Estimation and variable selection in high dimension in nonlinear mixed-effects models

Estelle Kuhn

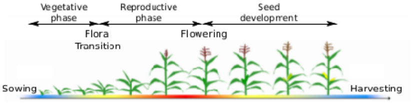
Applied Mathematics and Computer Science from Genome to Environment

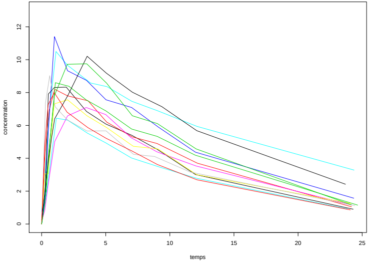
University Paris-Saclay, INRAE

Joint work with A. Caillebotte, S. Lemler and R. Rincent

Biological objectives

Two examples in plant growth and pharmacokinetic





- Understanding intra-subject processes
- Understanding variations of these processes across subjects

$\Rightarrow$ fundamental for developing individual strategies and guidelines

Our assumptions: access to

- Repeated measurements within individuals for a population
- Mechanistic model with interpretable parameters

$\Rightarrow$ Inference on mechanisms and variations in the population

Observations of growing process along time

[Pinheiro and Bates (2000)]

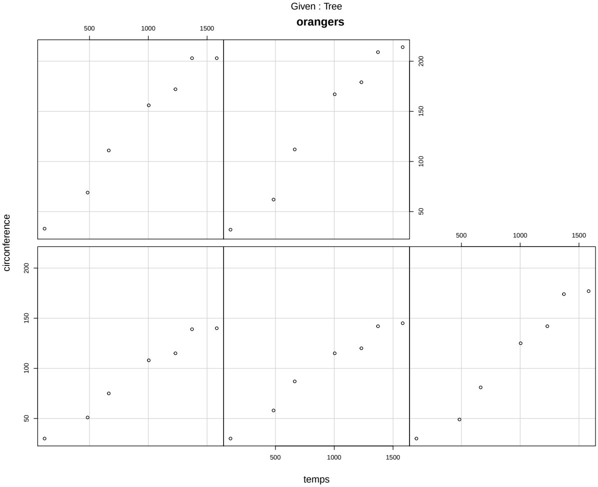
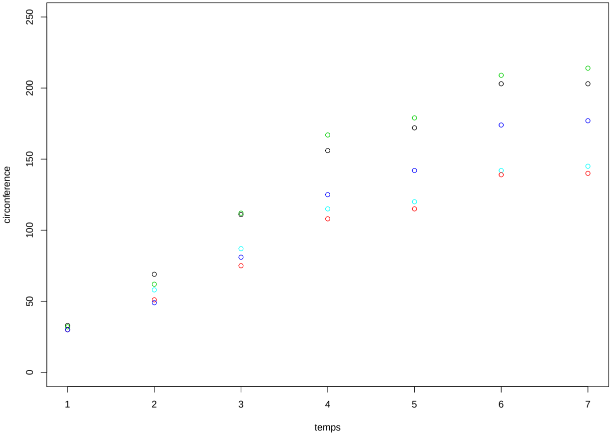


Figure: Circumferences of 5 orange trees measured at 7 times

Observations of growing process of five orange trees

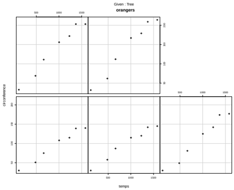


logistic model $y(t) = \frac{\varphi_{i1}}{1+\exp\left(-\frac{t-\varphi_{i2}}{\varphi_{i3}}\right)}$

Individual level approach versus population approach

- Ajusting N regression models each with J observations

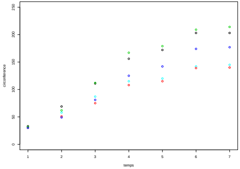
$$Y_{ij} = \frac{\varphi_{i1}}{1 + \exp\left(-\frac{t_j - \varphi_{i2}}{\varphi_{i3}}\right)} + \varepsilon_{ij}$$



Model parameters : $(\varphi_i, var_{\varepsilon_i}) \in \mathbb{R}^{3+1}$
$\Rightarrow N(3+1)$ parameters

