



## Statistical Shape Analysis of Large Datasets Based on Diffeomorphic Iterative Centroids

Claire Cury, Joan Glaunès, Roberto Toro, Marie Chupin, Gunter Schumann,  
Vincent Frouin, Jean-Baptiste Poline, Olivier Colliot

### ► To cite this version:

Claire Cury, Joan Glaunès, Roberto Toro, Marie Chupin, Gunter Schumann, et al.. Statistical Shape Analysis of Large Datasets Based on Diffeomorphic Iterative Centroids. *Frontiers in Neuroscience*, 2018, 12, pp.803. 10.3389/fnins.2018.00803 . hal-01920263v2

HAL Id: hal-01920263

<https://inria.hal.science/hal-01920263v2>

Submitted on 13 Nov 2018

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

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



Distributed under a Creative Commons Attribution 4.0 International License



# Statistical Shape Analysis of Large Datasets Based on Diffeomorphic Iterative Centroids

Claire Cury<sup>1,2,3,4,5,6\*</sup>, Joan A. Glaunès<sup>7</sup>, Roberto Toro<sup>8,9</sup>, Marie Chupin<sup>1,2,3,4,5</sup>, Gunter Schumann<sup>10</sup>, Vincent Frouin<sup>11</sup>, Jean-Baptiste Poline<sup>12</sup>, Olivier Colliot<sup>1,2,3,4,5</sup> and the Imagen Consortium<sup>†</sup>

<sup>1</sup> Institut du Cerveau et de la Moelle épinière, ICM, Paris, France, <sup>2</sup> Inserm, U 1127, Paris, France, <sup>3</sup> CNRS, UMR 7225, Paris, France, <sup>4</sup> Sorbonne Université, Paris, France, <sup>5</sup> Inria, Aramis project-team, Paris, France, <sup>6</sup> Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, VISAGES ERL U 1228, Rennes, France, <sup>7</sup> MAP5, Université Paris Descartes, Sorbonne Paris Cité, Paris, France, <sup>8</sup> Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France, <sup>9</sup> CNRS URA 2182 "Genes, Synapses and Cognition", Paris, France, <sup>10</sup> MRC-Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom, <sup>11</sup> Neurospin, Commissariat à l'Energie Atomique et aux Energies Alternatives, Paris, France, <sup>12</sup> Henry H. Wheeler Jr. Brain Imaging Center, University of California, Berkeley, California City, CA, United States

## OPEN ACCESS

### Edited by:

John Ashburner,  
University College London,  
United Kingdom

### Reviewed by:

James Fishbaugh,  
NYU Tandon School of Engineering,  
United States  
Ali R. Khan,  
University of Western Ontario, Canada

### \*Correspondence:

Claire Cury  
claire.cury.pro@gmail.com

<sup>†</sup><http://www.imagen-europe.com>

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

Received: 15 June 2018

Accepted: 16 October 2018

Published: 12 November 2018

### Citation:

Cury C, Glaunès JA, Toro R, Chupin M, Schumann G, Frouin V, Poline J-B, Colliot O and the Imagen Consortium (2018) Statistical Shape Analysis of Large Datasets Based on Diffeomorphic Iterative Centroids. *Front. Neurosci.* 12:803. doi: 10.3389/fnins.2018.00803

In this paper, we propose an approach for template-based shape analysis of large datasets, using diffeomorphic centroids as atlas shapes. Diffeomorphic centroid methods fit in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework and use kernel metrics on currents to quantify surface dissimilarities. The statistical analysis is based on a Kernel Principal Component Analysis (Kernel PCA) performed on the set of initial momentum vectors which parametrize the deformations. We tested the approach on different datasets of hippocampal shapes extracted from brain magnetic resonance imaging (MRI), compared three different centroid methods and a variational template estimation. The largest dataset is composed of 1,000 surfaces, and we are able to analyse this dataset in 26 h using a diffeomorphic centroid. Our experiments demonstrate that computing diffeomorphic centroids in place of standard variational templates leads to similar shape analysis results and saves around 70% of computation time. Furthermore, the approach is able to adequately capture the variability of hippocampal shapes with a reasonable number of dimensions, and to predict anatomical features of the hippocampus, only present in 17% of the population, in healthy subjects.

**Keywords:** Morphometry, shape analysis, atlas, computational anatomy, hippocampus, IHI, riemannian barycentres, LDDMM

## 1. INTRODUCTION

Statistical shape analysis methods are increasingly used in neuroscience and clinical research. Their applications include the study of correlations between anatomical structures and genetic or cognitive parameters, as well as the detection of alterations associated with neurological disorders. A current challenge for methodological research is to perform statistical analysis on large databases, which are needed to improve the statistical power of neuroscience studies.

A common approach in shape analysis is to analyse the deformations that map individuals to an atlas or template (e.g., Ashburner et al., 1998; Chung et al., 2001; Vaillant et al., 2004; Durrleman et al., 2008; Lorenzi, 2012). The three main components of these approaches are the underlying

deformation model, the template estimation method and the statistical analysis itself. The Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework (Trouvé, 1998; Beg et al., 2005; Younes, 2010) provides a natural setting for quantifying deformations between shapes or images. This framework provides diffeomorphic transformations which preserve the topology and also provides a metric between shapes. The LDDMM framework is also a natural setting for estimating templates from a population of shapes, because such templates can be defined as means in the induced shape space. Various methods have been proposed to estimate templates of a given population using the LDDMM framework (Vaillant et al., 2004; Glaunès and Joshi, 2006; Durrleman et al., 2008; Ma et al., 2008). All methods are computationally expensive due the complexity of the deformation model. This is a limitation for the study of large databases.

In this paper, we present a fast approach for template-based statistical analysis of large datasets in the LDDMM setting, and apply it to a population of 1,000 hippocampal shapes. The template estimation is based on iterative diffeomorphic centroid approaches, which two of them (IC1 and IC2) were introduced at the Geometric Science of Information conference GSI13 (Cury et al., 2013), while the PW approach was introduced in a chapter of the Geometric Theory of Information book (Cury et al., 2014b). The advantage of such method is the possibility of parallelising the process, easily adding new subjects and saving computation time without losing accuracy in the template estimation. Furthermore, the iterative nature of the methodology is useful when data are being continuously collected or pre-processed. Indeed, in such cases, there is no need to recompute the mean shape each time new subjects are added to the population, the mean shape is only refined. The main idea of these methods is to iteratively update a centroid shape by successive matchings to the different subjects. This procedure involves a limited number of matchings and thus quickly provides a template estimation of the population. Iterative centroid methods have already been used to estimate mean shapes like the hippocampus (Cury et al., 2014a, 2017) and the thalamus (Cury et al., 2018). We previously showed that these centroids can be used to initialize a variational template estimation procedure (Cury et al., 2014b), and that even if the ordering of the subject along iterations does affect the final result, all centres are very similar. Here, we propose to use these centroid estimations directly for template-based statistical shape analysis. The analysis is done on the tangent space to the template shape, either directly through Kernel Principal Component Analysis (Kernel PCA Schölkopf et al., 1997) or to approximate distances between subjects. We perform a thorough evaluation of the approach using three datasets: one synthetic dataset and two real datasets composed of 50 and 1,000 subjects respectively. In particular, we study extensively the impact of different centroids on statistical analysis, and compare the results to those obtained using a standard variational template method. We will also use the large database to predict, using the shape parameters extracted from a centroid estimation of the population, some anatomical variations of the hippocampus in the normal population, called Incomplete

Hippocampal Inversions and present in 17% of the normal population (Cury et al., 2015). IHI are interesting anatomical features also present in temporal lobe epilepsy with a frequency around 50% (Baulac et al., 1998), and is also involved in major depression disorders (Colle et al., 2016).

The paper is organized as follows. We first present in section 2 the mathematical frameworks of diffeomorphisms and currents, on which the approach is based, and then introduce the diffeomorphic centroid methods in section 3. Section 4 presents the statistical analysis. The experimental evaluation of the method is then presented in section 5, with the different data used in this paper.

## 2. MATHEMATICAL FRAMEWORKS

Our approach is based on two mathematical frameworks which we will recall in this section. The Large Deformation Diffeomorphic Metric Mapping framework is used to generate optimal matchings and quantify differences between shapes. Shapes themselves are modeled using the framework of currents which does not assume point-to-point correspondences and allows performing linear operations on shapes.

### 2.1. LDDMM Framework

Here we very briefly recall the main properties of the LDDMM setting (see Trouvé, 1998; Beg et al., 2005; Younes, 2010) for more details. The Large Deformation Diffeomorphic Metric Mapping framework allows analysing shape variability of a population using diffeomorphic transformations of the ambient 3D space. It also provides a shape space representation which means that shapes of the population are seen as points in an infinite dimensional smooth manifold, providing a continuum between shapes.

In the LDDMM framework, deformation maps  $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  are generated by integration of time-dependent vector fields  $v(x, \cdot)$ , with  $x \in \mathbb{R}^3$  and  $t \in [0, 1]$ . If  $v(x, t)$  is regular enough, i.e., if we consider the vector fields  $(v(\cdot, t))_{t \in [0, 1]}$  in  $L^2([0, 1], V)$ , where  $V$  is a Reproducing Kernel Hilbert Space (RKHS) embedded in the space of  $C^1$  vector fields vanishing at infinity, then the transport equation:

$$\begin{cases} \frac{d\phi_v}{dt}(x, t) = v(\phi_v(x, t), t) & \forall t \in [0, 1] \\ \phi_v(x, 0) = x & \forall x \in \mathbb{R}^3 \end{cases} \quad (1)$$

has a unique solution, and one sets  $\varphi_v = \phi_v(\cdot, 1)$  the diffeomorphism induced by  $v(x, \cdot)$ . The induced set of diffeomorphisms  $\mathcal{A}_V$  is a subgroup of the group of  $C^1$  diffeomorphisms. The regularity of velocity fields is controlled by:

$$E(v) := \int_0^1 \|v(\cdot, t)\|_V^2 dt. \quad (2)$$

The subgroup of diffeomorphisms  $\mathcal{A}_V$  is equipped with a right-invariant metric defined by the rules:  $\forall \varphi, \psi \in \mathcal{A}_V$ ,

$$\begin{cases} D(\varphi, \psi) = D(id, \psi \circ \varphi^{-1}) \\ D(id, \varphi) = \inf\{\int_0^1 \|v(\cdot, t)\|_V dt, \varphi = \phi_v(\cdot, 1)\} \end{cases} \quad (3)$$

i.e., the infimum is taken over all  $v \in L^2([0, 1], V)$  such that  $\varphi_v = \varphi$ .  $D(\varphi, \psi)$  represents the shortest length of paths connecting  $\varphi$  to  $\psi$  in the diffeomorphisms group.

## 2.2. Momentum Vectors

In a discrete setting, when the matching criterion depends only on  $\varphi_v$  via the images  $\varphi_v(x_p)$  of a finite number of points  $x_p$  (such as the vertices of a mesh) one can show that the vector fields  $v(x, t)$  which induce the optimal deformation map can be written via a convolution formula over the surface involving the reproducing kernel  $K_V$  of the RKHS  $V$ :

$$v(x, t) = \sum_{p=1}^n K_V(x, x_p(t))\alpha_p(t), \quad (4)$$

where  $x_p(t) = \phi_v(x_p, t)$  are the trajectories of points  $x_p$ , and  $\alpha_p(t) \in \mathbb{R}^3$  are time-dependent vectors called momentum vectors, which completely parametrize the deformation. Trajectories  $x_p(t)$  depend only on these vectors as solutions of the following system of ordinary differential equations:

$$\frac{dx_q(t)}{dt} = \sum_{p=1}^n K_V(x_q(t), x_p(t))\alpha_p(t), \quad (5)$$

for  $1 \leq q \leq n$ . This is obtained by plugging formula 4 for the optimal velocity fields into the flow Equation 1 taken at  $x = x_q$ . Moreover, the norm of  $v(\cdot, t)$  also takes an explicit form:

$$\|v(\cdot, t)\|_V^2 = \sum_{p=1}^n \sum_{q=1}^n \alpha_p(t)^T K_V(x_p(t), x_q(t))\alpha_q(t). \quad (6)$$

Note that since  $V$  is a space of vector fields, its kernel  $K_V(x, y)$  is in fact a  $3 \times 3$  matrix for every  $x, y \in \mathbb{R}^3$ . However we will only consider scalar invariant kernels of the form  $K_V(x, y) = h(\|x - y\|^2/\sigma_V^2)I_3$ , where  $h$  is a real function (in our case we use the Cauchy kernel  $h(r) = 1/(1+r)$ ), and  $\sigma_V$  a scale factor. In the following we will use a compact representation for kernels and vectors. For example Equation 6 can be written:

$$\|v(\cdot, t)\|_V^2 = \alpha(t)^T K_V(\mathbf{x}(t))\alpha(t), \quad (7)$$

where  $\alpha(t) = (\alpha_p(t))_{p=1 \dots n}, \in \mathbb{R}^{3 \times n}$ ,  $\mathbf{x}(t) = (x_p(t))_{p=1 \dots n}, \in \mathbb{R}^{3 \times n}$  and  $K_V(\mathbf{x}(t))$  the matrix of  $K_V(x_p(t), x_q(t))$ .

