Mixed effect model for the spatiotemporal analysis of longitudinal manifold valued data

Stéphanie Allassonnière

with J.B. Schiratti, J. Chevallier, I. Koval, V. Debavalaere and S. Durrleman

Université Paris Descartes & Ecole Polytechnique

Computational Anatomy

- Represent and analyse **geometrical** elements upon which **deformations** can **act**
- Describe the observed objects as **geometrical variations** of one or several representative elements
- **Quantify** this variability inside a population

Deformable template model from Grenander

- How does the deformation act?
- What is a representative element?
- How to quantify the geometrical variability ?

Computational Anatomy

One solution:

- Quantify the distance between observations using deformations
- Provide a **statistical model** to approximate the generation of the observed population from the atlas
- Propose a **statistical learning algorithm**
- Optimise the numerical estimation

Bayesian Mixed Effect model

- First model :
  - **One observation per subject**
  - **Image or shape (viewed as currents)**
  - **Deformations either linearized or diffeomorphic**
  - **Homogeneous or heterogeneous populations (mixture models)**

Bayesian Mixed Effect model

Template

Deformation

Observations

$$\begin{cases} y_i & \sim & \mathcal{N}(I_0 \circ \phi_i^{-1}, \sigma^2 Id) \\ \phi_i & = & f(\beta_i) \\ \beta_i & \sim & \mathcal{N}(0, \Gamma_g) \end{cases}$$

deterministic parametrisation

Normal distribution with full covariance matrix

For $1 \leqslant i \leqslant n$ subjects,

Bayesian Mixed Effect model

Template
- Grey level images
- Probability maps

Deformation
- Linearised
- Diffeomorphic

Observations
- T1
- DWIs
- fMRI
- T1+fMRI

$$\begin{cases} y_i & \sim & \mathcal{N}(I_0 \circ \phi_i^{-1}, \sigma^2 Id) \\ \phi_i & = & f(\beta_i) \\ \beta_i & \sim & \mathcal{N}(0, \Gamma_g) \end{cases}$$

deterministic parametrisation

Normal distribution with full covariance matrix

Mixtures of all these models

For $1 \leqslant i \leqslant n$ subjects,

Bayesian Mixed Effect model

- First model :
  - **One observation per subject**
  - **Image or shape (viewed as currents)**
  - **Deformations either linearized or diffeomorphic**
  - **Homogeneous or heterogeneous populations (mixture models)**

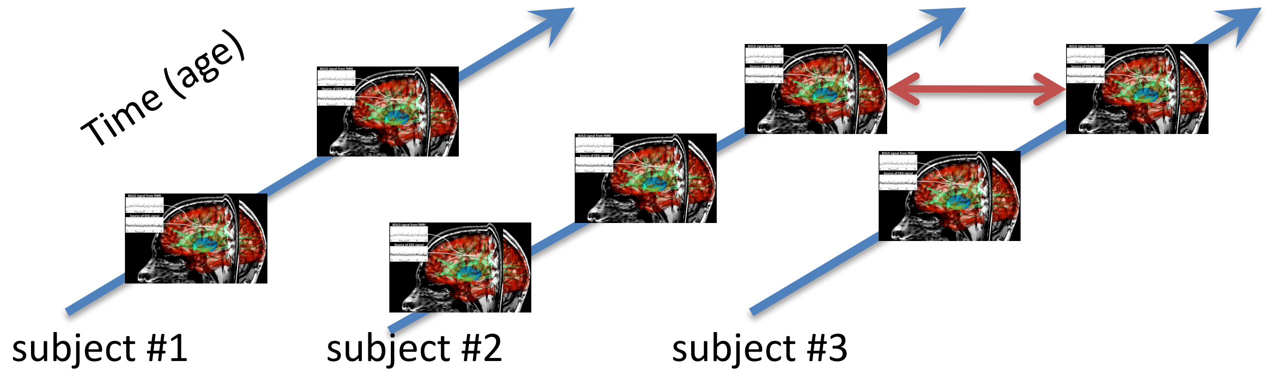
➢ **Limitations**

- ➢ **One observation per subject**
- ➢ **Corresponding acquistion time**

Longitudinal Data Analysis

- Longitudinal model :
  - **Several observation per subject**
  - **Image, shape, etc**
  - **Atlas = representative trajectory and population variability**

Longitudinal Data Analysis



How to learn representative trajectories of data changes from longitudinal data?

Temporal marker of progression
(e.g. time since drug injection, seeding, birth, etc..)

Regression
(e.g. compare measurements at same time-point)

vector-valued data

Linear mixed-effects models
[Laird&Ware'82, Diggle et al., Fitzmaurice et al.]

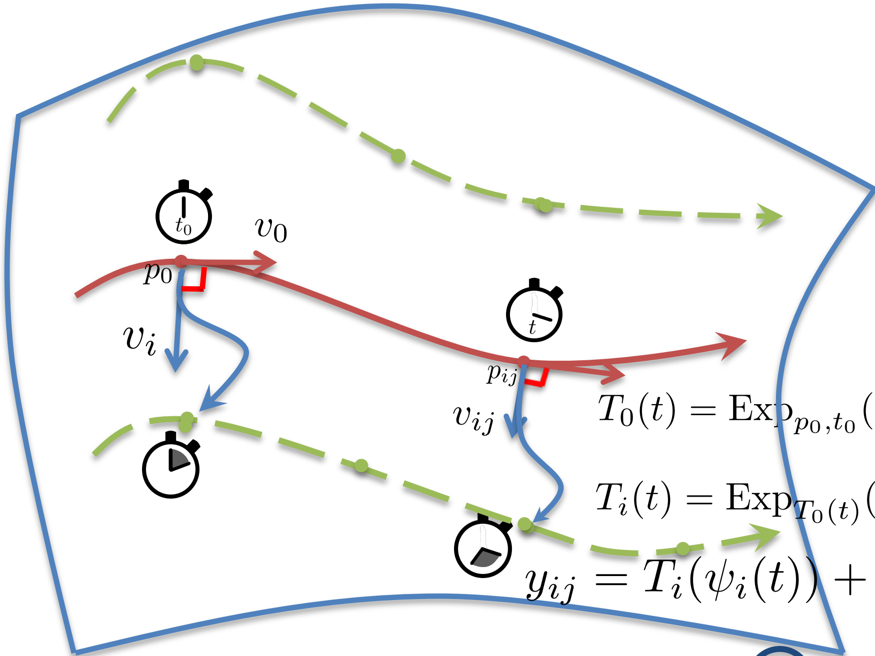
No temporal marker of progression
(e.g. in aging, neurodegenerative diseases, etc..)

Learning **spatiotemporal distribution** of trajectories
- *Find temporal correspondences*
- *Compare data at corresponding stages of progression*

manifold-valued data
(normalized data, positive matrices, shapes, etc.)

Needs to disentangle differences in:
- *Measurements*
- *Dynamics of measurement changes*

Spatiotemporal Statistical Model



$$T_0(t) = \mathrm{Exp}_{p_0,t_0}(v_0)(t)$$

$$T_i(t) = \mathrm{Exp}_{T_0(t)}(\mathrm{P}^{T_0}_{t_0,t}(v_i))$$

$$y_{ij} = T_i(\psi_i(t)) + \varepsilon_{ij}$$

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

- Statistical model inclinding:
  - a **representative trajectory** of data changes
  - **spatiotemporal variations** in:
    - measurement values
    - pace of measurement changes
- Orthogonality condition ensures identifiability (unique space/time decomposition)
- Time is not a covariate but a random variable

	Acceleration factor	Time-shift	Space-shift
Random effects:	$\alpha_i \sim log\,\mathcal{N}(0, \sigma_\alpha^2)$	$\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$	$v_i = (A_1 \vert \dots \vert A_K)\, s_i$ $A_k \perp v_0$

Fixed effects: (p_0, t_0, v_0) and $(\sigma_\alpha^2, \sigma_\tau^2, A_1, \dots A_K)$

[Schiratti et al. IPMI’15, NIPS’15]

Spatiotemporal Statistical Model

$$
\left\{
\begin{array}{l}
y_{ij} = T_i(\psi_i(t)) + \varepsilon_{ij} \\
T_i(t) = \mathrm{Exp}_{T_0(t)}(\mathrm{P}_{t_0,t}^{T_0}(v_i)) \\
T_0(t) = \mathrm{Exp}_{p_0,t_0}(v_0)(t) \\
\psi_i(t) = t_0 + \alpha_i(t - t_0 - \tau_i) \\
\alpha_i \sim log\, \mathcal{N}(0, \sigma_\alpha^2) \\
\tau_i \sim \mathcal{N}(0, \sigma_\tau^2) \\
v_i = (A_1|\ldots|A_K)\, s_i \\
A_k \perp v_0 \\
(p_0, t_0, v_0) \\
(\sigma_\alpha^2, \sigma_\tau^2, A_1, ...A_K)
\end{array}
\right.
$$

