

Méthodes de régularisation pour l'intégration de données de grande dimension dans les modèles mécanistes : application pour le développement de vaccins

Soutenance de thèse de Auriane Gabaut

Sous la direction de Mélanie Prague et Cécile Proust-lima

Université de Bordeaux
Inria, BPH

25 novembre 2025

Modélisation mécaniste pour la vaccinologie

La **modélisation mécaniste** décrit mathématiquement ces dynamiques à partir de **compartiments biologiques**.

$$\begin{cases} \frac{d}{dt} S(t) = -\delta_S S(t) \\ \frac{d}{dt} L(t) = -\delta_L L(t) \\ \frac{d}{dt} Ab(t) = \theta_S S(t) + \theta_L L(t) - \delta_{Ab} Ab(t) \end{cases}$$

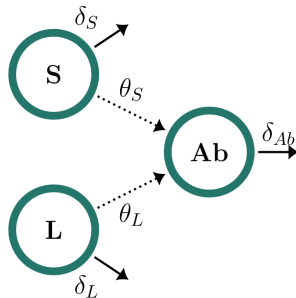


Figure 1 – Modèle de la réponse immunitaire humorale distinguant deux sous-populations de plasmocytes (Andraud *et al.*, 2012; Pasin *et al.*, 2019).

Modèle non-linéaire à effets mixtes (NLEM)

Modèle structurel

Décrit la dynamique intra-individuelle du phénomène d'intérêt.

$$\frac{d\mathbf{S}_i}{dt} = u(\mathbf{S}_i(t), \psi_i, t).$$

u est une fonction de lien fixe.

Modèle statistique

Introduit la variabilité inter-individuelle. Pour chaque individu i :

$$h(\psi_i) = h(\psi_0) + \mathbf{X}_i\beta + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \Omega).$$

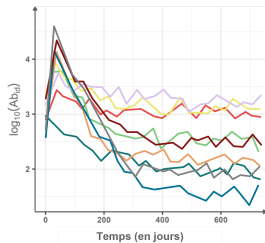
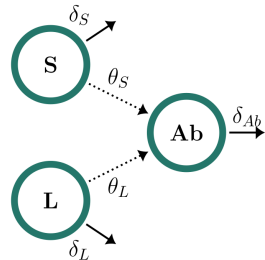
h est une fonction de lien fixe.

Modèle d'observation

Relie les prédictions du modèle structurel aux données observées $\mathbf{Y}_{ij}$.

$$\mathbf{Y}_{ij} = g(\mathbf{S}_i(t_{ij}), \psi_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2).$$

g est une fonction de lien fixe.



Estimation dans les modèles NLEM

Objectif : estimer les paramètres du modèle à partir des données longitudinales.

- Dans ces modèles, la vraisemblance s'écrit :

$$p(\mathbf{Y}_i|\boldsymbol{\theta}) = \int p(\mathbf{Y}_i|\boldsymbol{\eta}_i, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i|\boldsymbol{\theta}) d\boldsymbol{\eta}_i.$$

- Cette intégrale n'est **pas analytique** :
 - effets aléatoires latents ;
 - modèle non-linéaire.

Approches possibles

- Linéarisation locale : FOA, FOCE.
- Intégration numérique : quadrature de Gauss(-Hermite).
- Simulation : Monte Carlo ou MCMC.
- Algorithmes EM / SAEM : approche stochastique robuste et efficace.

Algorithme SAEM : une version stochastique de l'EM

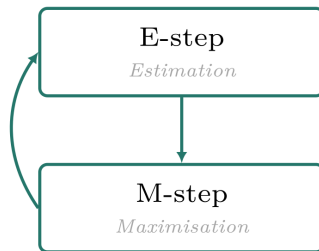
💡 Extension stochastique de l'algorithme EM (Delyon *et al.*, 1999).

