



# Mixed-effects model for the spatiotemporal analysis of longitudinal manifold-valued data

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

## ► To cite this version:

Jean-Baptiste Schiratti, Stéphanie Allasonnière, Olivier Colliot, Stanley Durrleman. Mixed-effects model for the spatiotemporal analysis of longitudinal manifold-valued data. 5th MICCAI Workshop on Mathematical Foundations of Computational Anatomy, Oct 2015, Munich, Germany. hal-01245905

HAL Id: hal-01245905

<https://hal.archives-ouvertes.fr/hal-01245905>

Submitted on 17 Dec 2015

**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.

Public Domain

# Mixed-effects model for the spatiotemporal analysis of longitudinal manifold-valued data

J.-B. Schiratti<sup>3,1</sup>, S. Allassonnière<sup>3</sup>, the Alzheimer’s Disease Neuroimaging Initiative, O. Colliot<sup>1,2</sup>, S. Durrleman<sup>1</sup>

<sup>1</sup> Inria Paris-Rocquencourt, Inserm U1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMRS 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France

<sup>2</sup> AP-HP, Pitié-Salpêtrière Hospital, Departments of Neurology and Neuroradiology, F-75013, Paris, France

<sup>3</sup> Ecole Polytechnique, Palaiseau, France

**Abstract.** In this work, we propose a generic hierarchical spatiotemporal model for longitudinal manifold-valued data, which consist in repeated measurements over time for a group of individuals. This model allows us to estimate a group-average trajectory of progression, considered as a geodesic of a given Riemannian manifold. Individual trajectories of progression are obtained as random variations, which consist in parallel shifting and time reparametrization, of the average trajectory. These spatiotemporal transformations allow us to characterize changes in the direction and in the pace at which trajectories are followed. We propose to estimate the parameters of the model using a stochastic approximation of the expectation-maximization (EM) algorithm, the Monte Carlo Markov Chain Stochastic Approximation EM (MCMC SAEM) algorithm.

This generic spatiotemporal model is used to analyze the temporal progression of a family of biomarkers. This progression model estimates a normative scenario of the progressive impairments of several cognitive functions, considered here as biomarkers, during the course of Alzheimer’s disease. The estimated average trajectory provides a normative scenario of disease progression. Random effects provide unique insights into the variations in the ordering and timing of the succession of cognitive impairments across different individuals.

## 1 Introduction

Neurodegenerative diseases such as Alzheimer’s disease (AD) or Parkinson’s disease are known to affect the metabolism, brain structure and cognitive functions. The effect of the disease can be quantified by observing cerebrospinal fluid (CSF), neuroimaging or neuropsychological biomarkers. In [9], Clifford R. Jack et al. proposed an hypothetical model to describe the temporal progression of these biomarkers during the course of the disease. However, there is still a need for data-driven models which could give experimental evidence of such patterns of disease progression. Statistical models for longitudinal data have been subject

to a growing interest in the last few years. In particular, mixed-effects models, which include fixed and random effects, have a hierarchical structure which allows us to describe the model at the group and subjects level.

Still, a statistical model of disease progression should take into account the fact that the age of a given individual is not an indicator of his stage of disease progression. Two individuals, observed at the same age, might actually be at very different stages of disease progression. As a consequence, trajectories should be registered in time to account for this variability in stages of disease progression. In [6], the concept of “time-warps” was introduced to allow for the registration in time of trajectories of shape changes. However, in order to combine the time-warps with the variability of shapes across individuals, the authors assumed that the variance of shapes does not depend on time whereas it should adapt to the average trajectory of shape changes. The set of the measurements of an individual at a given time-point is often a high-dimensional and nonlinear space. Building a model of disease progression therefore consists in estimating continuous subject-specific trajectories and an average trajectory in this space. At a given time point, the disease progression of two individuals will probably be described by two different trajectories. To construct the average trajectory, the individual trajectories need to be registered in *space*, where *space* may refer to the 2D or 3D space of spatial objects, or more generally to the space of measurements. In [15], time-warps were also used to define a metric between curves which has the property of being invariant under time-reparametrization. The authors did not spatially register the curves because of the small variability of the trajectories.

The framework of mixed-effects models provides tools to deal with this hierarchical problem. Mixed-effects models for longitudinal measurements were introduced in the seminal paper of Laird and Ware [11] and have been widely developed since then (see [4], [13] for instance). It should be pointed out that this kind of models suffers from two main drawbacks regarding our problem. These models describe the distribution of the measurements at a given reference time. In many situations, this reference time is given by the experimental setting: in plant growth studies, the point in time at which the plant was seeded is a natural choice, as well as the date of birth in developmental studies. However, in studies on neurodegenerative diseases, there is no natural choice of reference time as the disease-onset time is most probably different for each individual. Another limitation of usual mixed-effects models is that they are defined for data lying in Euclidean spaces. Although the development of statistical models for manifold-valued data is a blooming topic (see [14], [15]), the construction of statistical models for longitudinal data on a manifold remains an open problem.

In this paper, we propose a statistical model to describe the temporal progression of a family of biomarkers. This progression model can be seen as a particular case of a more general spatiotemporal model for longitudinal manifold-valued data. The Riemannian manifold and its metric are chosen a priori, which allows us to introduce anatomical, physiological constraints into the model. The definition of the generic spatiotemporal model requires no other choice. The models

which we introduce herein are based on the concept of parallel curves on a manifold. The random effects of the model allow to spatially and temporally register individual trajectories of progression. The generic spatiotemporal model belongs to a class of statistical models for which maximum likelihood estimates cannot be obtained in closed form. We address this issue by using a stochastic version of the Expectation-Maximization algorithm [3], namely the MCMC SAEM [1], for which theoretical results regarding the convergence have been proven in [2], [1].

