Learning spatio-temporal trajectories from manifold-valued longitudinal data

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

Learning spatio-temporal trajectories from manifold-valued longitudinal data

Jean-Baptiste Schiratti[1,2], Stéphanie Allassonnière[2], Olivier Colliot[1], Stanley Durrleman[1]

Inria Paris-Rocquencourt, Sorbonne Universités, UPMC Univ Paris 06 UMR S1127, CNRS UMR 7225, ICM, F-75013, Paris, France

CMAP, Ecole Polytechnique, Palaiseau, France

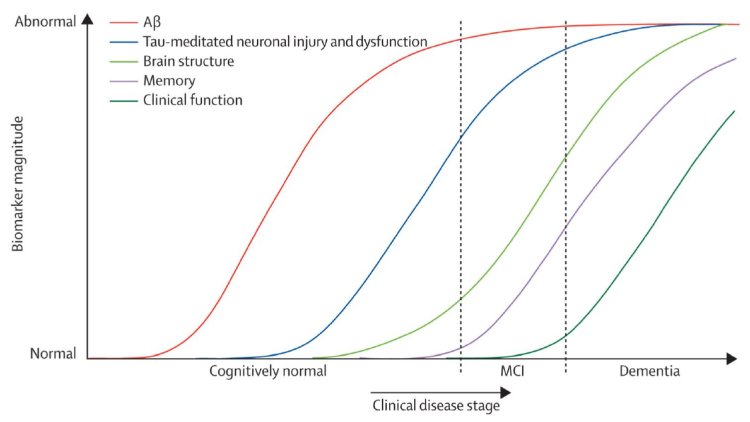
ARAMISLAB

www.aramislab.fr

Introduction

Aim : model the progression of neuro-degenerative diseases

- Understanding the progression of neuro-degenerative diseases, such as Alzheimer's Disease (AD) is necessary for early and acurate diagnosis and care planning.
- We need to validate experimentally hypothetical models of disease progression, such as [Clifford Jack et al, 2010].



Clifford Jack et al, Lancet Neurol., 2010

Working with longitudinal data in the context of neuro-degenerative diseases raises two difficulties :

- **Two individuals of the same age might be at very different stages of disease progression**

  ⟹ statistical models based on the regression of measurements with age are inadequate to model disease progression and age shoud not be treated as a covariate but as a random variable.

- **Longitudinal measurements sometimes belong to Riemannian manifolds** (non-Euclidean spaces).

  ⟹ statistical models for such longitudinal data should be defined for manifold-valued measurements. Linear mixed-effects models [Laird and Ware, 1982] are not defined for manifold-valued measurements

Generic spatio-temporal model for longitudinal data

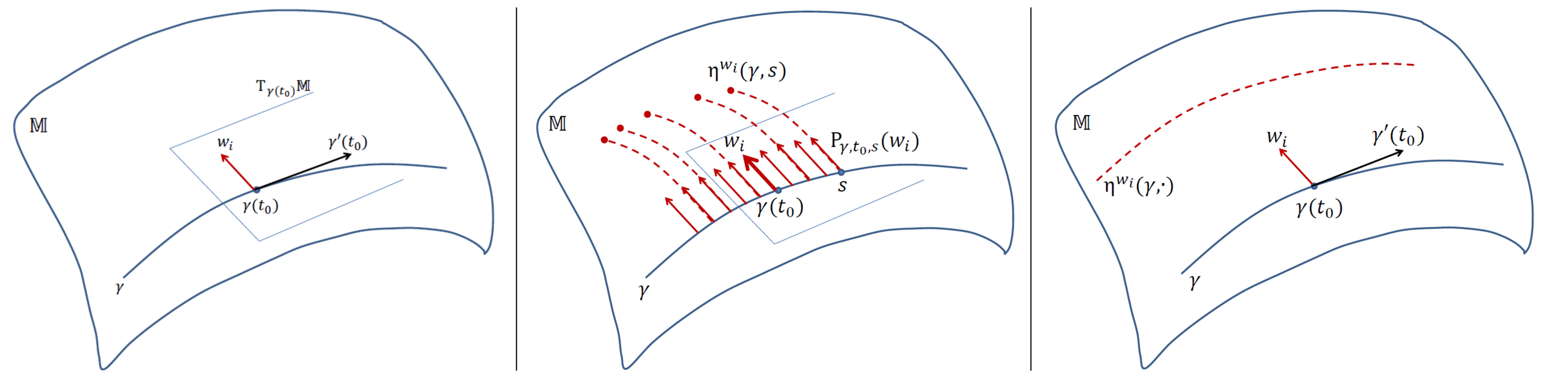
Summary : we propose a generic mixed-effects model for longitudinal manifold-valued data. The model allows to estimate an average trajectory as well as individual trajectories. Random effects allow to characterize changes in direction and pace at which individual trajectories are followed. This generic model is used to analyze the temporal progression of a family of univariate biomakers.