- Ajusting 1 model with NJ observations

$$\left\{ \begin{array}{rcl} Y_{ij} & = & \frac{\varphi_{i1}}{1+\exp\left(-\frac{t_j - \varphi_{i2}}{\varphi_{i3}}\right)} + \varepsilon_{ij} \\ \varphi_i & = & \mu_\varphi + \nu_i \end{array} \right.$$



Parameters: $\theta_{pop} = (\mu_\varphi, var_\nu, var_\varepsilon)$
$\Rightarrow 2 * 3 + 1$ parameters

Statistical issues raised up

Consider the following mixed effects model:

$$\begin{cases} Y_{ij} & = & h(\varphi_i, t_j) + \varepsilon_{ij} & 1 \le i \le N,\ 1 \le j \le J \\ \varphi_i & = & \mu_\varphi + b_i, & 1 \le i \le N \end{cases}$$

with $b_i \overset{iid}{\sim} \mathcal{N}(0; \Gamma)$ random effects, $\varepsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0; \sigma^2)$ noise term

Objectives:

- estimate model parameters $\theta = (\mu_\varphi, \Gamma, \sigma^2)$
- predict individual output as $\hat{\varphi}_i$ or $\hat{Y}_i$
- explain variabilities of individual parameters φ_i
- ...

Introducing high-dimensional covariates in the model

Consider the following mixed effects model:

$$\begin{cases} Y_{ij} & = & h(\varphi_i, t_j) + \varepsilon_{ij} & 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i & = & \mu_\varphi + b_i, & 1 \leq i \leq N \end{cases}$$

with $b_i \sim \mathcal{N}(0; \Gamma)$ random effects, $\varepsilon_{ij} \sim \mathcal{N}(0; \sigma^2)$ noise term

Idea : explain parameter variability with covariates

$$\varphi_i = \mu + \beta X_i + b_i, \quad 1 \leq i \leq N$$

with X_i covariates of size p large versus N

$\Rightarrow$ *Variable selection in high dimension in mixed model*

Inference through reguralized maximum likelihood estimate

[Caillebotte et al. (2025)]

Consider the following mixed effects model:

$$\begin{cases} Y_{ij} & = & h(\varphi_i, t_j) + \varepsilon_{ij} & 1 \leq i \leq N,\ 1 \leq j \leq J \\ \varphi_i & = & \mu + \beta X_i + b_i, & 1 \leq i \leq N \end{cases}$$

Consider the LASSO estimate [Tibshirani (1996)]

$$\hat{\theta}_{\lambda}^{LASSO} = \arg\max_{\theta \in \Theta} \left\{ \log L_{marg}(\theta; y) - \lambda \|\theta\|_1 \right\}.$$

with λ regularization parameter and

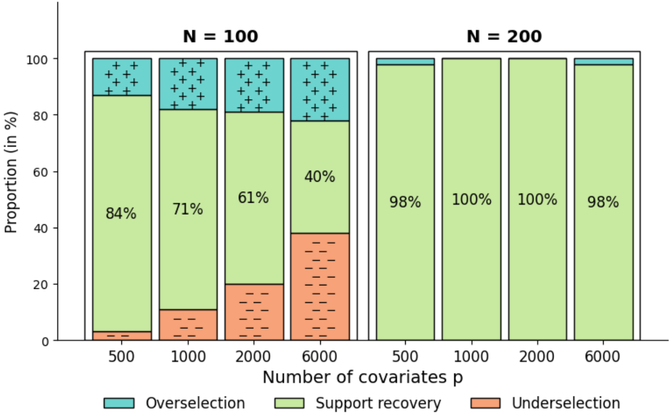
$$L_{marg}(\theta; y) \quad = \quad \int L_{comp}(\theta; y, b) db$$

$\Rightarrow$ weighted proximal stochastic gradient descent
and eBIC criteria [Chen and Chen (2009)]

