International Journal of Computer Vision
https://doi.org/10.1007/s11263-020-01343-w

Learning the spatiotemporal variability in longitudinal shape data sets

Alexandre Bône[1,2,3,4,5] · **Olivier Colliot**[1,2,3,4,5] · **Stanley Durrleman**[1,2,3,4,5] ·
for the Alzheimer's Disease Neuroimaging Initiative[†]

Received: 12 April 2019 / Accepted: 19 May 2020

Abstract In this paper, we propose a generative statistical model to learn the spatiotemporal variability in longitudinal shape data sets, which contain repeated observations of a set of objects or individuals over time. From all the short-term sequences of individual data, the method estimates a long-term normative scenario of shape changes and a tubular coordinate system around this trajectory. Each individual data sequence is therefore (i) mapped onto a specific portion of the trajectory accounting for differences in pace of progression across individuals, and (ii) shifted in the shape space to account for intrinsic shape differences across individuals that are independent of the progression of the observed process. The parameters of the model are estimated using a stochastic approximation of the expectation-maximization algorithm. The proposed approach is validated on a simulated data set, illustrated on the analysis of facial expression in video sequences, and applied to the modeling of the progressive atrophy of the hippocampus in Alzheimer's disease patients. These experiments show that one can use the method to reconstruct data at the precision of the noise, to highlight significant factors that may modulate the progression, and to simulate entirely synthetic longitudinal data sets reproducing the variability of the observed process.

Keywords Longitudinal data · Statistical shape analysis · Large deformation diffeomorphic metric mapping · Medical imaging · Disease progression modeling

[1] Institut du Cerveau, ICM, F-75013, Paris, France
[2] Inserm, U 1127, F-75013, Paris, France
[3] CNRS, UMR 7225, F-75013, Paris, France
[4] Sorbonne Université, F-75013, Paris, France
[5] Inria, Aramis project-team, F-75013, Paris, France

[†]Data used in preparation of this article were partly obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: adni.loni.usc.edu.

1 Introduction

1.1 Motivation

Video sequences of smiling faces, repeated measurements of growing plants or developing cells, medical images collected at multiple visits from a population of patients affected by a chronic disease: all these examples can be understood as data collections where individual instances of a common underlying process are observed at multiple timepoints. Such collections are called longitudinal data sets.

The individual processes are thought to result from random variations of a common underlying process (or few of them). Because of the dynamic nature of the observed processes, one might decompose the variability in two components: the dynamic or temporal variability on the one hand, and the time-independent or spatial variability on the other hand. In our examples, the variability in the pace of growth or in age at disease onset is understood as temporal variability. By contrast, there are also intrinsic inter-individual differences in height, weight or shape that are independent of the pace at which the plant grows or the disease progresses, which we call spatial variability. The main difficulty here is that growth or disease progression affect also height, weight or shape, so that the differences between two observations of two different samples are due to (i) the fact that the two individuals are observed at different stages of the process, and (ii) that they have different intrinsic characteristics. Disentangling these two sources of variability would not be possible if one had only one observation per individual. Having repeated observations of the individuals over time, as in longitudinal data sets, implies that one could decompose the changes due to the progression of the process from those due to intrinsic differences that are independent of the progression.

The goal of this paper is to propose a statistical learning method which can describe the spatiotemporal variability in a longitudinal data set. We focus here on shape data, where the shape may be encoded by an image, or by geometrical objects extracted from images such as curves, surface meshes or segmented volumes.

One of the main difficulties is that the experimental design often provides little control on the temporal sampling of the observations. We are interested here by processes for which there is no clear marker of progression, such as the progression of neurodegenerative diseases for which the age at disease onset is hard to determine. Therefore there is no easy way to re-align in time the individual data sequences to analyze the inter-individual variability at each stage of the process. By contrast, the method needs to learn how the individual data sequences position themselves in relation to each other. Furthermore, the follow-up period of the observations rarely covers the whole process, but often just a small part of it. In clinical studies for examples, patients may be followed for few years whereas the disease may progress over decades. Eventually, dealing with shape data raises the need for generic representations of such data that can be included in computational approaches.

1.2 Related work

Structured data like shapes can be advantageously represented as elements of curved spaces, such as Riemmanian manifolds, in order to account for the prior on their structure. Either defined by invariance [34, 56, 57] or topology-preserving properties [7, 15, 21, 33, 51], shape spaces define distance metrics adapted to the geometry of a well-identified class of objects, such as brain magnetic resonance images or segmented organs. These data representations allows the generalization of the mean-variance analysis [3, 28, 50, 65], which learns the geometrical distribution of a cross-sectional data set in terms of an average shape, and variability-encoding parameters. Typical healthy or pathological configurations can be summarized in this manner, thus opening the way to automatic diagnosis at the individual level. Time-series data sets, consisting in the repeated observation of the same object at successive time-points, can be described by generalized regression approaches on the same shape spaces [6, 26, 27, 29, 39, 49]. A time-continuous scenario of geometrical transformation is then estimated, offering in turn individualized interpolation and extrapolation methods. The statistical analysis of longitudinal data sets requires to extend the concept of generalized mean-variance for such time series. In other words, it requires the definition of a statistical distribution of curves drawn on a shape space.

Shape spaces are usually equipped with a differential structure of infinite dimension. In particular, the large deformation diffeomorphic metric mapping (LDDMM) approach defines shape spaces as orbits of template shapes under the action of an infinite-dimensional parametric group of diffeomorphisms of the 2D/3D ambient space [63]. With this approach, the geometrical differences between two objects are captured by estimating the diffeomorphic transformation that warps one into the other. More recent works propose finite-dimensional approaches built on the same principles: [64] uses truncated Fourier transforms to build a finite-dimensional Lie algebra, and [20] constructs a finite-dimensional Riemannian manifold based on a set of self-interacting particles.

Such structures are favorable to the analysis of longitudinal data sets because they naturally offer the parallel transport operator [40], which allows to compare tangent-space vectors at distant points in a relevant manner. This operator is key to compare trajectories on the manifold, and therefore to analyze longitudinal data. In [56, 57] for instance, trajectories on manifolds are compared by parallel-transporting their initial velocity vectors back to some privileged point of the manifold, thereby handling the spatial variability if a reference configuration and reference time-point is known. A similar approach is followed in [35] where medical images are analyzed in a voxel-wise fashion, or also in [54] with the co-adjoint transport instead of the parallel transport. In [16], the variability of a large longitudinal data set of thalamus shapes is analyzed by transporting individual residual deformations along a common and pre-computed trajectory, back to a baseline point. In [52, 53] the authors define the exp-parallelization operator which extends the notion of parallel lines to Riemannian manifolds. The works [10, 36] build on this operator to analyze dynamic networks and shape objects respectively. Other approaches propose to work on a space of trajectories, such as in [47] where the Sasaki metric is used to define distances between geodesic curves on a manifold, or in [12] which requires the same number of observations per subject.

If parallel transport allows to spatially align manifold-valued trajectories, a temporal alignment mechanism is also needed for data sets with variability in the individual progression dynamics. For instance, two patients developing the same neurological disease have no reason to reach the same disease stage at the same age, nor to have synchronous progressions. A solution is to use time-warp functions, which define a mapping between an abstract common reference time frame and the individual time lines [9, 10, 21, 35, 36, 45, 52, 53]. In [56, 57], the authors build on the square-root velocity fields framework to quotient the space of spatiotemporal paths by diffeomorphic time-warps. In [48], a monotonic Gaussian process is built from a set of temporal sources.

1.3 Contributions

In this paper, we propose a method that learns an average progression and its spatiotemporal variability from a longitudinal shape data set. The average progression takes the form of a geodesic curve in the finite-dimensional Riemannian approximation of the LDDMM framework of [22]. The concept of exp-parallelization introduced in [52, 53] is then applied in this context to define a tubular coordinate system, also called Fermi coordinates, around the average geodesic. The average trajectory and its coordinate system are automatically learned by the method, so that every individual data sequence is mapped to a specific portion of the average trajectory to account for the temporal variability, and shifted in the shape space to account for the spatial variability. The calibration of the resulting generative statistical model is done by adapting a stochastic approximation EM method. This paper extends the conference paper [10], with finer modeling of the variability in the individual paces of progression, and an original optimization method for accelerated model calibration. The proposed approach is validated on a simulated data set, illustrated on a facial expression recognition task, and applied to hippocampus shape progression modeling in Alzheimer's disease.

Section 2 defines the concept of shape spatiotemporal coordinate systems, which allows the introduction of the generative statistical model in Section 3. Section 4 details the calibration, personalization and simulation algorithms, which are evaluated and illustrated in Section 5. These experiments will evaluate the goodness-of-fit of the model, the relevance of the representation of the spatiotemporal variability for the identification of factors explaining this variability, and the ability of the model to generate synthetic data sets that reproduce the observed variability in the training data set.

2 Shape spatiotemporal reference frame

Within LDDMM frameworks, shape are positioned with respect to a reference shape, often called atlas or template. A coordinate system is defined in the tangent-space at the template shape. We propose here to replace the template (which is a single shape) by a curve (i.e. a shape trajectory), and the coordinate system by a tubular spatiotemporal coordinate system centered around the template trajectory. We first review the usual construction of a static template shape before extending it to the spatiotemporal case.

2.1 Positioning a shape with respect to a static atlas

Positioning a target shape y with respect to a static reference y_0 is called the registration problem. Deformation-based morphometry solves it by estimating a diffeomorphism ϕ_1 of the ambient space $\mathbb{R}^d$ ($d = 2$ or 3) that transforms y_0 into y, which we note $\phi_1 \star y_0 = y$. In the context of LDDMM, diffeomorphisms are constructed by following the streamlines of dynamic vector fields $t \to v_t \in C_0^\infty(\mathbb{R}^d, \mathbb{R}^d)$ over $[0, 1]$:

$$\partial_t \phi_t = v_t \circ \phi_t \quad \text{with} \quad \phi_0 = \text{Id}. \tag{1}$$

Following the approach in finite-dimension of [22], we further assume that any v_t writes as the Gaussian convolution of p momentum vectors $m_t = m_t^{(1)}, ..., m_t^{(p)} \in \mathbb{R}^d$ over a corresponding set of control points $c_t = c_t^{(1)}, ..., c_t^{(p)} \in \mathbb{R}^d$:

$$v_t : x \in \mathbb{R}^d \to \sum_{k=1}^{p} g\big[c_t^{(k)}, x\big] \cdot m_t^{(k)} \in \mathbb{R}^d \tag{2}$$

with $g : x, x' \in \mathbb{R}^d \to \exp \|x' - x\|^2_{\ell^2} / \sigma^2$ the Gaussian kernel function of kernel width $\sigma > 0$. Many other diffeomorphisms constructed in this manner might actually transform y_0 into $\phi_1 \star y_0$: we call solution of the registration problem the most regular transformation i.e. that minimizes its "kinetic" energy:

$$\frac{1}{2} \int_{t=0}^{1} \|v_t\|^2_{G_{c_t}} = \frac{1}{2} \int_{t=0}^{1} m_t^\top \cdot G_{c_t} \cdot m_t \tag{3}$$

where $\forall t \in [0, 1]$, G_{c_t} is the $p \times p$ "kernel" symmetric positive-definite matrix of general term $g[c_t^{(k)}, c_t^{(l)}]$, and $(.)^\top$ is the matrix transposition. Such energy-minimizing curves, also called geodesics, are such that the control points and the momentum vectors trajectories are fully determined by their initial values and the following Hamiltonian equations [46]:

$$\dot{c}_t = G_{c_t} \cdot m_t \quad ; \quad \dot{m}_t = -\frac{1}{2} \nabla_{c_t} \left\{ m_t^\top \cdot G_{c_t} \cdot m_t \right\} \tag{4}$$

where $\nabla_x(.)$ is the gradient operator with respect to x. Assuming that there exist a diffeomorphism ϕ_1 constructed according to equations (1, 2) such that $\phi_1 \star y_0 = y$, this last result allows to compactly represent the positioning of y with respect to y_0 with a set of p control points c_0 and attached momenta m_0. In other words, m_0 is the coordinate of y in the coordinate system defined by (c_0, y_0). In practice, the perfect registration constraint $\phi_1 \star y_0 = y$ is relaxed, and we call solution to the registration problem the extremal-path diffeomorphism ϕ_1 that warps y_0 as close as possible to y, for some extrinsic error measure $d_{\mathcal{E}}(y_0, y)$. In this paper, the following choices are considered for $d_{\mathcal{E}}$, depending on the nature of y_0 and y:

- the ℓ^2 metric for meshes with point-to-point correspondence (i.e. the sum of squared differences between point positions),
- the current metric [14, 59] for oriented surface meshes without point-to-point correspondence (details are given in appendix for the reader's convenience).

Noting y_0 as a collection $y_0^{(1)}, ..., y_0^{(K)}$ of K points of $\mathbb{R}^d$, ϕ_1 acts independently and directly on each point $y_0^{(k)}$ according to $\phi_1 \star y_0^{(k)} = \phi_1 \circ y_0^{(k)}$. Note that the methodology introduced in this section can be adapted straightforwardly to image data, by defining the action of the diffeomorphisms ϕ on the image I as $I \circ \phi^{-1}$ and using the sum of squared differences between image intensities as the error measure.

2.2 Riemannian structure

Let c_0 be a set of p control points. We define:

$$\mathcal{D}_{c_0} = \big\{\phi_1 \,|\, \partial_t \phi_t = v_t \circ \phi_t,\ \phi_0 = \mathrm{Id},\ v_t = \mathrm{Conv}(c_t, m_t) \\ (\dot{c}_t, \dot{m}_t) = \mathrm{Ham}(c_t, m_t),\ m_0 \in \mathbb{R}^{p\times d}\big\} \qquad (5)$$

where $\mathrm{Conv}(.,.)$ and $\mathrm{Ham}(.,.)$ are compact notations for the convolution operator defined by equation (2) and the Hamiltonian equations (4) respectively. Equipped at any $\phi \in \mathcal{D}_{c_0}$