In section 2, we will introduce our propagation model for a family of biomarkers and explain how this model appears as a particular case of a more general mixed-effects model for longitudinal manifold-valued data. We explain how the MCMC-SAEM was used in section 3. The last section consists of experimental results obtained on neuropsychological test scores.

## 2 Propagation model for a family of biomarkers

### 2.1 Riemannian geometry prerequisites

Let  $(\mathbb{M}, g^{\mathbb{M}})$  be a Riemannian manifold of dimension  $N$  equipped with a Riemannian metric  $g^{\mathbb{M}}$ , which we assume to be geodesically complete. Meaning that the geodesics of  $\mathbb{M}$  are defined on  $\mathbb{R}$ . The Riemannian metric  $g^{\mathbb{M}}$  defines a unique affine connexion on  $\mathbb{M}$ , namely the Levi-Civita connexion, denoted by  $\nabla^{\mathbb{M}}$ . Let  $\gamma$  denote a geodesic of  $\mathbb{M}$  and  $t_0 \in \mathbb{R}$ . We recall that, given a tangent vector  $\xi$  in  $T_{\gamma(t_0)}\mathbb{M}$ , the parallel transport of  $\xi$  along  $\gamma$ , denoted by  $X(s) = P_{\gamma, t_0, s}(\xi)$ , is a vector field along  $\gamma$  which satisfies :  $X(t_0) = \xi$  and  $\nabla^{\mathbb{M}}X(s) = 0$ . Let  $\mathbf{p} \in \mathbb{M}$ . The Riemannian exponential in  $\mathbb{M}$  at  $\mathbf{p}$  is denoted by  $\text{Exp}_{\mathbf{p}}^{\mathbb{M}}$ . For  $\mathbf{v} \in T_{\mathbf{p}}\mathbb{M}$ ,  $\text{Exp}_{\mathbf{p}}^{\mathbb{M}}(\mathbf{v})$  denotes the value at time 1 of the geodesic in  $\mathbb{M}$  issued from  $\mathbf{p}$  with initial velocity  $\mathbf{v}$ .

### 2.2 Model description

We are interested in the temporal progression of a family of  $N$  ( $N \geq 2$ ) scalar biomarkers. We consider a longitudinal dataset of the form  $(\mathbf{y}_{i,j}, t_{i,j})_{i,j}$ , obtained by observing  $p$  individuals at repeated time points. The vector  $\mathbf{y}_{i,j}$  denotes the  $j$ -th observation ( $1 \leq j \leq n_i$ ) of the  $i$ -th individual. The  $k$ -th coordinate of  $\mathbf{y}_{i,j}$ , denoted by  $y_{i,j,k}$ , corresponds to the measurement of the  $k$ -th biomarker, at time  $t_{i,j}$ . We will assume that each measurement  $y_{i,j,k}$  belongs to a one dimensional Riemannian manifold  $(M, g)$  which is geodesically complete. In this setting, the observations  $\mathbf{y}_{i,j} = (y_{i,j,1}, \dots, y_{i,j,N})$  can be considered as points in the product manifold  $\mathbb{M} = M^N$ . The average progression of this family of biomarkers is modeled by a geodesic trajectory on the manifold  $\mathbb{M}$ , which is equipped with the product metric, denoted by  $g^{\mathbb{M}}$ .

The statistical model is described for observations on a manifold which is a product of one-dimensional manifolds. This framework is particularly convenient to analyze the temporal progression of a family of biomarkers. The model can be seen as a particular case of a more generic spatiotemporal model (2).

### Choice of the average trajectory

In order to determine relative progression of the biomarkers among themselves, the average trajectory is chosen among the parametric family of geodesics  $: (t \mapsto (\gamma_0(t), \gamma_0(t + \delta_1), \dots, \gamma_0(t + \delta_{N-1})))_{\boldsymbol{\delta}}$ , where  $\boldsymbol{\delta} = (0, \delta_1, \dots, \delta_{N-1})$  and  $\gamma_0$  is a geodesic, of the one-dimensional manifold  $M$ , parametrized by a point  $p_0$  in  $M$ , a time  $t_0$  and a velocity  $v_0$  in  $T_{p_0}M$ . This parametrization of the geodesic  $\gamma_0$  is the natural parametrization  $: \gamma_0(t_0) = p_0$  and  $\dot{\gamma}_0(t_0) = v_0$ . By choosing the average trajectory among this parametrized family of geodesics, we assume that, on average, the biomarkers follow the same trajectory but shifted in time. The delay between the progression of the different biomarkers is measured by the vector  $\boldsymbol{\delta} = (0, \delta_1, \dots, \delta_{N-1}) \in \mathbb{R}^N$ . The parameters  $\delta_i$  ( $1 \leq i \leq N - 1$ ) measure a relative delay between two consecutive biomarkers. The parameter  $t_0$  plays the role of reference time as the trajectory of the first biomarker will reach the value  $p_0$  at time  $t_0$  whereas the other trajectories will reach the same value  $p_0$  at different points in time, shifted with respect to  $t_0$ .

