Longitudinal models for Parkinson's Disease

Raphael Couronné[1,3]
Stanley Durrleman[1,3], **Stéphane Lehéricy**[1,2], **Marie Vidhaillet**[1,2]

1. ARAMIS LAB, Inserm U1127, CNRS UMR 7225, UPMC, ICM, Inria, Paris, France
2. AP-HP, Departments of Neurology and Neuroradiology, Paris, France
3. INRIA Paris

Raphael.couronne@inria.fr

• Introduction

Aim : model the progression of Parkinson's disease

- Gain understanding the progression of Parkinson's disease
  - Infer a template of the average disease progression
- Infer each patient disease stage and future evolution
  - Two individuals at the same age might be at different stages of disease progression
    → Age is modeled as a random variable and not a covariable

Tools and Data

- We use a non linear mixed-effect model to estimate the temporal progression of the biomarkers
- We use multimodal data regrouping cognitive scores, motor scores and imaging from the international PPMI cohort on Parkinson's Disease.

• Methods

Generic spatio-temporal model for longitudinal data

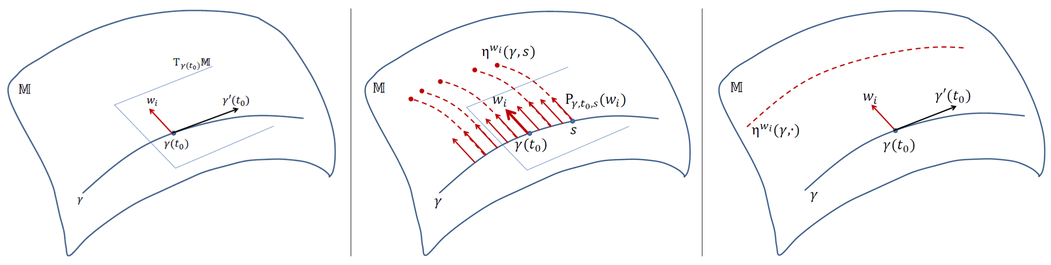
Summary : we use a generic mixed-effects model for longitudinal manifold-valued data [Schiratti et al, 2015]. The model allows to estimate an average trajectory as well as individual trajectories. Random effects characterize changes in direction and pace at which individual trajectories are followed. This generic model is used to analyze the temporal progression of a family of univariate biomakers.

➢ 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 :** $\boldsymbol{y}_{i,j} = \boldsymbol{\gamma}_i(t_{i,j}) + \boldsymbol{\varepsilon}_{i,j}$

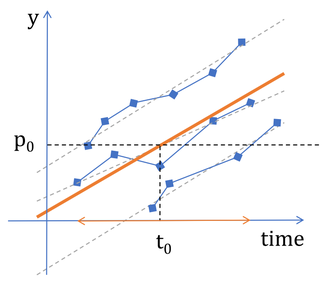
1. The average trajectory $t \mapsto \boldsymbol{\gamma}(t)$ is choosen to be the geodesic $\mathrm{t} \mapsto \mathrm{Exp}^{M}_{\boldsymbol{p_0},t_0}(\boldsymbol{v_0})(t)$, $\boldsymbol{p}_0 \in M$, $\boldsymbol{v}_0 \in \mathrm{T}_{\boldsymbol{p}_0}M$

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.

- $(\widetilde{M}, g^{\widetilde{M}})$ smooth Riemannian manifold included in $\mathbf{R}^n$
- (M, g^M) sub-Riemannian manifold of $\widetilde{M}$, 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 $\boldsymbol{\gamma}(t)$.
- $\mathrm{t} \mapsto \mathrm{Exp}^{M}_{\boldsymbol{p},t_0}(\boldsymbol{v})(t)$; geodesic of M which goes through $\boldsymbol{p}$ at time t_0 with velocity $\boldsymbol{v}$.



The model in practice for scalar longitudinal data

Univariate Linear Model
$M = \mathbb{R}$ (equipped with canonical metric)

$$y_{i,j} = p_0 + \alpha_i v_0\left(t_{i,j} - t_0 - \tau_i\right) + \epsilon_{i,j}$$

Univariate Logistic Model
$M =]0,1[$ with the metric $\mathrm{g} = \left(\mathrm{g_p}\right)_{(\mathrm{p}\in]0,1[}$, $\mathrm{g_p}(\mathrm{u},\mathrm{v}) = uv/p^2(1-\mathrm{p})^2$

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

Aim : 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 $t \mapsto \gamma(t)$ is choosen among a parametric family of geodesics of M : $\gamma(t) = \left(\gamma_0(t), \gamma_1(t), \ldots, \gamma_{N-1}(t)\right)$

Multivariate Linear Model

$$y_{i,j,k} = p_k + v_k \alpha_i \left(t_{i,j,k} - t_0 - \tau_i\right) + (\mathbf{A}\boldsymbol{s}_i)_k + \epsilon_{i,j,k}$$

