Morphometry of anatomical shape complexes with dense deformations and sparse parameters

Stanley Durrleman[a,b], Marcel Prastawa[c], Nicolas Charon[d], Julie R. Korenberg[e], Sarang Joshi[c], Guido Gerig[c], Alain Trouvé[d]

[a] *INRIA, project-team Aramis, Paris, France*
[b] *Institut du Cerveau et de la Moëlle épinière (ICM), Hôpital de la Pitié Salpêtrière, 75013 Paris, France*
[c] *Scientific Computing and Imaging (SCI) Institute, University of Utah, Salt Lake City, UT 84112,USA*
[d] *Centre de Mathématiques et Leurs Applications (CMLA), Ecole Normale Supérieure de Cachan, 94230 Cachan, France*
[e] *Brain Institute, University of Utah, Salt Lake City, UT 84112, USA*

Abstract

Shape statistics in neuroimaging studies consist in finding variations in brain structure from a set of segmented contours. Contours in brain images are indicative of numerous boundaries between tissues. Their intrinsic shape is not more important than the relative position between them, and with the background which may also contain boundaries, but that are not visible at the current resolution. We propose here a generic statistical model for such complex of surfaces, which is based on diffeomorphic deformations that warp the ambient 3D space while preserving the internal organization of tissues. The method estimates a template shape complex which integrates shape invariants in a given data set. Variability is quantified by space deformations that warp the template to each individual complex. Their parameters are used as shape descriptors for statistical purposes. The method estimates simultaneously shape variants from data and an optimal basis of fixed dimension for the decomposition of the shape descriptors, therefore alleviating the need for post-hoc dimension reduction technique. The approach is illustrated in a typical neuroimaging study of Down syndrome. It shows that a complex of deep brain structures has a statistically significant different shape according to the copy number of chromosome 21 ($p < 10^{-3}$). The model classifies subjects with very high specificity and sensitivity, thus showing important generalization capability, even with a very limited number of samples. These

results are stable if the dimension of the parameterization is drastically reduced, and might even suggest an increased statistical performance in lower dimension. The method highlights realistic and interpretable patterns of anatomical variations. It has therefore a great potential to find anatomical biomarkers of brain disorders or give insights into pathophysiology.

Keywords: morphometry, deformation, varifold, anatomy, shape, statistics

---

1. Introduction

The blooming of non-invasive imaging methods has made possible the analysis of anatomical phenotypic variations over larger and larger data collections. Magnetic Resonance Images (MRI) are used to reveal and quantify effects of pathologies on anatomy, such as hippocampal atrophy in neurodegenerative diseases or change in neuronal connectivity in neurodevelopmental disorders. In clinical trials, MRIs are used to assess the effectiveness of drugs or treatments on disease progression. Digital anatomical models are built from the extraction of structures of interest in images, which usually take the form of a set of 3D surfaces, called a shape complex. The quantification of phenotypic variations of shape complexes across individuals or populations is crucial to gives insights into pathophysiology, to detect early biomarker of disease onset or to correlate phenotypes with functional or genotypic variations, for instance. Not only the *quantification* is important, but also the *description* of the significant anatomical differences between individuals or groups, to interpret the findings and drive the search for hypothesis.

Geometric morphometric methods rely on carefully defined correspondences between homologous positions on surfaces called landmarks (Bookstein, 1991; Dryden and Mardia, 1998). Landmark-free metrics often make strong assumptions about the topology of the surface, whether they work only for genus zero surfaces (Chung et al., 2003; Boyer et al., 2010), use medial representations (Styner et al., 2005; Bouix et al., 2005; Gorczowski et al., 2010) or Laplace-Beltrami eigenfunctions (Reuter et al., 2006). Therefore, they can be rarely applied to raw surface meshes resulting from segmentation algorithms, as they might have small holes, irregular sampling, or may consist of different sheets.

More importantly, such metrics deal with individual shape components, without taking into account the relations among components within a complex. By contrast, we believe that it is crucial to define metric on shape

complexes, especially to study brain anatomy that consists in an intricate network of various structures with strong interrelationships. Metrics of individual components could be summed up, or correlations among components could be measured (Gorczowski et al., 2010), but these approaches do not constrain the preservation of the complex structure in the statistical analysis, and therefore may yield unrealistic anatomical configurations with collisions between components.

One way to address this problem is to consider surfaces as a set of contours embedded in the 3D space and to measure shape variations induced by the deformation of the underlying 3D space. This idea stems from Grenander's group theory for modeling objects (Grenander, 1994), which revisits morphometry by the use of 3D space deformations. Similarity between shape complexes is then quantified by the "amount" of deformation needed to warp one shape complex to another. Only smooth and invertible 3D deformations (i.e. diffeomorphisms) are used, so that the internal organization of the shape complex is preserved during deformation: neither surface intersection or shearing may occur. The coherence of the underlying biological tissue is preserved, which includes structures that might not have been segmented from the original images, or that might not be visible at image resolution. Such space deformations integrate correlations among surface components, while enforcing topological constraints.

This approach contrasts with most shape statistics techniques, which essentially build on the idea that points on surfaces correspond, such correspondences resulting either from manual indication of homologous points or from surface-to-surface mapping. Here, we assume point correspondence in the whole 3D volume instead, the surfaces being used as constraints in the search for those correspondences. As a result, surfaces do not perfectly match, thus making the method more robust to segmentation errors. This approach shares similarities with iconic methods that use image intensities as constraints. However, iconic methods give more weight to high-contrast contours, such as the ventricles boundaries, and less weight to low-contrast contours such as those of the deep brain nuclei. Using labeled surfaces adds also the information about the homology of surfaces. Knowing that homologous surfaces need to correspond, one alleviates the problem of structure mismatch that is typical of iconic methods.

In this paper, we propose a generic method for the statistical analysis of collections of shape complexes, which builds on the implementation of Grenander's theory in the LDDMM framework (Miller et al., 2006; Vaillant

et al., 2007; McLachlan and Marsland, 2007). The method estimates simultaneously a template shape complex, which averages out the shape characteristics of every individual, and 3D deformations which warp the template to each individual's complex. Template and deformations altogether is called an *atlas*.

The template is an ideal representation of the shape complex. The user freely chooses its structure and sampling: number of connected components, number of vertices and edges in each component. Then, the method automatically adjusts the positions of the vertices, integrating shape features captured in individuals, while preserving the template structure. In contrast to previous attempts (Glaunès and Joshi, 2006; Durrleman et al., 2009; Ma et al., 2008), the template can have one connected component per anatomical structures and its initialization is not biased toward a particular subject. To measure similarity between the template complex with a given topology and complexes of individuals which may have different sampling and mesh imperfections, we use the metric of varifolds that does not assume point correspondence across surfaces and is robust to variations in meshing (Charon and Trouvé, 2013). It extends the currents (Vaillant and Glaunès, 2005) in that it does not need consistent surface orientation across samples, thus further reducing the need for pre-processing and user intervention. It also enables to study folded surfaces, such as the cortex, whose folds cancel out in currents. Conversely, it also prevents smooth surfaces from folding during deformation instead of pushing their boundary.

This metric is used to drive the estimation of space deformations that warp the template complex to each individual complex. Deformation parameters give the relative position of each individual with respect to the template on a Riemannian manifold. They are used for the statistical analysis of shape variations within or between groups (Vaillant et al., 2004; Pennec, 2006). We adopt here a control point parameterization of space deformations. The number of control points is fixed by the user and the method automatically adjusts their position near the most variable parts of the shape complex. Each control point integrates information in a 3D neighborhood and is not linked to a particular surface as in (Vaillant and Glaunès, 2005). In this setting, the dimension of the shape descriptor used in statistics is controlled, and does not explode with the number of surfaces and their sampling (Vaillant and Glaunès, 2005). We show that statistical performance is not altered by this finite-dimensional approximation and that these parameters can robustly detect fine anatomical differences in a typical low sample size study. In some

cases, the statistical performance is even increased, as the ratio between the number of subjects and the number of parameters becomes more favorable.

The originality of the method lies in the combination of different techniques that have been developed in the recent past within the LDDMM community. It is the first time that one combines (i) the sparse parameterization of diffeomorphisms introduced in the LDDMM literature in Durrleman et al. (2013), (ii) the technique to control the topology of the template shape complex introduced in Durrleman et al. (2012) and (iii) the varifold metric to compare meshes without the need for consistent orientation introduced in Charon and Trouvé (2013). These techniques are now combined into a simplified theoretical framework thanks to a formulation inherited from optimal control theory. In the following, we give a self-contained presentation of the method, which can be accessible to readers who might not be familiar with the LDDMM literature. Then, we will follow with a typical anatomical study of Down syndrome, which will show the potential of the method.

2. Mathematical Framework

2.1. Kernel formulation of splines

In the spline framework, 3D deformations ϕ are of the form $\phi(x) = x + v(x)$, where v is a displacement field that is a sum of radial basis functions K located at control point positions $\{c_k\}_{k=1,\dots,N_{\text{cp}}}$:

$$v(x) = \sum_{k=1}^{N_{\text{cp}}} K(x, c_k)\alpha_k, \tag{1}$$

where $\alpha_1, \dots, \alpha_{N_{\text{cp}}}$ are vector weights and $K(x, y)$ is a scalar function on $\mathbb{R}^3 \times \mathbb{R}^3$, such as the Gaussian $K(x, y) = \exp(-|x - y|^2 / \sigma_V^2)$ that we will use in the applications.

It is beneficial to assume that K is a positive definite symmetric kernel, namely that K is continuous and that for any finite set of distinct points $\{x_i\}_i$ and vectors $\{\alpha_i\}_i$:

$$\sum_i \sum_j K(x_i, x_j){\alpha_i}^T \alpha_j \geq 0, \tag{2}$$

the equality holding only if all α_i vanish. Translation invariant kernels are of particular interest. According to Bochner's theorem, functions of the form

$K(x-y)$ are positive definite kernels if and only if their Fourier transform is a positive definite operator ((2) is then a discrete convolution). This enables to easily check that the previous Gaussian function is indeed a positive-definite kernel, among other possible choices.

Assuming K is a kernel allows us to define the pre-Hilbert space V as the set of any finite sums of terms $K(.,x)\alpha$ for vector weights α. Given two vector fields $v_1 = \sum_i K(.,x_i)\alpha_i$ and $v_2 = \sum_j K(.,y_j)\beta_j$, (2) ensures that the bilinear map

$$\langle v_1, v_2 \rangle_V = \sum_i \sum_j K(x_i, y_j) {\alpha_i}^T \beta_j \tag{3}$$

defines an inner-product on V. This also shows that any vector field $v \in V$ satisfies the reproducing property:

$$\langle v, K(.,x)\alpha \rangle_V = v(x)^T \alpha \tag{4}$$

As shown in Zeidler (1991) such vector spaces could be "completed" into a Hilbert space, for which (4) hold, called Reproducing Kernel Hilbert Space (RKHS).

Using matrix notations, we denote $\mathbf{x}$ and $\boldsymbol{\alpha}$ (resp. $\mathbf{y}$ and $\boldsymbol{\beta}$) in $\mathbb{R}^{3N}$ (resp. $\mathbb{R}^{3M}$) the concatenation of the 3D vectors x_i and α_i (resp. y_j and β_j), so that the dot product (3) writes $\langle v_1, v_2 \rangle_V = \boldsymbol{\alpha}^T \mathbf{K}(\mathbf{x}, \mathbf{y}) \boldsymbol{\beta}$, where $\mathbf{K}(\mathbf{x}, \mathbf{y})$ is the $N \times M$ matrix with entries $K(x_i, y_j)$.

2.2. Flows of diffeomorphisms

The main drawback of spline deformations is its non-invertibility, as soon as the magnitude of $v(x)$ or its Jacobian is "too" large. The idea to build diffeomorphisms is to use the vector field v as an instantaneous velocity field instead of a displacement field. Formally, one integrates the differential equation for $t \in [0,1]$:

$$\frac{\partial \phi_t}{\partial t} = v_t \circ \phi_t, \quad \phi_0 = \mathrm{Id}_{\mathbb{R}^3}, \tag{5}$$

where v_t is a time-varying velocity field of the form:

$$v_t(x) = \sum_{i=1}^{N_{\mathrm{cp}}} K(x, c_k(t)) \alpha_k(t) \tag{6}$$

It is shown in Miller et al. (2006) that (5) could be integrated over $t \in [0,1]$ for any square integrable time-varying vectors $\alpha_k(t)$, and that the solution

$x \to \phi_t(x)$ is a time-varying flow of 3D diffeomorphisms. The diffeomorphism at time $t = 1$ is the deformation of interest. Any point x_0 follows the trajectory $x(t) = \phi_t(x_0)$ which is the integral curve of $\dot{x}(t) = v_t(x(t))$ with $x(0) = x_0$. In particular, control points $\mathbf{c}_0 = \{c_{0,k}\}$ move along the integral curves of (using matrix notations),

$$\dot{\mathbf{c}}(t) = \mathbf{K}(\mathbf{c}(t), \mathbf{c}(t))\boldsymbol{\alpha}(t), \quad \mathbf{c}(0) = \mathbf{c}_0. \tag{7}$$

The time-varying vectors $\boldsymbol{\alpha}(t)$ which enables to reach a given deformation ϕ_1 from $\phi_0 = \mathrm{Id}_{\mathbb{R}^3}$ are not unique. According to Lagrangian principles, we choose the time-varying vectors that minimize the kinetic energy along the path:

$$\int_0^1 \|v_t\|_V^2 \, dt = \int_0^1 \boldsymbol{\alpha}(t)^T \mathbf{K}(\mathbf{c}(t), \mathbf{c}(t))\boldsymbol{\alpha}(t) dt, \tag{8}$$

considering $\mathbf{c}(0)$ and $\mathbf{c}(1)$ fixed. We show in Appendix A that minimizing vectors satisfy geodesic equations that write, combining them with (7):

$$\begin{cases} \dot{c}_k(t) = \displaystyle\sum_{p=1}^{N_{\mathrm{cp}}} K(c_k(t), c_p(t))\alpha_p(t) \\ \dot{\alpha}_k(t) = -\displaystyle\sum_{p=1}^{N_{\mathrm{cp}}} \alpha_k(t)^T \alpha_p(t) \nabla_1 K(c_k(t), c_p(t)) \end{cases} \tag{9}$$

One could easily check that $\|v_t\|_V$ is constant along such geodesic paths. Therefore, $\|v_0\|_V$ is the length of the geodesic path which connects $\phi_0 = \mathrm{Id}_{\mathbb{R}^3}$ to ϕ_1.

Denoting $\mathbf{S}(t) = \{\mathbf{c}(t), \boldsymbol{\alpha}(t)\}$ the state of the system of control points at time t, (9) could be written in short as

$$\dot{\mathbf{S}}(t) = F(\mathbf{S}(t)), \quad \mathbf{S}(0) = \{\mathbf{c}_0, \boldsymbol{\alpha}_0\}. \tag{10}$$

Given initial position of control points $\mathbf{c}_0$ and initial vectors $\boldsymbol{\alpha}_0$ (called momenta in this context), the solution of this coupled ODEs parameterizes a time-varying velocity field v_t via (6), and therefore a geodesic flow of diffeomorphisms $\phi_t(x)$ via (5). To be more precise, let $\mathbf{X}_0$ be the concatenation of points in 3D, then these points move in space to $\phi_t(\mathbf{X}_0)$ that we denote $\mathbf{X}(t)$. Given (6) and (5), $\mathbf{X}(t)$ is the solution of the ODE $\dot{\mathbf{X}}(t) = v_t(\mathbf{X}(t)) = \mathbf{K}(\mathbf{X}(t), \mathbf{c}(t))\boldsymbol{\alpha}(t)$ with $\mathbf{X}(0) = \mathbf{X}_0$. We write this ODE in short as:

$$\dot{\mathbf{X}}(t) = G(\mathbf{X}(t), \mathbf{S}(t)), \quad \mathbf{X}(0) = \mathbf{X}_0 \tag{11}$$

2.3. Varifold metric between surfaces

To estimate a diffeomorphism that best matches two surfaces, we need a similarity metric between any pair of surfaces $\mathcal{S}$ and $\mathcal{S}'$. We adopt the varifold model (Charon and Trouvé, 2013): a rectifiable surface is seen as continuous linear form on W, a subspace of smooth scalar fields on $\mathbb{R}^3 \times \overleftrightarrow{\mathbb{S}}$, where $\overleftrightarrow{\mathbb{S}}$ denotes the quotient of the unit sphere in $\mathbb{R}^3$ by the two elements group $\{\pm \mathrm{Id}_{\mathbb{R}^3}\}$ (i.e. the set of non oriented directions). In the sequel, $\overleftrightarrow{u}$ denotes the class of $u \in \mathbb{R}^3$ in $\overleftrightarrow{\mathbb{S}}$. For a rectifiable surface $\mathcal{S}$ seen as a vector in W^*, the dual of W, and $\omega \in W$, we define the dual bracket as:

$$\mathcal{S}(\omega) = \int_{\mathcal{S}} \omega(x, \overleftrightarrow{n(x)}) d\sigma(x), \tag{12}$$

where $n(x)$ denotes the normal of $\mathcal{S}$ at point x and $d\sigma(x)$ the volume form. This definition does not require the surface $\mathcal{S}$ to be orientable and is invariant under normal flipping.