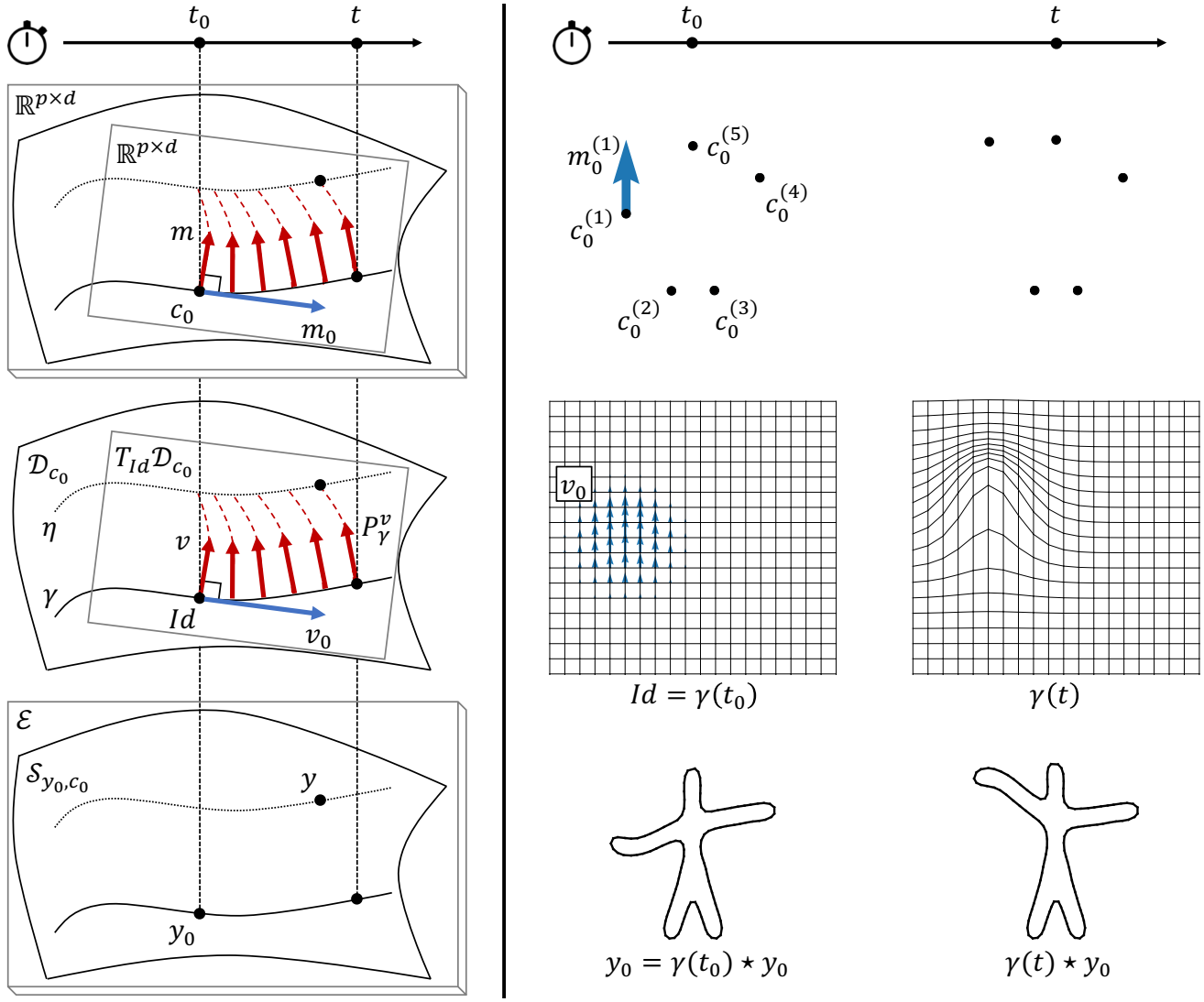


Fig. 1: **[Left].** Shape spatiotemporal reference frame y_0, c_0, m_0, t_0 with respect to which a shape y admits coordinates $t \in \mathbb{R}$, $v \in v_0^\perp \subset T_{\gamma(t_0)}\mathcal{D}_{c_0}$. Three spaces are involved: the manifold of control points $\mathbb{R}^{p\times d}$ (top), the manifold of diffeomorphisms $\mathcal{D}_{c_0}$ (middle), and the shape submanifold $\mathcal{S}_{y_0,c_0}$ of the extrinsic shape space $\mathcal{E}$ (bottom). The momenta m_0, m and the velocity fields v_0, v are in one-to-one correspondence. The velocity field v, also called space-shift, is parallel-transported along the geodesic γ by the operator $t \to P_\gamma^v(t)$. Figure 2 illustrates the effect of parallel transport on $\mathcal{D}_{c_0}$.
[Right]. Illustrations of the manifolds abstractly depicted on the left side of the figure. Each row displays two example elements of the corresponding geodesic (solid black lines on the left panel). The two columns correspond respectively to the times t_0 and t.

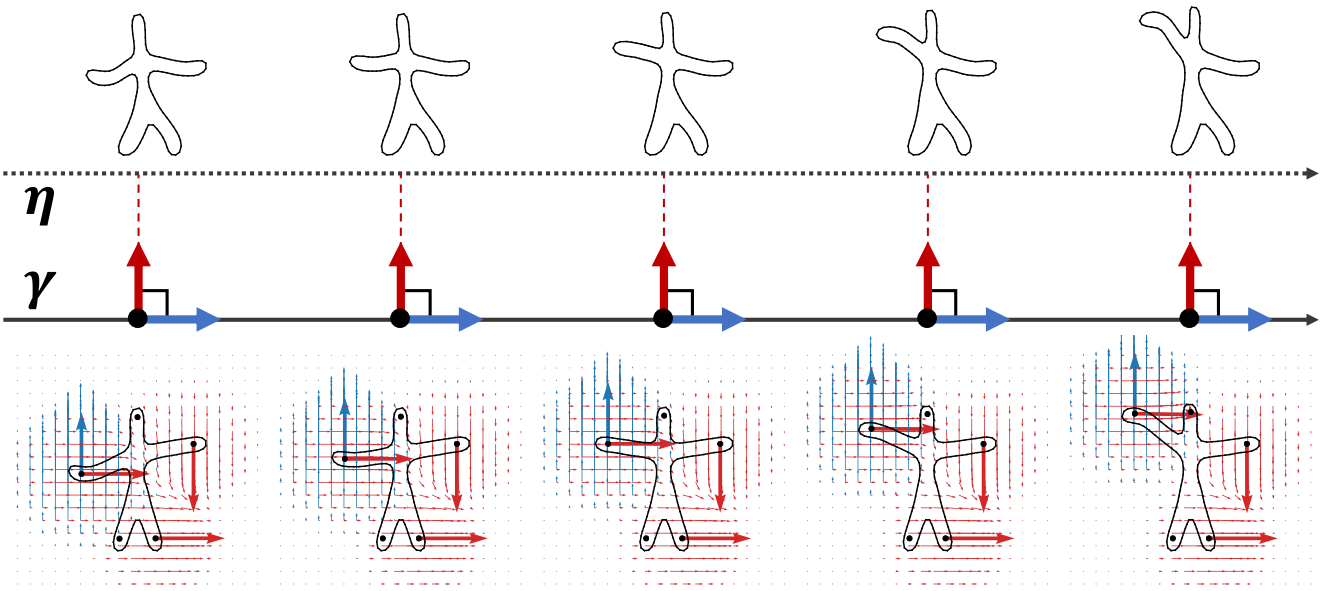


Fig. 2: **[Bottom].** Illustration of a shape geodesic $t \to \gamma(t) \star y_0$: the man-like shape (solid black contour) raises his left arm. This geodesic is parametrized by a single set of control points c_0 (black dots) and attached momentum vectors m_0 (bold blue arrows), to which corresponds the velocity field v_0 (light blue arrows). A second set of momentum vectors m (bold red arrows) attached to the same control points c_0 parametrizes the exp-parallelization of this shape geodesic.
[Top]. Exp-parallel shape curve $t \to \eta(t) \star y_0$ to the shape geodesic $\gamma \star y_0$: the exp-parallelization transfers the arm-raising motion from one man-like shape to another.

with the local metric $G^{-1}_{\phi(c_0)}$, $\mathcal{D}_{c_0}$ has the structure of a Riemannian manifold of dimension $p \times d$. The tangent-space at ϕ is the set of velocity fields obtained by convolving any set of momentum on $\phi(c_0)$:

$$T_\phi \mathcal{D}_{c_0} = \left\{ \mathrm{Conv}(\phi(c_0), m) \,|\, m \in \mathbb{R}^{p \times d} \right\}. \tag{6}$$

The geodesics of $\mathcal{D}_{c_0}$ are the curves $t \to \phi_t$ of constant kinetic energy (see equation (3)) i.e. such that the corresponding control points and momenta trajectories $t \to c_t, m_t$ satisfy the Hamiltonian equations (4). We define the exponential operator on $\mathcal{D}_{c_0}$:

$$\mathrm{Exp}_\phi^{t_0,t} : v_0 \in T_\phi \mathcal{D}_{c_0} \to \phi_t \in \mathcal{D}_{c_0} \tag{7}$$

where ϕ_t is the diffeomorphism reached at time t by the geodesic path obtained by integration from some reference time $t_0 \in \mathbb{R}$ with initial conditions $\phi(c_0)$, m_0 such that $v_0 = \mathrm{Conv}(\phi(c_0), m_0)$, and $\phi_0 = \phi$. The momentum vector m_0 is the dual of the velocity field v_0. The particular case $\mathrm{Exp}_\phi^{0,1}$ corresponds to the usual Riemannian exponential map and will be noted Exp_ϕ. Diffeomorphisms $\phi \in \mathcal{D}_{c_0}$ act on shapes of the ambient space y through the action $\star$ previously defined. Let y_0 be a reference shape. We define its orbit under the action $\star$:

$$\mathcal{S}_{y_0,c_0} = \mathcal{D}_{c_0} \star y_0 = \{ \phi \star y_0 \,|\, \phi \in \mathcal{D}_{c_0} \}. \tag{8}$$

$\mathcal{S}_{y_0,c_0}$ is a submanifold of the extrinsic shape space $\mathcal{E}$ in which is defined the distance $d_\mathcal{E}$.

2.3 Positioning a shape with respect to a dynamic atlas

Instead of positioning shapes with respect to a static atlas y_0, we aim now to position shapes with respect to a shape geodesic $t \to \gamma(t) \star y_0$, where γ is a geodesic of $\mathcal{D}_{c_0}$ of the form $\gamma(t) = \mathrm{Exp}_{\mathrm{Id}}^{t_0,t}(v_0)$ with $v_0 = \mathrm{Conv}(c_0, m_0)$. Similarly to cylindrical coordinates in Euclidian spaces, under some conditions (see [30, 43]) a shape $y \in \mathcal{S}_{y_0,c_0}$ admits a unique spatiotemporal coordinate, also known as Fermi coordinates, $t \in \mathbb{R}$ and $v \in T_{\gamma(t_0)} \mathcal{D}_{c_0}$ such that $v \perp \dot{\gamma}(t_0)$:

$$y = \mathrm{ExpP}_\gamma^v(t) \star y_0 \quad \text{with} \quad \mathrm{ExpP}_\gamma^v(t) = \mathrm{Exp}_{\gamma(t)}\left[\mathrm{P}_\gamma^v(t)\right] \tag{9}$$

where $\mathrm{P}_\gamma^v(t)$ denotes the parallel transport of v along γ from t_0 to t. The curves γ and $\eta : t \to \mathrm{ExpP}_\gamma^v(t)$ are said exp-parallel, and the mapping $\gamma \to \eta$ is called exp-parallelization along v [52, 53]. In other words, a choice of y_0, c_0, m_0, t_0 defines a spatiotemporal reference frame, with respect to which a shape y can be unambiguously positioned in terms of a time t and a velocity field v orthogonal to $v_0 = \dot{\gamma}(t_0)$. The time t is the temporal component of the coordinate which positions the shape along the reference trajectory given by the direction v_0. The velocity v is the spatial component of the coordinate, which positions the shape in the hyperplane that is orthogonal to v_0. This decomposition can also be understood as the orthogonal projection of y onto the one-dimensional submanifold $\gamma \star y_0$, hence the condition $v \perp v_0$.

Figures 1 and 2 illustrate this concept of spatiotemporal reference frame, in which any shape y admits the coordinates t, v. Note that the time-point t_0 does not play any particular role, in the sense that y can be described in the same manner for any other choice t'_0; a one-to-one transformation of the spatiotemporal reference frame can actually be derived as:

$$t' = t + t'_0 - t_0 \quad \text{and} \quad v' = \mathrm{P}^v_\gamma(t'_0). \tag{10}$$

In general, the target shape y might not exactly belong to $\mathcal{S}_{y_0,c_0}$. Similarly to the static atlas case, equation (9) is relaxed and we call solution to the longitudinal registration problem the pair t, v such that y_0 is warped as close as possible to y, in the sense of the extrinsic metric $d_\mathcal{E}$.

3 Statistical model for longitudinal data sets of shapes

3.1 Hierarchical generative model

Let $\{y_{i,j}, t_{i,j}\}_{i,j}$ be a longitudinal data set of shapes, which is the collection of repeated individual measurements $y_{i,1}$, ..., y_{i,n_i} for $i = 1, ..., n$, where each shape $y_{i,j}$ corresponds to a time $t_{i,j} \in \mathbb{R}$. Measurements are considered as sample points along individual trajectories, which are in turn considered exp-parallel to a reference geodesic curve, therefore having a constant spatial coordinate in the spatiotemporal reference frame centered around this reference geodesic. Noting y_0, c_0, m_0, t_0 the parameters of the spatiotemporal coordinate system, $v_0 = \mathrm{Conv}(c_0, m_0)$ and $\gamma : t \to \mathrm{Exp}^{t_0,t}_{c_0}(v_0)$ the reference geodesic, the statistical model writes:

$$\begin{aligned}
&\mathrm{ExpP}^{v_i}_\gamma\big[\psi_i(t_{i,j})\big] \star y_0 \overset{\text{iid}}{\sim} \mathcal{N}_\mathcal{E}\big(y_{i,j},\, \sigma^2_\epsilon\big),\\
&\text{where} \left|\begin{array}{l} \psi_i : t \to \alpha_i \cdot (t - \tau_i) + t_0, \\ v_i = \mathrm{Conv}(c_0, m_i), \quad m_i = A_{0,m_0^\perp} \cdot s_i, \end{array}\right. \\
&\text{and} \left|\begin{array}{l} \alpha_i \overset{\text{iid}}{\sim} \mathcal{N}_{[0,+\infty[}(1, \sigma^2_\alpha), \quad \tau_i \overset{\text{iid}}{\sim} \mathcal{N}(t_0, \sigma^2_\tau), \\ s_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1) \end{array}\right.
\end{aligned} \tag{11}$$

where the noise distribution $\mathcal{N}_\mathcal{E}\big(\mu,\, \sigma^2_\epsilon\big)$ is defined such that the likelihood is proportional to $p(y) \propto \exp(-d_\mathcal{E}(y-\mu)^2/2\sigma^2_\epsilon)$. Model (11) is hierarchical in the sense that individual trajectories $t \to \mathrm{ExpP}^{v_i}_\gamma \circ \psi_i(t)$ are independently defined as spatiotemporal transformations of a common, population-level geodesic $t \to \gamma(t)$.

The time-warp functions ψ_i encode the temporal variability of the observed individual trajectories in terms of pace of progression α_i and onset time τ_i. They map the index $t_{i,j}$ of the j-th shape of the i-th individual (e.g. the age of the subject at a given visit), to a time-point $\psi_i(t_{i,j})$ on the reference geodesic (e.g. the disease stage).

The spatial variability is encoded by the space-shifts $v_i \in v_0^\perp \subset T_{\gamma(t_0)}\mathcal{D}_{c_0}$ along which γ is exp-parallelized. Those space-shifts admit dual representations under the form of the momenta m_i, which are assumed to derive from q source parameters $s_i = s_i^{(1)}, ..., s_i^{(q)}$, in the spirit of independent component analysis (ICA) [31]. The orthogonality $v_i \perp v_0$, necessary for the identifiability of the model, is ensured by the projection of each column of the $(p \cdot d) \times q$ mixing matrix A_0 onto the hyperplane $m_0^\perp$ of $\mathbb{R}^{p \times d}$ for the cometric G_{c_0}, noted $A_{0,m_0^\perp}$. The individual parameters are modeled as independent samples from normal distributions:

- a truncated normal distribution with fixed mean for the acceleration factor α_i, allowing the identifiability of m_0;
- a normal distribution for the onset time τ_i;
- a normal distribution with fixed mean and variance for the sources s_i, allowing the identifiability of y_0 and $A_{0,m_0^\perp}$ respectively.

These parameters define individual trajectories as random spatiotemporal transformations of a common reference trajectory. The spatial and temporal transformations commute, in the sense that $\forall t \in \mathbb{R}$, $\mathrm{ExpP}^{v_i}_{\gamma \circ \psi_i} = \mathrm{ExpP}^{v_i}_\gamma \circ \psi_i$. The population trajectory is fully parameterized by the template shape y_0, the control points c_0, the momenta m_0 and the reference time t_0. The individual variability is unambiguously represented by two reduced sets of scalar parameters: the acceleration α_i and the onset time τ_i for the temporal part, and the sources $s_i^{(1)}, ..., s_i^{(q)}$ for the spatial part. In practice, it is possible to choose a number of sources $q \ll p \times d$ much lower than the dimension of the tangent-space $T_{\gamma(t_0)}\mathcal{D}_{c_0}$ while still capturing most of the geometrical variability in the data.

3.2 Mixed-effects and Bayesian modeling

We further specify the formulation of the model (11) to fit the framework of mixed-effects models. We distinguish:

- the fixed-effects $\theta = (\theta_1, \theta_2)$ with $\theta_1 = (t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$ and $\theta_2 = (y_0, c_0, m_0, A_0)$, also called the model parameters,
- the random effects $z = (z_i)_i$ where $z_i = (\alpha_i, \tau_i, s_i)$.