Submanifold value observations

Parallel curve

Representative trajectory

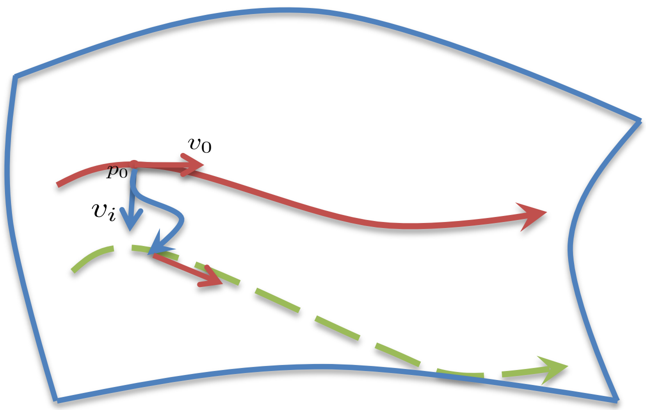
Linear time reparametrization

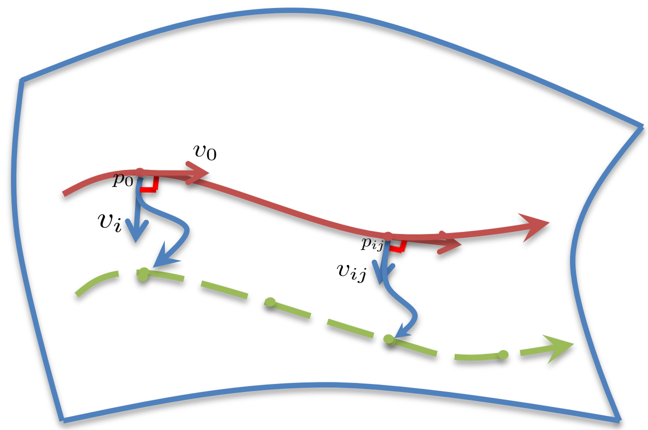
Hidden random variables:
Acceleration factor
Time shift
Space shift

Parameters:
Mean trajectory parametrization
and
prior parameter

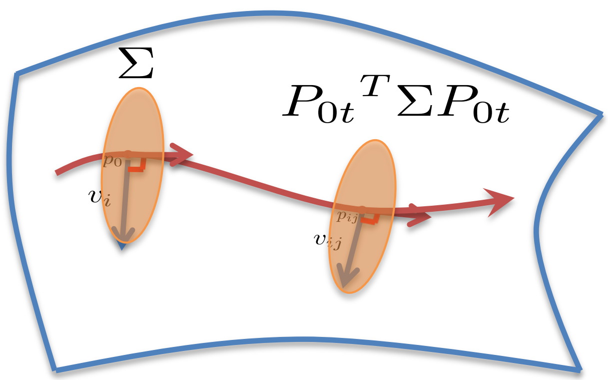
Spatiotemporal Statistical Model

Comparison with previous work:



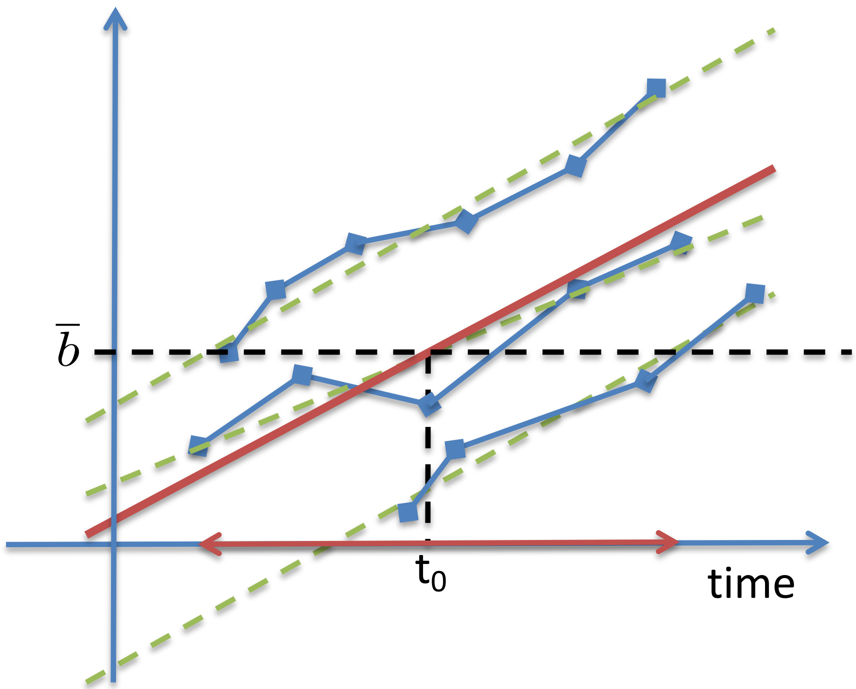


Interest: Parallel transport keep invariant the structure of the distribution, but updated it in time



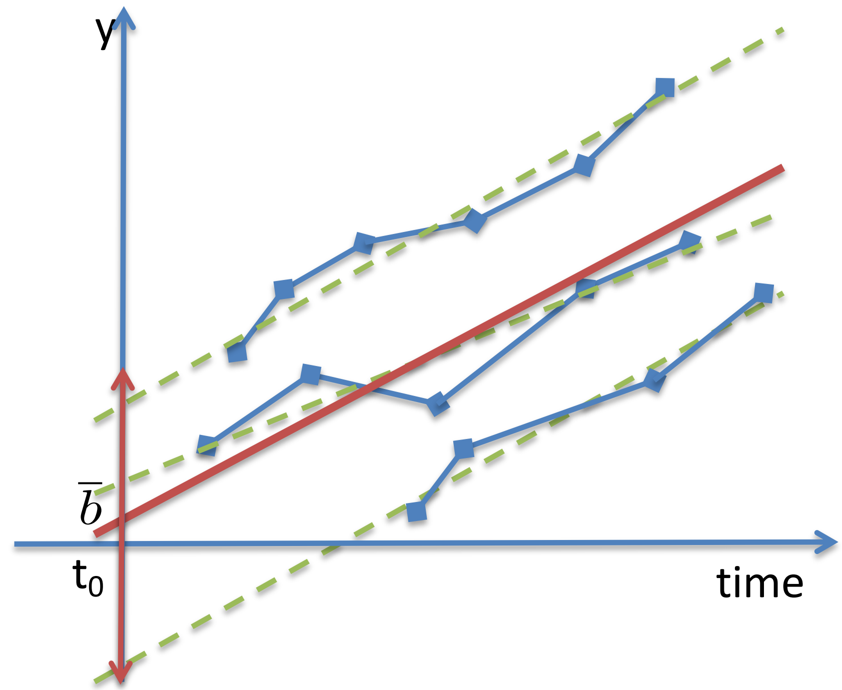
Spatiotemporal Statistical Model

- The straight line model $\mathbb{M} = \mathbb{R}$



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

Time at which measurement of the ith subject reaches $\bar{b}$



$$y_{ij} = (\bar{a} \ast a_i)(t_{i,j} - t_0) + \underbrace{\bar{b} + b_i} + \varepsilon_{i,j}$$

Measurement of the ith subject at time t_0

Schiratti et al. (2015)

Laird & Ware (1982)

Spatiotemporal Statistical Model

- The logistic curve model $\mathbb{M} =]0,1[, \quad g(p)(u,v) = \dfrac{uv}{p^2(1-p)^2}$

  - Geodesic are logistic curves

$$\gamma_0(t) = 1 + \frac{(1-p_0)/p_0}{\exp\left(-\frac{v_0}{p_0(1-p_0)}(t-t_0)\right)} \qquad\qquad y_{ij} = \gamma_0\Big(t_0 + \alpha_i(t - t_0 - \tau_i)\Big) + \varepsilon_{ij}$$

  - It is *not* equivalent to a linear model on the logit of the observations (i.e. the Riemannian log at p_0 = 0.5), since p_0 is estimated

  - If we fix p_0 = 0.5 in our model → end up with our previous linear case (different from Laird&Ware)

Spatiotemporal Statistical Model

- The propagation model $\mathbb{M} =]0,1[^N, \; g(p)(u,v) = \sum_{k=1}^{N} \frac{u_k v_k}{p_k^2(1-p_k)^2}$

  - Geodesics are logistic curves in each coordinate

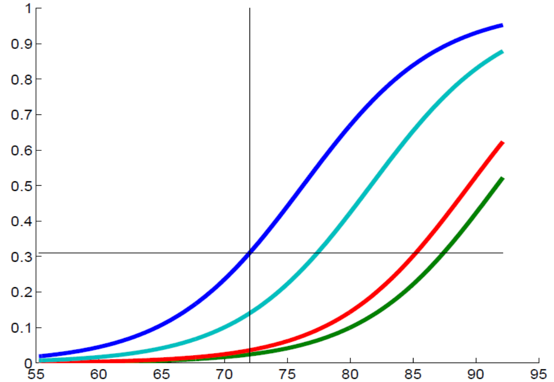
  - Parametric family of geodesics seen as a model of propagation of an effect

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

  - The parallel curve in the direction of the space-shift v_i writes

$$\left(\gamma_0\left(t + \frac{v_{i,1}}{v_0}\right), \gamma_0\left(t - \delta_1 + \frac{v_{i,2}}{v_0}\right), \ldots, \gamma_0\left(t - \delta_{N-1} + \frac{v_{i,N}}{v_0}\right)\right)$$

→ The parallel changes the *relative timing* of the effect onset across coordinates



Parameter Estimation

$$y = (y_1, ..., y_N),\ \ z = (z_1, ... z_N),\ \ \theta = (\sigma_z^2, \sigma_\varepsilon^2, A_1, ..., A_K, p_0, t_0, v_0)$$

- Maximum Likelihood:

$$max_\theta \, p(y|\theta) = \int p(y, z|\theta) dz$$

- EM: $\theta_{k+1} = argmax_\theta \sum_{i=1}^{N} \int \log \left(\underbrace{p(y_i, z_i|\theta)}_{p(y_i|z_i,\theta)p(z_i|\theta)} \right) p(z_i|y_i, \theta_k) dz_i$

- Distribution from the **curved exponential family**

$$\log p(y_i, z_i|\theta) = \phi(\theta)^T S(y_i, z_i) - log(C(\theta))$$

$$\theta_{k+1} = argmax_\theta \left\{ \phi(\theta)^T \sum_{i=1}^{N} \int S(y_i, z_i) p(z_i|y_i, \theta_k) dz_i - N \log(C(\theta)) \right\}$$