For computational reasons (Vaillant and Glaunès, 2005; Durrleman et al., 2009), we assume W to be a separable RKHS on $\mathbb{R}^3 \times \overleftrightarrow{\mathbb{S}}$ with kernel $\mathcal{K}$ chosen as:

$$\mathcal{K}\left((x, \overleftrightarrow{u}), (y, \overleftrightarrow{v})\right) = K^W(x,y) \left(\frac{u^T v}{|u|\,|v|}\right)^2 \tag{13}$$

where K^W is typically a Gaussian kernel of width σ_W.

The reproducing property (4) shows that:

$$\mathcal{S}(\omega) = \left\langle \omega, \textstyle\int_{\mathcal{S}} \mathcal{K}((x, \overleftrightarrow{n(x)}), (.,.))\, |n(x)|\, dx \right\rangle_W = \left\langle \omega, \mathcal{L}_W^{-1}(\mathcal{S}) \right\rangle_W,$$

where $\mathcal{L}_W^{-1}$ maps $\mathcal{S}$ to its Riesz representant in W.

Therefore, the inner-product between two rectifiable surfaces $\mathcal{S}$ and $\mathcal{S}'$ is $\langle \mathcal{S}, \mathcal{S}' \rangle_{W^*} = \mathcal{S}(\mathcal{L}_W^{-1}(\mathcal{S}')) =$

$$\int_{\mathcal{S}} \int_{\mathcal{S}'} K^W(x,x') \left(\frac{n(x)^T n(x')}{|n(x)|\,|n(x')|}\right)^2 d\sigma(x) d\sigma(x') \tag{14}$$

If $\mathcal{S}$ and $\mathcal{S}'$ are surface meshes, this inner-product could be approximated by (accounting for cells areas)

$$\langle \mathcal{S}, \mathcal{S}' \rangle_{W^*} = \sum_p \sum_q K^W(c_p, c'_q) \left(u_p^T u'_q\right)^2 \tag{15}$$

where c_p and n_p (resp. c'_q and n'_q) denotes the centers and normals of the faces of $\mathcal{S}$ (resp. $\mathcal{S}'$), $u_p = n_p / |n_p|^{1/2}$ and $u'_p = n'_p / |n'_p|^{1/2}$.

For $\mathcal{S}$ a rectifiable surface and ϕ a diffeomorphism, the surface $\phi(\mathcal{S})$ can still be seen as a varifold. Indeed, a change of variables shows that for $\omega \in W$, $\phi(\mathcal{S})(\omega) = \mathcal{S}(\phi \star \omega)$ where $\phi \star \omega(x, \overleftrightarrow{n}) = \left|(d_x\phi^{-1})^T n\right| \omega(\phi(x), \overleftrightarrow{(d_x\phi^{-1})^T n})$ (Charon and Trouvé, 2013). Therefore, the varifold metric can be used to search for the diffeomorphism ϕ that best matches $\mathcal{S}$ with $\mathcal{S}'$ by minimizing $d_W(\phi(\mathcal{S}), \mathcal{S}')^2 = \|\phi(\mathcal{S}) - \mathcal{S}'\|^2_{W^*}$, without the need to find point correspondence between meshes and in a way which is robust to mesh imperfections at scale smaller than kernel width σ_W, like noise, small holes, shearing or re-meshing.

2.4. Atlas construction method

Let $\{\mathcal{O}_i\}$ be a set of N_{su} surface complexes, each complex being made of meshes $\mathcal{S}_{i,1}, \ldots, \mathcal{S}_{i,N}$. This is typically a set of N anatomical structures segmented in N_{su} subjects. In the context of Fréchet means, the template complex $\mathcal{O}_0$ is defined as the minimizer of the sample variance: $\mathcal{O}_0 = \arg\min \sum_i d_\phi(\mathcal{O}_0, \mathcal{O}_i)^2$, where we use as metric d_ϕ the length of the geodesic flow of diffeomorphisms which connects $\mathcal{O}_0$ to $\mathcal{O}_i$, namely $\|v_0^i\|_V$ where the velocity v_0^i parameterize the flow $\phi_t^{v_0^i}$ such that $\phi_0^{v_0^i}(\mathcal{O}_0) = \mathcal{O}_0$ and $\phi_1^{v_0^i}(\mathcal{O}_0) = \mathcal{O}_i$.

Such a flow that exactly connects two surface meshes does not necessarily exist, and is not even desirable to avoid over-fitting. Therefore, we use the following relaxed formulation of the criterion to be minimized:

$$E(\mathcal{O}_0, \{v_0^i\}) = \sum_{i=1}^{N_{\text{su}}} \left\{ \sum_{k=1}^{N} \frac{1}{2\sigma_k^2} d_W(\phi_1^{v_0^i}(\mathcal{S}_{0,k}), \mathcal{S}_{i,k})^2 + \left\|v_0^i\right\|_V^2 \right\} \tag{16}$$

where σ_k^2 play the role of one Lagrange mutliplier per surface mesh, and where the metric on varifolds d_W is used to constrain the i*th* deformation to match the k*th* shape within the template complex $\mathcal{O}_0$ (i.e. $\mathcal{S}_{0,k}$) with its homologous shape within the i*th* subject's complex $\mathcal{O}_i$ (i.e. $\mathcal{S}_{i,k}$). The sum of $\|v_0^i\|_V^2$ is the sample variance: it attracts the template complex to the "mean" of the set.

To minimize (16), we make the following assumptions. The template object $\mathcal{O}_0$ consists of N surface meshes, whose number of vertices and edges are fixed by the user. The initial velocities v_0^i are parameterized by momenta $\alpha_{0,p}^i$ attached to control points $c_{0,p}$ whose number is also fixed by the user.

Control points are the same for all subjects, so that each velocity v_0^i share the same parameterization. Momenta $\alpha_{0,p}^i$ are subject-specific and parameterize N_{su} different deformations.

Let $\mathbf{S}_0^i = \{c_{0,p}, \alpha_{0,p}^i\}$ denote the parameters of v_0^i, and $\mathbf{X}_0$ the vertices of every template surface concatenated into a single vector. The flow of diffeomorphisms results from the integration of N_{su} differential equations, as in (10): $\dot{\mathbf{S}}^i(t) = F(\mathbf{S}^i(t))$ with $\mathbf{S}^i(0) = \mathbf{S}_0^i$. As in (11), $\mathbf{X}_0$ follows the integral curve of N_{su} differential equations: $\dot{\mathbf{X}}^i(t) = G(\mathbf{X}^i(t), \mathbf{S}^i(t))$ with $\mathbf{X}^i(0) = \mathbf{X}_0$. The final value being $\mathbf{X}^i(1) = \phi_1^{v_0^i}(\mathbf{X}_0)$, the criterion (16) becomes:

$$E(\mathbf{X}_0, \{\mathbf{S}_0^i\}) = \sum_{i=1}^{N_{\text{su}}} A(\mathbf{X}^i(1)) + L(\mathbf{S}_0^i) \tag{17}$$

where A denotes the weighted sum of data attachment terms $d_W(\mathbf{X}^i(1), \mathcal{S}_{i,k})^2$ and $L(\mathbf{S}_0^i) = \|v_0^i\|_V^2 = {\boldsymbol{\alpha}_0^i}^T K(\mathbf{c}_0, \mathbf{c}_0)\boldsymbol{\alpha}_0^i$.

As proven in Appendix B, the gradient of E is given as:

$$\begin{cases} \nabla_{\boldsymbol{\alpha}_0^i} E = \boldsymbol{\xi}^{\alpha,i}(0) + \nabla_{\boldsymbol{\alpha}_0^i} L(\mathbf{S}_0^i) \\ \nabla_{\mathbf{c}_0} E = \displaystyle\sum_{i=1}^{N_{\text{su}}} \left(\boldsymbol{\xi}^{c,i}(0) + \nabla_{\mathbf{c}_0} L(\mathbf{S}_0^i)\right) \end{cases}, \quad \nabla_{\mathbf{X}_0} E = \textstyle\sum_{i=1}^{N_{\text{su}}} \boldsymbol{\theta}^i(0),$$

where the auxiliary variables $\boldsymbol{\xi}^i(t) = \{\boldsymbol{\xi}^{c,i}(t), \boldsymbol{\xi}^{\alpha,i}(t)\}$ (of the same size as $\mathbf{S}^i(t)$) and $\boldsymbol{\theta}^i(t)$ (of the same size as $\mathbf{X}_0$) satisfy the linear ODEs (integrated backward in time):

$$\begin{cases} \dot{\boldsymbol{\theta}}^i(t) = -\Big(\partial_1 G(\mathbf{X}^i(t), \mathbf{S}^i(t))\Big)^T \boldsymbol{\theta}^i(t), \quad \boldsymbol{\theta}^i(1) = \nabla_{\mathbf{X}^i(1)} A \\ \dot{\boldsymbol{\xi}}^i(t) = -\big(\partial_2 G(\mathbf{X}^i(t), \mathbf{S}^i(t)\big)^T \boldsymbol{\theta}^i(t) - d_{\mathbf{S}^i(t)} F^T \boldsymbol{\xi}^i(t), \ \boldsymbol{\xi}^i(1) = 0 \end{cases}$$

Data come into play only in the gradient of the varifold metric with respect to the position of the *deformed* template $\nabla_{\mathbf{X}^i(1)} A$ (derivation is straightforward and given in Appendix C). This gradient shows in which direction the vertices of the deformed template have to move to decrease the criterion. This could be achieved in two ways, by optimizing the shape of the template complex and the deformations matching the template to each complex. The vector $\boldsymbol{\theta}^i$ transports the gradient back to $t = 0$ where it is used to update the position of the vertices of the template complex. The vector $\boldsymbol{\xi}^i$ interpolates at the control points the information in $\boldsymbol{\theta}^i$ that is located at the template

points, and is used at $t = 0$ to update deformation parameters. A striking advantage of this formulation is that one single gradient descent optimizes simultaneously the template shape and deformation parameters.

This gradient has been computed using the L^2 metric for template points $\mathbf{X}_0$. To prevent self-intersections during template evolution by the gradient descent $s \to \mathbf{X}_0(s)$, we compute the gradient for a metric given by a Gaussian kernel K^X with width σ_X, so that the gradient used in practice is:

$$\nabla^X_{x_{0,k}} E = \sum_{i=1}^{N_{\text{su}}} \sum_{p=1}^{N_{\text{x}}} K^X(x_{0,k}, x_{0,p}) \theta^i_p(0). \tag{18}$$

As shown in Appendix D, this new gradient $\nabla^X E$ is the restriction to $\mathbf{X}_0$ of a smooth vector field u_s. Therefore, at any step s of the gradient descent, $\mathbf{X}_0(s) = \psi_s(\mathbf{X}_0(0))$ where ψ_s is the family of diffeomorphisms integrating the flow of u_s.

Eventually, the criterion is minimized using a line search gradient descent method. The algorithm is initialized with template surfaces given as ellipsoids, control points located at the nodes of a regular lattice and momenta vectors set to zero (i.e. no deformation). At convergence, the method yields the final atlas: a template shape complex, optimized positions of control points and deformation momenta (more details in Appendix E). The vectors $\alpha^i_{0,k}$ that are all attached to the same points $c_{0,k}$ characterize the distribution of the data around the template. They are the tangent-space representations of the deformations that are used for statistics.

3. Application to a Down syndrome neuroimaging study

We evaluate our method on a dataset of 3 anatomical structures segmented from MRIs of 8 Down syndrome (DS) subjects and 8 control cases. The hippocampus, amygdala and putamen of the right hemisphere (respectively in green, cyan and orange in figures) form a complex of grey matter nuclei in the medial temporal lobe of the brain. This study aims to detect complex non-linear morphological differences between both groups, thus going beyond size analysis that already showed DS subjects to have smaller brain structures than controls (Korenberg et al., 1994). To discard any linear differences, including size, we co-register all shape complexes using affine transforms. We construct then an atlas from pooled data, setting $\sigma_V = 10\,\text{mm}$, $\sigma_W = 5\,\text{mm}$, $\sigma_X = \sigma_V/2$ and $\sigma_k = \sigma_V$ for all nuclei, and control

points initially located at the nodes of regular lattice of step σ_V, yielding a set of 105 points (see Appendix F). Rationale behind parameter setting and robustness of results with respect to their values is discussed in Sec. 3.6 and Appendix E.2.

3.1. Group differences

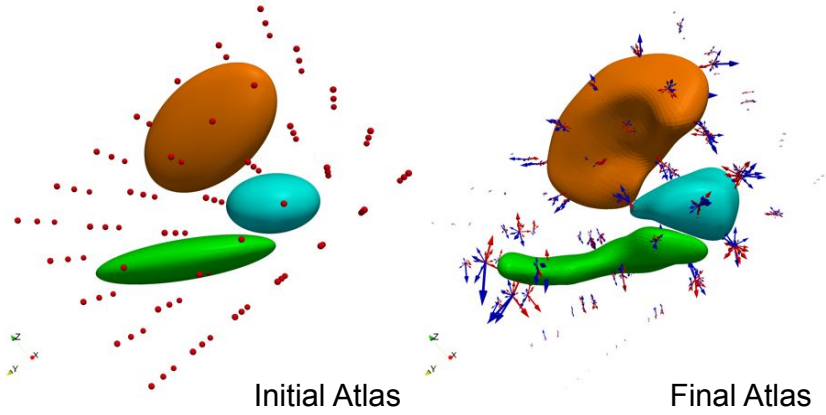
The template shape complex resulting from the atlas construction (Fig. 1-a) averages the shape characteristics of every individual in the data set. The position of each subject's anatomical configuration (either DS or controls) with respect of the template configuration is quantified by initial momentum vectors located at control point positions (arrows in Fig. 1). Each set of momentum vectors defines a 3D dense deformation that warp the template complex to each individual's complex via the integration of the geodesic shooting equations (10) followed by the flow equations (11).

To highlight group differences, we compute the sample mean of the momenta for each group separately. We then deform the template complex in the direction of both means, thus showing anatomical configurations that are typical of each group (Fig. 2). It shows that DS subjects nuclei are turned toward the left part of the brain, with an other torque that pushes the hippocampus tail (its posterior part) toward the superior part of the brain, and the head toward the inferior part. These two torques are more pronounced near the hippocampus/amygdala boundary than in the hippocampus tail or upper putamen region. DS subjects amygdala has also lesser lateral extension than controls.

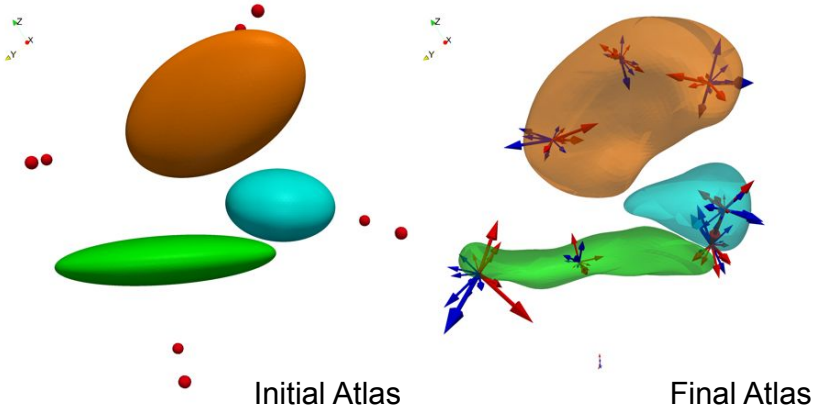
We perform a Linear Discriminant Analysis (LDA) to exhibit the most discriminative axis between both groups in the momenta space (more in Appendix G). The deformation of the template complex in the direction of the most discriminative axis (Fig. 3) reveals similar thinning effects and torques as in Fig. 2. It also shows that DS subjects putamens are more bended.