We choose to work in a Bayesian framework, in order to theoretically ensure the existence of the maximum a posteriori (MAP) estimate of the parameters θ^{m}. Such priors also regularize and guide the estimation procedure thanks to reasonable and mild prior assumptions on the optimal fixed effects values. The following standard conjugate distributions are selected as Bayesian priors on the model parameters:

$$\begin{aligned}
t_0 &\sim \mathcal{N}(\overline{t_0}, \varsigma^2_t), & y_0 &\sim \mathcal{N}(\overline{y_0}, \varsigma^2_y),\\
\sigma^2_\tau &\sim \mathcal{IG}(m_\tau, \varsigma^2_\tau), & c_0 &\sim \mathcal{N}(\overline{c_0}, \varsigma^2_c),\\
\sigma^2_\alpha &\sim \mathcal{IG}(m_\alpha, \varsigma^2_\alpha), & m_0 &\sim \mathcal{N}(\overline{m_0}, \varsigma^2_m),\\
\sigma^2_\epsilon &\sim \mathcal{IG}(m_\epsilon, \varsigma^2_\epsilon), & A_0 &\sim \mathcal{N}(\overline{A_0}, \varsigma^2_A),
\end{aligned}$$

where $\mathcal{IG}(.,.)$ denotes the inverse-gamma distribution.

4 Algorithms: calibration, personalization, simulation

4.1 Objectives

Given a longitudinal data set of shapes $\{y_{i,j}, t_{i,j}\}_{i,j}$ that we may note more compactly $\{y, t\}$, we formulate three algorithmic objectives:

- *Calibration*, which consists in computing the MAP parameters θ^{m}, unconditionally to any random effect z:

$$\theta^{\mathrm{m}} = \mathrm{argmax}_\theta \int p\big(\{y\}, z, \theta\,; \{t\}\big) \cdot \mathrm{d}z. \tag{12}$$

- *Personalization*, which consist in computing the MAP random effects z^{m} that best represent some longitudinal shape data set $\{y, t\}$ (which may or may not be the one used for calibration), given the calibrated model θ^{m}:

$$z^{\mathrm{m}} = \mathrm{argmax}_z\ p\big(\{y\}, z, \theta^{\mathrm{m}}\,; \{t\}\big). \tag{13}$$

- *Simulation*, which consist in generating a new data set $\{y^s\}$ that resembles the original data set $\{y\}$.

We give now the details of the algorithms to solve these optimization problems. Their implementation is freely available in the software Deformetrica (find the install instructions and the documentation at `www.deformetrica.org`).

4.2 Computation of the complete log-likelihood

Evaluating the joint log-likelihood $\log p(\{y\}, z, \theta\,; \{t\}) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \log p(y_{i,j}, z_i, \theta\,; t_{i,j})$ for some set of parameters θ and random effects $z = (z_i)_i$ is central for both calibration and personalization algorithms. The computationally most intensive part is the computation of the conditional log-likelihood $\log p(y_{i,j}|z_i, \theta\,; t_{i,j})$, which amounts to synthesize the candidate data for the current values of the fixed and random-effects (θ, z_i) and measure its discrepancy with the true observation $y_{i,j}$. The synthesis of the data follows the generative model introduced in Section 2 and essentially requires the integration of ordinary differential equations.

Algorithm 1 details the procedure, where $|\mathcal{E}|$ denotes the dimension of the extrinsic shape shape $\mathcal{E}$, and $F(.)$ the cumulative distribution function of the standard Gaussian. The "source index" refers to the ICA components.

Algorithm 1: Compute the complete log-likelihood.

input : Longitudinal data set of shapes $\{y, t\} = \{y_{i,j}, t_{i,j}\}_{i,j}$.
Population parameters $\theta = y_0, c_0, m_0, a_0, t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon$.
Individual parameters $z = (z_i)_i$ with $z_i = \alpha_i, \tau_i, s_i$.
output: The complete log-likelihood $Q = \log p(\{y\}, z, \theta\,; \{t\})$.
Set $Q = 0$. `// initialization`

`/* compute the squared residuals` $\epsilon^2_{i,j}$ `for each visit */`
Compute the initial velocity field $v_0 = \mathrm{Conv}(c_0, m_0)$.
Compute the geodesic $\gamma : t \to \mathrm{Exp}_{\mathrm{Id}}^{t_0,t}(v_0)$. `// see [22]`
for the source index $l = 1$ **to** q
  Compute the l-th column of $A_{0,m_0^\perp}$, projecting $\mathrm{Col}_l(A_0)$ on $m_0^\perp$.
  Compute the initial velocity field $w_l = \mathrm{Conv}\big[c_0,\ \mathrm{Col}_l(A_{0,m_0^\perp})\big]$.
  Compute the parallel transport $w_l : t \to \mathrm{P}_\gamma^{w_l}(t)$. `// see [41]`
end
for the individual index $i = 1$ **to** n
  for the visit index $j = 1$ **to** n_j
    Compute the time-warped age $\psi_{i,j} = \alpha_i \cdot (t_{i,j} - \tau_i) + t_0$.
    Compute the initial velocity field $v_{i,j} = \sum_{l=1}^{q} s_i^{(l)} \cdot w_l(\psi_{i,j})$.
    Compute $\phi_{i,j} = \mathrm{Exp}_{\gamma(\psi_{i,j})}(v_{i,j}) \circ \gamma(\psi_{i,j})$. `// see [22]`
    Compute the squared residual $\epsilon^2_{i,j} = d_{\mathcal{E}}(y_{i,j},\ \phi_{i,j} \star y_0)^2$.
    `/* add the model log-likelihood` $\log p(y_{i,j}|z_i, \theta\,; t_{i,j})$ `*/`
    Update $Q \leftarrow Q - \frac{1}{2}\big\{\, |\mathcal{E}| \cdot \log \sigma_\epsilon^2 + \epsilon^2_{i,j}/\sigma_\epsilon^2 \big\}$.
  end
  `/* add the random effects log-likelihood` $\log p(z_i|\theta)$ `*/`
  Update $Q \leftarrow Q - \frac{1}{2}\big\{\, \log \sigma_\tau^2 + (\tau_i - t_0)^2/\sigma_\tau^2 + \|s_i\|^2_{\ell^2} + \log \sigma_\alpha^2 + \log(1 - F(-1/\sigma_\alpha))^2 + (\alpha_i - 1)^2/\sigma_\alpha^2 \big\}$.
end
`/* add the Bayesian prior log-likelihood` $\log p(\theta)$ `*/`
Update $Q \leftarrow Q - \frac{1}{2}\big\{(t_0 - \overline{t_0})^2/\varsigma_t^2 + m_\tau\,(\log \sigma_\tau^2 + \varsigma_\tau^2/\sigma_\tau^2) + m_\alpha\,(\log \sigma_\alpha^2 + \varsigma_\alpha^2/\sigma_\alpha^2) + m_\epsilon\,(\log \sigma_\epsilon^2 + \varsigma_\epsilon^2/\sigma_\epsilon^2)\big\}$. `// log` $p(\theta_1)$
Update $Q \leftarrow Q - \frac{1}{2}\big\{\, \|y_0 - \overline{y_0}\|^2_{\ell^2}/\varsigma_y^2 + \|c_0 - \overline{c_0}\|^2_{\ell^2}/\varsigma_c^2 + \|m_0 - \overline{m_0}\|^2_{\ell^2}/\varsigma_m^2 + \big\|A_0 - \overline{A_0}\big\|^2_{\ell^2}/\varsigma_A^2 \big\}$. `// log` $p(\theta_2)$

4.3 Calibration

4.3.1 Initialization procedure for model calibration

A good choice of initial parameters $\theta^{[0]}$ and latent variables $z^{[0]}$ improves the convergence speed of the calibration algorithm. We propose in this section an initialization procedure that combines several elementary shape analysis tools. Given a longitudinal data set of shapes $\{y_{i,j}, t_{i,j}\}_{i,j}$:

1. estimate a Bayesian atlas model (see [28]) from the baseline shapes $\{y_{i,1}\}_i$, to get an approximate population-level average geometry y_0', c_0' as well as n space-shift momenta m_i' mapping this geometry to the baseline observations, and an estimate of the noise level σ_ϵ';
2. for $i = 1, ..., n$, estimate a geodesic regression model (see [26]) from the individual time-series $\{y_{i,j}\}_j$, then parallel transport (see [41]) the computed individual initial momenta back to the mean geometry y_0', c_0' along the corresponding space-shift m_i', and finally compute the Euclidean average of those w_i' to get an approximate population-level mean momenta $m_0' = \langle w_i' \rangle_i$;
3. for $i = 1, ..., n$, initialize the individual temporal parameters with $\tau_i = \langle t_{i,j} \rangle_j$, $\alpha_i^2 = \frac{w_i' \cdot G_{c_0'} \cdot m_0'}{m_0' \cdot G_{c_0'} \cdot m_0'}$ if this value is positive and $\alpha_i = 1$ otherwise, then compute σ_τ' and σ_α' according to equations (16) and (17) respectively;
4. solve a standard ICA problem with q components from the collection of space-shift momenta $w'_{i,m_0'^\perp}$ preliminarily projected on the orthogonal space to m_0', and set A_0' as the estimated mixing matrix;

5. shoot forward the mean geometry y_0', c_0' in the direction m_0' with length $t_0'' - t_0'$ where $t_0' = \langle t_{i,1} \rangle_i$ and $t_0'' = \langle t_{i,j} \rangle_{i,j}$ to get longitudinally centered estimates c_0'', y_0'', m_0'', and parallel-transport the q columns of A_0' along the same geodesic to obtain A_0'';
6. personalize the model given by the initial parameters $\theta^{[0]} = (y_0'', c_0'', m_0'', A_0'', t_0'', \sigma_\tau', \sigma_\alpha', \sigma_\epsilon')$ to obtain $z^{[0]}$.

4.3.2 The MCMC-SAEM-GD algorithm

Calibration is a computationally-intensive task for mostly two reasons. First, the optimized variable θ is of high dimension $|\theta| = 4 + |y_0| + d \cdot p \cdot (2+q)$ where d is the dimension of the ambient space, p the number of control points, q the number of sources, and $|y_0|$ the number of vertices necessary to describe the template mesh. Second, the optimized function requires the computation of the integral over the latent variables. The term $p(\{y\}, z, \theta\,; \{t\})$ can only be evaluated for some given random-effect values z, by solving sets of ordinary differential equations (see Algorithm 1). In this paper, we propose to address this computational challenge by combining the MCMC-SAEM algorithm with gradient descent (GD). The backbone of this algorithm is the SAEM algorithm [18], which is a stochastic approximation (SA) of the classical expectation-maximization (EM) algorithm [19]: are alternated a stochastic simulation step $z^{[k]} \sim p(z|\{y\}, \theta^{[k-1]}\,; \{t\})$ of the latent variables followed by a deterministic update of the model parameters $\theta^{[k]} \leftarrow \theta^\star(z^{[k]})$.

Algorithm 2: Calibration with MCMC-SAEM-GD.

input : Dataset y. Initial parameters $\theta^{[0]}$ and $z^{[0]}$.
Sequence of step-sizes $(\rho^{[k]})_k$. Sampling variances $(\sigma^{(b)})_b$.
output: Estimation of $\theta^{\text{m}} \approx \theta^{[k]}$.
Set $k = 0$ and $S_1^{[0]} = S_1(z^{[0]})$. `// initialization, eq. (14)`
repeat
  Set $k \leftarrow k + 1$.
  `/* block Gibbs symmetric random walk sampling */`
  foreach random variable $z^{(b)}$ **in** $(z^{(1)}, z^{(2)}, z^{(3)}) = (\alpha, \tau, s)$ **do**
    Draw a candidate $z^{(b)} \sim \mathcal{N}(z^{[k-1](b)}, \sigma_b^2)$.
    Let $z = (z^{[k](1)}, ..., z^{[k](b-1)}, z^{(b)}, z^{[k-1](b+1)}, ...)$.
    Compute the ratio $\omega = \log \frac{p(\{y\}, z, \theta^{[k-1]}; \{t\})}{p(\{y\}, z^{[k-1]}, \theta^{[k-1]}; \{t\})}$. `// alg. 1`
    Draw u according to the uniform distribution $u \sim \mathcal{U}(0,1)$.
    if $\log u < \omega$ **then** $z^{[k](b)} \leftarrow z^{(b)}$ **else** $z^{[k](b)} \leftarrow z^{[k-1](b)}$.
  end
  Adapt the proposal variances $(\sigma^{(b)})_b$. `// see [5]`
  `/* analytical update rule for` θ_1 `(classical SAEM) */`
  Set $S_1^{[k]} \leftarrow S_1^{[k-1]} + \rho^{[k]} \cdot \left[S_1(z^{[k]}) - S_1^{[k-1]}\right]$. `// eq. (14)`
  Set $\theta_1^{[k]} \leftarrow \theta_1^\star(S^{[k]})$ `// eqs. (15) - (18)`
  `/* gradient-descent-based update heuristic for` θ_2 `*/`
  Solve $\theta_2^\star = \underset{\theta_2}{\text{argmax}}\, p(\{y\}, z^{[k]}, \theta_1^{[k]}, \theta_2\,; \{t\})$ by GD. `// alg. 1`
  Set $\theta_2^{[k]} \leftarrow \theta_2^{[k-1]} + \rho^{[k]} \cdot \left[\theta_2^\star - \theta_2^{[k-1]}\right]$. `// heuristic`
until convergence

In [37], the authors introduce the MCMC-SAEM algorithm, where the simulation step is replaced by a Markov chain Monte-Carlo (MCMC) step while still preserving the theoretical convergence properties. In this paper, an analytical update rule $\theta^\star$ cannot be found for all the parameters θ: we use a gradient descent approach to overcome this difficulty, and we name MCMC-SAEM-GD the global resulting algorithm. Algorithm 2 gives a high-level pseudo-code of the proposed procedure. The sufficient statistics write:

$$S_t = \frac{1}{n}\sum_{i=1}^{n} \tau_i, \qquad S_\alpha = \frac{1}{n}\sum_{i=1}^{n} (\alpha_i - 1)^2, \tag{14}$$
$$S_\tau = \frac{1}{n}\sum_{i=1}^{n} \tau_i^2, \qquad S_\epsilon = \frac{1}{|\mathcal{E}| \cdot n \cdot \langle n_i \rangle_i} \sum_{i=1}^{n}\sum_{j=1}^{n_i} \epsilon_{i,j}^2,$$

where $\epsilon_{i,j}^2 = d_{\mathcal{E}}\{y_{i,j},\ \text{ExpP}_\gamma^{v_i}\left[\psi_i(t_{i,j})\right] \star y_0\}^2$ and $\langle n_i \rangle_i$ is the average number of longitudinal observations per subject. The update rules write:

$$t_0^\star = \left[\varsigma_t^2\, S_t + \frac{\sigma_\tau^{\star 2}}{n}\,\overline{t_0}\right] \cdot \left[\varsigma_t^2 + \frac{\sigma_\tau^{\star 2}}{n}\right]^{-1} \tag{15}$$

$$\sigma_\tau^\star = \left[S_\tau - 2\, t_0^\star\, S_t + t_0^{\star 2} + \frac{m_\tau}{n}\,\varsigma_\tau^2\right]^{\frac{1}{2}} \cdot \left[1 + \frac{m_\tau}{n}\right]^{-\frac{1}{2}} \tag{16}$$

$$\sigma_\alpha^\star = \left[S_\alpha + \frac{m_\alpha}{n}\,\varsigma_\alpha^2\right]^{\frac{1}{2}} \cdot \left[1 - \frac{f(-1/\sigma_\alpha^\star)/\sigma_\alpha^\star}{1 - F(-1/\sigma_\alpha^\star)} + \frac{m_\alpha}{n}\right]^{-\frac{1}{2}} \tag{17}$$

$$\sigma_\epsilon^\star = \left[S_\epsilon + \frac{m_\epsilon}{|\mathcal{E}|\ n\ \langle n_i \rangle_i}\,\varsigma_\epsilon^2\right]^{\frac{1}{2}} \cdot \left[1 + \frac{m_\epsilon}{|\mathcal{E}|\ n\ \langle n_i \rangle_i}\right]^{-\frac{1}{2}} \tag{18}$$

where $f(.)$ is the probability density function of the standard normal distribution. Both the coupled set of equations (15)-(16) and the implicit equation (17) can easily be solved by iterative update. Equation (18) is closed-form.