Simulation study in a logistic mixed model

Let $1 \leq i \leq N$ and $1 \leq j \leq J$

$$\begin{cases} Y_{ij} = \dfrac{\varphi_{i1}}{1 + \exp\left(-\dfrac{t_{ij} - \varphi_{i2}}{\tau}\right)} + \varepsilon_{ij} & , \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \\ \varphi_{i1} = \mu_1 + b_{i1} & , b_{i1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \gamma_1^2) \\ \varphi_{i2} = \mu_2 + \beta^t X_i + b_{i2} & , b_{i2} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \gamma_2^2) \end{cases}$$



Variable selection in pharmacokinetic mixed effects models

Pharmacokinetic model:
$N = 200$, $J = 12$ and $p = 500$; *volume* and *dose* are known

$$\begin{cases} Y_{ij} = \frac{dose\ \varphi_{i1}}{volume\ \varphi_{i1}-\varphi_{i2}} \left(\exp(-\varphi_{i2} t_{ij}/volume) - \exp\left(-\varphi_{i1} t_{ij}\right)\right) + \varepsilon_{ij} \\ \varphi_{i1} = \mu_1 + {\beta_1}^t X_i + b_{i1} \\ \varphi_{i2} = \mu_2 + {\beta_2}^t X_i + b_{i2} \\ b_{i1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \gamma_1^2)\ b_{i2} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \gamma_2^2)\ \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \end{cases}$$

$\Rightarrow$ comparison between 3 approaches:

- A. $(\hat{\varphi}_i)$ by N individual model + selection of (X_i) using $\hat{\varphi}_i$
- B. $(\hat{\varphi}_i)$ by one mixed model + selection of (X_i) using $\hat{\varphi}_i$
- C. direct selection by one mixed model integrating (X_i)

$\Rightarrow$ regarding robustness to partial observation
keep only the first 3 observations for a proportion ρ of individuals

Variable selection's results

$$Y_{ij} = \frac{dose\ \varphi_{i1}}{volume\ \varphi_{i1} - \varphi_{i2}} \left(\exp(-\varphi_{i2} t_{ij}/volume) - \exp\left(-\varphi_{i1} t_{ij}\right)\right) + \varepsilon_{ij}$$

- A. $(\hat{\varphi}_i)$ by N individual model + selection of (X_i) using $\hat{\varphi}_i$
- B. $(\hat{\varphi}_i)$ by one mixed model + selection of (X_i) using $\hat{\varphi}_i$
- C. direct selection by one mixed model integrating (X_i)

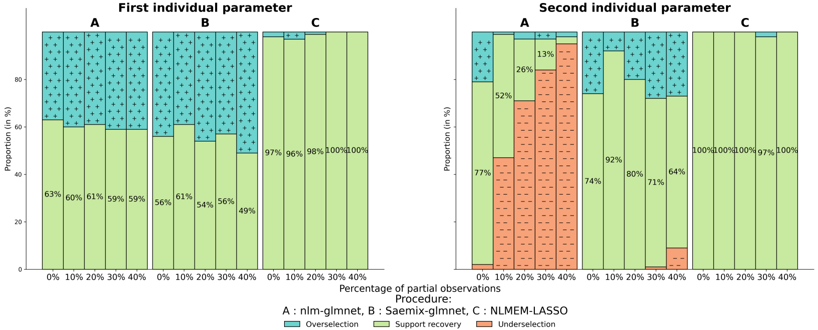


Figure: Proportions of support selection under increasing censoring rates.

Take home message and open questions

- modeling parameter variability in a mechanistic model
- more interpretability of parameter estimates
- explain parameter variability with HD covariates
- statistical tool to reduce parameter number
- population approach regularize variable selection

$\Rightarrow$ Many open questions :

? manage correlation between covariates

? optimize selection of the regularization parameter

? computational cost with complex mechanistic model

? post model selection inference after LASSO

? ...

$\Rightarrow$ visit maiage.inrae.fr/estelle-kuhn

Bibliography

Caillebotte, A., E. Kuhn, and S. Lemler (2025). Estimation and variable selection in high dimension in nonlinear mixed-effects models.

Chen, Z. and J. Chen (2009). Tournament screening cum ebic for feature selection with high-dimensional feature spaces. *Science in China Series A: Mathematics 52*(6), 1327–1341.

Pinheiro, J. and D. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology 58*(1), 267–288.