

A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations

Jean-Baptiste Schiratti, Stéphanie Allassonniere, Olivier Colliot, Stanley Durrleman

▶ **To cite this version:**

Jean-Baptiste Schiratti, Stéphanie Allassonniere, Olivier Colliot, Stanley Durrleman. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. Journal of Machine Learning Research, 2017, 18, pp.1-33. hal-01540367v2

HAL Id: hal-01540367
https://hal.science/hal-01540367v2

Submitted on 5 Dec 2017 (v2), last revised 31 Oct 2018 (v3)

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

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

Journal of Machine Learning Research 18 (2017) 1-29 Submitted 4/17; Revised 9/17; Published 9/17

A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations

Jean-Baptiste Schiratti JEAN-BAPTISTE.SCHIRATTI@CMAP.POLYTECHNIQUE.FR *CMAP, Ecole Polytechnique*
Route de Saclay, 91128 *Palaiseau*

Stéphanie Allassonnière STEPHANIE.ALLASSONNIERE@POLYTECHNIQUE.EDU
INSERM, UMRS 1138, team 22. Centre de recherche des Cordeliers, Université Paris 5 René Descartes

Olivier Colliot OLIVIER.COLLIOT@UPMC.FR

Stanley Durrleman STANLEY.DURRLEMAN@INRIA.FR
*ARAMIS Lab, INRIA Paris, Inserm U*1127*, CNRS UMR* 7225*, Sorbonne Universités, UPMC Univ Paris* 06 *UMRS* 1127
Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France

Editor: David Sontag

Abstract

We propose a generic Bayesian mixed-effects model to estimate the temporal progression of a biological phenomenon from observations obtained at multiple time points for a group of individuals. The progression is modelled by continuous trajectories in the space of measurements. Individual trajectories of progression result from spatiotemporal transformations of an average trajectory. These transformations allow for the quantification of changes in direction and pace at which the trajectories are followed. The framework of Riemannian geometry allows the model to be used with any kind of measurements with smooth constraints. A stochastic version of the Expectation-Maximization algorithm was used to produce maximum a posteriori estimates of the parameters. We evaluated our method using a series of neuropsychological test scores from patients with mild cognitive impairments, later diagnosed with Alzheimer's disease, and simulated evolutions of symmetric positive definite matrices. The data-driven model of impairment of cognitive functions showed the variability in the ordering and timing of the decline of these functions in the population. We also showed that the estimated spatiotemporal transformations effectively put into correspondence significant events in the progression of individuals.

Keywords: longitudinal model, spatiotemporal analysis, Riemannian geometry, stochastic expectation-maximization algorithm

1. Introduction

The study of the temporal progression of a biological or natural phenomenon is central to several scientific fields. For instance, the study of progressive diseases plays a crucial role in the diagnosis and prognosis of patients. In computer vision, the dynamics of facial ex-

©2017 J.-B. Schiratti et al..

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v18/jbs.html.

pressions in video sequences may be important in automatically detecting and charaterizing emotions.

For a given individual or object, the evolution of the observed phenomenon can be measured by several characteristics or features, which describe the state of the individual at a given time point. In medicine, these features may be blood markers, height, or weight, but also structured multivariate data such as medical images. The shape of a human face may be described by the position of characteristic points in the nose, mouth or brows. These features may be represented, at a given time point, as a point in a high-dimensional space. The temporal evolution of these features may be modelled therefore as a smooth parametric curve in the space of measurements, i.e. a spatiotemporal trajectory. These trajectories vary across individuals in two possible ways. Firstly, the position and direction of the trajectory differ because the measurements have intrinsically different values and different trajectory of changes for different individuals. Secondly, the pace at which the trajectory is followed (i.e. the way the curve is parameterized) varies because some individuals may follow the same progression pattern but at a different age and possibly at a different speed. We refer to the first type of variability as a *spatial* variability, and the second type as a *temporal* variability, leading together to the concept of *spatiotemporal* variability.

The goal of this paper was to automatically estimate the typical trajectory of changes and its spatiotemporal variability within a group of individuals. We aimed to infer such spatiotemporal patterns from longitudinal data sets, which consisted of repeated observations of the same biological phenomenon at several time points for a group of individuals. The time points and their number may have varied for different individuals.

In the literature, *mixed-effects models* (Eisenhart, 1947; Laird and Ware, 1982) and (Verbeke and Molenberghs, 2009) appear as a popular method for the analysis of longitudinal data. These statistical models include *fixed* and *random effects* which provide these models with a hierarchical structure, where fixed effects described the data at the population (or group) level, and the random effects at the individual level. By fitting a mixed-effects model, one can discern an average trajectory as well as individual-specific trajectories. Moreover, mixed-effects models enforce conditions on the distribution of the random effects, thus opening up the possibility to discern a distribution of trajectories in the space of observations.

Linear Mixed Effects (LME) models are the most simple mixed-effects models introduced in Laird and Ware (1982). A particular, but yet informative case of the LME models for analyzing longitudinal data is the *random slope and intercept model*. This model is written: $\mathbf{y}_{i,j} = (t_{i,j}-t_0)(\overline{\mathbf{A}}+\mathbf{A}_i)+(\overline{\mathbf{B}}+\mathbf{B}_i)+\boldsymbol{\varepsilon}_{i,j}$, where $t_0 \in \mathbb{R}$ and $(t_{i,j})_{1\leq j\leq k_i}$ denotes the time points at which the observations $\mathbf{y}_{i,j} \in \mathbb{R}^n$ of the ith individual were obtained. The population parameters (or fixed effects) of the model are the slope $\overline{\mathbf{A}}$ and the intercept $\overline{\mathbf{B}}$. The random effects are the subject-specific slopes $(\mathbf{A}_i)_{1\leq i\leq p}$ and intercepts $(\mathbf{B}_i)_{1\leq i\leq p}$, which are assumed to be normally distributed and independent of each other. This random slope and intercept model estimates an average trajectory $\overline{\mathbf{D}}(t) = (t - t_0)\overline{\mathbf{A}} + \overline{\mathbf{B}}$. The random effects of the model allows us to also estimate individual trajectories $\mathbf{D}_i(t) = (t-t_0)(\overline{\mathbf{A}}+\mathbf{A}_i)+(\overline{\mathbf{B}}+\mathbf{B}_i)$, which are obtained by adjusting the slope and intercept of the average trajectory. This model is essentially built on the idea of regressing the measurements against time. The parameter t_0 can be understood as a *reference time*. If the longitudinal dataset arises from animal breeding studies, developmental studies or pharmacological studies, the reference

time t_0 may be chosen to be the date of birth or the time at which a drug was administered. However, there are many situations in which there is no obvious reference time t_0 at which observations may be compared. In ageing, for instance, different individuals of the same age may be at different stages of ageing, or stages of disease progression. Therefore, it does not make sense to regress the measurements against age, or, in other words, to statistically compare measurements at a given age. In video sequences, there is no obvious way to find the frames corresponding to the same event in two different sequences. By contrast, we would like this temporal alignment of the trajectories to be automatically estimated from the data. Adding the reference time t_0 as a new parameter of the model is not a solution as the model becomes non-identifiable: an infinite number of triplets $(\overline{A}, \overline{B}, t_0)$ parameterize the same trajectory.

In Yang et al. (2011) and Delor et al. (2013), the authors addressed this problem by introducing *time shifts* in their statistical analysis. In Durrleman et al. (2009, 2013), time reparametrizations called *time warps* (smooth monotonic transformations of the real line) are considered to address this point in the context of longitudinal shape analysis, and parameters were estimated by optimizing an uncontrolled approximation of the likelihood. In Hong et al. (2014), the authors used parametric time warps with a regression model for shape analysis. In Lorenzi et al. (2015), the authors used Riemannian manifold techniques to estimate a model of normal brain ageing from MR images. The model was used to compute a time shift, called *morphological age shift*, which corresponded to the actual anatomical age of the subject with respect to an estimated reference age. In (Fonteijn et al., 2012; Young et al., 2015), the authors developed a statistical model called *the Event-Based Model*, which estimated an ordering of categorical variables. The model was used to estimate the progression of a series of events. However, these models do not allow estimation of the relative timing between two consecutive events. In Jedynak et al. (2012), the authors modelled the progression of biomarkers using a nonlinear mixed-effects model for univariate observations. This model estimated individual trajectories which are defined using individual-specific time reparametrizations of an average trajectory. However, the proposed model was not identifiable unless some conditions were imposed on the parameters of the model. Therefore, generalizing the model to multivariate observations is not straightforward. Also, the model was specific to univariate observations whereas our generic model, presented below, allows analysis of any kind of observations defined by smooth constraints. This work offers pragmatic solutions to include the idea of time reparameterization in the estimation of trajectories of changes for some specific applications. Nevertheless, we are still lacking a principle and generic approach to deal with the estimation of spatiotemporal variability in longitudinal data sets.

Structured multivariate data such as images, graphs, shapes, or positive definite matrices add further difficulty as these data do not lie in Euclidean spaces. Algebraic operations such as addition or scaling are not defined or do not yield an output of the same type. The spaces in which they live are defined by smooth constraints and may be considered in general as *Riemannian manifolds*. There is no natural extension of LME models on Riemannian manifolds. In Fletcher (2011), the authors proposed an extension of linear regression for Riemannian manifolds, which was later extended for longitudinal data in a group of diffeomorphisms (Singh et al., 2013, 2014). Nevertheless, this longitudinal model strongly depends on a choice of reference time-point to define random effects, therefore

making difficult its coupling with time reparameterization. In Su et al. (2014), trajectories were defined by the quotient with a time reparameterization group. This approach allowed the definition of statistics in the quotient space, but as a consequence did not yield any estimate of the temporal variability.

This paper proposes a Bayesian mixed-effects model, called *generic spatiotemporal model*, defined for any longitudinal observations on a Riemannian manifold. The fixed effects of the model were used to define an average trajectory and the random effects were used to define individual-specific trajectories. In order to define such individual trajectories, we introduced the notion of "exp-parallelization" of a curve on a Riemannian manifold, based on the idea of "variations of a curve". This construction allowed the definition of random effects to account for the spatial variability, the distribution of which (up to an isometric transform) does not depend on a reference time-point. It allowed us, therefore, to easily include random time re-parameterization to account for the temporal variability in the model. All in one, the model defines distributions of spatiotemporal trajectories for data on any Riemannian manifolds. It gives a systematic way to derive specific nonlinear mixed-effects models for a large variety of observations and Riemannian manifolds.

These models needed to then be fitted to given longitudinal data sets. Given their strong non-linear nature, we proposed to use a stochastic version of the Expectation-Maximization (EM) (Dempster et al., 1977) algorithm, called the Monte Carlo Markov Chain Stochastic Approximation EM (MCMC-SAEM) algorithm. Theoretical results regarding the convergence of the MCMC-SAEM have been proven in Kuhn and Lavielle (2004),
Allassonnière et al. (2010) and ensure that the algorithm maximizes the observed likelihood. This technique allowed us to propose a generic algorithm for the estimation of the model parameters. We will instantiate this method for a set of multivariate bounded measurements and for positive definite matrices.

The paper was organized as follows: in Section 2, we gave the key mathematical tools and defined the generic mixed-effects model with spatiotemporal transformations for manifold-valued measurements. Particular cases of the generic model were given and discussed in Section 3. Section 4 was focused on the MCMC-SAEM which was used to estimate the parameters of the statistical model. Finally, Sections 5.2 and 5 were dedicated to empirical and experimental validations of our generic model.

2. A Bayesian Mixed-Effects Model for Longitudinal Observations on a Riemannian Manifold

This section is aimed at introducing a notion of Riemannian geometry called "exp-parallelization". Given a group-average trajectory on a Riemannian manifold, the notion of exp-parallelization an be used to define individual trajectories. For a comprehensive review of basic concepts of Riemannian geometry, see Do Carmo Valero (1992); Petersen (2006). In this section, we assumed that $\mathbb{M}$ is an open subset of $\mathbb{R}^N$ equipped with a Riemannian metric $g^{\mathbb{M}}$.

2.1 Exp-parallelization on a Riemannian Manifold

This section introduces the notion of "exp-parallelization" of a curve on a Riemannian manifold $(\mathbb{M}, g^{\mathbb{M}})$. The notion of "variation of a differentiable curve" on a manifold is defined in Do Carmo Valero (1992) (Chapter 9). It allowed us to define *neighbouring* curves to a

given curve c. In the next section, this construction will be used to define individual trajectories. Let $(\mathbb{M}, g^{\mathbb{M}})$ denotes a *geodesically complete* Riemannian manifold equipped with its Levi-Civita connection $\nabla^{\mathbb{M}}$.

Definition 1 *Let $c : I \subset \mathbb{R} \to \mathbb{M}$ a differentiable curve on $\mathbb{M}$, $t_0 \in I$ and $\mathbf{w} \in \mathrm{T}_{c(t_0)}\mathbb{M}$ a tangent vector to $\mathbb{M}$ at $c(t_0)$. An* exp-parallelization of c in the direction of $\mathbf{w}$ *is a curve $\boldsymbol{\eta}^{\mathbf{w}}(c, \cdot) : I \to \mathbb{M}$ defined by:*

$$\forall t \in I,\ \boldsymbol{\eta}^{\mathbf{w}}(c, t) = \mathrm{Exp}^{\mathbb{M}}_{c(t)}\big(\mathrm{P}_{c,t_0,t}(\mathbf{w})\big). \tag{1}$$

This construction is illustrated in Fig. 1. Given $t \in I$, parallel transport carries the tangent vector $\mathbf{w}$ from $\mathrm{T}_{c(t_0)}\mathbb{M}$ to $\mathrm{T}_{c(t)}\mathbb{M}$ along the curve c. At the point $c(t)$, a new point on $\mathbb{M}$ is obtained by taking the Riemannian exponential of the tangent vector $\mathrm{P}_{c,t_0,t}(\mathbf{w})$. This new point is denoted by $\boldsymbol{\eta}^{\mathbf{w}}(c, t)$. As t varies, one describes a curve $\boldsymbol{\eta}^{\mathbf{w}}(c, \cdot)$ on $\mathbb{M}$, which can be understood as a "parallel" to the curve c. Note that if $\mathbb{M}$ is the Euclidean space $\mathbb{R}^N$, an exp-parallelization of a curve c, in the direction of a tangent vector $\mathbf{w}_i$, is the translation of c by the vector $\mathbf{w}_i$.