Parameter Estimation: stochastic algorithm

- **SA-EM**: replaces integration by one simulation of the hidden variable:
  sample $z_{i,k+1}$ from $p(z_i|y_i,\theta_k)$,
  and a stochastic approximation of the sufficient statistics

$$\overline{S}_{k+1} = (1-\Delta_k)\overline{S}_k + \Delta_k \left(\frac{1}{N} \sum_{i=1}^{N} S(y_i, z_{i,k+1}) \right)$$

  Maximization step (unchanged)

$$\theta_{k+1} = argmax_\theta \left\{ \phi(\theta)^T \overline{S}_{k+1} - \log(C(\theta)) \right\}$$

- **MCMC**-SAEM: replaces sampling by a single Markov Chain step

  - For each subject, sample the random effect w.r.t a transition kernel of a geometrically ergodic Markov chain targeting the conditional distribution

$$q(z_i|y_i,\theta_k)$$

[Delyon, Lavielle, Moulines.'99]
[Allassonnière et al.'10]

Parameter Estimation: stochastic algorithm

- **SA-EM**: replaces integration by one simulation of the hidden variable:
  sample $z_{i,k+1}$ from $p(z_i|y_i,\theta_k)$,
  and a stochastic approximation of the sufficient statistics

$$\overline{S}_{k+1} = (1-\Delta_k)\overline{S}_k + \Delta_k \left(\frac{1}{N} \sum_{i=1}^{N} S(y_i, z_{i,k+1}) \right)$$

  Maximization step (unchanged)

$$\theta_{k+1} = argmax_\theta \left\{ \phi(\theta)^T \overline{S}_{k+1} - \log(C(\theta)) \right\}$$

- **MCMC**-SAEM: replaces sampling by a single Markov Chain step

  - For each subject, sample the random effect w.r.t a transition kernel of a geometrically ergodic Markov chain targeting the conditional distribution

$$\tilde{q}_k(z_i, \theta_k)$$

As long as $\tilde{q}_k$ "converges towards" $q(z_i|y_i,\theta_k)$ as $k \to \infty$

[Chevallier & Allassonnière, preprint '19]

Parameter Estimation: stochastic algorithm

- **Theoretical properties of the sampler**:

  Under mild conditions:
  - Drift property
  - Small set
  - **Geometric ergodicity uniformly on any compact set** of the parameters

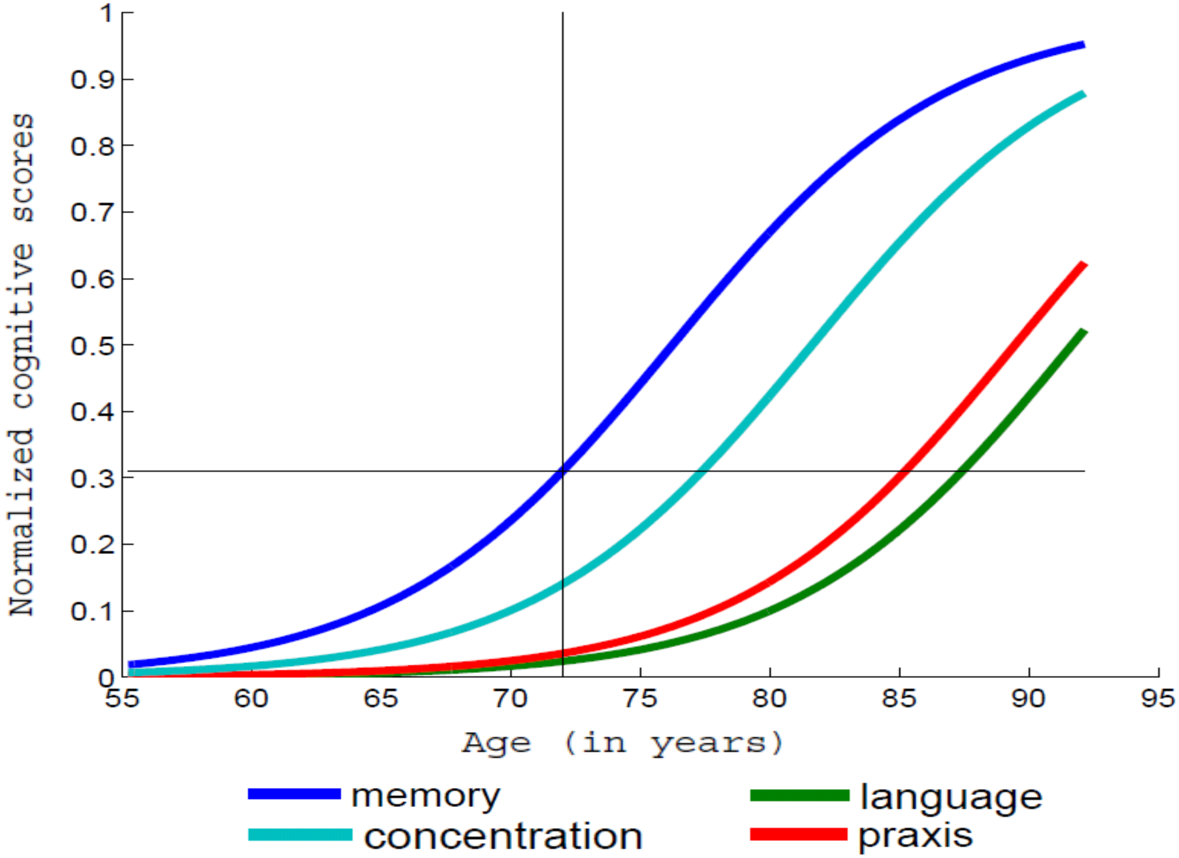
- **Theoretical properties of the estimation algorithm**:

  - **a.s. convergence** towards the MAP estimator
  - **Normal asymptotic behaviour**: speed $1/\sqrt{\Delta_k}$
  - Normal asymptotoc behaviour with optimal speed with averaging sequences $1/\sqrt{k}$

Model of Alzheimer's disease progression

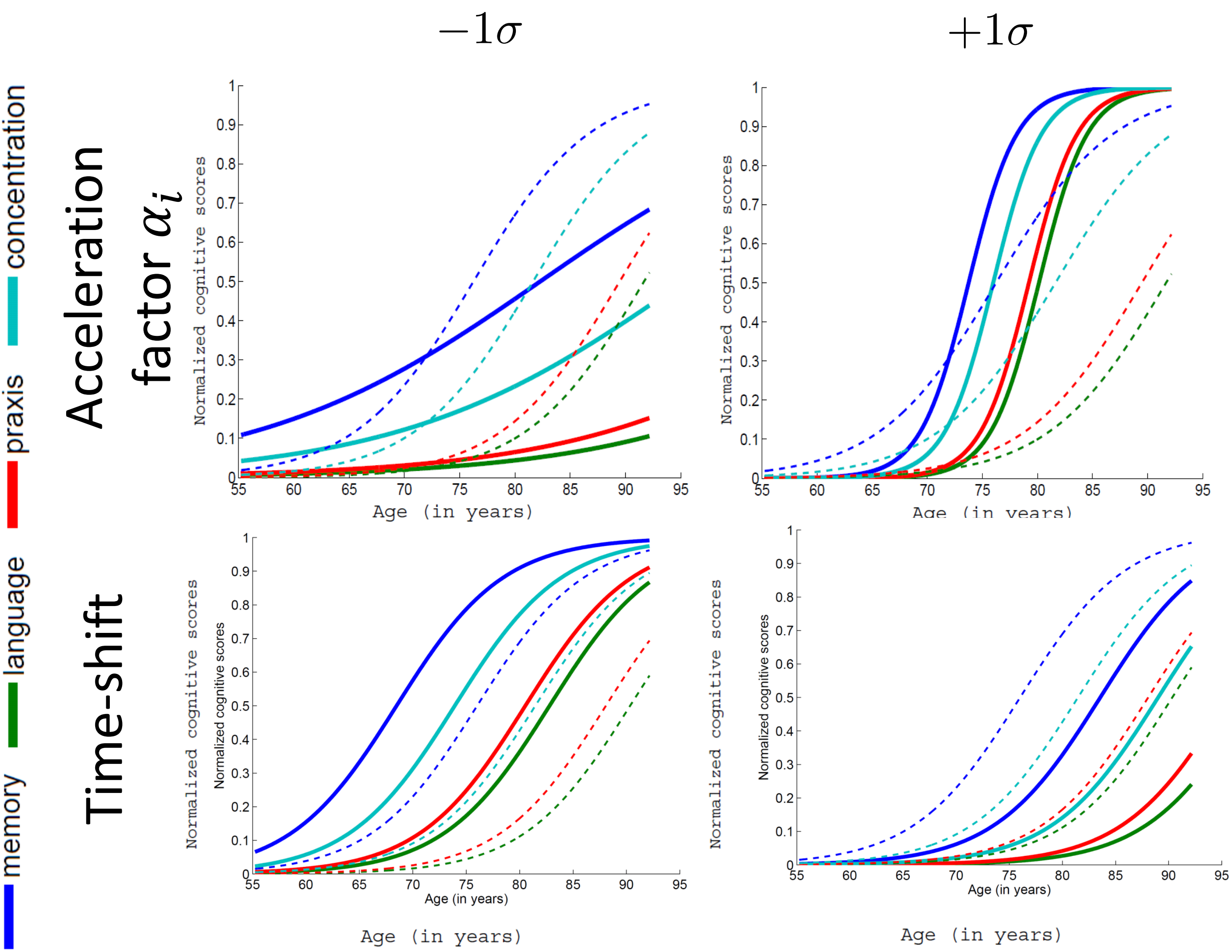
The average trajectory of data changes

- Neuropsychological tests ADAS-Gog from ADNI
- 248 subjects who converted from MCI to AD
- 6 time-points per subjects on average (min 3, max 11)
- Data points $y_{ij} \in]0,1[^4$ with propagation logistic model



[Schiratti et al. IPMI'15, NIPS'15]

