



HAL
open science

A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes

Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Cyril Poupon, Andreas Hartmann, Nicholas Ayache, Stanley Durrleman

► **To cite this version:**

Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Cyril Poupon, et al.. A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes. *Medical Image Analysis*, 2017, 35, pp.458-474. 10.1016/j.media.2016.08.011 . hal-01359423

HAL Id: hal-01359423

<https://inria.hal.science/hal-01359423>

Submitted on 2 Sep 2016

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

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

A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes

Pietro Gori^{a,b,c,d,e,*}, Olivier Colliot^{a,b,c,d,e,f}, Linda Marrakchi-Kacem^{a,b,c,d,e}, Yulia Worbe^{a,b,c,d,e,f}, Cyril Poupon^g,
Andreas Hartmann^{a,b,c,d,e,f}, Nicholas Ayache^h, Stanley Durrleman^{a,b,c,d,e}

^aAramis project-team, Inria, Paris, France

^bSorbonne Universités, UPMC Univ Paris 06, UM 75, ICM, F-75013 Paris, France

^cInserm, U1127, ICM, F-75013 Paris, France

^dCNRS, UMR 7225, ICM, F-75013 Paris, France

^eInstitut du Cerveau et de la Moelle Épinière, Paris, France

^fAP-HP, Pitié-Salpêtrière Hospital, Departments of Neurology and Neuroradiology, F-75013, Paris, France

^gNeuroSpin, CEA, Gif-Sur-Yvette, France

^hAsclepios project-team, Inria, Sophia Antipolis, France

Abstract

We present a Bayesian framework for atlas construction of multi-object shape complexes comprised of both surface and curve meshes. It is general and can be applied to any parametric deformation framework and to all shape models with which it is possible to define probability density functions (PDF). Here, both curve and surface meshes are modelled as Gaussian random varifolds, using a finite-dimensional approximation space on which PDFs can be defined. Using this framework, we can automatically estimate the parameters balancing data-terms and deformation regularity, which previously required user tuning. Moreover, it is also possible to estimate a well-conditioned covariance matrix of the deformation parameters. We also extend the proposed framework to data-sets with multiple group labels. Groups share the same template and their deformation parameters are modelled with different distributions. We can statistically compare the groups' distributions since they are defined on the same space. We test our algorithm on 20 Gilles de la Tourette patients and 20 control subjects, using three sub-cortical regions and their incident white matter fiber bundles. We compare their morphological characteristics and variations using a single diffeomorphism in the ambient space. The proposed method will be integrated with the Deformetrica software package, publicly available at www.deformetrica.org.

Keywords: shape, Bayesian, varifolds, fiber bundle, morphometry, complex, multi-object, atlas

1. Introduction

Morphological analysis of human organs based on Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) scans is an important field in medical imaging. An example of clinical application is the identification and quantification of the effect of a pathology on the anatomy of the brain (i.e., hippocampal atrophy in the Alzheimer's disease (Gerardin et al., 2009)) or of the heart (i.e., Tetralogy of Fallot (Mansi et al., 2011), stroke or myocardial ischemia). This provides precious insights into the pathological development giving the possibility to identify biomarkers of disease progression, quantitative rules for disease evaluation, therapy planning and monitoring.

The analysis can be carried out either directly on images or on selected structures which are previously segmented as 3D meshes. The second strategy has the drawback that it depends on an accurate segmentation, which can be challenging in some cases, but it has also some advantages. First, it is possible to focus only on certain structures belonging to a particular area of an organ. Second, it allows visualisation and direct analysis of the entire 3D geometry

*Corresponding author.

Address: ICM - équipe Aramis - Hôpital de la Pitié-Salpêtrière - 47, boulevard de l'hôpital, 75013, Paris, France.

Email address: pietro.gori@inria.fr (Pietro Gori)

of a structure without scrolling through a stack of images. And third, it becomes possible to combine different imaging modalities since one can analyse together meshes obtained from structural MRI images, CT scans and diffusion MRI images. Furthermore, the two strategies can be combined together in an iconic-geometric setting (Siless et al., 2012; Graciano Fouquier et al., 2014).

The structures of an organ can be modelled as 3D surfaces or 3D curves. In the brain, surfaces are used to model sub-cortical nuclei or cortical surface (Durrleman et al., 2009; Auzias et al., 2011). In the heart they are used as models of the left, right ventricle or of the entire myocardium (Mansi et al., 2011; Palit et al., 2015; Lombaert and Peyrat, 2013). Lungs and liver can also be modelled as 3D surfaces (Gorbunova et al., 2010). Curves are instead employed in the brain to model trajectories of ensembles of neural tracts, commonly called fiber bundles (O'Donnell et al., 2009; Durrleman et al., 2011a). In the heart they can be used as models of cardiac myofibers (Palit et al., 2015) whereas in the lungs or liver they may be employed to model blood vessels Gorbunova et al. (2010). The geometrical representation of an entire organ may thus combine surface and curve meshes into a single multi-object complex, which we call *shape complex*.

There are several examples of mesh based morphometry in the literature. Most of them use a single-object approach since they select and analyse only one particular structure of an organ (Golland et al., 2005; Niethammer et al., 2007; Davies et al., 2010; Hufnagel et al., 2009; Kurtek et al., 2011; Savadjiev et al., 2012; Cury et al., 2015). This strategy limits the extent of the clinical conclusion to the chosen object, thus neglecting the information given by the surrounding structures. It does not seem appropriate to study intricate and complex systems of inter-connected objects such as the human organs. There is a growing awareness that multi-object studies could be more befitting. Some examples are Gorcowski et al. (2010); Cates et al. (2008); Bossa et al. (2011); Qiu et al. (2010); Durrleman et al. (2014) where the authors have shown that one better discriminates controls from patients when also considering the relative position between structures.

In this paper, we focus on a general morphometric approach called atlas construction. It consists of estimating an average shape complex of a population called *template complex* and the deformations of the embedding space which warp the template complex to the shape complexes of every subject. Deformations put into correspondence the template with every shape complex. The template captures the common morphological characteristics of the population and the deformations capture the variations in shape across the population. Every deformation is a single 3D diffeomorphism (a smooth deformation with smooth inverse) which transforms the entire 3D space. All meshes are therefore deformed together whatever their number or kind. This kind of deformation preserves the spatial organization of the structures since no intersection, folding or shearing may occur. It is thus possible to study a human organ as a whole, and not as a set of independent components, analysing not only the shape of each structure, but also their relative position.

The deformations align each structure of the template complex with the corresponding ones of the subject shape complex. To this end, one needs to define a similarity measure for every structure. Different metrics have been proposed for surfaces and curves. Some of them assume that it is possible to find point-correspondences between homologous structures, called landmarks. This is usually hard (or even impossible) especially when working with groups of curves. In this paper, we have opted for a correspondence-free metric based on the framework of *varifolds* (Charon and Trounev, 2013) which can be used with both surfaces and curves. It requires neither point-correspondences nor curve-correspondences. Moreover, surfaces and curves are treated as instances of the same mathematical object. This simplifies the atlas construction and the following statistical analysis.

A standard procedure to estimate a multi-object atlas (Durrleman et al., 2014; Avants and Gee, 2004; Ma et al., 2010) involves an optimization scheme where all template structures are optimised together. Every structure is weighted by a scalar which balances the importance of the structure with respect to the other structures and with respect to a regularity term on the deformations. These weights are fixed by the user and the results are rather sensitive to them. Optimal values should be determined by cross-validation which is time consuming and computationally intensive. Moreover, the computational load increases with the number of structures and subjects under study, making sometimes infeasible the automatic estimate of the weights via cross-validation.

Another strategy consists in embedding the atlas procedure into a statistical setting. Using a generative model, every shape can be interpreted as the sum between a deformation of the template and noise (Allasonnière et al., 2007). Moreover, considering the deformation parameters, template and noise as random variables, it is possible to use Maximum Likelihood (ML) or Maximum A Posteriori (MAP) estimations to infer the atlas. If the noise follows a Gaussian distribution, it turns out that the weight of a structure can be interpreted as the variance of the noise and

that it can be automatically estimated with ML or MAP techniques (Allasonnière et al., 2007; Zhang et al., 2013; Simpson et al., 2012; Folgoc et al., 2014). Other authors also proposed to infer the level of regularisation from the data by modelling it as the variance of the deformation parameters or deformation field (Simpson et al., 2012; Folgoc et al., 2014; Risholm et al., 2013; Wassermann et al., 2014). However, most of these works are about registration and atlas construction of *images*. Few of them focus on *shapes*. Moreover, they are usually based on single-object complexes and the variance of the data noise as well as the regularity level are usually fixed by the user. Three examples are Ma et al. (2010) where the authors introduced a surface single-object diffeomorphic atlas procedure using a hyper-template and Durrleman et al. (2011a, 2014) where the authors employed a generative model adapted for complexes of only curves or only surfaces respectively.

Here, we propose to embed a diffeomorphic multi-object atlas construction into a Bayesian framework. Shape complexes can be composed of curves and surfaces and they are both modelled as varifolds. The noise of each structure is assumed to be a random Gaussian varifold which belongs to an infinite dimensional space. We define a finite-dimensional approximation space where the noise is projected to and where it is possible to define probability distributions. Furthermore, we use a prior distribution on the variance of the noise as in Allasonnière et al. (2007) adapted for shape complexes modelled as varifolds. This allows us to automatically estimate all balancing weights, one for each structure, in one single atlas construction. Moreover, we assume that also the deformation parameters follow a Gaussian distribution. We use a prior distribution on their covariance matrix similarly to Allasonnière et al. (2007) which allows us to estimate it automatically even with a small sample size. Furthermore, the estimated covariance matrix is always well-conditioned whatever the number of subjects and it can be directly employed to compute statistics about the population under study.

This paper extends our previous work Gori et al. (2013) by providing a more detailed methodological description and more comprehensive results. Furthermore, we take advantage of the proposed Bayesian framework to define a multi-population atlas construction. We use the subjects' clinical diagnosis to subdivide the data-set into different groups. Their deformation parameters are then modelled as different Gaussian distributions. Since all sets of deformation parameters are defined on the same space, we can quantitatively compare them and the moments of their distributions such as the covariance matrices. Thanks to the proposed Bayesian framework, this multi-population analysis can be performed at no additional cost, namely without tuning any new balancing weight.

The paper is organised as follows. In Section 2, we first formulate the proposed Bayesian framework in a rather general way, showing that it could be employed with different shape models and deformation settings. We then present how to compute a multi-population atlas construction. Afterwards, we describe how to model multi-object shape complexes with varifolds and how to define a varifold random variable. Then, we show the diffeomorphic framework used in the experiments and how to integrate it in our statistical setting. In the last part of Section 2, we present the gradient-descent scheme used to optimise the cost function and how to initialise the template of both surfaces and curves. Eventually, we test the robustness of our algorithm with respect to the hyperparameters of the prior distributions and we also show how the proposed method can be employed to assess morphological differences between two groups of subjects.

2. Methods

2.1. Bayesian atlas construction

We define the mesh of structure j belonging to subject i as S_{ij} . The total number of subjects is N and the number of analysed structures is M . Every subject shape complex S_i is defined, for the moment, in a generic way as the ensemble of meshes S_{ij} of subject i . We model it as a deformed template complex $\phi_i(\mathbf{T})$ plus a residual ϵ_i . Both \mathbf{T} and ϵ_i are also defined as the ensembles of templates T_j and residuals ϵ_{ij} . This formulation is known as forward model where we assume that all elements belong to an algebraic structure where addition is defined (Durrleman et al., 2011b; Allasonnière et al., 2007; Ma et al., 2008) :

$$S_i = \phi_i(\mathbf{T}) + \epsilon_i \quad (1)$$

The deformation ϕ_i , proper to subject i , can belong to any diffeomorphic framework present in the literature. The only requirement, for the scope of this paper, is that it deforms the entire ambient space, namely all structures of the template complex simultaneously. The goal of the atlas construction is to estimate the template complex and its

morphological variations within the population of shape complexes. The variations are described by the ensemble of deformations $\{\phi_i\}$ and each one of them is parametrised by a set of parameters $\alpha_i \in \mathbb{R}^q$. We assume that these parameters follow a Gaussian distribution with zero mean and covariance matrix Γ_α :

$$\alpha_i \sim N(0, \Gamma_\alpha) \quad p(\alpha_i | \Gamma_\alpha) \propto \frac{1}{|\Gamma_\alpha|^{1/2}} \exp \left[-\frac{1}{2} \alpha_i^T \Gamma_\alpha^{-1} \alpha_i \right] \quad (2)$$

From Eq.2, we can notice that the distribution of α_i is completely described by Γ_α . We can thus rephrase the goal of the atlas construction as estimating the template complex \mathbf{T} and the covariance matrix Γ_α knowing the shape complexes of the population $\{S_i\}$ or, more formally, maximizing their joint posterior distribution:

$$\{\mathbf{T}^*, \Gamma_\alpha^*\} = \operatorname{argmax}_{\mathbf{T}, \Gamma_\alpha} p(\mathbf{T}, \Gamma_\alpha | \{S_i\}) \quad (3)$$

This maximization is constrained by the fact that the deformed template $\phi_i(\mathbf{T})$ should resemble to the shape complex S_i i.e., the residual ϵ_i should be small. This is a common problem in statistical learning and it is usually tackled by assuming that the residual follows a Gaussian distribution centred at zero. Maximizing its likelihood is equivalent to minimizing its squared norm. We will start by modelling the structures of the shape complexes S_i and template complex \mathbf{T} with landmarks. Every structure j has a number of landmarks equal to Λ_j and the norm of the difference between two meshes of the same structure is defined as the square root of the sum of squared differences between pairs of landmarks (L^2 -norm, $\|\cdot\|_2$). The likelihood of the residuals modelled with landmarks is thus defined as:

$$\epsilon_{ij} \sim N(0, \sigma_j^2 Id_{\Lambda_j}) \quad p(\epsilon_{ij} | \sigma_j^2) \propto \frac{1}{|\sigma_j^2|^{\Lambda_j/2}} \exp \left[-\frac{1}{2\sigma_j^2} \|S_{ij} - \phi_i(T_j)\|_2^2 \right] \quad (4)$$

In section 2.3, we will make clear how to adapt this equation to varifolds by changing only the norm $\|\cdot\|$ and the definition of the parameter Λ_j . The variance of the noise of structure j depends only on the parameter σ_j^2 which is structure-dependent and modelled as a random variable.

It is important to notice that in Eq.1 a value of α_i equal to zero means no deformation. The mean of α_i is set to zero because the template is supposed to represent the average of the shape complexes (i.e. $S_{ij} \sim N(\phi_i(T_j), \sigma_j^2 Id_{\Lambda_j})$). In other words, using the terminology of mixed models, the deformation parameters α_i are the random effects associated to the fixed effect \mathbf{T} which is the expected value of the shape complexes.