### 2.2.1. Geodesic Shooting

The minimization of the energy  $E(v)$  in matching problems can be interpreted as the estimation of a length-minimizing path in the group of diffeomorphisms  $\mathcal{A}_V$ , and also additionally as a length-minimizing path in the space of point sets when considering discrete problems. Such length-minimizing paths obey geodesic equations (see Vaillant et al., 2004) which write as follows:

$$\begin{cases} \frac{d\mathbf{x}(t)}{dt} = K_V(\mathbf{x}(t))\alpha(t) \\ \frac{d\alpha(t)}{dt} = -\frac{1}{2}\nabla_{\mathbf{x}(t)} [\alpha(t)^T K_V(\mathbf{x}(t))\alpha(t)], \end{cases} \quad (8)$$

Note that the first equation is nothing more than Equation 5 which allows to compute trajectories  $x_p(t)$  from any time-dependent momentum vectors  $\alpha_p(t)$ , while the second equation gives the evolution of the momentum vectors themselves. This new set of ODEs can be solved from any initial conditions  $(x_p(0), \alpha_p(0))$ , which means that the initial momentum vectors  $\alpha_p(0)$  fully determine the subsequent time evolution of the system (since the  $x_p(0)$  are fixed points). As a consequence, these initial momentum vectors encode all information of the optimal diffeomorphism. For example, the distance  $D(id, \varphi)$  satisfies

$$D(id, \varphi)^2 = E(v) = \|v(\cdot, 0)\|_V^2 = \alpha(0)^T K_V(\mathbf{x}(0))\alpha(0), \quad (9)$$

We can also use geodesic shooting from initial conditions  $(x_p(0), \alpha_p(0))$  in order to generate any arbitrary deformation of a shape in the shape space.

## 2.3. Shape Representation: Currents

The use of currents (Schwartz, 1952; de Rham, 1955) in computational anatomy was introduced by Vaillant and Glaunès (2005) and Glaunès (2005) and subsequently developed by Durrleman (2010). The basic idea is to represent surfaces as currents, i.e., linear functionals on the space of differential forms and to use kernel norms on the dual space to express dissimilarities between shapes. Using currents to represent surfaces has some benefits. First it avoids the point correspondence issue: one does not need to define pairs of corresponding points between two surfaces to evaluate their spatial proximity. Moreover, metrics on currents are robust to different samplings and topological artifacts and take into account local orientations of the shapes. Another important benefit is that this model embeds shapes into a linear space (the space of all currents), which allows considering linear combinations such as means of shapes in the space of currents.

Let us briefly recall this setting. For sake of simplicity we present currents as linear forms acting on vector fields rather than differential forms which are an equivalent formulation in our case. Let  $S$  be an oriented compact surface, possibly with boundary. Any smooth vector field  $w$  of  $\mathbb{R}^3$  can be integrated over  $S$  via the rule:

$$[S](w) = \int_S \langle w(x), n(x) \rangle d\sigma_S(x), \quad (10)$$

with  $n(x)$  the unit normal vector to the surface,  $d\sigma_S$  the Lebesgue measure on the surface  $S$ , and  $[S]$  is called a 2-current associated to  $S$ .

Given an appropriate Hilbert space  $(W, \langle \cdot, \cdot \rangle_W)$  of vector fields, continuously embedded in  $C_0^1(\mathbb{R}^3, \mathbb{R}^3)$ , the space of currents we consider is the space of continuous linear forms on  $W$ , i.e., the dual space  $W^*$ . For any point  $x \in \mathbb{R}^3$  and vector  $\alpha \in \mathbb{R}^3$  one can consider the Dirac functional  $\delta_x^\alpha : w \mapsto \langle w(x), \alpha \rangle$  which belongs to  $W^*$ . The Riesz representation theorem states that there exists a unique  $u \in W$  such that for all  $w \in W$ ,  $\langle u, w \rangle_W = \delta_x^\alpha(w) = \langle w(x), \alpha \rangle$ .  $u$  is thus a vector field which depends on  $x$  and linearly on  $\alpha$ , and we write it  $u = K_w(\cdot, x)\alpha$ .  $K_w(x, y)$  is a  $3 \times 3$  matrix, and  $K_w : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^{3 \times 3}$  the mapping

called the reproducing kernel of the space  $W$ . Thus we have the rule

$$\langle K_W(\cdot, x)\alpha, w \rangle_W = \langle w(x), \alpha \rangle.$$

Moreover, applying this formula to  $w = K_W(\cdot, y)\beta$  for any other point  $y \in \mathbb{R}^3$  and vector  $\beta \in \mathbb{R}^3$ , we get

$$\begin{aligned} \langle K_W(\cdot, x)\alpha, K_W(\cdot, y)\beta \rangle_W &= \langle K_W(x, y)\beta, \alpha \rangle \\ &= \alpha^T K_W(x, y)\beta = \langle \delta_x^\alpha, \delta_y^\beta \rangle_{W^*} \end{aligned} \quad (11)$$

Using equation 11, one can prove that for two surfaces  $S$  and  $T$ ,

$$\langle [S], [T] \rangle_{W^*} = \int_S \int_T \langle n_S(x), K_W(x, y)n_T(y) \rangle d\sigma_S(x)d\sigma_T(y) \quad (12)$$

This formula defines the metric we use as data attachment term for comparing surfaces. More precisely, the difference between two surfaces is evaluated via the formula:

$$\| [S] - [T] \|_{W^*}^2 = \langle [S], [S] \rangle_{W^*} + \langle [T], [T] \rangle_{W^*} - 2 \langle [S], [T] \rangle_{W^*} \quad (13)$$

The type of kernel fully determines the metric and therefore will have a direct impact on the behavior of the algorithms. We use scalar invariant kernels of the form  $K_W(x, y) = h(\|x-y\|^2/\sigma_W^2)I_3$ , where  $h$  is a real function (in our case we use the Cauchy kernel  $h(r) = 1/(1+r)$ ), and  $\sigma_W$  a scale factor.

Note that the varifold (Charon and Trouv  , 2013) can be also used for shape representation without impacting the methodology. The shapes we used for this study are well represented by currents.

## 2.4. Surface Matchings

We can now define the optimal match between two currents  $[S]$  and  $[T]$ , which is the diffeomorphism minimizing the functional

$$J_{S,T}(\nu) = \gamma E(\nu) + \| [\varphi_\nu(S)] - [T] \|_{W^*}^2 \quad (14)$$

This functional is non convex and in practice we use a gradient descent algorithm to perform the optimization, which cannot guarantee to reach a global minimum. We observed empirically that local minima can be avoided by using a multi-scale approach in which several optimization steps are performed with decreasing values of the width  $\sigma_W$  of the kernel  $K_W$  (each step provides an initial guess for the next one). Evaluations of the functional and its gradient require numerical integrations of high-dimensional ordinary differential equations (see equation 5), which is done using Euler trapezoidal rule. Note that three important parameters control the matching process:  $\gamma$  controls the regularity of the map,  $\sigma_V$  controls the scale in the space of deformations and  $\sigma_W$  controls the scale in the space of currents.

## 2.5. GPU Implementation

To speed up the matchings computation of all methods used in this study (the variational template and the different centroid estimation algorithms), we use a GPU implementation for the computation of kernel convolutions. This computation constitutes the most time-consuming part of LDDMM methods. Computations were performed on a Nvidia Tesla C1060 card. The GPU implementation can be found here: <http://www.mi.parisdescartes.fr/~glaunes/>.

## 3. DIFFEOMORPHIC CENTROID METHODS

Computing a template in the LDDMM framework can be highly time consuming, taking a few days or some weeks for large real-world databases. Here we propose a fast approach which provides a centroid correctly centred among the population. Matlab codes are available here: [https://github.com/cclairec/Iterative\\_Centroid](https://github.com/cclairec/Iterative_Centroid), and use the GPU implementation mentioned above for the kernel convolutions.

### 3.1. General Idea

The LDDMM framework, in an ideal setting (exact matching between shapes), sets the template estimation problem as a centroid computation on a Riemannian manifold. The Fr  chet mean is the standard way for defining such a centroid and provides the basic inspiration of all LDDMM template estimation methods.

If  $\mathbf{x}^i, 1 \leq i \leq N$  are points in  $\mathbb{R}^d$ , then their centroid is defined as

$$\mathbf{b}^N = \frac{1}{N} \sum_{i=1}^N \mathbf{x}^i. \quad (15)$$

It also satisfies the following two alternative characterizations:

$$\mathbf{b}^N = \arg \min_{\mathbf{y} \in \mathbb{R}^d} \sum_{1 \leq i \leq N} \|\mathbf{y} - \mathbf{x}^i\|^2. \quad (16)$$

and

$$\begin{cases} \mathbf{b}^1 = \mathbf{x}^1 \\ \mathbf{b}^{k+1} = \frac{k}{k+1} \mathbf{b}^k + \frac{1}{k+1} \mathbf{x}^{k+1}, \quad 1 \leq k \leq N-1. \end{cases} \quad (17)$$

Now, when considering points  $\mathbf{x}^i$  living on a Riemannian manifold  $M$  (we assume  $M$  is path-connected and geodesically complete), the definition of  $\mathbf{b}^N$  cannot be used because  $M$  is not a vector space. However the variational characterization of  $\mathbf{b}^N$  as well as the iterative characterization, both have analogs in the Riemannian case. The Fr  chet mean is defined under some hypotheses (see Arnaudon et al., 2012) on the relative locations of points  $\mathbf{x}^i$  in the manifold:

$$\mathbf{b}^N = \arg \min_{\mathbf{y} \in M} \sum_{1 \leq i \leq N} d_M(\mathbf{y}, \mathbf{x}^i)^2. \quad (18)$$

Many mathematical studies (as for example Karcher, 1977; Kendall, 1990; Le, 2004; Afsari, 2011; Arnaudon et al., 2012;

Afsari et al., 2013), have focused on proving the existence and uniqueness of the mean, as well as proposing algorithms to compute it. However, these approaches are computationally expensive, in particular in high dimension and when considering non trivial metrics. An alternative idea consists in using the Riemannian analog of the second characterization:

$$\begin{cases} \tilde{\mathbf{b}}^1 = \mathbf{x}^1 \\ \tilde{\mathbf{b}}^{k+1} = \text{geod}(\tilde{\mathbf{b}}^k, \mathbf{x}^{k+1}, \frac{1}{k+1}), \quad 1 \leq k \leq N-1, \end{cases} \quad (19)$$

where  $\text{geod}(\mathbf{y}, \mathbf{x}, t)$  is the point located along the geodesic from  $\mathbf{y}$  to  $\mathbf{x}$ , at a distance from  $\mathbf{y}$  equal to  $t$  times the length of the geodesic. This does not define the same point as the Fréchet mean, and moreover the result depends on the ordering of the points. In fact, all procedures that are based on decomposing the Euclidean equality  $\mathbf{b}^N = \frac{1}{N} \sum_{i=1}^N \mathbf{x}^i$  as a sequence of pairwise convex combinations lead to possible alternative definitions of centroid in a Riemannian setting. However, this should lead to a fast estimation. We hypothesize that, in the case of shape analysis, it could be sufficient for subsequent template based statistical analysis. Moreover, this procedure has the side benefit that at each step  $\tilde{\mathbf{b}}^k$  is the centroid of the  $\mathbf{x}^i, 1 \leq i \leq k$ .