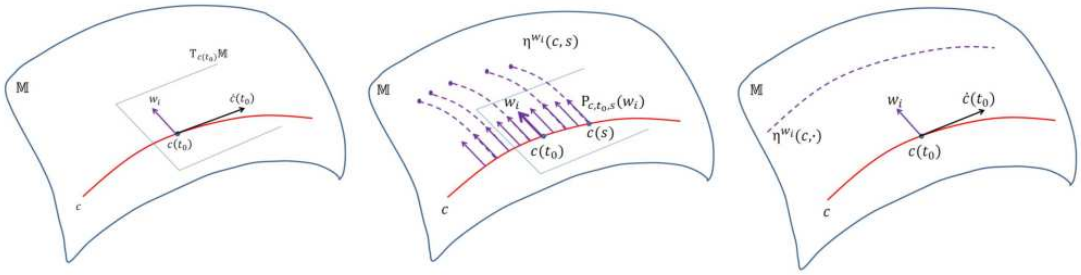


Figure 1: Exp-parallelization on a schematic manifold. Left: a non-zero vector $\mathbf{w}_i$ is chosen in $\mathrm{T}_{c(t_0)}\mathbb{M}$. Middle: the tangent vector $\mathbf{w}_i$ is transported continuously along the curve c. Then, a point $\eta^{\mathbf{w}_i}(c, s)$ is constructed at time s by use of the Riemannian exponential. Right: The curve $\eta^{\mathbf{w}_i}(c, \cdot)$ is the "parallel" resulting from the construction.

2.2 Hierarchical Structure of the Model

In this section, we considered a longitudinal dataset $(\mathbf{y}_{i,j})_{1\leq i\leq p,\, 1\leq j\leq k_i}$. The observations were obtained for a group of p individuals. For the ith individual, the observations $(\mathbf{y}_{i,j})_{1\leq j\leq k_i}$ were obtained at times $t_{i,1} < \ldots < t_{i,k_i}$. The number k_i of observations may have varied from one individual to another.

The *generic spatiotemporal* model is a nonlinear mixed-effects model. As emphasized in the introduction, mixed-effects models include fixed and random effects. The *fixed-effects* are parameters which are shared by all the individuals and allowed us to describe the model at the population level. *Random effects* are individual-specific random variable which describe the model at the individual level. These two types of effects provided the model with a hierarchical structure. The generic spatiotemporal model was constructed as follows. To

begin with, a group-average trajectory $\boldsymbol{\gamma}_0$ was defined on the manifold $\mathbb{M}$. Given the average trajectory, subject-specific trajectories were obtained by spatiotemporal transformations, which consisted *exp-parallelizations* of the average trajectory $\boldsymbol{\gamma}_0$ and *time reparametrization*. The data points $\mathbf{y}_{i,j}$ were seen as samples along these individual trajectories. If $\boldsymbol{\gamma}_i$ denoted the trajectory of the ith individual, the model wrote: $\mathbf{y}_{i,j} = \boldsymbol{\gamma}_i(t_{i,j}) + \varepsilon_{i,j}$, where $\varepsilon_{i,j}$ is Gaussian noise. The observation $\mathbf{y}_{i,j}$ was therefore considered as a small perturbation of a quantity which lies in a Riemannian manifold.

The group-average trajectory $\boldsymbol{\gamma}_0$ was chosen to be the unique geodesic $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}_{\mathbf{p}_0,t_0,\mathbf{v}_0}$ of $\mathbb{M}$ which goes through the point $\mathbf{p}_0 \in \mathbb{M}$ at time t_0 and with velocity $\mathbf{v}_0 \in T_{\mathbf{p}_0}\mathbb{M}$. Let $i \in \{1, \ldots, p\}$ denote the ith individual. The subject-specific trajectory $\boldsymbol{\gamma}_i$ was defined in two steps. The first step consisted of constructing the curve $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_0, \cdot)$, which is an *exp-parallelization* of the average trajectory $\boldsymbol{\gamma}_0$ in the direction of a tangent vector $\mathbf{w}_i \in T_{\mathbf{p}_0}\mathbb{M}$. This tangent vector was chosen to be orthogonal, for the inner product $g^{\mathbb{M}}_{\mathbf{p}_0}$, to $\dot{\boldsymbol{\gamma}}_0(t_0) = \mathbf{v}_0$. The tangent vectors $(\mathbf{w}_i)_{1\leq i\leq p}$ are random effects of the model, called *space shifts*. The orthogonality condition on the space shifts is discussed below. The second step consisted of *reparametrizing in time* the exp-parallelization $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_0, \cdot)$. We considered a subject-specific affine mapping ψ_i of the form $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$, where $\alpha_i > 0$ and $\tau_i \in \mathbb{R}$ are random effects of our model. The trajectory $\boldsymbol{\gamma}_i$ of the ith individual is $\boldsymbol{\gamma}_i(t) = \boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0, \psi_i(t))$. The mapping ψ_i was called *time reparametrization* and the random effects α_i (respectively τ_i) were called *acceleration factor* (respectively *time shift*).

2.3 Definition of the Space Shifts

As mentioned above, the space shifts $(\mathbf{w}_i)_{1\leq i\leq p}$ were required to be orthogonal to $\dot{\boldsymbol{\gamma}}_0(t_0) = \mathbf{v}_0$ for the inner product $g^{\mathbb{M}}_{\mathbf{p}_0}$ induced by the Riemannian metric on $\mathbb{M}$. This section discusses different methods which allowed the inclusion of this orthogonality condition on the space shifts into a statistical model. The methodological challenge raised by this section consisted of defining a (nonlinear) mixed-effects model with smooth constraints on some of the random effect of the model.

In order to ensure the interpretability of the space shifts, we considered an Independent Component Analysis (ICA) (Hyvärinen et al., 2004) decomposition of each tangent vector $\mathbf{w}_i$ as a linear combination of $N_s < N$ statistically independent tangent vectors $(\mathbf{A}_l)_{1\leq l\leq N_s}$ which are called *independent components* or *independent directions*. As a consequence, the space shifts $(\mathbf{w}_i)_{1\leq i\leq p}$ were defined as follows: $\forall i \in \{1, \ldots, p\}$, $\mathbf{w}_i = \mathbf{A}\mathbf{s}_i = \sum_{l=1}^{N_s} s_{l,i}\mathbf{A}_l$ where $\mathbf{A} = (\mathbf{A}_l)_{1\leq l\leq N_s}$ is such that each $\mathbf{A}_i$ is a vector in $\mathrm{T}_{\boldsymbol{\gamma}_0(t_0)}\mathbb{M}$. In this definition, the weights $(s_{l,i})_{1\leq l\leq N_s}$ are random effects of the model called *sources*. By defining the space shifts this way, the generic spatiotemporal model will estimate an ICA decomposition of the space shifts. However, this definition does not ensure the orthogonality of the space shifts. A possible solution to make the vectors $\mathbf{w}_i$ orthogonal to $\mathbf{v}_0 = \dot{\boldsymbol{\gamma}}_0(t_0)$ consisted of decomposing each vector $(\mathbf{A}_l)_{1\leq l\leq N_s}$ in an orthonormal basis of $\mathrm{Span}\big(\dot{\boldsymbol{\gamma}}_0(t_0)\big)^{\perp} \subset \mathrm{T}_{\mathbf{p}_0}\mathbb{M}$. Indeed, if $(\mathcal{B}_k)_{1\leq k\leq (N-1)N_s}$ is an orthonormal basis of $\mathrm{Span}\big(\dot{\boldsymbol{\gamma}}_0(t_0)\big)^{\perp}$, we assumed that: $\forall l \in \{1, \ldots, N_s\}$, $A_l = \sum_{k=1}^{(N-1)N_s} \beta_{l,k}\mathcal{B}_k$. By construction, each independent component $\mathbf{A}_l$ $(1 \leq l \leq N_s)$ is orthogonal, for the inner product $g^{\mathbb{M}}_{\mathbf{p}_0}$, to $\mathbf{v}_0$. Therefore, each space shift $(\mathbf{w}_i)_{1\leq i\leq p}$ is orthogonal to $\mathbf{v}_0$ since we assumed that it writes as a linear combination of

these independent components. In the following, the orthonormal basis was computed using the Gram-Schmidt algorithm or the Householder method (Coleman and Sorensen, 1984).

Moreover, it is important to note that the choice of the form of the distribution of the space-shifts did not depend on the reference time-point t_0. Indeed, the $\mathbf{w}_i = \mathbf{A}\mathbf{s}_i$ were defined in the tangent space of the curve at point $\mathbf{p}_0 = \boldsymbol{\gamma}_0(t_0)$. At another point $\mathbf{p}'_0 = \boldsymbol{\gamma}_0(t'_0)$, space-shifts become $\mathbf{w}'_i = \mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t'_0}\mathbf{w}_i$, where $\mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t'_0}$ is an orthogonal matrix. They were therefore distributed according to $\mathbf{w}'_i = P_{\boldsymbol{\gamma}_0,t_0,t}\mathbf{A}\mathbf{s}_i$: the distribution of the sources $\mathbf{s}_i$ did not change and the independent components (i.e. the columns of $\mathbf{A}$) were adjusted to the new position on the average trajectory. In particular, the variance of the $\mathbf{w}'_i$ was invariant. This property holds for isometric invariant distributions. For instance, if $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, then $\mathbf{w}'_i \sim \mathcal{N}(\mathbf{0}, \mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t'_0}\,\boldsymbol{\Sigma}\,\mathrm{P}^{\top}_{\boldsymbol{\gamma}_0,t_0,t'_0})$.

2.4 The Statistical Model

The generic spatiotemporal model assumed that the jth observation of the ith individual derives from:

$$\mathbf{y}_{i,j} = \eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_0, \psi_i(t_{i,j})) + \boldsymbol{\varepsilon}_{i,j}. \tag{2}$$

With the notations introduced above, let $\mathbf{z}_{\text{pop}} = (\mathbf{p}_0, t_0, \mathbf{v}_0, (\beta_{l,k})_{l,k})$ denote the *population variables* and $(\mathbf{z}_i)_{1\leq i\leq p}$ denote the set of *individual variables* with: $\mathbf{z}_i = (\xi_i, \tau_i, (s_{l,i})_{l,i})$. Both $\mathbf{z}_{\text{pop}}$ and $(\mathbf{z}_i)_{1\leq i\leq p}$ are *latent* (or random) variables assumed independent of each other and distributed as follows:

$$\begin{aligned} &\mathbf{p}_0 \sim \mathcal{N}(\overline{\mathbf{p}_0}, \sigma^2_{\mathbf{p}_0}), \quad t_0 \sim \mathcal{N}(\overline{t_0}, \sigma^2_{t_0}), \\ &\mathbf{v}_0 \sim \mathcal{N}(\overline{v_0}, \sigma^2_{v_0}), \quad \beta_{l,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\overline{\beta}_{l,k}, \sigma^2_{\beta}), \end{aligned} \tag{3}$$

and $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ with $\alpha_i = \exp(\xi_i)$ and :

$$\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_{\xi}), \quad \tau_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2_{\tau}), \quad s_{l,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1). \tag{4}$$

where $\sigma^2_{\mathbf{p}_0}$, $\sigma^2_{t_0}$, $\sigma^2_{\mathbf{v}_0}$ and σ^2_{β} are fixed variance parameters. The noise variables $(\boldsymbol{\varepsilon}_{i,j})_{i,j}$ are assumed independent of the other random variables and identically distributed:

$$\boldsymbol{\varepsilon}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2). \tag{5}$$

Let $\boldsymbol{\theta}_{\text{var}} = (\sigma^2_{\xi}, \sigma^2_{\tau}, \sigma^2)$ denote the variance parameters which are not fixed and $\boldsymbol{\theta} = \left(\overline{\mathbf{p}_0}, \overline{t_0}, \overline{\mathbf{v}_0}, (\overline{\beta}_{l,k}), \boldsymbol{\theta}_{\text{var}}\right)$ be the *parameters* of the model. The domain of $\boldsymbol{\theta}$ is denoted by Θ and defined by:

$$\begin{aligned} \Theta = \big\{\boldsymbol{\theta} = &(\overline{\mathbf{p}_0}, \overline{\mathbf{v}_0}, \overline{t_0}, (\overline{\beta}_{l,k})_{l,k}, \boldsymbol{\theta}_{\text{var}}) \,/ (\overline{\mathbf{p}_0}, \overline{\mathbf{v}_0}) \in \mathrm{T}\mathbb{M}, \\ &\overline{t_0} \in \mathbb{R}, (\overline{\beta}_{l,k})_{l,k} \in \mathbb{R}^{(N-1)N_s}, \boldsymbol{\theta}_{\text{var}} \in]0, +\infty[^3\big\}. \end{aligned} \tag{6}$$

2.4.1 Discussion

The additive, or *extrinsic*, noise model in Eq. (2) makes sense because we assumed that $\mathbb{M}$ is a subset of the Euclidean space $\mathbb{R}^N$. The term $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_0, \psi_i(t_{i,j}))$ belongs to the manifold $\mathbb{M}$ while the noise term $\boldsymbol{\varepsilon}_{i,j}$ is added in the underlying Euclidean space. However, the noise

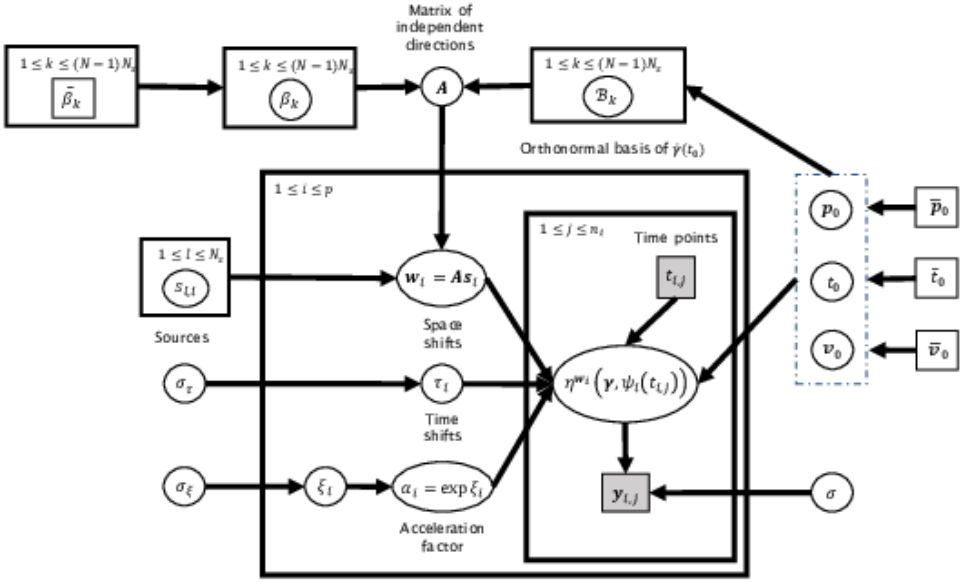


Figure 2: Graphical representation of the generic spatiotemporal model. Round shapes indicate latent variables of the model. Boxes with indexes in the upper left corner indicate a repetition. Shaded boxes indicate that the quantity is observed. This figure illustrates the dependence between the variables of the generic spatiotemporal model.

model is not *intrinsic* in the sense that the noise term $\boldsymbol{\varepsilon}_{i,j}$ is not added on the manifold. In Fletcher (2011), the author has considered an intrinsic noise model which would write: $\mathbf{y}_{i,j} = \text{Exp}_{\boldsymbol{\eta}^{\mathbf{w}_i}(\boldsymbol{\gamma}_0, \psi_i(t_{i,j}))}(\boldsymbol{\varepsilon}_{i,j})$. This noise model allowed for it to remain on the manifold. Still, obtaining maximum *a posteriori* estimates of the parameters with this intrinsic noise model is more difficult as the model likelihood might not be available in closed-form.