- $(\tilde{M}, g^{\tilde{M}})$ smooth Riemannian manifold included in $\mathbf{R}^n$
- (M, g^M) sub-Riemannian manifold of $\tilde{M}$, assumed to be geodesically complete
- $\boldsymbol{p} \in M$, $\boldsymbol{v} \in \mathrm{T}_{\boldsymbol{p}}M$, $\mathrm{Exp}_{\boldsymbol{p}}{}^M(\boldsymbol{v})$: Riemannian exponential in M at $\boldsymbol{p}$ of the tangent vector $\boldsymbol{v}$
- $\boldsymbol{\gamma} : \mathbf{R} \to M$: geodesic of M
- $t, t_0 \in \mathbf{R}$, $\mathrm{P}_{\boldsymbol{\gamma},t_0,t}(\cdot)$: parallel transport in M along $\boldsymbol{\gamma}$ from $\boldsymbol{\gamma}(t_0)$ to $\boldsymbol{\gamma}(t)$.
- $t \mapsto \mathrm{Exp}^M_{\boldsymbol{p},t_0}(\boldsymbol{v})(t)$; geodesic of M which goes through $\boldsymbol{p}$ at time t_0 with velocity $\boldsymbol{v}$.

A hierarchical model :

Average trajectory : $t \mapsto \boldsymbol{\gamma}(t)$ ⟶ Trajectory of the i-th individual : $t \mapsto \boldsymbol{\gamma}_i(t) = \eta^{\boldsymbol{w}_i}(\boldsymbol{\gamma}, \psi_i(t))$ ⟶ Observations : $\boldsymbol{y}_{i,j} = \boldsymbol{\gamma}_i(t_{i,j}) + \boldsymbol{\varepsilon}_{i,j}$

1. The average trajectory $t \mapsto \boldsymbol{\gamma}(t)$ is choosen to be the geodesic $t \mapsto \mathrm{Exp}^M_{\boldsymbol{p}_0,t_0}(\boldsymbol{v}_0)(t)$, $\boldsymbol{p}_0 \in M$, $\boldsymbol{v}_0 \in \mathrm{T}_{\boldsymbol{p}_0}M$
2. The trajectory of the i-th individual is obtained in two steps. We start by constructing the **parallel shift** of the average trajectory by using a tangent vector $\boldsymbol{w}_i$, which we choose orthogonal to $\boldsymbol{v}_0$.



3. The trajectory $t \mapsto \boldsymbol{\gamma}_i(t)$ is then obtained by **reparametrizing in time** the parallel shift $\eta^{\boldsymbol{w}_i}(\boldsymbol{\gamma}, \cdot)$ using the affine time reparametrization $\psi_i : t \mapsto \alpha_i(t - t_0 - \tau_i) + t_0$. This allows to account for the variability in stages of disease progression across the population.

The operation of parallel shifting, on the manifold M, using a tangent vector, is defined as follows :

Definition : $\boldsymbol{w} \in \mathrm{T}_{\boldsymbol{\gamma}(t_0)}M$, $\boldsymbol{w} \neq 0$. The curve $s \mapsto \eta^{\boldsymbol{w}}(\boldsymbol{\gamma}, s)$ defined by ::

$$\eta^{\boldsymbol{w}}(\gamma, s) = Exp_{\gamma(s)}\left(P_{\gamma,t_0,s}(\boldsymbol{w})\right), \; s \in \mathbf{R}.$$

is said to be the « parallel shift of γ » using $\boldsymbol{w}$.

By virtue of the tubular neighborhood theorem [Hirsch M.W., 2012], parallel shifting defines a local spatio-temporal coordinate system.