- À l'itération k , pour les paramètres courants $\theta^{(k-1)}$:

- **E-step** : calculer l'espérance conditionnelle

$$Q(\theta|\theta^{(k-1)}) = \mathbb{E}_{\psi|Y, \theta^{(k-1)}} [\log p(Y, \psi|\theta)].$$

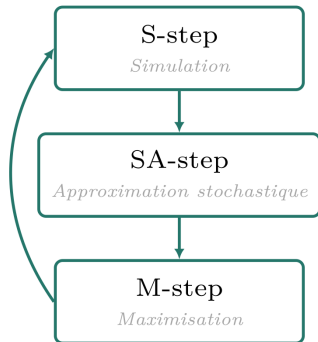
- **M-step** : maximisation pour obtenir $\theta^{(k)}$.



Algorithme SAEM : une version stochastique de l'EM

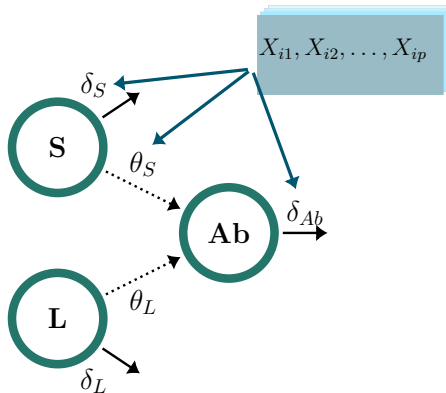
💡 Extension stochastique de l'algorithme EM (Delyon *et al.*, 1999).

- À l'itération k , pour les paramètres courants $\theta^{(k-1)}$:
 - **S-step** : simulation des effets aléatoires $\psi^{(k)} \sim p(\psi | \mathbf{Y}, \theta^{(k-1)})$;
 - **SA-step** : mise à jour stochastique de l'espérance conditionnelle $Q(\theta | \theta^{(k-1)})$;
 - **M-step** : maximisation pour obtenir $\theta^{(k)}$.



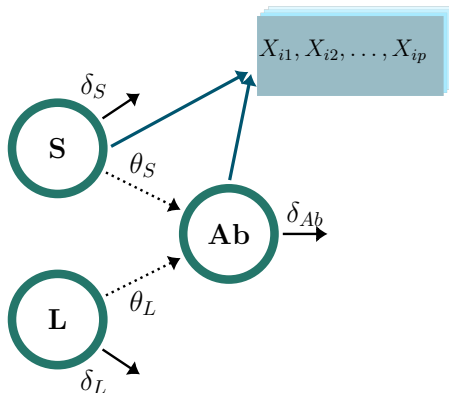
Problématique et objectifs de la thèse

- Les modèles non-linéaires à effets mixtes décrivent des **dynamiques mécanistes complexes et variables**.
- Les données modernes sont **de grande dimension**, et de plus en plus accessibles (Thiébaud *et al.*, 2025).



Problématique et objectifs de la thèse

- Les modèles non-linéaires à effets mixtes décrivent des **dynamiques mécanistes complexes et variables**.
- Les données modernes sont **de grande dimension**, et de plus en plus accessibles (Thiébaud *et al.*, 2025).



Problématique et objectifs de la thèse

Développer des méthodes d'estimation pénalisée pour intégrer des données de grande dimension dans les modèles mécanistes et identifier les biomarqueurs clés de la réponse vaccinale.

- **Lasso-SAMBA** : approche de **sélection Lasso stable** pour construire le **modèle de covariables**.
- **REMixed** : **intégration de biomarqueurs longitudinaux** comme **observations indirectes** de compartiments latents.