Multivariate Logistic Model

$$y_{i,j,k} = \left(1 + \left(\frac{1}{p_k} - 1\right)\exp\left(-\frac{\alpha_i v_k\left(t_{i,j} - t_0 - \tau_i\right) + (\mathbf{A}\boldsymbol{s}_i)_k}{p_k(1-p_k)}\right)\right)^{-1} + \epsilon_{i,j,k}$$

✓ Missing values can be disgarded as we have a generative model

Estimation of model parameters

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

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

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

- Initialisation :
  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}^{(0)}$, $\boldsymbol{z}^{(0)} \leftarrow$ random, $\mathbf{S} \leftarrow 0$, $\left(\varepsilon_p\right)_p$ sequence of positive step-sizes

  repeat until convergence

  Simulation (Hasting-Metropolis within Gibbs sampler) :

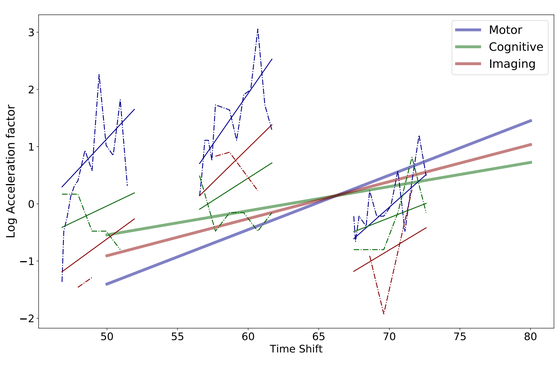
  $$\boldsymbol{z}^{(p+1)} \leftarrow \text{Gibbs sampler}(\boldsymbol{z}^{(p)}, \boldsymbol{y}, \boldsymbol{\theta}^{(p)})$$

  - Compute the sufficent statistics :
    $\mathbf{S}_1^{(p)} \leftarrow [\boldsymbol{y}_{i,j}^{(p)\top} \boldsymbol{f}_{i,j}]_{i,j}$, $\mathbf{S}_{22}^{(p)} \leftarrow [\|\boldsymbol{f}_{i,j}\|_2^2]_{i,j}$
    $\mathbf{S}_3^{(p)}, \mathbf{S}_4^{(p)} \leftarrow [\left(\boldsymbol{\xi}_i^{(p)}\right)]_i, [\left(\boldsymbol{\tau}_i^{(p)}\right)]_i$
    $\mathbf{S}_5^{(p)}, \mathbf{S}_6^{(p)}, \mathbf{S}_7^{(p)}, \mathbf{S}_7^{(p)}, \mathbf{S}_8^{(p)} \leftarrow p_0^{(p)}, t_0^{(p)}, v_0^{(p)}, [\beta_j^{(p)}]_j$
  - Stochastic approximation :
    $\mathbf{S}_{\mathrm{j}}^{(p+1)} \leftarrow \mathbf{S}_j^{(p)} + \varepsilon_k\left(\mathbf{S}(\boldsymbol{y}, \boldsymbol{z}^{(p)}) - \mathbf{S}_j^{(p)}\right)$ for all j
  - Maximization $(\boldsymbol{\theta}^{(k+1)} \leftarrow \mathrm{argmax}_{\boldsymbol{\theta}\in\Theta}\left[-\phi(\boldsymbol{\theta}) + \langle \mathbf{S}^{(p+1)}, \boldsymbol{\psi}(\boldsymbol{\theta})\rangle\right])$: closed-form updates

  end repeat

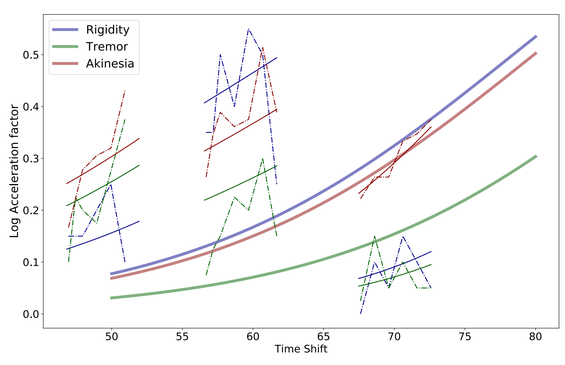
• Results

Multivariate Linear Model



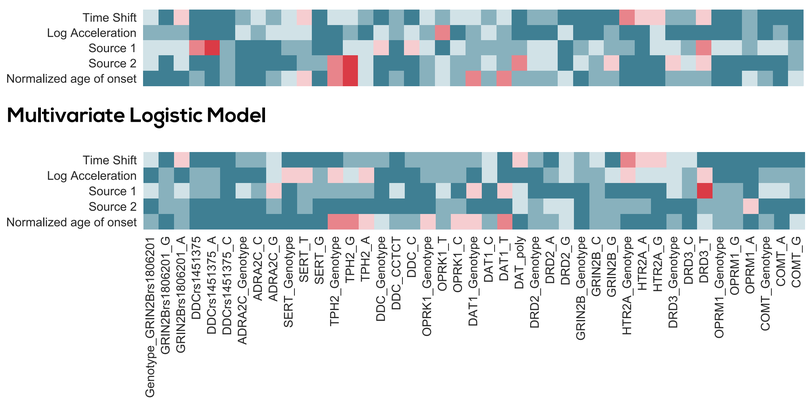
Multivariate linear model for motor, sognitive and imaging scores. In wide the mean template of the model, and for 3 patients the model and real scores respectively in plain and dotted lines.