The model :

$$\boxed{y_{i,j} = \eta^{w_i}\left(\boldsymbol{\gamma}, \psi_i(t_{i,j})\right) + \varepsilon_{i,j}.}$$

where $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ and :

- α_i is a subject-specific **acceleration factor**, $\alpha_i = \exp(\xi_i)$ and $\xi_i \sim N(0, \sigma_\xi{}^2)$
- τ_i is a subject-specific **time shift**, $\tau_i \sim N(0, \sigma_\tau{}^2)$
- $\boldsymbol{\varepsilon}_{i,j} \sim N(0, \sigma^2)$
- $\boldsymbol{w}_i$ is a subject-specific **space shift** :

$$\boldsymbol{w}_i = \mathbf{A}\boldsymbol{s}_i \text{ and : } \forall\, 1 \leq j \leq N_s, s_{l,i} \sim Laplace\left(\tfrac{1}{2}\right).$$

In the spirit of Independent Component Analysis, the space shift $\boldsymbol{w}_i$ appears as a linear combination of N_s independent components, namely the columns of the matrix $\mathbf{A}$.

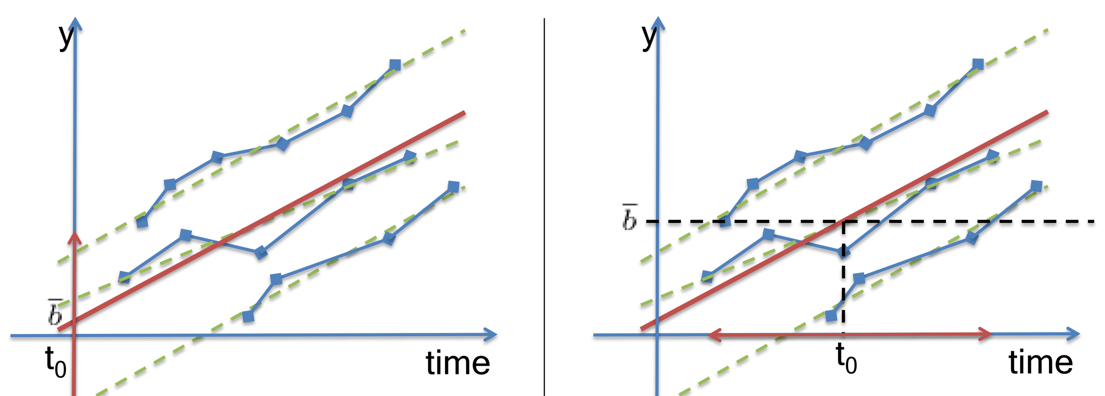
Three particular cases of our generic spatio-temporal model

- **« Straight lines » model** [Schiratti et al., IPMI 2015]

✓ $M = \mathbf{R}$ (equipped with the canonical metric)

✓ Geodesics are straight lines

$$y_{i,j} = p_0 + \alpha_i v_0(t_{i,j} - t_0 - \tau_i) + \varepsilon_{i,j}.$$



$y_{i,j} = (\bar{a} + a_i)(t_{i,j} - t_0) + \bar{b} + b_i + \varepsilon_{i,j}$ [Laird & Ware, 1982]

$y_{i,j} = (\bar{a} + a_i)(t_{i,j} - t_0 - \tau_i) + \bar{b} + \varepsilon_{i,j}$ Our approach

- **« Logistic curves » model** [Schiratti et al., IPMI 2015]

✓ $M =]0,1[$, equipped with the metric $g = (g_p)_{p \in]0,1[}$, $g_p(u,v) = uv \,/\, p^2(1-p)^2$

✓ Geodesics are logistic curves

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

Note that this model is not equivalent to a linear model on the logit of the observations : the logit transform corresponds to the Riemannian logarithm at $p_0 = 0{,}5$. The model written in the tangent space is still *not* linear due to the multiplication between the random effects $\alpha_i = \exp(\xi_i)$ and τ_i.

- **A progression model for a familiy of univariate biomarkers :**

Aim : we want to analyze the temporal progression of a family of N biomarkers.

✓ We assume that the measurements of each biomarker belong to a one-dimensional Riemannian manifold I, geodesically complete and included in $\mathbf{R}$. As a consequence, M is a product of one-dimensional manifolds : $M = I^N = I \times I \times \cdots \times I$.

✓ The average trajectory $t \mapsto \boldsymbol{\gamma}(t)$ is choosen among a parametric family of geodesics of M :

$$\gamma_\delta(t) = (\gamma_0(t), \gamma_0(t + \delta_1), \ldots, \gamma(t + \delta_{N-1}))$$

where the parameters δ_i $(1 \leq i \leq N-1)$ correspond to the relative delay between the biomarkers and $t \mapsto \gamma_0(t)$ is a geodesic of the one-dimensional Riemannian manifold I (straight line, logistic curve, …)

✓ If the space of observations is the open interval $]0,1[$ for all the biomarkers (we consider normalized measurements), the manifold M is $]0,1[^N$, equipped with the product metric (see « Logistic curves model »).