In the following, we present three algorithms that build on this idea. The two first methods are iterative, and the third one is recursive, but also based on pairwise matchings of shapes.

### 3.2. Direct Iterative Centroid (IC1)

The first algorithm roughly consists in applying the following procedure: given a collection of  $N$  shapes  $S_i$ , we successively update the centroid by matching it to the next shape and moving along the geodesic flow. More precisely, we start from the first surface  $S_1$ , match it to  $S_2$  and set  $B_2 = \phi_{v^1}(S_1, 1/2)$ .  $B_2$  represents the centroid of the first two shapes, then we match  $B_2$  to  $S_3$ , and set as  $B_3 = \phi_{v^2}(B_2, 1/3)$ . Then we iterate this process (see Algorithm 1).

---

#### Algorithm 1: Iterative Centroid 1 (IC1)

---

**Data:**  $N$  surfaces  $S_i$   
**Result:** 1 surface  $B_N$  representing the centroid of the population  
 $B_1 = S_1$ ;  
**for**  $i$  from 1 to  $N - 1$  **do**  
 $B_i$  is matched to  $S_{i+1}$  which results in a deformation map  $\phi_{v^i}(x, t)$ ;  
Set  $B_{i+1} = \phi_{v^i}(B_i, \frac{1}{i+1})$  which means that we transport  $B_i$  along the geodesic and stop at time  $t = \frac{1}{i+1}$ ;  
**end**

---

### 3.3. Centroid With Averaging in the Space of Currents (IC2)

Because matchings are not exact, the centroid computed with the IC1 method accumulates small errors which can have an impact on the final centroid. Furthermore, the final centroid is in fact

a deformation of the first shape  $S_1$ , which makes the procedure even more dependent on the ordering of subjects than it would be in an ideal exact matching setting. In this second algorithm, we modify the updating step by computing a mean in the space of currents between the deformation of the current centroid and the backward flow of the current shape being matched. Hence the computed centroid is not a surface but a combination of surfaces, as in the template estimation method. The algorithm proceeds as presented in Algorithm 2.

---

#### Algorithm 2: Iterative Centroid 2 (IC2)

---

**Data:**  $N$  surfaces  $S_i$   
**Result:** 1 current  $\mathcal{B}_N$  representing the centroid of the population  
 $\mathcal{B}_1 = [S_1]$ ;  
**for**  $i$  from 1 to  $N - 1$  **do**  
 $\mathcal{B}_i$  is matched to  $[S_{i+1}]$  which results in a deformation map  $\phi_{v^i}(x, t)$ ;  
Set  $\mathcal{B}_{i+1} = \frac{i}{i+1} \phi_{v^i}(\mathcal{B}_i, \frac{1}{i+1}) + \frac{1}{i+1} [\phi_{u^i}(S_{i+1}, \frac{i}{i+1})]$  which means that we transport  $\mathcal{B}_i$  along the geodesic and stop at time  $t = \frac{1}{i+1}$ ;  
where  $u^i(x, t) = -v^i(x, 1-t)$ , i.e.,  $\phi_{u^i}$  is the reverse flow map.  
**end**

---

The weights in the averaging reflect the relative importance of the new shape, so that at the end of the procedure, all shapes forming the centroid have equal weight  $\frac{1}{N}$ .

Note that we have used the notation  $\phi_{v^i}(\mathcal{B}_i, \frac{1}{i+1})$  to denote the transport (push-forward) of the current  $\mathcal{B}_i$  by the diffeomorphism. Here  $\mathcal{B}_i$  is a linear combination of currents associated to surfaces, and the transported current is the linear combination (keeping the weights unchanged) of the currents associated to the transported surfaces.

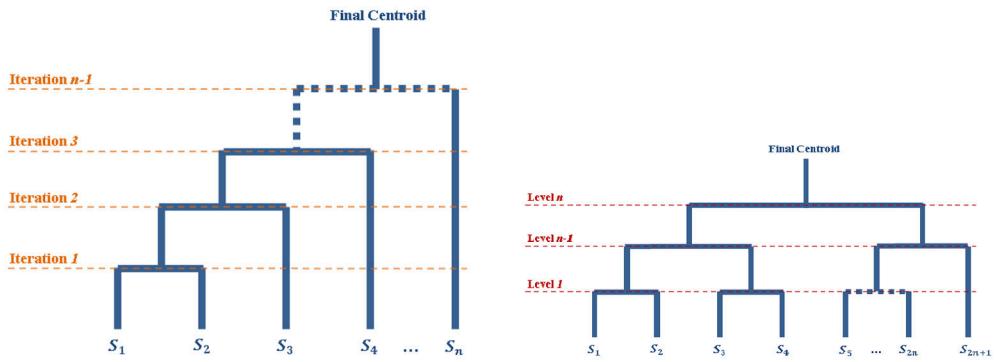
### 3.4. Alternative Method : Pairwise Centroid (PW)

Another possibility is to recursively split the population in two parts until having only one surface in each group (see Figure 1), and then going back up along the dyadic tree by computing pairwise centroids between groups, with appropriate weight for each centroid (Algorithm 3).

#### 3.4.1. Remarks on the IC Algorithms

These three methods depend on the ordering of subjects. In a previous work (Cury et al., 2014b), we showed empirically that different orderings result in very similar final centroids. Here we focus on the use of such centroid for statistical shape analysis.

Regarding memory performances of the algorithm, computing a mean shape with the algorithm IC1 on meshes composed by 500 vertices, requires 1.2 MB. For the algorithm IC2, which combines surfaces at each iterations, requires more memory, for 50 meshes IC2 needs 20 MB. We used the real dataset RD50 (see section 5.1 for further details) to estimate memory performances.



**FIGURE 1 |** Diagrams of the iterative processes which lead to the centroids computations. The tops of the diagrams represent the final centroid. The diagram on the left corresponds to the Iterative Centroid algorithms (IC1 and IC2). The diagram on the right corresponds to the pairwise algorithm (PW).

### Algorithm 3: Pairwise Centroid (PW)

```

Data: N surfaces  $S_i$ 
Result: 1 surface  $B$  representing the centroid of the
population
if  $N \geq 2$  then
     $B_{left}$  = Pairwise Centroid ( $S_1, \dots, S_{[N/2]}$ );
     $B_{right}$  = Pairwise Centroid ( $S_{[N/2]+1}, \dots, S_N$ );
     $B_{left}$  is matched to  $B_{right}$  which results in a deformation
    map  $\phi_v(x, t)$ ;
    Set  $B = \phi_v(B_{left}, \frac{[N/2]+1}{N})$  which means we transport  $B_{left}$ 
    along the geodesic and stop at time  $t = \frac{[N/2]+1}{N}$ ;
end
else
     $| B = S_1$ 
end
```

### 3.5. Comparison With a Variational Template Estimation Method

In this study, we will compare our centroid approaches to a variational template estimation method proposed by Glaunès et al (Glaunès and Joshi, 2006). This variational method estimates a template given a collection of surfaces using the framework of currents. It is posed as a minimum mean squared error estimation problem. Let  $S_i$  be  $N$  surfaces in  $\mathbb{R}^3$  (i.e., the whole surface population). Let  $[S_i]$  be the corresponding current of  $S_i$ , or its approximation by a finite sum of vectorial Diracs. The problem is formulated as follows:

$$\{\hat{v}_i, \hat{T}\} = \arg \min_{v_i, T} \sum_{i=1}^N \{\|T - [\varphi_{v_i}(S_i)]\|_{W^*}^2 + \gamma E(v_i)\}, \quad (20)$$

The method uses an alternated optimization i.e., surfaces are successively matched to the template, then the template is updated and this sequence is iterated until convergence. One can observe that when  $\varphi_i$  is fixed, the functional is minimized when  $T$  is the average of  $[\varphi_i(S_i)]$ :  $T = \frac{1}{N} \sum_{i=1}^N [\varphi_i(S_i)]$ , which makes the optimization with respect to  $T$  straightforward. This

optimal current is the union of all surfaces  $\varphi_{v_i}(S_i)$ . However, all surfaces being co-registered, the  $\hat{\varphi}_{v_i}(S_i)$  are close to each other, which makes the optimal template  $\hat{T}$  close to being a true surface. Standard initialization consists in setting  $T = \frac{1}{N} \sum_{i=1}^N [S_i]$ , which means that the initial template is defined as the combination of all unregistered shapes in the population. Alternatively, if one is given a good initial guess  $T$ , the convergence speed of the method can be improved. In particular, the initialisation can be provided by iterative centroids; this is what we will use in the experimental section.

Regarding the computational complexity, the different centroid approaches perform  $N - 1$  matchings while the variational template estimation requires  $N \times iter$  matchings, where  $iter$  is the number of iterations. Moreover the time for a given matching depends quadratically on the number of vertices of the surfaces being matched. It is thus more expensive when the template is a collection of surfaces as in IC2 and in the variational template estimation.

### 4. STATISTICAL ANALYSIS

The proposed iterative centroid approaches can be used for subsequent statistical shape analysis of the population, using various strategies. A first strategy consists in analysing the deformations between the centroid and the individual subjects. This is done by analysing the initial momentum vectors  $\alpha^i(0) = (\alpha_p^i(0))_{p=1 \dots n} \in \mathbb{R}^{3 \times n}$  which encode the optimal diffeomorphisms computed from the matching between a centroid and the subjects  $S_i$ . Initial momentum vectors all belong to the same vector space and are located on the vertices of the centroid. Different approaches can be used to analyse these momentum vectors, including Principal Component Analysis for the description of populations, Support Vector Machines or Linear Discriminant Analysis for automatic classification of subjects. A second strategy consists in analysing the set of pairwise distances between subjects. Then, the distance matrix can be entered into analysis methods such as Isomap Tenenbaum et al. (2000), Locally Linear Embedding Roweis and Saul (2000), and Yang X. et al. (2011) or spectral clustering algorithms (Von

Luxburg, 2007). Here, we tested two approaches: (i) the analysis of initial momentum vectors using a Kernel Principal Component Analysis for the first strategy; (ii) the approximation of pairwise distance matrices for the second strategy. These tests allow us both to validate the different iterative centroid methods and to show the feasibility of such analysis on large databases.

#### 4.1. Principal Component Analysis on Initial Momentum Vectors

The Principal Component Analysis (PCA) on initial momentum vectors from the template to the subjects of the population is an adaptation of PCA in which Euclidean scalar products between observations are replaced by scalar products using a kernel. Here the kernel is  $K_V$  the kernel of the R.K.H.S  $V$ . This adaptation can be seen as a Kernel PCA (Schölkopf et al., 1997). PCA on initial momentum vectors has previously been used in morphometric studies in the LDDMM setting (Vaillant et al., 2004; Duralleman et al., 2009) and it is sometimes referred to tangent PCA.

We briefly recall that, in standard PCA, the principal components of a dataset of  $N$  observations  $\alpha^i \in \mathbb{R}^p$  with  $i \in \{1, \dots, N\}$  are defined by the eigenvectors of the covariance matrix  $C$  with entries:

$$C(i,j) = \frac{1}{N-1}(\alpha^i - \bar{\alpha})^t(\alpha^j - \bar{\alpha}) \quad (21)$$

with  $\alpha^i$  given as a column vector,  $\bar{\alpha} = \frac{1}{N} \sum_{i=1}^N \alpha^i$ , and  $x^t$  denotes the transposition of a vector  $x$ .

