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

Stéphanie Allassonnière

with J.B. Schiratti, O. Colliot 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?
Important issues in atlas estimation:

- Register any new data in the «coordinates» of the reference shape:
  - Transport the available information from the representative element
  - «Registration» penalised as a function of its «normality»

- Quantify anatomical structure variability in different sub-groups

Targetted applications:

- Pathology effects
- Classification of new patients
- Early diagnostic
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

\[
\begin{align*}
  y_i & \sim \mathcal{N}(I_0 \circ \phi_i^{-1}, \sigma^2 I_d) \\
  \phi_i & = f(\beta_i) \\
  \beta_i & \sim \mathcal{N}(0, \Gamma_g)
\end{align*}
\]

For $1 \leq i \leq n$ subjects,
Bayesian Mixed Effect model

Observations
- T1
- DWIs
- fMRI
- T1+fMRI

Template
- Grey level images
- Probability maps

Deformation
- Linearised
- Diffeomorphic

\[ y_i \sim \mathcal{N}(I_0 \circ \phi_i^{-1}, \sigma^2 I_d) \]

\[ \phi_i = f(\beta_i) \]

\[ \beta_i \sim \mathcal{N}(0, \Gamma_g) \]

deterministic parametrisation

Mixture of all these models

Normal distribution with full covariance matrix

For \(1 \leq i \leq n\) 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 acquisition 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)

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

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

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

- Needs to disentangle differences in manifold-valued data
  - (normalized data, positive matrices, shapes, etc..)
  - Dynamics of measurement changes
Spatiotemporal Statistical Model

- Statistical model including:
  - 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

**Random effects:**
\[ \alpha_i \sim \log \mathcal{N}(0, \sigma^2_\alpha) \]
\[ \tau_i \sim \mathcal{N}(0, \sigma^2_\tau) \]

**Fixed effects:**
\[ (p_0, t_0, v_0) \quad \text{and} \quad (\sigma^2_\alpha, \sigma^2_\tau, A_1, \ldots A_K) \]

\[ v_i = (A_1 \mid \ldots \mid A_K) s_i \]
\[ A_k \perp v_0 \]

[Schiratti et al. IPMI’15, NIPS’15]
Spatiotemporal Statistical Model

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

\[ T_i(t) = \text{Exp}_{T_0(t)}(P_{t_0,t}^{T_0}(v_i)) \]

\[ T_0(t) = \text{Exp}_{p_0,t_0}(v_0)(t) \]

\[ \psi_i(t) = t_0 + \alpha_i(t - t_0 - \tau_i) \]

\[ \alpha_i \sim \log N(0, \sigma^2_\alpha) \]

\[ \tau_i \sim N(0, \sigma^2_\tau) \]

\[ v_i = (A_1|\ldots|A_K) s_i \]

\[ A_k \perp v_0 \]

\[ (p_0, t_0, v_0) \]

\[ (\sigma^2_\alpha, \sigma^2_\tau, A_1, \ldots A_K) \]

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 \( M = \mathbb{R} \)

\[
y_{ij} = (\bar{a} \times a_i)(t_{i,j} - t_0 - \tau_i) + \bar{b} + \varepsilon_{i,j}
\]

Time at which measurement of the \( i^{th} \) subject reaches \( \bar{b} \)

\[
y_{ij} = (\bar{a} \times a_i)(t_{i,j} - t_0) + \bar{b} + b_i + \varepsilon_{i,j}
\]

Measurement of the \( i^{th} \) subject at time \( t_0 \)

Spatiotemporal Statistical Model

• The logistic curve model:

\[ g(p)(u, v) = \frac{uv}{p^2(1 - p)^2} \]

\[ \gamma_0(t) = 1 + \frac{(1 - p_0)/p_0}{\exp\left(-\frac{v_0}{p_0(1-p_0)} (t - t_0)\right)} \]

\[ y_{ij} = \gamma_0\left(t_0 + \alpha_i(t - t_0 - \tau_i)\right) + \varepsilon_{ij} \]