Model of Alzheimer's disease progression

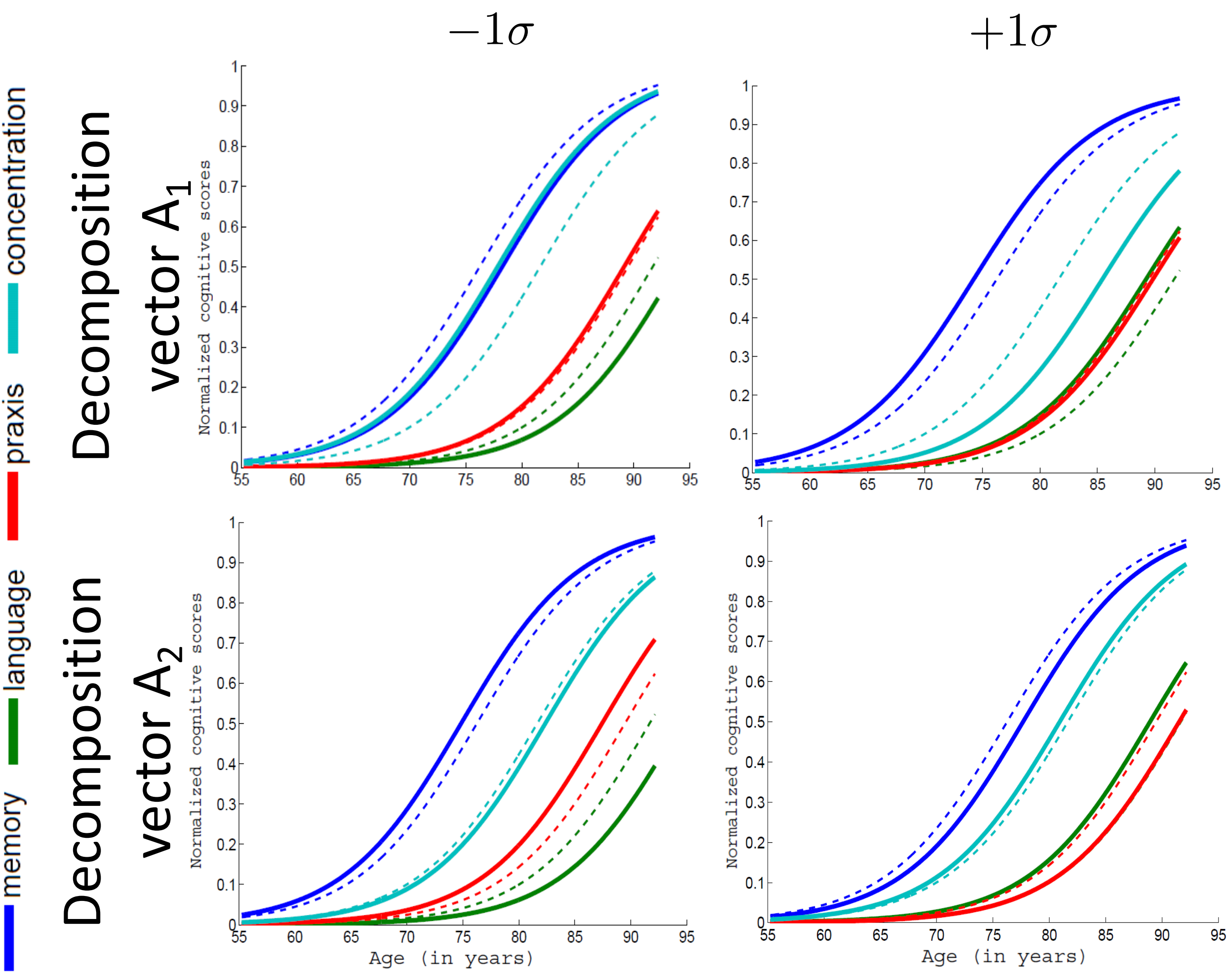


Distinguish **fast** vs. **slow** progressers

Distinguish **early** vs. **late** onset individuals

[Schiratti et al. IPMI'15, NIPS'15]

Model of Alzheimer’s disease progression



Variability in the **relative timing** and **ordering** of the events

[Schiratti et al. IPMI’15, NIPS’15]

Computational comparisons

- **Comparison of: MCMC-SAEM – STAN – MONOLIX**

  - Number of iterations:
    - MCMC-SAEM: 1 000 000 (6s / 1 000 iterations)
    - STAN: 15 000 (25min / 1 000 iterations)
    - MONOLIX: 20 000 (3,5 min / 1 000 iterations)