Writing the generic spatio-temporal model in this case leads to a progression model for normalized biomarkers named « multivariate logistic curves model ». This model is given by :

$$y_{i,j,k} = \left(1 + \left(\frac{1}{p_0} - 1\right)\exp\left(-\frac{\alpha_i v_0(t_{i,j} - t_0 - \tau_i) + v_0\delta_k + v_0\frac{(\mathbf{A}s_i)_k}{\gamma'(t_0 + \delta_k)}}{p_0(1-p_0)}\right)\right)^{-1} + \varepsilon_{i,j,k}.$$

where $y_{i,j,k}$ = measurement of the k-th biomarker for individual i, at time $t_{i,j}$.

Estimation of the parameters of the model

The parameters of the generic spatio-temporal model are $\boldsymbol{\theta} = (p_0, t_0, v_0, \boldsymbol{\delta}, \sigma_\xi, \sigma_\tau, \sigma, \mathrm{vec}(\mathbf{A}))$.

Summary : the parameters are estimated using a stochastic version of the EM algorithm [Dempster, Laird, Rubin, 1977]. This algorithm is the Monte Carlo Markov Chain Stochastic Approximation EM algorithm (MCMC-SAEM) [Delyon et al.,1999 ; Allassonnière et al.,2010].Theoretical results regarding the convergence of the algorithm have been proved in [Delyon et al.,1999 ; Allassonnière et al.,2010].

Note that the MCMC-SAEM requires that the model belongs to the *curved exponential family* . However, the multivariate logistic curves model does not belong to this family. The model can be made exponential by considering each parameters as realizations of independent Gaussian random variables.

Overview of the MCMC-SAEM for the multivariate logistic curves model :

$\boldsymbol{z}^{(k)}$ (resp. $\boldsymbol{\theta}^{(k)}$) denotes the vector of hidden variables (resp. parameters) at the k-th iteration.

- Initialisation :

$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}^{(0)}$, $\boldsymbol{z}^{(0)} \leftarrow$ random, $\mathbf{S} \leftarrow 0$, $(\varepsilon_k)_k$ sequence of positive step-sizes

repeat until convergence

Simulation (Hasting-Metropolis within Gibbs sampler) :

$$\boldsymbol{z}^{(k+1)} \leftarrow \text{Gibbs sampler}(\boldsymbol{z}^{(k)}, \boldsymbol{y}, \boldsymbol{\theta}^{(k)})$$

- Compute the sufficent statistics : $\mathbf{S_1}^{(k)} \leftarrow [\boldsymbol{y}_{i,j}{}^\top \boldsymbol{f}_{i,j}]_{i,j}$, $\mathbf{S_2}^{(k)} \leftarrow [\|\boldsymbol{f}_{i,j}\|^2]_{i,j}$, $\mathbf{S_3}^{(k)} \leftarrow [(\boldsymbol{\xi}_i{}^{(k)})^2]_i$, $\mathbf{S_4}^{(k)} \leftarrow [(\boldsymbol{\tau}_i{}^{(k)})^2]_i$, $\mathbf{S_5}^{(k)} \leftarrow p_0{}^{(k)}$, $\mathbf{S_6}^{(k)} \leftarrow t_0{}^{(k)}$, $\mathbf{S_7}^{(k)} \leftarrow v_0{}^{(k)}$, $\mathbf{S_8}^{(k)} \leftarrow [\delta_j{}^{(k)}]_j$, $\mathbf{S_9}^{(k)} \leftarrow [\beta_j{}^{(k)}]_j$
- **Stochastic approximation :**