3.2. Statistical significance

We estimate the statistical significance of the above group differences using permutation tests. Each subject's data is characterized by a shape descriptor of dimension 3 times the number of control points (namely the concatenation of the momentum vectors), so that we can use multivariate statistics and then avoid the problem of correction for multiple comparisons. In our experiments, however, the number of subjects is always smaller than



a - Atlas construction with 105 control points



b- Atlas construction with 8 control points

Figure 1: Atlas estimated from different initial conditions. Top: 105 control points with initial spacing equal to the deformation kernel width $\sigma_V = 10$ mm, Bottom: 8 control points. Arrows are the momentum vectors of DS subjects (red) and controls (blue)

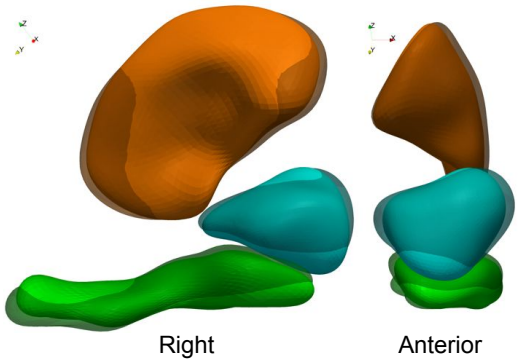


Figure 2: Template complex deformed using the mean deformation of controls (transparent shapes) and DS subjects (opaque shapes)

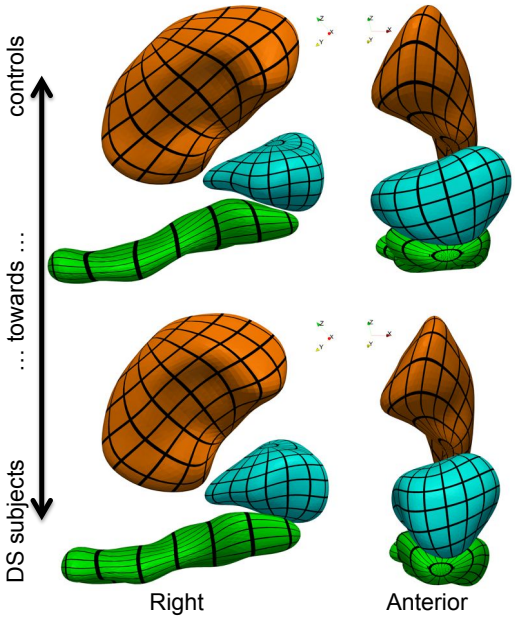


Figure 3: Most discriminative deformation axis. Differences are amplified, since the distance between the two configurations is twice the distance between the means. (black grids have been mapped to the surface for visualization only)

the dimension of the descriptor, in which case the distribution Hotelling T^2 statistics cannot be computed.

We address this issue by selecting the first modes of the average of the two within-class covariance matrices (See Appendix G.3 for details). For a number of modes N_{modes} that is smaller than the number of subjects, the directions of modes define a subspace of N_{modes} dimension in the parameter space, onto which individual shape descriptors are projected. The Hotelling T^2 statistics is then computed in this subspace.

To estimate the distribution of the Hotelling statistics under the null hypothesis of equal means, we repeat the above procedure by permuting the subjects' indices 10^5 times. Each permutation changes the empirical means and within-class covariance matrices, and then the selected subspace and the statistics. The p-value which corresponds to the number of modes that explains 95% of the within-class covariance is $p = 10^{-3.58}$, thus showing that anatomical differences highlighted in Fig. 2 are statistically significant.

3.3. Sensitivity and specificity using cross-validation

Over-fitting is a common problem of statistical estimations in high dimension low sample size setting. We perform leave-out experiments to evaluate the generalization errors of our model, namely its sensitivity and specificity.

We compute an atlas with the same parameters and initial conditions but with one control and one DS subject data out, yielding $8^2 = 64$ gradient descents (note that this is a design choice: one does not necessarily need to have balanced groups to apply the method). For each experiment, we register the template shape complex to each of the left-out complex (by minimizing (16) for $N_{\text{su}} = 1$ and considering template and control points of the atlas fixed). The resulting momentum vectors are compared with those of the atlas. We classify based on Maximum Likelihood (ML) ratios and LDA (details in Appendix G). The very high sensitivity and specificity reported in Table 1 (first row) show that the anatomical differences between DS and controls that were captured by the model are not specific to this particular data set, but are likely to generalize well to independent data sets.

3.4. Shape complexes versus bags of shapes

In this section, we aim to emphasize the differences between using a single model of shape complex and using different models for each component.

We perform the same experiments as described above, but for each of the three structures independently. The atlas of each structure has its own set

	LDA		ML	
	specificity	sensitivity	specificity	sensitivity
Shape complex	98 (63/64)	100 (64/64)	100 (64/64)	100 (64/64)
Hippocampus	97 (62/64)	87 (56/64)	92 (59/64)	100 (64/64)
Amygdala	98 (63/64)	100 (64/64)	91 (58/64)	100 (64/64)
Putamen	75 (48/64)	100 (64/64)	98 (63/64)	100 (64/64)
Composite	97 (62/64)	100 (64/64)	100 (64/64)	100 (64/64)

Table 1: Classification with 105 control points using LDA or ML classifier. Scores (in percentage) are computed using our descriptor for shape complex (first row), only one structure at a time (rows 2-4) or a composite descriptor (fifth row)

of control points and momentum vectors. The hypothesis of equal means for DS and control subjects is rejected with probability of false positive of $p = 10^{-2.45}$ for the hippocampus, $p = 10^{-2.33}$ for the putamen and $p = 10^{-3.92}$ for the amygdala. The statistical significance is lower for the hippocampus and the putamen than for the shape complex ($p = 10^{-3.58}$), and higher for the amygdala. The classification scores reported in Table 1 (rows 2 to 4) show that none of the structures alone could predict subject's status as well as the shape complex. Although the model for amygdala has a higher statistical significance, this is at the cost of lower specificity in the Maximum Likelihood approach.

We deform each structure along their most discriminative axis. Because the 3 deformations are not combined into a single space deformation, intersections between surfaces occur (Fig. 4). The deformation of the amygdala, though highly significant, is not compatible with the deformation of the hippocampus. From an anatomical point of view, both parts of the amygdala/hippocampus boundary should vary in the same direction: they could not intersect and are very unlikely separated by a variable amount of white matter.

The previous shape complex analysis showed that the most discriminative effects involve deformations of specific subregions, and in particular the most lower-anterior part of the complex, where the amygdala is located (Fig. 3). Therefore, it is not surprising that this structure shows higher statistical performance than the hippocampus and putamen in an independent analysis of each structure. However, the most discriminative deformations of each structure are not consistent among themselves, thus misleading the interpretation

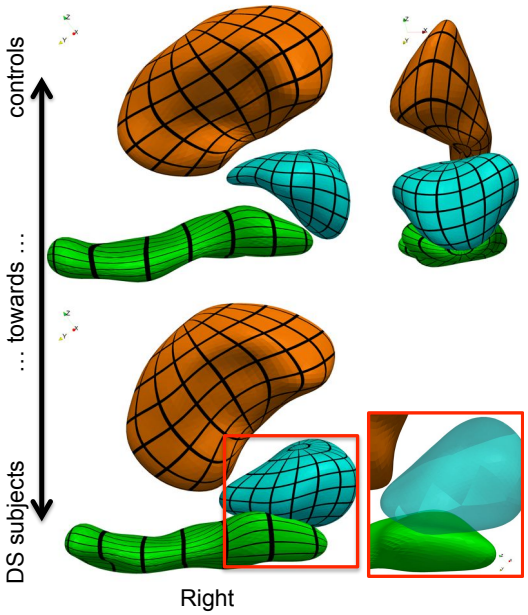


Figure 4: Most discriminative deformation axis computed for each surface independently. Surface intersection occurs in absence of a global diffeomorphic constraint. (black grids have been mapped to the surface for visualization only)

of the findings. By contrast, the shape complex analysis shows that the discriminative effect is not specific to the amygdala but to the whole lower anterior part of the medial temporal lobe with strong correlations between parts of the structures within the region. The shape complex model may be slightly less significant, but it highlights more informative effects that could be interpreted.

Another way to account for correlations between structures is to build a composite shape descriptor by concatenating control points and momentum vectors of each atlas (see Appendix G.5). This closely mimics what usual shape analysis methods would do. This composite descriptor enables to achieve nearly as good classification as with the single atlas method (Table 1, last row) with a very high statistical significance $p < 10^{-5}$. However, this composite descriptor does not parameterize one single space deformation, but still three of them, and surfaces still intersect during deformation (Supplementary Data Fig. S1). By contrast, the single atlas method integrates topology constraints in the analysis, and therefore could only measure correlations that preserve the internal organization of the anatomical complex.

3.5. Effects of dimension reduction

Our approach offers the possibility to control the dimension of the shape descriptor by choosing the number of control points given as input of the method. In 3D, the dimension of the shape descriptor is 3 times the number of control points. We evaluate here the impact of this dimension for the atlas construction and statistical estimations in such a low sample size setting.

A sensible initialization of control points is to place them a the nodes of a regular lattice. This fixes their number, while their positions are optimized by the gradient descent. An adequate initial spacing between control points is given by deformation kernel width σ_V, in order to avoid too much redundancy in the deformation parameters (Durrleman et al., 2009). It gives 105 control points for $\sigma_V = 10$ mm. However, it might be that the variations of the shapes are constrained enough, so that they belong to small dimensional space and therefore could be described by much fewer parameters.

We successively down-sample the initial grid of 105 control points. With only 8 points, the number of deformation parameters is decreased by more than one order of magnitude and the initial ellipsoidal shapes still converge to a similar template shape complex (Fig. 1-b). The main reason is that control points were able to move to the most strategic places, noticeably at the tail of the hippocampus and the anterior part of the amygdala where the variability is the greatest. Qualitatively, the most discriminant axis is stable when the dimension is varied (Supplementary Data Fig. S2), as well as the spectrum of sample covariance matrices of the momentum vectors (Supplementary Data Fig. S3). The method was able to optimize the "amount" of variability captured for a given dimension of deformations parameters. Nevertheless, the residual data term at convergence increases. The initial data term (i.e. varifold norm) was decreased by 97.8% for 105 points, and only by 93.3 for 8 points, thus showing that the sparsest model captured less variability in the data set (Table 2).

Optimizing deformation parameterization in a infinite dimensional setting amounts to place a continuum of control points on the surface meshes, and in practice, one control points at each vertex (Vaillant and Glaunès, 2005; Ma et al., 2008). At the given resolution of the template meshes, such a parameterization would involve 23058 control points. Nonetheless, this number can be arbitrarily increased or decreased by up/down sampling the initial ellipsoids, regardless of the variability in the data set! We increase the number of control points to 650 and notice that the estimated template shapes are the same as with 105 control points (results not shown), and that

Number of CP	8	12	16	24	36	105	600
Decrease of data term (in % of initial value)	93.3	94.8	94.6	95.8	96.7	97.9	97.8

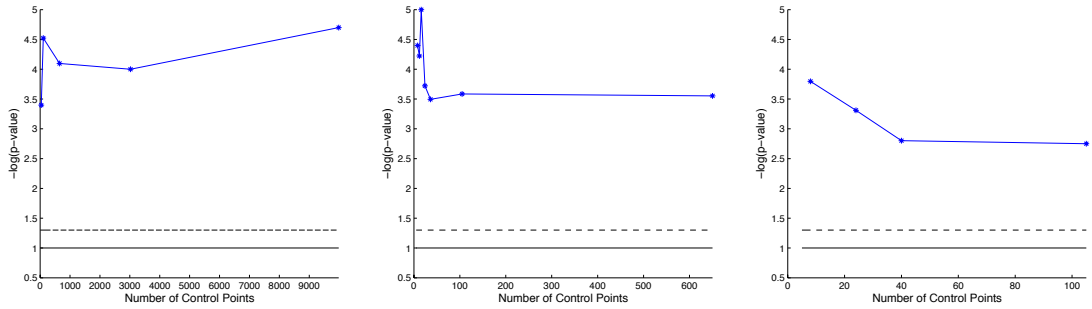
Table 2: Decrease of the data term during optimization for different number of control points and $\sigma_V = 10$ mm

the atlas explains the same proportion of the initial data term (Table 2). Therefore, increasing the number of control points does not allow to capture more information (which is essentially determined by the deformation kernel size σ_V, see below), but distributes this information over a larger number of parameters.

Now, we estimate the effect of this dilution of information in higher dimension on the statistical performance of the model. We build an atlas for different number of control points that are always located initially at the nodes of a regular lattice. We compute the p-value associated to the Hotelling T^2 statistics for each atlas as explained in Sec. 3.2, using the number of modes that explains 95% of the variance in the post-hoc Principal Component Analysis (PCA). We notice that the statistical significance is not increased with higher dimension (Fig. 5-b). It is even smaller than in small dimension, the maximum being reached for 16 control points ($p < 10^{-5}$).

We also compute the generalization errors of the model for each set of control points. Leave-2-out experiments give 100% specificity and sensitivity using the ML approach, regardless of the number of control points used. To highlight differences, we performed classification using the hippocampus shape only. Again, the performance of the classifier does not necessarily decrease with the number of control points (Table 3). ROC curves in Fig. 6 show that the atlases with 48 and 18 control points have poorer performance than atlases with 12 and 8 control points.

These results suggest that using atlases of small dimension could have greater statistical power, especially in a such small sample size setting. Nevertheless, two different dimension reduction techniques compete with each other in these experiments. The use of a small set of control points is a built-in dimension reduction technique, which has the advantage to optimize *simultaneously* the information captured in the data and the encoding of this information in a space of fixed dimension. The post-hoc dimension reduction using PCA projects shape descriptors on the subspace explaining most of the



a- $\sigma_V = 5$ mm b- $\sigma_V = 10$ mm c- $\sigma_V = 15$ mm

Figure 5: Statistical significance of the group means difference for varying number of control points. The solid (resp. dashed) line corresponds to the 0.1 (resp. 0.05) significance threshold

variance captured. The variation of the p-values when the number of modes selected in the PCA is varied shows that this post-hoc dimension reduction technique has more effects on the statistical significance than the variation of the number of control points (Fig. G.1). Despite the erratic aspects of the plots, it is worth noticing that in all cases, the maximum p-value is never achieved for number of control points greater than 105.

Even with only 8 control points (one point at each corner of a bounding box), the dimension of the atlas is of $3 * 8 = 24$, thus still larger than the number of subjects in this study. This means that a post-hoc dimension reduction technique is applied in all cases, making difficult a fair comparison between the two approaches. With a few dozen of subjects, we could estimate full-rank covariance matrices and not rely on the post-hoc reduction technique at all. Our hypothesis is that, in this regime, the trend of increased statistical significance when the number of control points is decreased would be amplified. Indeed, the ratio between the number of variables to estimate and the number of subjects is more favorable in this case, thus making the statistical estimations more stable. This would be even more crucial, would we estimate means and covariance matrices during optimization in a Bayesian framework, in the spirit of Allassonnière et al. (2007); Allassonnière and Kuhn (2009).