Computational comparisons

- **Comparison of: MCMC-SAEM – STAN – MONOLIX**

True values

p_0	t_0	v_0	σ_ξ	σ_τ	σ
0.24	70	0.034	0.5	7	0.01

MCMC-SAEM

p_0	t_0	v_0	σ_ξ	σ_τ	σ
0.23	69.93	0.0317	0.52	6.75	0.01

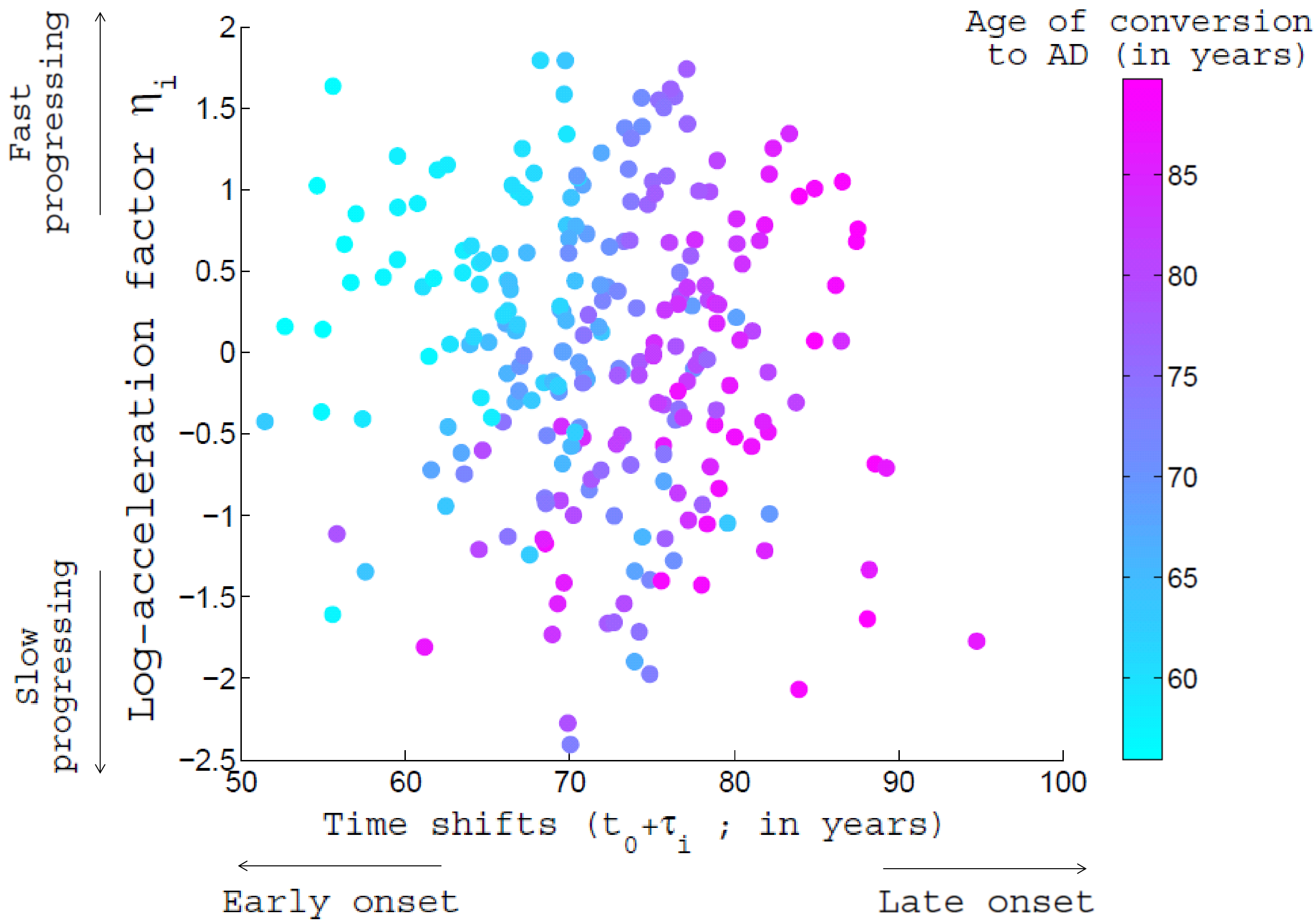
STAN

p_0	t_0	v_0	σ_ξ	σ_τ	σ
0.218	68.66	0.0305	0.53	6.73	0.098

Monolix

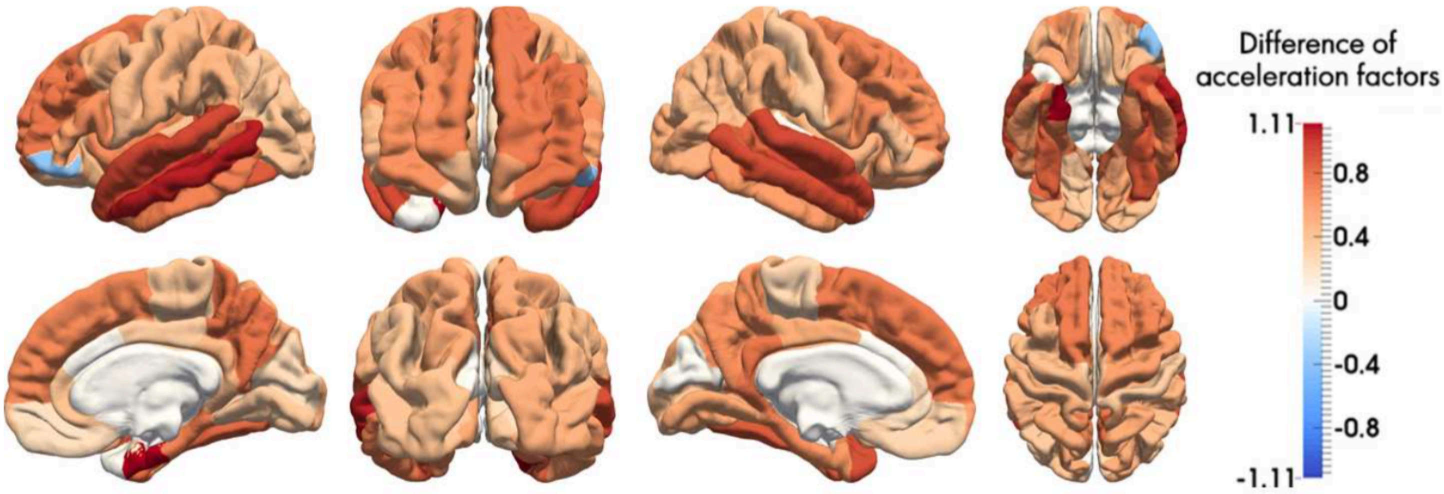
p_0	t_0	v_0	σ_ξ	σ_τ	σ
0.37	71.6	0.0406	0.52	6.8	0.01

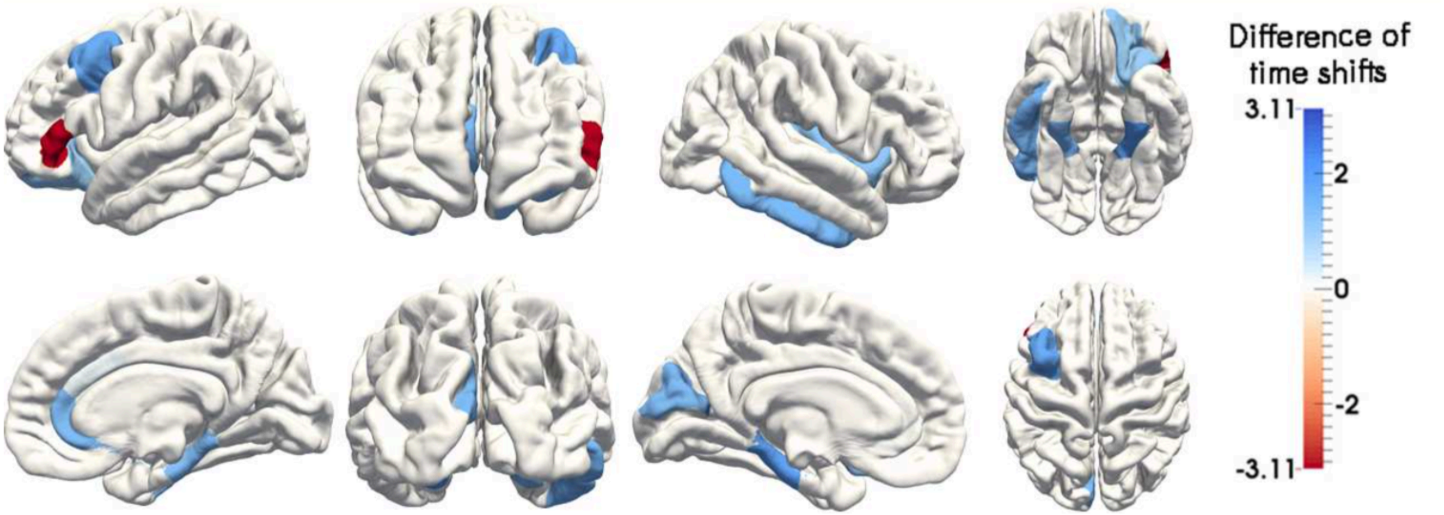
Model of Alzheimer's disease progression



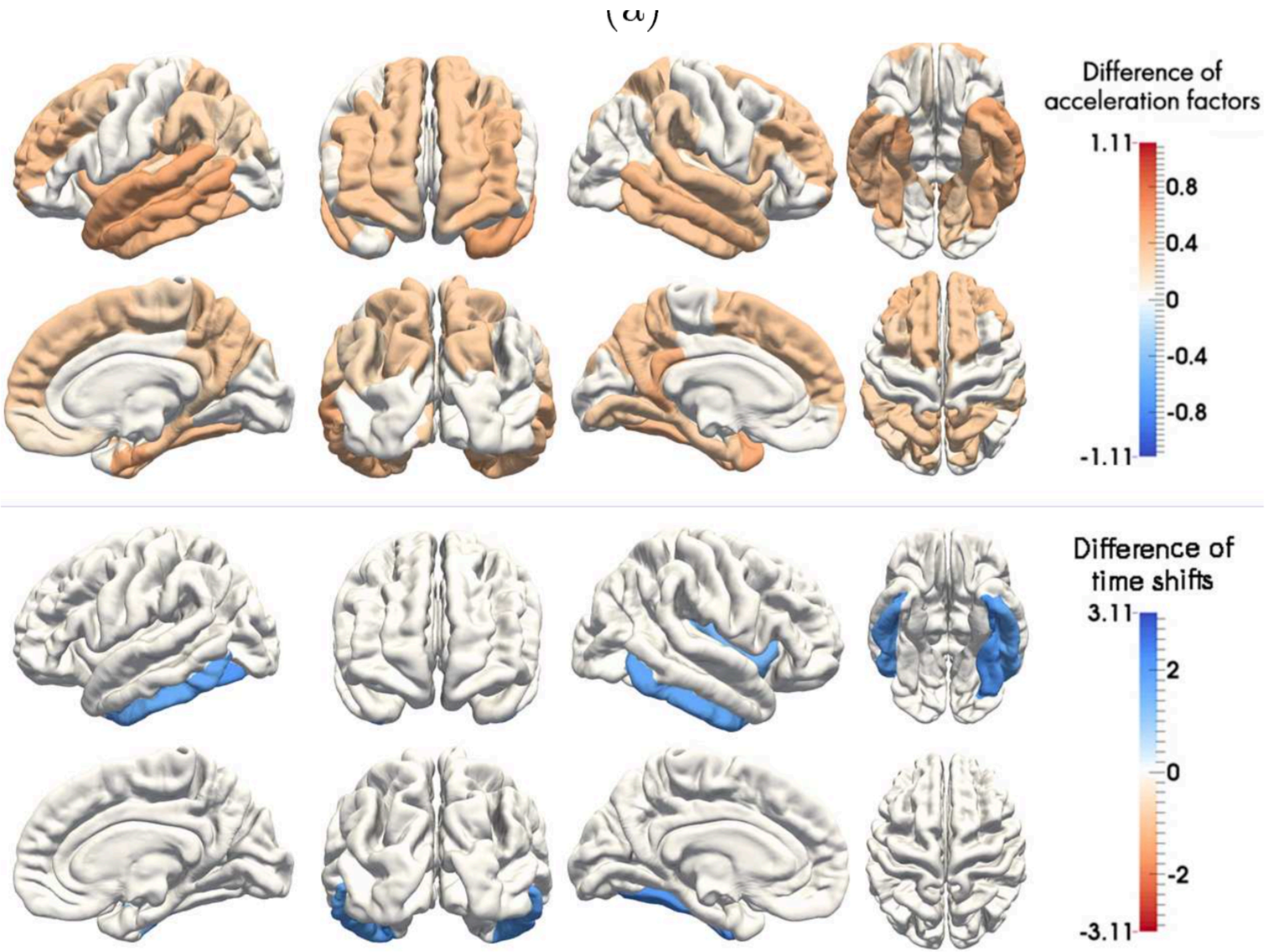
[Schiratti et al. IPMI'15, NIPS'15]