### Construction of subject-specific trajectories of disease progression with time reparametrization

The model proposed herein is a hierarchical model : data points are assumed to be sampled from subject-specific trajectories of progression. These individual trajectories are derived from the average trajectory  $\gamma_{\boldsymbol{\delta}}$ . The subject-specific trajectory of the  $i$ -th individual is constructed by considering a non-zero tangent vector  $\mathbf{w}_i$  in  $T_{\gamma_{\boldsymbol{\delta}}(t_0)}\mathbb{M}$ , orthogonal to  $\dot{\gamma}_{\boldsymbol{\delta}}(t_0)$  for the inner product defined by the metric  $(\langle \cdot, \cdot \rangle_{\gamma_{\boldsymbol{\delta}}(t_0)} = g_{\gamma_{\boldsymbol{\delta}}(t_0)}^{\mathbb{M}})$ . This tangent vector  $\mathbf{w}_i = (w_{1,i}, \dots, w_{N,i})$  is a *space shift* which allows us to register the individual trajectories in the space of measurements. The tangent vector  $\mathbf{w}_i$  is transported along the geodesic  $\gamma_{\boldsymbol{\delta}}$  from time  $t_0$  to time  $s$  using parallel transport. This transported tangent vector is denoted by  $P_{\gamma_{\boldsymbol{\delta}}, t_0, s}(\mathbf{w}_i)$ . At the point  $\gamma_{\boldsymbol{\delta}}(s)$ , a new point in  $\mathbb{M}$  is obtained by taking the Riemannian exponential of  $P_{\gamma_{\boldsymbol{\delta}}, t_0, s}(\mathbf{w}_i)$ . This new point is denoted by  $\eta^{\mathbf{w}_i}(\gamma_{\boldsymbol{\delta}}, s)$ . As  $s$  varies, this point describes the curve  $s \mapsto \eta^{\mathbf{w}_i}(\gamma_{\boldsymbol{\delta}}, s)$ , which is considered as a “parallel” to the curve  $\gamma_{\boldsymbol{\delta}}$  (Fig. 1). The orthogonality condition on the tangent vectors  $\mathbf{w}_i$  is an important hypothesis which ensures that a point  $\eta^{\mathbf{w}_i}(\gamma_{\boldsymbol{\delta}}, s)$  on a parallel moves at the same pace on this parallel than on the average trajectory. This hypothesis ensures the uniqueness of the decomposition between spatial and temporal components.

The trajectory  $\gamma_i$  of the  $i$ -th individual is obtained by reparametrizing the parallel  $\eta^{\mathbf{w}_i}(\gamma_{\boldsymbol{\delta}}, \cdot) : \gamma_i(s) = \eta^{\mathbf{w}_i}(\gamma_{\boldsymbol{\delta}}, \psi_i(s))$ , where the mapping  $\psi_i(s) = \alpha_i(s - t_0 - \tau_i) + t_0$  is a subject-specific affine reparametrization which allows us to register in time the different individual trajectories of progression. This time-warp was introduced in [12] to define subject-specific trajectories of progression from an average trajectory, in the case of univariate manifold-valued data. In this univariate work, because the manifold is one-dimensional, no random effect is associated to the fixed effect  $p_0$ . Here, the tangent vector  $\mathbf{w}_i$  can be considered, in the light of the univariate model, as a random effect associated to the point  $p_0$ . The effect  $\alpha_i$  is an *acceleration factor* which encodes whether the  $i$ -th individual

is progressing faster or slower than the average individual. Whereas the effect  $\tau_i$  encodes the advance or delay of the  $i$ -th individual with respect to the average. Both are assumed to be random, non observed, variables as are also the space shifts  $\mathbf{w}_i$ .

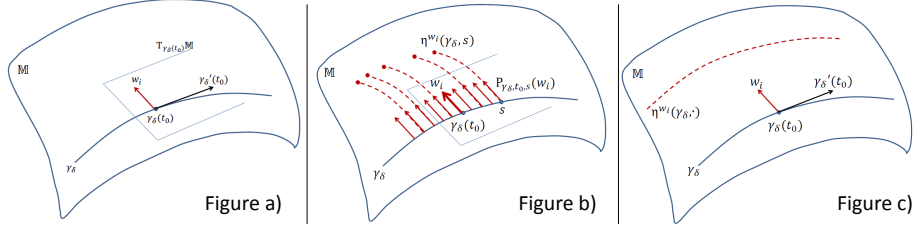


Fig. 1: Schematic description of parallel shifting. Figure a) (left) : A non-zero tangent vector  $\mathbf{w}_i$  is chosen in  $T_{\gamma(t_0)}\mathbb{M}$ . Figure b) (middle) : the tangent vector  $\mathbf{w}_i$  is transported along  $\gamma$  from  $\gamma(t_0)$  to  $\gamma(s)$  using parallel transport on  $\mathbb{M}$ . A new point,  $\eta^{\mathbf{w}_i}(\gamma, s)$  is obtained at time  $s$  by shooting with the Riemannian exponential. Figure c) (right) : the curve  $\eta^{\mathbf{w}_i}(\gamma, \cdot)$  is the “parallel” to  $\gamma$  constructed from  $\mathbf{w}_i$ .