4.3.3 Implementation details

The sequence of $\rho^{[k]}$ required by Algorithm 2 is chosen to be constantly equal to 1 in a preliminary "burn-in" phase of the calibration procedure, and then decreases with the iterations with an exponential decay. The fanning numerical scheme is used to compute the parallel transport along geodesics in a scalable manner [41,42,62]. A block Metropolis-Hasting-within-Gibbs approach is used for the MCMC sampling step, where each variable α_i, τ_i and s_i are successively sampled. Several transition kernels can be chained in order to decrease the correlation between $z^{[k-1]}$ and $z^{[k]}$. Proposal variances are dynamically adapted during the iterations to ensure that the acceptation rates remain close to 30% [5]. The optimization problem for the update of θ_2 is solved by steepest gradient descent. The gradients of the complete log-likelihood are obtained by autodifferentiation using the PyTorch library.

The PyKeops library [13] implements smart autodifferentiation methods for convolution intensive computations to avoid memory overflows.

4.4 Personalization

Once the model is calibrated using a training data set, any individual data sequence $\{y_{i,j}, t_{i,j}\}_j$ can be reconstructed by fitting the model (whether it was part of the training set or not, i.e. $i \leq n$ or $i > n$ respectively). This procedure, called here personalization, consists in solving the optimization problem defined by equation (13). Note that all individuals can be treated independently i.e. equation (13) is equivalent to solving several sub-problems of the form $z_i^{\mathrm{m}} = \mathrm{argmax}_{z_i} \log p(\{y_{i,j}\}_j, z_i, \theta^{\mathrm{m}}; \{t_{i,j}\}_j)$. The computed optimal latent variables z_i^{m} give in turn the spatiotemporal coordinates of the individual trajectory in the reference frame of the calibrated model θ^{m}. We use the L-BFGS optimization method [38], where gradients are automatically computed using the PyTorch autodifferentiation library.

4.5 Simulation

The purpose of the simulation is to take advantage of the generative nature of the model to generate an entirely synthetic data set that reproduces the characteristics of the original training data set.

Given a longitudinal data set, the calibration followed by the personalization to the training data yields a normative model of progression, a spatiotemporal coordinate system (both being encoded by the parameters θ^{m}) and the coordinates of each individual in this reference frame (i.e. $z^{\mathrm{m}} = (z_i^{\mathrm{m}})_i$). We denote $\widetilde{p}(z^{\mathrm{m}}, \{t\})$ the empirical joint distribution of those individual parameters and of the corresponding time-indices $\{t\}$. We simulate synthetic data $\{y^{\mathrm{s}}\}$ by sampling random variables from this empirical distribution, and generate data following the generative model (11).

We use statistic functions ζ (most often not sufficient, similarly to [44]) to evaluate to which extend the simulated data resemble the training data:

$$\zeta(\{y^{\mathrm{s}}\}) \approx \zeta(\{y\}), \text{ with } \begin{cases} \{y^{\mathrm{s}}\} \overset{\text{iid}}{\sim} p(\{y\}|z^{\mathrm{s}}, \theta^{\mathrm{m}}; \{t^{\mathrm{s}}\}) \\ (z^{\mathrm{s}}, \{t^{\mathrm{s}}\}) \overset{\text{iid}}{\sim} \widetilde{p}(z^{\mathrm{m}}, \{t\}) \end{cases}. \tag{19}$$

For visualization purposes, we may choose to ignore the calibrated variance of noise $\sigma_\epsilon^{\mathrm{m}}$ and replace it with the degenerated value $\sigma_\epsilon = 0$. This choice will generate smoother shapes, and we will call such simulations "without noise".

5 Experiments

5.1 Validation on synthetic shape data

In this section, the calibration, personalization and simulation algorithms are validated on a synthetic data set in 2D. The simulation algorithm is first used in Section 5.1.1 to generate a synthetic shape data set from a chosen ground truth model. We use then the calibration method to infer the model parameters from the synthetic data set. The performance and the stability of the calibration algorithm is evaluated in various settings in Section 5.1.2. The calibrated model is personalized in Section 5.1.3, and the learned individual parameters are compared to the true values. Eventually, we re-simulate a synthetic data set from the calibrated model, and assess in Section 5.1.4 to which extend this new synthetic data set has similar statistics as the original data set.

5.1.1 Simulating synthetic data from a ground truth model

We choose values of fixed effects $\theta = (y_0, c_0, m_0, A_0, t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$, which specifies a normal distribution of shape trajectories. The chosen geometrical parameters y_0, c_0, m_0, A_0 are shown in Figure 3. In addition, we choose $t_0 = 0$, $\sigma_\tau = 2$, $\sigma_\alpha = 0.2$ and $\sigma_\epsilon \in \{0.00, 0.01, 0.02, 0.03, 0.05\}$.

We use the generative model to simulate a total of $n \in \{50, 100, 200\}$ individual trajectories and to sample them at several time-points $\{t_{i,j}\}_{j=1}^{n_i}$. We draw the number of observations for each individual $n_i \geq 2$ according to a shifted Poisson distribution with parameter $E(n_i) \in \{3, 5, 7, 9\}$. We

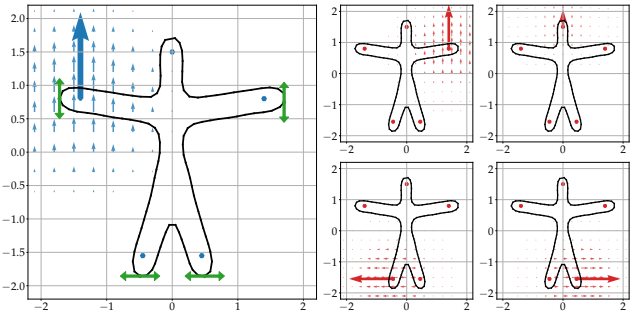


Fig. 3: Visualization of the parameters $\theta_1 = (y_0, c_0, m_0, A_0)$. The template shape y_0 is in solid black, the control points c_0 are the five dot points in either blue or red, the momenta m_0 is the bold blue arrow, and the four columns of A_0 are the bold red arrows. The velocity fields corresponding to the momenta m_0 or the geometrical components of A_0 are respectively represented with light blue or light red arrows. The green dots and arrows on the top figure mark the four landmark positions that will be considered for the statistic ζ, in order to validate the simulation algorithm in Section 5.1.4.

finally impose that the individual time-points $\{t_{i,j}\}_j$ are uniformly distributed in the observation interval $[t_{i,1}, t_{i,n_i}] = [t_{i,0} - \Delta t_i/2, t_{i,0} + \Delta t_i/2]$, where both the observation time window $\Delta t_i = t_{i,n_i} - t_{i,1}$ and the mid-point $t_{i,0}$ are drawn according to normal distributions: $\Delta t_i \overset{\text{iid}}{\sim} \mathcal{N}\big(E(n_i) - 2, \sigma_\tau^2\big)$ and $t_{i,0} \overset{\text{iid}}{\sim} \mathcal{N}(t_0, \sigma_\tau^2)$. Figure 4 displays some generated data in the reference case where $\sigma_\epsilon = 0.02$ and $E(n_i) = 7$.

5.1.2 Model calibration

The model calibration outputs are the estimated population parameters $\theta = (\theta_1, \theta_2)$ with $\theta_1 = (t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$ and $\theta_2 = (y_0, c_0, m_0, A_0)$. They are expected to be close to the MAP θ^{m} defined by equation (12).

Computing the MAP. Because only a finite number of data points are available and a Bayesian prior $p(\theta)$ is assumed on the parameters θ, the MAP θ^{m} does not correspond exactly to the ground truth parameters θ^{t}. The calibrated parameters θ are expected to converge towards the corresponding MAP parameters θ^{m} when the number of iterations goes to infinity – and this section experimentally verifies it, when the MAP parameters θ^{m} are known to converge towards the ground truth parameters θ^{t} when the number of observations goes to infinity. For each configuration of ground truth parameters θ^{t} and particular random sampling z^{t} of the generative model they define, θ_1^{m} can be analytically computed with equations (15-18) and θ_2^{m} approximated by a steepest gradient descent approach initialized with θ^{t} and z^{t} (see Algorithm 2). The calibration error between θ and θ^{m} is analyzed and discussed in details in the rest of this section, when the statistical error between θ^{m} and θ^{t} is only computed in the reference configuration. Using the performance metrics introduced in the following paragraph, the second line of Table 1 (in italicized text) gives the corresponding quantitative normalized distances, which remain below 6% in all cases.

Normalized error metrics. The error for the scalar parameters $\theta_1 = (t_0, \sigma_\tau, \sigma_\alpha, \sigma_\epsilon)$ is measured by the absolute difference between estimated and MAP values. The error for t_0 is normalized by the characteristic population observation window, that we define as $2 \cdot (1 + \sigma_\alpha) \cdot [E(\Delta t_i)/2 + \sigma_\tau]$. The remaining errors for $\sigma_\tau, \sigma_\alpha, \sigma_\epsilon$ are respectively normalized by the true standard deviations $\sigma_\tau = 2$, $\sigma_\alpha = 0.2$, and the estimated noise level by the Bayesian atlas model [28] computed during the initialization pipeline described in Section 4.3.1. The error on the template shape y_0 is assessed as the maximum point-to-point residual distance, and normalized by the conservative value of 3 spatial units for the characteristic size of the considered shape (see Figure 3). The control points c_0 and the momenta m_0 are jointly evaluated through the ℓ^2 distance of the estimated velocity field $v_0 = \text{Conv}(c_0, m_0)$ to the MAP value, normalized by the ℓ^2 norm of this MAP velocity field. Finally, the convergence

	y_0 (%)	c_0, m_0 (%)	c_0, A_0 (%)	t_0 (%)	σ_τ (%)	σ_α (%)	σ_ϵ (%)
reference	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
statistical error	*0.0*	*0.4*	*0.1*	*2.5*	*5.6*	*3.0*	*0.0*
$\sigma_\epsilon = 0.00$	2.7 ± 0.01	5.9 ± 0.17	2.5 ± 0.05	7.4 ± 0.24	2.6 ± 0.60	8.1 ± 5.24	36.4 ± 0.03
$\sigma_\epsilon = 0.01$	6.0 ± 0.01	2.5 ± 0.09	1.7 ± 0.01	2.7 ± 0.07	0.7 ± 0.21	0.8 ± 0.44	15.1 ± 0.03
$\sigma_\epsilon = 0.02$	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
$\sigma_\epsilon = 0.03$	1.4 ± 0.01	1.9 ± 0.09	1.5 ± 0.03	4.8 ± 0.08	2.3 ± 0.38	3.8 ± 1.23	5.1 ± 0.01
$\sigma_\epsilon = 0.05$	1.8 ± 0.02	4.1 ± 0.21	1.9 ± 0.04	1.5 ± 0.10	1.1 ± 0.34	0.9 ± 0.33	1.7 ± 0.01
$E(n_i) = 3$	1.6 ± 0.02	5.7 ± 0.29	2.1 ± 0.05	5.2 ± 0.22	3.9 ± 1.12	7.1 ± 2.03	6.5 ± 0.03
$E(n_i) = 5$	4.4 ± 0.01	5.9 ± 0.32	2.7 ± 0.03	1.1 ± 0.15	5.8 ± 0.56	0.7 ± 0.36	6.7 ± 0.02
$E(n_i) = 7$	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
$E(n_i) = 9$	2.4 ± 0.01	3.2 ± 0.05	1.7 ± 0.04	7.2 ± 0.03	1.8 ± 0.11	2.7 ± 0.17	6.3 ± 0.01
$q = 2$	2.5 ± 0.04	18.0 ± 0.28	51.3 ± 0.06	10.6 ± 0.16	9.3 ± 0.52	3.1 ± 0.84	172.4 ± 0.07
$q = 4$	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
$q = 6$	2.5 ± 0.01	6.8 ± 0.10	1.9 ± 0.03	8.9 ± 0.07	1.7 ± 0.20	9.3 ± 1.80	7.0 ± 0.02
$n = 50$	3.4 ± 0.01	3.6 ± 0.16	2.2 ± 0.03	4.7 ± 0.04	0.2 ± 0.16	2.0 ± 0.17	7.6 ± 0.02
$n = 100$	2.5 ± 0.01	6.2 ± 0.10	2.1 ± 0.02	8.8 ± 0.06	1.7 ± 0.28	7.0 ± 2.73	7.7 ± 0.01
$n = 200$	1.8 ± 0.00	3.9 ± 0.08	1.5 ± 0.02	2.6 ± 0.04	0.8 ± 0.12	3.7 ± 0.19	6.9 ± 0.01

Table 1: Final average normalized performance metrics and associated standard deviations, obtained after 10 independent runs of the MCMC-SAEM algorithm in varied configurations. The algorithm is run for 200 iterations in all configurations. The reference configuration corresponds to a noise level $\sigma_\epsilon = 0.02$, an average number of visits per subject $E(n_i) = 7$, $q = 4$ allowed components of geometrical variability, and $n = 100$ input subjects. The second line gives the discrepancy between the ground truth (used for generating the data) and the MAP (used for evaluating the calibration performance), in the reference case. We call this discrepancy the statistical error, by opposition to the calibration error.

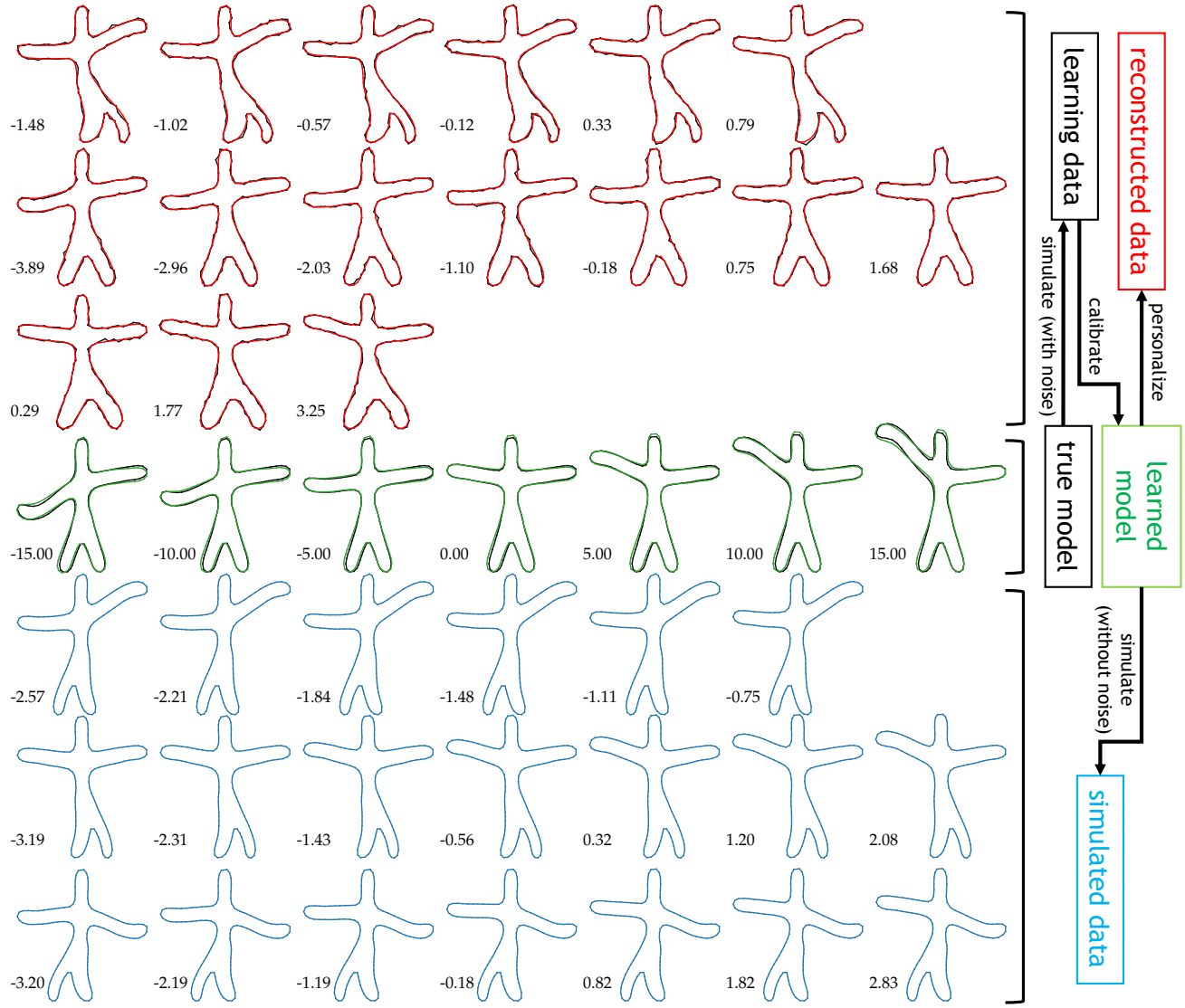


Fig. 4: Illustration of the evaluation procedure for model calibration, and subsequent personalization or simulation from the learned model. The ground truth population geodesic is plotted in black on the central line. From this model are simulated $n = 100$ individual spatiotemporal trajectories: three randomly-picked samples are plotted in black on the top lines. The population geodesic of the calibrated model is plotted in green on the central line, superimposed with the ground truth geodesic. This calibrated model can then be personalized to the training observations as plotted in red, or leveraged to simulate new spatiotemporal trajectories that resemble the original data set as plotted in blue.

of modulation matrix A_0 is assessed by measuring the mismatch between the sets of space-shifts that can be generated from the pair of parameters (c_0, A_0): (i) the estimated modulation matrix is first re-projected on the MAP control points, (ii) the linear subspaces generated by the columns of the MAP and estimated modulation matrices are defined, (iii) the matrix representations of the projectors over those subspaces are computed, (iv) the average of the four greatest eigenvalues of their difference captures the mismatch between those projectors, (v) the result is normalized by the largest eigenvalue of the MAP projector.