We assumed a centered log-normal distribution for the acceleration factors α_i. Indeed, this choice of probability distribution ensured the positiveness of the acceleration factors. With this assumption, the individual time reparametrizations do not reverse time. Other probability distributions, such as the exponential distribution, could have been considered.

3. Particular Cases of the Generic Spatiotemporal Model

The generic spatiotemporal model, introduced in the previous section, is a statistical tool which allowed us, given a Riemannian manifold $\mathbb{M}$ equipped with a Riemannian metric $g^{\mathbb{M}}$, to instantiate a large variety of nonlinear mixed-effects models. This section is aimed at describing the generic spatiotemporal model for classical Riemannian manifolds. The models

for one-dimensional geodesically complete Riemannian manifolds given in Section 3.1 were introduced in Schiratti et al. (2015c). The progression models, given in Section 3.3 were introduced in Schiratti et al. (2015b).

3.1 The case of One-Dimensional Geodesically Complete Riemannian Manifolds

Let M be an open interval of $\mathbb{R}$ equipped with a Riemannian metric g^M, for which it is geodesically complete. The case of one dimensional manifolds is particular because, for all $p_0 \in M$, $\mathrm{T}_p M \simeq \mathbb{R}$ and given $v_0 \in \mathrm{T}_{p_0} M$, there is only one tangent vector w at p_0 which is orthogonal (for the inner product $g^M_{p_0}$) to v_0 : $w = 0$. As a result, if γ_0 is a geodesic of M, $t_0 \in \mathbb{R}$ and $w = 0$, then for all $s \in \mathbb{R}$, $\eta^w(\gamma_0, s) = \gamma_0(s)$. Therefore, the generic spatiotemporal model writes: $y_{i,j} = \gamma_0 \circ \psi_i(t_{i,j}) + \varepsilon_{i,j}$, with, for all $i \in \{1, \ldots, p\}$, $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ and $\alpha_i = \exp(\xi_i)$.

We have shown that, in this one-dimensional framework, a different presentation of the generic spatiotemporal model is possible. This presentation provided a different insight on the role of the latent variables $(\alpha_i, \tau_i)_{1 \leq i \leq p}$. Let $p_0 \in M$, $t_0 \in \mathbb{R}$ and $v_0 \in \mathrm{T}_{p_0} M \simeq \mathbb{R}$. Let γ_0 be the group-average trajectory defined as the geodesic which goes through the point p_0 at time t_0 and with velocity v_0. Let $1 \leq i \leq p$. The trajectory γ_i of the ith individual is defined as the geodesic γ_i which goes through the point p_0 at time $t_0 + \tau_i$ and with velocity $\alpha_i v_0$. Having defined individual trajectories of progression, the observations were seen as random samples along these trajectories: $y_{i,j} = \gamma_i(t_{i,j}) + \varepsilon_{i,j}$. In this definition, the acceleration factor α_i allowed us to characterize whether the ith individual is progressing faster ($\alpha_i > 1$) or slower ($\alpha_i < 1$) than the average trajectory. The time shift τ_i allowed us to determine whether the ith individual is evolving ahead ($\tau_i < 0$) or behind ($\tau_i > 0$) the average trajectory. Moreover, it followed from a unicity property of the geodesics that, for all $i \in \{1, \ldots, p\}$, $\gamma_i(t) = \gamma_0\big(\psi_i(t)\big)$. This result legitimised the choice of affine time reparametrizations of the form $\psi_i : t \mapsto \alpha_i(t - t_0 - \tau_i) + t_0$.

3.1.1 The "Straight Lines Model"

Unbounded observations can be considered as points on the real line. The real line $M = \mathbb{R}$ equipped with its canonical metric is a geodesically complete one-dimensional Riemannian manifold. For the canonical metric, the geodesics are of the form $t \in \mathbb{R} \mapsto at + b$ with $(a, b) \in \mathbb{R}^2$. The generic spatiotemporal model writes: $y_{i,j} = p_0 + \alpha_i v_0(t_{i,j} - t_0 - \tau_i) + \varepsilon_{i,j}$. This model is referred to as the *univariate straight lines model*. Note that, even though the average and individual trajectories are straight lines, the model is *not* linear due to the multiplication between the random effects α_i and τ_i.

We proposed comparison of the nonlinear straight lines model to the random slope and intercept model, as discussed in the introduction. This linear mixed-effects model writes: $y_{i,j} = (\overline{a} + a_i)(t_{i,j} - t_0) + (\overline{b} + b_i) + \varepsilon_{i,j}$, where $(A_i, B_i)_{1 \leq i \leq p}$ are random effects of the model which are assumed to be independent of each other and normally distributed with mean 0 and variance-covariance matrix $\mathbf{D}$. The fixed effects of this model are $(\overline{A}, \overline{B}, t_0)$. This linear model analyzed the distribution of the observations at a fixed reference time t_0. In

comparison, the straight lines model analyzed the distribution of the times at which the observations reached a given value of the measurements. These two different approaches are illustrated in Fig. 3.

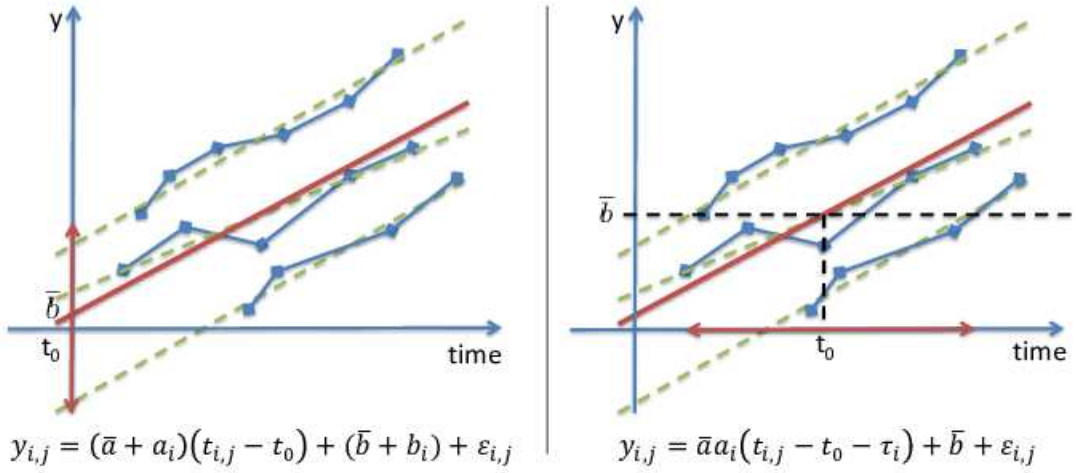


Figure 3: Schematic example of a random slope and intercept linear mixed-effects model (left) and straight lines model (right).

3.2 The "Logistic Curves Model"

If the observations are bounded, such as percentages or scores to a test, the measurements can be normalized to produce new observations in the open interval $M =]0,1[$. We considered that this open interval of the real line was equipped with the Riemannian metric $g = (g_p)_{p\in]0,1[}$ where : $\forall p \in \mathbb{M} =]0,1[$, $\forall (u,v) \in T_p\mathbb{M} \times T_p\mathbb{M}$, $g_p(u,v) = uM(p)v$, where $M(p) = 1/\big(p^2(1-p)^2\big)$. This Riemannian metric on $]0,1[$ was obtained as the push-forward of the Euclidean metric on $\mathbb{R}$ by the logit transform. In Schiratti et al. (2015c), it was proven that $\mathbb{M} =]0,1[$ is a geodesically complete Riemannian manifold and that the generic spatiotemporal model writes:

$$y_{i,j} = \left(1 + \left(\frac{1}{p_0} - 1\right)\exp\Big(-\frac{v_0\alpha_i(t_{i,j} - t_0 - \tau_i)}{p_0(1-p_0)}\Big)\right)^{-1} + \varepsilon_{i,j}. \tag{7}$$

In this framework, the Riemannian logarithm at $p = 1/2$, which corresponds to the inflexion point of the logistics, is given by: $\forall q \in]0,1[$, $\mathrm{Log}_{1/2}(q) = (1/4)\mathrm{logit}(q)$. However, in (7), the point p_0 is not fixed to $1/2$, but is estimated as a fixed effect. The model estimates the p_0, and therefore the best tangent space, which best describes the observations. Furthermore, even if one fixes $p = 1/2$, the model lifted up on the tangent space remains nonlinear due to the multiplication between the random effects α_i and τ_i. Therefore, the logistic curves model is not equivalent to a linear model on the logit transform of the observations.

3.3 A Progression Model

The generic spatiotemporal model can be used to study the temporal progression of a family of features which characterize the evolution of a biological phenomenon. We assumed that each feature is described by repeated *univariate* observations, which are random perturbations of quantities lying in a one-dimensional geodesically complete Riemannian manifold (M, g^M), open subset of $\mathbb{R}$. For each individual, at each time point, the observations $(\mathbf{y}_{i,j})_{1\leq i\leq p,\, 1\leq j\leq k_i}$ consisted of a N-dimensional vector of univariate features. Hence, for this progression model, the observations $(\mathbf{y}_{i,j})_{1\leq i\leq p,\, 1\leq j\leq k_i}$ were considered as random perturbations of quantities which belong to the product manifold $\mathbb{M} = M \times \ldots \times M = M^N$. Since each Riemannian manifold (M, g^M) is geodesically complete, $\mathbb{M}$ equipped with the product metric is also geodesically complete.

On the product manifold $\mathbb{M} = M^N$, equipped with the product metric, a geodesic is of the form $t \mapsto \big(\gamma_1(t), \ldots, \gamma_N(t)\big)$, where $\gamma_1, \ldots, \gamma_N$ are geodesics of the one-dimensional Riemannian manifold M. Because we would like to model the joint temporal progression of N features, we proposed to choose the group-average trajectory among a parametric family of geodesics of $\mathbb{M}$. This family was of the form:

$$\left\{\boldsymbol{\gamma}_{0,\boldsymbol{\delta}} : t \in \mathbb{R} \mapsto \big(\gamma_0(t), \gamma_0(t+\delta_1), \ldots, \gamma_0(t+\delta_{N-1})\big)\right\}, \tag{8}$$

with $\boldsymbol{\delta} = (0, \delta_1, \ldots, \delta_{N-1})^\top$, $\delta_i \in \mathbb{R}$ and γ_0 denoted a geodesic of the one-dimensional Riemannian manifold g^M which goes through a point $p_0 \in M$ at time t_0 with velocity v_0. The relative delay between two consecutive biomarkers was given by the parameters δ_i $(1 \leq i \leq N-1)$. The vector $\boldsymbol{\delta}$ was to be estimated as a fixed effect of the model. The first component of the vector $\boldsymbol{\delta}$ was chosen to be equal to zero to ensure the identifiability of the model. Note that assuming that the group average belongs to this parametric family of geodesics is equivalent to assuming that the progression of each feature is described by trajectories which have the same shape but are shifted in time.

Lemma 2 *Let $\boldsymbol{\gamma}$ be a geodesic of the product manifold $\mathbb{M} = M^N$ and let $t_0 \in \mathbb{R}$. If $\eta^{\mathbf{w}}(\boldsymbol{\gamma}, \cdot)$ denotes an exp-parallelization of the geodesic $\boldsymbol{\gamma}$ using $\mathbf{w} = (w_1, \ldots, w_N) \in \mathrm{T}_{\boldsymbol{\gamma}(t_0)}\mathbb{M}$ and with $\boldsymbol{\gamma}(t) = (\gamma_1(t), \ldots, \gamma_N(t))$, we have $\eta^{\mathbf{w}}(\boldsymbol{\gamma}, s) = \big(\gamma_1\big(\frac{w_1}{\dot{\gamma}_1(t_0)} + s\big), \ldots, \gamma_N\big(\frac{w_N}{\dot{\gamma}_N(t_0)} + s\big)\big)$, $s \in \mathbb{R}$.*

In this framework, an exp-parallelization of the group-average trajectory $\boldsymbol{\gamma}_{0,\boldsymbol{\delta}}$ can be computed using the result given in Lemma 2. Indeed, it follows from Lemma 2 that the generic spatiotemporal model 2 writes :

$$y_{i,j,k} = \gamma_0\left(\frac{w_{i,k}}{\dot{\gamma}_0(t_0+\delta_{k-1})} + \psi_i(t_{i,j}) + \delta_{k-1}\right) + \varepsilon_{i,j,k}, \tag{9}$$

where, for all $k \in \{1, \ldots, N\}$, $(\mathbf{y}_{i,j})_k$ denotes the kth component of $\mathbf{y}_{i,j}$. In other words, $(\mathbf{y}_{i,j})_k$ is the observation associated to the kth biomarker, for the ith individual, at the jth time point. Similarly, $(\mathbf{w}_i)_k$ denotes the kth component of the space shift $\mathbf{w}_i$. For all $i \in \{1, \ldots, p\}$, $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ is the individual specific time reparametrization. This model is referred to as the *progression model*. For this model, the latent variables are: $\mathbf{z}_{\text{pop}} = \big(p_0, t_0, v_0, (\delta_k)_{1\leq k\leq N-1}, (\beta_{l,k})_{l,k}\big)$ and, for all $i \in \{1, \ldots, p\}$, $\mathbf{z}_i = \big(\xi_i, \tau_i, (s_{l,i})_{l,i}\big)$. The

definition of the individual latent variables $(\mathbf{z}_i)_{1\leq i\leq p}$ remains unchanged. For the population latent variables $\mathbf{z}_{\text{pop}}$, the variables $(\delta_k)_{1\leq k\leq N-1}$ were added. We assumed that the latent variables $\mathbf{z}_{\text{pop}}$ were distributed as follows:

$$p_0 \sim \mathcal{N}(\overline{p_0}, \sigma^2_{p_0}),\ t_0 \sim \mathcal{N}(\overline{t_0}, \sigma^2_{t_0}),\ v_0 \sim \mathcal{N}(\overline{v_0}, \sigma^2_{v_0}), \tag{10}$$

and

$$\beta_{l,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\overline{\beta}_{l,k}, \sigma^2_\beta),\ \delta_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\overline{\delta}_k, \sigma^2_\delta), \tag{11}$$

where $\sigma^2_{p_0}$, $\sigma^2_{t_0}$, $\sigma^2_{v_0}$ and σ^2_δ are *fixed* variance parameters. Similarly to the generic spatiotemporal model, the latent variables were assumed independent of each other and independent of the noise variables $\boldsymbol{\varepsilon}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \mathrm{I}_N)$.