Notations et contexte

Modèle structurel

Décrit la dynamique intra-individuelle du phénomène d'intérêt.

$$\frac{d\mathbf{S}_i}{dt} = u(\mathbf{S}_i(t), \boldsymbol{\psi}_i, t).$$

Modèle statistique

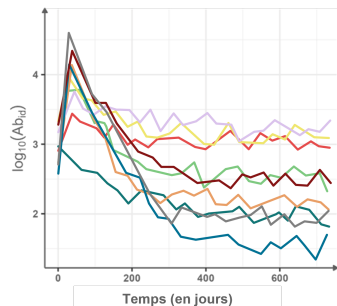
Introduit la variabilité inter-individuelle. Pour chaque individu i :

$$h(\boldsymbol{\psi}_i) = h(\boldsymbol{\psi}_0) + \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim \mathcal{N}(0, \boldsymbol{\Omega}).$$

Modèle d'observation

Relie les prédictions du modèle structurel aux données observées $\mathbf{Y}_{ij}$.

$$\mathbf{Y}_{ij} = g(\mathbf{S}_i(t_{ij}), \boldsymbol{\psi}_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2).$$



Notations et contexte

Modèle structurel

Décrit la dynamique intra-individuelle du phénomène d'intérêt.

$$\frac{d\mathbf{S}_i}{dt} = u(\mathbf{S}_i(t), \psi_i, t).$$

Modèle statistique

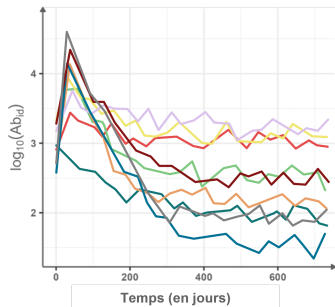
Introduit la variabilité inter-individuelle. Pour chaque individu i :

$$h(\psi_i) = h(\psi_0) + \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim \mathcal{N}(0, \boldsymbol{\Omega}).$$

Modèle d'observation

Relie les prédictions du modèle structurel aux données observées $\mathbf{Y}_{ij}$.

$$\mathbf{Y}_{ij} = g(\mathbf{S}_i(t_{ij}), \psi_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2).$$



Intégrer des covariables de grande dimension dans le modèle statistique, afin d'**identifier celles expliquant la variabilité inter-individuelle** des paramètres du modèle mécaniste.

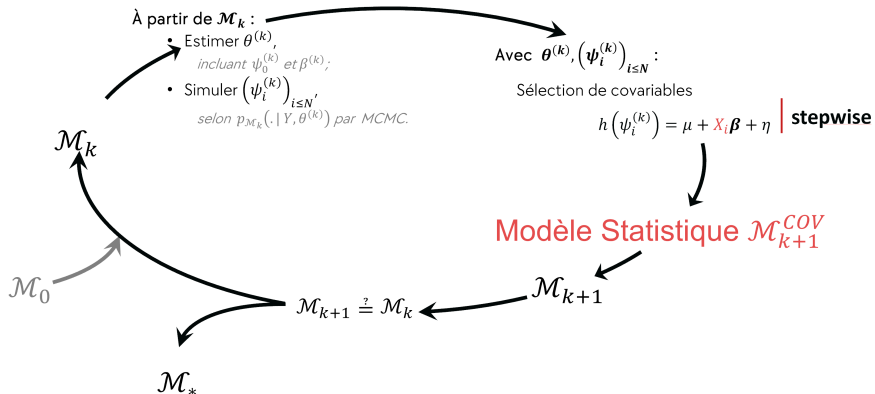
Approches existantes de sélection de covariables

	$p \gg N$	EDOs sans solution analytique
SCM — Stepwise Covariate Modeling (Jonsson et Karlsson, 1998)	✗	✓
COSSAC — COnditional Sampling use for Stepwise Approach based on Correlation tests (Ayrat et al., 2021)	✗	✓
SAEMVS — SAEM Variable Selection (Naveau et al., 2023)	✓	✗
SAMBA — Stochastic Approximation for Model Building Algorithm (Prague et Lavielle, 2022)	✗	✓

Algorithme SAMBA

(Prague et Lavielle, 2022)