Evaluation setups and results. In addition to the previously introduced setups, configurations with varying allowed number of sources $q \in \{2, 4, 6\}$ are also evaluated. We call the configuration with $\sigma_\epsilon = 0.02$, $q = 4$, $n = 100$, and $E(n_i) = 7$ the reference one. Augmented with the 11 configurations differing from this reference by a single parameter, a total of 12 calibration problems are defined. Each is solved 10 times by running the stochastic MCMC-SAEM-GD algorithm.

Figure 5 plots the evolution of the error metrics across the allowed 200 iterations for the reference configuration: the black lines correspond to the 10 different runs, and in green is represented their mean and standard deviation. The algorithm is stable i.e. converges to similar results at each run: the final standard deviation of the error is smaller than 10% of the maximal (initial) error, for all parameters. The two regimes of the algorithm can be identified: the burn-in phase for the first half of the iterations where the step-sizes $\rho^{[k]}$ remained fixed to 1, followed by the concentration phase where the step-sizes decrease geometrically. We can finally notice that σ_α is estimated with more variance than other parameters, and than σ_τ in particular. This suggests that adding higher-order components to the time-warp functions ψ_i would come with estimation challenges.
Table 1 gives the average final error metrics and the associated standard deviations. Those standard deviations remain in all but one case below 3%, underlining the stability of the estimation algorithm. In most cases, the parameters are estimated with less than 10% of error: exceptions only appear in the configurations with very low noise levels $\sigma_\epsilon \leq 0.01$ or underestimated number of geometrical sources $q = 2$. Interestingly, a higher noise level σ_ϵ does not necessarily correlate with a degraded estimation of the parameters. The presence of noise can actually help the algorithm to better explore the space of parameters: it seemed that local minima may be harder to escape for very low levels of noise. In particular, the estimation performance of the noise variance σ_ϵ^2 improves when the true value increases. Table 1 studies also the impact of the length of the observation period $E(n_i)$. Long periods generally favour more accurate estimation of the parameters $v_0 = \text{Conv}(c_0, m_0)$ which encodes the direction of the progression, and $(\sigma_\tau, \sigma_\alpha)$ which capture its dynamical variability. However, because of compensation mechanisms that may take place at the individual level between α_i and τ_i, it is rather the joint quantity $\sigma_\tau + \sigma_\alpha$ that is clearly better estimated when $E(n_i)$ increases than σ_τ and σ_α independently.

The same table compares also the estimation quality when the true number of sources is underestimated ($q = 2$), perfectly chosen ($q = 4$) or overestimated ($q = 6$). The reconstruction ability of our model, measured by σ_ϵ, increases with q, and seems to saturate once the optimal number of sources has been reached. The large estimation error made on σ_ϵ in the case $q = 2$ comes from the fact that data was simulated from exactly four geometrical components of comparable importance (see Figure 3), thus creating a strong reconstruction performance thresholding effect when choosing $q < 4$. One can expect smoother variations of the reconstructive performance on real data sets, which do not result from the exact simulation of the generative model. Parameters are in majority less well estimated in the $q = 2$ configuration, and at comparable distance to the MAP in the two remaining ones. Finally, Table 1 shows that the number of

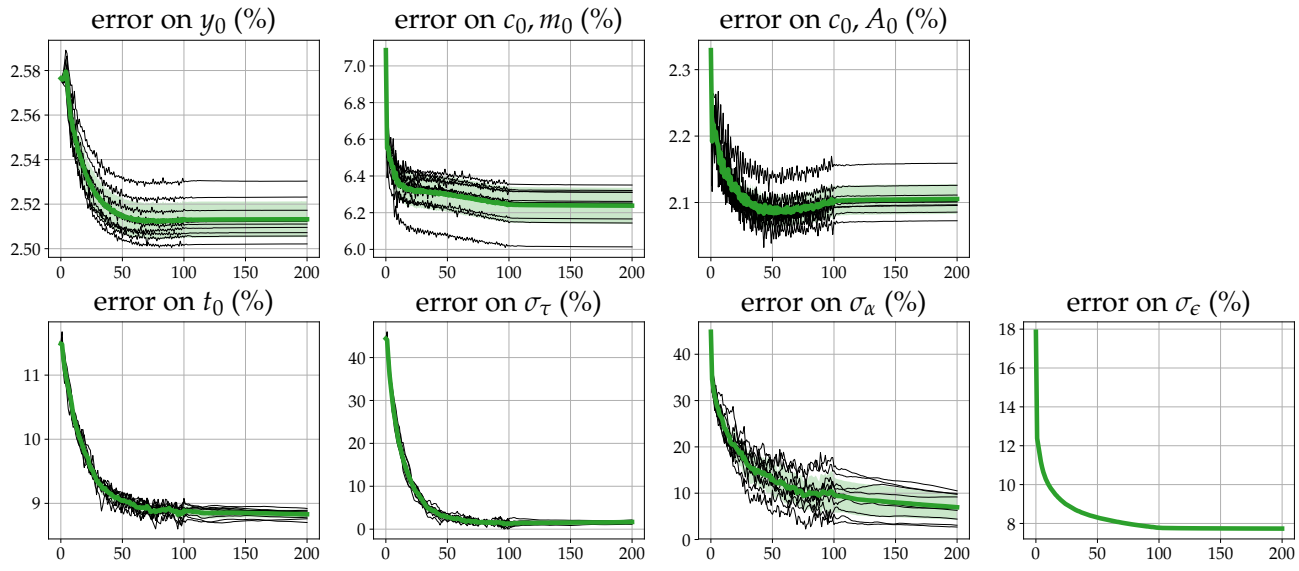


Fig. 5: Evolution of the error metrics across the 200 allowed iterations of the MCMC-SAEM algorithm, for the reference configuration: noise standard deviation $\sigma_\epsilon = 0.02$, $q = 4$ estimated components of geometrical variability, learning on a data set composed of $n = 100$ with on average $E(n_i) = 7$ visits per subject, spanning 5 time units. The 10 solid black curves correspond to 10 independent runs of the same – stochastic – MCMC-SAEM algorithm; the bold green curve is their average and the light green region indicates the associated standard deviation. The algorithm consistently converges towards similar parameters at each run, and those estimated parameters are satisfyingly close to the MAP estimate.

training subjects n has a major influence over the quality of the estimation. Almost all metrics are improved in the configuration with $n = 200$ subjects.

In conclusion, the proposed MCMC-SAEM-GD algorithm successfully solves our model calibration problem in varied configurations. The stochastic procedure is stable across independent repetitions. The presence of noise in the training data is well-handled, and actually seems to act as a good regularizer for the estimation procedure. An underestimated number of sources does not harm the convergence of the procedure, but mostly impairs the reconstruction ability of the learned model. This number should therefore be gradually increased to meet the reconstruction goals of the experimenter, keeping in mind that an intrinsic optimal performance will be reached when q is large enough. Finally, increasing the number of subjects or the number of visits are both beneficial for model calibration.

5.1.3 Personalization after calibration

Once calibrated, the longitudinal shape models are personalized to the training data. The estimated individual parameters α_i, τ_i, s_i are compared to their true value. In order to be comparable with the true sources, the estimated sources are first brought back to the cotangent space defined by the true control points c_0^{t} by solving $\mathrm{Conv}(c_0^{\mathrm{t}}, m_i^{\mathrm{t}}) = \mathrm{Conv}(c_0, m_i)$.

Figure 6 plots the estimated z_i against the true corresponding values. The acceleration factors are well aligned on the bisector. The onset ages and sources are also estimated with a low variance, but with a non-negligible bias.

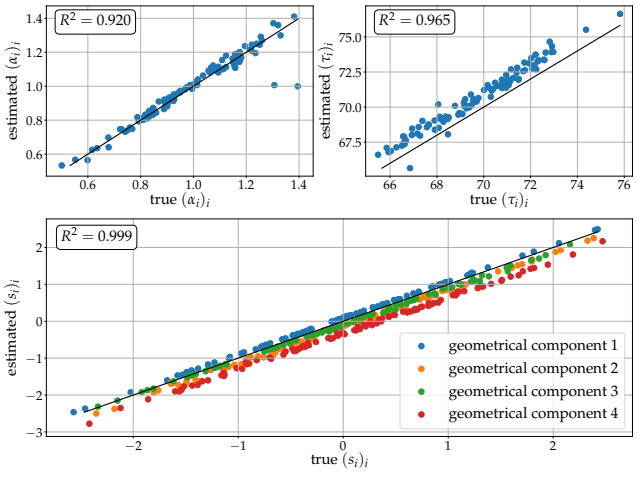


Fig. 6: Comparison of the estimated individual parameters $z_i = (\alpha_i, \tau_i, s_i)$ after personalization of the mean calibrated model to the simulated observations, in the reference scenario. In each scatter plot, the identity is represented by the solid black line. The R^2 value for the sources is an average over the four geometrical components.

This effect is due to the fact that the estimation of individual parameters during personalization may compensate for some error made during the estimation of population parameters during calibration. Time-shifts τ_i may compensate an error on the reference time t_0. Acceleration factors α_i may compensate for an error in the norm of v_0. Sources s_i may compensate for an error in the norm of the columns of A_0.

These effects do not question the identifiability of the model, but rather suggest that, for a finite number of observations, the likelihood may have a rather flat maximum, for which a range of parameter values may reconstruct data almost equally well. Finally, two outliers can be noticed in Figure 6 for the pace of progression α_i, as well as for the onset age τ_i. These outliers correspond to extremely reduced windows of observation $t_{i,n_i} - t_{i,1}$, respectively equal to 0.07 and 0.20, when the theoretical mean is equal to 5.

Table 2 summarizes the results in all configurations, giving for each of the twelve considered setups the median error and associated median absolute deviation on $z_i = (\alpha_i, \tau_i, s_i)$ when personalizing the average calibrated model. The median is reported instead of the mean because it is more robust to outliers. Focusing on the estimation variability, it appears that the sources s_i are the best estimated parameters, followed by the pace of progression α_i and the onset ages τ_i.

	$\Delta\alpha_i$ (%)	$\Delta\tau_i$ (%)	Δs_i (%)
reference	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$\sigma_\epsilon = 0.00$	-3.8 ± 4.3	42.4 ± 24.8	-5.4 ± 6.8
$\sigma_\epsilon = 0.01$	-2.1 ± 6.4	11.8 ± 9.4	-11.1 ± 12.0
$\sigma_\epsilon = 0.02$	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$\sigma_\epsilon = 0.03$	-7.6 ± 15.7	25.5 ± 15.0	-6.7 ± 2.5
$\sigma_\epsilon = 0.05$	-21.8 ± 23.4	4.4 ± 14.7	-9.1 ± 3.0
$E(n_i) = 3$	9.3 ± 32.7	-12.1 ± 19.3	-8.8 ± 3.2
$E(n_i) = 5$	2.9 ± 13.0	3.2 ± 20.8	-3.6 ± 7.5
$E(n_i) = 7$	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$E(n_i) = 9$	-8.2 ± 7.3	46.0 ± 16.1	-7.6 ± 2.3
$q = 2$	-1.6 ± 18.0	48.3 ± 50.7	-2.4 ± 69.4
$q = 4$	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$q = 6$	9.8 ± 12.3	45.3 ± 15.8	-11.7 ± 3.3
$n = 50$	-12.9 ± 8.1	24.8 ± 18.3	-6.9 ± 6.1
$n = 100$	4.4 ± 11.9	44.2 ± 15.3	-11.7 ± 3.3
$n = 200$	9.8 ± 10.5	13.2 ± 10.9	-13.1 ± 2.4

Table 2: Median of the residual errors and associated median absolute deviation times 1.4826 for the estimated individual parameters, expressed in percentage of the corresponding ground-truth standard deviations $\sigma_\alpha = 0.2$, $\sigma_\tau = 2$ and $\sigma_s = 1$. The results are given for the reference scenario plus eleven perturbed scenarii, where either the noise level σ_ϵ, the average number of visits per subject $E(n_i)$, the allowed number of geometrical components q or the number of subjects n is varied.

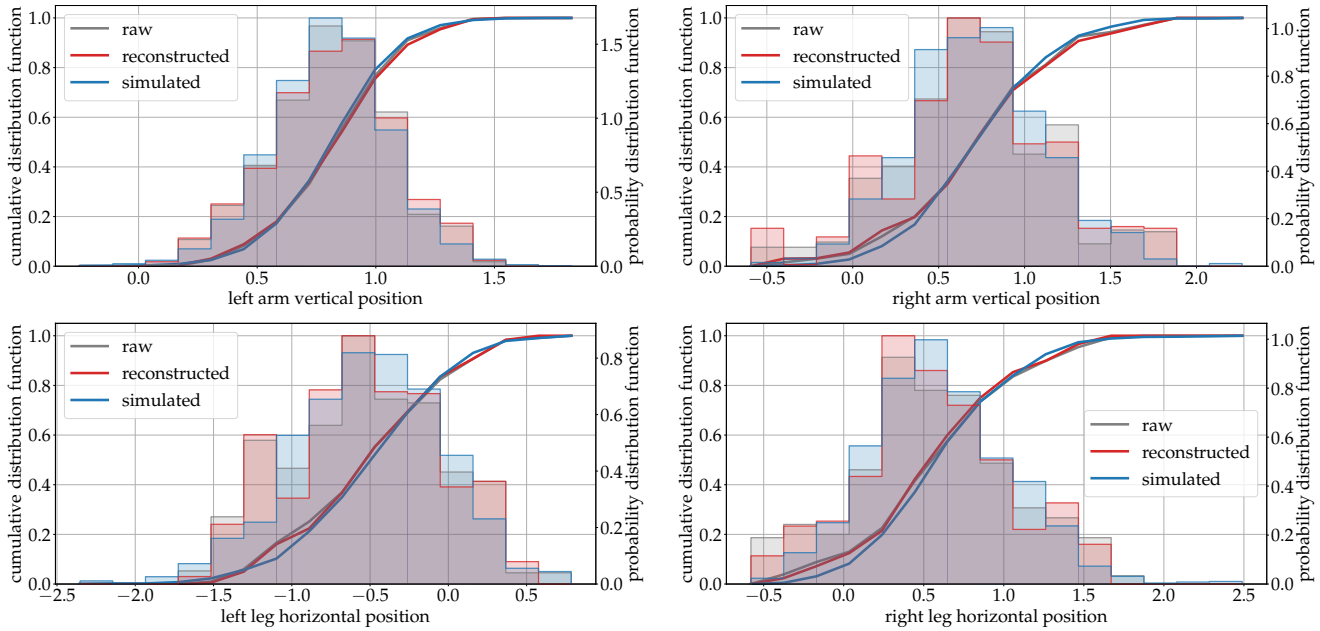


Fig. 7: Distribution of the position of landmarks of interest in the raw (i.e. original), reconstructed (by personalization of the calibrated model) and simulated data sets, for the reference configuration. Those landmarks of interest are indicated by green dots on Figure 3. The simulated distributions are similar to the corresponding raw ones, suggesting that the spatiotemporal variability of the original data set has been successfully captured.