3.4 The Symmetric Positive Definite (SPD) Matrices Model

In this section, we described how the generic spatiotemporal model can be used to analyze longitudinal datasets of symmetric positive definite matrices. Such datasets may arise in Diffusion Tensor Imaging (DTI) or when observing the temporal evolution of covariance matrices.

The space of 3×3 symmetric positive definite matrices is usually denoted by SDP(3), which is an open subset of the vector space of $(3, 3)$ symmetric real matrices, denoted by Sym(3). By identifying Sym(3) with $\mathbb{R}^6$, $\mathbb{M} = \text{SDP}(3)$ can be considered as an open submanifold of $\mathbb{R}^6$. Indeed, $\mathbb{M}$ can be equipped with a Riemannian metric. In Pennec et al. (2006), the authors defined an *affine-invariant* Riemannian metric on SDP(3). Equipped with this metric, the space of symmetric positive definite matrices is a geodesically complete Riemannian manifold, without boundaries (matrices with null eigenvalues are at infinity). The results presented below were obtained with the affine-invariant metric on SDP(3).

If $\boldsymbol{\Sigma} \in \text{SDP}(3)$, $\boldsymbol{\Sigma}^{1/2}$ denotes the unique symmetric positive definite matrix $\mathbf{S}$ such that : $\mathbf{S}^2 = \boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1/2}$ denotes its inverse. Let $\mathrm{T}_{\boldsymbol{\Sigma}}\mathbb{M}$ denote the tangent space to $\mathbb{M}$ at the point $\boldsymbol{\Sigma}$. $\mathrm{T}_{\boldsymbol{\Sigma}}\mathbb{M}$ can be identified with Sym(3) and is equipped with the inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{\Sigma}}$ defined by : $\forall (\mathbf{W}_1, \mathbf{W}_2) \in \mathrm{T}_{\boldsymbol{\Sigma}}\mathbb{M}$, $\langle \mathbf{W}_1, \mathbf{W}_2 \rangle_{\boldsymbol{\Sigma}} = \mathrm{tr}\big(\boldsymbol{\Sigma}^{-1/2}\mathbf{W}_1^\top \boldsymbol{\Sigma}^{-1}\mathbf{W}_2\boldsymbol{\Sigma}^{-1/2}\big)$.

In order to describe exp-parallelization in the Riemannian manifold SPD(n) equipped with the affine invariant metric, we gave a closed-form expression of the parallel transport. The result given in Lemma 3 is based in the work of Lenglet et al. (2006). For $\mathbf{P}_0 \in \text{SPD}(n)$, $t_0 \in \mathbb{R}$, $\mathbf{V}_0 \in \mathrm{T}_{\mathbf{P}_0}\text{SPD}(n) \simeq \text{SYM}(n)$, the geodesic $\boldsymbol{\gamma}_0$, defined by $\boldsymbol{\gamma}_0(t) = \mathrm{Exp}_{\mathbf{P}_0,t_0}(\mathbf{V}_0)(t)$ for all $t \in \mathbb{R}$, is given by : $\boldsymbol{\gamma}_0(t) = \mathbf{P}_0^{1/2} \exp\big(t\mathbf{P}_0^{-1/2}\mathbf{V}_0\mathbf{P}_0^{-1/2}\big)\mathbf{P}_0^{1/2}$, where $\mathbf{P}_0^{1/2}$ (respectively $\mathbf{P}_0^{-1/2}$) denotes the unique symmetric positive definite square root of $\mathbf{P}_0$ (respectively its inverse). The proof of Lemma 3 can be adapted to obtain the expression of the geodesics. One can also obtain this expression by noting that the geodesic starting at I_n with velocity $\mathbf{V} \in \text{SYM}(n)$ is given by $\exp(t\mathbf{V})$ and use the invariance of the affine-invariant metric under congruent transformations. Finally, the expression of the parallel transport along such a geodesic is given by the following lemma.

Lemma 3 *Let* $\mathbf{P}_0 \in \mathrm{SPD}(n)$, $t_0 \in \mathbb{R}$ *and* $\mathbf{V}_0 \in \mathrm{T}_{\mathbf{P}_0}\mathrm{SPD}(n) \simeq \mathrm{SYM}(n)$. *Let* $\boldsymbol{\gamma}_0$ *be the geodesic defined as above. If* $\mathbf{W}$ *is a tangent vector in* $\mathrm{T}_{\mathbf{P}_0}\mathrm{SPD}(n)$, *the parallel transport* $\mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t}(\mathbf{W})$ *is given by:*

$$\forall t \in \mathbb{R},\ \mathrm{P}_{\boldsymbol{\gamma}_0,t_0,t}(\mathbf{W}) = \exp\Big(\frac{t-t_0}{2}\mathbf{V}_0\mathbf{P}_0^{-1}\Big)\mathbf{W}\exp\Big(\frac{t-t_0}{2}\mathbf{P}_0^{-1}\mathbf{V}_0\Big). \tag{12}$$

The proof of Lemma 3 is given in Appendix A.2. It follows from this lemma that the generic spatiotemporal model writes :

$$\mathbf{Y}_{i,j} = \mathbf{P}_i(t_{i,j})^{1/2}\exp\big(\mathbf{P}_i(t_{i,j})^{-1/2}\mathbf{V}_i(t_{i,j})\mathbf{P}_i(t_{i,j})^{-1/2}\big)\mathbf{P}_i(t)^{1/2} + \boldsymbol{\varepsilon}_{i,j}, \tag{13}$$

with, for all $t \in \mathbb{R}$,

$$\mathbf{P}_i(t) = \mathbf{P}_0^{1/2}\exp\big(\alpha_i(t-t_0-\tau_i)\mathbf{P}_0^{-1/2}\mathbf{V}_0\mathbf{P}_0^{-1/2}\big)\mathbf{P}_0^{1/2}, \tag{14}$$

and:

$$\mathbf{V}_i(t) = \exp\Big(\frac{\alpha_i(t-t_0-\tau_i)}{2}\mathbf{V}_0\mathbf{P}_0^{-1}\Big)\mathbf{W}_i\exp\Big(\frac{\alpha_i(t-t_0-\tau_i)}{2}\mathbf{P}_0^{-1}\mathbf{V}_0\Big). \tag{15}$$

The prior distribution for the matrices $\mathbf{P}_0$, $\mathbf{V}_0$ and $(\boldsymbol{\varepsilon}_{i,j})_{i,j}$ are defined as follows: $\mathbf{P}_0 \sim \mathcal{SN}(\overline{\mathbf{P}_0}, \sigma^2_{P_0})$, $\mathbf{V}_0 \sim \mathcal{SN}(\overline{\mathbf{V}_0}, \sigma^2_{V_0})$ and $\boldsymbol{\varepsilon}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{SN}(\mathbf{0}, \sigma^2)$, where $\mathcal{SN}$ denotes the Gaussian distribution on the vector space $\mathrm{Sym}(n)$. Given $\overline{\mathbf{M}} \in \mathrm{SYM}(n)$, the probability distribution $\mathcal{SN}(\overline{\mathbf{M}}, \sigma^2)$ on $\mathrm{Sym}(n)$ is defined by the density function q such that: $q(\mathbf{M}) = \frac{1}{(2\pi)^{m/2}\sigma^m}\exp\big(-\frac{1}{2\sigma^2}\mathrm{tr}\big[(\overline{\mathbf{M}}-\mathbf{M})^2\big]\big)$ with $\mathbf{M} \in \mathrm{Sym}(n)$ and with $m = n(n+1)/2$. The "standard" distribution $\mathcal{SN}(\mathbf{0}, 1)$ is used in physics and in the theory of random matrices. It is sometimes called *Gaussian Orthogonal Ensemble*. The probability distribution of the other random effects of the model were defined as in Section 2.2. This model will be referred to as the *symmetric positive definite matrices model* or $\mathrm{SPD}(n)$ *matrices model*.

In Arsigny et al. (2006), the authors considered the space $\mathrm{SPD}(n)$ equipped with the log-Euclidean metric. This metric provides the space of symmetric positive definite matrices with a structure of Riemannian manifold. Unlike with the affine-invariant metric, the space $\mathrm{SPD}(n)$ endowed with the Log-Euclidean metric is a *flat* Riemannian manifold, meaning that its sectional curvature is null everywhere. By contrast, the space $\mathrm{SPD}(n)$ equipped with the affine-invariant metric is a Riemannian manifold of non-positive curvature (Skovgaard, 1984; Moakher and Zéraï, 2011) with no cut-locus. Within the Log-Euclidean framework, the geodesics are of the form: $\exp(\mathbf{V}_1 + t\mathbf{V}_2)$ with $\mathbf{V}_1, \mathbf{V}_2 \in \mathrm{SYM}(n)$. As expected, the geodesics are the image of a straight line in $\mathrm{SYM}(n)$ by the matrix exponential map. Future developments should include comparisons with the log-euclidean metric on the space SPD(3).

4. Parameters Estimation

The generic spatiotemporal model (2) is a nonlinear mixed-effects model for which the *observed* likelihood is not available in closed-form. Indeed, it writes as an intractable integral

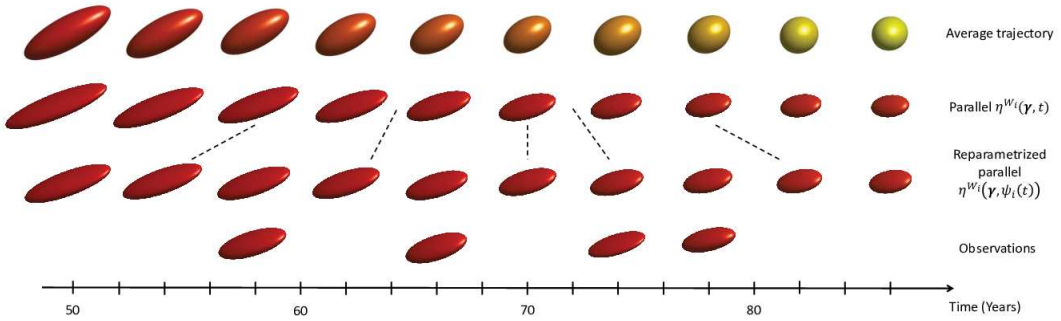


Figure 4: Simulated evolutions of diffusion tensors. First row : average trajectory from a highly anisotropic diffusion tensor to a sphere. Second row : a parallel to the average trajectory obtained using a random space shift $\mathbf{W}_i$. Third row : the reparametrization of the exp-parallelization with $\alpha_i = 0.7$ and $\tau_i = -4$ (years). Fourth row : the observations are samples from the reparametrized exp-parallelization. The samples are obtained using the noise model described in 3.4. Each diffusion tensor is colored according to its fractional anisotropy (red for a highly anisotropic tensors, yellow for a sphere).

which could only be approximated. In order to produce maximum likelihood estimates, we could use the Expectation Maximization (EM) algorithm (Dempster et al., 1977). The first step of the EM algorithm, usually called "E-step", requires us to compute the expectation of the log-complete likelihood (the likelihood of the observations $\mathbf{y} = (\mathbf{y}_{i,j})_{1 \leq i \leq p,\, 1 \leq j \leq k_i}$ and the latent variables $\mathbf{z}$) with respect to the conditional distribution of the latent variables knowing the observations and the current values of the parameters. In the case of our model, this expectation cannot be computed in closed-form.

Therefore, we chose to estimate the parameters of the generic spatiotemporal model by using a stochastic version of the EM algorithm, in which this step was replaced by a stochastic approximation. This algorithm is the Monte Carlo Markov Chains (MCMC) Stochastic Approximation EM (MCMC-SAEM) algorithm (Allassonnière et al., 2010). The MCMC-SAEM iterates, until convergence, between three steps: simulation, stochastic approximation and maximization.

4.1 Simulation Step

If $\tilde{\boldsymbol{\theta}}^{(k-1)}$ denotes the current estimate of $\tilde{\boldsymbol{\theta}}$ at the beginning of the kth iteration of the MCMC-SAEM, the simulation step consists of drawing a sample $\mathbf{z}^{(k)}$ from the transition kernel $\boldsymbol{\pi}_{\boldsymbol{\theta}^{(k-1)},\mathbf{y}}(\mathbf{z}^{(k-1),\cdot})$ of an ergodic Markov chain whose stationary distribution is the conditional distribution $q(\mathbf{z} \mid \mathbf{y}, \tilde{\boldsymbol{\theta}}^{(k-1)})$, the distribution of the latent variables $\mathbf{z}$ knowing the observations $\mathbf{y}$ and $\tilde{\boldsymbol{\theta}}^{(k-1)}$. This step is achieved using a Monte Carlo Markov Chain (MCMC) sampler. We chose to use a Block Metropolis-Hastings-within-Gibbs (Block MHwG) sampler for the sampling step of the MCMC-SAEM. Each Metropolis-Hastings step of the algorithm consists in a multivariate symmetric random walk. The Block MHwG sampler updates simultaneously *block* (or sets) of latent variables then, at each iteration, each block is updated

conditionally on the others. Even though the latent variables can be grouped in several ways, we chose to group the latent variables as follows: $\{\mathbf{z}_{\text{pop}}\}$ and $\{\mathbf{z}_i\}_{1\leq i\leq p}$. This grouping being given by the hierarchical structure of the model. Note that the latent variables also could have been grouped as follows: $\{\mathbf{p}_0, t_0, \mathbf{v}_0\}$, $\{(\beta_{l,k})_{l,k}\}$ and $\{\mathbf{z}_i\}_{1\leq i\leq p}$. In the case of the progression models, the delay variables $(\delta_k)_{1\leq k\leq N-1}$ were grouped with $\mathbf{z}_{\text{pop}}$, although they could also be considered as a block in itself.