• Geodesic are logistic curves

• 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 \( \rightarrow \) end up with our previous linear case (different from Laird & Ware)
• The propagation model $M = ]0, 1[^N$, $g(p)(u, v) = \sum_{k=1}^{N} \frac{u_k v_k}{p_k^2 (1 - p_k)^2}$

• Geodesics are logistic curves in each coordinate

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

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

• 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

\[\text{The effect onset across coordinates}\]
Parameter Estimation

\[ y = (y_1, \ldots, y_N), \quad z = (z_1, \ldots z_N), \quad \theta = (\sigma_z^2, \sigma_e^2, A_1, \ldots, A_K, p_0, t_0, v_0) \]

- **Maximum Likelihood:**

\[
\max_{\theta} p(y|\theta) = \int p(y, z|\theta)dz
\]

- **EM:**

\[
\theta_{k+1} = \operatorname{argmax}_{\theta} \sum_{i=1}^{N} \int \log \left( \frac{p(y_i, z_i|\theta)}{p(y_i|z_i, \theta)p(z_i|\theta)} \right) p(z_i|y_i, \theta_k)dz_i
\]

- **Distribution from the curved exponential family**

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

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

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

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

- For each coordinate $p$ (Gibbs sampler) sample $z_i \sim p(z_i^p | z_i^g \neq p, \theta)$
- Set $z_{i,k+1}^p = \tilde{z}_i^p$ with probability $\frac{1}{\sqrt{N}} \frac{p(y_i | \tilde{z}_i, \theta)}{p(y_i | z_i, \theta)}$
- $z_{i,k+1}^p = z_{i,k}^p$ otherwise

[Delyon, Lavielle, Moulines.'99] [Allassonnière et al. 10]
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 $\frac{1}{\sqrt{\Delta_k}}$
  – Normal asymptotic behaviour with optimal speed with averaging sequences $\frac{1}{\sqrt{k}}$
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 $y_{ij} \in [0, 1]$ with propagation logistic model

[Schiratti et al. IPMI’15, NIPS’15]
Model of Alzheimer’s disease progression

-1σ +1σ

Distinguish fast vs. slow progressers

Distinguish early vs. late onset individuals

[Schiratti et al. IPMI’15, NIPS’15]
Model of Alzheimer’s disease progression

Decomposition vector $A_1$

Decomposition vector $A_2$

Variability in the relative timing and ordering of the events

[Schiratti et al. IPMI’15, NIPS’15]
Model of Alzheimer’s disease progression

[Schiratti et al. IPMI’15, NIPS’15]
Model of Alzheimer’s disease progression

[Schiratti et al. IPMI’15, NIPS’15]
Model of Alzheimer’s disease progression
Model of diffusion tensors

- Geodesic in the Riemannian manifold of positive definite matrices
- Parallel transport the tensors
- Reparametrize in time
- Sample this curse
Model of diffusion tensors

- Synthetic data not generated from the model but imitating a non smooth evolution
- 100 subjects
- 5 time points in average
Model of diffusion tensors

• Fitting the model to a new patient
Comparison AD vs Controls
Comparison MCI vs Controls
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

<table>
<thead>
<tr>
<th></th>
<th>$p_0$</th>
<th>$t_0$</th>
<th>$v_0$</th>
<th>$\sigma_\xi$</th>
<th>$\sigma_\tau$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True values</td>
<td>0.24</td>
<td>70</td>
<td>0.034</td>
<td>0.5</td>
<td>7</td>
<td>0.01</td>
</tr>
<tr>
<td>MCMC-SAEM</td>
<td>0.23</td>
<td>69.93</td>
<td>0.0317</td>
<td>0.52</td>
<td>6.75</td>
<td>0.01</td>
</tr>
<tr>
<td>STAN</td>
<td>0.218</td>
<td>68.66</td>
<td>0.0305</td>
<td>0.53</td>
<td>6.73</td>
<td>0.098</td>
</tr>
<tr>
<td>Monolix</td>
<td>0.37</td>
<td>71.6</td>
<td>0.0406</td>
<td>0.52</td>
<td>6.8</td>
<td>0.01</td>
</tr>
</tbody>
</table>
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.* spatially distributed data, shape data, etc..)
Thank you!