In our case, our observations are initial momentum vectors  $\alpha^i \in \mathbb{R}^{3 \times n}$  and instead of computing the Euclidean scalar product in  $\mathbb{R}^{3 \times n}$ , we compute the scalar product with matrix  $K_V$ , which is a natural choice since it corresponds to the inner product of the corresponding initial vector fields in the space  $V$ . The covariance matrix then writes:

$$C_V(i,j) = \frac{1}{N-1}(\alpha^i - \bar{\alpha})^t K_V(x)(\alpha^j - \bar{\alpha}) \quad (22)$$

with  $\bar{\alpha}$  the vector of the mean of momentum vectors, and  $x$  the vector of vertices of the template surface. We denote  $\lambda_1, \lambda_2, \dots, \lambda_N$  the eigenvalues of  $C$  in decreasing order, and  $v^1, v^2, \dots, v^N$  the corresponding eigenvectors. The  $k$ -th principal mode is computed from the  $k$ -th eigenvector  $v^k$  of  $C_V$ , as follows:

$$m^k = \bar{\alpha} + \sum_{j=1}^N v_j^k (\alpha^j - \bar{\alpha}). \quad (23)$$

The cumulative explained variance  $CEV_k$  for the  $k$  first principal modes is given by the equation:

$$CEV_k = \frac{\sum_{h=1}^k \lambda_h}{\sum_{h=1}^N \lambda_h} \quad (24)$$

We can use geodesic shooting along any principal mode  $m^k$  to visualize the corresponding deformations.

#### Remark

To analyse the population, we need to know the initial momentum vectors  $\alpha^i$  which correspond to the matchings from the centroid to the subjects. For the IC1 and PW centroids, these initial momentum vectors were obtained by matching the centroid to each subject. For the IC2 centroid, since the mesh structure is composed of all vertices of the population, it is too computationally expensive to match the centroid toward each subject. Instead, from the deformation of each subject toward the centroid, we used the opposite vector of final momentum vectors for the analysis. Indeed, if we have two surfaces  $S$  and  $T$  and need to compute the initial momentum vectors from  $T$  to  $S$ , we can estimate the initial momentum vectors  $\alpha^{TS}(0)$  from  $T$  to  $S$  by computing the deformation from  $S$  to  $T$  and using the initial momentum vectors  $\tilde{\alpha}^{TS}(0) = -\alpha^{ST}(1)$ , which are located at vertices  $\phi_{ST}(x^S)$ .

#### 4.2. Distance Matrix Approximation

Various methods such as Isomap (Tenenbaum et al., 2000) or Locally Linear Embedding (Roweis and Saul, 2000; Yang X. et al., 2011) use as input a matrix of pairwise distances between subjects. In the LDDMM setting, it can be computed using diffeomorphic distances:  $\rho(S_i, S_j) = D(id, \varphi_{ij})$ . However, for large datasets, computing all pairwise deformation distance is computationally very expensive, as it involves  $O(N^2)$  matchings. An alternative is to approximate the pairwise distance between two subjects through their matching from the centroid or template. This approach has been introduced in Yang X. F. et al. (2011). Here we use a first order approximation to estimate the diffeomorphic distance between two subjects:

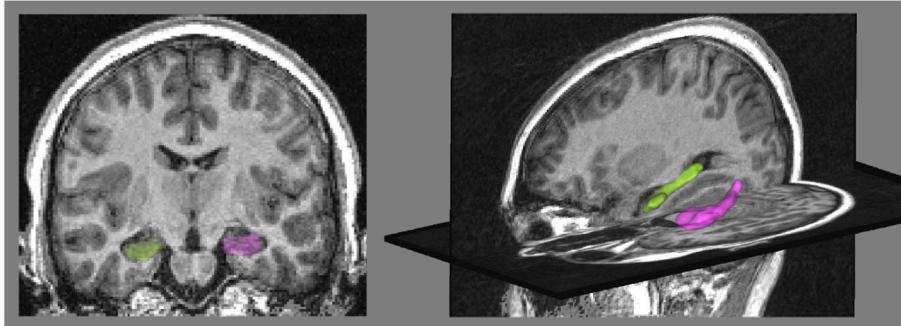
$$\tilde{\rho}(S_i, S_j) = \sqrt{\langle \alpha^j(0) - \alpha^i(0), K_V(x(0))(\alpha^j(0) - \alpha^i(0)) \rangle}, \quad (25)$$

with  $x(0)$  the vertices of the estimated centroid or template and  $\alpha^i(0)$  is the vector of initial momentum vectors computed by matching the template to  $S_i$ . Using such approximation allows to compute only  $N$  matchings instead of  $N(N-1)$ .

Note that  $\rho(S_i, S_j)$  is in fact the distance between  $S_i$  and  $\varphi_{ij}(S_i)$ , and not between  $S_i$  and  $S_j$  due to the not exactitude of matchings. However we will refer to it as a distance in the following to denote the dissimilarity between  $S_i$  and  $S_j$ .

## 5. EXPERIMENTS AND RESULTS

In this section, we evaluate the use of iterative centroids for statistical shape analysis. Specifically, we investigate the centring of the centroids within the population, their impact on population analysis based on Kernel PCA and on the computation of distance matrices. For our experiments, we used three different datasets: two real datasets and a synthetic one. In all datasets shapes are hippocampi. The hippocampus is an anatomical structure of the temporal lobe of the brain, involved in different memory processes.



**FIGURE 2 | (Left)** coronal view of the MRI with the binary masks of hippocampi segmented by the SACHA software (Chupin et al., 2009), the right hippocampus is in green and the left one in pink. **(Right)** 3D view of the hippocampus meshes.

## 5.1. Data

The two real datasets are from the European database IMAGEN (Schumann et al., 2010)<sup>1</sup> composed of young healthy subjects. We segmented the hippocampi from T1-weighted Magnetic Resonance Images (MRI) of subjects using the SACHA software (Chupin et al., 2009) (see Figure 2). The synthetic dataset was built using deformations of a single hippocampal shape of the IMAGEN database.

### 5.1.1. The Synthetic Dataset SD

SD is composed of synthetic deformations of a single shape  $S_0$ , designed such that this single shape becomes the exact center of the population. We will thus be able to compare the computed centroids to this exact center. We generated 50 subjects for this synthetic dataset from  $S_0$ , along geodesics in different directions. We randomly chose two orthogonal momentum vectors  $\beta_1$  and  $\beta_2$  in  $\mathbb{R}^{3 \times n}$ . We then computed momentum vectors  $\alpha^i, i \in \{1, \dots, 25\}$  of the form  $k_1^i \beta_1 + k_2^i \beta_2 + k_3^i \beta_3$  with  $(k_1^i, k_2^i, k_3^i) \in \mathbb{R}^3, \forall i \in \{1, \dots, 25\}, k_j^i \sim \mathcal{N}(0, \sigma_j)$  with  $\sigma_1 > \sigma_2 \gg \sigma_3$  and  $\beta_3$  a randomly selected momentum vector, adding some noise to the generated 2D space. We computed momentum vectors  $\alpha^j, j \in \{26, \dots, 50\}$  such as  $\alpha^j = -\alpha^{j-25}$ . We generated the 50 subjects of the population by computing geodesic shootings of  $S_0$  using the initial momentum vectors  $\alpha^i, i \in \{1, \dots, 50\}$ . The population is symmetrical since  $\sum_i^5 \alpha^i = 0$ . It should be noted that all shapes of the dataset have the same mesh structure composed of  $n = 549$  vertices. Data can be found here: <https://doi.org/10.5281/zenodo.1420395>.

### 5.1.2. The Real Dataset RD50

RD50 is composed of 50 left hippocampi from the IMAGEN database. We applied the following preprocessing steps to each individual MRI. First, the MRI was linearly registered toward the MNI152 atlas, using the FLIRT procedure (Jenkinson et al., 2002) of the FSL software<sup>2</sup>. The computed linear transformation was then applied to the binary mask of the hippocampal segmentation. A mesh of this segmentation was then computed

from the binary mask using the BrainVISA software<sup>3</sup>. All meshes were then aligned using rigid transformations to one subject of the population. For this rigid registration, we used a similarity term based on measures (as in Glaunès et al., 2004). All meshes were decimated in order to keep a reasonable number of vertices: meshes have on average 500 vertices.

### 5.1.3. The Real Database RD1000

RD1000 is composed of 1,000 left hippocampi from the IMAGEN database. We applied the same preprocessing steps to the MRI data as for the dataset RD50. This dataset has also a score of Incomplete Hippocampal Inversion (IHI) (Cury et al., 2015), which is an anatomical variant of the hippocampus, present in 17% of the normal population.

## 5.2. Experiments

For the datasets SD and RD50 (which both contain 50 subjects), we compared the results of the three different iterative centroid algorithms (IC1, IC2, and PW). We also investigated the possibility of computing variational templates, initialized by the centroids, based on the approach presented in section 3.5. We could thus compare the results obtained when using the centroid directly to those obtained when using the most expensive (in term of computation time) template estimation. We thus computed 6 different centres: IC1, IC2, PW and the corresponding variational templates T(IC1), T(IC2), T(PW). For the synthetic dataset SD, we could also compare those 6 estimated centres to the exact centre of the population. For the real dataset RD1000 (with 1,000 subjects), we only computed the iterative centroid IC1.

For all computed centres and all datasets, we investigated: (1) the computation time; (2) whether the centres are close to a critical point of the Fréchet functional on the manifold discretised by the population; (3) the impact of the estimated centres on the results of Kernel PCA; (4) their impacts on approximated distance matrices.

Computation times are given for the different template estimation methods. For the variational template initialized with an estimated centroid, the computation time depends on the

<sup>1</sup><http://www.imagen-europe.com/>

<sup>2</sup><http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslOverview>

<sup>3</sup><http://brainvisa.info/web/index.html>

number of iteration needed to converge. We fixed this number to 5 *iter* (see section 3.5), which is usually, for the kind of shape we are using here, the number of iteration needed by the variational template to converge. For the variational template with standard initialisation (all objects of the population are considered, see section 3.5), we used 8 iterations.

To assess how close an estimated centre is to a critical point of the Fréchet functional, for the different centroids and variational templates, we computed a ratio  $R$  using the initial momentum vectors from centres to subjects. The ratio  $R$  takes values between 0 and 1:

$$R = \frac{\|\frac{1}{N} \sum_{i=1}^N v^i(\cdot, 0)\|_V}{\frac{1}{N} \sum_{i=1}^N \|v^i(\cdot, 0)\|_V}, \quad (26)$$

with  $v^i(\cdot, 0)$  the vector field of the deformation from the estimated centre to the subject  $S_i$ , corresponding to the vector of initial momentum vectors  $\alpha^i(0)$ . This ratio gives information about the centering of the centroid. In fact, in a pure riemannian setting, i.e., disregarding inaccuracies of matchings, and under some reasonable assumptions about curvature of the shape space, the Fréchet functional (see equation 18) would have a unique critical point corresponding to the Fréchet mean, which can be interpreted as the theoretical centre of the population. Now a null ratio  $R$  would mean that  $\sum_{i=1}^N v^i(\cdot, 0) = 0$ , which is precisely the equation satisfied by critical points of the Fréchet functional. In practice, this ratio is expected to be close to zero but not null due to matchings inaccuracies.

We compared the results of Kernel PCA computed from these different centres by comparing the principal modes and the cumulative explained variance for different number of dimensions.

Finally, we compared the approximated distance matrices to the direct distance matrix.

For the RD1000 dataset, we will try to predict an anatomical variant of the normal population, the Incomplete Hippocampal Inversion (IHI), present in only 17% of the normal population.

### 5.3. Synthetic Dataset SD

#### 5.3.1. Computation Time

All the centroids and variational templates have been computed with  $\sigma_V = 15$ , which represents roughly half of the shapes length. Computation times for IC1 took 31 min, 85 min for IC2, and 32 min for PW. The corresponding variational template initialized by these estimated centroids took 81 min (112 min in total), 87 min (172 min in total) and 81 min (113 min in total). As a reference, we also computed a template with the standard initialisation whose computation took 194 min. Computing a centroid saved between 56 and 84% of computation time over the template with standard initialization and between 50 and 72% over the template initialized by the centroid.

For the following analysis we need additional matching estimations to extract the initial momentum vectors from the mean shape to individual subjects. Those extra matchings are needed for tangent space analysis, and are not mandatory if one just wants to estimate a mean shape of a population (Cury et al., 2018). Estimating the initial momentum vectors from each

**TABLE 1 |** Synthetic dataset SD.

Ratio	C	T(C)
IC1	$0.07 \pm 0.03$	$0.05 \pm 0.02$
IC2	$0.07 \pm 0.03$	$0.05 \pm 0.02$
PW	$0.11 \pm 0.05$	$0.07 \pm 0.02$

Ratio  $R$  (equation 26) computed for the 3 centroids C, and the 3 variational templates initialized via these centroids (T(C)).

centroids (IC1, IC2, and PW) added in average 22 min. No extra time for the variational template, since the method optimize those initial momentum vectors. Estimating initial momentum vectors and the centre of the population using IC1 took 53 min, using IC2 took 107 min and using PW took 54 min, which is still significantly faster than using a variational template.