Comparison AD vs Controls

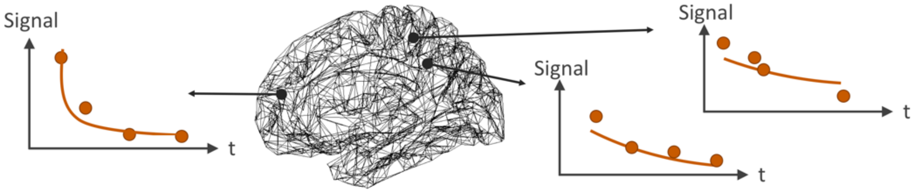




Comparison MCI vs Controls

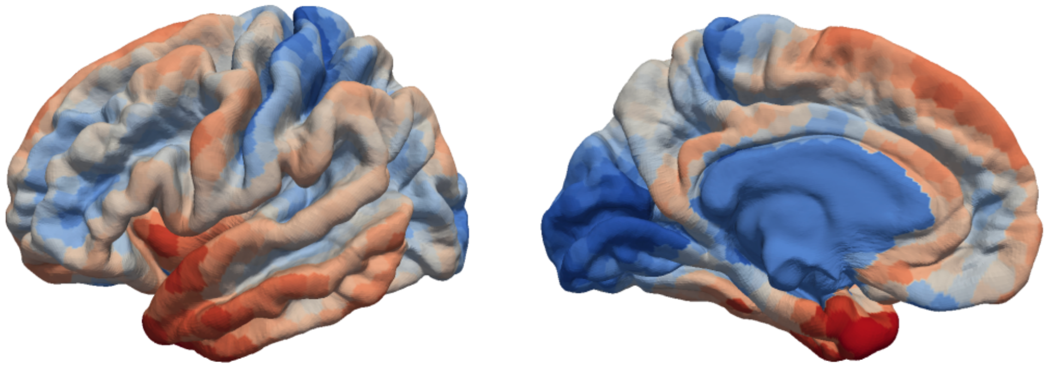


Model of propagation on a graph

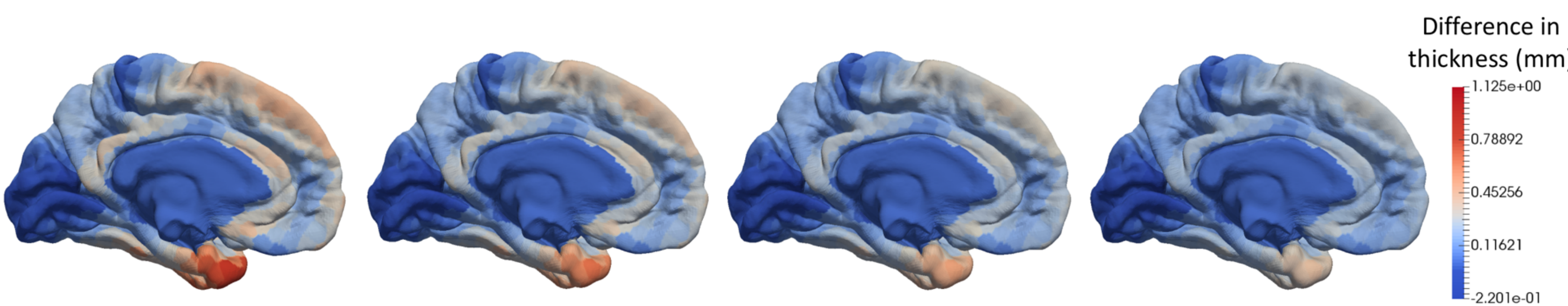


ADNI Data → Alzheimer's Disease cohort

Measures of the cortical thickness for MCI converters



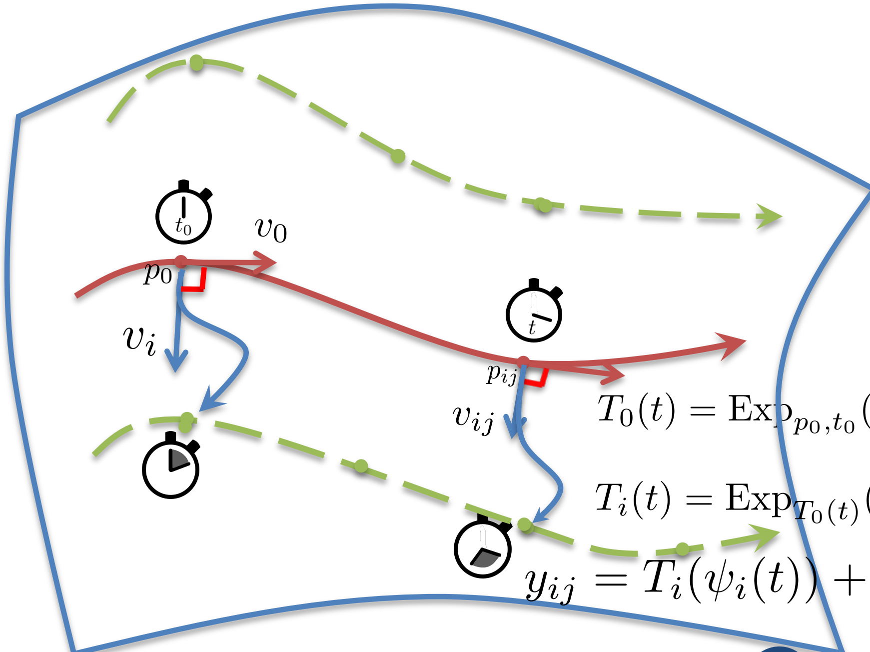
Neuronal loss within 10 years



Each snapshot corresponds to the neuronal loss within 5 years (from 70-75 to 85-90)

[Koval et al, Frontiers in Neurosciences'17]

Spatiotemporal Statistical Model



$$T_0(t) = \mathrm{Exp}_{p_0,t_0}(v_0)(t)$$

$$T_i(t) = \mathrm{Exp}_{T_0(t)}(\mathrm{P}^{T_0}_{t_0,t}(v_i))$$

$$y_{ij} = T_i(\psi_i(t)) + \varepsilon_{ij}$$

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

- **Geodesic** representative trajectory

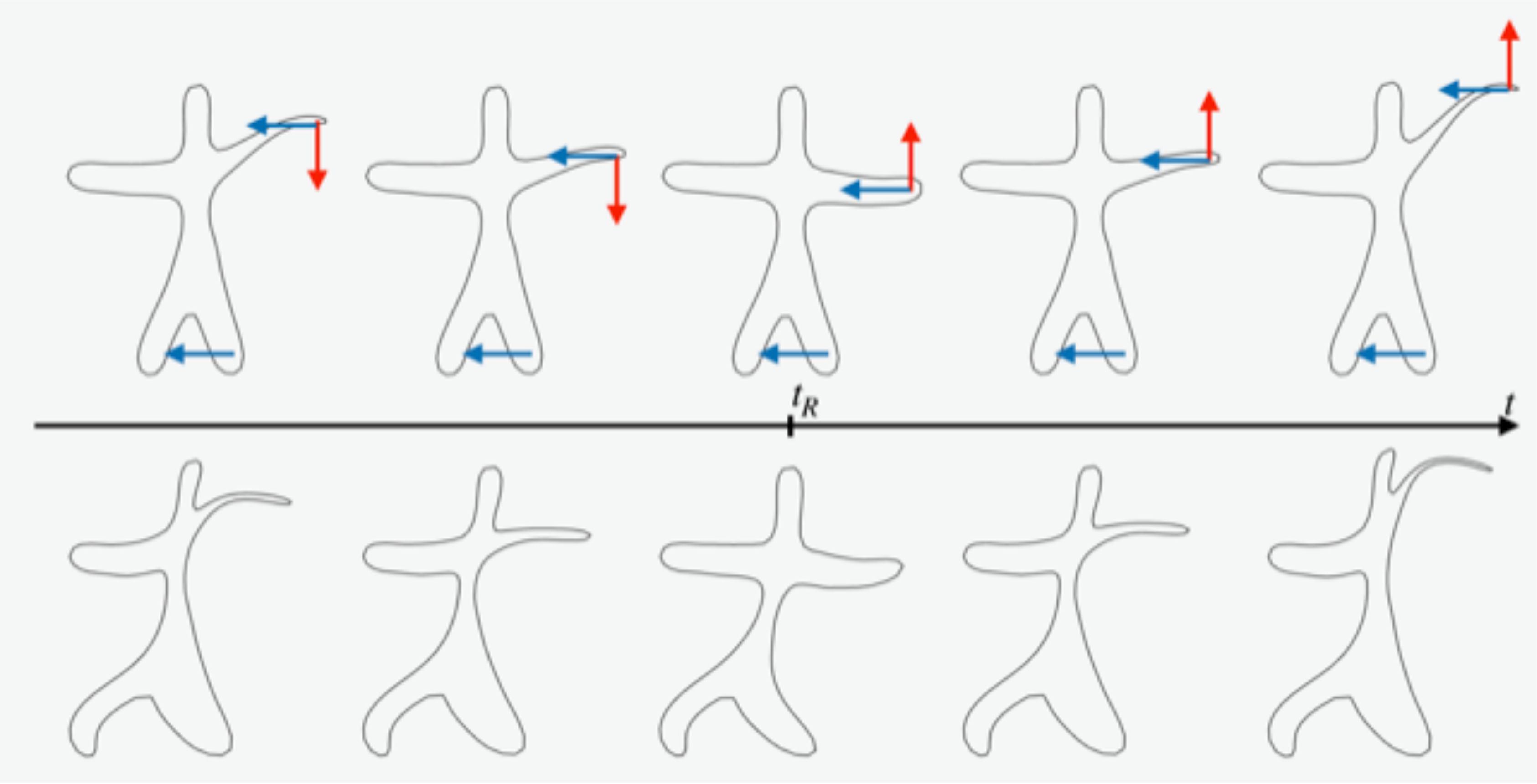
  -> What for **treated** pathologies?

	Acceleration factor	Time-shift	Space-shift
Random effects:	$\alpha_i \sim log\, \mathcal{N}(0, \sigma_\alpha^2)$	$\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$	$v_i = (A_1 \vert \dots \vert A_K)\, s_i$ $A_k \perp v_0$