3.6. Effects of parameters

We assess here the robustness of the results with respect to parameter values. The method depends on the deformation parameter σ_V, the var-

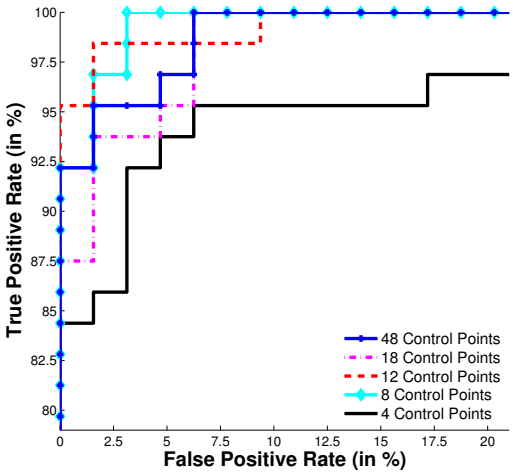


Figure 6: ROC curves for hippocampus classification using different number of control points in the atlas and ML classifier

ifold parameter σ_W, the weights σ_k of each structure in the cost function and the gradient smoothing parameter σ_X. The role of each parameter and rules of thumb for setting their values is explained in E.2. We perform the same experiments as above, but with different values of the parameters. Statistical estimates, and classification scores in particular, vary very smoothly with changes in deformation and varifold parameters, thus showing the reproducibility of our results without fine parameter tuning (Fig. 5 and Table H.1).

The most important effect is the one of the deformation parameter, which determines the scale of the variants that can be detected. Deformations build essentially on elementary deformations of patches of the domain of radius σ_V. With smaller values of σ_V, the deformation captured more local shape variations. With larger values, it averages variations over larger regions. This effect is visible on the estimated template shapes. With $\sigma_V = 5$ mm, the template shapes capture finer details in the data, resulting in less smooth surface than with $\sigma_V = 15$ mm (Fig. H.1). With fewer details in the atlas, classification scores decrease (Table H.1).

The dimension of the atlas is intrinsically linked with the deformation parameter. Deformations with smaller σ_V need more control points to potentially deform every small regions of the shape complex. Deformations with larger σ_V have fewer degrees of freedom and could be decomposed using fewer control points. Placing one control point at the nodes of lattice of step σ_V yields 15 control points for $\sigma_V = 15$ mm, 105 control points for $\sigma_V = 10$ mm and 650 control points for $\sigma_V = 5$ mm. We build an atlas for each of these values of σ_V and with down/up sampling the set of con-

	# Control Points	48	18	12	8	4
LDA	specificity	97 (62/64)	91 (58/64)	92 (59/64)	95 (61/64)	78 (50/64)
	sensitivity	87 (56/64)	89 (57/64)	89 (57/64)	89 (57/64)	81 (52/64)
ML	specificity	92 (59/64)	92 (59/64)	97 (69/64)	97 (62/64)	84 (54/64)
	sensitivity	100 (64/64)	100 (64/64)	98 (63/64)	100 (64/64)	97 (62/64)

Table 3: Classification ratios based solely on hippocampus shape. LDA and ML classification are performed with various number of control points in the atlas. Ratios are in percentage

trol points. All these atlases show a good significance level, far below the usual threshold of 0.05. On average, the statistical significance is decreased with increasing σ_V, as the atlas gives a coarser and coarser description of variability within the data set (Fig. 5). With $\sigma_V = 15$ mm (Fig. 5-c), the maximum significance is reached for 8 control points, and the significance is decreased with increasing dimension. With $\sigma_V = 5$ mm (Fig. 5-a), the same trend is observed, except an unexpected increase in statistical significance at very high dimension.

The results for the largest deformation parameter ($\sigma_V = 15$ mm) tend to confirm our hypothesis that the built-in dimension reduction technique yields increased statistical power. The number of control points that maximizes the p-value for any given number of modes is almost never the largest one (Supplementary Data Fig. S4). The results for the smallest deformation kernel ($\sigma_V = 5$ mm) tend also to show the same trend, though to a lesser extent (Supplementary Data Fig. S5). In this case, even important downsampling of the initial set of 650 control points does not decrease the dimension of the space to the point where it could compare with the number of subjects, thus making less visible the effect of the built-in dimension reduction technique compared to larger deformation parameters. This highlights the fact that the level of anatomical details that one would like to capture is limited by the number of subjects in the data set. The more local the shape variations we want to estimate, the more observations we need. The increase in statistical significance (type I error) if σ_V is decreased may be misleading in this respect, as it is likely to decrease the statistical power (type II error) for a given effect size and number of parameters.

4. Discussion and Conclusion

We presented here a comprehensive method for the statistical analysis of shape complexes extracted from 3D anatomical images. It can deal with raw surfaces resulting from nearly any segmentation methods thanks to its robustness to noise, mesh imperfections and inconsistencies in mesh orientation. The method estimates a template shape complex with a fixed topology that is representative of the anatomy, and modes of deformations that preserve template structure and capture variability in data. Such topology constraints lead to modes of variations that are anatomically realistic and interpretable. This approach contrasts therefore with the study of correlations between shape models that are estimated independently for each component within a shape complex. A typical neuroimaging study of a complex of deep brain structures in Down syndrome subjects shows that the method could find discriminative anatomical features with high statistical significance and small generalization errors, even with a very limited number of observations. We show the robustness of these results in various experimental setting, demonstrating the effectiveness of the method without fine parameter tuning.

The method uses control points that are automatically placed at their optimal positions. This approach has several advantages. It decreases the complexity of the algorithm, since the optimization in position of control points allows to decrease the number of points needed to describe the variability of a particular data set. It links the dimension of the shape descriptor with the deformation scale σ_V, namely the scale of the anatomical details that could be captured by the model, which seems more sensible than linking this dimension with the resolution of the template meshes. Last but not least, the choice of the number of control points offers a built-in dimension reduction technique. As compared with a post-hoc dimension reduction technique such as PCA, the dimension of the descriptors is known prior to any experiments. This eases power calculations and sample size estimates required in clinical trials. It may lead to increased statistical performance as suggest by our results, although they should be strengthened using more subjects and independent data sets. It also paves the way for estimating mean and covariance matrices during the optimization in a Bayesian framework, in the spirit of Allassonnière et al. (2007); Allassonnière and Kuhn (2009). Constraining such a statistical inference to take place in a small dimensional space is likely to increase the convergence speed of the statistical estimates, as compared to performing the inference in very high dimension and then perform dimension

reduction. Indeed, the direction of first modes in very high dimension, and hence the selected subspace, are not very stable as the number of subjects is increased, and introduce noise in the statistical estimation.

At the moment, the choice of the number of control points is left to the user, with a good heuristic that consists in placing one point every deformation parameter σ_V. Automatic selection of control points in the spirit of Durrleman et al. (2012, 2013) did not show increased statistical power due to its penalization in magnitude of momentum vectors. We believe that true model selection should be investigated instead.

Cross-validation showed a very good prediction capability of our model. Prediction of Down syndrome based on neuroimaging data has little clinical interest, since subjects are characterized by their genotype and especially the copy number of chromosome 21, which is known with very high confidence. Nonetheless, our model is generic and could be applied to different pathologies for which the clinical status may be more difficult to assess. This prediction capability of the method shows its potential for computer-aided diagnosis or prognosis in studies where subjects' status is based only on clinical diagnosis with limited reproducibility, such as in neurodegenerative diseases. The shape descriptors, which encode the shape variability in a small number of parameters, are also pieces of choice to study correlations between anatomical phenotypes and genotype, in the spirit of Korbel et al. (2009) where they can take the place of clinical variables.

Acknowledgments. This work has been partly supported by NIH grants U54 EB005149 (NA-MIC), 1R01 HD067731, 5R01 EB007688 and 2P41 RR0112553-12.

Appendix A Geodesic equations

We derive here the minimum action principle of Lagrangian mechanics. A variation $\delta\boldsymbol{\alpha}(t)$ of the time-varying momentum vectors $\boldsymbol{\alpha}(t)$ induces a variation of the control point positions $\delta\mathbf{c}(t)$, which in turn induces a variation δE of the kinetic energy $E = \int_0^1 \boldsymbol{\alpha}(t)^T \mathbf{K}(\mathbf{c}(t), \mathbf{c}(t))\boldsymbol{\alpha}(t)dt$.

Since $\dot{\mathbf{c}} = \mathbf{K}(\mathbf{c}, \mathbf{c})\boldsymbol{\alpha}$, we have

$$\delta\dot{\mathbf{c}} = \mathbf{K}(\mathbf{c}, \mathbf{c})\delta\boldsymbol{\alpha} + d_{\mathbf{c}}\left(\mathbf{K}(\mathbf{c}, \mathbf{c})\boldsymbol{\alpha}\right)\delta\mathbf{c}, \tag{19}$$

and

$$E = \int_0^1 \alpha^T \dot{\mathbf{c}} dt. \tag{20}$$

Therefore, we have

$$\dot{\mathbf{c}}^T \delta\boldsymbol{\alpha} = \boldsymbol{\alpha}^T \mathbf{K}(\mathbf{c},\mathbf{c})\delta\boldsymbol{\alpha} = \boldsymbol{\alpha}^T \delta\dot{\mathbf{c}} - \boldsymbol{\alpha}^T d_{\mathbf{c}} \left(\mathbf{K}(\mathbf{c},\mathbf{c})\boldsymbol{\alpha}\right) \delta\mathbf{c} \tag{21}$$

and

$$\begin{aligned} \delta E &= \int_0^1 \dot{\mathbf{c}}^T \delta\boldsymbol{\alpha} + \boldsymbol{\alpha}^T \delta\dot{\mathbf{c}} dt \\ &= \int_0^1 2\boldsymbol{\alpha}^T \delta\dot{\mathbf{c}} - \boldsymbol{\alpha}^T d_{\mathbf{c}} \left(\mathbf{K}(\mathbf{c},\mathbf{c})\boldsymbol{\alpha}\right) \delta\mathbf{c} dt. \end{aligned} \tag{22}$$

Assuming $\delta\mathbf{c}(0) = \delta\mathbf{c}(1) = 0$, integration by parts yields:

$$\delta E = -\int_0^1 \left(2\dot{\boldsymbol{\alpha}} + d_{\mathbf{c}} \left(\mathbf{K}(\mathbf{c},\mathbf{c})\boldsymbol{\alpha}\right)^T \boldsymbol{\alpha}\right)^T \delta\mathbf{c} dt \tag{23}$$

The linear ODE with source term (19) shows that there is a one-to-one relationship between $\delta\mathbf{c}$ and $\delta\boldsymbol{\alpha}$. Since $\delta\boldsymbol{\alpha}$ is arbitrary, so is $\delta\mathbf{c}$ and

$$\dot{\boldsymbol{\alpha}} = -\frac{1}{2} d_{\mathbf{c}} \left(\mathbf{K}(\mathbf{c},\mathbf{c})\boldsymbol{\alpha}\right)^T \boldsymbol{\alpha} \tag{24}$$

along extremal paths.

$\mathbf{K}(\mathbf{c},\mathbf{c})\boldsymbol{\alpha}$ is a $3N_{\text{cp}}$ vector, whose kth coordinate is the 3D vector: $\sum_{p=1}^{N_{\text{x}}} K(c_k, c_p)\alpha_p$. Therefore,

$$d_{c_i}(\mathbf{K}(\mathbf{c},\mathbf{c})\boldsymbol{\alpha})_k = \sum_{p=1}^{N_{\text{cp}}} \alpha_p \nabla_1 K(c_k, c_p)^T \delta(i-k) + \alpha_i \nabla_2 K(c_k, c_i)^T \tag{25}$$

Using the fact that K is symmetric (hence $\nabla_1 K(x,y) = \nabla_2 K(y,x)$) we have:

$$\dot{\alpha}_i = -\frac{1}{2} \sum_{k=1}^{N_{\text{cp}}} (d_{c_i}(\mathbf{K}(\mathbf{c},\mathbf{c})\boldsymbol{\alpha})_k)^T \alpha_k = -\left(\sum_{k=1}^{N_{\text{cp}}} \nabla_1 K(c_i, c_k)\alpha_k^T\right) \alpha_i \tag{26}$$

Appendix B Gradient of the atlas criterion

We provide here the differentiation of the criterion for atlas construction:

$$E(\mathbf{X}_0, \mathbf{c}_0, \{\boldsymbol{\alpha}_0^i\}) = \sum_{i=1}^{N_{\text{su}}} A(\mathbf{X}^i(1)) + L(\mathbf{S}_0^i) \tag{27}$$

subject to:

$$\begin{cases} \dot{\mathbf{S}}^i(t) = F(\mathbf{S}^i(t)) & \mathbf{S}^i(0) = \{\mathbf{c}_0, \boldsymbol{\alpha}_0^i\} \\ \dot{\mathbf{X}}^i(t) = G(\mathbf{X}^i(t), \mathbf{S}^i(t)) & \mathbf{X}^i(0) = \mathbf{X}_0 \end{cases} \tag{28}$$

where

$$L(\mathbf{S}_0^i) = {\boldsymbol{\alpha}_0^i}^T K(\mathbf{c}_0, \mathbf{c}_0) \boldsymbol{\alpha}_0^i \tag{29}$$

$\mathbf{X}$ is a vector of length $3N_{\mathrm{x}}$, where N_{x} is the number of points in the template shape, $\mathbf{c}$ and $\boldsymbol{\alpha}$ are two vectors of length $3N_{\mathrm{cp}}$ each, where N_{cp} is the number of control points, so that $\mathbf{S}$ is a vector of length $6N_{\mathrm{cp}}$.

$F(\mathbf{S}) = \begin{pmatrix} F^c(\mathbf{c}, \boldsymbol{\alpha}) \\ F^\alpha(\mathbf{c}, \boldsymbol{\alpha}) \end{pmatrix}$ is a vector of length $6N_{\mathrm{cp}}$, which is decomposed into two vectors of size $3N_{\mathrm{cp}}$. The k*th* coordinate (among N_{cp}) of F^c and F^α is the 3D vector:

$$\begin{aligned} F^c(\mathbf{S})_k &= \sum_{p=1}^{N_{\mathrm{cp}}} K(c_k(t), c_p(t)) \alpha_p(t) \\ F^\alpha(\mathbf{S})_k &= - \sum_{p=1}^{N_{\mathrm{cp}}} \alpha_k(t)^T \alpha_p(t) \nabla_1 K(c_k(t), c_p(t)) \end{aligned} \tag{30}$$

$G(\mathbf{X}, \mathbf{S})$ is a vector of size $3N_{\mathrm{x}}$. Its k*th* coordinate (among N_{x}) is the 3D vector:

$$G(\mathbf{X}, \mathbf{S})_k = \sum_{p=1}^{N_{\mathrm{cp}}} K(x_k(t), c_p(t)) \alpha_p(t) \tag{31}$$

Similarly,

$$L(\mathbf{S}_0^i) = \sum_{p=1}^{N_{\mathrm{cp}}} \sum_{q=1}^{N_{\mathrm{cp}}} {\alpha_{0,p}^i}^T K(c_{0,p}, c_{0,q}) \alpha_{0,q}^i \tag{32}$$

B.1 Gradient in matrix form

The differentiation of the criterion can be done for each subject i independently. Therefore, we differentiate only one term of the sum in (27) for a generic subject's index i that we omit in the following for clarity purposes.

A small perturbation $\delta\mathbf{S}_0$ induces a perturbation of the motion of the control points and momenta $\delta\mathbf{S}(t)$, which, in turn, induces a perturbation

of the template points' trajectory $\delta\mathbf{X}(t)$ and then of the criterion δE, which writes thanks to the chain rule:

$$\delta E = \left(\nabla_{\mathbf{X}(1)} A\right)^T \delta\mathbf{X}(1) + \left(\nabla_{\mathbf{S}_0} L\right)^T \delta\mathbf{S}_0. \tag{33}$$

According to (28), the perturbations $\delta\mathbf{S}(t)$ and $\delta\mathbf{X}(t)$ satisfy the linearized ODEs:

$$\begin{aligned} \delta\dot{\mathbf{S}}(t) &= d_{S(t)} F \delta\mathbf{S}(t) & \delta\mathbf{S}(0) &= \delta\mathbf{S}_0 \\ \delta\dot{\mathbf{X}}(t) &= \partial_1 G \delta\mathbf{X}(t) + \partial_2 G \delta\mathbf{S}(t) & \delta\mathbf{X}(0) &= \delta\mathbf{X}_0 \end{aligned}$$

The first ODE is linear. Its solution is given by:

$$\delta\mathbf{S}(t) = \exp\left(\int_0^t d_{\mathbf{S}(u)} F du\right) \delta\mathbf{S}_0. \tag{34}$$

The second ODE is linear with source term. Its solution is given by:

$$\delta\mathbf{X}(t) = \int_0^t \exp\left(\int_u^t \partial_1 G ds\right) \partial_2 G(u) \delta\mathbf{S}(u) du + \exp\left(\int_0^t \partial_1 G(s) ds\right) \delta\mathbf{X}_0 \tag{35}$$

Plugging (34) into (35) and then into (33) leads to:

$$\begin{cases} \nabla_{\mathbf{S}_0} E = \displaystyle\int_0^1 {R_{0t}}^T \partial_2 G(\mathbf{X}(t), \mathbf{S}(t))^T {V_{t1}}^T \nabla_{\mathbf{X}(1)} A dt + \nabla_{\mathbf{S}_0} L \\ \nabla_{\mathbf{X}_0} E = {V_{01}}^T \nabla_{\mathbf{X}(1)} A \end{cases}, \tag{36}$$

where we denoted $R_{st} = \exp\left(\int_s^t d_{\mathbf{S}(u)} F du\right)$ and $V_{st} = \exp\left(\int_s^t \partial_1 G(\mathbf{X}(u), S(u)) du\right)$.