Assuming independence between all random variables and considering α_i as an unobserved nuisance variable, we can rewrite Eq.3 as:

$$\{\mathbf{T}^*, \Gamma_\alpha^*, \sigma_j^{2*}\} = \operatorname{argmax}_{\mathbf{T}, \Gamma_\alpha, \sigma_j^2} \left[\prod_i^N \prod_j^M \int p(T_j, \Gamma_\alpha, \sigma_j^2, \alpha_i, S_{ij}) d\alpha_i \right] \quad (5)$$

The maximization of Eq.5 is not tractable analytically. A possible solution is to employ an iterative algorithm like the EM (Expectation Maximization) and to approximate the conditional distribution of the E step with a Dirac distribution at its mode. Let $\Theta = \{\Gamma_\alpha, \mathbf{T}, \{\sigma_j^2\}\}$ be the parameters of interest and $\mathbf{Z} = \{\{\alpha_i\}\}$ the unobserved nuisance variables, it results:

$$\{\Theta^*, \mathbf{Z}^*\} = \operatorname{argmin}_{\Theta, \mathbf{Z}} -\log [p(\{S_i\} / \mathbf{Z}, \Theta)] - \log [p(\mathbf{Z} / \Theta)] - \log [p(\Theta)] \quad (6)$$

This algorithm is an approximation of the MAP estimator which does not converge when applied to a dataset with low signal-to-noise ratio (SNR), as demonstrated in Allasonnière et al. (2007). In order to ensure the convergence of the iterative estimation scheme one may use sampling algorithms like MCMC (Markov Chain Monte Carlo) (Allasonnière et al., 2010). Unfortunately, these methods require a great computational load and execution time.

Not using priors for σ_j^2 and Γ_α is equivalent to a ML estimation and can produce degenerate estimates. Instead, as demonstrated in Allasonnière et al. (2007), the introduction of inverse Wishart distributions makes it possible to obtain good estimates even when the number of subjects is small. They regularize the estimates of σ_j^2 and Γ_α . As in Allasonnière et al. (2007), the probability density functions employed in this paper are adapted versions of the standard inverse Wishart distributions:

$$\sigma_j^2 \sim \mathcal{W}^{-1}(P_j, w_j) \quad p(\sigma_j^2; P_j, w_j) \propto (\sigma_j^2)^{-\frac{w_j}{2}} \exp\left[-\frac{1}{2} \frac{w_j P_j}{\sigma_j^2}\right] \quad (7)$$

$$\Gamma_\alpha \sim \mathcal{W}^{-1}(P_\alpha, w_\alpha) \quad p(\Gamma_\alpha; P_\alpha, w_\alpha) \propto |\Gamma_\alpha|^{-\frac{w_\alpha}{2}} \exp\left[-\frac{1}{2} w_\alpha \text{Tr}(P_\alpha^T \Gamma_\alpha^{-1})\right] \quad (8)$$

The scalars w_j , P_j and w_α are strictly positive and P_α is a positive symmetric matrix. The entire probabilistic model is graphically described in Fig.1. Using these priors for σ_j^2 and Γ_α , and assuming that the template \mathbf{T} has a non-informative prior distribution, Eq.6 becomes:

$$\sum_{j=1}^M \sum_{i=1}^N \frac{1}{2\sigma_j^2} \left(\|S_{ij} - \phi_i(T_j)\|_2^2 + \frac{P_j w_j}{N} \right) + \sum_{j=1}^M \frac{1}{2} (w_j + \Lambda_j N) \log(\sigma_j^2) + \frac{1}{2} \sum_{i=1}^N (\alpha_i)^T (\Gamma_\alpha)^{-1} \alpha_i + \frac{1}{2} (w_\alpha + N) \log(|\Gamma_\alpha|) + \frac{w_\alpha}{2} \text{tr}((\Gamma_\alpha)^{-1} P_\alpha) \quad (9)$$

150 The framed terms refer respectively to the data-terms and to the regularity of the deformations. If we fix both σ_j^2 and Γ_α and we set $w_j = 0$, their sum is the cost function minimised in standard atlas construction procedures (Avants and Gee, 2004; Durrleman et al., 2014) where a statistical framework is not employed. This minimisation is often referred to as the computation of the Fréchet mean. It can be noticed that σ_j^2 weights the contribution of structure j as mentioned in the Introduction.

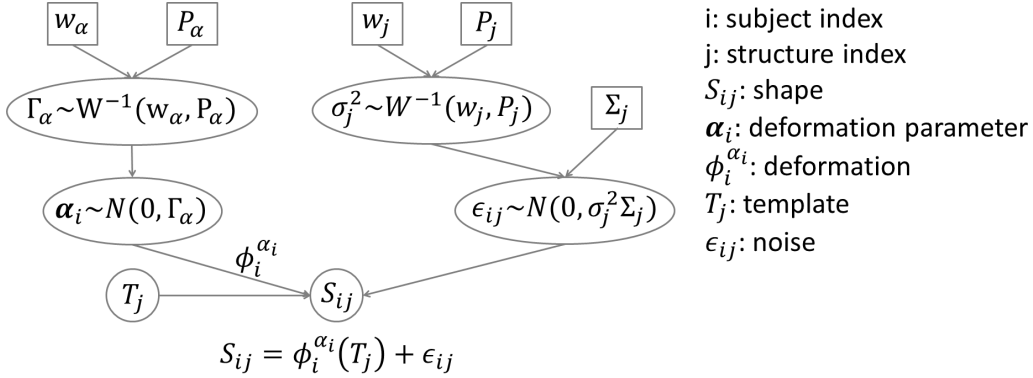


Figure 1: Graphical description of the proposed model. The variables in square boxes and circles are constants and random variables respectively. The shape S_{ij} is generated by a deformation ϕ_i of the template T_j parametrised by α_i . The noise ϵ_{ij} is parametrised by a scalar value σ_j^2 and a constant matrix Σ_j which depends on the computational model employed for S_{ij} .

155 2.2. Multi-population atlas construction

In the previous subsection we assumed that all subjects belonged to the same group. In a clinical application, a subject can be characterised by a label which can be for instance his/her clinical status. Depending on this label, subjects can be divided into different groups. Here we propose to take advantage of this information computing a multi-population atlas construction.

160 Let \mathbf{L} be the set of labels that can be given to a subject shape complex S_i . We assume that \mathbf{L} has only two elements for simplicity: $\mathbf{L} = \{c, p\}$, they can be interpreted as “control” and “patient”. The extension to more than two labels is straightforward and it will not be considered here. With this notation, we can rewrite Eq.1 for both groups as:

$$S_i^c = \phi_i^c(\mathbf{T}) + \epsilon_i^c \quad S_i^p = \phi_i^p(\mathbf{T}) + \epsilon_i^p \quad (10)$$

where we assume that the template complex T is shared between the two populations. Instead, the deformation parameters α_i are divided into two classes which can be modelled as distinct Gaussian distributions:

$$\alpha_i^c \sim N(\mu^c, \Gamma_\alpha^c) \quad \alpha_i^p \sim N(\mu^p, \Gamma_\alpha^p) \quad (11)$$

We can quantitatively compare α_i^c with α_i^p since we assume that they have the same dimensionality. To avoid identifiability issues, we force the sum of the means μ^c and μ^p to equal zero: $\mu^c = \mu$ and $\mu^p = -\mu$. This constraint forces the estimated template complex to lie halfway between the two groups' averages.

We can then assume either that the two groups share the same variability ($\Gamma_\alpha^c = \Gamma_\alpha^p = \Gamma_\alpha$) or that their within-group variations are different ($\Gamma_\alpha^c \neq \Gamma_\alpha^p$). The cost functions for both cases can be found in Appendix A. Using the second assumption, we could test the hypothesis of equality between the two covariance matrices or quantitatively compare their respective modes. On the contrary, if we ignored the labels of the subjects computing a single-population atlas, we would implicitly assume that the two different groups belong to the same population and that they share the same variability. Furthermore, we could also adopt another strategy by constructing an atlas for each group and obtaining therefore two different template complexes. However, in this case, we could neither quantitatively compare the deformation parameters of the two groups nor directly compare the resulting template complexes.

The multi-population extension is computationally feasible only when using the proposed Bayesian framework. As previously shown, we use the log-likelihood of the deformation parameters as regularity term of the deformations. Dividing the deformation parameters into two different groups entails the addition of another regularity term into the cost function (see Appendix A). This new term should be balanced by a trade-off parameter tuned using cross-validation. If we increased the number of groups, we would also augment the number of regularity terms and balancing values to tune. With the proposed statistical setting, there is no need to tune these weights since they are automatically taken into account in the estimates of the covariance matrices.

The proposed statistical framework, for both single- and multi-population, is general and it could be applied to any shape model and parametric diffeomorphic setting, provided that one defines a probability density function for the noise. Modelling shapes with landmarks simplifies the computations and it also eases the definition of random variables. Unfortunately, the choice of corresponding points between structures is not always an easy task and in some cases, like for fiber bundles in the brain, it is almost impossible. It is better to opt for a correspondence-free metric like the one based on the framework of varifolds (Charon and Trounev, 2013).

2.3. Varifolds

The framework of varifolds is an extension of the one of currents (Vaillant and Glaunès, 2005) and it can be used for both surfaces and curves. It does not require point-correspondence between structures or curve-correspondence between ensemble of tracts. This means that the same structure may have a different number of vertices among subjects. It is robust to mesh imperfections such as holes, spikes or tract interruptions. Moreover, it does not require a consistent orientation of the normal or tangent vectors among the population since the distance is invariant to a change of orientation of some normals or tangent vectors. Another important characteristic is that varifolds prevent what is known as the ‘‘currents cancelling effect’’. This happens when two surface/curve elements with opposite orientation cancel reciprocally their effect on the estimation of the template. For a more detailed discussion about varifolds the user is referred to Durrleman et al. (2014); Charon and Trounev (2013).

Let X and Y be two curves (oriented or unoriented rectifiable curves), they can be modelled as varifolds based on:

$$V_X(\omega) = \int_X \omega(x, \overleftarrow{\beta}(x)) |\beta(x)|_2 dx \quad V_Y(\omega) = \int_Y \omega(y, \overleftarrow{\gamma}(y)) |\gamma(y)|_2 dy \quad (12)$$

where $V_X(\omega)$ and $V_Y(\omega)$ are scalars, $\beta(p)$ and $\gamma(y)$ are the tangents of X and Y at the point x and y respectively and $\overleftarrow{\beta}_p$ is the unoriented unit vector with the same direction of β_p (respectively $\overleftarrow{\gamma}_q$ for γ_q). The set of unoriented unit vectors $\overleftarrow{\beta} \in \mathbb{RP}^2$ can be formally defined as the quotient of the unit sphere in \mathbb{R}^3 by the two elements group $\{\pm \text{Id}_3\}$. Instead, the test field $\omega \in W$ is a function of both the points in the ambient space \mathbb{R}^3 and the unoriented unit tangent vectors in \mathbb{RP}^2 .

The curves X and Y can be seen as polygonal lines of P and Q segments respectively. Every segment of X is completely described by its center point $x_p \in \mathbb{R}^3$ and tangent vector $\beta_p \in \mathbb{R}^3$ centred at x_p (respectively $y_q \in \mathbb{R}^3$ and

$\gamma_q \in \mathbb{R}^3$ for Y). Let c_p be the length of the tangent vector β_p (respectively d_q for γ_q), one can approximate Eq.12 with:
 $V_X(\omega) \approx \sum_{p=1}^P \omega(x_p, \vec{\beta}_p) c_p$ and $V_Y(\omega) \approx \sum_{q=1}^Q \omega(y_q, \vec{\gamma}_q) d_q$. It is important to notice that, even if this approximation seems reasonable, it is still an open question how to ensure its convergence when the number of segments tends to infinite, i.e., with a more accurate sampling of the curve (Charlier et al., 2014).

Furthermore, Eq.12 shows that the space of varifolds W^* is a linear functional from W to \mathbb{R} . This makes W^* a vector space and therefore the union of meshes is equal to a sum in the space of varifolds. But, contrary to currents, inverting the orientation of a mesh does not change the varifold representation. One chooses W to be a separable Reproducible Kernel Hilbert Space (RKHS) on $\mathbb{R}^3 \times \mathbb{R}\mathbb{P}^2$ in order to have an explicit definition of distance between two meshes modelled as varifolds. More precisely, since W is a product space, one defines its kernel K_W as the tensor product between two kernels k_x , Gaussian, and k_β , Cauchy-Binet, defined on \mathbb{R}^3 and on $\mathbb{R}\mathbb{P}^2$ respectively. Thus, the inner product between X and Y modelled as varifolds is defined as:

$$\begin{aligned} \langle V_X, V_Y \rangle_{W^*} &= \sum_{p=1}^P \sum_{q=1}^Q k_x(x_p, y_q) k_\beta(\vec{\beta}_p, \vec{\gamma}_q) c_p d_q \\ &= \sum_{p=1}^P \sum_{q=1}^Q \exp\left(-\frac{\|x_p - y_q\|^2}{\lambda_W^2}\right) \left(\frac{\beta_p^T \gamma_q}{c_p d_q}\right)^2 c_p d_q \end{aligned} \quad (13)$$

The distance between V_X and V_Y is therefore: $\|V_X - V_Y\|_{W^*}^2 = \langle V_X, V_X \rangle_{W^*} + \langle V_Y, V_Y \rangle_{W^*} - 2 \langle V_X, V_Y \rangle_{W^*}$. This metric is completely parametrized by the standard deviation λ_W of the Gaussian kernel k_x which is fixed by the user. This framework can be easily extended also to surface meshes by replacing x_p (y_q) and β_p (γ_q) with the centres and normals of the faces of the mesh, as shown in Charon and Trouvé (2013). The computations of the gradient with respect to the points of the mesh can be found in Charon and Trouvé (2013); Durrleman et al. (2014).