#### 5.3.2. "Centring" of the Estimated Centres

Since in practice a computed centre is never at the exact centre, and its estimation may vary accordingly to the discretisation of the underlying shape space, we decided to generate another 49 synthetic populations as detailed in section 5.1, so we have 50 different discretisations of the shape space. For each of these populations, we computed the 3 centroids and the 3 variational templates initialized with these centroids. We calculated the ratio  $R$  described in the previous section for each estimated centre. Table 1 presents the mean and standard deviation values of the ratio  $R$  for each centroid and template, computed over these 50 populations.

In a pure Riemannian setting (i.e., disregarding the fact that matchings are not exact), a zero ratio would mean that we are at a critical point of the Fréchet functional, and under some reasonable assumptions on the curvature of the shape space in the neighborhood of the dataset (which we cannot check however), it would mean that we are at the Fréchet mean. By construction, the ratio computed from the exact centre using the initial momentum vectors  $\alpha^i$  used for the construction of subjects  $S_i$  (as presented in section 5.1) is zero.

Ratios  $R$  are close to zero for all centroids and variational templates, indicating that they are close to the exact centre. Furthermore, the value of  $R$  may be partly due to the non-exactitude of the matchings between the estimated centres and the subjects. To become aware of this non-exactitude, we matched the exact centre toward all subjects of the SD dataset. The resulting ratio is  $R = 0.05$ . This is of the same order of magnitude as the ratios obtained in Table 1, indicating that the estimated centres are indeed very close to the exact centre.

#### 5.3.3. PCA on Initial Momentum Vectors

We performed a PCA computed with the initial momentum vectors (see section 4 for details) from our different estimated centres (3 centroids, 3 variational templates and the exact centre).

We computed the cumulative explained variance for different number of dimensions of the PCA. Results are presented in Table 2. The cumulative explained variances are very similar for the different centres for any number of dimensions.

**TABLE 2 |** Synthetic dataset SD.

	1st mode	2nd mode	3rd mode
Centre	0.829	0.989	0.994
IC1	0.829	0.990	0.995
IC2	0.833	0.994	0.996
PW	0.829	0.990	0.995
T(IC1)	0.829	0.995	0.999
T(IC2)	0.829	0.995	0.999
T(PW)	0.829	0.995	0.999

Proportion of cumulative explained variance of kernel PCA computed from different centres.

We wanted to take advantage of the construction of this synthetic dataset to answer the question: Do the principal components explain the same deformations? The SD dataset allows to visualize the principal component relatively to the real position of the generator vectors and the population itself. Such visualization is not possible for real dataset since shape spaces are not generated by only 2 vectors. For this synthetic dataset, we can project principal components on the 2D space spanned by  $\beta_1$  and  $\beta_2$  as described in the previous paragraph. This projection allows displaying in the same 2D space subjects in their native space, and principal axes computed from the different Kernel PCAs. To visualize the first component (respectively the second one), we shot from the associated centre in the direction  $k * \mathbf{m}^1$  (resp.  $\mathbf{m}^2$ ) with  $k \in [-2\sqrt{\lambda_1}; +2\sqrt{\lambda_1}]$  (resp.  $\sqrt{\lambda_2}$ ). Results are presented in **Figure 3**. The deformations captured by the 2 principal axes are extremely similar for all centres. The principal axes for the 7 centres, have all the same position within the 2D shape space. So for a similar amount of explained variance, the axes describe the same deformation.

Overall, for this synthetic dataset, the 6 estimated centres give very similar PCA results.

### 5.3.4. Distance Matrices

We then studied the impact of different centres on the approximated distance matrices. We computed the seven approximated distance matrices corresponding to the seven centres, and the direct pairwise distance matrix computed by matching all subjects to each other. Computation of the direct distance matrix took 1,000 min (17 h) for this synthetic dataset of 50 subjects. In the following, we denote as  $aM(C)$  the approximated distance matrix computed from the centre  $C$ .

To quantify the difference between these matrices, we used the following error  $e$ :

$$e(M_1, M_2) = \frac{1}{N^2} \sum_{i,j=1}^N \frac{|M_1(i,j) - M_2(i,j)|}{\max(M_1(i,j), M_2(i,j))} \quad (27)$$

with  $M_1$  and  $M_2$  two distance matrices. Results are reported in **Table 3**. For visualization of the errors against the pairwise computed distance matrices, we also computed the error between the direct distance matrix, by computing pairwise deformations (23h h of computation per population), for 3 populations

randomly selected. **Figure 4** shows scattered plots between the pairwise distances matrices and the approximated distances matrices of IC1 and T(IC1) for the 3 randomly selected populations. The errors between  $aM(C)$  matrices and the pairwise distance matrices of each of the populations are 0.17 0.16 and 0.14, respectively 0.11 0.08 and 0.07 for the errors with the corresponding  $aM(T(C))$ . We can observe a subtle curvature of the scatter-plot, which is due to the curvature of the shape space. This figure illustrates the good approximation of the distances matrices, regarding to the pairwise estimation distance matrix. The variational templates are getting slightly closer to the identity line, which is expected (as for the better ratio values) since they are extra iterations to converge toward a centre of the population, however the estimated centroids from the different algorithms, still provide a good approximation of the pairwise distances of the population. In conclusion for this set of synthetic population, the different estimated centres have also a little impact on the approximation of the distance matrices.

## 5.4. The Real Dataset RD50

We now present experiments on the real dataset RD50. For this dataset, the exact center of the population is not known, neither is the distribution of the population and meshes have different numbers of vertices and different connectivity structures.

### 5.4.1. Computation Time

We estimated our 3 centroids IC1 (75 min) IC2 (174 min) and PW (88 min), and the corresponding variational templates, which took respectively 188, 252, and 183 min. The total computation time for  $T(IC1)$  is 263 min, 426 min for  $T(IC2)$  and 271 min for  $T(PW)$ .

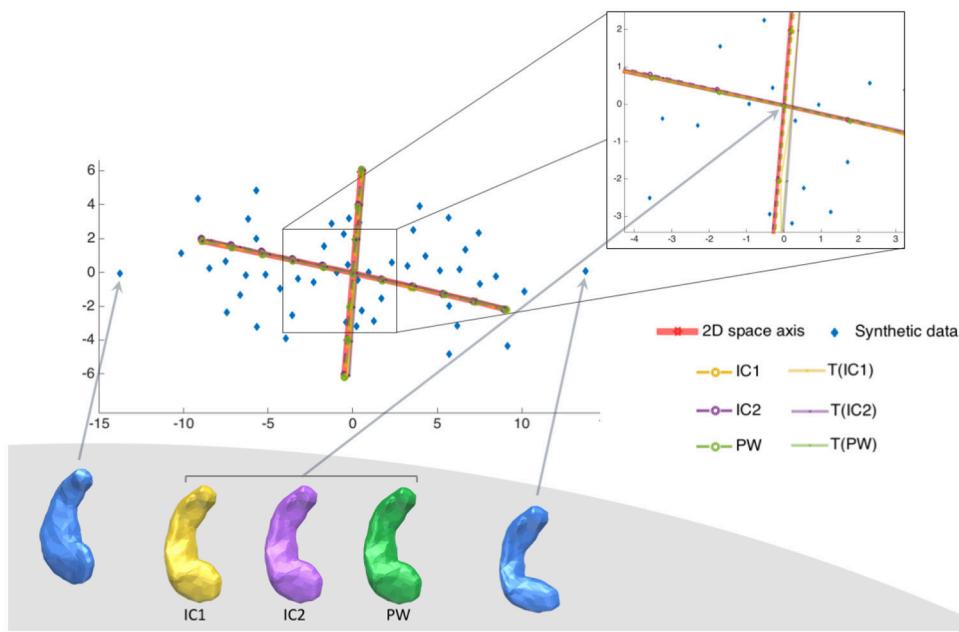
For comparison of computation time, we also computed a template using the standard initialization (the whole population as initialisation) which took 1220 min (20.3 h). Computing a centroid saved between 85 and 93% of computation time over the template with standard initialization and between 59 and 71% over the template initialized by the centroid.

Extra computation time was needed to estimate initial momentum vectors from IC1, IC2, and PW to each subject of the population, for subsequent analysis in the tangent spaces of each centres. These additional matchings took 19 min for IC1, 54 min for IC2, and 17 min for PW. The computation time is thus 94 min for IC1, 228 for IC2, and 105 min for PW which is considerably faster than estimating the variational template (1,220 min).

### 5.4.2. Centring of the Centres

As for the synthetic dataset, we assessed the centring of these six different centres. To that purpose, we first computed the ratio  $R$  of equation (26), for the centres estimated via the centroids methods, IC1 has a  $R = 0.25$ , for IC2 the ratio is  $R = 0.33$  and for PW it is  $R = 0.32$ . For centres estimated via the variational templates initialized by those centroids, the ratio for  $T(IC1)$  is  $R = 0.21$ , for  $T(IC2)$  is  $R = 0.31$  and for  $T(PW)$  is  $R = 0.26$ .

The ratios are higher than for the synthetic dataset indicating that centres are less centred. This was predictable since the population is not built from one surface via geodesic shootings as the synthetic dataset. In order to better understand these values,



**FIGURE 3 |** Synthetic dataset SD. Illustration of the two principal components of the 6 centres projected into the 2D space of the SD dataset. Synthetic population is in blue, and the real centre is at position (0,0). There are two principal components, projected into the 2D space of the population, computed from each of the 7 different estimated centres (marked in red for the exact centre, in yellow for IC1, in purple for IC2 and in green for PW). Each axis goes from  $-2\sigma$  to  $+2\sigma$  of the variability of the corresponding axe. This figure shows that the 6 different tangent spaces projected into the 2D space of shapes, are very similar, even if the centres have different positions.

**TABLE 3 |** Synthetic dataset SD.

$e(aM(\cdot), aM(\cdot))$	IC2	PW	T(IC1)	T(IC2)	T(PW)
IC1	$0.001 \pm 0.001$	$0.002 \pm 0.002$	$0.084 \pm 0.02$	$0.084 \pm 0.02$	$0.084 \pm 0.02$
IC2	0	$0.003 \pm 0.002$	$0.084 \pm 0.02$	$0.084 \pm 0.02$	$0.084 \pm 0.02$
PW		0	$0.084 \pm 0.02$	$0.084 \pm 0.02$	$0.084 \pm 0.02$
T(IC1)			0	$0.001 \pm 0.001$	$0.002 \pm 0.001$
T(IC2)				0	$0.002 \pm 0.001$

Mean  $\pm$  standard deviation errors  $e$  (equation 27) between the six different approximated distance matrices for each of the generated data sets.

we computed the ratio for each subject of the population (after matching each subject toward the population), as if each subject was considered as a potential centre. For the whole population, the average ratio was 0.6745, with a minimum of 0.5543, and a maximum of 0.7626. These ratios are larger than the one computed for the estimated centres and thus the 6 estimated centres are closer to a critical point of the Frechet functional than any subject of the population.

#### 5.4.3. PCA on Initial Momentum Vectors

As for the synthetic dataset, we performed six PCAs from the estimated centres.

