F., Miller, M.I., Trounev, A., Younes, L., 2005. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* 61 (2), 139–157.
- Bertsekas, D.P., 1999. *Nonlinear Programming*. Athena scientific Belmont.
- Bhatia, K.K., Aljabar, P., Boardman, J.P., Srinivasan, L., Murgasova, M., Counsell, S.J., Rutherford, M.A., Hajnal, J.V., Edwards, A.D., Rueckert, D., 2007. Groupwise combined segmentation and registration for atlas construction. In: Proc. International Conference on Medical Image Computing and Computer-assisted Intervention, MICCAI 2007. Springer, pp. 532–540.
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*, vol. 1. Springer New York.
- Blaiotta, C., Cardoso, M.J., Ashburner, J., 2016. Variational inference for medical image segmentation. *Comput. Vis. Image Underst.* 151, 14–28.
- Bowden, A.E., Rabbitt, R.D., Weiss, J.A., 1998. Anatomical registration and segmentation by warping template finite element models. In: BIOS'98 International Biomedical Optics Symposium. International Society for Optics and Photonics, pp. 469–476.
- Brett, M., Johnsrude, I.S., Owen, A.M., 2002. The problem of functional localization in the human brain. *Nat. Rev. Neurosci.* 3 (3), 243–249.
- Brex, P.A., Ciccarelli, O., O'riordan, J.I., Sailer, M., Thompson, A.J., Miller, D.H., 2002. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *N. Engl. J. Med.* 346 (3), 158–164.
- Chapelle, O., Schölkopf, B., Zien, A., et al., 2006. *Semi-supervised Learning*. MIT press Cambridge.
- Chetelat, G., Landeau, B., Eustache, F., Mezenge, F., Viader, F., De La Sayette, V., Desgranges, B., Baron, J.-C., 2005. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study. *Neuroimage* 27 (4), 934–946.
- Christensen, G.E., 1999. Consistent linear-elastic transformations for image matching. In: Biennial International Conference on Information Processing in Medical Imaging. Springer, pp. 224–237.
- Christensen, G.E., Rabbitt, R.D., Miller, M.I., 1996. Deformable templates using large deformation kinematics. *IEEE Trans. Image Process.* 5 (10), 1435–1447.
- Chui, H., Win, L., Schultz, R., Duncan, J., Rangarajan, A., 2001. A unified feature registration method for brain mapping. In: Biennial International Conference on Information Processing in Medical Imaging. Springer, pp. 300–314.
- Cocosco, C.A., Kollokian, V., Kwan, R.K.-S., Pike, G.B., Evans, A.C., 1997. Brainweb: online interface to a 3D MRI simulated brain database. In: *Neuroimage*. Citeseer.
- Denton, E.R., Sonoda, L.I., Rueckert, D., Rankin, S.C., Hayes, C., Leach, M.O., Hill, D.L., Hawkes, D.J., 1999. Comparison and evaluation of rigid, affine, and nonrigid registration of breast MR images. *J. Comput. Assisted Tomogr.* 23 (5), 800–805.
- Dieudonné, J., 2013. *Foundations of Modern Analysis*. Read Books Ltd.
- Evans, A., Kamber, M., Collins, D., MacDonald, D., 1994. An MRI-based probabilistic atlas of neuroanatomy. In: *Magnetic Resonance Scanning and Epilepsy*. Springer, pp. 263–274.
- Falk, J.E., 1967. Lagrange multipliers and nonlinear programming. *J. Math. Anal. Appl.* 19 (1), 141–159.
- Filipovych, R., Davatzikos, C., Initiative, A.D.N., et al., 2011. Semi-supervised pattern classification of medical images: application to mild cognitive impairment (MCI). *Neuroimage* 55 (3), 1109–1119.
- Fillmore, P.T., Phillips-Meek, M.C., Richards, J.E., 2015. Age-specific MRI brain and head templates for healthy adults from 20 through 89 years of age. *Front. Aging Neurosci.* 7, 44.
- Fonov, V.S., Le Troter, A., Taso, M., De Leener, B., Lévesque, G., Benhamou, M., Sdika, M., Benali, H., Pradat, P.-F., Collins, D.L., et al., 2014. Framework for integrated mri average of the spinal cord white and gray matter: the mni-poly-amu template. *Neuroimage* 102, 817–827.
- Fox, N.C., Cousins, S., Scahill, R., Harvey, R.J., Rossor, M.N., 2000. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects. *Arch. Neurol.* 57 (3), 339–344.
- Friston, K.J., Holmes, A., Poline, J.-B., Price, C.J., Frith, C.D., 1996. Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage* 4 (3), 223–235.
- Good, C.D., Johnsrude, I., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *Neuroimage* 14 (3), 685–700.
- Greenspan, H., Ruf, A., Goldberger, J., 2006. Constrained gaussian mixture model framework for automatic segmentation of mr brain images. *IEEE Trans. Med. Imaging* 25 (9), 1233–1245.
- Grenander, U., Miller, M.I., 1998. Computational anatomy: an emerging discipline. *Q. Appl. Math.* 56 (4), 617–694.
- Helms, G., Dathé, H., Dechent, P., 2008. Quantitative FLASH MRI at 3t using a rational approximation of the Ernst equation. *Magnetic Reson. Med.* 59 (3), 667–672.
- Hernandez, M., Olmos, S., 2008. Gauss-Newton optimisation in diffeomorphic registration. In: 5th IEEE International Symposium on Biomedical Imaging: from Nano to Macro, 2008. IEEE, pp. 1083–1086.
- Iglesias, J.E., Sabuncu, M.R., Van Leemput, K., 2012. A generative model for multi-atlas segmentation across modalities. In: 9th IEEE International Symposium on Biomedical Imaging, 2012. IEEE, pp. 888–891.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5 (2), 143–156.