The estimation of the pace of progression α_i quickly deteriorates with increasing levels of noise σ_ϵ, reaching almost 25% of the true standard deviation $\sigma_\alpha = 0.2$ in the most noisy configuration. The estimation of the onset ages τ_i and sources seems more robust, with no clear tendency. The estimation of the pace α_i improves when the number of visits per subject $E(n_i)$ increases. The same trend can be noticed for the onset age τ_i, although with a reduced amplitude. The sources s_i remain well-estimated in all scenarii. No clear difference can be noticed between the reference $q = 4$ and the over-estimated number of geometrical components case $q = 6$, suggesting that adding components does not hamper the personalization of a calibrated model. However, underestimating this number of components with $q = 2$ deteriorates the estimation of the sources s_i, and the dynamical parameters α_i and τ_i to a lesser extend. As in the previous section, we interpret this large performance drop due to the fact that data was simulated according to exactly four geometrical sources of similar magnitude (see Figure 3): in real data sets, one may expect the estimation performance to change more smoothly with q. Finally, an increased number of subjects n allows a better performance of the personalization algorithm, especially for the onset age τ_i and source s_i parameters.

5.1.4 Simulation after calibration and personalization

After calibration and personalization, the learned model and empirical distribution of the random effects can be used to simulate entirely synthetic shape trajectories. Figure 4 gives some randomly selected samples from such simulated trajectories for the reference scenario, where (see equation (19)):

- the fixed effects θ^{m} are averages over the 10 calibrations;
- the random effects z^{s} are drawn according to independent normal distributions with mean and standard deviations equal to the values given by Table 2;
- the visit ages t^{s} are drawn according to the true procedure based on the average calibrated values for t_0, σ_τ, and the empirical average $\langle n_i \rangle_i$ for $E(n_i)$ (see Section 5.1.1).

Figure 7 compares the distribution of vertical or horizontal positions of the tips of the original (see Section 5.1.1), reconstructed (see Section 5.1.3) and simulated observations. Those landmarks of interest are indicated by green dots and arrows on Figure 3, and form the statistic ζ introduced in equation (19). A total of 1,000 subjects are simulated, when only 100 were available for model calibration. The three distributions largely overlap, indicating that the learned distribution of shape trajectories reproduces the true distribution.

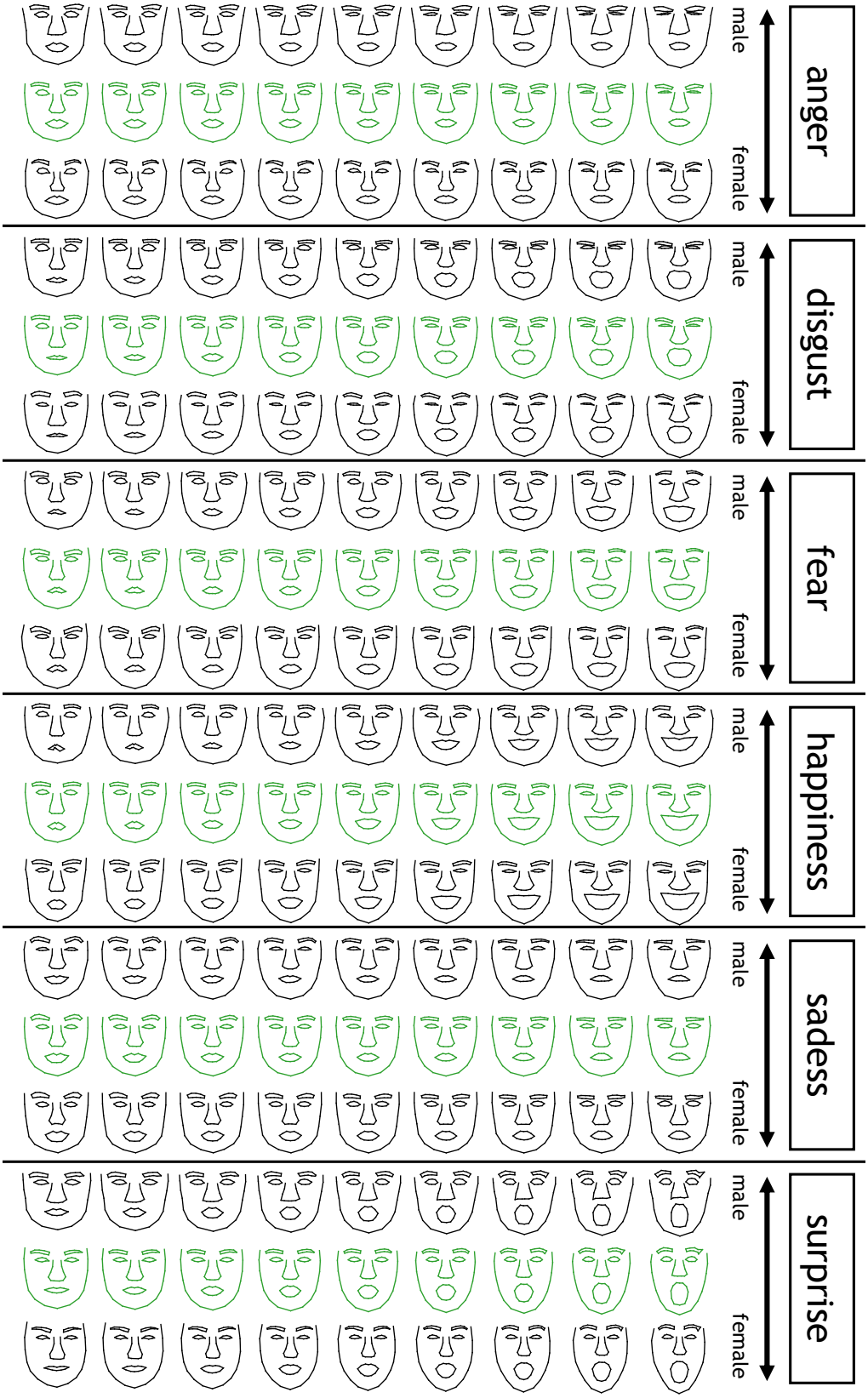


Fig. 8: Learned emotion spatiotemporal models. The population geodesic is plotted in green, and the shifted progressions along the gender mode of geometrical variability are plotted in black.

5.2 Dynamic facial expression

5.2.1 Data and preprocessing

The Birmingham University 3D dynamic facial expression database [61] gathers short video sequences from 101 subjects (of which 58 female, 43 male). Each subject mimics in 6 distinct sequences basic emotions which are Anger, Disgust, Fear, Happiness, Sadness and Surprise. For each of those 606 sequences we uniformly extract 8 frames spanning from the first to the 36-th one, which correspond to a subsampling of the first 1.4 seconds of each video. We do not work directly with the images, but with a set of 75 semi-automatically extracted landmarks, which come with this data set. Every set of 3D landmarks is registered to a reference one by similarity-based Procrustes alignment.

5.2.2 Model calibration: learned emotion models

We learn 6 distinct longitudinal atlas models: one per emotion, calibrated on the $n = 101$ sequences of $n_i = 8$ frames for all subjects i. We choose $q = 10$ sources. Figure 8 shows in green the estimated average scenario for each emotion. Qualitatively, those average scenarii show a typical pattern of facial expression. The Disgust, Fear, Happiness and Surprise models feature large displacements in the area of the mouth in particular. The Sadness expression is more mute, with a subtle displacement of the eyebrows. The Anger model shows a combined displacement of both eyes and eyebrows.

5.2.3 Gender-specific emotion patterns

The estimated models are personalized to the corresponding training data sets, giving for each sequence an optimal $z_i = (\alpha_i, \tau_i, s_i)$. We only focus on exploiting the individual source parameters $s_i \in \mathbb{R}^q = \mathbb{R}^{10}$ in this section. For each model, we fit a 1D partial least square regression model for predicting the gender from a linear combination of the sources variables s_i [1]. We then test whether the linear combination of the sources are significantly different between men and women using a Student t-test. All p-values are smaller than 10^{-5}, thus showing significant differences in the geometry of the face between genders that are independent of the pattern of expression.

Figure 8 shows the typical scenario for men and women, which are built by translating the mean scenario in the direction of the average of the sources for each gender (in black). For all emotion models, male subjects tend to have wider faces than females, as it can very clearly be seen in the area of the cheeks or of the nose for the Anger and Surprise models.

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	64.3	7.0	8.1	4.0	16.6	-
Disgust	13.7	55.1	12.4	14.8	1.9	2.0
Fear	1.0	16.6	58.6	13.9	7.0	3.0
Happy	1.9	6.0	13.0	79.1	-	-
Sad	16.5	2.0	14.2	1.1	66.2	-
Surprise	1.0	3.0	16.0	-	1.0	79.1

Table 3: Average confusion matrix across 5-fold linear discriminant classification. The sequence features consist in a 12-scalar vector that stacks the 6 pairs of dynamical parameters α_i, τ_i obtained by personalizing the 6 emotion models. The average accuracy is $67.08\,\%$.

5.2.4 Application to classification

We propose to automatically recognize the emotion from a sequence based on the personalization of each facial expression model to the sequence. We propose here to use the dynamic variables α_i, τ_i for classification.

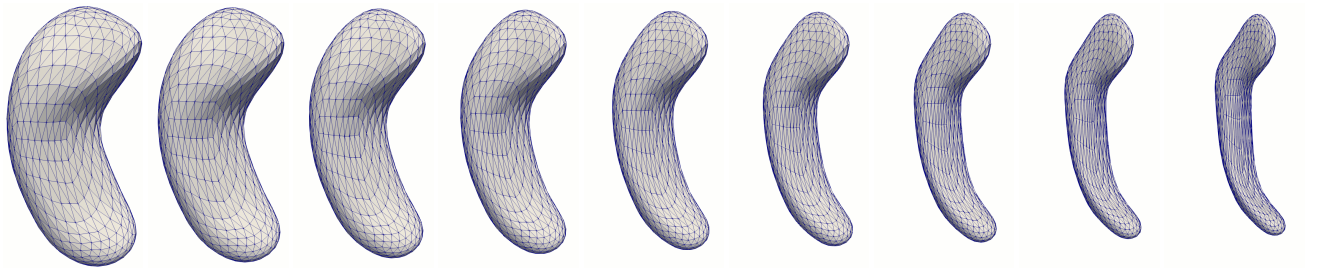
More precisely, we perform a 5-fold cross-validation ensuring that each group is gender-balanced. For each split:

- six longitudinal shape models are learned on the training sequences for each emotion;
- these models are personalized to all the 606 sequences: for each sequence a total of 6 z_i vectors are therefore estimated;
- for each sequence, the estimated temporal parameters α_i, τ_i are stacked into vectors of $6 \times 2 = 12$ scalars;
- these feature vectors are used to train and test a simple linear discriminant classifier on the corresponding train and test sequences.

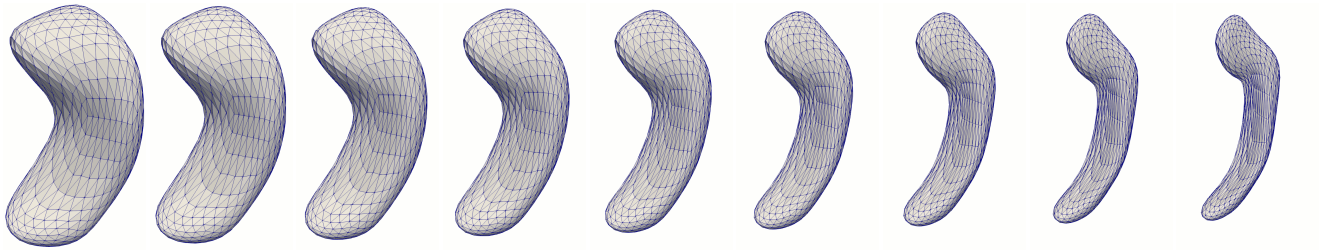
Table 3 gives the confusion matrix obtained with this procedure, averaged over the 5 folds. The average classification accuracy is $67.08\,\%$, above the chance level which amounts to $16.67\,\%$. For comparison, [4] reported an average accuracy of almost $100\,\%$, [58] of $90.44\,\%$, and [23] of $74.63\,\%$. We emphasize however that our performance is achieved:

- using the default linear discriminant analysis from the `sklearn` library, without any hyperparameter tuning as in [4] with random forest, in [58] with hidden markov model or in [23] with radial support vector machine;
- on all the 606 available sequences, without any manual selection of a subset of 60/101 subjects as it is done in [4, 58] or of 507/606 sequences as done in [23];
- based only on 12 intuitive scalar features per sequence, that encode how an individual emotional pattern dynamically compares to population models of basic emotions.

From this experiment that has not been particularly tuned to achieve best classification performance, we conclude that our model captured shape characteristics that are specific to



(a) Left hippocampus mean progression.



(b) Right hippocampus mean progression.

Fig. 9: Typical model of hippocampus atrophy from MCI to Alzheimer's disease stage. Physiological ages (from left to right, in years): 58.6, 63.0, 67.4, 71.8, 76.2, 80.7, 85.1, 89.5, 93.9.

each emotion. It is worth noting that we used here only dynamic parameters that capture how fast or slow the face is changing in the sequence, or with which delay.

5.3 Hippocampal atrophy in Alzheimer's disease

5.3.1 Data and preprocessing

Data used in the preparation of this section were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

We select all the T1-weighted MRIs of subjects that were diagnosed as presenting mild cognitive impairements at some visit, and diagnosed as converted to Alzheimer's disease at some later visit. See Table 4 for summary statistics. This data set amounts to a total of 1993 visits from $n = 322$ subjects. Second-take "re-test" MR images are available for 1838 of those visits and will be used to estimate the noise in the data. All those $1993 + 1838 = 3831$ images are pre-processed exactly in the same manner, starting with the longitudinal pipeline of FreeSurfer[1] (version 5.3.0) [24,25]. The skull-stripped brains are then aligned with an affine 12-degrees-of-freedom transformation onto the Colin27 average brain[2] with FSL 5.0[3] [60]. Meshes of the left and right hippocampus are obtained from the original images as follow:

- the volumetric segmentations of the hippocampus computed with FreeSurfer are transformed into meshes using the aseg2srf script of July 2009[4],
- the resulting meshes are decimated by a 88% factor using Paraview 5.4.1[5] [2],
- they are aligned using the previously-computed global affine transformation estimated with the FSL software,
- residual pose differences among subjects are removed by rigidly aligning the meshes from the baseline image of each subject to the corresponding hippocampus mesh in the Colin27 atlas image, this transformation with 6 degrees of freedom being computed with the GMMReg script of June 2008[6] [32],
- the same transformation is finally used to align the meshes from the follow-up images of the same subject.

Number of subjects	322
Number of visits	1993
Average number of visits per subject ($\pm$ std)	5.8 ($\pm$ 2.4)
Average age ($\pm$ std)	74.0 ($\pm$ 6.7)
Sex ratio (F/M in %)	41.2 / 58.8
Amyloid status (+/-/unknown in %)	73.2 / 7.1 / 19.7
APOE carriership (%)	65.2
Education (mean $\pm$ std, in years)	15.9 ($\pm$ 2.8)
Marital status (married/not married in %)	80.9 / 19.1

Table 4: Summary statistics of the medical data set of Alzheimer's disease patients.

[1] available at: https://surfer.nmr.mgh.harvard.edu
[2] available at: http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27
[3] available at: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
[4] available at: https://brainder.org
[5] available at: www.paraview.org
[6] available at: https://github.com/bing-jian/gmmreg

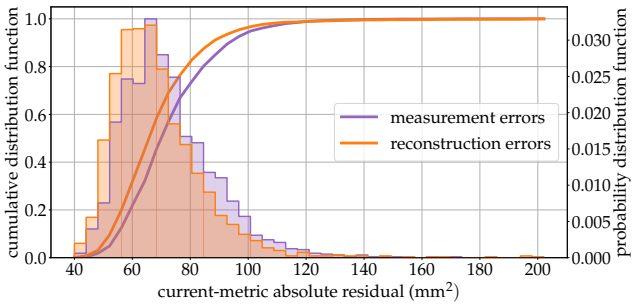
5.3.2 Models of atrophy of the hippocampus

We calibrate two longitudinal shape models on all the 1993 meshes of the left and right hippocampus respectively, choosing in both cases $q = 8$ sources. The deformation kernel width is set to $\sigma = 10\,mm$. The current distance is used to compute distances between meshes without point correspondence, with a kernel width of $\sigma_{\mathcal{E}} = 5\,mm$ [14,59].