Let us denote $\boldsymbol{\theta}(t) = {V_{t1}}^T \nabla_{\mathbf{X}(1)} A$, $g(t) = \partial_2 G(t)^T \boldsymbol{\theta}(t)$ and $\boldsymbol{\xi}(t) = \int_t^1 {R_{ts}}^T g(s) ds$, so that the gradient (36) can be re-written as:

$$\begin{cases} \nabla_{\mathbf{S}_0} E = \displaystyle\int_0^1 {R_{0s}}^T g(s) ds + \nabla_{\mathbf{S}_0} L = \boldsymbol{\xi}(0) + \nabla_{\mathbf{S}_0} L \\ \nabla_{\mathbf{X}_0} E = \boldsymbol{\theta}(0) \end{cases}.$$

Now, we need to make explicit the computation of the auxiliary variables $\boldsymbol{\theta}(t)$ and $\boldsymbol{\xi}(t)$. By definition of V_{t1}, we have $V_{11} = \mathrm{Id}$ and $dV_{t1}/dt = V_{t1}\partial_1 G(t)$, which implies that $\boldsymbol{\theta}(1) = \nabla_{X(1)} A$ and $\dot{\boldsymbol{\theta}}(t) = -\partial_1 G(t)^T \boldsymbol{\theta}(t)$.

For $\boldsymbol{\xi}(t)$, we notice that $R_{ts} = \mathrm{Id} - \int_t^s \frac{dR_{us}}{du} du = \mathrm{Id} + \int_t^s R_{us} d_{\mathbf{S}(u)} F(u) du$. Therefore, using Fubini's theorem, we get:

$$\begin{aligned}
\boldsymbol{\xi}(t) &= \int_t^1 {R_{ts}}^T g(s) ds \\
&= \int_t^1 g(s) + d_{\mathbf{S}(s)} F^T \int_s^1 {R_{su}}^T g(u) du ds \\
&= \int_t^1 g(s) + d_{\mathbf{S}(s)} F^T \boldsymbol{\xi}(s) ds.
\end{aligned}$$

This last equation is nothing but the integral form of the ODE given in the main text.

Given the actual values of $\mathbf{S}_0$ and $\mathbf{X}_0$, one needs to integrate the geodesic shooting equations and the flow equation in (28) forward in time to give the full path of parameters $\mathbf{S}(t)$ and template shape points $\mathbf{X}(t)$. Then, one needs to compute the gradient of the data term $\nabla_{\mathbf{X}(1)} A$, which is given in Appendix C. This term indicates in which direction one has to move the vertices of the deformed template shape in order to better match the observations. This term is transported back to time $t = 0$ by the coupled linear equations satisfied by $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$. The values of time $t = 0$ of these auxiliary variables are used to update the deformation parameters (position of control points and momenta) and the position of the vertices of the template surfaces.

B.2 Gradient in coordinates

Expanding the variables $\mathbf{S}^i(t) = \{c_{0,k}(t), \alpha^i_{0,k}(t)\}$, $\mathbf{X}^i(t) = \{X^i_k(t)\}$, $\boldsymbol{\theta}^i(t) = \{\theta^i_k(t)\}$ and $\boldsymbol{\xi}^i(t) = \{\xi^{c,i}_k(t), \xi^{\alpha,i}_k(t)\}$, we have

$$\begin{aligned}
\nabla_{c_{0,k}} E &= \sum_{i=1}^{N_{\mathrm{su}}} \xi^{c,i}_k(0) + \nabla_{c_{0,k}} L(\mathbf{S}^i_0) \\
\nabla_{\alpha^i_{0,k}} E &= \sum_{i=1}^{N_{\mathrm{su}}} \xi^{\alpha,i}_k(0) + \nabla_{\alpha^i_k} L(\mathbf{S}^i_0) \\
\nabla_{x_{0,p}} E &= \sum_{i=1}^{N_{\mathrm{su}}} \theta^i_p(0)
\end{aligned}$$

where the gradient of L is given as (from now on, we omit the subject's index i for clarity purposes):

$$\nabla_{\alpha_{0,k}} L = 2 \sum_{p=1}^{N_{\text{cp}}} K(c_{0,k}, c_{0,p}) \alpha_{0,p}$$
$$\nabla_{c_{0,k}} L = 2 \sum_{p=1}^{N_{\text{cp}}} {\alpha_{0,p}}^T \alpha_{0,k} \nabla_1 K(c_{0,k}, c_{0,p})$$

The term $\partial_1 G(\mathbf{X}(t), \mathbf{S}(t))$ is a block-matrix of size $3N_{\text{cp}} \times 3N_{\text{x}}$ whose $(k,p)th$ 3×3 block is given as:

$$d_{X_k} G(\mathbf{X}(t), \mathbf{S}(t))_p = \sum_{j=1}^{N_{\text{cp}}} \alpha_j(t) \nabla_1 K(X_p(t), c_j(t))^T \delta(p-k)$$

so that the vector $\boldsymbol{\theta}(t)$ is updated according to:

$$-\dot{\theta}_k(t) = \sum_{p=1}^{N_{\text{cp}}} \alpha_p(t)^T \theta_k(t) \nabla_1 K(X_k(t), c_p(t)) \tag{37}$$

The terms $\partial_{\mathbf{c}^{\text{g}}} G(\mathbf{X}(t), \mathbf{S}(t))$ and $\partial_{\boldsymbol{\alpha}} G(\mathbf{X}(t), \mathbf{S}(t))$ are both matrices of size $3N_{\text{x}} \times 3N_{\text{cp}}$, whose (k,p) block is given respectively by:

$$d_{c_k} G_p = \alpha_k \left(\nabla_1 K(c_k, X_p) \right)^T$$
$$d_{\alpha_k} G_p = K(c_k, X_p) \mathrm{I}_3$$

The differential of the function $F(\mathbf{S}) = \begin{pmatrix} F^c(\mathbf{c}, \boldsymbol{\alpha}) \\ F^\alpha(\mathbf{c}, \boldsymbol{\alpha}) \end{pmatrix}$ can be decomposed into 4 blocks as follows:

$$d_{\mathbf{S}(t)} F = \begin{pmatrix} \partial_{\mathbf{c}} F^c & \partial_{\boldsymbol{\alpha}} F^c \\ \partial_{\mathbf{c}} F^\alpha & \partial_{\boldsymbol{\alpha}} F^\alpha \end{pmatrix} \tag{38}$$

Therefore, the update rules for the auxiliary variables $\boldsymbol{\xi}^c(t)$ and $\boldsymbol{\xi}^\alpha(t)$ are given as:

$$\begin{cases} -\dot{\xi}^c_k(t) = \displaystyle\sum_{p=1}^{N_{\mathrm{x}}} \alpha_k(t)^T \theta_p(t) \nabla_1 K(c_k(t), X_p(t)) \\ \qquad\qquad + (\partial_{\mathbf{c}} F^c)^T \xi^c(t)_k + (\partial_{\mathbf{c}} F^\alpha)^T \xi^\alpha(t)_k \\ -\dot{\xi}^\alpha_k(t) = \displaystyle\sum_{p=1}^{N_{\mathrm{x}}} K(c_k(t), X_p(t)) \theta_p(t) \\ \qquad\qquad + (\partial_\alpha F^c)^T \xi^c(t)_k + (\partial_\alpha F^\alpha)^T \xi^\alpha(t)_k \end{cases}$$

with

$$\begin{aligned} (\partial_c F^c)^T \xi^c(t)_k = &\quad \sum_{p=1}^{N_{\mathrm{cp}}} \Big(\alpha_p(t)^T \xi^c_k(t) + \alpha_k(t)^T \xi^c_p(t) \Big) \nabla_1 K(c_k(t), c_p(t)) \\ (\partial_c F^\alpha)^T \xi^\alpha(t)_k = &\quad \sum_{p=1}^{N_{\mathrm{cp}}} \alpha_k(t)^T \alpha_p(t) \nabla_{1,1} K(c_k(t), c_p(t))^T \Big(\xi^\alpha_p(t) - \xi^\alpha_k(t) \Big) \\ (\partial_\alpha F^c)^T \xi^c(t)_k = &\quad \sum_{p=1}^{N_{\mathrm{cp}}} K(c_k(t), c_p(t)) \xi^c_j(t) \\ (\partial_\alpha F^\alpha)^T \xi^\alpha(t)_k = &\quad \sum_{p=1}^{N_{\mathrm{cp}}} \nabla_1 K(c_k(t), c_p(t))^T \Big(\xi^\alpha_p(t) - \xi^\alpha_k(t) \Big) \alpha_p(t) \end{aligned}$$

In these equations, we supposed the kernel symmetric: $K(x,y) = K(y,x)$. If the kernel is a scalar isotropic kernel of the form $K = f(\|x-y\|^2)\mathrm{I}_3$, then we have:

$$\begin{aligned} &\nabla_1 K(x,y) = 2f'(\|x-y\|^2)(x-y) \\ &\nabla_{1,1} K(x,y) = 4f''(\|x-y\|^2)(x-y)(x-y)^T + 2f'(\|x-y\|^2)\mathrm{I}_3 \end{aligned}$$

Appendix C Gradient of the varifold metric for meshes

We derive here the gradient of the varifold metric with respect to the position of the vertex of the mesh. Let $\mathcal{S}$ be a triangular mesh. For each face f_k, we denote n_k its normal, p_k its center and $u_k = n_k / |n_k|^{1/2}$. Let $\mathcal{T}$ be another triangular mesh, m_k its normal, q_k its center and $v_k = m_k / |m_k|^{1/2}$.

Our goal is to compute the gradient of $d(\mathcal{S},\mathcal{T})^2$ with respect to x_i, a given vertex of $\mathcal{S}$. The chain rule gives:

$$\nabla_{x_i} d(\mathcal{S},\mathcal{T})^2 = \sum_{f_k \ni x_i} (d_{x_i} n_k)^T (d_{n_k} u_k)^T \nabla_{u_k} d(\mathcal{S},\mathcal{T})^2 + (d_{x_i} p_k)^T \nabla_{p_k} d(\mathcal{S},\mathcal{T})^2, \tag{39}$$

where we sum over all the faces that have x_i among their vertices.

Given the inner-product between varifolds (see main text), we have:

$$\nabla_{u_k} d(\mathcal{S},\mathcal{T})^2 = 4 \left(\sum_{i=1}^{N_\mathcal{S}} K^W(p_k,p_i) u_i u_i^T - \sum_{j=1}^{N_\mathcal{T}} K^W(p_k,q_j) v_j v_j^T \right) u_k, \tag{40}$$

and denoting $p_{k,d}$ the dth coordinate of the 3D vector p_k,

$$\left(\nabla_{p_k} d(\mathcal{S},\mathcal{T})^2\right)_d = 2 u_k^T \left(\sum_{i=1}^{N_\mathcal{S}} \frac{\partial K^W(p_k,p_i)}{\partial_{p_{k,d}}} u_i u_i^T - \sum_{j=1}^{N_\mathcal{T}} \frac{\partial K^W(p_k,q_j)}{\partial_{p_{k,d}}} v_j v_j^T \right) u_k \tag{41}$$

Finally, for a face f_k, we have $n_k = \frac{1}{2}(X_1 - X_0) \times (X_2 - X_0)$ and $p_k = \frac{1}{3}(X_0 + X_1 + X_2)$, where we denote X_0, X_1, and X_2 the vertices of the face. If we denote e the edge opposite to the vertex x_i (i.e. $e = X_2 - X_1$ if $x_i = X_0$), we have for a generic 3D-vector V:

$$(d_{x_i} n_k)^T V = \frac{1}{2} e \times V \quad \text{and} \quad (d_{x_i} p_k)^T V = \frac{1}{3} V. \tag{42}$$

and since $u_k = n_k / |n_k|^{1/2}$,

$$d_{n_k} u_k = \frac{1}{|n_k|^{1/2}} \left(\mathrm{I}_3 - \frac{1}{2} \frac{n_k n_k^T}{|n_k|^2} \right) = \frac{1}{|u_k|} \left(\mathrm{I}_3 - \frac{1}{2} \frac{u_k u_k^T}{|u_k|^2} \right) \tag{43}$$

The gradient is computed by plugging (40), (41), (42) and (43) into (39). The gradient is computed by scanning each face of the mesh $\mathcal{S}$ and adding the contribution of this face to each of its vertex.

One can easily check that (39) is independent of the ordering of the vertices, thus showing its invariance with respect to the local orientation of the mesh.

Appendix D Diffeomorphic template evolution

The purpose of this section is to prove that no self-intersection may occur during the optimization of the template shape, by showing that the updates of the template follows a geodesic flow of diffeomorphisms. Using notations of the main text, $\nabla E_{x_{0,p}}$ is the gradient of the criterion with respect to the position of the vertex $x_{0,p}$ of the current template using the L^2 metric, and $\nabla^X E_{x_{0,p}}$ its smoothed version using a metric given by a Gaussian kernel with width $\sigma_X > 0$, K^X, so that:

$$\nabla^X_{x_{0,k}} E = \sum_{p=1}^{N_X} K^X(x_{0,k}, x_{0,p}) \nabla E_{x_{0,p}} = -u_s(x_{0,k}) \,,$$

where u_s is a vector field in V^X, the RKHS associated with the Gaussian kernel K^X. In particular, if ψ_s is the flow associated with integration of u_s, we get $X_0(s) = \psi_s(X_0(0))$. An important point to be verified here is that this flow exists and generates a continuous curve $s \to \psi_s$ of C^1 diffeomorphisms so that the template components cannot degenerate or self-intersect. Let Ω_X be the *open set* of the configurations $\mathbf{X}_0$ such that all the mesh faces associated with $\mathbf{X}_0$ are non degenerated (positive area) and that any pairs of distinct vertices do not coincide in space. The total energy $E(\mathbf{X}_0, \{\mathbf{S}_0^i\})$ is C^1 on an open set $\Omega_X \times \mathbb{R}^{N_S}$ so that the local existence of the gradient descent follows from the Cauchy-Lipschitz theorem. Now, if we consider a maximal solution on $[0, s_f[$, we will prove below (and this is the key estimate) that

$$\int_0^{s_f} |u_s|^2_{V^X} ds \leq E_0 \doteq E(\mathbf{X}_0(0), \{\mathbf{S}_0^i(0)\}) < \infty \tag{44}$$

so that the flow ψ_s is a flow of C^1 diffeomorphisms staying at a bounded distance $d_X(\mathrm{Id}, \psi_s) \leq \sqrt{E_0}$ from the identity and $\mathbf{X}_0(s) = \psi_s(\mathbf{X}_0(0))$ stays in a compact set of Ω_X. In particular, since the differential $d\psi$ and $d\psi^{-1}$ can be controlled uniformly by $d_X(\mathrm{Id}, \psi)$, we get that no face can degenerate during the gradient descent, that the distance between two distinct vertices or two surface patches (up to the continuous limit) cannot vanish.