Because  $\mathbb{M}$  is equipped with the product metric, the parallel transport of the tangent vector  $\mathbf{w}_i \in T_{\gamma_\delta(t_0)}\mathbb{M}$  is a  $N$ -dimensional vector whose  $k$ -th ( $1 \leq k \leq N$ ) component is equal to the parallel transport of the tangent vector  $w_{k,i} \in T_{\gamma_0(\cdot + \delta_{k-1})}M$  along the curve  $s \mapsto \gamma_0(s + \delta_{k-1})$  in the one-dimensional manifold  $M$ . It follows that  $P_{\gamma_\delta, t_0, s}(\mathbf{w}_i) = (\frac{w_{1,i}}{\dot{\gamma}_0(t_0)}\dot{\gamma}_0(s), \dots, \frac{w_{N,i}}{\dot{\gamma}_0(t_0 + \delta_{N-1})}\dot{\gamma}_0(s + \delta_{N-1}))$ . Taking the Riemannian exponential, in  $\mathbb{M}$ , of the tangent vector  $P_{\gamma_\delta, t_0, s}(\mathbf{w}_i)$  boils down to taking the Riemannian exponential, in  $M$ , of each component of the vector. If  $\text{Exp}^M$  denotes the Riemannian exponential map in  $M$ , the  $k$ -th ( $1 \leq k \leq N$ ) component of  $\eta^{\mathbf{w}_i}(\gamma_\delta, s)$  is given by :  $\text{Exp}_{\gamma_0(s + \delta_{k-1})}^M[\frac{w_{k,i}}{\dot{\gamma}_0(t_0 + \delta_{k-1})}\dot{\gamma}_0(s + \delta_{k-1})] = \gamma_0(s + \delta_{k-1} + \frac{w_{k,i}}{\dot{\gamma}_0(t_0 + \delta_{k-1})})$ . For the longitudinal dataset  $(\mathbf{y}_{i,j}, t_{i,j})$  ( $1 \leq i \leq p, 1 \leq j \leq n_i$ ), our hierarchical model writes :  $\mathbf{y}_{i,j} = \gamma_i(t_{i,j}) + \boldsymbol{\varepsilon}_{i,j}$ . In particular, for the  $k$ -th biomarker, this hierarchical model writes :

$$y_{i,j,k} = \gamma_0\left(\frac{w_{k,i}}{\dot{\gamma}_0(t_0 + \delta_{k-1})} + t_0 + \alpha_i(t_{i,j} - t_0 - \tau_i) + \delta_{k-1}\right) + \varepsilon_{i,j,k}. \quad (1)$$

with  $\alpha_i = \exp(\eta_i)$ ,  $\mathbf{w}_i = (w_{1,i}, \dots, w_{N,i})$ ,  $\mathbf{w}_i = A\mathbf{s}_i$  and :

$$\eta_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\eta^2), \tau_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\tau^2), \boldsymbol{\varepsilon}_{i,j} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_i}) \text{ and } s_{j,i} \stackrel{\text{i.i.d.}}{\sim} \text{Laplace}(1/2).$$

The parameters of the model are  $\boldsymbol{\theta} = (p_0, t_0, v_0, \boldsymbol{\delta}, \sigma_\eta, \sigma_\tau, \sigma, \text{vec}(A))$  and the random effects of the model are  $(\alpha_i, \tau_i, \mathbf{w}_i)$  ( $1 \leq i \leq p$ ). Note that the first two random effects are scalars. The acceleration factor is assumed to follow a

log-normal distribution to ensure its positiveness (the affine reparametrization must not reverse time). The time shifts follow a Gaussian distribution with zero mean. The space shifts are vectors of dimension  $N - 1$  which belong to the vector space  $\{\dot{\gamma}_\delta(t_0)\}^\perp$ . In the spirit of independent component analysis [8], we assume that the tangent vectors  $\mathbf{w}_i$  are a linear combination of  $N_s < N$  statistically independent components. This writes  $\mathbf{w}_i = \mathbf{A}\mathbf{s}_i$  where  $\mathbf{A}$  is a  $N \times N_s$  matrix of rank  $N_s$  whose columns are vectors in  $\mathbb{T}_{\gamma_\delta(t_0)}\mathbb{M}$  and  $\mathbf{s}_i$  is a vector of  $N_s$  independent sources following a Laplace distribution with parameter  $1/2$ . To ensure the orthogonality condition on the tangent vectors  $\mathbf{w}_i$ , we assume that, for all  $j \in \{1, \dots, N_s\}$ ,  $\langle \mathbf{A}_j, \dot{\gamma}_\delta(t_0) \rangle_{\gamma_\delta(t_0)} = 0$ , where  $\mathbf{A}_j$  denotes the  $j$ -th column of  $\mathbf{A}$ .

The model given in (1) can be used to analyze longitudinal observations on any geodesically complete Riemannian manifold. The *generic spatiotemporal model* writes:

$$\mathbf{y}_{i,j} = \eta^{\mathbf{w}_i}(\gamma_\delta, \psi_i(t_{i,j})) + \varepsilon_{i,j}. \quad (2)$$

where the parallel  $s \mapsto \eta^{\mathbf{w}_i}(\gamma_\delta, s)$  is given by:

$$\eta^{\mathbf{w}_i}(\gamma_\delta, s) = \text{Exp}_{\gamma_\delta(s)}^{\mathbb{M}}(\mathbf{P}_{\gamma_\delta, t_0, s}(\mathbf{w}_i)), \quad s \in \mathbb{R}.$$

It should be pointed out that a parallel  $s \mapsto \eta^{\mathbf{w}_i}(\gamma_\delta, s)$  to the geodesic  $\gamma_\delta$  is not, in general a geodesic. In the Euclidean case, a parallel to  $\gamma_\delta$  is just a translation of  $\gamma_\delta$  :  $\eta^{\mathbf{w}_i}(\gamma_\delta, s) = \gamma_\delta(s) + \mathbf{w}_i$ .