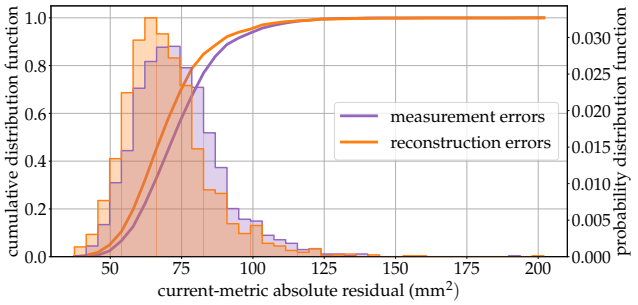
Figure 9 shows the estimated average progression, which consists in an overall atrophy of both the left and right hippocampus with a specific deformation of their shape. It is worth noting that we reconstruct here the progressive atrophy of the hippocampus over more than 30 years of disease progression although patients have never been observed for more than few years. This can be achieved because the method automatically re-aligns in time the data of patients that are at different, but unknown, disease stage.

5.3.3 Personalization to unseen data

We assess the reconstruction performance of the calibrated models using a 5-fold cross-validation. The $n = 322$ subjects are split into 5 groups; 2×5 distinct shape models are calibrated on the training sets for the left and right hippocampus. Those models are then personalized to the unseen test subjects. To assess the goodness of fit, we measure the residual errors and compared the distribution of such errors with the noise distribution. This noise distribution is determined by measuring the distance between the two meshes extracted from the "test" and "re-test" images acquired from the same patient the same day, thus capturing all the variability due to varying image quality and its consequence in the processing. Figure 10 shows the superimposition of the distribution of the residual errors with the distribution of the differences between the meshes of the test and re-test images. The reconstruction errors are on average smaller than the intrinsic uncertainty on the data, and with a lower variance as well. The model allows therefore to reconstruct individual data at the precision of the noise. It is worth noting that this could be achieved using a reduced set of 2×10 scalars, which are for each hippocampus the pace of progression α_i, the onset age τ_i, and the eight sources s_i.



(a) Left hippocampus. The mean error is $68.5 \pm 15.9\,mm^2$ for the shape model, and $83.2 \pm 36.0\,mm^2$ for the re-test measurement.



(b) Right hippocampus. The mean error is $69.8 \pm 15.0\,mm^2$ for the shape model, and $85.2 \pm 40.1\,mm^2$ for the re-test measurement.

Fig. 10: Comparison of the generalization error to unseen data of the learned shape models and the intrinsic measurement error. The discrepancies between meshes are computed with the current metric with $\sigma_{\mathcal{E}} = 5\,mm$, without assuming any point-to-point correspondence.

5.3.4 Association with co-factors

We calibrate and personalize the models on whole data set, and aim to study how some genetic, biological and environmental co-factors may modulate the progression of Alzheimer's

				left hippocampus		right hippocampus	
genetic	gender (female vs. male)	α_i		$\times 1.23$	[**]	$\times 1.21$	[**]
		τ_i		-12.4 months	[**]	-8.7 months	[*]
		s_i		± 0.54	[***]	± 0.57	[****]
	APOE-ϵ4 (carrier vs. non-carrier)	α_i		$\times 1.22$	[*]		
		τ_i		-35.8 months	[***]	-32.5 months	[**]
		s_i					
biological	amyloid (positive vs. negative)	α_i		$\times 1.52$	[**]	$\times 1.67$	[*]
		τ_i					
		s_i					
environmental	marital (married vs. non-married)	α_i		$\times 1.14$	[*]		
		τ_i		-42.5 months	[***]	-36.3 months	[**]
		s_i					
	education (nb. of years of education)	α_i					
		τ_i		-3.7 months/y	[**]	-5.1 months/y	[***]
		s_i					

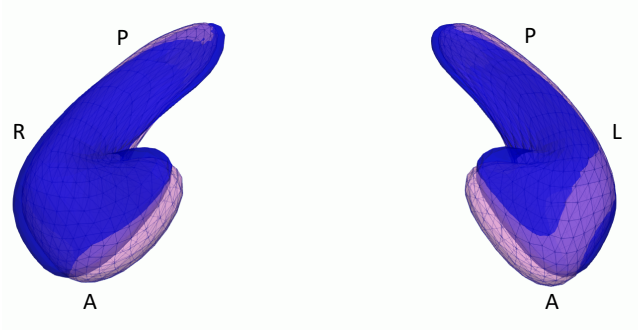
Table 5: Significant associations of individual parameters with genetic, biological and environmental factors: effect sizes and significance levels of the adjusted p-values (thresholds 5%, 1%, 0.1%, 0.01%). Time-shifts τ_i are in months, others have no units. Directions of space-shifts are not signed. The 23 subjects (out of $n = 322$) without amyloid information have been discarded.

disease in patients. We therefore aim to find correlations between individual variables $z_i = (\alpha_i, \tau_i, s_i)$ and the following factors: gender, APOE-ϵ4 carriership, presence of amyloid plaques, education level and marital status.

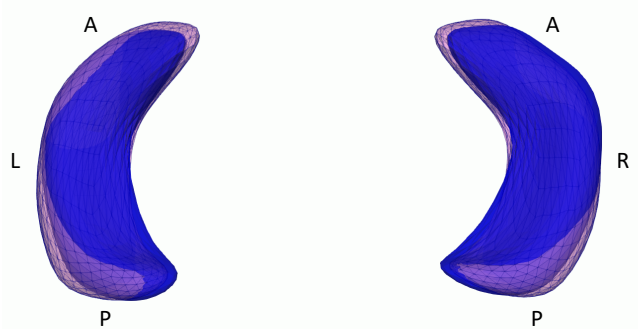
To this end, the parameters α_i and τ_i are regressed against the five considered cofactors, and two-tailed t-tests are performed on the coefficients. A 2-block partial least square regression model [1] is used to regress the eight sources s_i against the five cofactors in a one-dimensional projection space. A two-tailed t-test is then performed on the weights of the multivariate regression of the linear combination of sources against the cofactors. For each case, the obtained five p-values are corrected with the Benjamini-Hochberg false discovery rate procedure [8].

The obtained correlations for both left and right hippocampus are summarized in Table 5. The two first rows indicate that the atrophy of the hippocampus develops faster and starts earlier in female subjects. Male and female subjects present significantly different shape of their hippocampus regardless of its atrophy due to aging or disease progression. Figure 11 presents the corresponding mode of geometrical variability. Hippocampal atrophy also starts earlier in carriers of at least one ϵ4 allele of the APOE gene, with an effect size of almost three years. The atrophy occurs at an accelerated pace in amyloid-positive subject, as well as for APOE-ϵ4 allele carriers and married subjects but only in a significant manner in the left hemisphere of the brain. Finally, the atrophy occurs earlier in married subjects, as well as in educated subjects.

The results obtained by correlating the estimated individual parameters z_i with the genetic and biological factors are in line with current knowledge. The results obtained with respect to the marital status are more surprising, and should probably be taken with care as the non-married group, which represents less than 20% of the considered 299 subjects (see Table 5) is very heterogeneous. It gathers widowed, divorced, or never married subjects. Finally, we show that the atrophy starts earlier also in subjects with higher level of education. This fact is not as counter-intuitive as it appears, and actually is in line with the cognitive reserve theory [55], which supports the idea that education can help to compensate damaged brain anatomy at the clinical level, maintaining unaltered cognitive capacities for a period of time. In other words, cognitive decline would be delayed with respect to the onset of brain atrophy in educated subjects. Since, in addition, the age at diagnosis is not correlated with the number of years of education in our dataset ($r = -0.02$ and $p = 0.70$ according to a two-tailed test based on Pearson's correlation coefficient), this explains why the subjects present an increased atrophy of their hippocampi for an increased education: they enrolled with a more advanced stage of anatomical pathology, after some years of compensation.



(a) Coronal view.



(b) Axial view.

Fig. 11: Superposition of the male-like (in blue) and the female-like (in pink) hippocampus geometries, in two standard views. The letters L, R, A, P respectively indicate the left, right, anterior and posterior directions.

5.3.5 Simulation of hippocampus atrophy due to AD

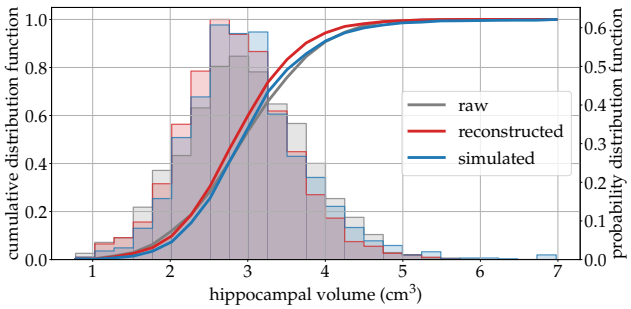
The calibrated models and the empirical distribution of random effects z_i estimated by their personalization to the training data are used to simulate synthetic progressions of the hippocampus. In order to validate such a simulation method, the simulated trajectories are sampled at several ages, and the empirical distribution of the volumes of the simulated hippocampus are compared to the distribution of the original hippocampus. The volume is commonly used as a biomarker in clinical studies, and we aim to assess if the simulated cohort could be used instead of the original one.

To do so, we simulate the same number of subjects as in the training cohort ($n = 322$) with the same number of time-points and same time interval between visits. Note that we do not use the age at baseline, so that the sequence of observation time-points in the synthetic subjects may be shifted in time compared to the real ones. We simulate according to the empirical distribution of the individual parameters z_i and the age at baseline. There exists indeed a correlation between the estimated time-shift τ_i and the baseline age of the enrolled subjects $t_{i,1}$, as they tend to be included in the study at similar disease stage. To be more precise:

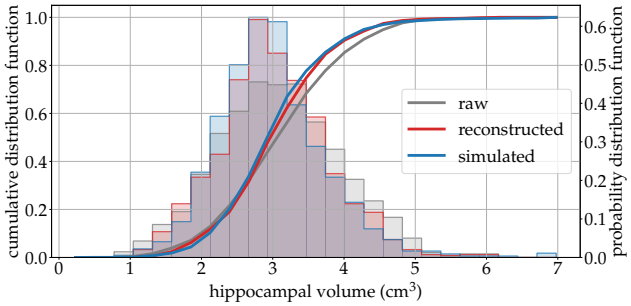
– the empirical joint distribution of the time-related parameters α_i and τ_i augmented with the age at baseline $t_{i,1}$ is computed using a kernel density estimation method;
– the empirical joint distribution of the time-related parameters augmented with the sources s_i is captured by fitting a multivariate Gaussian distribution.

A simulated data set is then created by applying 322 times the following procedure:

– draw the acceleration factor α_i, the onset age τ_i and the baseline age $t_{i,1}$ from the corresponding kernel density;
– draw the sources s_i from the multivariate Gaussian conditional distribution with respect to its already-drawn time-related parameters;
– draw without replacement the sequence of visits of one subject i.e. the number of visits and the time intervals between them;



(a) Left hippocampus: the mean volume is $2958 \pm 779\, mm^3$ for raw data, $2863 \pm 693\, mm^3$ for reconstructed data, and $2865 \pm 746\, mm^3$ for simulated data.



(b) Right hippocampus: the mean volume is $3081 \pm 862\, mm^3$ for raw data, $3014 \pm 754\, mm^3$ for reconstructed data, and $3063 \pm 763\, mm^3$ for simulated data.

Fig. 12: Distribution of the left and right hippocampal volume in the raw, reconstructed and simulated data set. The simulated volume distribution is very close to the volume distribution of the reconstructed data set. The remaining bias between those two distributions and the one corresponding to the raw data comes from the smoothing behavior of the current noise model, leveraged to deal with noisy meshes without point correspondence. See Figure 13.

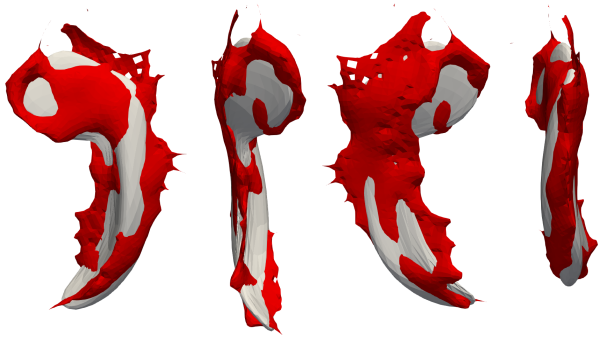


Fig. 13: Several views of a single example of the reconstruction of a right hippocampus structure by the longitudinal shape model. The reconstruction is the smooth white structure, and the raw data point is plotted in red.

– sample the individual hippocampus trajectory defined by $z_i = (\alpha_i, \tau_i, s_i)$ at the baseline age $t_{i,1}$ and the follow-up visits.

This protocol is repeated for both the left and right hippocampus, and for men and women (meaning that the estimation of the empirical distributions is done for both genders separately).

Figure 12 shows the volume distributions of the raw, reconstructed and simulated data. The cumulative distribution functions associated to the simulated and reconstructed distributions of hippocampal volumes are superimposed. This result suggests that for this volume statistic, the simulated and true data set could be used interchangeably. Raw and reconstructed distributions does not superimpose so well, because the model reconstructs smooth shapes whereas raw meshes often have small protrusion pointing outward of the surface which tend to bias volume computation (see Figure 13). This volume difference between the raw and reconstructed meshes amounts on average to 84.5 mm^3 for the left hippocampus and 67.3 mm^3 for the right hippocampus.

Now validated, the simulation algorithm could be used to synthesize a data set of left and right hippocampus of any number of subjects, with any desired visit sampling. The proposed gender-wise split further allows to achieve any desired male-female balance.

6 Conclusion

We proposed a statistical modeling approach that represents individual data sequences as samples along continuous trajectories, these trajectories being considered as spatiotemporal perturbations of a population-average progression. The spatial warp is defined thanks to the exp-parallelization operator on manifolds. The time warps are affine time-reparameterizing functions. The spatial and temporal individual parameters

position the progression of each subject in a spatiotemporal reference frame centered around the average trajectory of the population.

We proposed calibration, personalization and simulation algorithms to address different statistical questions. The calibration algorithm combines the MCMC-SAEM stochastic approach with gradient descent to estimate the underlying common process and its spatiotemporal variability from a longitudinal data set of shapes. It does not require a common time reference to be available across individual processes, which furthermore may be observed each for only short periods of time. Personalizing such calibrated models to a new individual data yields quantitative, low-dimensional and interpretable measures of how the progression of an individual deviates from a normative scenario. These parameters include an acceleration factor and a time-shift on the one hand, and geometrical sources of variability on the other hand. Such individual parameters offer relevant features for classification or correlation tasks, in a post-processing step. The generative nature of the proposed model naturally offers a simulation algorithm, which can generate entirely synthetic data sets. Such data set may be sampled at any desired temporal frequency, for any number of subjects and with a full control over the population characteristics, for instance in terms of gender balance.

We emphasize that the proposed modeling approach is able to deal with meshes without any assumption on their topology, in particular without assuming point-to-point correspondence. It may be extended easily to deal with images or other geometric primitives, provided that one can define a metric between such objects.

The three proposed algorithms were validated in varied simulated configurations, demonstrating their ability to retrieve the true parameters or reproduce the original data distribution. They were illustrated on a data set of facial expressions, showing the relevance of the learned normative scenarios and the potential of the spatiotemporal parameters for classification. We apply the method also to large medical data set of patients that develop Alzheimer's disease. The average scenarii of atrophy for the hippocampus subcortical structures are in line with current medical knowledge. Individual sequences are successfully parametrized by 10 scalar spatiotemporal coordinates in the calibrated reference frames. Correlating these coordinates with genetic, biological and environmental factors gives valuable insights into protective factors influencing age at onset or pace of progression. We also evidence typical shape differences across sub-groups, which are independent of the shape changes due to ageing or disease progression.