Now, we prove (44). From the RHKS property of the kernel we get

$$
\begin{aligned}
|u_s|^2_{V^X} &= \sum_{p=1}^{N_{\mathrm{x}}} \left(\nabla E_{x_{0,p}}\right)^T \left(\sum_{q=1}^{N_{\mathrm{x}}} K^X(x_{0,p}, x_{0,q}) \nabla E_{x_{0,q}} \right) \\
&= -\sum_{p=1}^{N_{\mathrm{x}}} \left(\nabla E_{x_{0,p}}\right)^T u_s(x_{0,p}) \\
&\leq -\sum_{p} \left(\nabla E_{x_{0,p}}\right)^T \frac{dx_{0,p}}{ds} \underbrace{-\sum_{i=1}^{N_{\mathrm{su}}} \left(\nabla_{\mathbf{S}_0^i} E\right)^T \frac{d\mathbf{S}_0^i}{ds}}_{\geq 0} = -\frac{dE}{ds}
\end{aligned}
$$

so that $\int_0^{s_f} |u_s|^2_{V^X} ds \leq E(X_0(0)) - E(X_0(s_f)) \leq E(X_0(0))$ (we use here that $E \geq 0$) and $\int_0^{s_f} |u_s|^2_{V^X} ds < \infty$.

Appendix E Computational aspects

E.1 Numerical schemes

The criterion for atlas estimation is minimized using a line search gradient descent method combined with Nesterov's scheme (Nesterov, 1983). Differential equations are integrated using a Euler scheme with prediction correction, also known as Heun's method, which has the same accuracy of Runge-Kutta scheme of order 2. Sums over the control points or over template points are computed using projections on regular lattices and FFTs using the method in Durrleman (2010) for efficiency purposes.

We initialize the gradient descent with template surfaces as ellipsoidal meshes, control points located at the nodes of a regular lattice and momenta set to zero (i.e. no deformation). During the optimization, the number of vertices of the template and the number of control points are preserved. The connectivity between the vertices is also preserved, thus avoiding changes in meshing of the template. The smoothness over the gradient also guarantees that no self-intersection of the template meshes occur since the template evolution is carried by diffeomorphic transport, as shown in Appendix D.

E.2 Parameters

The method depends on the kernel width for the deformation σ_V, for the varifolds σ_W and for the gradient σ_X, as well as the weights σ_k that

balance each data term against the sum of squared geodesic distances between template and observations.

The kernel widths σ_V and σ_W compare with the shape sizes. The varifold kernel width needs to be large enough in order to be sensitive to differences in relative position of a given structure in different subjects (Durrleman, 2010, Ch. 1), otherwise too small values tend to make the shapes orthogonal. In our experiment, the 5 mm value is to account for twist of the hippocampus tail. Would one use the currents metric, the thin tail of the hippocampus would tend to cancel out at this scale (upper and lower part of the tail have opposite orientation), which would make the estimation the twist more challenging. This problem is alleviated here using the varifold metric, in which the contributions of the upper and lower part of the tail are added and not subtracted. The deformation kernel width σ_V compares with the scale of shape variations that one expects to capture. Deformations are built essentially by integrating small translations acting on neighborhood of radius σ_V. It integrates any shape information in such neighborhoods and extrapolates it within the whole neighborhood. As we expect to see local shape variations, we set this parameter to $\sigma_V = 10\,\text{mm}$. With smaller values, the model considers more independent local variations and does not integrate so well the information in larger anatomical regions. With larger values, the model captures more rigid shape differences.

The value of σ_X is essentially a fraction of σ_V: σ_V or $0.5\sigma_V$ work well in practice. In our experiments, we chose $\sigma_X = 0.5\sigma_V$. The weights σ_k are chosen so that data terms have the same order of magnitude as the sum of squared geodesic lengths. Too small values over-weight the importance of data term and prevent the template to converge to the "mean" of the shape set, as the criterion becomes insensitive to sum of squared geodesic distances to each observation. Too large values alter matching accuracy and thus shape features captured by the model. For $\sigma_V = 10\,\text{mm}$ and $\sigma_W = 5\,\text{mm}$, the values of $\sigma_k = 10\,\text{mm}$ for each structure allow the data terms to have roughly the same order of magnitude than the sum of squared geodesic lengths.

A reasonable sampling of control points is reached for a distance between two control points being equal to the deformation kernel width σ_V. Finer sampling often induces a redundant parameterization of the velocity fields as shown in Durrleman (2010). Nonetheless, coarser sampling may be also sufficient for the description of the observed variability, as shown in the presented experiments.

Kernel widths were chosen after few trials to register a pair of shape

complex. The weights σ_k were then assessed while building an atlas with 3 subjects. The initial distribution of the control points was always chosen as the nodes of a regular lattice with step σ_V or down-sampled version of it. We always keep $\sigma_X = 0.5\sigma_V$. We show in Appendix H that similar results are obtained with other values of kernel widths and weights resulting from the same rationale. A qualitative discussion about the effects of these parameters could be found also in Durrleman (2010).

Appendix F Material and Methods

F.1 Data Acquisition

We use data from a project that seeks to link structural changes to the defects in cognition and to the copy number variation on chromosome 21. Inclusion criteria for full trisomy are as follows: clinical DS aged 12-25 with cytogenetic trisomy 21, and inability to complete cognitive studies.

Anatomic T2-weighted images were acquired using a turbo-spin-echo sequence with field-of-view 25.6 mm, imaging matrix 256x256, TR 3.2 sec, TE 499 ms, slice thickness 1mm, 160 axial slices with 10 slice oversampling, signal averages = 1, GRAPPA acceleration factor = 2, bandwidth 574 Hz/Pixel and echo spacing 3.58 ms. Anatomic T1-weighted images were acquired using an MPRAGE sequence with field-of-view 25.6 cm, imaging matrix 256x256, repetition time TR = 2 sec, inversion time TI = 1.1 sec, TE 3.38 ms, slice thickness 1 mm, flip angle 8 degrees, signal averages = 1, 160 axial slices with 20% slice oversampling, water-excitation RF pulse, GRAPPA acceleration factor = 2, bandwidth 300 Hz/pixel and echo spacing 6.3ms.

F.2 Mesh generation and affine alignment

Segmentations of the structures of interest are initialized using an automatic MRI brain tissue segmentation tool[1] based on an expectation maximization algorithm (Van Leemput et al., 1999; Prastawa et al., 2005) that uses registration of a probabilistic brain atlas as priors for segmentation of tissue probabilities from multi-parameter MRI. The tool incorporates coregistration of multi-contrast MRI using mutual information, brain stripping, bias inhomogeneity correction, and tissue classification. The segmentation is then further refined into segmentations of subcortical and lobar

[1] `http://www.nitrc.org/projects/abc`

structures by mapping a parcellation template using non-linear deformation, where we subdivide the brain of each subject via a template that undergoes high-dimensional fluid deformation (Joshi et al., 2004). Reliability tests on repeated MRI of test subjects showed less than 1% variability in resulting regional volumes, which is far better than manual expert segmentations.

The surface meshes representing the subcortical shapes used in this paper was obtained from the final image segmentations using the marching cube surface construction method (Lorensen and Cline, 1987). We extract surface meshes for amygdala, hippocampus, and caudate that are standardized to have the same center of mass and the same general alignment by ensuring that moments of each point cloud in the 3D space are identical. This standardization was performed to isolate local changes across subjects by removing global transformations.

F.3 Atlas construction

For each of the three structures, we perform a Principal Component Analysis (PCA) on the pooled data set. We then map a spherical mesh model with 2562 vertices and 5120 triangle cells (with isotropic sampling) to an ellipsoidal mesh, whose center of mass coincides with the one of the data set, whose axes are aligned with the principal axes of the PCA, and whose semi-axes length equals 1.5 times the square root of the eigenvalues of the PCA.

Input control points are located at the nodes of a regular lattice within data bounding box. For a lattice step of 10 mm, 105 control points are generated (see Appendix E.2). We also consider regular down-sampling of this set of 105 control points yielding sets of 36, 24, 16, 12 and 8 control points.

The atlases are estimated using the deformation kernel width $\sigma_V = 10$ mm, the varifold kernel width $\sigma_W = 5$ mm, the gradient kernel width $\sigma_X = 5$ mm and the weights $\sigma_k = 10$ mm for the three structures, according to the rationale given in Appendix E.2.

We also build atlases considering only one structure at a time. For these experiments, exact same parameters are used, except that the initial lattice number of control points is restricted to a bounding box adjusted to each structure.

Appendix G Statistical analysis

For a given atlas, we denote $\mathbf{c} = \{c_k\}_{k=1,\dots,N_{\text{cp}}}$ the optimized position of the control points (concatenated in a $3N_{\text{cp}}$ dimensional vector) and $\boldsymbol{\alpha}^i = \{\alpha_k^i\}_{k=1,\dots,N_{\text{cp}}}$ the estimated initial momentum vectors (also concatenated in a $3N_{\text{cp}}$ dimensional vector) that parametrize the deformation of the template to the ith subject's shape complex ($i = 1, \dots, N_{\text{su}}$).

The velocities associated to the momentum vectors are given as $\mathbf{v}^i = \mathbf{K}\boldsymbol{\alpha}^i$, where $\mathbf{K}$ is $3N_{\text{su}} \times 3N_{\text{su}}$ block-matrix whose (p,q)-block is given by $K^V(c_p, c_q)\mathrm{I}_3$.

G.1 Means and covariance matrices

The sample mean of the controls (resp. the DS subjects) is defined by $\overline{\boldsymbol{\alpha}}^{HC} = \frac{1}{N_{\text{su}}^{HC}} \sum_{i \in HC} \boldsymbol{\alpha}^i$ (resp. $\overline{\boldsymbol{\alpha}}^{DS} = \frac{1}{N_{\text{su}}^{DS}} \sum_{i \in DS} \boldsymbol{\alpha}^i$) where HC (resp. DS) denotes the set of indices corresponding to healthy controls (resp. Down syndrome subjects). The typical anatomical configuration of the controls (resp. the DS subjects) is obtained by solving the geodesic shooting equations along the direction of $\overline{\boldsymbol{\alpha}}^{HC}$ (resp. $\overline{\boldsymbol{\alpha}}^{DS}$) and deforming the template complex with the corresponding flow of diffeomorphisms.

The estimated template complex is "centered", in the sense that the mean of all momentum vectors $\overline{\boldsymbol{\alpha}} = (\overline{\boldsymbol{\alpha}}^{DS} + \overline{\boldsymbol{\alpha}}^{HC})/2$ is small with respect to the means of each sub-group and to the variance. Denoting σ^2 the total variance explained (i.e. the trace of the matrix $(\langle \boldsymbol{\alpha}^i - \overline{\boldsymbol{\alpha}}, \boldsymbol{\alpha}^j - \overline{\boldsymbol{\alpha}} \rangle_{W^*} / N_{\text{su}} = (\boldsymbol{\alpha}^i - \overline{\boldsymbol{\alpha}})^T (\mathbf{v}^i - \overline{\mathbf{v}})/N_{\text{su}})_{i,j}$), we have:

$$\frac{\|\overline{\boldsymbol{\alpha}}\|_{W^*}}{\sigma} = 0.19 \tag{45}$$

for the atlas using 105 control points and 0.08 for the atlas using 8 control points. In comparison, the means of each sub-group satisfy:

$$\frac{\left\|\overline{\boldsymbol{\alpha}}^{HC}\right\|_{W^*} + \left\|\overline{\boldsymbol{\alpha}}^{DS}\right\|_{W^*}}{2\sigma} = 0.51 \quad \frac{\left\|\overline{\boldsymbol{\alpha}}^{DS} - \overline{\boldsymbol{\alpha}}^{HC}\right\|_{W^*}}{\sigma} = 0.94 \tag{46}$$

for the atlas using 105 control points and $0.51, 0.99$ for the atlas using 8 control points. This shows that the pooled mean is nearly zero, the two subgroup means are pointing in almost opposite direction and the distance between both means compare with the standard deviation of the data set.

We estimate the covariance matrix as:

$$\Sigma = \frac{1}{N_{\text{su}}} \left(\sum_{i \in HC} (\mathbf{v}^i - \overline{\mathbf{v}}^{HC}) \left(\mathbf{v}^i - \overline{\mathbf{v}}^{HC}\right)^T + \sum_{i \in DS} (\mathbf{v}^i - \overline{\mathbf{v}}^{DS})(\mathbf{v}^i - \overline{\mathbf{v}}^{DS})^T \right).$$

Note that if the number of observations is smaller than 3 times the number of control points, then Σ is not invertible, and we use instead the regularized matrix $\Sigma + \varepsilon \mathrm{I}_3$. In practice, we used $\varepsilon = 10^{-2}$, which leads to a condition number of the covariance matrix of order 1000. We show in Appendix H the robustness of the classification scores when this value is increased.

Remark G.1. Note that we perform the statistical analysis using the velocity field sampled at the control points $\mathbf{v} = \mathbf{K}\boldsymbol{\alpha}$, whereas it would seem more natural to use the RKHS metric on V instead. Using the RKHS metric amounts to use $\tilde{\mathbf{v}} = \mathbf{K}^{1/2}\boldsymbol{\alpha}$ so that one computes the inner-product as $(\tilde{\mathbf{v}}^i)^T\tilde{\mathbf{v}}^j = \boldsymbol{\alpha}^{iT}\mathbf{K}\boldsymbol{\alpha}^j$ which is the inner-product between the two continuous velocity fields in V. One can easily check that without regularization ($\varepsilon = 0$), the most discriminant axis is the same in both cases, as will be the LDA and ML classification criteria introduced in the sequel. Using the identity matrix as a regularizer for the sample covariance matrix above amounts to use the matrix $\mathbf{K}^{-1}$ as regularizer in the RKHS space. More precisely, the matrix $\Sigma + \varepsilon \mathrm{I}_3$ becomes $\tilde{\Sigma} + \varepsilon \mathbf{K}^{-1}$ where $\tilde{\Sigma}$ is the sample covariance matrix of the $\tilde{\mathbf{v}}^i$'s. It is natural to use this regularizer, since the criterion for atlas construction precisely assumes the momentum vector to be distributed with a zero-mean Gaussian distribution with covariance matrix $\mathbf{K}^{-1}$ (which leads to $\|v_0^i\|_V^2 = \boldsymbol{\alpha}_0^{iT}\mathbf{K}\alpha_0^i$ in (17)). For this reason, the same matrix is used in Allassonnière et al. (2007) as a prior in a Bayesian estimation framework.

G.2 Most Discriminative Axis

According to linear discriminant analysis, the two directions of the most discriminative axis in the velocity space is defined as $\mathbf{v}_{\pm}^{LDA} = \overline{\mathbf{v}} \pm \Sigma^{-1}(\overline{\mathbf{v}}^{HC} - \overline{\mathbf{v}}^{DS})$ where $\overline{\mathbf{v}} = \frac{1}{2}(\overline{\mathbf{v}}^{HC} + \overline{\mathbf{v}}^{DS})$. The associated momentum vectors are given as: $\boldsymbol{\alpha}_{\pm}^{LDA} = \mathbf{K}^{-1}\mathbf{v}_{\pm}^{LDA}$. The anatomical configurations are generated by solving for the geodesic equations in the two directions $\boldsymbol{\alpha}_{\pm}^{LDA}$ and deforming the template complex along these flows of diffeomorphisms. We normalize the direction, so that $\left\|\boldsymbol{\alpha}_{\pm}^{LDA}\right\|_{W^*} = \sqrt{(\boldsymbol{\alpha}_{\pm}^{LDA})^T\mathbf{K}(\boldsymbol{\alpha}_{\pm}^{LDA})} = \sqrt{(\boldsymbol{\alpha}_{\pm}^{LDA})^T(\mathbf{v}_{\pm}^{LDA})}$ equals the norm between the means: $\left\|\overline{\boldsymbol{\alpha}}^{HC} - \overline{\boldsymbol{\alpha}}^{DS}\right\|_{W^*}$. Therefore, the sum of the geodesic distance between the template complex and each of the deformed complex is twice the norm between the means.