### 2.3 The logistic propagation model

If the measurements of the biomarkers can be *normalized*, we can consider these measurements as points in the one-dimensional manifold  $M = ]0, 1[$ . For example, neuropsychological test scores are bounded above by a maximum score and can therefore be normalized to produce measurements in  $]0, 1[$ . In this case, the model given in (1) can be used to analyze these measurements. We consider that  $M = ]0, 1[$  is equipped with the Riemannian metric  $g$  given by : for  $p \in ]0, 1[$ ,  $(u, v) \in \mathbb{T}_p M \times \mathbb{T}_p M$ ,  $g_p(u, v) = uG(p)v$  with  $G(p) = 1/(p^2(1-p)^2)$ . For this Riemannian metric, the geodesics are the logistic curves of the form :  $\gamma_0(t) = (1 + (\frac{1}{p_0} - 1) \exp(-\frac{v_0}{p_0(1-p_0)}(t - t_0)))^{-1}$ . In this setting, (1) writes:

$$y_{i,j,k} = \left( 1 + \left( \frac{1}{p_0} - 1 \right) \exp \left( - \frac{v_0 \alpha_i(t_{i,j} - t_0 - \tau_i) + v_0 \delta_k + v_0 \frac{(\mathbf{A}\mathbf{s}_i)_k}{\gamma_0(t_0 + \delta_k)}}{p_0(1-p_0)} \right) \right)^{-1} + \varepsilon_{i,j,k}, \quad (3)$$

where  $(\mathbf{A}\mathbf{s}_i)_k$  denotes the  $k$ -th component of the vector  $\mathbf{w}_i = \mathbf{A}\mathbf{s}_i$ .

## 3 Parameters estimation

A stochastic version of the Expectation-Maximization (EM) algorithm [3] is used to estimate the parameters  $\boldsymbol{\theta} = (p_0, t_0, v_0, \boldsymbol{\delta}, \sigma_\eta, \sigma_\tau, \sigma, \text{vec}(\mathbf{A}))$  of the model.

Because of the nonlinearity of the model, the E step of the EM algorithm is intractable. As a consequence, we consider a stochastic version of the EM algorithm, namely the Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization (MCMC-SAEM) algorithm [1], based on [2].

In order to ensure the theoretical convergence of the MCMC SAEM algorithm, the model must belong to the curved exponential family. Equivalently, the complete log-likelihood of the model shall write :  $\log q(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}) = -\phi(\boldsymbol{\theta}) + S(\mathbf{y}, \mathbf{z})^\top \psi(\boldsymbol{\theta})$ , where  $S(\mathbf{y}, \mathbf{z})$  is a sufficient statistic of the model. Note that the *logistic propagation model* does not belong to the curved exponential family. A usual workaround consists in regarding the parameters of the model as realizations of independents Gaussian random variables ([10]) :  $\boldsymbol{\theta} \sim \mathcal{N}(\bar{\boldsymbol{\theta}}, \mathbf{D})$  where  $\mathbf{D}$  is a diagonal matrix with very small diagonal entries and the estimation now targets  $\bar{\boldsymbol{\theta}}$ . This yields:  $p_0 \sim \mathcal{N}(\bar{p}_0, \sigma_{p_0}^2)$ ,  $t_0 \sim \mathcal{N}(\bar{t}_0, \sigma_{t_0}^2)$ ,  $v_0 \sim \mathcal{N}(\bar{v}_0, \sigma_{v_0}^2)$  and, for all  $k$ ,  $\delta_k \sim \mathcal{N}(\bar{\delta}_k, \sigma_{\delta}^2)$ . The matrix  $A$  is also considered as a realization of a Gaussian random variable and, in order to ensure the orthogonality condition on the columns of  $A$ , we assume that  $A$  follows a normal distribution on the space  $\boldsymbol{\Sigma} = \{A = (\mathbf{A}_1, \dots, \mathbf{A}_{N_s}) \in (\mathbb{T}_{\gamma_\delta(t_0)}\mathbb{M})^{N_s}; \forall j, \langle \mathbf{A}_j, \dot{\gamma}_\delta(t_0) \rangle_{\gamma_\delta(t_0)} = 0\}$ . Equivalently, we assume that the matrix  $A$  writes :  $A = \sum_{k=1}^{(N-1)N_s} c_k \mathcal{A}_k$  where, for all  $k$ ,  $c_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\bar{c}_k, \sigma_c^2)$  and  $(\mathcal{A}_1, \dots, \mathcal{A}_{(N-1)N_s})$  is an orthonormal basis of  $\boldsymbol{\Sigma}$  obtained by application of the Gram-Schmidt process to a basis of  $\boldsymbol{\Sigma}$ . The random variables  $c_1, \dots, c_{(N-1)N_s}$  are considered as new hidden variables of the model. The parameters of the model are  $\boldsymbol{\theta} = (\bar{p}_0, \bar{t}_0, \bar{v}_0, (\bar{\delta}_k)_{1 \leq k \leq N-1}, (\bar{c}_k)_{1 \leq k \leq (N-1)N_s}, \sigma_\eta, \sigma_\tau, \sigma)$  whereas the hidden variables of the model are  $\mathbf{z} = (p_0, t_0, v_0, (\delta_k)_{1 \leq k \leq N-1}, (c_k)_{1 \leq k \leq (N-1)N_s}, (\eta_i)_{1 \leq i \leq p}, (\tau_i)_{1 \leq i \leq p}, (s_{j,i})_{1 \leq j \leq N_s, 1 \leq i \leq p})$ .