For each block, the proposal in the Metropolis-Hastings step was chosen to be a multivariate Gaussian distribution centered at the current state of the block. Each variance-covariance matrix of a proposal distribution was chosen to be diagonal matrix: $\mathbf{D}_{\text{pop}} = \text{Diag}\big(\zeta^2_{\mathbf{p}_0}\text{I}_N, \zeta^2_{t_0}, \zeta^2_{\mathbf{v}_0}\text{I}_N, \zeta^2_{\boldsymbol{\beta}}\text{I}_{(N-1)N_s}\big)$ for the proposal distribution associated to $\mathbf{z}_{\text{pop}}$ and $\mathbf{D}_{\text{indiv}} = \text{Diag}\big(\zeta^2_\xi, \zeta^2_\tau, \zeta^2_s\big)$ for the proposal distribution associated to $\mathbf{z}_i$ $(1 \leq i \leq p)$. The variances parameters $\zeta^2_{\mathbf{p}_0}, \zeta^2_{t_0}, \zeta^2_{\mathbf{v}_0}, \zeta^2_{\boldsymbol{\beta}}$ and $\zeta^2_\xi, \zeta^2_\tau$ were adjusted by hand to ensure an average acceptance rate for each block around 23% (Roberts et al., 1997). The Block MHwG sampler is described in Algorithm 1.

Algorithm 1 The Block Metropolis-Hastings-within-Gibbs sampler

Require: Set of latent variables $\mathbf{z}^{(k-1)} = \big(\mathbf{z}^{(k-1)}_{\text{pop}}, (\mathbf{z}^{(k-1)}_i)_{1\leq i\leq p}\big)$, current estimate of the parameters $\boldsymbol{\theta}^{(k-1)}$, variance-covariance matrices $\mathbf{D}_{\text{pop}}$ and $(\mathbf{D}_i)_{1\leq i\leq p}$ and $\boldsymbol{\theta}_{\text{hyper}}$

Ensure: Set of latent variables $\mathbf{z}^{(k)}$

Block $\mathbf{z}^{(k)}_{\text{pop}}$ of population latent variables:

Draw a candidate $\mathbf{z}^*_{\text{pop}} \sim \mathcal{N}\big(\mathbf{z}^{(k-1)}_{\text{pop}}, \mathbf{D}_{\text{pop}}\big)$

Compute the acceptance ratio $\alpha(\mathbf{z}^{(k-1)}_{\text{pop}}, \mathbf{z}^*_{\text{pop}})$ defined by:

$$\alpha(\mathbf{z}^{(k-1)}_{\text{pop}}, \mathbf{z}^*_{\text{pop}}) = \frac{q(\mathbf{y} \mid \mathbf{z}^*_{\text{pop}}, (\mathbf{z}^{(k-1)}_i)_{1\leq i\leq p}, \boldsymbol{\theta}^{(k-1)}) q_{\text{pop}}(\mathbf{z}^*_{\text{pop}} \mid \boldsymbol{\theta}^{(k-1)})}{q(\mathbf{y} \mid \mathbf{z}^{(k-1)}_{\text{pop}}, (\mathbf{z}^{(k-1)}_i)_{1\leq i\leq p}, \boldsymbol{\theta}^{(k-1)}) q_{\text{pop}}(\mathbf{z}^{(k-1)}_{\text{pop}} \mid \boldsymbol{\theta}^{(k-1)})} \wedge 1.$$

Draw $U \sim \text{Uniform}\big([0,1]\big)$

Set: $\mathbf{z}^{(k)}_{\text{pop}} = \mathbf{z}^*_{\text{pop}}$ if $U \leq \alpha(\mathbf{z}^{(k-1)}_{\text{pop}}, \mathbf{z}^*_{\text{pop}})$ and $\mathbf{z}^{(k)}_{\text{pop}} = \mathbf{z}^{(k-1)}_{\text{pop}}$ otherwise.

for $i = 1 \dots p$ **do**

    Blocks $(\mathbf{z}^{(k)}_i)_{1\leq i\leq p}$ of individual latent variables:

    Draw a candidate $\mathbf{z}^*_i \sim \mathcal{N}\big(\mathbf{z}^{(k-1)}_i, \mathbf{D}_{\text{indiv}}\big)$

    Compute the acceptance ratio $\alpha(\mathbf{z}^{(k-1)}_i, \mathbf{z}^*_i)$ defined by:

$$\alpha(\mathbf{z}^{(k-1)}_i, \mathbf{z}^*_i) = \frac{q(\mathbf{y} \mid \mathbf{z}^{(k)}_{\text{pop}}, \mathbf{z}^{(k-1),(k)}_{-i}, \mathbf{z}^*_i, \boldsymbol{\theta}^{(k-1)}) q_i(\mathbf{z}^*_i \mid \boldsymbol{\theta}^{(k-1)})}{q(\mathbf{y} \mid \mathbf{z}^{(k)}_{\text{pop}}, \mathbf{z}^{(k-1),(k)}_{-i}, \mathbf{z}^{(k-1)}_i, \boldsymbol{\theta}^{(k-1)}) q_i(\mathbf{z}^{(k-1)}_i \mid \boldsymbol{\theta}^{(k-1)})} \wedge 1.$$

    Draw $U \sim \text{Uniform}([0,1])$

    Set: $\mathbf{z}^{(k)}_i = \mathbf{z}^*_i$ if $U \leq \alpha(\mathbf{z}^{(k-1)}_i, \mathbf{z}^*_i)$ or $\mathbf{z}^{(k)}_i = \mathbf{z}^{(k-1)}_i$ otherwise.

end for

Return: $\mathbf{z}^{(k)} = \big(\mathbf{z}^{(k)}_{\text{pop}}, (\mathbf{z}^{(k)}_i)_{1\leq i\leq p}\big)$.

Let $\boldsymbol{\theta}_{\text{hyper}} = (\sigma^2_{\mathbf{p}_0}, \sigma^2_{t_0}, \sigma^2_{\mathbf{v}_0}, \sigma^2_{\boldsymbol{\beta}})$ denote the fixed hyperparameters which appear in the probability distribution of the latent variables in $\mathbf{z}_{\text{pop}}$. Let $i \in \{1, \dots, p\}$ and $q_{\text{pop}}\big(\cdot \mid \boldsymbol{\theta}\big)$ (respectively $q_i(\cdot \mid \boldsymbol{\theta})$ denote the density function of the joint distribution of the latent variables $\mathbf{z}_{\text{pop}}$ (respectively $\mathbf{z}_i$) as specified in the generative model (equations (3) and (4)):

$$\begin{aligned} q_{\text{pop}}(\mathbf{z}_{\text{pop}} \mid \boldsymbol{\theta}) \propto \exp\Big(-\frac{1}{2\sigma^2_{\mathbf{p}_0}}\|\mathbf{p}_0 - \overline{\mathbf{p}_0}\|^2\Big)\exp\Big(-\frac{1}{2\sigma^2_{t_0}}(t_0 - \overline{t_0})^2\Big) \\ \exp\Big(-\frac{1}{2\sigma^2_{\mathbf{v}_0}}\|\mathbf{v}_0 - \overline{\mathbf{v}_0}\|^2\Big)\exp\Big(-\frac{1}{2\sigma^2_{\boldsymbol{\beta}}}\|\boldsymbol{\beta} - \overline{\boldsymbol{\beta}}\|^2\Big), \end{aligned} \tag{16}$$

and

$$q_i(\mathbf{z}_i \mid \boldsymbol{\theta}) \propto \exp\Big(-\frac{1}{2\sigma^2_{\xi}}\xi_i^2\Big)\exp\Big(-\frac{1}{2\sigma^2_{\tau}}\tau_i^2\Big)\exp\Big(-\frac{1}{2}\|\mathbf{s}_i\|^2\Big), \tag{17}$$

with: $\boldsymbol{\beta} = [\beta_{l,k}]_{1\leq l\leq N_s,\, 1\leq k\leq N-1}$ and for all $i \in \{1, \dots, p\}$, $\mathbf{s}_i = [s_{l,i}]_{1\leq l\leq N_s}$. The probability distributions q_{pop} and q_i $(1 \leq i \leq p)$ are given up to a constant. Indeed, the normalizing constant of q_{pop} or q_i $(1 \leq i \leq p)$ depends only on the parameters $\boldsymbol{\theta}$. Therefore, these constants can be omitted for the computation of the acceptance ratio in Algorithm 1.

4.1.1 Discussion

In order to avoid tuning by hand the parameters $\zeta^2_{\mathbf{p}_0}$, $\zeta^2_{t_0}$, $\zeta^2_{\mathbf{v}_0}$ and $\zeta^2_{\boldsymbol{\beta}}$ of the proposal distribution in the Block MHwG sampler, a possible solution would consist of using an *adaptive* (Atchadé, 2006) version of the Block MHwG sampler, where the algorithm automatically adjusts these variance parameters.

4.2 Stochastic Approximation

The convergence of the MCMC-SAEM was proven, in Kuhn and Lavielle (2004) (for bounded latent variables) and in Allassonnière et al. (2010) (for unbounded latent variables), for statistical models which belong to the *curved exponential family*. That is to say, models for which the log complete likelihood $q(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta})$ writes: $\forall \boldsymbol{\theta} \in \Theta$, $\log q(\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) = -\Phi(\boldsymbol{\theta}) + \langle \mathbf{S}(\mathbf{y}, \mathbf{z}), \Psi(\boldsymbol{\theta})\rangle$, where Φ, Ψ are smooth functions of the parameters, $\mathbf{S}(\mathbf{y}, \mathbf{z})$ is a measurable function of the observations and latent variables called *sufficient statistic of the model* and $\langle \cdot, \cdot \rangle$ is an inner product on a product space. The generic spatiotemporal model belongs to the curved exponential family.

Because the generic spatiotemporal model belongs to the curved exponential family, the stochastic approximation can be done on the sufficient statistics of the model. At the kth iteration of the MCMC-SAEM, we have: $\mathbf{S}_k = \mathbf{S}_{k-1} + \varepsilon_k\big(\mathbf{S}(\mathbf{y}, \mathbf{z}^{(k)}) - \mathbf{S}_{k-1}\big)$, where $(\varepsilon_k)_{k\geq 1}$ is a sequence of positive step sizes such that $\sum_k \varepsilon_k = +\infty$ and $\sum_k \varepsilon_k^2 < +\infty$. If $\varepsilon_k = 1$, then $\mathbf{S}_k$ does not depend on $\mathbf{S}_{k-1}$. Intuitively, the sequence $(\mathbf{S}_k)_{k\geq 0}$ has "no memory" as long as $\varepsilon_k = 1$ and the MCMC-SAEM freely explores the parameters space during this period. In practice, we chose $\varepsilon_k = 1$ as long as $k \leq N_b$ and $\varepsilon_k = (k - N_b)^{-0.65}$ if $k > N_b$.

4.2.1 Maximization Step

The maximization step consists of solving the following optimization problem : $\boldsymbol{\theta}^{(k)} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}}\,\big(-\Phi(\boldsymbol{\theta}) + \langle \mathbf{S}_k, \Psi(\boldsymbol{\theta})\rangle\big)$, where $\mathbf{S}_k$ denotes the stochastic approximation on the suffi-

cient statistics of the model, obtained in the "stochastic approximation step" of the algorithm. For the generic spatiotemporal model, this optimization problem was solved in closed-form.

4.3 Computational Aspects

For the progression model (9), the MCMC-SAEM would have to estimate the following parameters : $\boldsymbol{\theta} = (\overline{p_0}, \overline{t_0}, \overline{v_0}, \overline{\delta}_1, \ldots, \overline{\delta}_{N-1}, \overline{\beta}_1, \ldots, \overline{\beta}_{(N-1)N_s}, \sigma_\eta, \sigma_\tau, \sigma)$. In this example, we saw that the number of parameters to estimate was $6 + (N-1)(N_s+1)$. As the dimension N of the manifold $\mathbb{M}$ increases, the number of parameters increased linearly. Moreover, as N increased, the number N_s of independent sources had a greater impact on the number of parameters to estimate.

The number p of individuls also impacted on the runtime of the MCMC-SAEM. As the number p of individuals increased, the cost of a single computation of the observed likelihood increased. This step is the most expensive step of the MCMC-SAEM algorithm. The overall runtime of the MCMC-SAEM could be improved by sampling the blocks $(\mathbf{z}_i)_{1 \leq i \leq p}$ (in the Block MHwG sampler) in parallel.

5. Experiments

The section on numerical experiments begins with a comparison of our implementation of the MCMC-SAEM algorithm with other state-of-the-art algorithms. It is followed by a section entitled "Validation Procedure" which presents an evaluation criteria designed to quantify how well the time reparametrizations of the generic spatiotemporal model allowed to temporally align the progression of individuals. Finally, results obtained on a synthetic dataset of Symmetric Positive Definite (SPD) matrices and on a real dataset of neuropsychological tests scores are given. These datasets were analyzed using the particular cases of the generic spatiotemporal model described in Section 3.

5.1 Convergence of the Algorithm

In this section, we aimed to compare our implementation of the MCMC-SAEM with state of the art algorithms. Our algorithm, implemented in MATLAB, was compared with STAN and MONOLIX. STAN is a R/C++ library which implements an adaptive Hamiltonian Monte Carlo sampler called "the No U-Turns Sampler" (NUTS, Hoffman and Gelman (2014)).
MONOLIX is software developed by Marc Lavielle and the Lixoft company. It implements the MCMC-SAEM algorithm with some technical improvements (such as a simulated annealing scheme). Note that our implementation of the MCMC-SAEM algorithm differed from MONOLIX in the sense that it can be used with any particular case of the generic spatiotemporal model presented in Section 3. In particular, our implementation can be used to analyze univariate, as well as multivariate (such as covariance matrices) longitudinal observations. In its current version ($2016R1$), MONOLIX does not allow observations to be passed as matrices or vectors, which is not convenient for the analysis multivariate longitudinal observations.