G.3 Statistical significance using permutation tests

With the notations of the previous sections, we denote Σ the estimate of the within-class covariance matrix (without regularization in this section, i.e. $\varepsilon = 0$). We denote (u_k, λ_k^2) the eigenvectors and eigenvalues sorted in decreasing order of the matrix Σ. We truncate the matrix up to the N_{modes} largest eigenvalues: $\tilde{\Sigma} = \sum_{k=1}^{N_{\text{modes}}} \lambda_k^2 u_k u_k^T$ whose inverse is given by: $\tilde{\Sigma}^{-1} = \sum_{k=1}^{N_{\text{modes}}} \frac{1}{\lambda_k^2} u_k u_k^T$. We compute then the T^2 Hotelling statistics as:

$$T^2 = \frac{N_{\text{su}} - 2}{4} (\overline{\mathbf{v}}^{HC} - \overline{\mathbf{v}}^{DS})^T \tilde{\Sigma}^{-1} (\overline{\mathbf{v}}^{HC} - \overline{\mathbf{v}}^{DS})$$

To estimate the distribution of the statistics under the null hypothesis of equal means, we compute the statistics for 10^5 permutations of the subjects indices i. This implies that the truncation of the matrix Σ selects a different small-dimensional space for each permutation. The corresponding p-values are shown in Fig. G.1 for each model (i.e. different number of control points) and different number N_{modes} of modes. Note that the covariance matrix Σ has only N_{su} non-zero eigenvalues and in all our cases has a bad conditioning for $N_{\text{modes}} \geq N_{\text{su}} - 2$. This is the reason why plots in Fig. G.1 stop at $N_{\text{modes}} = 14$.

We repeat the procedure for two other values of the deformation parameter $\sigma_V = 15$ mm and $\sigma_V = 5$ mm while keeping other parameters fixed. Results are reported in Supplementary Data Fig. S4 and S5 respectively.

The plots in Fig. 5 are built by picking among the p-values the ones that correspond to the number of modes which explains 95% of the trace of the matrix Σ for each number of control points. In all experiments, this number of modes was either 8 or 9.

G.4 Classification in a leave-2-out setting

With the notations of the previous sections, we compute the regularized sample means and covariance matrix Σ from the atlas built with 7 controls and 7 DS subjects (among 8 controls and 8 DS subjects). We then register the template complex to the left-out shape complexes. Let us denote $\boldsymbol{\alpha}^{test}$ the initial momenta parameterizing the deformation of the template shape complex to a given left-out shape complex (seen as a test data), and $\mathbf{v}^{test} = \mathbf{K}\boldsymbol{\alpha}^{test}$. In the framework of the Linear Discriminant Analysis, the classification criterion writes:

$$C(\mathbf{v}^{test}) = (\mathbf{v}^{test} - \overline{\mathbf{v}})^T \Sigma^{-1} (\overline{\mathbf{v}}^{HC} - \overline{\mathbf{v}}^{DS}) \tag{47}$$

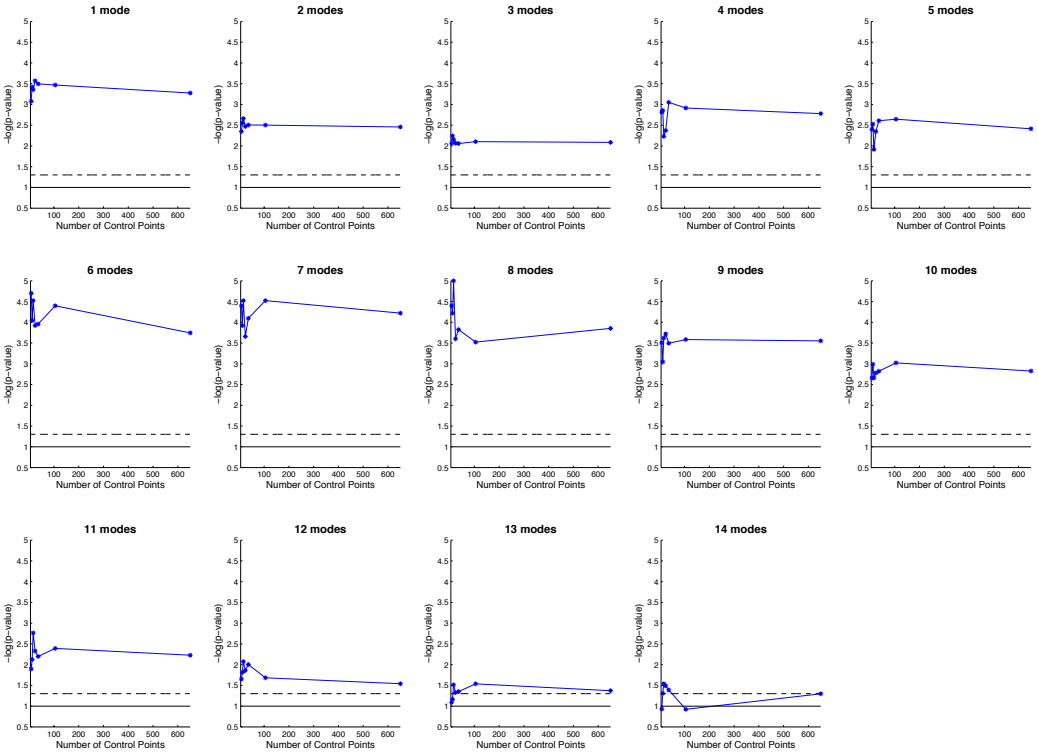


Figure G.1: P-values computed for different number of control points and different number of selected modes for deformation parameter $\boldsymbol{\sigma_V} = \mathbf{10\ mm}$. Solid (resp. dashed) lines corresponds to the p-value of 10% (resp 5%)

For a threshold η, the test data is classified as healthy control if $C(\mathbf{v}^{test}) > \eta$ and Down syndrome subject otherwise. ROC curves are built when the threshold η is varied. For estimating classification scores, we estimate the threshold η on the training data set so that the best separating hyperplane (orthogonal to the most discriminative axis $\Sigma^{-1}(\overline{\mathbf{v}}^{HC} - \overline{\mathbf{v}}^{DS})$) is positioned at equal distance to the two classes. This threshold value is used for classifying the test data.

For classifying in a Maximum Likelihood framework, we compute the sample covariance matrices $\Sigma_{DS} = \frac{1}{N_{\text{su}}^{Ds}} \sum_{i \in DS} (\mathbf{v}^i - \overline{\mathbf{v}}^{DS})(\mathbf{v}^i - \overline{\mathbf{v}}^{DS})^T$ and $\Sigma_{HC} = \frac{1}{N_{\text{su}}^{HC}} \sum_{i \in HC} (\mathbf{v}^i - \overline{\mathbf{v}}^{HC})(\mathbf{v}^i - \overline{\mathbf{v}}^{HC})^T$. The classification criterion, also called Mahalanobis distance, is given by:

$$C(\mathbf{v}^{test}) = (\mathbf{v}^{test} - \overline{\mathbf{v}}^{DS})^T \Sigma_{DS}^{-1} (\mathbf{v}^{test} - \overline{\mathbf{v}}^{DS}) - (\mathbf{v}^{test} - \overline{\mathbf{v}}^{HC})^T \Sigma_{HC}^{-1} (\mathbf{v}^{test} - \overline{\mathbf{v}}^{HC}) \tag{48}$$

and the classification rule remains the same.

Note that in both cases, we used regularized covariance matrices by adding $\varepsilon \mathrm{I}_3$ for $\varepsilon = 10^{-2}$ (see Remark G.1).

G.5 Composite Analysis

In the composite analysis, we compute an atlas for each of the three structures. We end up with 3 sets of control points $\mathbf{c}_s$, and $3 \times N_{\text{su}}$ sets of initial momenta $\boldsymbol{\alpha}_s^i$ for $s = 1, 2, 3$. We denote $\mathbf{v}_s^i = \mathbf{K}_s \boldsymbol{\alpha}_s^i$ the associated velocities and build a composite shape descriptor $\mathbf{v}^i$ by concatenating $\mathbf{v}_1^i$, $\mathbf{v}_2^i$ and $\mathbf{v}_3^i$. We use this composite descriptor to compute means, sample covariance matrices, most discriminative axis and classification criteria as above.

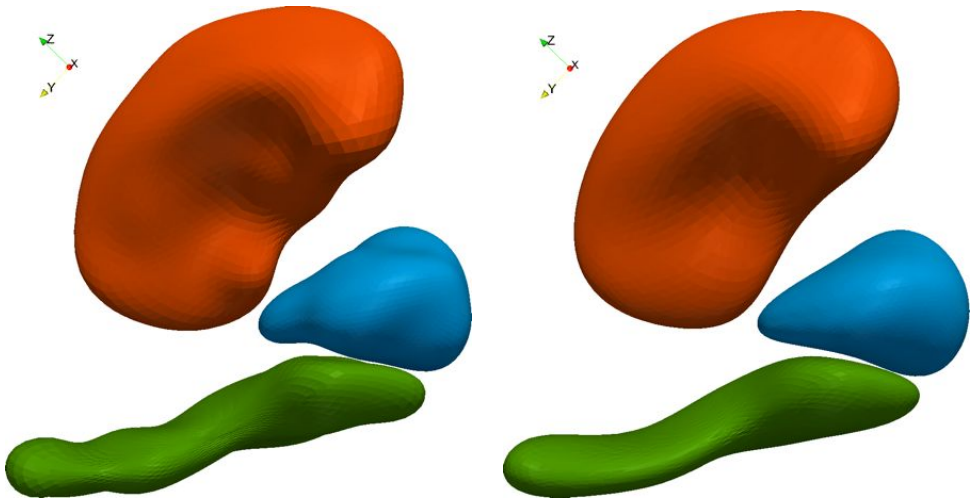
The composite vector $\mathbf{v}$ does not parameterize a single diffeomorphism, but each of its three component does. For instance, the direction of the most discriminative axis $\mathbf{v}^{LDA}$ takes into account the correlations between the variability of each of the components. To display these correlations, we compute the initial momentum vectors associated to each component: $\boldsymbol{\alpha}_s^{LDA} = \mathbf{K}_s^{-1} \mathbf{v}_s^{LDA}$ for $s = 1, 2, 3$, and then solve three geodesic shooting equations and deform each structure using a different diffeomorphism.

Appendix H Robustness of the results with respect to parameter values

We assess the robustness of the results with respect to parameter values. The values were first chosen given the rationale in Appendix E.2. Then, we change the values of the deformation and varifold kernel widths by $\pm 50\%$, namely by setting $\sigma_V = 5, 10$ or $15\,\text{mm}$ and $\sigma_W = 2.5, 5,$ or $7.5\,\text{mm}$. Other parameters are kept fixed, namely the weights $\sigma_k = 10\,\text{mm}$, the gradient kernel width $\sigma_X = 0.5\sigma_V$ and the initial distance between control points equal to σ_V. Classification scores are reported in Table H.1 and show a great robustness of the statistical estimates, noticeably for the ML method. We note a decrease in the specificity in the LDA classifier for the large deformation kernel width $\sigma_V = 15\,\text{mm}$. With large deformation kernel widths, the atlas captures more global shape variations, which might not be as discriminative as more local changes. This effect is more pronounced with increased varifold width σ_W, as surface matching accuracy decreases, thus reducing the variability captured in the atlas. The number of generated control points was equal to 650 for $\sigma_V = 5\,\text{mm}$, 105 for $\sigma_V = 10\,\text{mm}$ and 15 for $\sigma_V = 40\,\text{mm}$.

We show in Fig. H.1 two template complexes, one estimated with $\sigma_V = 5\,\text{mm}$ and the other one with $\sigma_V = 15\,\text{mm}$, while fixing the other parameters to $\sigma_W = 7.5\,\text{mm}$, $\sigma_k = 10\,\text{mm}$ for the three structures and $\sigma_X = 0.5\sigma_V$. One notices that the atlas captures more local shape variations with the smaller parameter, whereas it captures more global and rigid variations with the larger parameter, thus leading to much smoother surfaces.

We also assess the influence of the amount of regularization in the covariance matrices ε, which otherwise are singular. In the previous experiments, we chose $\varepsilon = 10^{-2}$, for which the regularized covariance matrices had a condition number of order 1000. We increase the regularization to $\varepsilon = 0.1$ and $\varepsilon = 1$, for which the condition number is of order 100 and 10 respectively. Classification scores still show the robustness of the method with respect to this parameter (Supplementary Data Table S1 and S2). The sensitivity in the LDA approach is also decreased if the amount of regularization is increased. This is not surprising, since the effect of the regularization, like increase in deformation kernel width, tends to reduce the sensitivity of the statistical model. For $\varepsilon = 10$, the condition number is of order 1, meaning that the regularization is of the same order as the largest eigenvalues of the sample covariance matrix, which obviously alters classification scores (Supplementary Table S3).



a - $\sigma_V = 5\,\text{mm}$ b- $\sigma_V = 15\,\text{mm}$

Figure H.1: Template shape complex estimated with two different deformation kernel widths σ_V. The smaller the width, the more local the variations captured by the model. The larger the width, the more global and rigid variations captured by the model, resulting in much smoother surfaces

		LDA		ML	
		specificity	sensitivity	specificity	sensitivity
$\sigma_V = 5$	$\sigma_W = 2.5$	98 (63/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	98 (63/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	98 (63/64)	100 (64/64)	100 (64/64)	100 (64/64)
$\sigma_V = 10$	$\sigma_W = 2.5$	98 (63/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	98 (63/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	94 (60/64)	100 (64/64)	100 (64/64)	100 (64/64)
$\sigma_V = 15$	$\sigma_W = 2.5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	83 (53/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	84 (54/64)	100 (64/64)	100 (64/64)	100 (64/64)

Table H.1: Classification scores when deformation and varifold kernel sizes are varied. Regularization of the covariance matrices $\varepsilon = 10^{-2}$.

It is clear that the weights σ_k's should have been also adjusted. As noted in Akin and Mumford (2012), adjusting the weights could increase matching accuracy, and possibly increase statistical performance when deformation

smoothness is varied. We did not adapt these values for several reasons. First, not adjusting values is the least favorable case, thus strengthening our conclusions regarding robustness of the method with respect to parameter values. Second, our rationale for setting these weights, as given in Appendix E.2, requires to perform few tests on a reduced data set to assess the relative importance of the data term with respect to the sum of squared geodesic distances. We mimic here a use of the method on a routine basis where one could not finely tune parameters. Last but not least, it is clear from a statistical point of view, that these weights are noise variance in data. For a given model (σ_V and σ_W fixed), there is an optimal value of this noise variance that should be estimated from data, and not fixed by the user. The estimation of such parameter for image data has been proposed in Allassonnière et al. (2007, 2010) using a Bayesian framework. Extending this framework for surfaces modeled as varifolds is not straightforward. Nonetheless, we believe that the determination of these weights should be done in this framework.

References

Akin, A., Mumford, D., 2012. "You laid out the lands:" georeferencing the Chinese Yujitu [Map of the Tracks of Yu] of 1136. Cartography and Geographic Information Science 39, 154–169.

Allassonnière, S., Amit, Y., Trouvé, A., 2007. Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society Series B 69, 3–29.

Allassonnière, S., Kuhn, E., 2009. Stochastic algorithm for bayesian mixture effect template estimation. ESAIM Probability and Statistics In Press.

Allassonnière, S., Kuhn, E., Trouvé, A., 2010. Construction of bayesian deformable models via a stochastic approximation algorithm: A convergence study. Bernoulli Journal 16, 641–678.

Bookstein, F., 1991. Morphometric tools for landmark data: geometry and biology. Cambridge University Press.

Bouix, S., Pruessner, J.C., Collins, D.L., Siddiqi, K., 2005. Hippocampal shape analysis using medial surfaces. NeuroImage 25, 1077 – 1089. doi:10.1016/j.neuroimage.2004.12.051.

Boyer, D.M., Lipman, Y., Clair, E.S., Puente, J., Patel, B.A., Funkhauser, T., Jernvall, J., Daubechies, I., 2010. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. Proc of Natl Acad Sci USA 108, 18221–18226.

Charon, N., Trouvé, A., 2013. The varifold representation of non-oriented shapes for diffeomorphic registration. CoRR abs/1304.6108 URL: http://adsabs.harvard.edu/abs/2013arXiv1304.6108C.

Chung, M.K., Worsley, K.J., Robbins, S., Paus, T., Taylor, J., Giedd, J.N., Rapoport, J.L., Evans, A.C., 2003. Deformation-based surface morphometry applied to gray matter deformation. NeuroImage 18, 198 – 213. doi:10.1016/S1053-8119(02)00017-4.

Dryden, I., Mardia, K., 1998. Statistical Shape Analysis. Wiley.