**Figure 5** and **Table 4** show the proportion of cumulative explained variance for different number of modes. We can note that for any given number of modes, all PCAs result have really similar proportions of explained variance. The highest differences in cumulative explained variance, at a given number

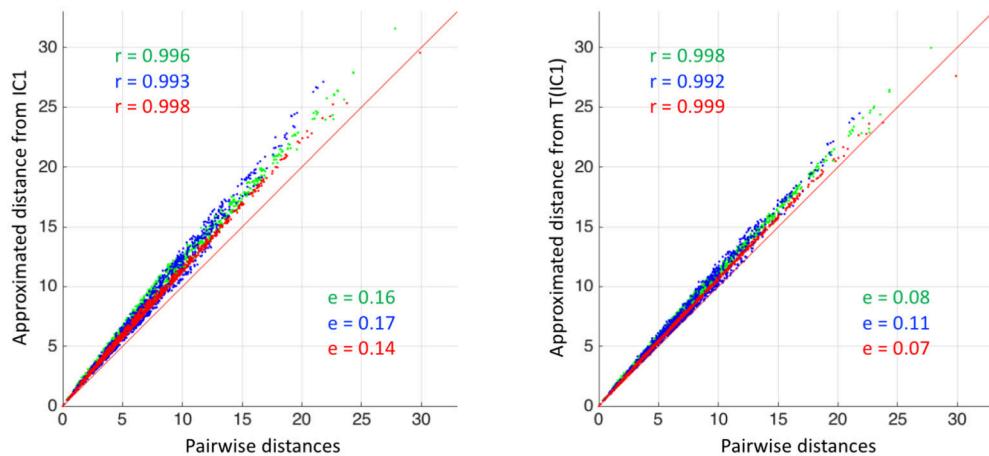
of component, between all 6 centres is 0.034 ( $>4\%$  of the total variance) and it is between IC2 and T(PW) for the 20th mode of variation (see **Table 4**). To reach 30% of total variance, a KPCA computed from  $T(C)$  for any  $C \in \{IC1, IC2, PW\}$  needs 3 components while  $C$  needs 4 components. To reach 50% KPCA computed from all 6 centres need 7 components. Finally to reach 90% of total variance,  $T(C)$  needs 21 components while  $C$  needs 22.

#### 5.4.4. Distance Matrices

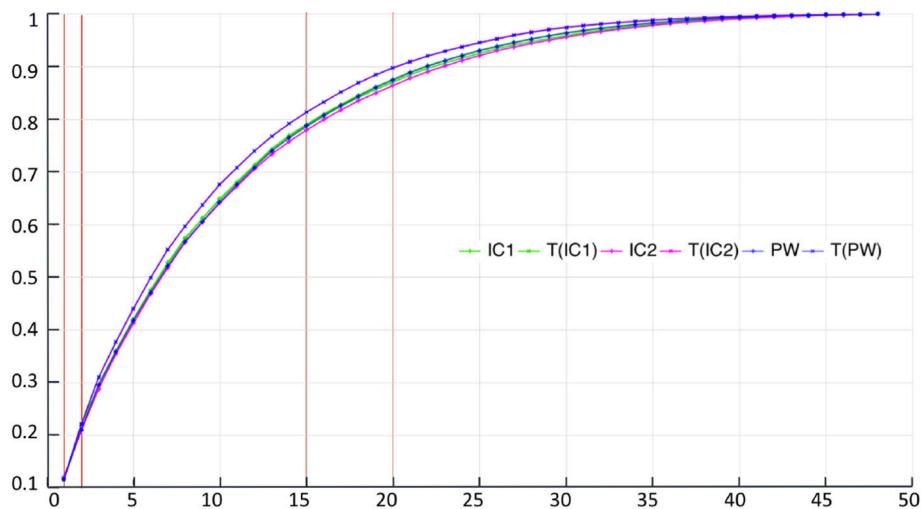
As for the synthetic dataset, we then studied the impact of these different centres on the approximated distance matrices. A direct distance matrix was also computed (around 90 h of computation time). We compared the approximated distance matrices of the different centres to: (i) the approximated matrix computed with IC1; (ii) the direct distance matrix.

We computed the errors  $e(M_1, M_2)$  defined in equation 27. Results are presented in **Table 5**. Errors are small and with the same order of magnitude.

**Figure 6-left** shows scatter plots between the direct distance matrix and the six approximated distance matrices. Interestingly, we can note that the results are similar to those obtained by Yang X. F. et al. (2011), **Figure 2**). **Figure 6-right** shows scatter plots between the approximated distance matrix from IC1 and the five others approximated distance matrices. The approximated matrices thus seem to be largely independent of the chosen centre.



**FIGURE 4 |** Synthetic dataset SD. **(Left)** Scatter plot between the approximated distance matrices  $aM(I/C1)$  of 3 different populations, and the pairwise distances matrices of the corresponding populations. **(Right)** Scatter plot between the approximated distance matrices  $aM(T(I/C1))$  of 3 different populations, and the pairwise distances matrices of the corresponding populations. For both plots, the matrix error (equation 27) and Pearson correlation coefficients (only here to support the visualization of the scatter plots) are indicated in color. The red line corresponds to the identity.



**FIGURE 5 |** Real dataset RD50. Proportion of cumulative explained variance for Kernel PCAs computed from the 6 different centres, with respect to the number of dimensions. Vertical red lines indicate values displayed at **Table 4**.

## 5.5. Real Dataset RD1000

Results on the real dataset RD50 and the synthetic SD showed that results were highly similar for the 6 different centres. In light of these results and because of the large size of the real dataset RD1000, we only computed IC1 for this last dataset. The computation time was about 832 min (13.8 h) for the computation of the centroid using the algorithm IC1, and 12.6 h for matching the centroid to the population.

The ratio  $R$  of equation 26 computed from the IC1 centroid was 0.1011, indicating that the centroid is well centred within the population.

We then performed a Kernel PCA on the initial momentum vectors from this IC1 centroid to the 1,000 shapes of the population. The proportions of cumulative

explained variance from this centroid are 0.07 for the 1st mode, 0.12 for the 2nd mode, 0.48 for the 10th mode, 0.71 for the 20th mode, 0.85 for the 30th mode, 0.93 for the 40th mode, 0.97 for the 50th mode and 1.0 from the 100th mode. In addition, we explored the evolution of the cumulative explained variance when considering varying numbers of subjects in the analysis. Results are displayed in Figure 7. We can first note that about 50 dimensions are sufficient to describe the variability of our population of hippocampal shapes from healthy young subjects. Moreover, for large number of subjects, this dimensionality seems to be stable. When considering increasing number of subjects in the analysis, the dimension increases and converges around 50.

Finally, we computed the approximated distance matrix. Its histogram is shown in **Figure 8**. It can be interesting to note that, as for RD50, the average pairwise distance between the subject is around 12, which means nothing by itself, but the points cloud on **Figure 6-left** and the histogram on **Figure 8**, show no pairwise distances below 6, while the minimal pairwise distance for the SD50 dataset - generated by a 2D space - is zero. This corresponds to the intuition that, in a space of high dimension, all subjects are relatively far from each other.

**TABLE 4 |** Real dataset RD50.

	<b>1st mode</b>	<b>2nd mode</b>	<b>15th mode</b>	<b>20th mode</b>
IC1	0.118	0.214	0.793	0.879
IC2	0.121	0.209	0.780	0.865
PW	0.117	0.209	0.788	0.875
T(IC1)	0.117	0.222	0.815	0.899
T(IC2)	0.115	0.220	0.814	0.898
T(PW)	0.116	0.221	0.814	0.898

Proportion of cumulative explained variance for kernel PCAs computed from the 6 different centres, for different number of dimensions.

**TABLE 5 |** Real dataset RD50.

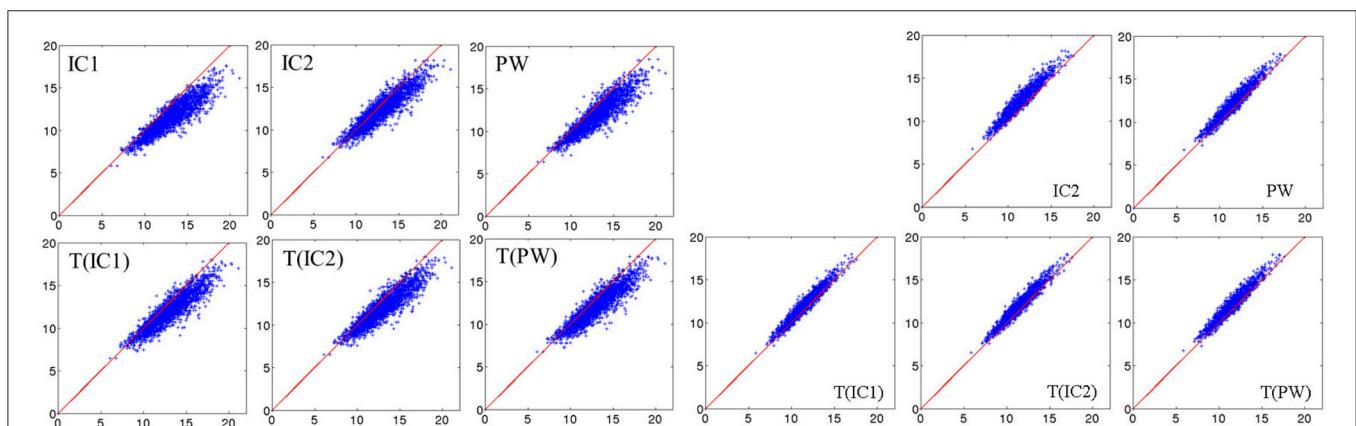
	<b><math>e(., aM(IC1))</math></b>		<b><math>e(., dM)</math></b>	
	<b>C</b>	<b>T(C)</b>	<b>C</b>	<b>T(C)</b>
IC1	0	0.04	0.10	0.08
IC2	0.06	0.04	0.06	0.08
PW	0.03	0.04	0.08	0.07

Errors  $e$  (equation 27) between the approximated distance matrices of each estimated centre and: (i) the approximated matrix computed with IC1 (left columns); (ii) the direct distance matrix (right columns).

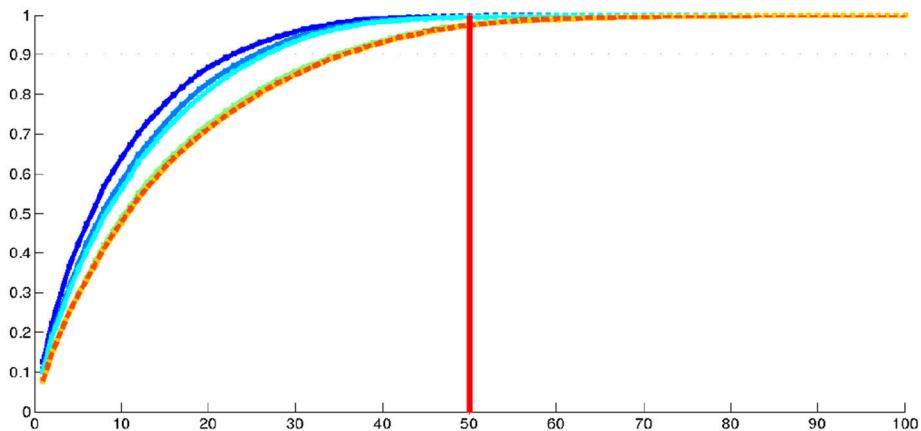
### 5.5.1. Prediction of IHI Using Shape Analysis

Incomplete Hippocampal Inversion (IHI) is an anatomical variant of the hippocampal shape present in 17% of the normal population. All those 1,000 subjects have an IHI score (ranking from 0 to 8) (Cury et al., 2015), which indicates the strength of the incomplete inversion of the hippocampus. A null score means there is no IHI, 8 means there is a strong IHI. IHI scores were assessed on a T1 weighted MRI, in coronal view after registration of all MRI in the standard MNI space. The IHI score is based on the roundness of the hippocampal body, the medial positionning of the hippocampus within the brain, and the depth of the colateral sulcus and the occipito-temporal sulcus, which both are close to the hippocampus. Here we selected only hippocampi with good segmentation, we therefore removed strong IHI which could not be properly segmented. We now apply our approach to predict incomplete hippocampal inversions (IHI) from hippocampal shape parameters. Specifically, we predict the visual IHI score, which corresponds to the sum of the individual criteria as defined in Cury et al. (2015). Only 17% of the population have an IHI score higher than 3.5. We studied whether it is possible to predict the IHI score using statistical shape analysis on the RD1000 dataset composed of 1,000 healthy subjects (left hippocampus).