### Overview of the MCMC-SAEM algorithm

The MCMC-SAEM iterates, until convergence, between three steps : simulation, stochastic approximation and maximization. Let  $k$  be an integer greater than 1 and  $\boldsymbol{\theta}^{(k-1)}$  (respectively  $\mathbf{z}^{(k-1)}$ ) denote the parameters (respectively the hidden variables) at the  $k - 1$ -th iteration of the algorithm. The  $k$ -th iteration is summarized as follows :

- **Simulation** :  $\mathbf{z}^{(k)}$  is sampled from the transition kernel of an ergodic Markov Chain whose stationary distribution is the conditional distribution of the hidden variables knowing the observations  $\mathbf{y} = (\mathbf{y}_{i,j})_{i,j}$  and the current estimates of the parameters  $\boldsymbol{\theta}^{(k-1)}$ . This sampling is done by using the Hasting-Metropolis within Gibbs sampler.
- **Compute the sufficient statistics** : we compute the sufficient statistics  $\mathbf{S}^{(k)}$ .
- **Stochastic approximation** : because the model belongs to the curved exponential family, the stochastic approximation is done on the sufficient statistics as follows :  $\mathbf{S}^{(k+1)} \leftarrow \mathbf{S}^{(k)} + \varepsilon_k (\mathbf{S}(\mathbf{y}, \mathbf{z}^{(k)}) - \mathbf{S}^{(k)})$ , where  $(\varepsilon_k)_k$  is a decreasing sequence of positive step sizes.
- **Maximization** : parameters updates are obtained in closed form from the stochastic approximation on the sufficient statistics.



## 4 Experiments

### 4.1 Data

We use the neuropsychological assessment test “ADAS-Cog 13” from the ADNI1, ADNIGO or ADNI2 cohorts of the Alzheimer’s Disease Neuroimaging Initiative (ADNI). The “ADAS-Cog 13” consists of 13 questions, which allow to test the impairment of several cognitive functions. To each of the 13 items of the test a cognitive function was associated as follows : memory (items 1, 4, 7, 8 and 9), language (items 2, 5, 10, 11 and 12), praxies (items 3 and 6), concentration (item 13). The score of each item was normalized by the maximum possible score. Consequently, each data point of each individual consists in thirteen normalized scores, which can be seen as a point on the manifold  $\mathbb{M} = ]0, 1[^{13}$ .

We choose to consider 248 individuals who were diagnosed with mild cognitive impairment (MCI) at their first visit and whose diagnosis changed to AD before their last visit. Among this population, we have an average of 6 visits per individual (with a minimum of 3 and a maximum of 11) and a typical duration of either 6 or 12 months between consecutive visits.

### 4.2 Experimental results

In this situation where  $\mathbb{M} = ]0, 1[^{13}$ , the number of independent sources  $N_s$  can be any integer between 1 and 12. The choice of the number of independent sources influences the number of parameters to be estimated, which equals  $9 + 12N_s$ . In order to keep a reasonable runtime, we conducted 3 experiments with  $N_s$  equal to 1, 2 and 3. For each experiment, the MCMC-SAEM was run five times with different initial parameters. Only the experiment which returned the smallest residual noise variance was kept. Increasing the number of sources allowed to decrease the residual noise among the experiments :  $\sigma^2 = 0.02$  for  $N_s = 1$ ,  $\sigma^2 = 0.0162$  for  $N_s = 2$  and  $\sigma^2 = 0.0159$  for  $N_s = 3$ . Because the residual noise was almost similar for  $N_s = 2$  and  $N_s = 3$  sources, we choose to report here the results obtained with the less complex model. As a consequence, we report the results obtained with 2 independent sources.

The average trajectory  $\gamma_{\delta}$  is given in Fig. 2, where each curve represents the temporal progression of one specific item of the ADAS-Cog test. The estimated fixed effects are  $p_0 = 0.74$ ,  $t_0 = 79.88$  years,  $v_0 = 0.047$  unit per year, and  $\delta = [0; -14; -11; 4.6; -13; -14; -7.7; -0.9; -14.4; -14.05; -11.80; -15.3292]$  years. This means that, on average, the memory-related items (items 1, 4, 7, 8, 9) reach the value  $p_0 = 0.74$  at respectively  $t_0, t_0 - \delta_4, t_0 - \delta_7, t_0 - \delta_8$  and  $t_0 - \delta_9$  years, which corresponds to respectively 79.88, 75.2, 87.6, 80.7 and 94.3 years. The concentration item reaches the same value at  $t_0 - \delta_{13} = 86.1$  years. The progression of the concentration item is followed by praxis and language items.

Random effects show the variability of this average trajectory within the studied population. The standard deviation of the time-shift equals  $\sigma_{\tau} = 8.3$  years, meaning that the disease progression model in Fig. 2 is shifted in time by

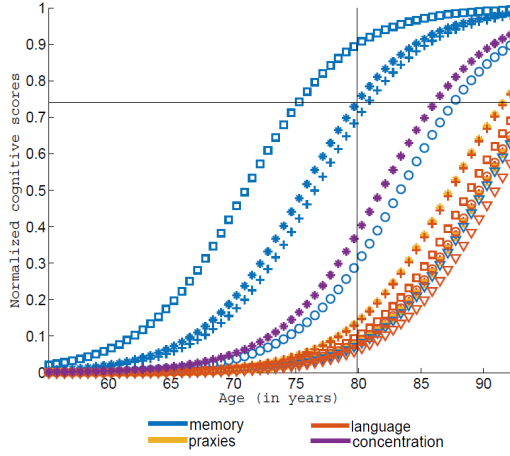


Fig. 2: The estimated average trajectory. In blue: the average trajectory of progression for the 5 memory-related items (item 1:\*, item 4:□, item 7:○, item 8:+ and item 9:△). In orange: average trajectory for the 5 language-related items (item 2:\*, item 5:□, item 10:○, item 11:+ and item 12:△). In yellow: average progression trajectory for the 2 praxies-related items (item 3:\* and item 6:□). In purple: average progression trajectory for the concentration-related item (item 13:\*).