Varifolds random variable. In Eq.4, we defined the residual ϵ as a multivariate Gaussian random variable on the finite-dimensional Euclidean space of landmarks. Here we extend this definition to the framework of varifolds following the same line of reasoning presented in Durrleman (2010) for the framework of currents. We define a random Gaussian varifold as a linear map between every test field $\omega \in W$ to a real random Gaussian variable $G(\omega)$ such that, given two test fields ω_1 and ω_2 , $E[G(\omega_1)] = 0$ and $E[G(\omega_1), G(\omega_2)] = \langle \omega_1, \omega_2 \rangle_W$. This shows that the kernel K_W of the space W completely defines the covariance matrix of the Gaussian varifold. However, since it is infinite-dimensional, it has no probability density function and therefore we can not simulate instances of it. To tackle this problem, we define for each structure j a finite-dimensional space $W_{\Lambda_j}^*$ on which we project the template T_j and all the shapes $\{S_{ij}\}_{i=1\dots N}$ modelled as varifolds. This finite dimensional space $W_{\Lambda_j}^*$ is defined as the span of a set of delta Dirac varifolds: $\text{Span}\{\delta_{(x_u, \vec{\beta}_k)}\}$ where both the points $\{x_u\}$ and the unoriented unit vectors $\{\vec{\beta}_k\}$ are constrained to belong to two predefined grids, respectively Υ_x and Υ_β . The first one is a linearly spaced grid in the ambient space and Υ_β is a regular sampling of the half unit sphere in \mathbb{R}^3 . The number of points of the two grids is respectively Λ_x^j and Λ_β and Λ_j is their product where the index j refers to the j -th structure.

For every finite-dimensional space $W_{\Lambda_j}^*$, it is possible to define a block matrix K_{W, Λ_j} whose blocks are the RKHS kernel $K_{W, \Lambda_j} \left((x_u, \vec{\beta}_k), (\cdot, \cdot) \right)_{\substack{u=1, \dots, \Lambda_x^j \\ k=1, \dots, \Lambda_\beta}}$ between every possible combination of the couples $\{x_u, \vec{\beta}_k\}$. The matrix K_{W, Λ_j} has a dimension of $[\Lambda_j \times \Lambda_j]$. To project a delta Dirac varifold $\delta_{(y, \vec{\alpha})} |\alpha|$ onto $W_{\Lambda_j}^*$, using for instance the closest neighbour projection, we look for the closest point to y of the grid Υ_x and for the closest direction to $\vec{\alpha}$ among the ones given by Υ_β and we assign the scalar $|\alpha|$ to that particular couple of grid points. The projection is therefore completely parametrised by the vector $\{|\alpha|\}$ of size Λ_j . Using this scheme, the squared norm of a projected varifold $\|\sum_{l=1}^L \delta_{(y_l, \vec{\alpha}_l)} |\alpha_l|\|_{W_{\Lambda_j}^*}^2$ is equal to $\sum_{u \in \Upsilon_x} \sum_{p \in \Upsilon_x} \sum_{k \in \Upsilon_\beta} \sum_{q \in \Upsilon_\beta} c_{uk} K_{W, \Lambda_j} \left((x_u, \vec{\beta}_k), (x_p, \vec{\beta}_q) \right) c_{pq}$ where the scalars c_{uk} and c_{pq} refer to the values obtained at the end of the projection of the varifold in $\{x_u, \vec{\beta}_k\}$ and $\{x_p, \vec{\beta}_q\}$ respectively.

Scaling the matrix K_{W, Λ_j} with a scalar σ_j^2 , we can define the likelihood of the residuals modelled as varifolds in the following way:

$$p(\epsilon_{ij}|\sigma_j^2) \propto \frac{1}{|\sigma_j^{2|\Lambda_j|/2}} \exp \left[-\frac{1}{2\sigma_j^2} \|S_{ij} - \phi_i(T_j)\|_{W_{\Lambda_j}^*}^2 \right] \quad (14)$$

where it is important to notice that $\mathbf{K}_{W_{\Lambda_j}}$ is a constant matrix that is computed only once at the beginning of the atlas construction. This equation has the same formulation of the likelihood of the residuals modelled as landmarks in Eq.4. The only two differences are the definitions of the norm and of the parameter Λ_j . These are also the only changes to apply to the cost function in Eq.9 when using the framework of varifolds instead of landmarks.

In practice, due to the heavy computational load and time required to calculate $\|\cdot\|_{W_{\Lambda_j}^*}^2$, we choose to compute the exact expression of the norm $\|(\cdot)\|_{W_{\Lambda_j}^*}^2$ using a fast GPU (CUDA) implementation. Moreover, we define Υ_x^j as a regular linearly-spaced grid containing both shapes and template of structure j . The distance between every couple of points is given by λ_W . The resulting number of points Λ_x^j is used to compute $\Lambda_j = \Lambda_x^j$. In this way, Λ_j depends only on λ_W which is fixed by the user and it is related to the distance at which two shapes can be compared and at the same time to the degree of detail of the anatomical features one wants to detect. In Sec.3.3, we show that the number of grid points, if above a reasonable threshold, does not affect the results and the proposed choice seems a good one.

2.4. Diffeomorphic Transformations

We define here how to compute the diffeomorphic deformations of the template complex. Our approach relies on the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework based on the control point formulation presented in Durrleman et al. (2011b). Template transformations are built by integrating a time-varying vector field $v_t(x)$ over $t \in [0, 1]$ where $v_t(x)$ represents the instantaneous velocity of every point x belonging to the ambient space at time t . Calling $\phi_t(x)$ the position of a point at time t which was located in x at time $t = 0$, its evolution is given by: $\frac{\partial \phi_t(x)}{\partial t} = v_t(\phi_t(x))$ with $\phi_0(x) = x$, which produces a flow of deformations $\{\phi_t\}_{t \in [0,1]}$. Furthermore, v_t belongs to a RKHS with Gaussian kernel K_V . In Glaunès (2005) it is shown that if such a vector field is square integrable in the interval $[0,1]$, then every deformation of the flow $\{\phi_t\}_{t \in [0,1]}$ is a diffeomorphism. Here, we define v_t using a dynamical system of C_p control points $\mathbf{c} = \{c_k \in \mathbb{R}^3\}$ scattered in the ambient space and a set of time-varying vectors called momenta $\alpha = \{\alpha_k \in \mathbb{R}^3\}$ linked to each control point. The velocity of every point in the space is given by:

$$\dot{x}(t) = v_t(\phi_t(x)) = \sum_{p=1}^{C_p} K_V(x(t), c_p(t)) \alpha_p(t) \quad \phi_0(x) = x(0) = x \quad (15)$$

where $\phi_t(x) = x(t)$. With this definition, every ϕ_t is a diffeomorphism provided that all $\alpha(t)$ are square integrable for every t . The evolution of a point depends only on its initial position and on the system $\mathbf{L}(t) = \{\mathbf{c}(t), \alpha(t)\}$. In an atlas construction, the deformation of the template complex is the last one (ϕ_1) of the flow of diffeomorphisms $\{\phi_t\}_{t \in [0,1]}$. Among all the possible paths connecting ϕ_0 to ϕ_1 we use the geodesic one, which means the one that minimizes the total kinetic energy along the path:

$$\int_0^1 \|v_t\|_V^2 dt = \int_0^1 \sum_{k=1}^{C_p} \sum_{p=1}^{C_p} \alpha_k(t)^T K_V(c_k(t), c_p(t)) \alpha_p(t) dt \quad (16)$$

It has been demonstrated in Durrleman et al. (2014) that the extremal paths are such that the system $\mathbf{L}(t)$ satisfies:

$$\begin{cases} \dot{c}_k(t) = \sum_{p=1}^{C_p} K(c_k(t), c_p(t)) \alpha_p(t) \\ \dot{\alpha}_k(t) = - \sum_{p=1}^{C_p} \alpha_k(t)^T \alpha_p(t) \nabla_1 K(c_k(t), c_p(t)) \end{cases} \quad (17)$$

Denoting $\mathbf{L}(0) = \mathbf{L}_0 = \{\mathbf{c}(0), \alpha(0)\}$ the initial condition of the system, the previous set of ODEs can be rewritten as: $\dot{\mathbf{L}}(t) = F[\mathbf{L}(t)]$ with $\mathbf{L}(0) = \mathbf{L}_0$. This equation shows that the entire flow of diffeomorphisms, and more precisely the last one, is completely determined by the initial state of the system \mathbf{L}_0 . Thus, given $\mathbf{c}(0)$ and $\alpha(0)$, one first integrates Eq.17 obtaining the values of $\mathbf{c}(t)$ and $\alpha(t)$ at every time t . Then, every point x in the ambient space is deformed by integrating Eq. 15.

In geometrical terms, the value L_0 defines the initial velocity of the geodesic path in the tangent space of our group of diffeomorphisms at identity. The parameters $\{c(0), \alpha(0)\}$ can be used to perform the so-called tangent-space statistics (Vaillant et al., 2004).

2.5. Optimization procedure

This diffeomorphic setting can be easily included into the proposed Bayesian framework. We choose to employ a unique set of control points $c(t)$ for all subjects (and groups) and N subject-specific sets of momenta $\{\alpha_i(t)\}_{i=1,\dots,N}$. It is important to notice that initial momenta of different groups are defined on the same control points and therefore they can be compared among each other. Initial control points and momenta $\{c(0), \alpha_i(0)\} = \{c_0, \alpha_{i0}\}$ represent the deformation parameters that warp the template complex towards the shape complex of subject i . Controls points c_0 are considered as parameters of our model since they are a fixed effect of the entire population. Initial momenta are instead considered as hidden variables (i.e., $\mathbf{Z} = \{\{\alpha_{i0}\}_{i=1,\dots,N}\}$). Assuming that all random variables are independent, the expression of the cost function in Eq.9 does not change except for the second framed term where we simply substitute α_i with α_{i0} : $\frac{1}{2} \sum_{i=1}^N (\alpha_{i0})^T (\Gamma_\alpha)^{-1} \alpha_{i0}$.

The variables \mathbf{T} , $\{\alpha_0^i\}$ and c_0 are minimised using an accelerated version of the line search gradient descent method based on the Nesterov's scheme (Nesterov, 1983). Instead, due to the use of conjugate priors, the optimal values for $\{\sigma_j^2\}$ and Γ_α can be computed in closed form:

$$\hat{\Gamma}_\alpha = \frac{\sum_{i=1}^N [(\alpha_{i0})(\alpha_{i0})^T] + w_\alpha P_\alpha^T}{(w_\alpha + N)} \quad \hat{\sigma}_j^2 = \frac{\sum_{i=1}^N \|S_{ij} - \phi_i(T_j)\|_{W_{\Lambda_j}^*}^2 + w_j P_j}{(w_j + N\Lambda_j)} \quad (18)$$

The first parameter $\hat{\Gamma}_\alpha$ is equal to a weighted sum between the sample covariance matrix of the initial momenta α_{i0} and the prior P_α . A good choice for the prior seems to be: $P_\alpha = K_V^{-1}$, where K_V is a block matrix whose blocks are 3D Gaussian kernels between two different control points. This is the kernel of the RKHS to which the velocity field v_t belongs to. If the number of subjects N is small with respect to w_α then $\hat{\Gamma}_\alpha \propto K_V^{-1}$ and this means that the deformation regularity in Eq.9 becomes $\sum_{i=1}^N (\alpha_{i0})^T K_V \alpha_{i0}$, which is the sum of the geodesic distances between the template complex and all its transformations. This is the deformation regularity term used in previous atlas constructions not based on a statistical setting (Durrleman et al., 2014). Moreover, the use of this prior makes the inversion of Γ_α possible even when the number of subjects N is smaller than the number of deformation parameters.

The second parameter $\hat{\sigma}_j^2$ is equal to a weighted sum between the data-term of the j -th structure and the prior. It balances the importance of structure j with respect to the other structures and at the same time it is a trade-off between the data-term and the other parts of the cost function (Eq.9). The automatic estimate of this parameter is rather important as shown in Sec.3.2. The use of an inverse Wishart prior prevents over-fitting. In fact, using a maximum likelihood estimator, the minimization process might focus only on structure k reducing its residuals almost to zero and consequently $\log(\sigma_k^2) \rightarrow -\infty$. This could also result in almost ignoring all the other structures. The role played by the prior is to impose a minimum value to σ_k^2 in order to avoid such a situation.

It is important to notice that the closed-form solutions for both $\hat{\Gamma}_\alpha$ and $\hat{\sigma}_j^2$ in Eq.18 would be different if using the standard Inverse Wishart distributions instead than the ones proposed in Eq.7 and Eq.8. They would not be equal to a weighted average between prior and empirical estimates, making therefore more difficult their interpretation and use.

The gradients of the cost function E (Eq.9) with respect to $\mathbf{T}, \{\alpha_0^i\}$ and c_0 are respectively equal to:

$$\begin{aligned} \nabla_{T_k} E &= \sum_{i=1}^N \frac{1}{2\sigma_k^2} \nabla_{T_k} [D_{ik}] \\ \nabla_{\alpha_{s0}} E &= \sum_{j=1}^M \frac{1}{2\sigma_j^2} \nabla_{\alpha_{s0}} [D_{sj}] + \Gamma_\alpha^{-1} \alpha_{s0} \\ \nabla_{c_0} E &= \sum_{i=1}^N \sum_{j=1}^M \frac{1}{2\sigma_j^2} \nabla_{c_0} [D_{ij}] + \frac{w_\alpha}{2} \nabla_{c_0} [\text{tr}((\Gamma_\alpha)^{-1} P_\alpha)] \end{aligned} \quad (19)$$

where $D_{ij} = \|S_{ij} - \phi_i(T_j)\|_{W_{\lambda_j}}^2$. The differentiation of the prior P_α with respect to \mathbf{c}_0 is not taken into account since its norm is negligible with respect to the one of the data term.

In order to compute the three derivatives of the data-terms D_{ij} , we need first to compute the deformed template complex $\phi_i(\mathbf{T})$ where the deformation ϕ_i is the final diffeomorphism of the flow generated by integrating forward in time Eq.17 and then Eq.15 which can be rewritten in matrix notation as: $\dot{\mathbf{T}}_i(t) = G[\mathbf{T}_i(t), \mathbf{L}_i(t)]$ with $\mathbf{T}_i(0) = \mathbf{T}(0) = \mathbf{T}$. After that, we can compute the gradient of D_{ij} with respect to the deformed template complex $\mathbf{T}_i(1) = \phi_i(\mathbf{T})$ and then bring back this information to $t = 0$ in order to update the *initial* template points \mathbf{T} , control points \mathbf{c}_0 and momenta α_{i0} . This is done by integrating backward in time from $t=1$ to $t=0$ a set of linearised ODEs called adjoint equations like in Durrleman et al. (2014). It results:

$$\begin{aligned}\nabla_{\mathbf{T}} E &= \sum_{i=1}^N \boldsymbol{\theta}_i(0) \\ \nabla_{\alpha_{i0}} E &= \boldsymbol{\xi}_i^\alpha(0) + \Gamma_\alpha^{-1} \alpha_{i0} \\ \nabla_{\mathbf{c}_0} E &= \sum_{i=1}^N \boldsymbol{\xi}_i^c(0)\end{aligned}\tag{20}$$

where the auxiliary variables $\boldsymbol{\xi}_i(t) = \{\boldsymbol{\xi}_i^\alpha(t), \boldsymbol{\xi}_i^c(t)\}$ and $\boldsymbol{\theta}_i(t)$ satisfy the linearised ODEs:

$$\dot{\boldsymbol{\theta}}_i(t) = -(\partial_{\mathbf{T}_i} G[\mathbf{T}_i(t), \mathbf{L}_i(t)])^T \boldsymbol{\theta}_i(t) \quad \boldsymbol{\theta}_i(1) = \frac{1}{2\sigma^2} \nabla_{\mathbf{T}_i(1)} [\mathbf{D}_i]\tag{21}$$

$$\dot{\boldsymbol{\xi}}_i(t) = -(\partial_{\mathbf{L}_i} G[\mathbf{T}_i(t), \mathbf{L}_i(t)])^T \boldsymbol{\theta}_i(t) + d_{\mathbf{L}_i} F[\mathbf{L}_i(t)]^T \boldsymbol{\xi}_i(t) \quad \boldsymbol{\xi}_i(1) = 0\tag{22}$$

ODEs are integrated with the Euler's method using 10 steps. More details can be found in the Appendix of Durrleman et al. (2014). The computations of the gradients for the multi-population extension can be found in Appendix A.