Durrleman, S., 2010. Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Thèse de sciences (phd thesis). Université de Nice-Sophia Antipolis.

Durrleman, S., Allassonnière, S., Joshi, S., 2013. Sparse adaptive parameterization of variability in image ensembles. Int J Comput Vision 101, 161–183. URL: http://dx.doi.org/10.1007/s11263-012-0556-1, doi:10.1007/s11263-012-0556-1.

Durrleman, S., Pennec, X., Trouvé, A., Ayache, N., 2009. Statistical models of sets of curves and surfaces based on currents. Med Image Anal 13, 793–808.

Durrleman, S., Prastawa, M., Korenberg, J.R., Joshi, S., Trouvé, A., Gerig, G., 2012. Topology preserving atlas construction from shape data without correspondence using sparse parameters, in: Ayache, N., Delingette, H., Golland, P., Mori, K. (Eds.), Med Image Comput Comput Assist Interv., Springer. pp. 223–230.

Glaunès, J., Joshi, S., 2006. Template estimation from unlabeled point set data and surfaces for computational anatomy.

Gorczowski, K., Styner, M., Jeong, J.Y., Marron, J.S., Piven, J., Hazlett, H.C., Pizer, S.M., Gerig, G., 2010. Multi-object analysis of volume, pose,

and shape using statistical discrimination. IEEE Trans. Pattern Anal. Mach. Intell. 32, 652 – 661.

Grenander, U., 1994. General Pattern Theory: a Mathematical Theory of Regular Structures. Oxford University Press.

Joshi, S., Davis, B., Jomier, M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160.

Korbel, J.O., Tirosh-Wagner, T., Urban, A.E., Chen, X.N., Kasowski, M., Dai, L., Grubert, F., Erdman, C., Gao, M.C., Lange, K., Sobel, E.M., Barlow, G.M., Aylsworth, A.S., Carpenter, N.J., Clark, R.D., Cohen, M.Y., Doran, E., Falik-Zaccai, T., Lewin, S.O., Lott, I.T., McGillivray, B.C., Moeschler, J.B., Pettenati, M.J., Pueschel, S.M., Rao, K.W., Shaffer, L.G., Shohat, M., Van Riper, A.J., Warburton, D., Weissman, S., Gerstein, M.B., Snyder, M., Korenberg, J.R., 2009. The genetic architecture of down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. Proc of Natl Acad Sci USA 106, 12031–12036. doi:10.1073/pnas.0813248106.

Korenberg, J.R., Chen, X.N., Schipper, R., Sun, Z., Gonsky, R., Gerwehr, S., Carpenter, N., Daumer, C., Dignan, P., Disteche, C., 1994. Down syndrome phenotypes: the consequences of chromosomal imbalance. Proc of Natl Acad Sci USA 91, 4997–5001. URL: `http://www.pnas.org/content/91/11/4997.abstract`.

Lorensen, W.E., Cline, H.E., 1987. Marching cubes: A high resolution 3d surface construction algorithm, in: ACM Siggraph Computer Graphics, ACM. pp. 163–169.

Ma, J., Miller, M.I., Trouvé, A., Younes, L., 2008. Bayesian template estimation in computational anatomy. NeuroImage 42, 252 – 261. doi:DOI: 10.1016/j.neuroimage.2008.03.056.

McLachlan, R.I., Marsland, S., 2007. Discrete mechanics and optimal control for image registration. ANZIAM Journal 48, C1–C16. URL: `http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/82`.

Miller, M., Trouvé, A., Younes, L., 2006. Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision 24, 209–228.

Nesterov, Y.E., 1983. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. Soviet Math. Dokl. 27. Translation by A. Rosa.

Pennec, X., 2006. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. Journal of Mathematical Imaging and Vision 25, 127–154.

Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G., 2005. Automatic segmentation of mr images of the developing newborn brain. Medical Image Analysis 9, 457–466.

Reuter, M., Wolter, F.E., Peinecke, N., 2006. Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. Comput. Aided Des. 38, 342–366.

Styner, M., Lieberman, J.A., McClure, R.K., Weinberger, D.R., Jones, D.W., Gerig, G., 2005. Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. Proc of Natl Acad Sci USA 102, 4872–4877.

Vaillant, M., Glaunès, J., 2005. Surface matching via currents, pp. 381–392.

Vaillant, M., Miller, M., Younes, L., Trouvé, A., 2004. Statistics on diffeomorphisms via tangent space representations. NeuroImage 23, 161–169.

Vaillant, M., Qiu, A., Glaunès, J., Miller, M., 2007. Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus. NeuroImage 34, 1149–1159.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based tissue classification of mr images of the brain. Medical Imaging, IEEE Transactions on 18, 897–908.

Zeidler, E., 1991. Applied Functional Analysis: Application to Mathematical Physics. Springer.

Supplementary Data

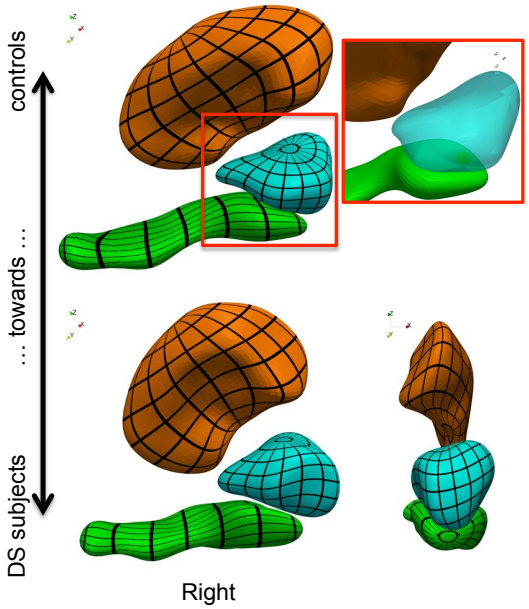


Figure S1: Most Discriminative Axis computed using a composite descriptor. The direction takes into account the correlations among the three structures. However, it does not parameterize a single space deformation, but three of them. Therefore, intersections between surfaces occur. Moreover the patterns of shape variations are rather different from the results using a single atlas of shape complex, in particular the relative position of the amygdala (in blue) with respect to the hippocampus (in green)

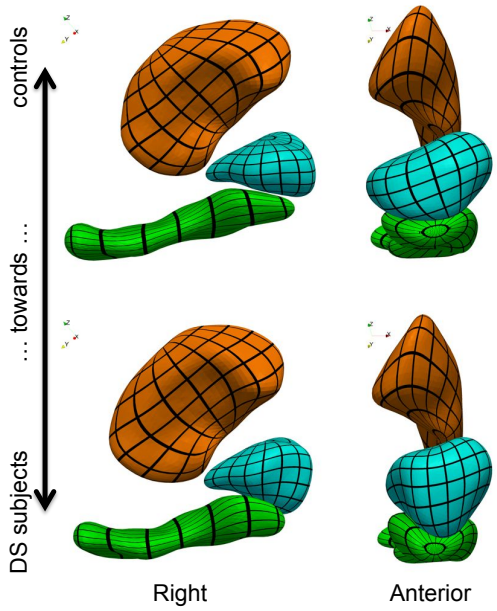


Figure S2: Most Discriminative Axis in the atlas with 8 control points. The patterns of shape variations are qualitatively similar with the axis shown using 105 control points, especially for the hippocampus and amygdala (in green and cyan), and to a lesser extend for the putamen. This shows the robustness of the findings with respect to different initial conditions

		LDA		ML	
		specificity	sensitivity	specificity	sensitivity
$\sigma_V = 5$	$\sigma_W = 2.5$	98 (63/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	97 (62/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	97 (62/64)	100 (64/64)	100 (64/64)	100 (64/64)
$\sigma_V = 10$	$\sigma_W = 2.5$	98 (63/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	97 (62/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	92 (59/64)	100 (64/64)	100 (64/64)	100 (64/64)
$\sigma_V = 15$	$\sigma_W = 2.5$	83 (53/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	83 (53/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	84 (54/64)	100 (64/64)	100 (64/64)	100 (64/64)

Table S1: Classification scores when deformation and varifold kernel sizes are varied. Regularization of the covariance matrices $\varepsilon = 0.1$.

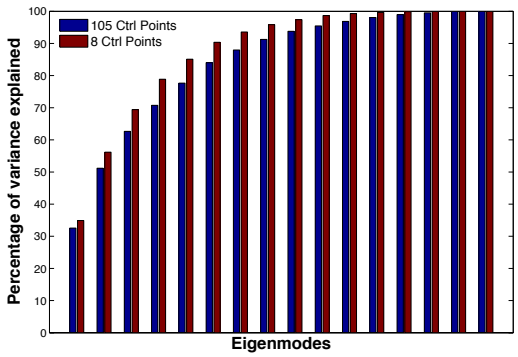


Figure S3: Cumulative variance explained using the sample covariance matrix of the momentum vectors. The spectrum is slightly more concentrated with 8 control points than with 105. The total variance explained in both cases is similar: $\sigma^2 = 27.1$ for 105 points and $\sigma^2 = 23.6$ for 8 points

		LDA		ML	
		specificity	sensitivity	specificity	sensitivity
$\sigma_V = 5$	$\sigma_W = 2.5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
$\sigma_V = 10$	$\sigma_W = 2.5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
$\sigma_V = 15$	$\sigma_W = 2.5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)
	$\sigma_W = 7.5$	89 (57/64)	100 (64/64)	100 (64/64)	100 (64/64)

Table S2: Classification scores when deformation and varifold kernel sizes are varied. Regularization of the covariance matrices $\varepsilon = 1$.

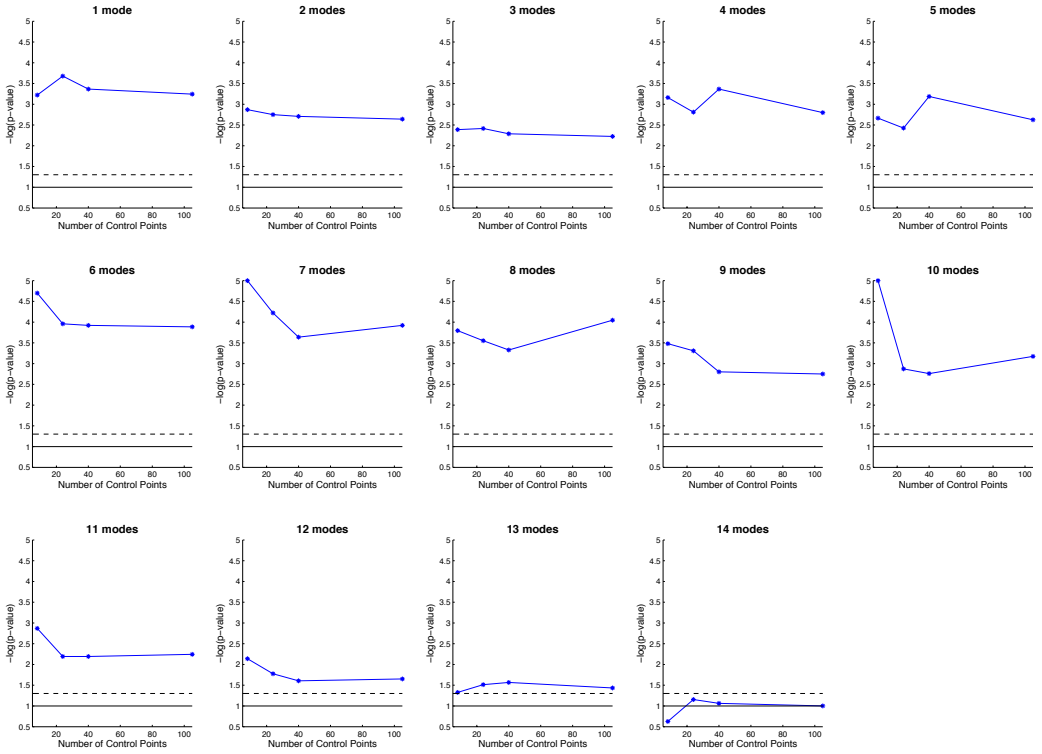


Figure S4: P-values computed for different number of control points and different number of selected modes for deformation parameter $\boldsymbol{\sigma_V = 15}$ **mm**. Solid (resp. dashed) lines corresponds to the p-value of 10% (resp 5%)

		LDA		ML	
		specificity	sensitivity	specificity	sensitivity
$\sigma_V = 5$	$\sigma_W = 2.5$	88 (56/64)	92(59/64)	88 (56/64)	97(62/64)
	$\sigma_W = 5$	88 (56/64)	92(59/64)	88 (56/64)	97(62/64)
	$\sigma_W = 7.5$	88 (56/64)	95(61/64)	88 (56/64)	98(63/64)
$\sigma_V = 10$	$\sigma_W = 2.5$	88 (56/64)	92(59/64)	88 (56/64)	98(63/64)
	$\sigma_W = 5$	88 (56/64)	92(59/64)	88 (56/64)	98(63/64)
	$\sigma_W = 7.5$	88 (56/64)	98(63/64)	88 (56/64)	100(64/64)
$\sigma_V = 15$	$\sigma_W = 2.5$	88 (56/64)	89(57/64)	88 (56/64)	100(64/64)
	$\sigma_W = 5$	88 (56/64)	94(60/64)	88 (56/64)	100(64/64)
	$\sigma_W = 7.5$	88 (56/64)	94(60/64)	88 (56/64)	100(64/64)

Table S3: Classification scores when deformation and varifold kernel sizes are varied. Regularization of the covariance matrices $\varepsilon = 10$.

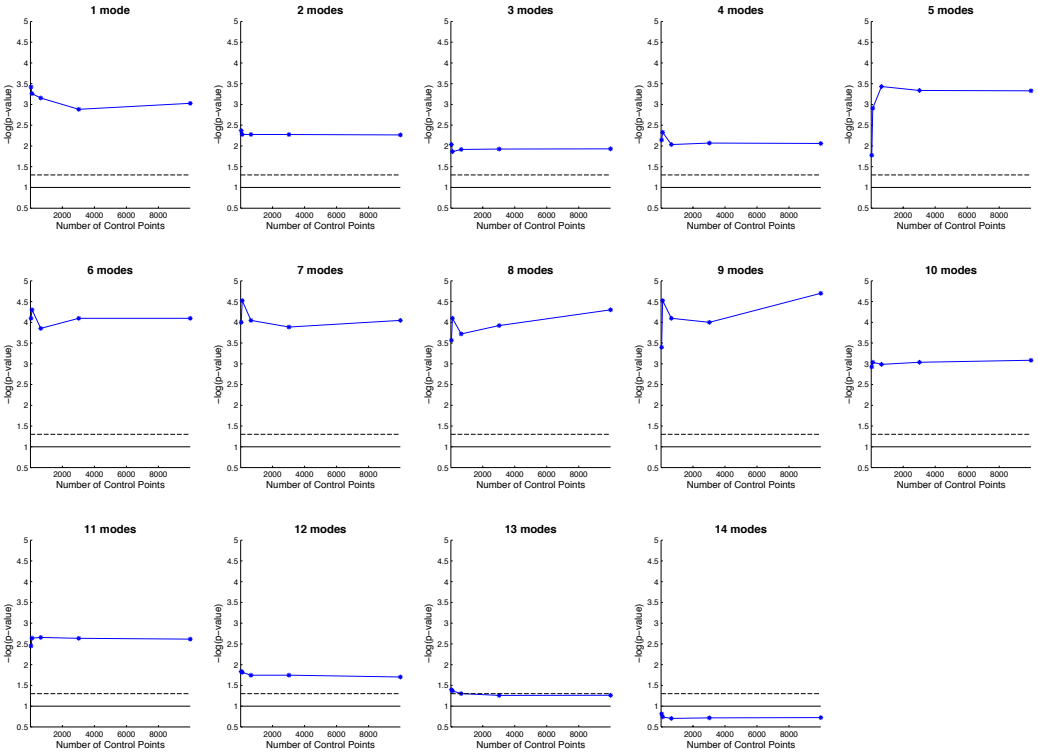


Figure S5: P-values computed for different number of control points and different number of selected modes for deformation parameter $\boldsymbol{\sigma_V} = \mathbf{5\ mm}$. Solid (resp. dashed) lines corresponds to the p-value of 10% (resp 5%)