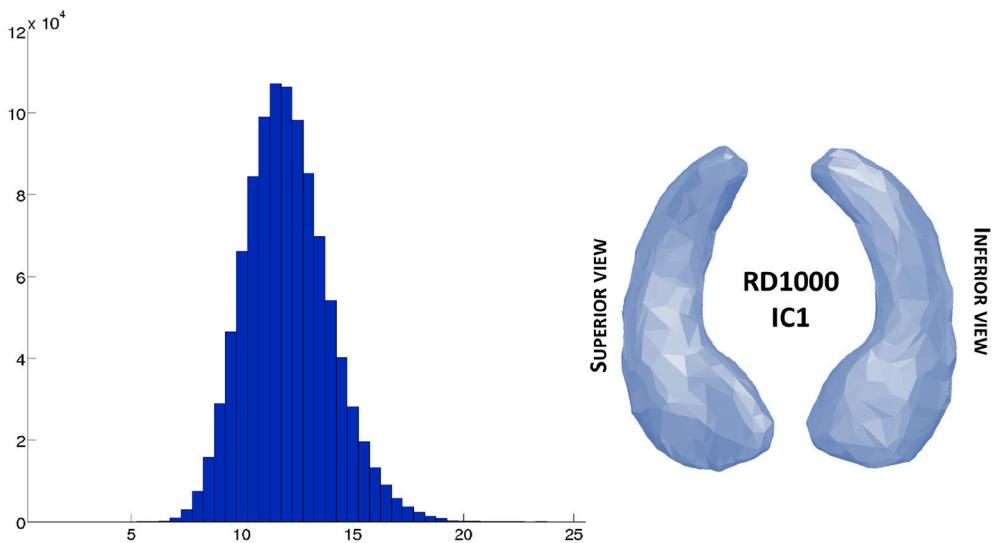
The deformation parameters characterizing the shapes of the population are the eigenvectors computed from the centroid IC1, and they are the independent variables we will use to predict the IHI scores. As we saw in the previous step, 40 eigenvectors are enough to explain 93% of the total anatomical variability of the population. We use the centred and normalized principal eigenvectors  $X_{1,i}, \dots, X_{40,i}$  computed from the RD1000 database with  $i \in \{1, \dots, 1000\}$  to predict the IHI score  $Y$ . We simply used a multiple linear regression model (Hastie et al., 2009) which is written as  $f(X) = \beta_0 + \sum_{i=1}^{40} X_{ji}\beta_i$  where  $\beta_0, \beta_1, \dots, \beta_{40}$  are the regression coefficients to estimate. The standard method to estimate the regression coefficients is the least squares estimation method in which the coefficients  $\beta_i$  minimize the residual sum of squares  $RSS(\beta) = \sum_{j=1}^N (y_j - \beta_0 - \sum_{i=1}^p x_{ji}\beta_i)^2$ , which leads



**FIGURE 6 |** Real dataset RD50. **(Left)** 6 scatter plots between direct distance matrix and approximated distance matrices from the six centres. The red line corresponds to the identity. **(Right)** 5 scatter plots between the approximated distance matrix  $aM(IC1)$  computed from IC1 and the 5 others approximated distance matrices. The red line corresponds to the identity.



**FIGURE 7 |** Real dataset RD1000. Proportion of cumulative explained variance of K-PCA as a function of the number of dimensions (in abscissa) and considering varying number of subjects. The dark blue curve was made using 100 subjects, the blue 200, the light blue 300, the green curve 500 subjects, the yellow one 800, very close to the dotted orange one which was made using 1,000 subjects.



**FIGURE 8 |** Real dataset RD1000. Left side shows histogram of the approximated distances of the large database RD1000 estimated from the computed centroid IC1 shown on the right side.

to the estimated  $\hat{\beta}$  (with matrix notations)  $\hat{\beta} = (X^T X)^{-1} X^T Y$ . For each number of dimensions  $p \in \{1, \dots, 40\}$  we validated the quality of the computed model with the adjusted coefficient of determination  $R_{adj}^2$ , which expresses the part of explained variance of the model with respect to the total variance:

$$R_{adj}^2 = 1 - \frac{SSE/(N-p)}{SST/(N-1)} \quad (28)$$

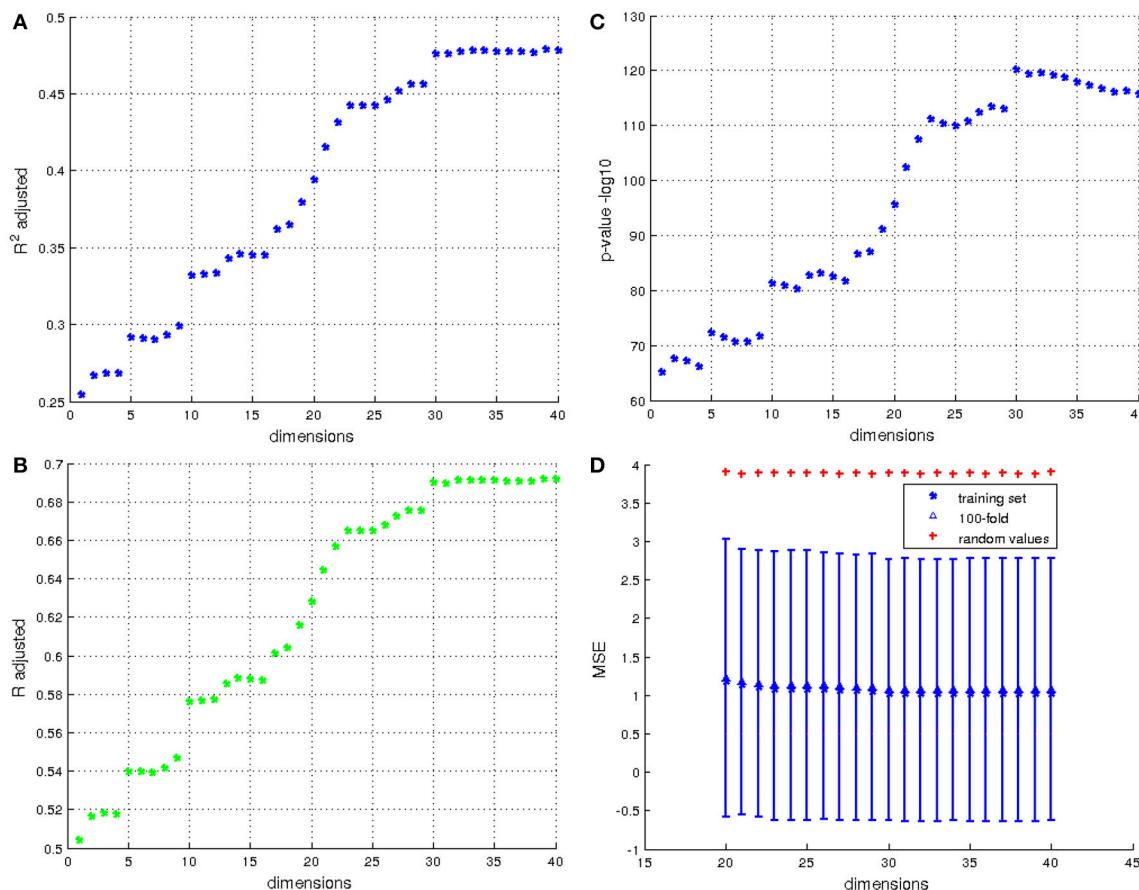
with  $SSE = \sum_i^N (y_i - (X_{1\dots p,i}^T \hat{\beta}))^2$  the residual variance due to the model and  $SST = \sum_{i=1}^N (y_i - \bar{Y})^2$  the total variance of the model. The  $R_{adj}^2$  coefficient, unlike  $R^2$ , takes into account the number of variables and therefore does not increase with the number of variables. One can note that  $R$  is the coefficient of correlation

of Pearson. We then tested the significance of each model by computing the F statistic

$$F = \frac{R^2/p}{(1-R^2)/(N-p-1)} \quad (29)$$

which follows a F-distribution with  $(p, n-p-1)$  degrees of freedom. So for each number of variables (i.e., dimensions of the PCA space) we computed the adjusted coefficient of determination to evaluate the model and the  $p$ -value to evaluate the significance of the model.

Then we used the  $k$ -fold cross validation method which consists in using  $1000-k$  subjects to predict the  $k$  remaining ones. To quantify the prediction of the model, we used the traditional



**FIGURE 9 |** Results for prediction of IHI scores. **(A)** Values of the adjusted coefficient of determination using from 1 to 40 eigenvectors resulting from the PCA. **(B)** the coefficient correlation corresponding to the coefficient of determination of A. **(C)** The  $p$ -values in  $-\log_{10}$  of the corresponding coefficient of determination. **(D)** Cross validation of the models using 20 to 40 dimensions by 100-fold. The red cross indicates the  $MSE$  of the model predicted using random values, and the errorbar corresponds to the standard deviation of  $MSE$  computed from 10,000 cross validations for each model, the triangle corresponds to the average  $MSE$ .

mean square error  $MSE = SSE/N$  which corresponds to the unexplained residual variance. For each model, we computed 10,000  $k$ -fold cross validation and displayed the mean and the standard deviation of  $MSE$  corresponding to the model.

Results are given at Figure 9, and display the coefficient of determination of each model. The cross validation is only computed on models with a coefficient of correlation higher than 0.5, so models using at least 20 dimensions. For the  $k$ -fold cross validation, we chose  $k = 100$  which represents 10% of the total population. Figure 9D presents results of cross validation; for each model computed from 20 to 40 dimensions we computed the mean of the 10,000  $MSE$  of the 100-fold and its standard deviation. To have a point of comparison, we also computed the  $MSE$  between the IHI scores and random values which follow a normal distribution with the same mean and standard deviation as the IHI scores (red cross on the Figure). The  $MSE$  of the cross validation are similar to the  $MSE$  of the training set. This result shows that using the first 30 to 40 principal components of initial momentum vectors computed from a centroid of the population, it is possible to predict the IHI score with a correlation of 69%.

The firsts principal components (between 1 and 20) represent general variance maybe characteristic of the normal population, the shape differences related to IHI appear after. It is indeed expected that the principal (i.e., the first once) modes of variation does not capture a modification of the anatomy present in only 17% of the population.

## 6. DISCUSSION AND CONCLUSION

In this paper, we proposed a method for template-based shape analysis using diffeomorphic centroids. This approach leads to a reasonable computation time making it applicable to large datasets. It was thoroughly evaluated on different datasets including a large population of 1,000 subjects. Codes can be found here: [https://github.com/cclairec/Iterative\\_Centroid](https://github.com/cclairec/Iterative_Centroid) and synthetic data can be found here: <https://doi.org/10.5281/zenodo.1420395>.

The results demonstrate that the method adequately captures the variability of the population of hippocampal shapes with a

reasonable number of dimensions. In particular, Kernel PCA analysis showed that the large population of left hippocampi of young healthy subjects can be explained, for the metric we used, by a relatively small number of variables (around 50). Moreover, when a large enough number of subjects was considered, the number of dimensions was independent of the number of subjects.

The comparisons performed on the two small datasets show that the different centroids or variational templates lead to very similar results. This can be explained by the fact that in all cases the analysis is performed on the tangent space to the template, which correctly approximates the population in the shape space. Moreover, we showed that the different estimated centres are all close to the Frechet mean of the population.

While all centres (centroids or variational templates) yield comparable results, they have different computation times. IC1 and PW centroids are the fastest approaches and can save between 70 and 90% of computation time over the variational template. Thus, for the study of hippocampal shape, IC1 or PW algorithms seem to be more adapted than IC2 or the variational template estimation. However, it is not clear whether the same conclusion would hold for more complex sets of anatomical structures, such as an exhaustive description of cortical sulci (Auzias et al., 2011). Besides, as mentioned above in section 5.3, unlike with the variational template estimation, centroid computations do not directly provide transformations between the centroid and the population which must be computed afterwards to obtain momentum vectors. This requires  $N$  more matchings, which almost doubles the computation time. Even with this additional step, centroid-based shape analysis stills leads to a competitive computation time (about 26 h for the complete procedure on the large dataset of 1,000 subjects).

Here we computed for all our datasets the initial momentum vectors from the mean shape to each subject of the population to apply a kernel PCA on the mean shape. But this is not a mandatory step, for example, in Cury et al. (2018) the iterative centroid method have been applied to provide a mean shape used afterwards as a baseline shape for an other analysis. It is also possible, for visualization purposes, to estimate different templates within a population (controls vs. patients for example) and identify where shape differences are located between templates.