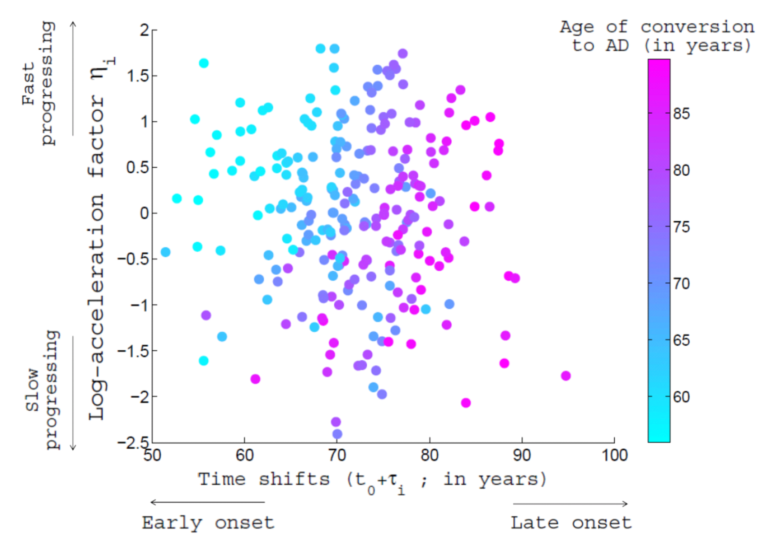
$\mathbf{S}_j{}^{(k+1)} \leftarrow \mathbf{S}_j{}^{(k)} + \varepsilon_k\left(\mathbf{S}(\boldsymbol{y}, \boldsymbol{z}^{(k)}) - \mathbf{S}_j{}^{(k)}\right)$ for all j

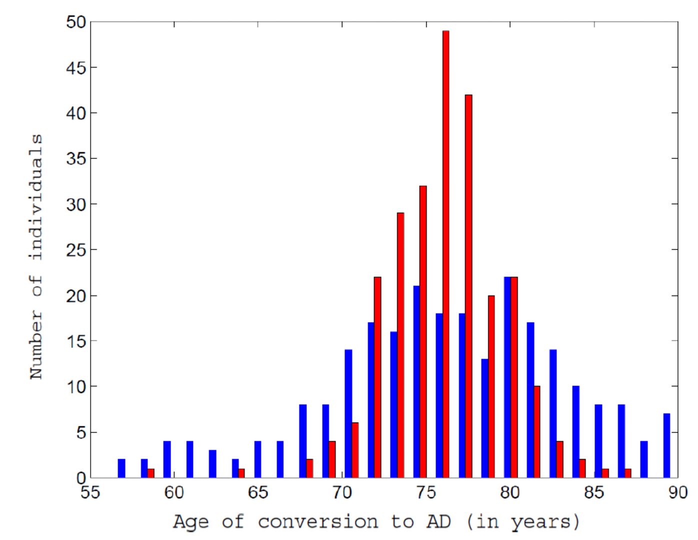
- **Maximization** $(\boldsymbol{\theta}^{(k+1)} \leftarrow \mathrm{argmax}_{\boldsymbol{\theta} \in \Theta}\left[-\phi(\boldsymbol{\theta}) + \langle \mathbf{S}^{(k+1)}, \boldsymbol{\psi}(\boldsymbol{\theta}) \rangle\right])$: closed-form updates