2.6. Template initialisation

Since we use a gradient descent scheme to minimize the cost function, we need an initial template. This should not be the mesh of a subject, as otherwise it would inherit all the imperfections due to segmentation errors and data noise, biasing consequently the atlas construction.

For surfaces, we use a regular sphere. It is first centred with respect to the ensemble of subject meshes. Then, it is scaled to an ellipsoid using the three main modes of variability of the meshes considered as points cloud.

For bundles of curves, we propose instead to initialise the template as the ‘‘most relevant’’ curves in the population. We first gather the curves F of every subject bundle into a raw initial template $T = \cup_{h=1}^P F_h$, where P is the total number of curves. Then, we select the set of H curves which most resembles the raw template T . Calling $\{F\}_k$ one of the possible $\binom{P}{H}$ combinations of H curves out of P , one writes:

$$\operatorname{argmin}_k \|T - \{F\}_k\|_{W^*}^2 \quad k \in \binom{P}{H}\tag{23}$$

In many applications the number of curves of the raw template can be huge, thus we have developed a greedy algorithm which does not test all the possible combinations of curves but it is based on an iterative scheme. At each step, we divide T in subsets of Z curves, we look for the best one and we remove it from the raw template T saving it in a new template. This process is repeated until the new template has H curves, where H is the average number of curves in the population and Z is a submultiple of H . At each iteration, the best subset of Z curves is selected minimising:

$$\operatorname{argmin}_k \|T - \sum_{h=(kZ)+1}^{(k+1)Z} F_h\|_{W^*}^2 \quad k = 0, \dots, (H/Z) - 1\tag{24}$$

after having randomized the order of the curves in T . The new template is the initial template used in the atlas construction. In Sec.3.4, we will show that the proposed initialisation method for templates of fiber bundles is stable and that it produces similar results even when varying the value of Z .

Since the number of points per curve and the number of curves in the bundles can be different among the population, we have implemented this algorithm with the metric of varifolds, but any metric could be used.

An example of template initialisation applied to three brain fiber bundles is shown in Fig.2. The last row represents the initial templates used in the following experiments.

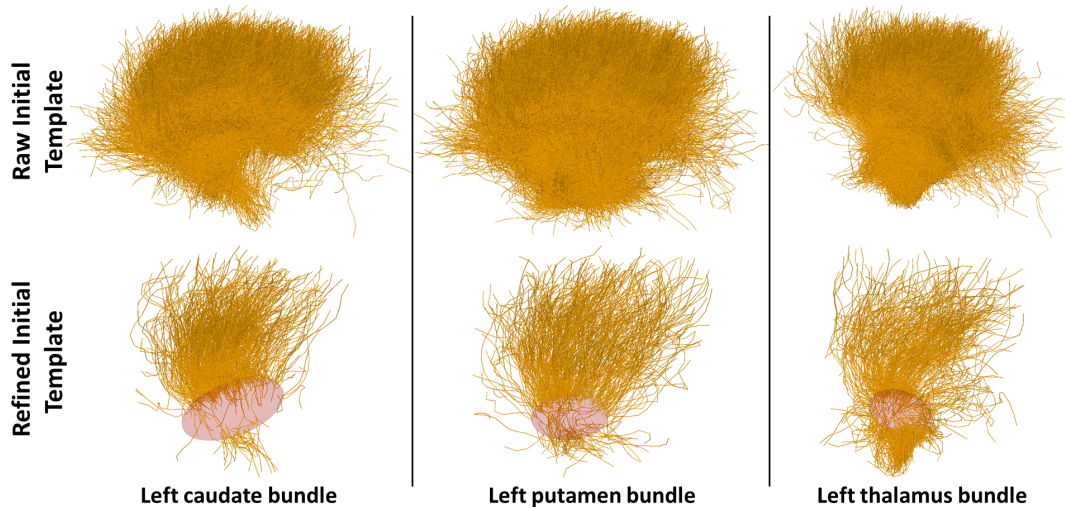


Figure 2: Template initialisation for three brain fiber bundles: left caudate, left putamen and left thalamus bundle. The top row contains the three raw initial templates where the fibers of every subject present in the population have been gathered together (see 3.1). The second row presents the initial templates obtained at the end of the proposed initialisation process cut at the intersection with their respective sub-cortical initial templates, transparent and coloured in red.

3. Experiments

In this section, we first present the dataset used in the following experiments. Then, we evaluate the robustness of our Bayesian framework with respect to the values of the hyperparameters, and we compare it with the robustness of a previous atlas procedure without automatic estimate of σ_j^2 . Furthermore, we also analyse the stability of the Bayesian atlas construction, as a function of the number of points of the varifold grid Λ_j and of the initial template of fiber bundles T_j . Eventually, we present how the proposed multi-population atlas construction approach can be used to highlight morphological differences between two groups.

3.1. Materials

The dataset used throughout this article contains 40 subjects: 20 controls and 20 patients subject to Gilles de la Tourette syndrome. This neurodevelopmental disorder is thought to be associated with dysfunctions of the cortico-striato-pallido-thalamic circuits which are composed of sub-cortical structures linked to the cortical surface by fiber bundles. We will investigate the morphological characteristics of these structures which distinguish the group of patients from the one of controls and which might be due to the Gilles de la Tourette syndrome.

For each subject, we consider three sub-cortical structures (left caudate, left putamen and left thalamus) and the fiber bundles connecting them to the left hemisphere of the cortical surface. The three sub-cortical structures are segmented with FSL (Patenaude et al., 2011) from T1-weighted images (3T). Since we want to analyse as a single structure nucleus accumbens and caudate, we merge these two segmentations together. We use the marching cubes algorithm to create the 3D meshes of all structures (AimsMesh function of Brainvisa 4.3.0). Fiber bundles result from a probabilistic tractography method using 8 seeds (Perrin et al., 2005) applied to diffusion-weighted images (50

directions, B-factor=1000) using an analytical Q-Ball model to estimate the local underlying orientation distribution function (ODF) (Descoteaux et al., 2007). Every subject bundle is downsampled by randomly selecting 10% of its fibers and it is then approximated as proposed in Durrleman et al. (2009) using a Matching Pursuit Algorithm. For more details about the acquisition and tractography, the reader is referred to Worbe et al. (2015).

3.2. Robustness with respect to the hyperparameters

We evaluate here the robustness of the proposed algorithm with respect to the hyperparameters $\{w_j, P_j, w_\alpha\}$. We do not consider P_α since, as explained in Sec.2.5, we fix it to $P_\alpha = K_V^{-1}$. We compute 18 different atlases changing every time only one of the hyperparameters and keeping fixed the others at a certain value. Both w_j and w_α are previously normalised in order to use the same range of values, namely $w_\epsilon = w'_\epsilon N \Lambda$ and $w_\alpha = w'_\alpha N$, giving as result:

$$\hat{\Gamma}_\alpha = \frac{\frac{1}{N} \sum_{i=1}^N [(\alpha_{i0})(\alpha_{i0})^T] + w'_\alpha P_\alpha^T}{(1 + w'_\alpha)} \quad \hat{\sigma}_j^2 = \frac{\frac{1}{N\Lambda_j} \sum_{i=1}^N \|S_{ij} - \phi_i(T_j)\|_{W_j}^2 + w'_j P_j}{(1 + w'_j)} \quad (25)$$

In order to understand a plausible range of testing values, we compute the residuals $\frac{1}{N\Lambda_j} \sum_{i=1}^N \|S_{ij} - \phi_i(T_j)\|_{W_j}^2$ using the aforementioned dataset and initial templates and we notice that the maximum value considering either only the sub-cortical structures or only the fiber bundles is never above 10 and 100 respectively. This means that the product $w'_j P_j$ should not surpass this limit, otherwise a reduction of the data-term would impact very little the estimate of $\hat{\sigma}_j^2$ and consequently the minimization of the cost function. Noting that, we choose as fixed value 10^{-2} and as testing range for all hyperparameters the interval $[10^{-2} - 10^3]$ for sub-cortical structures and $[10^{-2} - 10^4]$ for fiber bundles.

In order to test the robustness of the results, we compute the norm of the difference between the resulting template complex of all 18 atlases and a reference template complex. Moreover, we also analyse the norm of the residuals obtained at the end of the atlas constructions.

To show the importance of an automatic estimate of σ_j^2 and Γ_α , we compare our results with the ones obtained using a previous atlas procedure (Fixed) where both σ_j^2 and Γ_α are fixed. In order to be consistent with the proposed Bayesian algorithm, we set $\Gamma_\alpha = K_V^{-1}$, which is the value of the prior P_α , and we let σ_j^2 change along the same range of values as the hyperparameters. This means building 6 other atlases changing only the value of σ_j^2 . In all 24 atlas procedures, the other parameters, like λ_W and λ_V , are kept equal. Results using only left caudates or left caudate bundles of the 20 controls are shown in Fig.3 where we have used as reference template the one obtained using the Fixed method with $\sigma_j^2=10^{-1}$.

In Fig.4 we compare the robustness of the results using simultaneously two structures in each atlas construction: left caudate and left caudate bundle. This means that we need to fix two σ_j^2 values with the Fixed method and 5 hyperparameters in the Bayesian framework (w_α , two w_j and two P_j). Using as range of values $[10^{-2} - 10^3]$, the number of possible combinations for the parameters of the Fixed technique is 36 whereas for the hyperparameters of the Bayesian framework is 7776. This makes infeasible an extensive and complete analysis of the robustness of the results. Nevertheless, we compute 30 atlases (15 Fixed + 15 Bayesian) picking randomly the values of the hyperparameters and σ_j^2 in the range $[10^{-2} - 10^2]$. The employed values for both the hyperparameters and σ_j^2 are listed in Table B.3. The reference template is the one obtained at the end of the 9-th case using the Fixed method.

It is possible to conclude that the Fixed method leads to much more variable results than using Bayesian priors. The choice of the hyperparameters in our Bayesian framework is therefore easier than the one of σ_j^2 . We decide to set them at: $w_j=0.01\Lambda_j N$ and $P_j=\frac{0.05R_j^0}{w_j}$ where R_j^0 is the initial data-term of structure j . This brings to $\sigma_j^2=\frac{R_j+0.05R_j^0}{N\Lambda_j(1+0.01)}$ which means that the minimum value of σ_j^2 is equal to about 5% of $\frac{R_j^0}{N\Lambda_j}$. This choice reduces the risk of over-fitting since a value of R_j smaller than $\frac{0.05R_j^0}{N\Lambda_j}$ would almost not affect the estimate of σ_j^2 . We also fix w_α to 10^{-3} . This makes the contribution of P_α negligible for the estimate of $\hat{\Gamma}_\alpha$, especially when N is big, but it still permits the inversion of $\hat{\Gamma}_\alpha$.

3.3. Robustness with respect to the number of points Λ of the varifold grid

Another parameter of our algorithm is the number of points Λ of the varifold grid where template and shape complexes are projected to. We decide to make it dependent on λ_W , which is the bandwidth of the varifold kernel, as

explained in Sec.2.3. In Fig.5, we show that the results are not so sensitive to a change of the number of grid points and that our choice seems a good one. We compute 9 different atlases using the left-caudate surfaces of the 20 controls. All atlas constructions share the same set of parameters except for Λ which is tested in the range: $[1-10^5]$. As before, we analyse the norm of the difference between a reference template ($\Lambda=1$) and the estimated templates at the end of the atlas procedures. We also show the norm of the residuals and the logarithm of the determinant of the covariance matrix of the initial momenta ($\log(|\Gamma_\alpha|)$). As it is possible to notice, there is a minimum value of grid points (≈ 1000) above which results are very similar and also more satisfactory.

In order to explain these results, it can be noticed that by replacing σ_j^2 with $\hat{\sigma}_j^2$ into Eq.9, one obtains a new data-term which is equal to $(w + \Lambda N) \log(\hat{\sigma}^2)$. When Λ is small, the reduction of the residuals, and consequently of $\log(\hat{\sigma}^2)$, is not enough amplified by the multiplication with Λ to compensate a great increase in the regularity term. Thus, the estimated initial momenta keep a norm close to zero throughout the atlas construction. The deformations of the template are therefore insignificant and the covariance matrix Γ_α depends only on the prior, which explains why $\log(|\Gamma_\alpha|)$ does not vary. The number of grid points obtained with the proposed method is equal to: 3456.

3.4. Robustness with respect to the initial template of fiber bundles

The proposed initialisation procedure for templates of fiber bundles is iterative. At each iteration, a raw template is randomly subdivided in subsets of Z fibers. The best subset is then removed from the raw template and added to the employed initial template, as explained in Sec.2.6. Here, we show that the initial templates obtained using different Z values are very similar to each other. Moreover, we also verify that the results of the atlas construction are rather stable when using different initial templates.

First, we compute 40 initial templates for each fiber bundle considered in this paper: left-caudate bundle (LCD), left-putamen bundle (LPU) and left-thalamus bundle (LTH). We use four different values of Z : 25,50,100 and 150. For each value, we compute 10 different initial templates. Then, we compare all the resulting initial templates T with the one shown in Fig.2, called T_0 . We use two dissimilarity measures. The first one is the varifold norm of the difference between every T and T_0 divided by the varifold norm of T_0 . The second one is the absolute value of the difference between the data-terms of every template $D(T)$ and $D(T_0)$ divided by $D(T_0)$, where $D(T) : \frac{1}{N} \sum_{i=1}^N \|S_i - \phi_i(T)\|_{W^*}^2$. Results are shown in Fig.6. They are very stable and the difference between $D(T)$ and $D(T_0)$ is always very small.