In order to compare these algorithms, we considered a synthetic longitudinal dataset of observations in $]0, 1[$. The open set $]0, 1[$ was equipped with the Riemannian metric defined in Section 3.2, which generated logistic shaped geodesics. This dataset was generated for

Table 1: Relative error on the parameters estimated with the different algorithms. First row: results obtained with our implementation of the MCMC-SAEM. Second row: results obtained with STAN. Third row: results obtained with MONOLIX.

$\lvert\widehat{p_0} - p_0^*\rvert/p_0^*$	$\lvert\widehat{t_0} - t_0^*\rvert/t_0^*$	$\lvert\widehat{v_0} - v_0^*\rvert/v_0^*$	$\lvert\widehat{\sigma_\xi} - \sigma_\xi^*\rvert/\sigma_\xi^*$	$\lvert\widehat{\sigma_\tau} - \sigma_\tau^*\rvert/\sigma_\tau^*$	$\lvert\widehat{\sigma} - \sigma^*\rvert/\sigma^*$
0.0150	0.0050	0.0176	0.0600	0.0545	0.010
0.0917	0.0191	0.1088	0.0600	0.0386	0.010
0.0417	0.0086	0.0412	0.0400	0.0286	0.008

Table 2: Number of iterations and runtimes corresponding to the experimental results given in Table 1. First row: results obtained with our implementation of the MCMC-SAEM. Second row: results obtained with STAN. Third row: results obtained with MONOLIX.

Iterations	Runtime (for 1000 iterations)	Overall runtime
3500	30 s	90 s
15000	15 min	$\simeq 3,75$ h
400	110 s	45 s

$p = 250$ individuals, with an average of 5 time points per individual. Each algorithm was run with the same initialization and the logistic curves model (7) was used to generate the data.

The experimental results given in Table 1 consist of relative errors on the parameters estimated with the different algorithms. The runtime and number of iteration needed for each method to converge are reported in Table 2. The results presented in the first table show that all the different methods succeeded in estimating the parameters which were used to generate the data. The corresponding number of iterations show that STAN is, by far, the most computationally intensive method. Even though our implementation of the MCMC-SAEM required more iterations to converge than MONOLIX, the overall runtime was similar. The fact that MONOLIX required fewer iterations to converge can be explained by the fact that the MCMC-SAEM was coupled with a simulated annealing procedure Lavielle and Mentré (2007), which allowed for a better and faster exploration of the parameters space.

The results presented in these tables show that the performance of our implementation of the MCMC-SAEM is similar to one of state-of-the-art methods. Still, the number of iterations needed to converge could be further reduced, for example, by combining several MCMC samplers in the sampling step of the MCMC-SAEM.

5.2 Validation Procedure

Here we proposed a validation procedure to see if the model is a fit for the data in which the optimal values of the parameters are not known.

The idea was to assess how well the estimated individual time reparametrizations put in correspondence the progression of the individuals. To this end, we would use additional information which was not used in the model: the time at which a particular event occurred in the life of an individual. For example, the event could be the time at which a disease is diagnosed, or at which a metabolic change occurred, for instance. Such an event occured at a different time point (or age) for each individual.

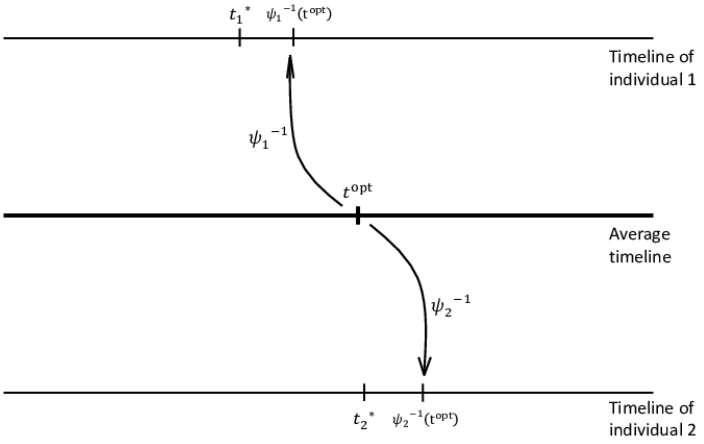


Figure 5: The average time of event t^{opt} was mapped to the individual timelines using ψ_i^{-1}.

The individual time reparametrization ψ_i are supposed to precisely put into correspondence the time at which similar spatiotemporal patterns were found in the individual data. To assess how well this was achieved, we tested whether the time at which a particular event occured in the life of the individuals was mapped to the same time-point in the average trajectory of the model.

For the ith individual, ψ_i mapped the timeline of this individual to the "average timeline", namely the one of the average trajectory. Let t_i^* be the time point at which the event occured in the timeline of the ith individual. We estimated the time-point t^{opt} that corresponded to the occurrence of the event in the average trajectory $\boldsymbol{\gamma}_0$ by minimizing the sum of errors $E(t) = \sum_i |t_i^* - \psi_i^{-1}(t)|$. Note that t^{opt} can be interpreted as a median of the normalized ages $(\psi_i(t_i^*))_i$, and could therefore not be unique. Then we mapped t^{opt} back to the individual timelines by using the mappings ψ_i^{-1}, as illustrated in Fig. 5. The value $\psi_i^{-1}(t^{\text{opt}})$ may be thought of as a prediction of the model of the time-point (or age) at which the event occured for the i*th* individual. Without errors, this time-point would be exactly t_i^*. In practice, the difference $|t_i^* - \psi_i^{-1}(t^{\text{opt}})|$ allowed us to quantify how well the events the timeline of the ith individual and the average timeline have been put into correspondence.

In the following experiments, the median t^{opt} of $\left(\psi_i(t_i^*)\right)_{1\leq i\leq p}$ was computed unambiguously. To assess how well the individual trajectories and the average trajectory are put into correspondence, we plotted a histogram of the errors $\left(|t_i^* - \psi_i^{-1}(t^{\text{opt}})|\right)_{1\leq i\leq p}$.

5.3 Tensors : a Synthetic Dataset

We started by presenting the synthetic dataset of SPD matrices used for this experiment and then presented the results obtained with the particular case of the generic spatiotemporal model for SPD matrices (see Section 3.4).

5.3.1 Data

We considered a synthetic dataset, in which we simulated repeated observations of a a symmetric definite positive matrix (also called a diffusion tensor in medical imaging) for one hundred individuals. The observations were not generated from the model. The observations were obtained instead by prescribing an adhoc hierarchical model on the eigenvalues of the diffusion tensors. At the level of the population, the eigenvalues of the diffusion tensors followed a decreasing piecewise linear evolution with a change point at 50 years old. Observations for a given individual were simulated by randomly shifting the change point (time at which the change occurred in the speed at which eigenvalues decrease) and randomly increasing or decreasing the slopes of each eigenvalue (see Fig.6, left). In this synthetic dataset, the individuals had, on average, five time points.

5.3.2 Results

The results presented below were obtained with $N_s = 1$ source. A greater number of independent sources would have been possible but many more iterations would have been necessary for the MCMC-SAEM to converge. The Bayesian tensor model with the MCMC-SAEM allowed us to estimate an average trajectory of progression in the space SDP(3). This average trajectory was the geodesic which goes through the point $\overline{\mathbf{P}_0}$, at time $\overline{t_0}$, with velocity $\overline{\mathbf{V}_0}$, given by :

$$\overline{\mathbf{P}_0} = \begin{pmatrix} 11.30 & 0.96 & 0.68 \\ 0.96 & 9.53 & 1.21 \\ 0.68 & 1.21 & 10.19 \end{pmatrix}, \quad \overline{t_0} = 53.83 \text{ years},$$

and

$$\overline{\mathbf{V}_0} = \begin{pmatrix} -0.99 & -0.17 & -0.20 \\ -0.17 & -0.75 & -0.27 \\ -0.20 & -0.27 & -0.85 \end{pmatrix} \text{ unit per year.}$$

The evolution of the eigenvalues of the average trajectory, plotted in Fig. 6, was similar to the model used to generate the observations. However, the MCMC-SAEM tended to underestimate the first eigenvalue and overestimate the third eigenvalue. The variability in speed and delay of progression was captured by the estimated parameters $\sigma_\eta = 0.07$ and $\sigma_\tau = 0.5$ year. Fig. 6 (left) shows that the eigenvalues of each individual decreased at a similar pace before and after the change point. This may explain why the model captured small variations in speed of progression. The standard deviation σ_τ of the parameter t_0 is much smaller. The individual acceleration factor, time shift and space shift allowed us to fit the average trajectory to the observations of an individual. As shown in Fig. 6 (right), the estimated individual trajectory was well adjusted to the observations of the individual.

The eigenvalues of the average estimated trajectory were smooth functions of time. Therefore, it would not have been possible to obtain a piecewise-linear progression of the

eigenvalues for the average trajectory. Nevertheless, we could still validate the ability of the tensor model to put into correspondence the dynamic of each individual following Sec. 5.2 by using the individual change point t_i^*. For this dataset, the sum of errors $\sum_i |t_i^* - \psi_i^{-1}(t)|$ had a unique minimum at $t^{\text{opt}} = 49.73$ years. This minimum t^{opt} is close to 50 years, the time at which the change point occurred in the average model used to generate the data. Fig. 7 shows that the model made an error of less than 2 years for almost 60% of the population by predicting the individual change point with $\psi_i^{-1}(t^{\text{opt}})$, and less than 4 years for 90% of the population. The change point was generated using a Gaussian distribution centered at 50 years with a standard deviation of 2 years. Therefore, the error was of the same order as the standard deviation of the change point.

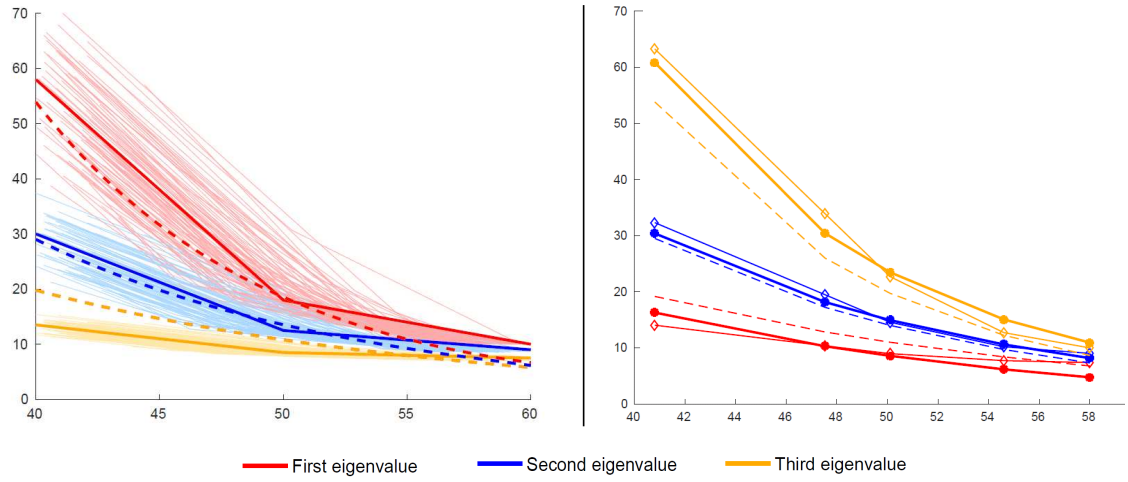


Figure 6: Left : In solid bold lines, the average model of eigenvalues evolution for the synthetic dataset of tensors. In solid lines, the evolution of the eigenvalues for all the individuals in the dataset. In dotted lines, the evolution of the eigenvalues of the average trajectory, given by the MCMC-SAEM. Right : the evolution of the eigenvalues of an individual. In dotted lines, the eigenvalues of the average trajectory estimated by the MCMC-SAEM. With square markers, the eigenvalues of the observations for this individual. With round markers, the eigenvalues of the estimated individual trajectory.

5.4 Neuropsychological Tests

This section is started by presenting the real dataset of neuropsychological tests scores from the ADNI database. Then, we presented results obtained using the progression model (see Eq. (9)) with logistic curves.

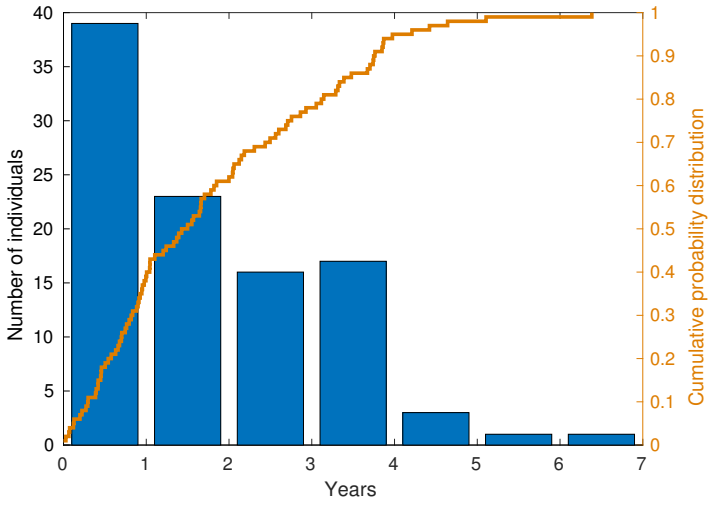


Figure 7: Histogram of $\left(|t_i^* - \psi_i^{-1}(t^{\text{opt}})|\right)_{1\leq i\leq 100}$ superimposed with the cumulative distribution of this error. Here, t_i^* represents the age of the change point for the ith individual.

5.4.1 Data

The dataset consisted of scores for the modified "ADAS-Cog" test (Mohs et al., 1997) obtained from the ADNI1, ADNIGO and ADNI2 cohorts of the Alzheimer's Disease Neuroimaging Initiative. The 13 items were grouped into 4 categories according to the cognitive function they assess: memory, language, concentration, and praxis. For each cognitive function, the scores were added and normalized by the maximum possible value therefore producing measurements in $]0, 1[$. As a consequence, each observation was a point on the manifold $\mathbb{M} =]0, 1[^4$ (note that results without item pooling are presented in Schiratti et al. (2015a)). We used 248 individuals who were included in the study as mild cognitive impaired (MCI) subjects and later converted to Alzheimer's disease (AD). Each individual was observed on average 6 times.