Multivariate Logistic Model



Multivariate logistic model for 3 different motor score linked with the MDS-UPDRS normalized between 0 and 1. In wide the mean template of the model, and for 3 patients the model and real scores

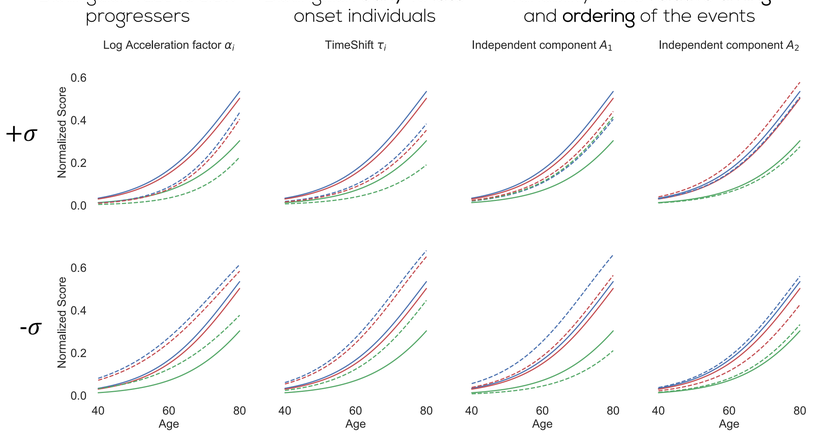
Multivariate Linear Model



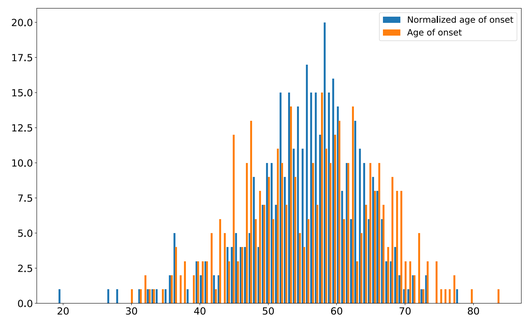
Left : Correlation between both computed models and gene data. We seek to find correlations between covariables of interest and model output to gain understanding of the disease.

Multivariate Logistic Model

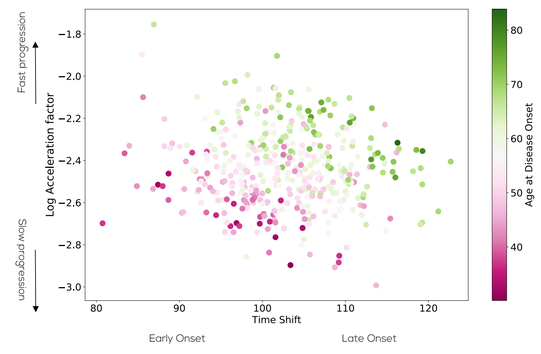
Distinguish **fast** vs **slow** progressers | Distinguish **early** vs **late** onset individuals | Variability in the **relative timing** and **ordering** of the events



Right: The figure illustrates the effect of the variance of the acceleration factors, and the two estimated independent components ($N_s = 2$). Individual space shifts along the first (or second) independent components may change the relative order between the motor modalities..



Histogram of the ages of conversion to PD ($(t_i^{\mathrm{conv}})_i$) in orange and histogram of the normalized ages $\left(\left(\psi_i\left(t_i^{\mathrm{conv}}\right)\right)_i\right)$; ages of conversion mapped from the individual timeline to the reference timeline using the subject-specific affine reparametrization.



Plot of temporal individual parameters against the age of conversion to PD. The Time shifts correlates well with the age of conversion ($\rho = 0.6$).

• Conclusion

We use a mixed-effect model which is able to evaluate a group-average spatiotemporal propagation of biomarkers. Moreover, individual parameters characterizing personalized patterns of propagation as variations of the mean scenario are estimated. The evaluation of this model is made with the MCMC-SAEM algorithm.
Applied on Parkinson's disease, we model a mean template of the disease evolution, while being customizable to fit individual data, predicting stage of the disease or time to symptom onset. We wish now to use this information to be able to correlate the model ouput with covariables of interest such as gene data, and separate patients in distincts clusters caracterizing different subtypes of the disease.

• References

[1] J.-B. Schiratti, S. Allassonnière, O. Colliot, S. Durrleman. Learning spatiotemporal trajectories from manifold-valued longitudinal data. *Advances in Neural Information Processing Systems* (2015)
[2] S. Durrleman, X. Pennec, A. Trouvé, J. Braga, G. Gerig, N. Ayache. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *International Journal of Computer Vision* (2013)
[3] B. Delyon, M. Lavielle, E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics* (1999)
[4] S. Allassonnière, E. Kuhn, A. Trouvé. Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli* (2010)