Furthermore, we also compute 4 multi-population atlases with $\Gamma_\alpha^c = \Gamma_\alpha^p$ using the left caudates and left-caudate bundles of all 40 subjects (20 controls and 20 patients). For each atlas, we use a different initial template for the left-caudate bundles, one for each group of Z . We then compare the results of the atlas constructions with respect to the ones of the first atlas, chosen as reference. We use 5 dissimilarity measures shown in Table 1. The subscript 0 refers to the reference atlas. Results are very stable and therefore we can conclude that they are not very sensitive to a change of initial template of fiber bundles.

	Average [%]	Standard deviation [%]
$\frac{\ CP-CP_0\ _2}{\ CP_0\ _2}$	0.50	0.04
$\frac{ D(T^c)-D(T_0^c) }{D(T_0^c)}$	0.51	0.41
$\frac{ D(T^b)-D(T_0^b) }{D(T_0^b)}$	0.41	0.02
$\frac{ \sigma_c^2-\sigma_{c0}^2 }{\sigma_{c0}^2}$	0.51	0.21
$\frac{ \sigma_b^2-\sigma_{b0}^2 }{\sigma_{b0}^2}$	1.50	0.61

Table 1: Comparison between the results of four multi-population atlases computed using left caudate surfaces and left caudate bundles. Each atlas is characterised by a different initial template for the left caudate bundle. The subscript 0 refers to the first atlas, which is used as reference. The variable CP refers to control points, the superscript c to the left caudate surface, the superscript b to the left caudate bundle and $D(T)$ to the data-term.

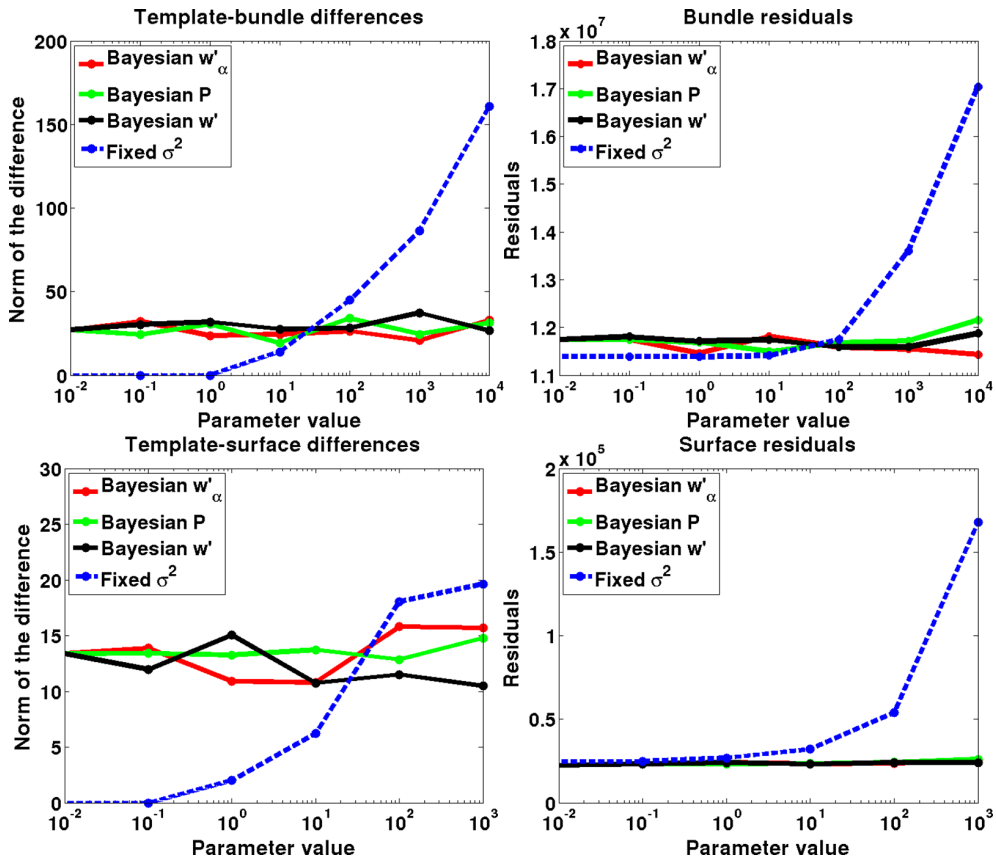


Figure 3: Analysis of the robustness of the results when changing the hyper-parameters of the proposed algorithm (Bayesian) and the value of σ_j^2 in an atlas without automatic estimates (Fixed). Every dot represents the result of an atlas construction. In all Bayesian estimations, it has been changed only one of the hyperparameters, fixing the others to 0.01. “Surfaces” are the left caudates and “Bundles” are the left-caudate bundles. Figures on the left represent the norms of the differences between the templates obtained at the end of the atlas constructions and a reference template (Fixed method with $\sigma_j^2=0.01$). Figures on the right refer to the norms of the residuals obtained at the end of the atlas procedures.

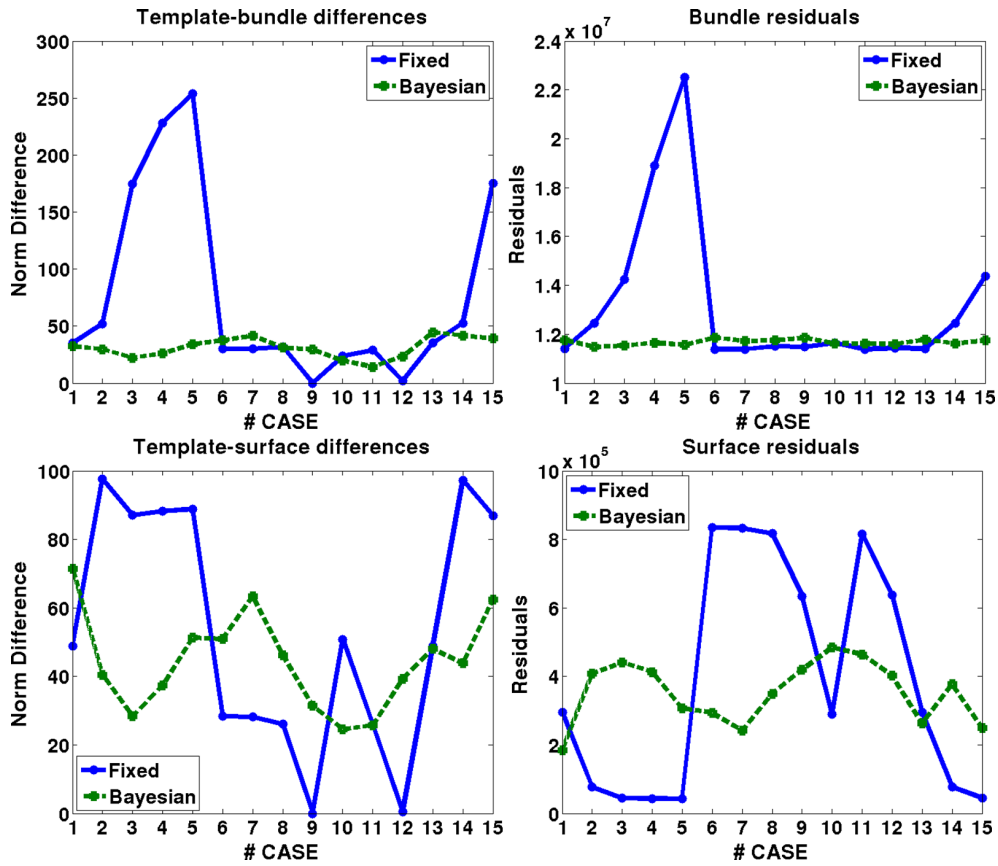


Figure 4: Analysis of the robustness of the results when changing the hyper-parameters of the proposed algorithm (Bayesian) and the value of σ_j^2 in an atlas without automatic estimates (Fixed). Every dot represents the result of an atlas construction. In all atlas estimations the template complex is composed of two structures: left caudate and left caudate bundle. The values of the hyperparameters for the Bayesian estimation and the values of the two σ_j^2 for the Fixed method are listed in Table B.3. Figures on the left represent the norms of the differences between the templates obtained at the end of the atlas constructions and a reference template (9-th case of the Fixed method). Figures on the right refer to the norms of the residuals obtained at the end of the atlas procedures.

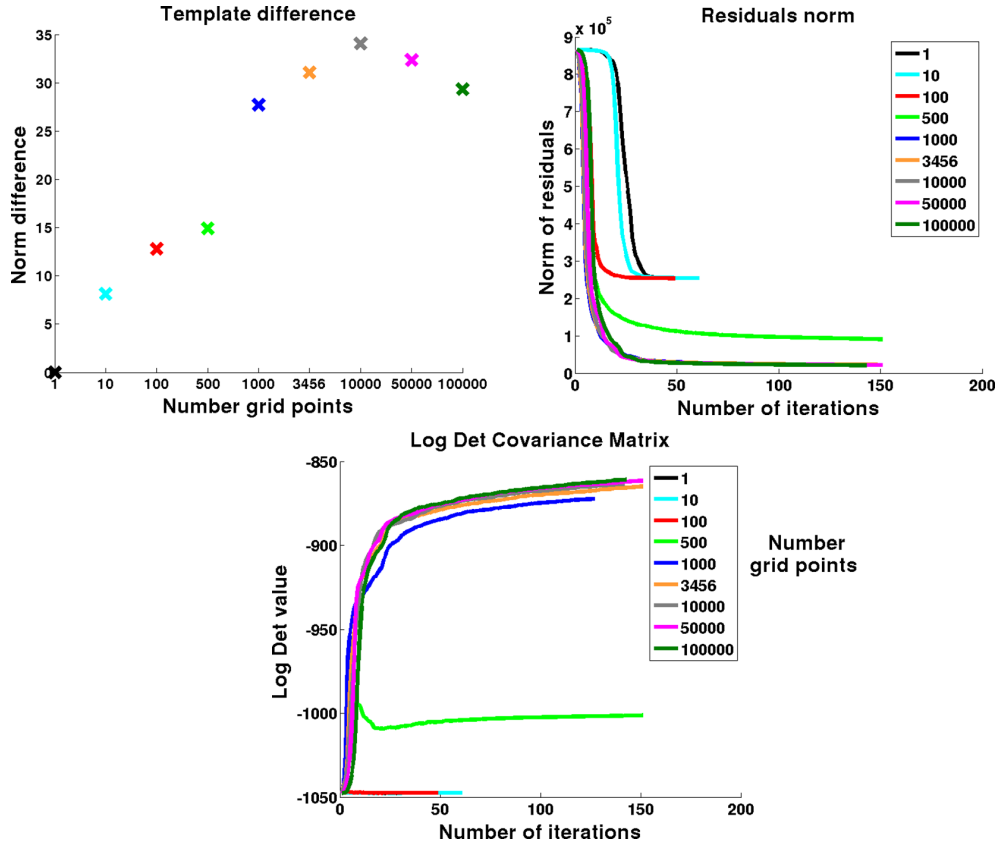


Figure 5: Robustness of the results with respect to the number of points Λ of the varifold grid. We compute 9 atlases using 20 left-caudate surfaces and changing only Λ within the range $[1-10^5]$. Top-left figure represents the norm of the difference between a reference template ($\Lambda=1$) and the templates obtained at the end of the atlas constructions. Top-right figure shows the norm of the residuals during the atlas constructions. Bottom figure is about the evolution of $\log(|\Gamma_\sigma|)$ during the atlas procedures. The maximum number of iterations in every atlas construction is 150. Colors refer to the different grid sizes.

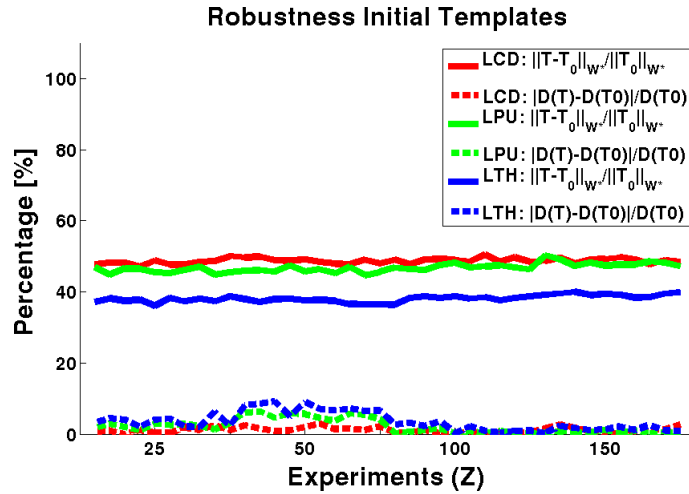


Figure 6: Robustness analysis with respect to the initial template of fiber bundles. We compute 40 initial templates for each fiber bundle employed in this paper: left-caudate bundle (LCD), left-putamen bundle (LPU) and left-thalamus bundle (LTH). Experiments are divided into four groups depending on the value of Z : 25 (10 templates), 50 (10 templates), 100 (10 templates) and 150 (10 templates). We use two dissimilarity measures, as explained in Sec.3.4, highlighted here with a continuous and dotted line respectively and with a different color for each bundle.

3.5. Population differences

In all the following experiments, we use the same parameters. All structures, except for the caudate, have a varifold parameter λ_W equal to 2mm. For the caudate we choose a value of 3mm. About the bandwidth of the diffeomorphic kernel K_V , we use a value of 7mm with a consequent number of control points equal to 1080. The maximum number of iterations is 120 and all the computations are performed on a Intel Xeon, 32 cores, CPU E5-2650, 2.60GHz with a graphic card NVIDIA Quadro 5000. All shape complexes are previously rigidly registered to a reference shape complex.

Multi-population Atlas. ($\Gamma_\alpha^c = \Gamma_\alpha^p = \Gamma_\alpha$) We compute a multi-population atlas using the groups of controls and patients. The template complex and control points are shared between the two groups. Initial momenta are assumed to follow two Gaussian distributions with opposite mean and equal covariance matrix. The overall processing time is about 100 hours based on a single-CUDA implementation.