5.4.2 Results

This data set was analyzed using the *progression model* given in Eq. (9) with logistic curves. The number of independent components could be either $1, 2$ or 3, as the manifold is of dimension 4. The model with one independent component estimated a residual noise variance $\sigma^2 = 0.012$ and explained 79% of the total variance. The model with two (resp. three) independent components estimated a noise variance $\sigma^2 = 0.008$ (resp. $\sigma^2 = 0.0084$) and explained 84% (resp. 85%) of the total variance. Because the results obtained with three independent components are similar to the results obtained with two independent components, we chose, for the sake of clarity, to report the results obtained with two components ($N_s = 2$).

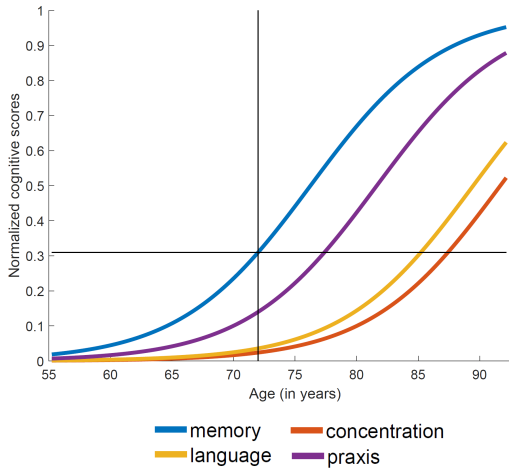


Figure 8: The estimated average trajectory. The estimated parameters p_0 (resp. t_0) are represented by an horizontal (resp. vertical line) at $p_0 = 0.3$ (resp. $t_0 = 72$ years).

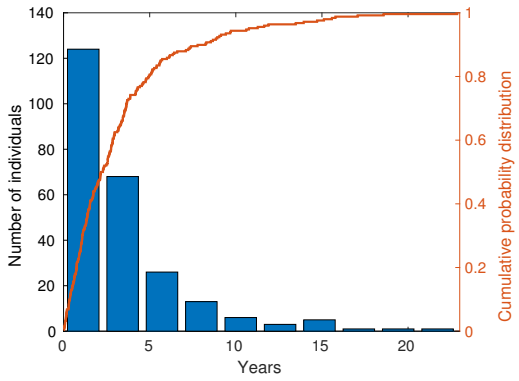


Figure 9: Histogram of $\left(|t_i^{\text{conv}} - \psi_i^{-1}(t^{\text{opt}})|\right)_{1\leq i\leq 248}$ superimposed with the cumulative distribution of this error.

The average trajectory estimated by the progression model, plotted in Fig. 8, is characterized by the fixed effects $p_0 = 0.3$, $t_0 = 72$ years, $v_0 = 0.04$ unit per year and $\boldsymbol{\delta} = [0; -15; -13; -5]$ years. The first biomarker (memory) reached the value $p_0 = 0.3$ at 72 years on average, the second one (concentration), at $t_0 + 5 = 77$ years, followed by praxis and language. The fixed effects provided an ordering of the biomarkers and the relative delay between them. The random effects characterized the spatiotemporal variability of the average trajectory among the population. The estimated standard deviation at the time-shift is $\sigma_\tau = 7.5$ years, meaning that age of disease onset ranged between 72 ± 7.5 years for 95% of the individuals. A positive (resp. negative) time-shift meant that the individual was evolving behing (resp. ahead) the average trajectory. The estimated standard deviation of the acceleration factors was $\sigma_\eta = 0.9$. As a consequence, most of the individuals were progressing between $e^{\sigma_\eta} \simeq 2.4$ times faster or $e^{-\sigma_\eta} \simeq 0.4$ times slower than the average trajectory (see Fig. 11, first row). Estimates of the individual time-shits and log-acceleration factors are plotted in Fig. 10. This figure shows a clear correspondence between the time shifts and the estimated age at which individuals were diagnosed with the disease. This fact shows that the normalized age $\psi_i(t)$ is a better temporal marker of disease progression than age. This was confirmed by our validation procedure (Fig. 9), which showed an error in the prediction of age at diagnosis of less than 2.5 years in 50% of the cases.

In this multivariate setting, random effects also included space shifts, which are a combination of two independent components denoted here $\mathbf{c}_1(\mathbf{A})$ and $\mathbf{c}_2(\mathbf{A})$. As shown in

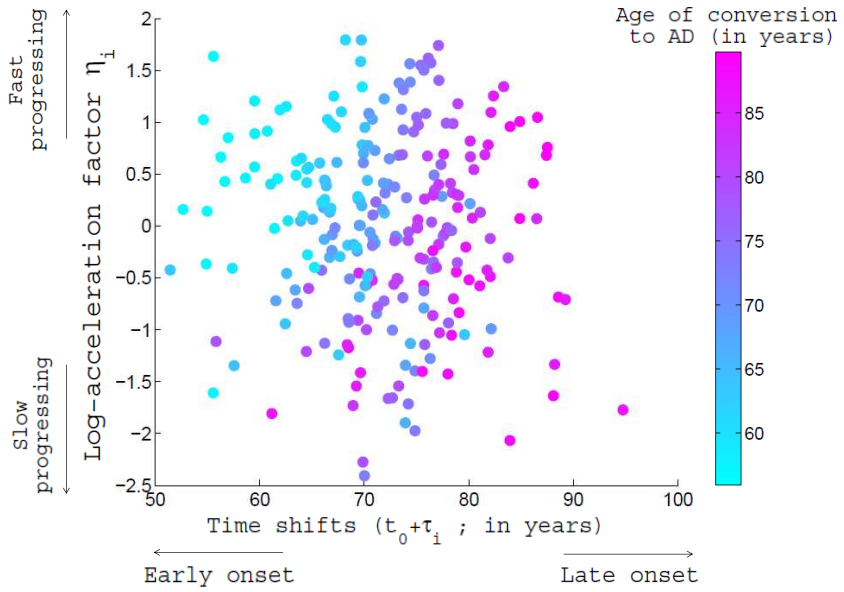


Figure 10: Plot of $t_0 + \tau_i$ with respect to the log-acceleration factor η_i. Each point is colored with respect to the estimated age of conversion to AD.

lemma 2, these space-shifts perturbed the relative delay and the ordering in the progression of biomarkers. Fig. 11 shows that individuals with a space shift of the form $\mathbf{w}_i = \sigma_{s_i}\mathbf{c}_1(\mathbf{A})$ had memory and concentration impaired at nearly the same time, while the language and praxis remained nearly constant. In the opposite direction, impairment in language and praxis nearly coincided for individuals with a space-shift of the form $\mathbf{w}_i = -\sigma_{s_i}\mathbf{c}_1(\mathbf{A})$. The second independent component almost did not change memory and concentration but changed the delay and the ordering between language and praxis. These results showed that the biomarkers tended to evolve in pairs : memory & concentration, language & praxis. Space shifts captured here the variability in the profile of cognitive decline at the individual level during the onset of the disease.

5.5 Body Mass Index in Adolescent Girls

We analyzed a longitudinal dataset of body fat percentages from 162 adolescent girls. This dataset was taken from the MIT Growth and Development Study Bandini et al. (2002); Phillips et al. (2003). The data was analyzed using the univariate logistic model (see supplementary material for the details and results). The analysis of this data in Fitzmaurice et al. (2012) required the use of the time at menarche to temporally align the data across individuals before the statistical analysis. By contrast, our approach was able to include such an alignment as a random effect of the model. The estimated parameters produced an error of less than one year for 50% of the individuals in the alignment of the age at which menarche occured.

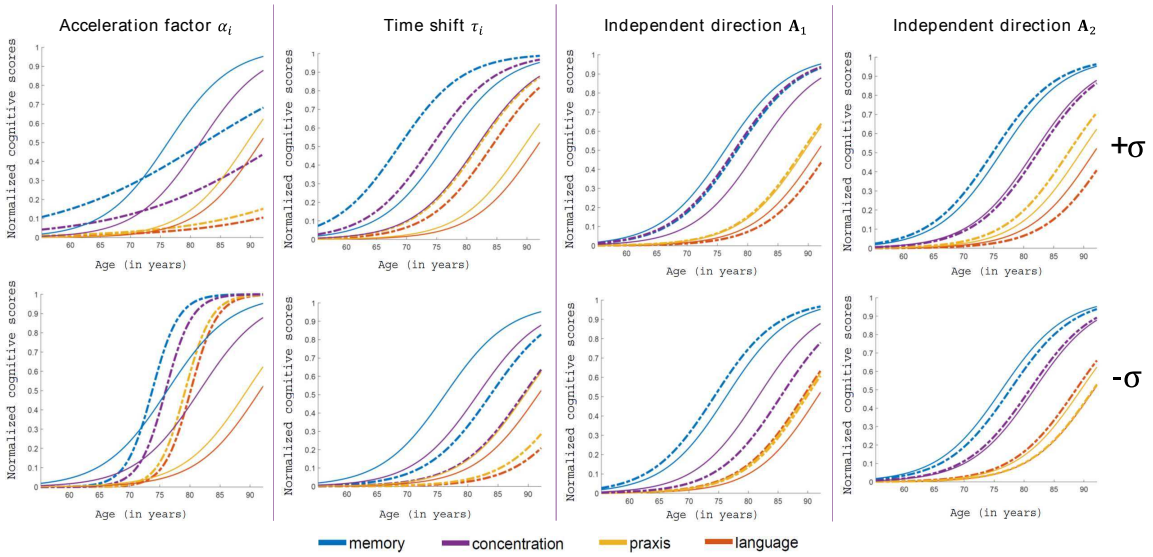


Figure 11: Variability of the average trajectory γ_δ in terms of space shift, time shift and acceleration factor. The solid lines represent the average trajectory, while the dotted lines represent the variability of this average trajectory among the population.

6. Conclusion

We proposed a mixed-effects model for the spatiotemporal analysis of manifold-valued measurements. This generic model can be used to automatically learn the temporal progression of a biological phenomenon from repeated observations of several samples. The model accounts for the fact that each sample had a different appearance, had different trajectory of changes and different pace of changes. This was enabled by the introduction of a hierarchical statistical model whose fixed effects defined a group-average trajectory in the space of measurements and random effects accounted for the spatiotemporal variability of the trajectories of changes at the individual level.

Building the model in the framework of Riemannian geometry allowed us to identify the key orthogonality conditions to uniquely decompose temporal and spatial variability across trajectories of changes. Our particular use of parallel transport ensured the invariance of the form of the distribution of spatial random effects in time. Finally, it allowed the instantiation of the model for a large variety of data types, as shown by our experiments with univariate and multivariate bounded measurements, as well as symmetric definite positive matrices. This framework was also well suited for analysing data with smooth constraints or highly structured data such as images, networks(Koval et al. (2017)), or shapes, for instance.

We relied on the stochastic approximation of the Expectation-Maximization algorithm to estimate model parameters. This algorithm is proven to converge asymptotically, and our experiments confirmed this convergence in several practical situations. This algorithm is very generic and allowed a modular implementation of the method which eases its instanciation

for any kind of manifold-valued observations. Note, nevertheless, that specific optimizations were needed to match, and even outperform state-of-the-art implementations for real-valued measurements.

The model was particularly useful to analyze biological phenomenon for which there is no obvious way to temporally align individual data time series. Temporal re-alignement, introduced as random-effect, allowed us to automatically predict the age at which patients were diagnosed with Alzheimer's disease or at which menarche occured in adolescent girls. Although the model assumed a monotonic progression of body fat with age, this hypothesis could be relaxed by replacing the distribution of the acceleration factors and space shifts with a mixture of Gaussian distributions. This could also be achieved by considering non-affine time reparametrizations of the average trajectory. Improvements in prediction and goodness of fit might result from extensions of the model, for instance by introducing mixture models to identify population clusters or by adding a drift in the parallel transport so that individual trajectories do not always remains parallel to the average trajectory.

Acknowledgments

This work has been partly funded by the European Research Council (ERC) under grant agreement No 678304, European UnionâĂŹs Horizon 2020 research and innovation program under grant agreement No 666992, and the program âĂIJInvestissements dâĂŹavenirâĂİ ANR-10-IAIHU-06.

Appendix A. Proof of Lemma 2 and Lemma 3

In this appendix, we have first given a proof of Lemma 2 and we have then given a proof of Lemma 3.

A.1 Proof of Lemma 2

The definitions and properties given above are useful to prove the result given in 2. Before proving this result, we recalled its context. We considered a product manifold $\mathbb{M} = M^N$ where $M \subset \mathbb{R}$ is an open interval. M is equipped with a Riemannian metric g and is geodesically complete. The product manifold $\mathbb{M}$ is equipped with the product metric.
Proof Let $t_0 \in \mathbb{R}$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_N)$ be a geodesic of $\mathbb{M}$ and $\mathbf{w} = (w_1, \dots, w_N) \in \mathrm{T}_{\boldsymbol{\gamma}(t_0)}\mathbb{M}$. Using the previous proposition on product manifolds, the computation of $\eta^{\mathbf{w}}(\boldsymbol{\gamma}, t)$ boiled down to the computation of $\mathrm{Exp}_{\gamma_i(t)}\big(\mathrm{P}_{\gamma_i,t_0,t}(w_i)\big)$. This term was computed in three steps.

Let $i \in \{1, \dots, N\}$. The parallel transport $\mathrm{P}_{\gamma_i,t_0,t}(w_i)$ was computed as follows. First, note that M is an open interval of $\mathbb{R}$. Therefore, for all $p \in M$, $\mathrm{T}_pM \simeq \mathbb{R}$. The Riemannian metric g of M is necessarily of the form $p \in M \mapsto g_p$ with : $\forall (u, v) \in \mathrm{T}_pM$, $g_p(u, v) = uvf(p)$ where $f : M \to]0, +\infty[$ is a smooth function.