In future work, this approach could be improved by using a discrete parametrisation of the LDDMM framework (Durrleman et al., 2011), based on a finite set of control points. The control points number and position are independent from the shapes being deformed as they do not require to be aligned with the shapes' vertices. Even if the method accepts any kind of topology, for more complex and heavy meshes like the cortical surface (which can have more than 20000 vertices per subjects), we could also improve the method presented here by using a multiresolution approach (Tan and Qiu, 2016). An other interesting point would be to study the impact of the choice of parameters on the number

of dimensions needed to describe the variability population (in this study the parameters were selected to optimize the matchings). Finally we can note that this template-based shape analysis can be extended to data types such as images or curves.

## ETHICS STATEMENT

London: Psychiatry, Nursing and Midwifery (PNM) Research Ethics Subcommittee (RESC), Waterloo Campus, Kings College London. Nottingham: University of Nottingham Medical School Ethics Committee. Mannheim: Medizinische Fakultaet Mannheim, Ruprecht Karl Universitaet Heidelberg and Ethik-Kommission II an der Fakultaet fuer Klinische Medizin Mannheim. Dresden: Ethikkommission der Medizinischen Fakultaet Carl Gustav Carus, TU Dresden Medizinische Fakultaet. Hamburg: Ethics board, Hamburg Chamber of Physicians. Paris: CPP IDF VII (Comit de protection des personnes Ile de France), ID RCB: 2007-A00778-45 September 24th 2007. Dublin: TCD School of Psychology REC. Berlin: ethics committee of the Faculty of Psychology. And Mannheims ethics committee approved the whole study.

## AUTHOR CONTRIBUTIONS

CC guarantor of integrity of entire study. Study concepts and design or data acquisition or data analysis and interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors. CC, JG, and OC literature research. MC, RT, GS, VF, and J-BP Clinical studies and pre-processings. CC, JG, RT, MC, and OC Statistical analysis and manuscript editing, all authors.

## ACKNOWLEDGMENTS

The research leading to these results has received funding from ANR (project HM-TC, grant number ANR-09-EMER-006, and project KaraMetria, grant number ANR-09-BLAN-0332), from the CATI Project (Fondation Plan Alzheimer) and from the program Investissements d'avenir ANR-10-IA IHU-06.

IMAGEN was supported by the European Union-funded FP6 (LSHM-CT-2007-037286), the FP7 projects IMAGEMEND (602450) and MATRICS (603016), and the Innovative Medicine Initiative Project EU-AIMS (115300-2), Medical Research Council Programme Grant Developmental pathways into adolescent substance abuse (93558), the NIHR Biomedical Research Centre Mental Health as well as the Swedish funding agency FORMAS. Further support was provided by the Bundesministerium für Bildung und Forschung (eMED SysAlc; AERIAL; 1EV0711).

It should be noted that a part of this work has been presented for the first time in CC's Ph.D. thesis (Cury, 2015).

## REFERENCES

- Afsari, B. (2011). Riemannian  $L^p$  center of mass: Existence, uniqueness, and convexity. *Proc. Am. Math. Soc.* 139, 655–673. doi: 10.1090/S0002-9939-2010-10541-5
- Afsari, B., Tron, R., and Vidal, R. (2013). On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM J. Control Optimizat.* 51, 2230–2260. doi: 10.1137/12086282X
- Arnaudon, M., Dombry, C., Phan, A., and Yang, L. (2012). Stochastic algorithms for computing means of probability measures. *Stochast. Process. Their Appl.* 122, 1437–1455. doi: 10.1016/j.spa.2011.12.011
- Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K., et al. (1998). Identifying global anatomical differences: deformation-based morphometry. *Hum. Brain Mapp.* 6, 348–357.
- Auzias, G., Colliot, O., Glaunes, J. A., Perrot, M., Mangin, J.-F., Trouvé, A., et al. (2011). Diffeomorphic brain registration under exhaustive sulcal constraints. *Med. Imaging IEEE Trans.* 30, 1214–1227. doi: 10.1109/TMI.2011.2108665
- Baulac, M., Grissac, N. D., Hasboun, D., Oppenheim, C., Adam, C., Arzimanoglou, A., et al. (1998). Hippocampal developmental changes in patients with partial epilepsy: Magnetic resonance imaging and clinical aspects. *Ann. Neurol.* 44, 223–233.
- Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision* 61, 139–157. doi: 10.1023/B:VISI.0000043755.93987.aa
- Charon, N., and Trouvé, A. (2013). The varifold representation of nonoriented shapes for diffeomorphic registration. *SIAM J. Imaging Sci.* 6, 2547–2580. doi: 10.1137/130918885
- Chung, M., Worsley, K., Paus, T., Cherif, C., Collins, D., Giedd, J., et al. (2001). A unified statistical approach to deformation-based morphometry. *NeuroImage* 14, 595–606. doi: 10.1006/nimg.2001.0862
- Chupin, M., Hammers, A., Liu, R. S. N., Colliot, O., Burdett, J., Bardinet, E., et al. (2009). Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation. *NeuroImage* 46, 749–761. doi: 10.1016/j.neuroimage.2009.02.013
- Colle, R., Cury, C., Chupin, M., Deflesselle, E., Hardy, P., Nasser, G., et al. (2016). Hippocampal volume predicts antidepressant efficacy in depressed patients without incomplete hippocampal inversion. *NeuroImage Clin.* 12(Suppl. C), 949–955. doi: 10.1016/j.nicl.2016.04.009
- Cury, C. (2015). *Statistical Shape Analysis of the Anatomical Variability of the Human Hippocampus in Large Populations*. Theses, Université Pierre et Marie Curie, Paris VI.
- Cury, C., Durrleman, S., Cash, D., Lorenzi, M., Nicholas, J., Bocchetta, M., et al. (2018). Spatiotemporal analysis for detection of pre-symptomatic shape changes in neurodegenerative diseases: applied to GENFI study. *Biorxiv[Pre print]*. doi: 10.1101/385427
- Cury, C., Glaunès, J., Chupin, M., and Colliot, O. (2017). Analysis of anatomical variability using diffeomorphic iterative centroid in patients with Alzheimer's disease. *Comput. Methods Biomed. Eng. Imaging Visual.* 5, 350–358. doi: 10.1080/21681163.2015.1035403
- Cury, C., Glaunès, J. A., Chupin, M., and Colliot, O. (2014a). “Fast template-based shape analysis using diffeomorphic iterative centroid,” in *MIUA 2014-Medical Image Understanding and Analysis 2014* (London, UK: Constantino Carlos Reyes-Aladasor and Grefe Slabaugh), 39–44.
- Cury, C., Glaunès, J. A., and Colliot, O. (2013). “Template estimation for large database: a diffeomorphic iterative centroid method using currents,” in *GSI Vol. 8085, Lecture Notes in Computer Science*, eds F. Nielsen and F. Barbaresco (Paris: Springer), 103–111.
- Cury, C., Glaunès, J. A., and Colliot, O. (2014b). “Diffeomorphic iterative centroid methods for template estimation on large datasets,” in *Geometric Theory of Information, Signals and Communication Technology*, ed F. Nielsen (Springer International Publishing), 273–299.
- Cury, C., Toro, R., Cohen, F., Fischer, C., Mhaya, A., Samper-González, J., et al. (2015). Incomplete Hippocampal Inversion: A Comprehensive MRI Study of Over 2000 Subjects. *Front. Neuroanat.* 9:160. doi: 10.3389/fnana.2015.00160
- de Rham, G. (1955). *Variétés différentiables. Formes, courants, formes harmoniques*. Hermann: Inst. Math. Univ. Nancago III.
- Durrleman, S. (2010). *Statistical Models of Currents for Measuring the Variability of Anatomical Curves, Surfaces and Their Evolution*. Ph. D. thesis, University of Nice-Sophia Antipolis.
- Durrleman, S., Pennec, X., Trouvé, A., and Ayache, N. (2008). “A forward model to build unbiased atlases from curves and surfaces,” in *2nd Medical Image Computing and Computer Assisted Intervention. Workshop on Mathematical Foundations of Computational Anatomy* (New York, NY), 68–79.
- Durrleman, S., Pennec, X., Trouvé, A., and Ayache, N. (2009). Statistical models of sets of curves and surfaces based on currents. *Med. Image Anal.* 13, 793–808. doi: 10.1016/j.media.2009.07.007
- Durrleman, S., Prastawa, M., Gerig, G., and Joshi, S. (2011). Optimal data-driven sparse parameterization of diffeomorphisms for population analysis. In eds D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, et al., *Information Processing in Medical Imaging, volume 6801* (Berlin; Heidelberg: Springer Berlin Heidelberg), 123–134.
- Glaunès, J. (2005). *Transport Par Diffeomorphismes de Points, de Mesures et de Courants Pour la Comparaison de Formes et l'anatomie numérique*. Ph. D. thesis, Université Paris 13.
- Glaunès, J. and Joshi, S. (2006). “Template estimation from unlabeled point set data and surfaces for Computational Anatomy,” in *Proceedings of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2006)* eds X. Pennec and S. Joshi (Copenhagen: LNCS Springer) 29–39.
- Glaunès, J., Trouve, A., and Younes, L. (2004). “Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004*, Vol. 2, II–712–II–718.
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The Elements of Statistical Learning*. Springer.
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1016/S1053-8119(02)91132-8
- Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Commun. Pure Appl. Math.* 30, 509–541.
- Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. Lond. Math. Soc.* 3, 371–406.
- Le, H. (2004). Estimation of Riemannian barycentres. *LMS J. Comput. Math.* 7, 193–200. doi: 10.1112/S1461157000001091
- Lorenzi, M. (2012). *Deformation-Based Morphometry of the Brain for the Development of Surrogate Markers in Alzheimer's Disease*. Ph. D. thesis, Université de Nice Sophia-Antipolis.
- Ma, J., Miller, M. I., Trouvé, A., and Younes, L. (2008). Bayesian template estimation in computational anatomy. *NeuroImage* 42, 252–261. doi: 10.1016/j.neuroimage.2008.03.056
- Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 2323–2326. doi: 10.1126/science.290.5500.2323
- Schölkopf, B., Smola, A., and Müller, K.-R. (1997). “Kernel principal component analysis,” in *Artificial Neural Networks—ICANN'97* (Lausanne: Springer), 583–588.
- Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., et al. (2010). The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry* 15, 1128–1139. doi: 10.1038/mp.2010.4
- Schwartz, L. (1952). Théorie des distributions. *Bull. Amer. Math. Soc.* 58, 78–85.
- Tan, M. and Qiu, A. (2016). Large deformation multiresolution diffeomorphic metric mapping for multiresolution cortical surfaces: a coarse-to-fine approach. *IEEE Transactions on Image Processing*, 25, 4061–4074. doi: 10.1109/TIP.2016.2574982
- Tenenbaum, J., Silva, V., and Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323. doi: 10.1126/science.290.5500.2319
- Trouvé, A. (1998). Diffeomorphisms groups and pattern matching in image analysis. *Int. J. Comput. Vision* 28, 213–221.
- Vaillant, M. and Glaunès, J. (2005). “Surface matching via currents,” in *Information Processing in Medical Imaging* Glenwood (Springs, CO: Springer), 381–392.

- Vaillant, M., Miller, M. I., Younes, L., and Trouvé, A. (2004). Statistics on diffeomorphisms via tangent space representations. *NeuroImage* 23, S161–S169. doi: 10.1016/j.neuroimage.2004.07.023
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* 17, 395–416. doi: 10.1007/s11222-007-9033-z
- Yang, X., Goh, A., and Qiu, A. (2011). Locally Linear Diffeomorphic Metric Embedding (LLDME) for surface-based anatomical shape modeling. *NeuroImage* 56, 149–161. doi: 10.1016/j.neuroimage.2011.01.069
- Yang, X. F., Goh, A., and Qiu, A. (2011). Approximations of the diffeomorphic metric and their applications in shape learning. *Inf. Proc. Med. Imaging* 22, 257–270. doi: 10.1007/978-3-642-22092-0\_22
- Younes, L. (2010). *Shapes and Diffeomorphisms*, Vol. 171. Springer.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Cury, Glaunès, Toro, Chupin, Schumann, Frouin, Poline, Colliot and the Imagen Consortium. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.