In Fig.7 we show a scheme of the process. On the left, there are the initial template complex and the initial set of control points, which is initialised as an equally-distanced grid covering the entire ambient space. The distance between couples of initial control points is equal to the bandwidth of the diffeomorphic kernel: λ_V . Momenta α_{i0} are initially set to zero. At the end of the atlas construction we obtain: a final template complex, an updated set of control points, optimised subject-specific initial momenta, the average of the initial momenta μ and their common covariance matrix Γ_α . The final template complex shows the morphological features common to both groups. At the top and at the bottom, we present the average of the initial momenta for each group: μ and $-\mu$. Then, on the right, we show the final common template complex deformed according to μ and $-\mu$. The two complexes represent the anatomical configurations typical of each group and they can be directly compared since they stem from the same final template complex. We take advantage of that by computing the absolute value of the difference between the displacements from the final template complex along the two average directions (i.e. μ and $-\mu$). This is shown in Fig.8 for sub-cortical structures and in Fig.9 for fiber bundles. From the first figure, it is clear that the main differences are in the dorso-lateral part of the three sub-cortical structures, especially for the caudate. About the bundles, the differences are mainly in the central part of the caudate-cortico and putamen-cortico bundles.

A question that naturally arises is whether the shape dissimilarities shown in Fig.8 and Fig.9 are statistically significant. To check that, we compute two statistical tests: a global and a local one. The former is a permutation test which assesses whether the initial momenta of the two groups (controls and patients) come from the same distribution. We employ as statistics the Mahalanobis norm of the difference between the average initial momenta of the two groups. Calling α_1 and α_2 the initial momenta of the two groups at each permutation, $\bar{\alpha}_1$ and $\bar{\alpha}_2$ their respective averages and Γ_{12} their pooled covariance matrix (computed as in Eq.A.3), the employed statistics is equal to: $(\bar{\alpha}_1 - \bar{\alpha}_2)^T \Gamma_{12}^{-1} (\bar{\alpha}_1 - \bar{\alpha}_2)$. We perform 10.000 permutations and the resulting p-value is lower than 0.05. Thus, we can reject the null hypothesis of identical distribution. This shows that the average anatomical complexes of the two groups have a statistically significantly different global shape.

In addition to a multivariate test, we also perform mass univariate tests to produce local maps of significance. First, we compute the initial velocities at each vertex of the template complex (Eq.15) using the estimated initial momenta of both groups. For each vertex and for each group, we obtain 20 3D vectors which are normally distributed. This comes from the fact that initial momenta follow a prior Gaussian distribution. At each vertex, we can then use a Hotelling's two-sample T-squared test for assessing the statistical significance of the difference between the average initial velocities of the two groups. The ensemble of resulting p-values, one for each vertex, is corrected for multiple comparisons using a false discovery rate (FDR) method (Benjamini and Hochberg, 1995). Results are shown in Fig.10 and in Fig.11 for sub-cortical structures and fiber bundles respectively. After the FDR correction, the only structures which still have statistically significant differences at the 5% level are the caudate (dorsal part), the thalamus (ventral and dorsal part) and the caudate bundle (central part). This is probably due to the relatively small number of subjects employed in the experiments.

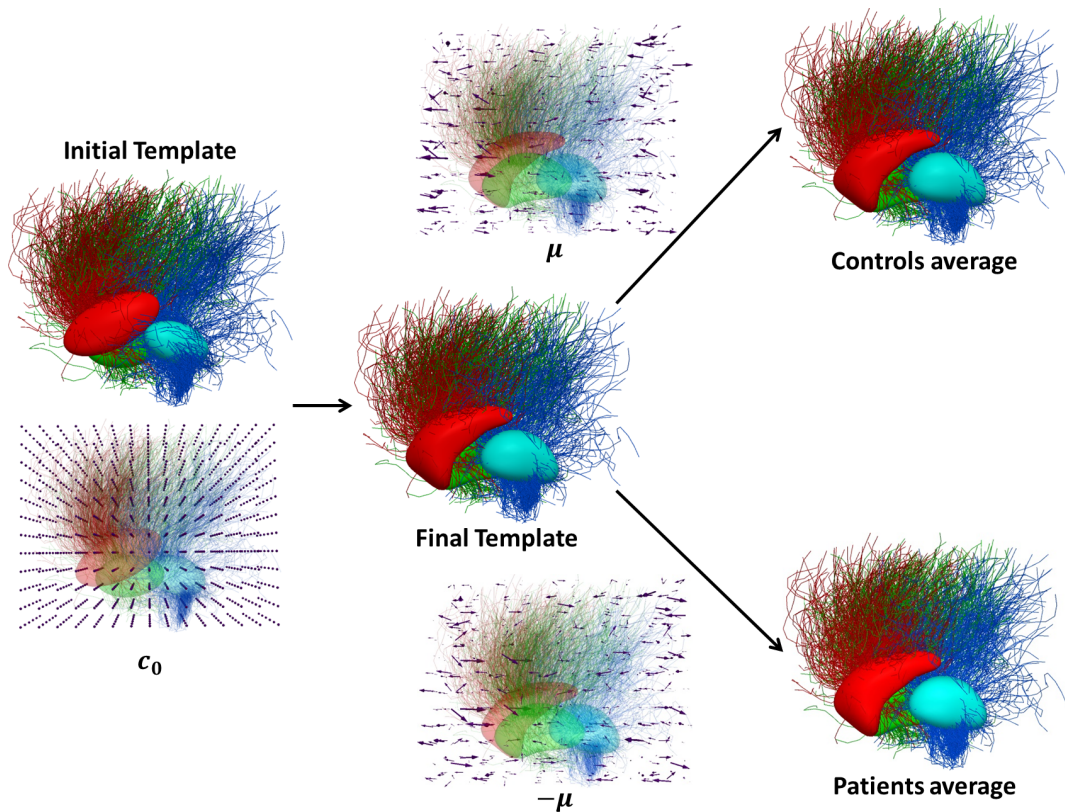


Figure 7: Atlas construction procedure. From left to right, we present first the initial template complex and the initial set of control points. Then, we show the final template complex obtained at the end of the atlas construction. The top and bottom arrows point to the final template complex deformed accordingly to the average initial momenta of controls μ (top) and patients $-\mu$ (bottom). The averages of initial momenta (μ and $-\mu$) are shown respectively above and below the two arrows.

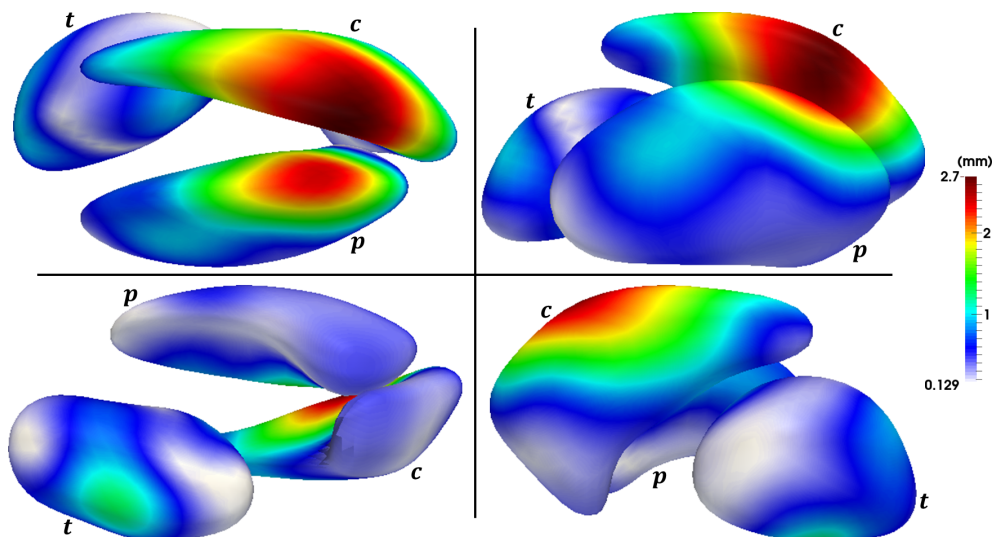


Figure 8: Shape dissimilarities of the sub-cortical structures between the two groups. Colors refer to the absolute value of the difference between the displacements from the final template complex to the average configurations of patients and controls. The two average configurations are obtained by deforming the final template complex with μ and $-\mu$. The four frames represent the same three structures from different points of view. Letters $\{c, p, t\}$ refer to caudate, putamen and thalamus respectively.

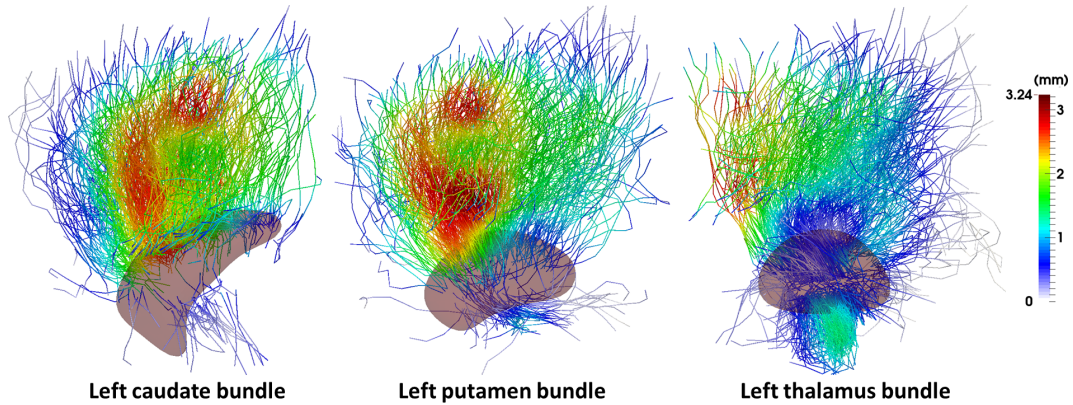


Figure 9: Shape dissimilarities of the fiber bundles between the two groups. Colors refer to the absolute value of the difference between the displacements from the final template complex to the average configurations of patients and controls. The two average configurations are obtained by deforming the final template complex with μ and $-\mu$.

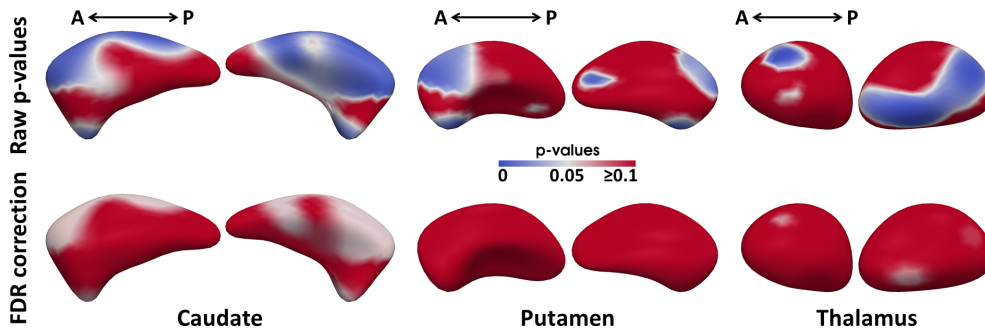


Figure 10: Statistically significant differences between the average initial velocities of the two groups (controls and patients) computed at each vertex of the templates of the sub-cortical structures. P-values result from a Hotelling's two-sample T-squared test. In the second row, p-values are corrected using a false discovery rate (FDR) method. Letters *A* and *P* refer to the Anterior and Posterior part of the brain respectively.

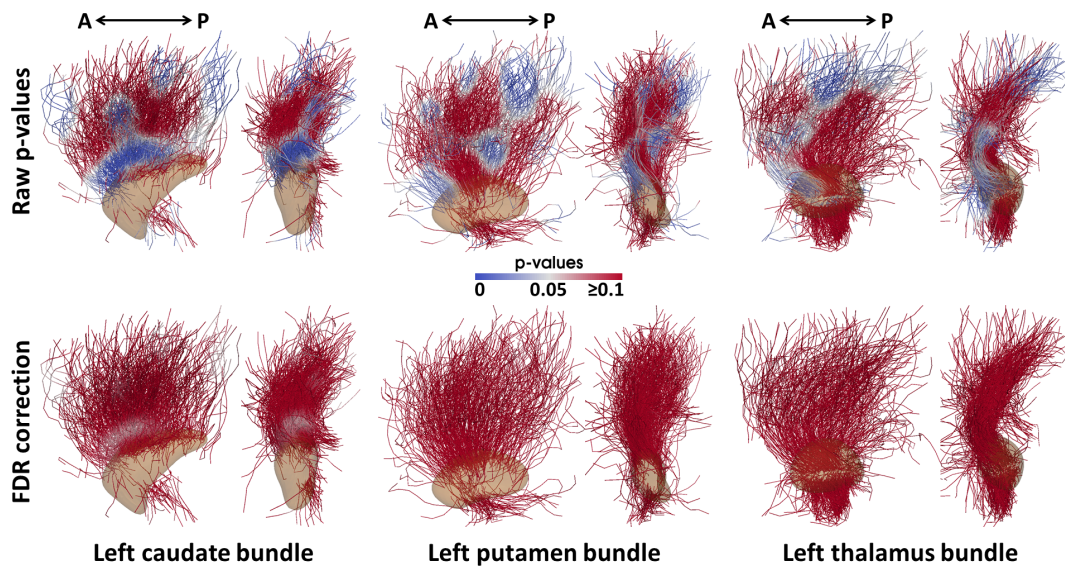


Figure 11: Statistically significant differences between the average initial velocities of the two groups (controls and patients) computed at each vertex of the templates of fiber bundles. P-values result from a Hotelling's two-sample T-squared test. In the second row, p-values are corrected using a false discovery rate (FDR) method. Letters *A* and *P* refer to the Anterior and Posterior part of the brain respectively.

490 *Multi-population Atlas.* ($\Gamma_\alpha^c \neq \Gamma_\alpha^p$) Here, we assume that the initial momenta of patients and controls do not share the same covariance matrix. To test this assumption, we use the test proposed in Srivastava and Yanagihara (2010). The p-value (0.3461) is too big to reject the null hypothesis: $\Gamma_\alpha^c = \Gamma_\alpha^p$. This might be due to the fact that the sample size (20 subjects per group) is too small with respect to the number of deformation parameters (3240). Another possible explanation could be that the two groups share most of their morphological variability which means that the first modes of variation of both covariance matrices are similar. To test this hypothesis, we first compute a Principal Component Analysis (PCA) for both covariance matrices. Then, we calculate the angles between the modes of variation of the two groups. The results for the first 5 modes are shown in Table 2. As it is possible to notice, the first mode of the group of controls is almost parallel to the first mode of the group of patients. This means that they produce similar morphological changes in the template complex. Moreover, they explain almost 35% and 45% of the total variability of their groups respectively (see Fig.12). This might explain why the previous test fails in rejecting the hypothesis of equality between the two covariance matrices.