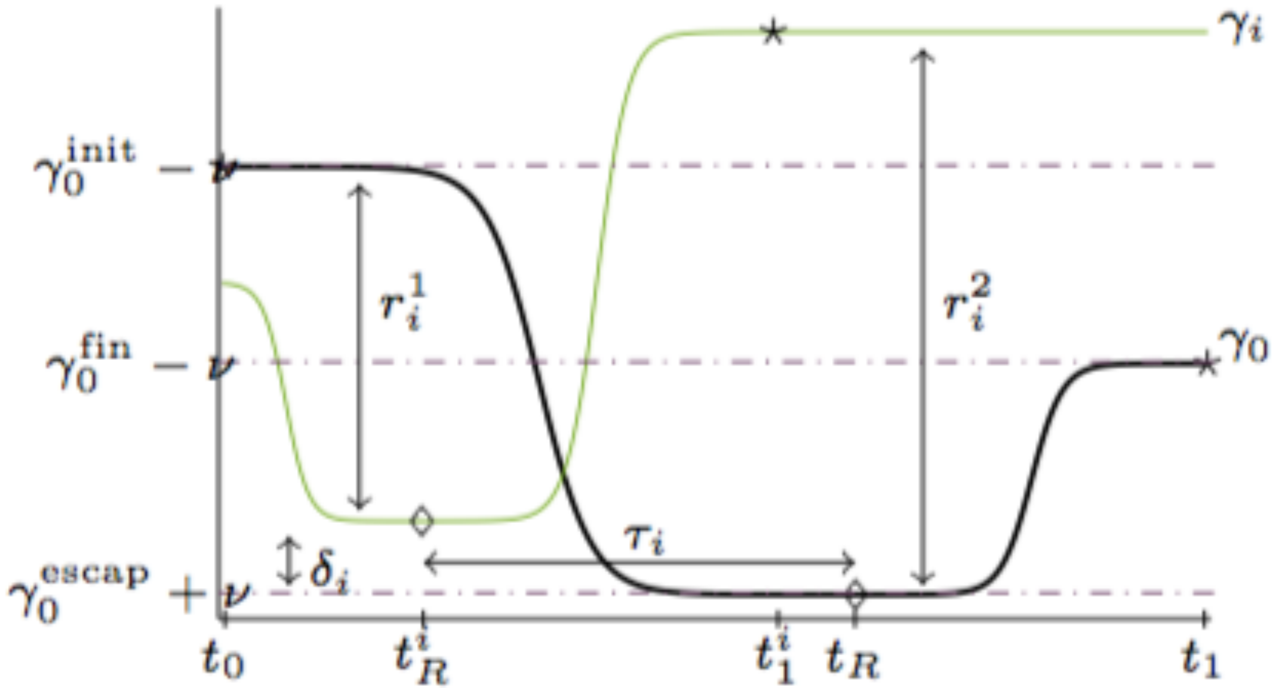
Fixed effects: (p_0, t_0, v_0) and $(\sigma_\alpha^2, \sigma_\tau^2, A_1, ...A_K)$

[Schiratti et al. IPMI'15, NIPS'15]

Spatiotemporal Statistical Model
Piecewise geodesic trajectories



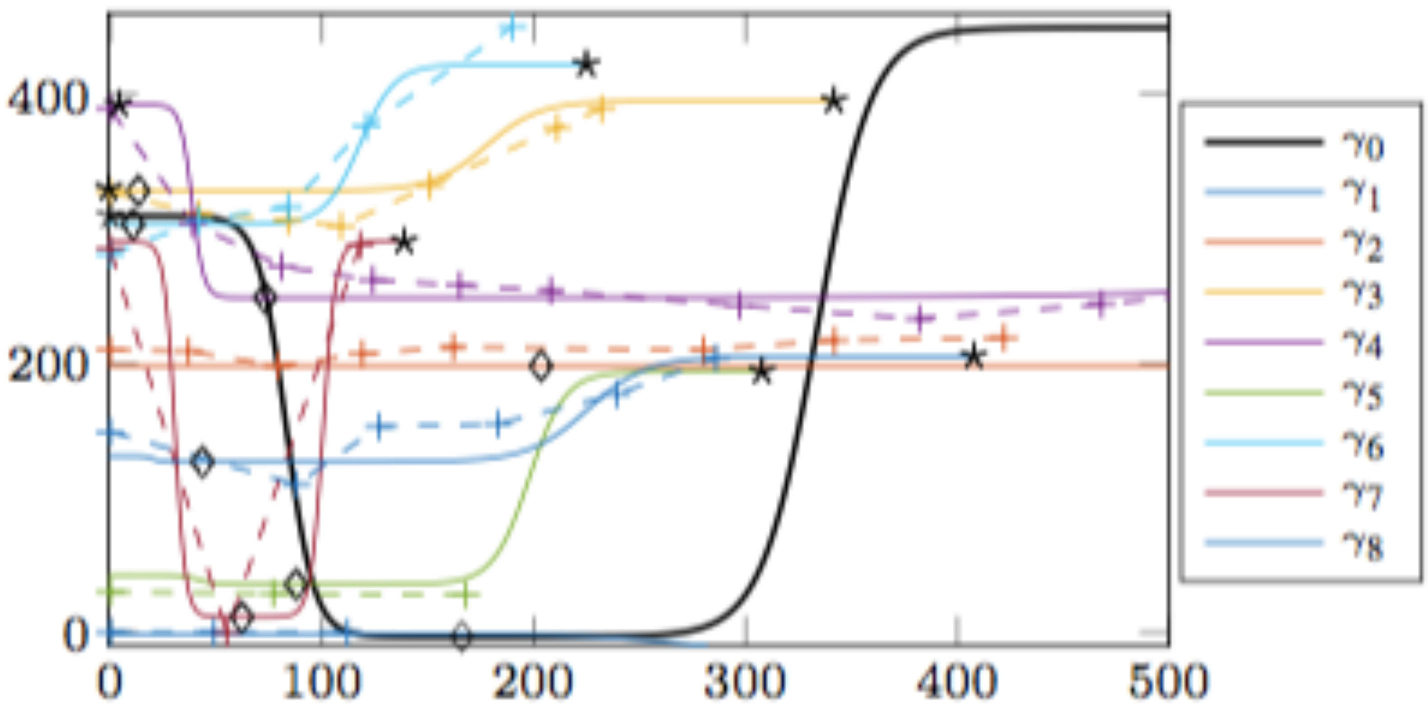
Spatiotemporal Statistical Model
Piecewise geodesic trajectories



(b) From average to individual trajectory.

[Chevallier et al, NIPS'17]

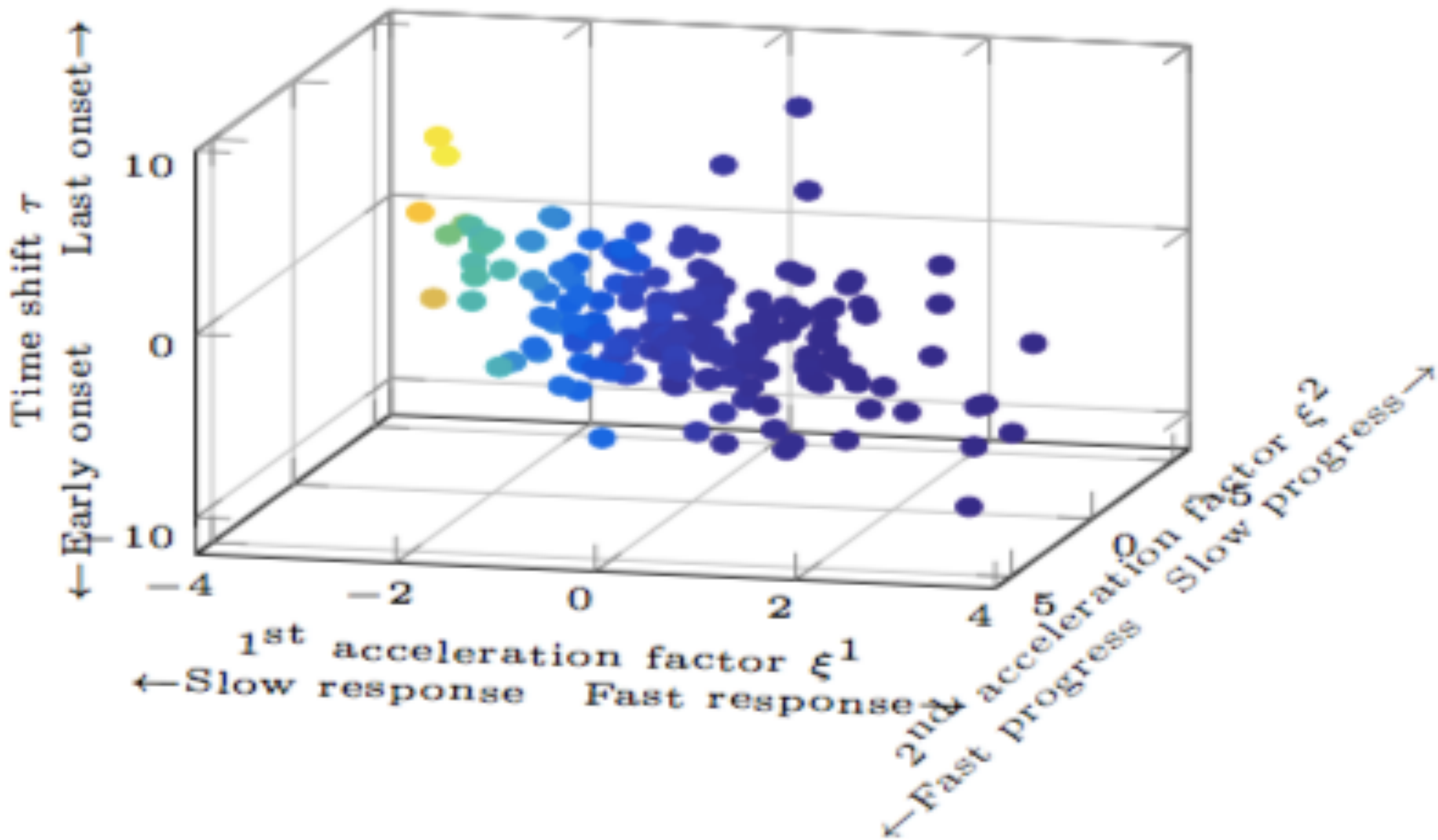
Spatiotemporal Statistical Model
Piecewise geodesic trajectories



(a) After 600 iterations.