- Joshi, S., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage* 23, S151–S160.
- Joshi, S.C., Miller, M.I., 2000. Landmark matching via large deformation diffeomorphisms. *IEEE Trans. Image Process.* 9 (8), 1357–1370.
- Khan, A.R., Wang, L., Beg, M.F., 2008. Freesurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping. *Neuroimage* 41 (3), 735–746.
- Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46 (3), 786–802.
- Landman, B.A., Warfield, S., 2012. MICCAI 2012: grand challenge and workshop on multi-atlas labeling. In: Proc. International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2012.
- Lester, H., Arridge, S.R., 1999. A survey of hierarchical non-linear medical image registration. *Pattern Recognit.* 32 (1), 129–149.
- Lorenzen, P., Davis, B., Gerig, G., Bullitt, E., Joshi, S., 2004. Multi-class posterior atlas formation via unbiased Kullback-Leibler template estimation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. Springer, pp. 95–102.
- Mahapatra, D., Sun, Y., 2012. Integrating segmentation information for improved MRF-based elastic image registration. *IEEE Trans. Image Process.* 21 (1), 170–183.
- Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2007. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cognitive Neurosci.* 19 (9), 1498–1507.
- Meng, X.-L., Rubin, D.B., 1993. Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80 (2), 267–278.
- Miller, M.I., 2004. Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *Neuroimage* 23, S19–S33.
- Miller, M.I., Trounev, A., Younes, L., 2006. Geodesic shooting for computational anatomy. *J. Math. Imaging Vis.* 24 (2), 209–228.
- Moderstizki, J., 2004. *Numerical Methods for Image Registration*. Oxford University Press on Demand.
- Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L., et al., 1998. A comparison of similarity measures for use in 2-D-3-D medical image registration. *IEEE Trans. Med. Imaging* 17 (4), 586–595.
- Pluta, J., Avants, B.B., Glynn, S., Awate, S., Gee, J.C., Detre, J.A., 2009. Appearance and incomplete label matching for diffeomorphic template based hippocampus segmentation. *Hippocampus* 19 (6), 565–571.
- Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.M., 2006. A Bayesian model for joint segmentation and registration. *Neuroimage* 31 (1), 228–239.
- Powell, M.J., 1978. *A fast algorithm for nonlinearly constrained optimisation calculations*. In: *Numerical Analysis*. Springer, pp. 144–157.
- Ribbens, A., Hermans, J., Maes, F., Vandermeulen, D., Suetens, P., 2014. Unsupervised segmentation, clustering, and groupwise registration of heterogeneous populations of brain MR images. *IEEE Trans. Med. Imaging* 33 (2), 201–224.
- Shen, D., Davatzikos, C., 2004. Measuring temporal morphological changes robustly in brain MR images via 4-dimensional template warping. *Neuroimage* 21 (4), 1508–1517.
- Stroman, P.W., Figley, C.R., Cahill, C.M., 2008. Spatial normalization, bulk motion correction and coregistration for functional magnetic resonance imaging of the human cervical spinal cord and brainstem. *Magn. Reson. Imaging* 26 (6), 809–814.
- Snyder, M., Brechbuhler, C., Szckely, G., Gerig, G., 2000. Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Trans. Med. Imaging* 19 (3), 153–165.
- Tang, Y., Hojatkashani, C., Dinov, I.D., Sun, B., Fan, L., Lin, X., Qi, H., Hua, X., Liu, S., Toga, A.W., 2010. The construction of a Chinese MRI brain atlas: a morphometric comparison study between Chinese and Caucasian cohorts. *Neuroimage* 51 (1), 33–41.
- Thompson, P.M., Woods, R.P., Mega, M.S., Toga, A.W., 2000. Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. *Hum. Brain Mapp.* 9 (2), 81–92.
- Trounev, A., 1998. Diffeomorphisms groups and pattern matching in image analysis. *Int. J. Comput. Vis.* 28 (3), 213–221.
- van Tulder, G., de Bruijne, M., 2015. Why does synthesized data improve multi-sequence classification?. In: Proc. International Conference on Medical Image Computing and Computer-assisted Intervention, MICCAI 2015. Springer, pp. 531–538.
- Vialard, F.-X., Risser, L., Rueckert, D., Holm, D.D., 2012. Diffeomorphic atlas estimation using geodesic shooting on volumetric images. *Ann. BMVA* 2012.
- Warfield, S., Robatino, A., Dengler, J., Jolesz, F., Kikinis, R., 1999. Nonlinear registration and template driven segmentation. *Brain Warping* 4, 67–84.
- Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23 (7), 903–921.
- Warner, F.W., 2013. *Foundations of Differentiable Manifolds and Lie Groups*, vol. 94. Springer Science & Business Media.

- Weiskopf, N., Suckling, J., Williams, G., Correia, M.M., Inkster, B., Tait, R., Ooi, C., Bullmore, E.T., Lutti, A., 2013. Quantitative multi-parameter mapping of R, PD\*, MT, and R2\* at 3T: a multi-center validation. *Front. Neurosci.* 7, 95.
- Yezzi, A., Zollei, L., Kapur, T., 2001. A variational framework for joint segmentation and registration. In: *Mathematical Methods in BioMedical Image Analysis, MMBIA 2001*, IEEE Workshop on. IEEE, pp. 44–51.
- Younes, L., 2010. *Shapes and Diffeomorphisms*, vol. 171. Springer Science & Business Media.
- Younes, L., Arrate, F., Miller, M.I., 2009. Evolutions equations in computational anatomy. *Neuroimage* 45 (1), S40–S50.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. imaging* 20 (1), 45–57.