It follows from the definition of parallel transport along the curve $t \mapsto \gamma_i$ that : $\forall t$, $\mathrm{P}_{\gamma_i,t_0,t}(w_i) \in \mathrm{T}_{\gamma_i(t)}M$. Since, for all t, $\mathrm{T}_{\gamma_i(t)}M$ is a one-dimensional vector space, the tangent vector $\dot{\gamma}_i(t) \neq 0$ spans this space. As a consequence, there exists a smooth function $\xi_i : \mathbb{R} \to \mathbb{R}$ such that : $\forall t \in \mathbb{R}$, $\mathrm{P}_{\gamma_i,t_0,t}(w_i) = \xi_i(t)\dot{\gamma}_i(t)$. Because the parallel transport is an isometry and because γ_i is a geodesic, we have :

$$\forall t \in \mathbb{R},\ g_{\gamma_i(t)}\big(\mathrm{P}_{\gamma_i,t_0,t}(w_i), \dot{\gamma}_i(t)\big) = g_{\gamma_i(t_0)}(w_i, \dot{\gamma}_i(t_0)). \tag{18}$$

The bilinearity of $g_{\gamma_i(t)}$ gives :

$$\begin{aligned} g_{\gamma_i(t)}\big(\mathrm{P}_{\gamma_i,t_0,t}(w_i), \dot{\gamma}_i(t)\big) &= g_{\gamma_i(t)}\big(\xi_i(t)\dot{\gamma}_i(t), \dot{\gamma}_i(t)\big) \\ &= \xi_i(t) g_{\gamma_i(t)}\big(\dot{\gamma}_i(t), \dot{\gamma}_i(t)\big). \end{aligned} \tag{19}$$

Using that $\dot{\gamma}_i$ is parallel along γ_i, we have :

$$\forall t \in \mathbb{R},\ g_{\gamma_i(t)}\big(\dot{\gamma}_i(t), \dot{\gamma}_i(t)\big) = g_{\gamma_i(t_0)}\big(\dot{\gamma}_i(t_0), \dot{\gamma}_i(t_0)\big). \tag{20}$$

As a consequence, (18), (19) and (20) give :

$$\forall t \in \mathbb{R},\ \xi_i(t) g_{\gamma_i(t_0)}\big(\dot{\gamma}_i(t_0), \dot{\gamma}_i(t_0)\big) = g_{\gamma_i(t_0)}\big(w_i, \dot{\gamma}_i(t_0)\big). \tag{21}$$

Using the form of the metric on M, (21) writes :

$$\forall t \in \mathbb{R},\ \xi_i(t)\big(\dot{\gamma}_i(t_0)\big)^2 f\big(\gamma_i(t_0)\big) = w_i \dot{\gamma}_i(t_0) f\big(\gamma_i(t_0)\big). \tag{22}$$

This last equation gives : $\forall t$, $\xi_i(t) = w_i/\dot{\gamma_i}(t_0)$. Finally,

$$\forall i \in \{1, \dots, N\},\ \forall t \in \mathbb{R},\ \mathrm{P}_{\gamma_i,t_0,t}(w_i) = \frac{w_i}{\dot{\gamma}_i(t_0)} \dot{\gamma}_i(t). \tag{23}$$

The last step consists of computing $\mathrm{Exp}_{\gamma_i(t)}\big(\mathrm{P}_{\gamma_i,t_0,t}(w_i)\big)$ with $t \in \mathbb{R}$ fixed. This is done by introducing the curves

$$c \,:\, s \in [0,1] \,\mapsto\, \mathrm{Exp}_{\gamma_i(t)}\big(s\mathrm{P}_{\gamma_i,t_0,t}(w_i)\big),$$

and

$$\tilde{c} : s \in [0,1] \mapsto \gamma_i\Big(t + s\frac{w_i}{\dot{\gamma}_i(t_0)}\Big).$$

Both curves c and $\tilde{c}$ are geodesics of M which satisfy to : $c(0) = \tilde{c}(0) = \gamma_i(t)$ and $\dot{c}(0) = \dot{\tilde{c}}(0) = \frac{w_i}{\dot{\gamma}_i(t_0)}\dot{\gamma}_i(t)$. By unicity, the two curves are equal. As a consequence, for all $i \in \{1,\ldots,N\}$ and all $t \in \mathbb{R}$,

$$\mathrm{Exp}_{\gamma_i(t)}\big(\mathrm{P}_{\gamma_i,t_0,t}(w_i)\big) = \gamma_i\Big(t + \frac{w_i}{\dot{\gamma}_i(t_0)}\Big). \tag{24}$$

This last equation completes the proof of the lemma. ■

A.2 Proof of Lemma 3

First, some notations need to be introduced. Let $m = n(n+1)/2$ denote the dimension of the linear space SYM(n), $(\mathbf{E}_i)_{1\leq i\leq m}$ be the canonical basis of SYM(n) and $(\mathbf{E}_i^*)_{1\leq i\leq m}$ its *dual basis*. The matrices are indexed by a single index, which corresponds to an enumeration of the pairs of integers $\{(k,l),\ 1 \leq k,l \leq n,\, k \leq l\}$. A matrix $\mathbf{V}$ in SYM(n) will be identified to the vector $(v_1,\ldots,v_m)$ of its coefficients from the upper triangular part. Using the expression of the Christoffel symbols in terms of the canonical basis of SYM(n) and its dual basis, Lenglet and collaborators proved that if $t \mapsto \mathbf{\Sigma}(t) = (\sigma_1(t),\ldots,\sigma_m(t))$ is a smooth curve in SPD(n) and $t \mapsto \mathbf{V}(t) = (v_1(t),\ldots,v_m(t))$ a vector field along Σ, the *covariant derivative* of $\mathbf{V}$ along $\mathbf{\Sigma}$ is given by the expression:

$$D\mathbf{V}/dt = \sum_i^m \frac{dv_i}{dt}(t)\mathbf{E}_i + \sum_{i,j=1}^m v_i(t)\frac{d\sigma_j}{dt}(t)\nabla_{\mathbf{E}_i}\mathbf{E}_j. \tag{25}$$

Taking the Frobenius inner product of the previous equality (on both sides) with $\mathbf{E}_k^*$ ($1 \leq k \leq m$), together with the expression of the Christoffel symbols (see Lenglet et al. (2006), Eq.(3) and Eq.(4)), one gets that the vector field $\mathbf{V}$ is parallel along the curve $\mathbf{\Sigma}$ if and only if:

$$\frac{d\mathbf{V}}{dt}(t) - \frac{1}{2}\mathbf{V}(t)\mathbf{\Sigma}(t)^{-1}\frac{d\mathbf{\Sigma}}{dt}(t) - \frac{1}{2}\frac{d\mathbf{\Sigma}(t)}{dt}(t)\mathbf{\Sigma}(t)^{-1}\mathbf{V}(t) = 0. \tag{26}$$

We can now give the proof of Lemma 3.

Proof With the expression of the geodesic $\boldsymbol{\gamma}_0$, given in Section 3.4, one can easily see that:

$$\forall t \in \mathbb{R},\ \frac{d\boldsymbol{\gamma}_0(t)}{dt}\boldsymbol{\gamma}_0^{-1}(t) = \mathbf{V}_0\mathbf{P}_0^{-1}. \tag{27}$$

Let $t \mapsto \mathbf{V}(t)$ be a vector field parallel along $\boldsymbol{\gamma}_0$. Eq. (26) is equivalent to:

$$\frac{d\mathbf{V}(t)}{dt} = \frac{1}{2}\mathbf{V}(t)\mathbf{P}_0^{-1}\mathbf{V}_0 - \frac{1}{2}\mathbf{V}_0\mathbf{P}_0^{-1}\mathbf{V}(t). \tag{28}$$

Note that Eq. (28) is a *differential Lyapunov equation*. It can be solved by considering the matrix-valued function $t \mapsto \exp(-t\mathbf{M}^\top)\mathbf{R}(t)\exp(-t\mathbf{M})$, with $\mathbf{M} = -(1/2)\mathbf{P}_0^{-1}\mathbf{V}_0$ and $\mathbf{R}$, any differentiable matrix-valued function. Given that $\mathbf{V}(t_0) = \mathbf{W}$, one has that the parallel transport in $\mathrm{SPD}(n)$ is given by:

$$\mathbf{V}(t) = \exp\Big(\frac{t-t_0}{2}\mathbf{V}_0\mathbf{P}_0^{-1}\Big)\mathbf{W}\exp\Big(\frac{t-t_0}{2}\mathbf{P}_0^{-1}\mathbf{V}_0\Big). \tag{29}$$

■

References

Stéphanie Allassonnière, Estelle Kuhn, and Alain Trouvé. Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli*, 16(3):641–678, 2010.

V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

Y. F. Atchadé. An adaptive version for the metropolis adjusted langevin algorithm with a truncated drift. *Methodology and Computing in Applied Probability*, 8(2):235–254, 2006.

L.G. Bandini, A. Must, J.L. Spadano, and W.H. Dietz. Relation of body composition, parental overweight, pubertal stage, and race-ethnicity to energy expenditure among premenarcheal girls. *The American Journal Of Clinical Nutrition*, 76(5):1040–1047, 2002.

T. F. Coleman and D. C. Sorensen. A note on the computation of an orthonormal basis for the null space of a matrix. *Mathematical Programming*, 29(2):234–242, 1984.

I. Delor, JE. Charoin, R. Gieschke, S. Retout, and P. Jacqmin. Modeling alzheimer's disease progression using disease onset time and disease trajectory concepts applied to cdr-sob scores from adni. *CPT: Pharmacometrics & systems Pharmacology*, 2(10):e78, 2013.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

M. P. Do Carmo Valero. *Riemannian Geometry*. Birkhäuser, 1992.

S. Durrleman, X. Pennec, A. Trouvé, G. Gerig, and N. Ayache. Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 297–304. Springer Berlin Heidelberg, 2009.

Stanley Durrleman, Xavier Pennec, Alain Trouvé, José Braga, Guido Gerig, and Nicholas Ayache. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *International Journal of Computer Vision*, 103(1):22–59, 2013.

C. Eisenhart. The assumptions underlying the analysis of variance. *Biometrics*, 3(1):1–21, 1947.

G.M. Fitzmaurice, N.M. Laird, and J.H. Ware. *Applied Longitudinal Analysis*, volume 998. John Wiley & Sons, 2012.

T. Fletcher. Geodesic regression on riemannian manifolds. In *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy-Geometrical and Statistical Methods for Modelling Biological Shape Variability*, pages 75–86, 2011.

Hubert M Fonteijn, Marc Modat, Matthew J Clarkson, Josephine Barnes, Manja Lehmann, Nicola Z Hobbs, Rachael I Scahill, Sarah J Tabrizi, Sebastien Ourselin, Nick C Fox, et al.

An event-based model for disease progression and its application in familial alzheimer's disease and huntington's disease. *NeuroImage*, 60(3):1880–1889, 2012.

M. D. Hoffman and A. Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

Yi Hong, Nikhil Singh, Roland Kwitt, and Marc Niethammer. Time-warped geodesic regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 105–112. Springer, 2014.

Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*, volume 46. John Wiley & Sons, 2004.

B. M. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jedynak, B. Caffo, J. L. Prince, et al. A computational neurodegenerative disease progression score: method and results with the alzheimer's disease neuroimaging initiative cohort. *NeuroImage*, 63(3):1478–1486, 2012.

I. Koval, J.-B. Schiratti, A. Routier, M. Bacci, O. Colliot, S. Allassonnière, and S. Durrleman. Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In *Medical Image Computing and Computer-Assisted Intervention*, 2017.

E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probability and Statistics*, 8:115–131, 2004.

Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

M. Lavielle and F. Mentré. Estimation of population pharmacokinetic parameters of saquinavir in hiv patients with the monolix software. *Journal of Pharmacokinetics and Pharmacodynamics*, 34(2):229–249, 2007.

C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor mri processing. *Journal of Mathematical Imaging and Vision*, 25(3):423–444, 2006.

M. Lorenzi, X. Pennec, G. B Frisoni, and N. Ayache. Disentangling normal aging from alzheimer's disease in structural magnetic resonance images. *Neurobiology of Aging*, 36: S42–S52, 2015.

M. Moakher and M. Zéraï. The riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *Journal of Mathematical Imaging and Vision*, 40(2):171–187, 2011.

R. C. Mohs, D. Knopman, R. C Petersen, S. H. Ferris, C. Ernesto, M. Grundman, M. Sano, L. Bieliauskas, D. Geldmacher, C. Clark, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the alzheimer's disease assessment scale that broaden its scope. *Alzheimer Disease & Associated Disorders*, 11:13–21, 1997.

X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

P. Petersen. *Riemannian Geometry*, volume 171. Springer, 2006.

S.M. Phillips, L.G. Bandini, D.V. Compton, E.N. Naumova, and A. Must. A longitudinal comparison of body composition by total body water and bioelectrical impedance in adolescent girls. *The Journal of Nutrition*, 133(5):1419–1425, 2003.

G. O. Roberts, A. Gelman, W. R. Gilks, et al. Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.

J.-B. Schiratti, S. Allassonnière, O. Colliot, and S. Durrleman. Mixed-effects model for the spatiotemporal analysis of longitudinal manifold-valued data. In *5th MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, 2015a.

J.-B. Schiratti, S. Allassonnière, O. Colliot, and S. Durrleman. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *Advances in Neural Information Processing Systems*, pages 2404–2412, 2015b.

J.-B. Schiratti, S. Allassonnière, A. Routier, O. Colliot, S. Durrleman, and the ADNI. A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data. In *International Conference on Information Processing in Medical Imaging*, pages 564–575. Springer, 2015c.

N. Singh, J. Hinkle, S. Joshi, and P. T. Fletcher. An efficient parallel algorithm for hierarchical geodesic models in diffeomorphisms. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 341–344. IEEE, 2014.

Nikhil Singh, Jacob Hinkle, Sarang Joshi, and P Thomas Fletcher. A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In *Information Processing in Medical Imaging*, pages 560–571. Springer, 2013.

L. T. Skovgaard. A riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, pages 211–223, 1984.

J. Su, S. Kurtek, E. Klassen, A. Srivastava, 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.

G. Verbeke and G. Molenberghs. *Linear Mixed Models for Longitudinal Data*. Springer Science & Business Media, 2009.

Eric Yang, Michael Farnum, Victor Lobanov, Tim Schultz, Nandini Raghavan, Mahesh N Samtani, Gerald Novak, Vaibhav Narayan, and Allitia DiBernardo. Quantifying the pathophysiological timeline of alzheimer's disease. *Journal of Alzheimer's Disease*, 26(4):745–753, 2011.

A. L. Young, N. P. Oxtoby, J. Huang, R. V. Marinescu, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, D. C. Alexander, et al. Multiple orderings of events in disease progression. In *International Conference on Information Processing in Medical Imaging*, pages 711–722. Springer, 2015.