end repeat

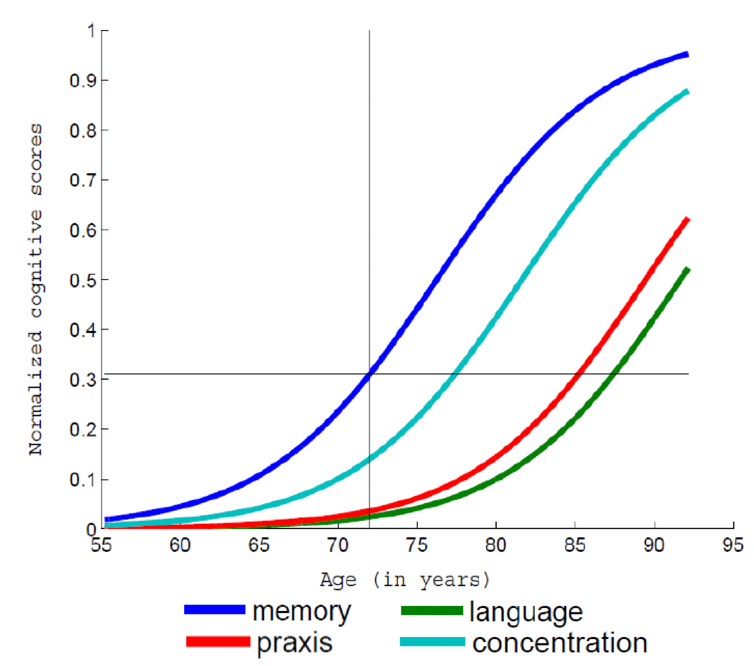
Experimental results

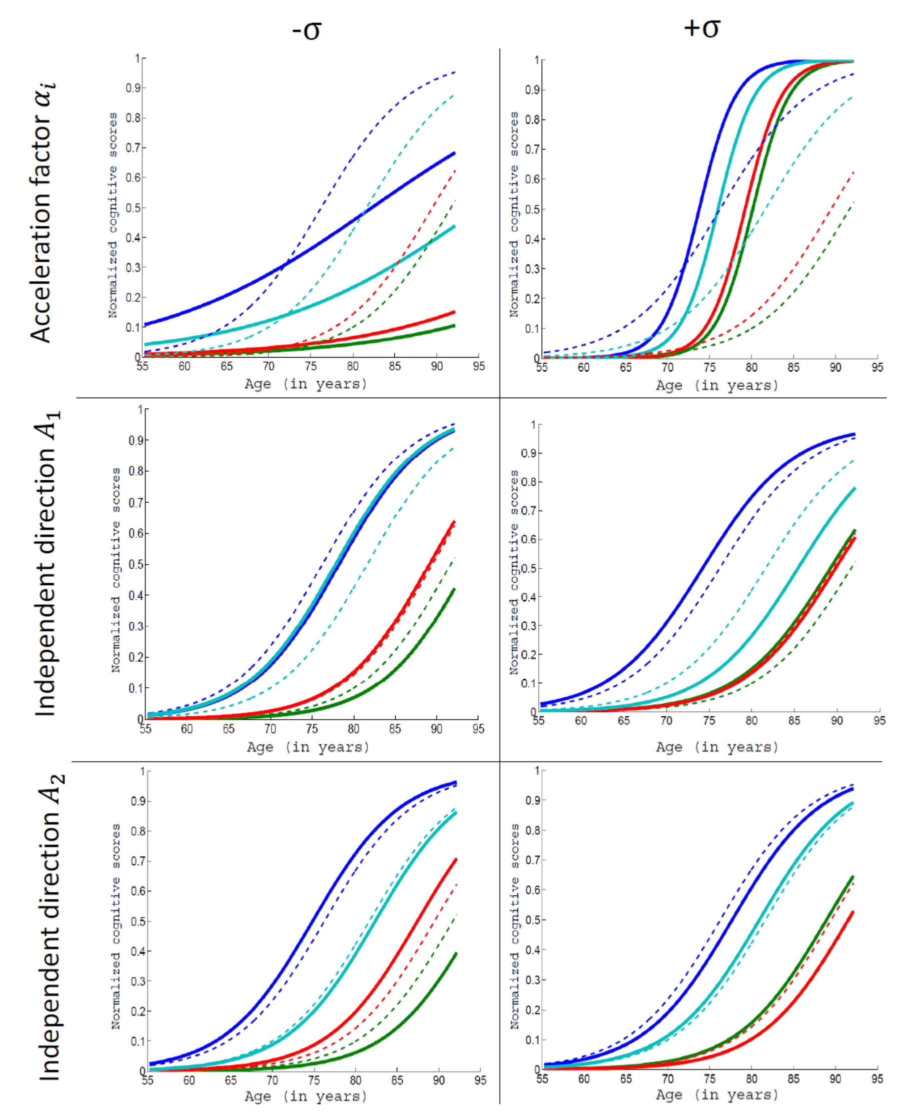
- **Data** : Normalized cognitive scores grouped into four categories (biomarkers) : memory (5 items), language (5 items), praxis (2 items), concentration (1 item). Data collected from the ADNI database for 248 MCI patients who converted to AD. Each observation is a point in $M =]0,1[^4$.





Left : the plot of individual random effects show that the time-shifts correspond well with the age at which individuals converted to AD. **Right** : histogram of the ages of conversion to AD ($(t_i{}^{\mathrm{conv}})_i$) in blue and histobram of the normalized ages ($(\psi_i(t_i{}^{\mathrm{conv}}))_i$; ages of conversion mapped from the individual timeline to the reference timeline using the subject-specific affine reparametrization. The individual time reparametrizations correctly register the dynamics of individual trajectories.





Right: The figure illustrates the effect of the variance of the acceleration factors, and the two estimated independent components ($N_s = 2$). Among the studied population, individuals evolve from 7 times faster or slower than the average trajectory. Individual space shifts along the first (or second) independent components may change the relative order between the cognitive functions.

Left : the estimated avergae trajectory i($p_0 = 0{,}3$ and $t_0 = 72$) The score associated to memory will reach the value 0,3 at 72 years old, followed by concentration which reaches the same value 5 years later and then praxis and language at 85 and 87 years old respectively.