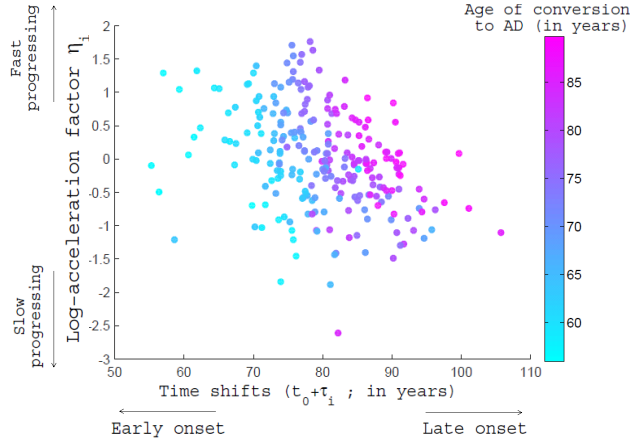


Fig. 3: Plot of the subject-specific random effects: the log-acceleration factor  $\eta_i$  is plotted against the time-shifts  $t_0 + \tau_i$ . Each point is colored according to the age of conversion to AD.

at most  $\pm 8.3$  years for 95% of the population. This accounts for the variability in the age of disease onset among the population. The effects of the variance of the acceleration factors, and the two independent components of the space-shifts are illustrated in Fig. 4. The first column of Fig. 4 illustrates the variability in pace of disease progression (the time-shifts are assumed to be zero in order to illustrate the effect of acceleration factor only). This variability is encoded by the variance

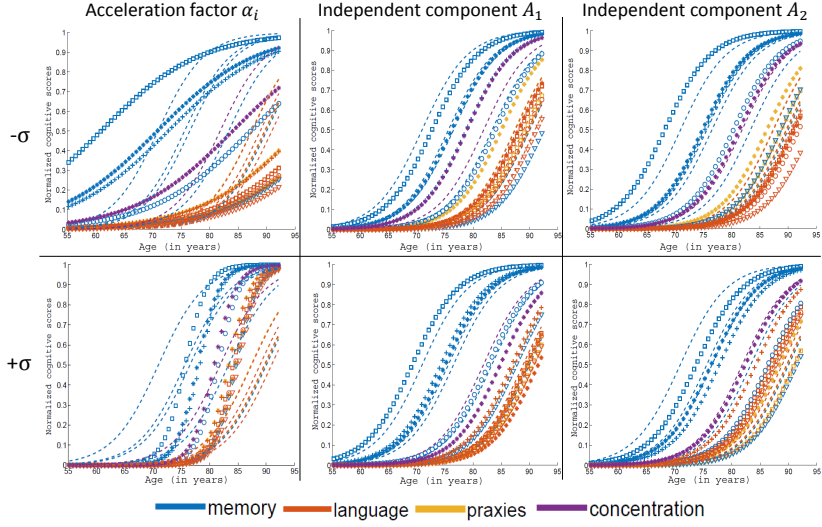


Fig. 4: First column : plot of  $t \mapsto \gamma_\delta(\alpha(t - t_0) + t_0)$  with  $\alpha = \exp(\pm\sigma_\eta)$  with  $\sigma_\eta = 0.8$ . Second (resp. third) column : plots of parallels  $\eta^{\pm\sigma_{s_i}} \mathbf{A}_i(\gamma_\delta, \cdot)$  in the direction given by the independent components.

$\sigma_\eta = 0.8$  of the acceleration factor. The first and second independent components illustrates the variability in the relative timing of the cognitive impairments. The first independent direction shows that some memory items and language items are shifted in time with respect to the other ones, especially for memory item 4 ( $\square$ ) and item 7 ( $\circ$ ). The ordering of the memory item 7 ( $\circ$ ) and the concentration item is inverted for individuals with a space shift  $\mathbf{w}_i = -\sigma_{s_{i,1}} \mathbf{A}_1$ . For those individuals, praxies items are impaired later, after the language items 2 ( $*$ ), items 12 ( $\Delta$ ) and item 5 ( $\square$ ). The second independent component shows a greater variability for the memory-related items than for the first independent components, in particular for memory item 9 ( $\Delta$ ) and item 4 ( $\square$ ). For individuals with a space shift  $\mathbf{w}_i = \sigma_{s_{i,2}} \mathbf{A}_2$ , language-related items might be impaired later than the average individual, especially for the language item 12 ( $\Delta$ ).

The subject-specific random effects estimated for each individual are obtained from the sampling step of the last iteration of the MCMC-SAEM and are plotted in Fig. 3. The figure shows that the individuals who have a positive (respectively negative) time shift (they are evolving ahead, respectively behind, the average trajectory) are the individuals who converted late (respectively early) to AD. This means that the individual time-shifts correspond well to the age at which a given individual was diagnosed with AD. We also note that there is a negative correlation, equal to  $-0.4$ , between the estimated log-acceleration factors and time shifts. There is a tendency for early onset patients to be fast progressers.