505 The first mode of both groups indicates the shared morphological variability between controls and patients. The second mode should instead produce shape variations which are more characteristic of each group (angle $\sim 60^\circ$) and which contribute also substantially ($\sim 10\%$) to the total variability. In Fig.13, we show the second mode of variation of both groups. Colors refer to the displacement from the final common template complex, shown in gray in the middle of the figure. It is possible to notice that the dorso-lateral part of the putamen varies almost in the same way between the two groups. Instead, the posterior-lateral part of the caudate and the ventral part of the thalamus present morphological variations which are different between the two groups.

		Patients				
		1	2	3	4	5
Controls	1	25.71	82.49	87.10	87.77	83.32
	2	87.70	60.39	72.57	83.62	80.30
	3	80.60	62.59	67.17	84.39	79.76
	4	86.74	75.15	82.60	88.98	77.35
	5	83.70	81.24	89.51	85.39	83.80

Table 2: Angles in degrees between the first 5 modes of variation of the group of controls and the first five modes of the group of patients. The average angle between all modes is 85.39° .

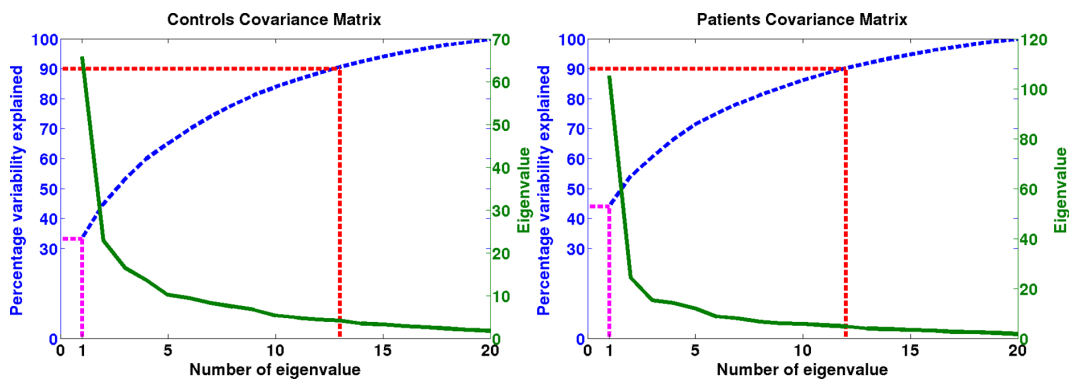


Figure 12: PCA of the two covariance matrices of initial momenta based on the assumption $\Gamma_\alpha^c \neq \Gamma_\alpha^p$. The magenta line represents the percentage of variability explained by the first mode. The red line shows the number of components needed to explain 90% of the total variability. The green line indicates the magnitude of the eigenvalues and the blue line shows the explained variability using all the previous components.

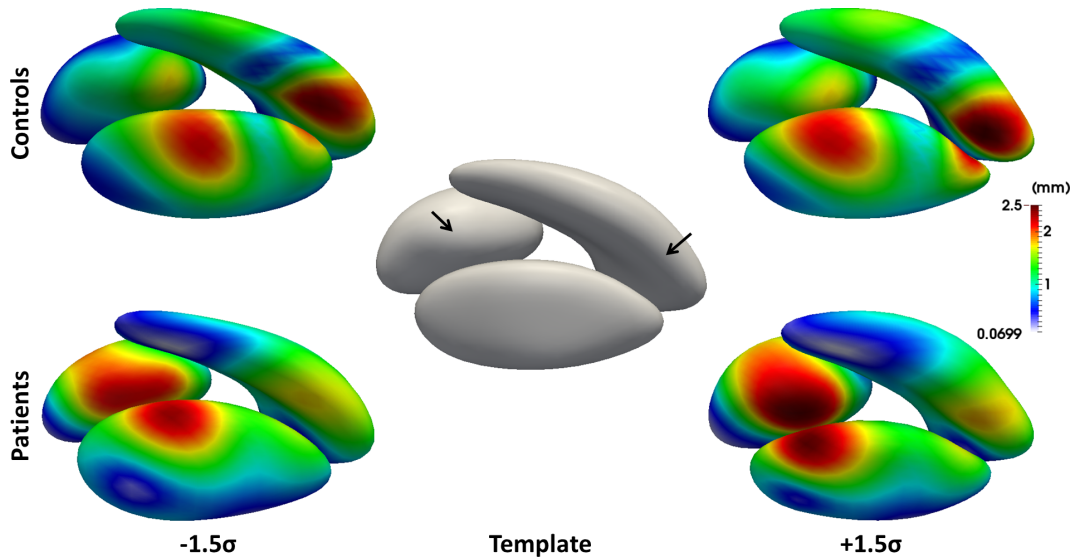


Figure 13: Second mode of variation at ± 1.5 standard deviation (σ) of two PCAs based on the two estimated covariance matrices of initial momenta Γ_α^c and Γ_α^p . Colors refer to the displacement from the final common template complex shown in the middle of the figure in gray. Arrows indicate the areas of the template complex which vary in a different way between the two groups.

4. Discussion and Conclusions

This work provides a Bayesian framework to embed a multi-object atlas construction into a statistical setting. It is general and it can be applied to any parametric deformation framework and to all shape models with which it is possible to define probability density functions. We can automatically estimate important balancing parameters which were fixed by the user in previous methods, namely the noise variances of every structure. We demonstrated that these parameters influence more the results than the hyper-parameters introduced with the proposed Bayesian priors. This statistical setting makes therefore multi-object atlas constructions more feasible and reproducible since the user is not obliged to tune or fix balancing weights which grow with the number of analysed structures.

With the proposed method, it is also possible to estimate a well-conditioned covariance matrix of the deformation parameters. In previous works (Durrleman et al., 2014; Avants and Gee, 2004; Ma et al., 2010), the authors proposed to use the sample covariance matrix computed at the end of the atlas procedure. Since the number of subjects is usually smaller than the number of deformation parameters, the sample covariance matrix might be ill-conditioned. A standard solution is to regularize it a posteriori using an identity matrix multiplied by a small scalar. This strategy is not satisfactory since it means that the estimates of the deformation parameters are based on a degenerate sample covariance matrix or on a wrong covariance matrix. Moreover, the regularization seems also too simplistic. In our approach, we directly estimate a well-conditioned covariance matrix using a more natural and coherent regularisation term given by the kernel of the RKHS to which belongs the velocity field used to compute the diffeomorphisms. Furthermore, the estimate of the covariance matrix does not change if the regularity term, and its related prior, are multiplied by a scalar. This means that we automatically take into account also the so-called regularity level.

We also extend the proposed statistical setting to a multi-population strategy, where we take into consideration the affiliation of a subject to a group. We can thus employ more pertinent models at no additional cost, meaning without fixing or tuning new parameters. One of the main advantages is that we can estimate different distributions of initial momenta, one for each group. We can then quantitatively compare their averages and modes of variations. It is interesting to notice that one could also compute the sample covariance matrices of different groups at the end of a single-population atlas and then compare their modes. However, this would not be consistent with the fact that in a single-population atlas the initial momenta of different groups are estimated as belonging to the same Gaussian distribution.

We propose to model both curve and surface meshes as Gaussian random varifolds. They can be seen therefore as instances of the same mathematical object. Since the space of varifolds is of infinite dimension, we create a grid

for every structure j on which we project both shapes and template. We decide to make the number of grid points Λ_j dependent on the varifold bandwidth λ_w^j . This choice does not influence the results in a single-object atlas, as demonstrated in Sec.3.3. In a multi-object analysis, the importance of structure j is highly determined by λ_w^j . In fact, the contribution of structure j mainly depends on the residuals and on Λ_j , which are both related to λ_w^j . On the one hand, this choice simplifies the atlas construction, since the user has to fix only one parameter. On the other hand, it makes the choice of λ_w^j even more crucial. In future works, we will investigate how to automatically estimate λ_w^j .

Another important parameter fixed by the user is the diffeomorphic kernel bandwidth λ_V . It is chosen based on the desired registration accuracy and it defines the number of control points. Its estimate, together with the one of λ_w^j , would make the atlas construction completely automatic whatever the number or kind of meshes. A possible strategy could be to select a range of suitable values of λ_V and then use sparse multi-scale methods (Sommer et al., 2012). Otherwise, one could also estimate the best deformation modules from a dictionary using sparse techniques in order to disentangle the single complicated diffeomorphism into interpretable transformations (Gris et al., 2015). This would augment the computational load and execution time but it would also make the analysis more objective.

Another improvement for our method would be the employment of sampling algorithms, like MCMC, to ensure the convergence of the EM scheme whatever the quality of the data. As already said, these methods require a great computational burden. A possible solution would be to use a GPU implementation based on Multi-Graphics Processing Units (Multi-GPU) as in Ha et al. (2009). The authors showed that a Multi-GPU implementation gives a computational gain up to sixty times faster than a single CPU implementation.

We tested the proposed algorithm comparing two populations of shape complexes, one of controls and one of Gilles de la Tourette patients. Shapes consisted of three left sub-cortical nuclei and the fiber bundles connecting them to the left cortical surface. Results proved the effectiveness of our method in detecting morphological differences between the two populations. They could indicate atypical connections resulting from abnormal brain development due to Gilles de la Tourette syndrome. Future works will aim to confirm this study by adding more structures, like the cortical surface, and possibly more subjects. Moreover, we plan to apply our algorithm to other human organs such as the heart or the liver.

Acknowledgements The research leading to these results has received funding from the program Investissements d'avenir ANR-10-IAIHU-06.

Appendix A. Multi-population atlas construction

Here we compute the cost function and gradients for the multi-population extension of the atlas construction.

Appendix A.1. $\Gamma_\alpha^c = \Gamma_\alpha^p = \Gamma_\alpha$

Let N_c and N_p be the number of subjects of the two groups and $N = N_c + N_p$. Moreover, we define $\alpha_{i0}^c \sim N(\boldsymbol{\mu}, \Gamma_\alpha)$ and $\alpha_{i0}^p \sim N(-\boldsymbol{\mu}, \Gamma_\alpha)$. Eq.9 becomes:

$$\begin{aligned}
& \sum_{j=1}^M \sum_{i=1}^{N_c} \frac{1}{2\sigma_j^2} \left(\|S_{ij}^c - \phi_i^c(T_j)\|^2 + \frac{P_j w_j}{N} \right) + \frac{1}{2} \sum_{i=1}^{N_c} (\alpha_{i0}^c - \boldsymbol{\mu})^T (\Gamma_\alpha)^{-1} (\alpha_{i0}^c - \boldsymbol{\mu}) + \\
& \sum_{j=1}^M \sum_{i=1}^{N_p} \frac{1}{2\sigma_j^2} \left(\|S_{ij}^p - \phi_i^p(T_j)\|^2 + \frac{P_j w_j}{N} \right) + \frac{1}{2} \sum_{i=1}^{N_p} (\alpha_{i0}^p + \boldsymbol{\mu})^T (\Gamma_\alpha)^{-1} (\alpha_{i0}^p + \boldsymbol{\mu}) + \quad (A.1) \\
& \sum_{j=1}^M \frac{1}{2} (w_j + \Lambda_j N) \log(\sigma_j^2) + \frac{1}{2} (w_\alpha + N) \log(|\Gamma_\alpha|) + \frac{w_\alpha}{2} \text{tr}((\Gamma_\alpha)^{-1} P_\alpha)
\end{aligned}$$

where the norm $\|\cdot\|$ can be both the L^2 -norm and the varifold one. The gradients with respect to \mathbf{T} , \mathbf{c}_0 and $\{\sigma_j^2\}$ are exactly the same as in Eq.19 and Eq.18. The one with respect to α_{i0} depends on the group and it is equal to:

$$\nabla_{\alpha_{s_0}^c} E = \sum_{j=1}^M \frac{1}{2\sigma_j^2} \nabla_{\alpha_{s_0}^c} D_{sj}^c + \Gamma_{\alpha}^{-1} (\alpha_{s_0}^c - \mu) \quad \nabla_{\alpha_{s_0}^p} E = \sum_{j=1}^M \frac{1}{2\sigma_j^2} \nabla_{\alpha_{s_0}^p} D_{sj}^p + \Gamma_{\alpha}^{-1} (\alpha_{s_0}^p + \mu) \quad (\text{A.2})$$

570 There is a closed form solution for both μ and Γ_{α} . They are equal to:

$$\hat{\Gamma}_{\alpha} = \frac{\sum_{i=1}^{N_c} [(\alpha_{i_0}^c - \mu)(\alpha_{i_0}^c - \mu)^T] + \sum_{i=1}^{N_p} [(\alpha_{i_0}^p + \mu)(\alpha_{i_0}^p + \mu)^T] + w_{\alpha} P_{\alpha}^T}{(N + w_{\alpha})}$$

$$\hat{\mu} = \frac{\sum_{i=1}^{N_c} \alpha_{i_0}^c - \sum_{i=1}^{N_p} \alpha_{i_0}^p}{N} \quad (\text{A.3})$$

Appendix A.2. $\Gamma_{\alpha}^c \neq \Gamma_{\alpha}^p$

As before, let N_c and N_p be the number of subjects of the two groups and $N=N_c+N_p$. Moreover, we define $\alpha_{i_0}^c \sim N(\mu, \Gamma_{\alpha}^c)$ and $\alpha_{i_0}^p \sim N(-\mu, \Gamma_{\alpha}^p)$. Priors on Γ_{α}^c and Γ_{α}^p are both equal to $\mathcal{W}^{-1}(P_{\alpha}, w_{\alpha})$. In this case Eq.9 becomes:

$$\boxed{\sum_{j=1}^M \sum_{i=1}^{N_c} \frac{1}{2\sigma_j^2} \left(\|S_{ij}^c - \phi_i^c(T_j)\|^2 + \frac{P_j w_j}{N} \right)} + \boxed{\frac{1}{2} \sum_{i=1}^{N_c} (\alpha_{i_0}^c - \mu)^T (\Gamma_{\alpha}^c)^{-1} (\alpha_{i_0}^c - \mu)} +$$

$$\boxed{\sum_{j=1}^M \sum_{i=1}^{N_p} \frac{1}{2\sigma_j^2} \left(\|S_{ij}^p - \phi_i^p(T_j)\|^2 + \frac{P_j w_j}{N} \right)} + \boxed{\frac{1}{2} \sum_{i=1}^{N_p} (\alpha_{i_0}^p + \mu)^T (\Gamma_{\alpha}^p)^{-1} (\alpha_{i_0}^p + \mu)} + \quad (\text{A.4})$$

$$\sum_{j=1}^M \frac{1}{2} (w_j + \Lambda_j N) \log(\sigma_j^2) + \frac{1}{2} (w_{\alpha} + N_c) \log(|\Gamma_{\alpha}^c|) + \frac{w_{\alpha}}{2} \text{tr}((\Gamma_{\alpha}^c)^{-1} P_{\alpha}) +$$

$$+ \frac{1}{2} (w_{\alpha} + N_p) \log(|\Gamma_{\alpha}^p|) + \frac{w_{\alpha}}{2} \text{tr}((\Gamma_{\alpha}^p)^{-1} P_{\alpha})$$