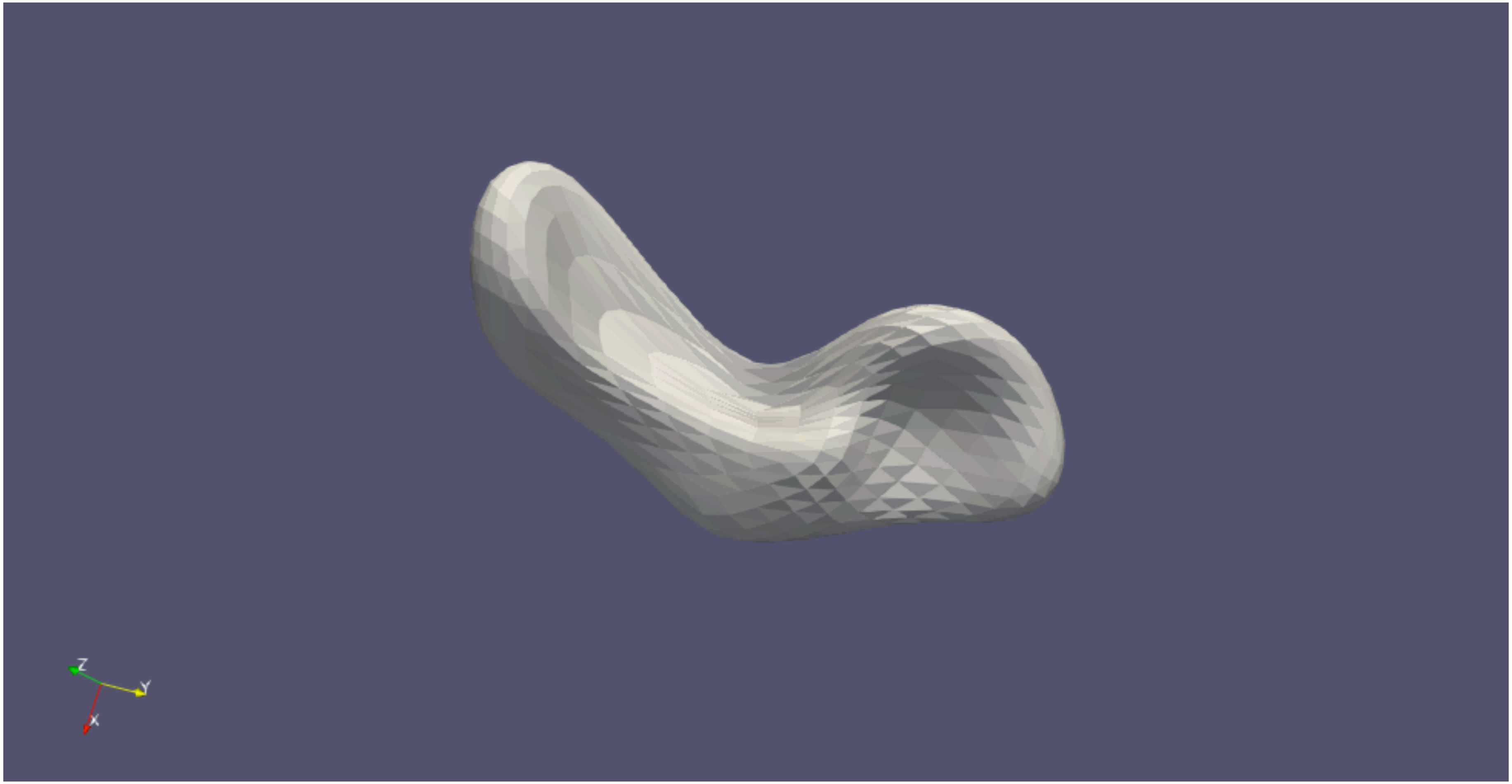
[Chevallier et al, NIPS'17]

Spatiotemporal Statistical Model
Piecewise geodesic trajectories

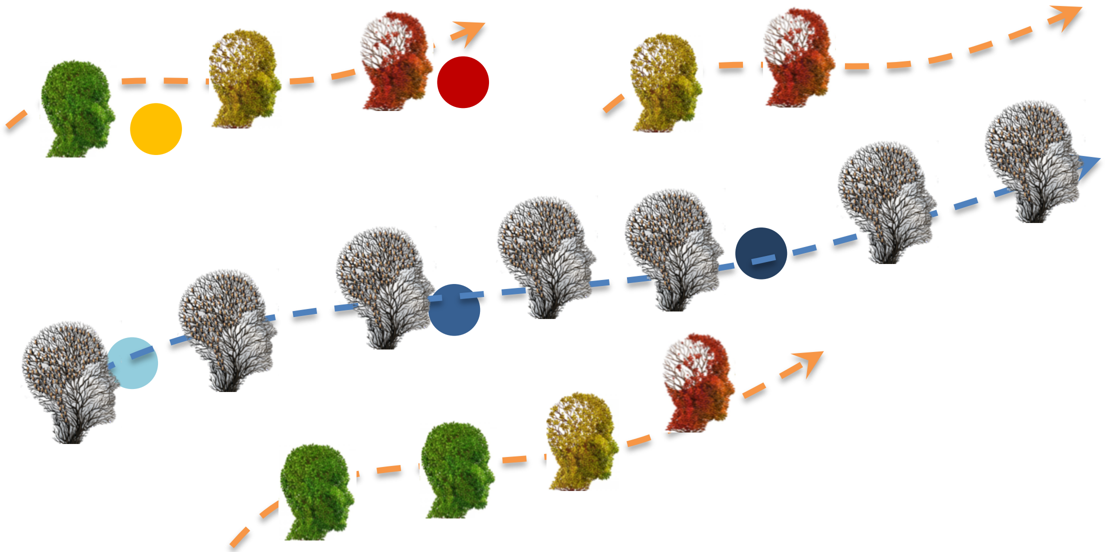


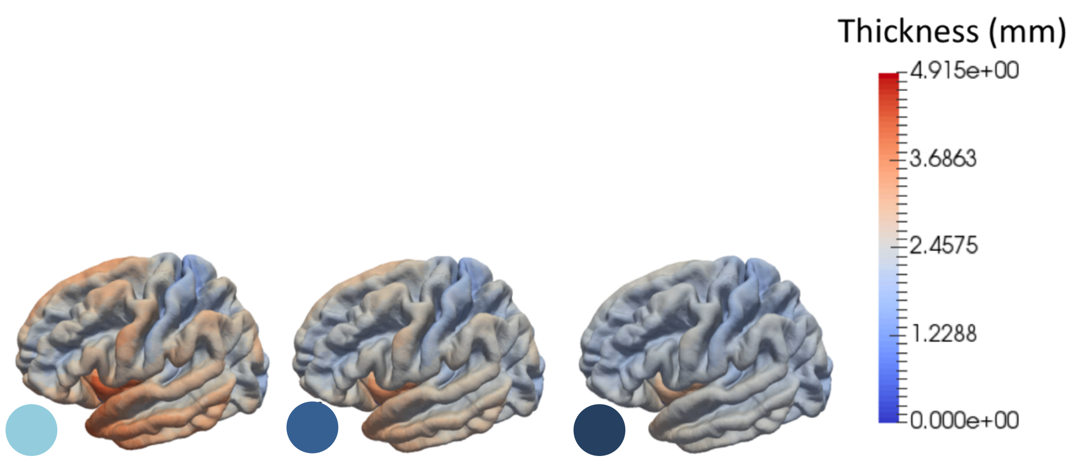
[Chevallier et al, NIPS'17]

Spatiotemporal Statistical Model
Piecewise geodesic trajectories



Prediction tool





The individual parameters are related to the real age of conversion of the individuals

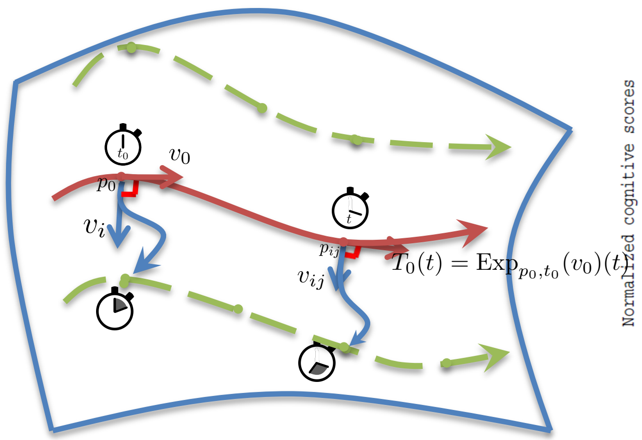
Conclusion

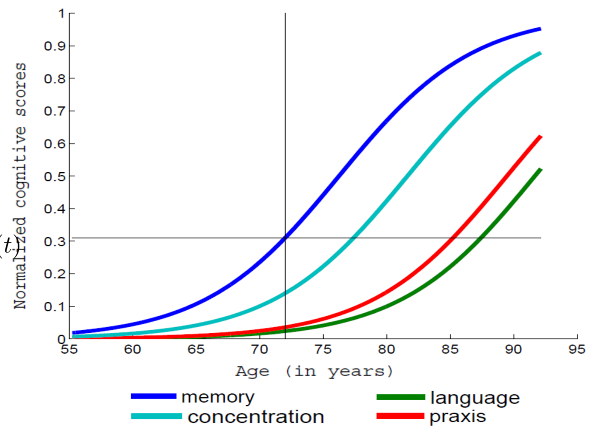
- Generic statistical model to learn **spatiotemporal distribution of trajectories** on manifolds:
  - Calibrated on **longitudinal** data sets using MCMC-SAEM
  - Automatically finds **temporal correspondences** among similar events that may happen at different age/time
  - Estimates the **variability** of the data at the corresponding events

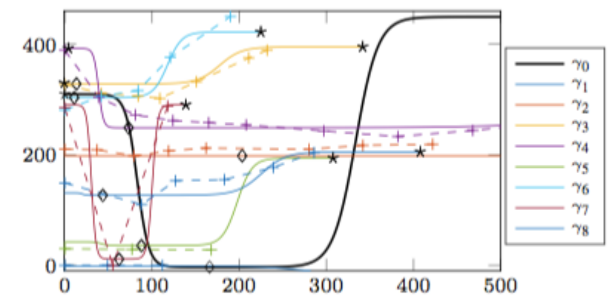
- It allows us to position disease progression within the life and history of the patient

- **Future work:**
  - Derive instances of the model for more complex manifold-valued data *(e.g. textured shapes data), metamorphosis and mixtures of all of these!*

Thank you!







(a) After 600 iterations.