Through its subject-specific affine reparametrization, the age of a given individual is registered to the common timeline of the average scenario. In figure 5 (left),  $t \mapsto \sum_i |t_i^{\text{diag}} - \psi_i^{-1}(t)|$  (where  $t_i^{\text{diag}}$  corresponds to the age at which the

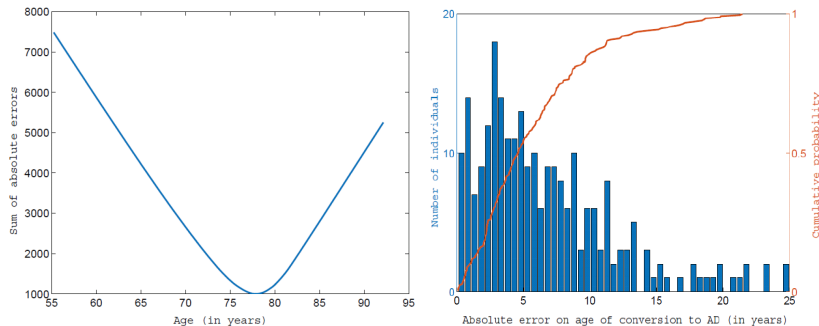


Fig. 5: Left : Sum of  $|t_i^{\text{diag}} - \psi_i^{-1}(t)|$  across the 248 individuals as a function of the variable  $t$  (time). Right : histogram of the absolute errors  $|t_i^{\text{diag}} - 77.45|$ .

$i$ -th individual converted to AD) shows a unique minimum at  $t^* = 77.45$  years. This age can be understood as the age of symptoms onset in the timeline of the normative scenario of disease progression. The histogram in 5 shows that the age  $t^*$  is a prediction of the true age of conversion : the error of prediction is less than 5 years for 50% of the population. This prediction is obtained by analyzing cognitive scores, which are inherently noisy and the reproducibility of these scores is questionable. We believe that the prediction can be improved by analyzing other types of biomarkers which are more objective and indicative of the progression of the disease.

### 4.3 Discussion and perspectives

We proposed a mixed-effects model to analyze the temporal progression of a family of biomarkers. This model appears as a particular case of a generic spatiotemporal model which can be used to analyze longitudinal manifold-valued measurements. These two models allow to estimate an average trajectory of disease progression. Individual trajectories of disease progression are constructed from the average trajectory by using subject-specific space shifts, acceleration factors and time shifts, which allow to spatially and temporally register the individual trajectories of progression.

The model for a family of biomarkers was used to estimate a scenario of Alzheimer's disease progression from neuropsychological tests. We validated the estimates of the spatiotemporal registration between individual trajectories by the fact that they put into correspondence the age at which patients were diagnosed with Alzheimer's disease. Alternatives to estimate model of disease progression include the event-based model proposed in [7], which estimates the ordering of categorical variables. The combination of spatial and temporal sources of variations in longitudinal data can be further improved by use of methods such as in [5]. In this work, we introduced the methodological background to construct models of disease progression based on longitudinal manifold-valued measurements. Improvements to the model we introduced above consist in an-

alyzing multimodal biomarkers. By doing so, we could experimental evidence, based on a data-driven model, of temporal progression of biomarkers as in [9].

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

## References

1. Allasonnière, S., Kuhn, E., Trouvé, A.: Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli* 16(3), 641–678 (2010)
2. Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics* pp. 94–128 (1999)
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)* pp. 1–38 (1977)
4. Diggle, P., Heagerty, P., Liang, K.Y., Zeger, S.: *Analysis of longitudinal data*. Oxford University Press (2002)
5. Donohue, M.C., Jacqmin-Gadda, H., Le Goff, M., Thomas, R.G., Raman, R., Gamst, A.C., Beckett, L.A., Jack, C.R., Weiner, M.W., Dartigues, J.F., Aisen, P.S., the Alzheimer’s Disease Neuroimaging Initiative: Estimating long-term multivariate progression from short-term data. *Alzheimer’s & Dementia* 10(5), 400–410 (2014)
6. Durrleman, S., Pennec, X., Trouvé, A., Braga, J., Gerig, G., Ayache, N.: Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *International Journal of Computer Vision* 103(1), 22–59 (2013)
7. Fonteijn, H.M., Modat, M., Clarkson, M.J., Barnes, J., Lehmann, M., Hobbs, N.Z., Scahill, R.I., Tabrizi, S.J., Ourselin, S., Fox, N.C., 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)
8. Hyvärinen, A., Karhunen, J., Oja, E.: *Independent component analysis*, vol. 46. John Wiley & Sons (2004)
9. Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q.: Hypothetical model of dynamic biomarkers of the alzheimer’s pathological cascade. *The Lancet Neurology* 9(1), 119–128 (2010)
10. Kuhn, E., Lavielle, M.: Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis* 49(4), 1020–1038 (2005)
11. Laird, N.M., Ware, J.H.: Random-effects models for longitudinal data. *Biometrics* pp. 963–974 (1982)
12. Schiratti, J.B., Allasonnière, S., Routier, A., the ADNI, Colliot, O., Durrleman, S.: A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data. In: *Information Processing in Medical Imaging*. pp. 564–576. Springer (2015)
13. Singer, J.D., Willett, J.B.: *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press (2003)
14. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: *Information Processing in Medical Imaging*. pp. 560–571. Springer (2013)
15. Su, J., Kurtek, S., Klassen, E., Srivastava, A., et al.: Statistical analysis of trajectories on riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *The Annals of Applied Statistics* 8(1), 530–552 (2014)