575 where the norm $\|\cdot\|$ can be both the L^2 -norm and the varifold one. The gradients with respect to T , c_0 and $\{\sigma_j^2\}$ are exactly the same as in Eq.19 and Eq.18. The one with respect to α_{i_0} depends on the covariance matrix of the group and its formulation is equal to the one in Eq.A.2. Furthermore, there is a closed form solution for μ , Γ_{α}^c and Γ_{α}^p . For the two covariance matrices, it is equal to the one in Eq.18 with a sum over N_c and N_p respectively. Instead, the optimal value for the average of the initial momenta is:

$$\hat{\mu} = \frac{\sum_{i=1}^{N_c} (\Gamma_{\alpha}^c)^{-1} \alpha_{i_0}^c - \sum_{i=1}^{N_p} (\Gamma_{\alpha}^p)^{-1} \alpha_{i_0}^p}{N_c (\Gamma_{\alpha}^c)^{-1} + N_p (\Gamma_{\alpha}^p)^{-1}} \quad (\text{A.5})$$

Appendix B. Robustness Analysis

580 We present in Table B.3 the values of the parameters and hyperparameters used in the experiments shown in Fig.4.

Case	Fixed		Bayesian				
	$\sigma_{surface}^2$	σ_{bundle}^2	w_α	$w'_{surface}$	$P_{surface}$	w'_{bundle}	P_{bundle}
1	0.01	0.01	1	10	0.1	10	0.1
2	0.01	0.1	0.01	0.01	0.01	0.01	0.01
3	0.01	1	0.1	0.1	0.1	0.1	0.1
4	0.01	10	1	1	1	1	1
5	0.01	100	10	10	10	10	10
6	100	0.01	100	100	100	100	100
7	100	0.1	10	1	0.1	0.01	100
8	100	1	0.1	10	1	10	0.1
9	100	10	0.01	100	10	1	0.01
10	100	100	100	0.1	0.1	10	1
11	1	0.01	0.1	0.1	100	1	10
12	1	0.1	0.1	10	10	1	1
13	1	1	0.01	1	10	0.1	100
14	1	10	10	10	0.1	100	1
15	1	100	10	1	0.01	0.1	1

Table B.3: Values of the hyperparameters for the Bayesian estimation and values of the two σ_j^2 for the Fixed method used in the robustness analysis shown in Fig.4.

References

- Allasonnière, S., Amit, Y., Trouvé, A., Jan. 2007. Towards a Coherent Statistical Framework for Dense Deformable Template Estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 69 (1), 3–29. 2, 3, 4
- Allasonnière, S., Kuhn, E., Trouvé, A., 2010. Construction of Bayesian deformable models via stochastic approximation algorithm: a convergence study. *Bernoulli* 16 (3), 641–678. 4
- Auzias, G., Colliot, O., Glaunès, J., Perrot, M., Mangin, J.-F., Trouvé, A., Baillet, S., Jun. 2011. Diffeomorphic Brain Registration Under Exhaustive Sulcal Constraints. *IEEE Transactions on Medical Imaging* 30 (6), 1214–1227. 2
- Avants, B., Gee, J. C., 2004. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *NeuroImage* 23, Supplement 1, S139–S150. 2, 5, 21
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1), 289–300. 17
- Bossa, M., Zacur, E., Olmos, S., Apr. 2011. Statistical analysis of relative pose information of subcortical nuclei: Application on ADNI data. *NeuroImage* 55 (3), 999–1008. 2
- Cates, J., Fletcher, P. T., Styner, M., Hazlett, H. C., Whitaker, R., 2008. Particle-Based Shape Analysis of Multi-object Complexes. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2008*. No. 5241 in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 477–485. 2
- Charlier, B., Charon, N., Trouv, A., Apr. 2014. The fshape framework for the variability analysis of functional shapes. arXiv:1404.6039 [cs, math]ArXiv: 1404.6039. 7
- Charon, N., Trouvé, A., 2013. The varifold representation of non-oriented shapes for diffeomorphic registration. *SIAM Journal on Imaging Sciences* 6 (4), 2547–2580. 2, 6, 7
- Cury, C., Glauns, J., Chupin, M., Colliot, O., May 2015. Analysis of anatomical variability using diffeomorphic iterative centroid in patients with Alzheimer’s disease. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 1–9. 2
- Davies, R., Twining, C., Cootes, T., Taylor, C., Apr. 2010. Building 3-D Statistical Shape Models by Direct Optimization. *IEEE Transactions on Medical Imaging* 29 (4), 961–981. 2
- Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R., Sep. 2007. Regularized, fast, and robust analytical Q-ball imaging. *Magn. Reson. Med.* 58 (3), 497–510. 12
- Durrleman, S., 2010. Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Ph.D. thesis, University of Nice-Sophia Antipolis. 7
- Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., Ayache, N., Apr. 2011a. Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NeuroImage* 55 (3), 1073–1090. 2, 3
- Durrleman, S., Pennec, X., Trouvé, A., Ayache, N., Oct. 2009. Statistical models of sets of curves and surfaces based on currents. *Medical Image Analysis* 13 (5), 793–808. 2, 12
- Durrleman, S., Prastawa, M., Charon, N., Korenberg, J. R., Joshi, S., Gerig, G., Trouvé, A., Nov. 2014. Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage* 101, 35–49. 2, 3, 5, 6, 7, 8, 9, 10, 21
- Durrleman, S., Prastawa, M., Gerig, G., Joshi, S., 2011b. Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis. In: Székely, G., Hahn, H. K. (Eds.), *Information Processing in Medical Imaging*. No. 6801 in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 123–134. 3, 8
- Folgot, L. L., Delingette, H., Criminisi, A., Ayache, N., Sep. 2014. Sparse Bayesian Registration. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2014*. No. 8673 in *Lecture Notes in Computer Science*. Springer International Publishing, pp. 235–242. 3
- Gerardin, E., Chtelat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.-S., Niethammer, M., Dubois, B., Leclercy, S., Garnero, L., Eustache, F., Colliot, O., Oct. 2009. Multidimensional classification of hippocampal shape features discriminates Alzheimer’s disease and mild cognitive impairment from normal aging. *NeuroImage* 47 (4), 1476–1486. 1
- Glaunès, J., 2005. Transport par diffeomorphismes de points, de mesures et de courants pour la comparaison de formes et l’anatomie numérique. Ph.D. thesis, Université Paris 13. 8
- Golland, P., Grimson, W. E. L., Shenton, M. E., Kikinis, R., Feb. 2005. Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis* 9 (1), 69–86. 2
- Gorbunova, V., Durrleman, S., Lo, P., Pennec, X., de Bruijne, M., Apr. 2010. Lung CT registration combining intensity, curves and surfaces. In: 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 340–343. 2
- Gorcowski, K., Styner, M., Jeong, J. Y., Marron, J., Piven, J., Hazlett, H., Pizer, S., Gerig, G., Apr. 2010. Multi-Object Analysis of Volume, Pose, and Shape Using Statistical Discrimination. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (4), 652–661. 2
- Gori, P., Colliot, O., Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., Durrleman, S., 2013. Bayesian Atlas Estimation for the Variability Analysis of Shape Complexes. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*. No. 8149 in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 267–274. 3
- Graciano Fouquier, A. B., Durrleman, S., Yelnik, J., Fernandez-Vidal, S., Bardinet, E., Sep. 2014. Iconic-Geometric Nonlinear Registration of a Basal Ganglia Atlas for Deep Brain Stimulation Planning. In: 2nd International MICCAI Workshop on Deep Brain Stimulation Methodological Challenges (DBSMC). 2
- Gris, B., Durrleman, S., Trouvé, A., 2015. A sub-riemannian modular approach for diffeomorphic deformations. In: *Proceedings of the 2nd SEE Conference on Geometric Science of Information*. 22
- Ha, L. K., Krüger, J., Fletcher, P. T., Joshi, S., Silva, C. T., 2009. Fast Parallel Unbiased Diffeomorphic Atlas Construction on Multi-Graphics Processing Units. In: *Proceedings of the 9th Eurographics Conference on Parallel Graphics and Visualization*. EG PGV’09. pp. 41–48. 22
- Hufnagel, H., Pennec, X., Ehrhardt, J., Ayache, N., Handels, H., Jun. 2009. Computation of a Probabilistic Statistical Shape Model in a Maximum-a-posteriori Framework. *Methods of Information in Medicine* 48 (4), 314–319. 2

- 645 Kurtek, S., Klassen, E., Ding, Z., Jacobson, S. W., Jacobson, J. B., Avison, M., Srivastava, A., Mar. 2011. Parameterization-Invariant Shape Comparisons of Anatomical Surfaces. *IEEE Transactions on Medical Imaging* 30 (3), 849–858. 2
- Lombaert, H., Peyrat, J.-M., 2013. Joint Statistics on Cardiac Shape and Fiber Architecture. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*. No. 8150 in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 492–500. 2
- 650 Ma, J., Miller, M. I., Trounev, A., Younes, L., Aug. 2008. Bayesian template estimation in computational anatomy. *NeuroImage* 42 (1), 252–261. 3
- Ma, J., Miller, M. I., Younes, L., 2010. A Bayesian Generative Model for Surface Template Estimation. *Int J Biomed Imaging* 2010. 2, 3, 21
- Mansi, T., Voigt, I., Leonardi, B., Pennec, X., Durrleman, S., Sermesant, M., Delingette, H., Taylor, A., Boudjemline, Y., Pongiglione, G., Ayache, N., Sep. 2011. A Statistical Model for Quantification and Prediction of Cardiac Remodelling: Application to Tetralogy of Fallot. *IEEE Transactions on Medical Imaging* 30 (9), 1605–1616. 1, 2
- 655 Nesterov, Y. E., 1983. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR* 269, 543–547. 9
- Niethammer, M., Reuter, M., Wolter, F.-E., Bouix, S., Peinecke, N., Koo, M.-S., Shenton, M. E., 2007. Global Medical Shape Analysis Using the Laplace-Beltrami Spectrum. *Med Image Comput Comput Assist Interv* 10 (Pt 1), 850–857. 2
- O'Donnell, L. J., Westin, C.-F., Golby, A. J., Apr. 2009. Tract-based morphometry for white matter group analysis. *NeuroImage* 45 (3), 832–844. 2
- 660 Palit, A., Bhudia, S. K., Arvanitis, T. N., Turley, G. A., Williams, M. A., Feb. 2015. Computational modelling of left-ventricular diastolic mechanics: Effect of fibre orientation and right-ventricle topology. *Journal of Biomechanics* 48 (4), 604–612. 2
- Patenaude, B., Smith, S. M., Kennedy, D. N., Jenkinson, M., Jun. 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* 56 (3), 907–922. 11
- 665 Perrin, M., Poupon, C., Cointepas, Y., Rieul, B., Golestani, N., Pallier, C., Rivire, D., Constantinesco, A., Bihan, D. L., Mangin, J.-F., 2005. Fiber Tracking in q-Ball Fields Using Regularized Particle Trajectories. In: Christensen, G. E., Sonka, M. (Eds.), *Information Processing in Medical Imaging*. No. 3565 in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 52–63. 11
- Qiu, A., Brown, T., Fischl, B., Ma, J., Miller, M., Jun. 2010. Atlas Generation for Subcortical and Ventricular Structures With Its Applications in Shape Analysis. *IEEE Transactions on Image Processing* 19 (6), 1539–1547. 2
- 670 Risholm, P., Janoos, F., Norton, I., Golby, A. J., Wells III, W. M., Jul. 2013. Bayesian characterization of uncertainty in intra-subject non-rigid registration. *Medical Image Analysis* 17 (5), 538–555. 3
- Savadjiev, P., Strijkers, G. J., Bakermans, A. J., Piuze, E., Zucker, S. W., Siddiqi, K., Dec. 2012. Heart wall myofibers are arranged in minimal surfaces to optimize organ function. *PNAS* 109 (24), 9248–9253. 2
- Siless, V., Glaunès, J., Guevara, P., Mangin, J.-F., Poupon, C., Bihan, D., Thirion, B., Fillard, P., 2012. *Medical Image Computing and Computer-Assisted Intervention - MICCAI*. Springer Berlin Heidelberg, Ch. Joint T1 and Brain Fiber Log-Demons Registration Using Currents to Model Geometry, pp. 57–65. 2
- 675 Simpson, I. J. A., Schnabel, J. A., Groves, A. R., Andersson, J. L. R., Woolrich, M. W., Feb. 2012. Probabilistic inference of regularisation in non-rigid registration. *NeuroImage* 59 (3), 2438–2451. 3
- Sommer, S., Lauze, F., Nielsen, M., Pennec, X., Dec. 2012. Sparse Multi-Scale Diffeomorphic Registration: The Kernel Bundle Framework. *J Math Imaging Vis* 46 (3), 292–308. 22
- 680 Srivastava, M. S., Yanagihara, H., Jul. 2010. Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis* 101 (6), 1319–1329. 20
- Vaillant, M., Glaunès, J., 2005. Surface Matching via Currents. In: Christensen, G. E., Sonka, M. (Eds.), *Information Processing in Medical Imaging*. No. 3565 in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 381–392. 6
- 685 Vaillant, M., Miller, M. I., Younes, L., Trounev, A., 2004. Statistics on diffeomorphisms via tangent space representations. *NeuroImage* 23, Supplement 1, S161–S169. 9
- Wassermann, D., Toews, M., Niethammer, M., Wells III, W., Jul. 2014. Probabilistic Diffeomorphic Registration: Representing Uncertainty. In: Ourselin, S., Modat, M. (Eds.), *Biomedical Image Registration*. No. 8545 in *Lecture Notes in Computer Science*. Springer International Publishing, pp. 72–82. 3
- 690 Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Valabregue, R., Poupon, F., Guevara, P., Tucholka, A., Mangin, J.-F., Vidailhet, M., Lehericy, S., Hartmann, A., Poupon, C., Feb. 2015. Altered structural connectivity of cortico-striato-pallido-thalamic networks in Gilles de la Tourette syndrome. *Brain* 138 (2), 472–482. 12
- Zhang, M., Singh, N., Fletcher, P. T., 2013. Bayesian Estimation of Regularization and Atlas Building in Diffeomorphic Image Registration. In: Gee, J. C., Joshi, S., Pohl, K. M., Wells, W. M., Zillei, L. (Eds.), *Information Processing in Medical Imaging*. No. 7917 in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 37–48. 3
- 695