The calibration algorithm is computationally intensive: estimating a model of hippocampus progression took around a day. Our code is already parallel, combines CPU and GPU together, and offers a fine-grained initialization pipeline. Further pure optimization of our code (among which multi-GPU support, fast Fourier transforms for convolutions) is planned, as well as evaluating the performance of variational methods for calibration – which are not trivial to implement in a longitudinal context without a fixed number of observations per individual.

As for any modeling approach, our model relies on some assumptions. For instance, subjects are considered to follow trajectories that are parallel to the population average. This hypothesis may be alleviated by introducing drift parameters to model a progressive deviation from the average scenario. Such a development would add to the complexity of the model, which may require to have even more data to be calibrated. Further extensions would consider also to estimate not only one representative trajectory at the population level but several of them, for instance by estimating a mixture model along the lines of [17]. Nonetheless, it is worth noting that in its current form the model is able to reconstruct data at the precision of the noise.

The model also builds on the LDDMM framework for modeling shape variability. This framework relies also on some assumptions on the geometry of the shape space. Future work will consider to learn such geometry from the data instead of relying on prior assumptions, along the lines of [11] for instance. Learning other parameters such as the number of sources, using automatic model selection methods for instance, would also add to the usability of the method.

Acknowledgements The research leading to this publication has been funded in part by the European Research Council (ERC) under grant agreement No 678304 (LEASP), European Union's Horizon 2020 research and innovation programme under grant agreement No 666992 (EuroPOND) and No 826421 (TVB-Cloud), and the program "Investissements d'avenir" ANR-10-IAIHU-06 (IHU ICM) and ANR-19-P3IA-0001 (PRAIRIE 3IA Institute)

The facial expression data set at the basis of Section 5.2 was built and shared by the Binghamton University. The authors warmly thank Pr. Lijun Yin for granting data access, and Peng Liu for his help in downloading the data set.

Regarding Section 5.3, data collection and sharing was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California

Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

1. Abdi, H.: Partial least square regression (pls regression). Encyclopedia for research methods for the social sciences **6**(4), 792–795 (2003)
2. Ahrens, J., Geveci, B., Law, C.: Paraview: An end-user tool for large data visualization. The visualization handbook **717** (2005)
3. Allassonnière, S., Durrleman, S., Kuhn, E.: Bayesian mixed effect atlas estimation with a diffeomorphic deformation model. SIAM Journal on Imaging Science **8**, 1367–1395 (2015)
4. Amor, B.B., Drira, H., Berretti, S., Daoudi, M., Srivastava, A.: 4-d facial expression recognition by learning geometric deformations. IEEE Trans. Cybernetics **44**(12), 2443–2457 (2014)
5. Atchade, Y.F.: An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. Methodology and Computing in applied Probability **8**(2), 235–254 (2006)
6. Banerjee, M., Chakraborty, R., Ofori, E., Okun, M.S., Viallancourt, D.E., Vemuri, B.C.: A nonlinear regression technique for manifold valued data with applications to medical image analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4424–4432 (2016)
7. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. IJCV **61**(2), 139–157 (2005)
8. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) **57**(1), 289–300 (1995)
9. Bilgel, M., Prince, J.L., Wong, D.F., Resnick, S.M., Jedynak, B.M.: A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. Neuroimage **134**, 658–670 (2016)
10. Bône, A., Colliot, O., Durrleman, S.: Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9271–9280 (2018)
11. Bône, A., Louis, M., Colliot, O., Durrleman, S., Initiative, A.D.N., et al.: Learning low-dimensional representations of shape data sets with diffeomorphic autoencoders. In: International Conference on Information Processing in Medical Imaging, pp. 195–207. Springer (2019)
12. Chakraborty, R., Banerjee, M., Vemuri, B.C.: Statistics on the space of trajectories for longitudinal data analysis. In: Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, pp. 999–1002. IEEE (2017)
13. Charlier, B., Feydy, J., Glaunès, J.A., Trouvé, A.: An efficient kernel product for automatic differentiation libraries, with applications to measure transport (2017)
14. Charon, N., Charlier, B., Glaunès, J., Gori, P., Roussillon, P.: Fidelity metrics between curves and surfaces: currents, varifolds, and normal cycles. In: Riemannian Geometric Statistics in Medical Image Analysis, pp. 441–477. Elsevier (2020)
15. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable templates using large deformation kinematics. IEEE transactions on image processing **5**(10), 1435–1447 (1996)
16. Cury, C., Durrleman, S., Cash, D.M., Lorenzi, M., Nicholas, J.M., Bocchetta, M., van Swieten, J.C., Borroni, B., Galimberti, D., Masellis, M., et al.: Spatiotemporal analysis for detection of presymptomatic shape changes in neurodegenerative diseases: Initial application to the genfi cohort. NeuroImage **188**, 282–290 (2019)
17. Debavelaere, V., Bône, A., Durrleman, S., Allassonnière, S., Initiative, A.D.N., et al.: Clustering of longitudinal shape data sets using mixture of separate or branching trajectories. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 66–74. Springer (2019)
18. Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the em algorithm. Annals of statistics pp. 94–128 (1999)
19. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the royal statistical society. Series B (methodological) pp. 1–38 (1977)
20. Durrleman, S., Allassonnière, S., Joshi, S.: Sparse adaptive parameterization of variability in image ensembles. IJCV **101**(1), 161–183 (2013)
21. Durrleman, S., Pennec, X., Trouvé, A., Braga, J., Gerig, G., Ayache, N.: Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. International Journal of Computer Vision **103**(1), 22–59 (2013). DOI 10.1007/s11263-012-0592-x. URL `https://doi.org/10.1007/s11263-012-0592-x`
22. Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trouvé, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. NeuroImage (2014)
23. Fang, T., Zhao, X., Shah, S.K., Kakadiaris, I.A.: 4d facial expression recognition. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pp. 1594–1601. IEEE (2011)
24. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences **97**(20), 11,050–11,055 (2000)
25. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron **33**(3), 341–355 (2002)
26. Fishbaugh, J., Prastawa, M., Gerig, G., Durrleman, S.: Geodesic regression of image and shape data for improved modeling of 4D trajectories. In: ISBI 2014 - 11th International Symposium on Biomedical Imaging, pp. 385 – 388 (2014)
27. Fletcher, T.: Geodesic regression and the theory of least squares on riemannian manifolds. IJCV **105**(2), 171–185 (2013)
28. Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., Poupon, C., Hartmann, A., Ayache, N., Durrleman, S.: A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes. Medical Image Analysis **35**, 458–474 (2017). DOI 10.1016/j.media.2016.08.011. URL `https://hal.inria.fr/hal-01359423`
29. Hinkle, J., Muralidharan, P., Fletcher, P.T., Joshi, S.: Polynomial regression on riemannian manifolds. In: European Conference on Computer Vision, pp. 1–14. Springer (2012)
30. Hirsch, M.W.: Differential topology, vol. 33. Springer Science & Business Media (2012)
31. Hyvärinen, A., Karhunen, J., Oja, E.: Independent component analysis, vol. 46. John Wiley & Sons (2004)
32. Jian, B., Vemuri, B.C.: Robust point set registration using gaussian mixture models. IEEE transactions on pattern analysis and machine intelligence **33**(8), 1633–1645 (2011)
33. Joshi, S.C., Miller, M.I.: Landmark matching via large deformation diffeomorphisms. IEEE Transactions on Image Processing **9**(8), 1357–1370 (2000)
34. Kendall, D.G.: Shape manifolds, procrustean metrics, and complex projective spaces. Bulletin of the London Mathematical Society **16**(2), 81–121 (1984)
35. Kim, H.J., Adluru, N., Suri, H., Vemuri, B.C., Johnson, S.C., Singh, V.: Riemannian nonlinear mixed effects models: Analyzing longitudinal deformations in neuroimaging. In: Proceedings

of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
36. Koval, I., Schiratti, J.B., Routier, A., Bacci, M., Colliot, O., Allassonnière, S., Durrleman, S., Initiative, A.D.N., et al.: Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 451–459. Springer (2017)
37. Kuhn, E., Lavielle, M.: Coupling a stochastic approximation version of em with an mcmc procedure. ESAIM: Probability and Statistics **8**, 115–131 (2004)
38. Liu, D.C., Nocedal, J.: On the limited memory bfgs method for large scale optimization. Mathematical programming **45**(1-3), 503–528 (1989)
39. Lorenzi, M., Ayache, N., Frisoni, G., Pennec, X.: 4D registration of serial brain's MR images: a robust measure of changes applied to Alzheimer's disease. Spatio Temporal Image Analysis Workshop (STIA), MICCAI (2010)
40. Lorenzi, M., Ayache, N., Pennec, X.: Schild's ladder for the parallel transport of deformations in time series of images. In: Biennial International Conference on Information Processing in Medical Imaging, pp. 463–474. Springer (2011)
41. Louis, M., Bône, A., Charlier, B., Durrleman, S.: Parallel transport in shape analysis: a scalable numerical scheme. In: International Conference on Geometric Science of Information, pp. 29–37. Springer (2017)
42. Louis, M., Charlier, B., Jusselin, P., Pal, S., Durrleman, S.: A fanning scheme for the parallel transport along geodesics on riemannian manifolds. SIAM Journal on Numerical Analysis **56**(4), 2563–2584 (2018)
43. Manasse, F., Misner, C.W.: Fermi normal coordinates and some basic concepts in differential geometry. Journal of mathematical physics **4**(6), 735–745 (1963)
44. Marin, J.M., Pudlo, P., Robert, C.P., Ryder, R.J.: Approximate bayesian computational methods. Statistics and Computing **22**(6), 1167–1180 (2012)
45. Marinescu, R.V., Eshaghi, A., Lorenzi, M., Young, A.L., Oxtoby, N.P., Garbarino, S., Shakespeare, T.J., Crutch, S.J., Alexander, D.C., Initiative, A.D.N., et al.: A vertex clustering model for disease progression: application to cortical thickness images. In: International Conference on Information Processing in Medical Imaging, pp. 134–145. Springer (2017)
46. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision **24**(2), 209–228 (2006)
47. Muralidharan, P., Fletcher, P.T.: Sasaki metrics for analysis of longitudinal data on manifolds. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 1027–1034. IEEE (2012)
48. Nader, C.A., Ayache, N., Robert, P., Lorenzi, M.: Monotonic gaussian process for spatio-temporal trajectory separation in brain imaging data. arXiv preprint arXiv:1902.10952 (2019)
49. Niethammer, M., Huang, Y., Vialard, F.X.: Geodesic regression for image time-series. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 655–662. Springer (2011)
50. Pennec, X.: Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. Journal of Mathematical Imaging and Vision **25**(1), 127–154 (2006)
51. Pennec, X., Fillard, P., Ayache, N.: A riemannian framework for tensor computing. International Journal of Computer Vision **66**(1), 41–66 (2006)
52. Schiratti, J.B., Allassonnière, S., Colliot, O., Durrleman, S.: Learning spatiotemporal trajectories from manifold-valued longitudinal data. In: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (eds.) NIPS 28, pp. 2404–2412. Curran Associates, Inc. (2015)
53. Schiratti, J.B., Allassonniere, S., Colliot, O., Durrleman, S.: A bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. The Journal of Machine Learning Research **18**(1), 4840–4872 (2017)
54. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: Hierarchical geodesic models in diffeomorphisms. IJCV **117**(1), 70–92 (2016)
55. Stern, Y.: Cognitive reserve and alzheimer disease. Alzheimer Disease & Associated Disorders **20**(2), 112–117 (2006)
56. Su, J., Kurtek, S., Klassen, E., Srivastava, A., et al.: Statistical analysis of trajectories on riemannian manifolds: bird migration, hurricane tracking and video surveillance. The Annals of Applied Statistics **8**(1), 530–552 (2014)
57. Su, J., Srivastava, A., de Souza, F.D., Sarkar, S.: Rate-invariant analysis of trajectories on riemannian manifolds with application in visual speech recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 620–627 (2014)
58. Sun, Y., Yin, L.: Facial expression recognition based on 3d dynamic range model sequences. In: European Conference on Computer Vision, pp. 58–71. Springer (2008)
59. Vaillant, M., Glaunès, J.: Surface matching via currents. In: Information processing in medical imaging, pp. 1–5. Springer (2005)
60. Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M.: Bayesian analysis of neuroimaging data in fsl. Neuroimage **45**(1), S173–S186 (2009)
61. Yin, L., Chen, X., Sun, Y., Worm, T., Reale, M.: A high-resolution 3d dynamic facial expression database. In: Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on, pp. 1–6. IEEE (2008)
62. Younes, L.: Jacobi fields in groups of diffeomorphisms and applications. Quarterly of Applied Mathematics **65**(1), 113–134 (2007)
63. Younes, L.: Shapes and Diffeomorphisms. Applied Mathematical Sciences. Springer Berlin Heidelberg (2010). URL `https://books.google.fr/books?id=SdTBtMGgeAUC`
64. Zhang, M., Fletcher, P.T.: Finite-dimensional lie algebras for fast diffeomorphic image registration. In: International Conference on Information Processing in Medical Imaging, pp. 249–260. Springer (2015)
65. Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: IPMI, vol. 23, pp. 37–48 (2013)

A Background: meshes represented as currents

The theory of currents has been introduced in [59], and is used in this paper to define a distance metric between pairs of meshes without any assumption on their topology, and in particular without assuming point-to-point correspondence. See also [14] for more details.

A.1 Continuous theory

Let y be a surface mesh, that we represent as an infinite set of tuples $(x, n(x))$ where x is a point of $\mathbb{R}^3$, and $n(x)$ the normal vector of y at this point. Let $g_{\mathcal{E}} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ be a positive-definite kernel operator, and $\mathcal{E}$ the associated reproducing kernel Hilbert space. We define the current transform $\mathcal{C}(y) : \mathbb{R}^3 \to \mathbb{R}^3 \in \mathcal{E}$ of y as:

$$\mathcal{C}(y)(.) = \int_y g_{\mathcal{E}}(x, .) \cdot n(x) \cdot d\sigma(x)$$

where $d\sigma(x)$ denotes an infinitesimal surface element of y. The inner product of $\mathcal{E}$ on currents therefore writes:

$$\left\langle \mathcal{C}(y), \mathcal{C}(y') \right\rangle_{\mathcal{E}} = \int_y \int_{y'} n'(x')^\top \cdot g_{\mathcal{E}}(x, x') \cdot n(x) \cdot \sigma(x) \cdot d\sigma'(x')$$

where $(.)^\top$ is the transposition operator. This inner product defines in turn a distance metric on currents:

$$d_{\mathcal{E}}(\mathcal{C}, \mathcal{C}') = \left\langle \mathcal{C}, \mathcal{C} \right\rangle_{\mathcal{E}} + \left\langle \mathcal{C}', \mathcal{C}' \right\rangle_{\mathcal{E}} - 2 \cdot \left\langle \mathcal{C}, \mathcal{C}' \right\rangle_{\mathcal{E}}.$$

A.2 Practical discrete case

In practice, y is described by a finite set of T triangles in $\mathbb{R}^3$ of centers $c_1, ..., c_T$ and corresponding surface normal vectors $n_1, ..., n_T$. We further assume that $g_{\mathcal{E}}$ is a Gaussian kernel of radius $\sigma_{\mathcal{E}}$. The current transform equation then writes:

$$\mathcal{C}(y)(x) = \sum_{k=1}^{T} \exp \frac{-\|x - c_k\|_{\ell^2}^2}{\sigma_{\mathcal{E}}^2} \cdot n_k$$

for any $x \in \mathbb{R}^3$. Similarly, the inner product formula becomes:

$$\left\langle \mathcal{C}(y), \mathcal{C}(y') \right\rangle_{\mathcal{E}} = \sum_{k=1}^{T} \sum_{l=1}^{T'} \exp \frac{-\|c'_l - c_k\|_{\ell^2}^2}{\sigma_{\mathcal{E}}^2} \cdot n_k^\top n'_l$$

which fully specifies the distance metric $d_{\mathcal{E}}$ that can be implemented in practice to measure the discrepancy between any pair of currents.