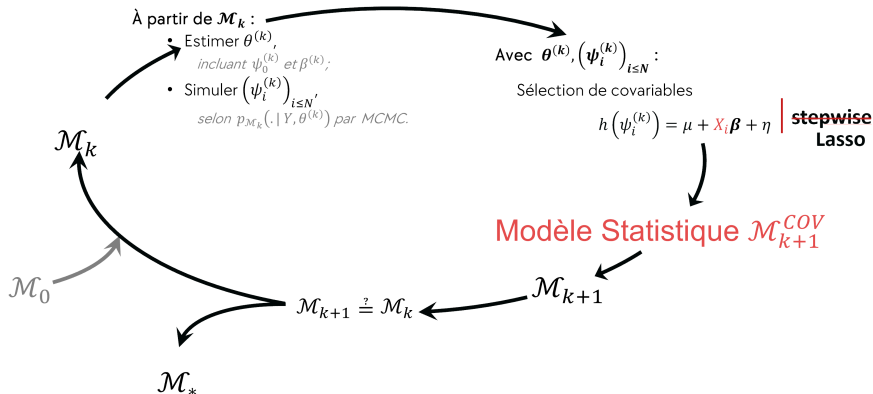
SAMBA est un algorithme itératif, permettant la sélection conjointe du **modèle de covariables**, du modèle d'erreur et de la structure de corrélation.



Algorithme SAMBA

(Prague et Lavielle, 2022)

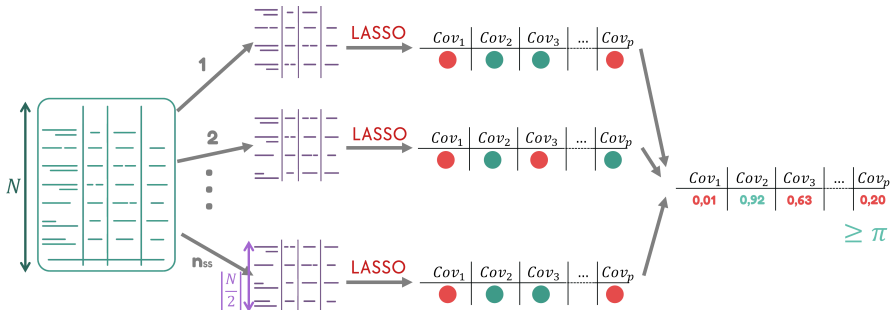
SAMBA est un algorithme itératif, permettant la sélection conjointe du **modèle de covariables**, du modèle d'erreur et de la structure de corrélation.



Algorithme de sélection stable

(Meinshausen et Bühlmann, 2010)

Pour le paramètre de pénalisation λ fixé.



Calibration des paramètres

Cette procédure introduit deux hyperparamètres, λ et π fixés.

Meinshausen et Bühlmann (2010)

$$\mathbb{E}[FDP(\hat{S}^{\lambda,\pi})] \leq U_{\lambda,\pi} = \frac{1}{|\hat{S}^{\lambda,\pi}|} \times \frac{q_{\lambda}^2}{(2\pi - 1)p},$$

où q_{λ} est le nombre moyen de covariables sélectionnées dans les batches pour la pénalisation λ et $\hat{S}^{\lambda,\pi}$ est l'ensemble des covariables sélectionnées par Lasso stable calibré avec λ et π .

- On explore un espace $\Lambda \times \Pi$ d'hyperparamètres (λ, π) .
- On conserve les couples (λ, π) vérifiant $U_{\lambda,\pi} \leq \alpha$,
- Puis on choisit parmi eux celui qui minimise le critère d'information :

$$(\lambda^*, \pi^*) = \arg \min_{U_{\lambda,\pi} \leq \alpha} BIC_{\lambda,\pi}.$$

- Le modèle associé à (λ^*, π^*) n'est retenu comme $\mathcal{M}_{k+1}^{\text{COV}}$ que s'il améliore le critère d'information.

Simulation de réponse humorale

(Pasin *et al.*, 2019)

Modèle de dynamique anticorps :

- Anticorps mesurés pour $N = 100$ individus en 6 temps.

$$\frac{d}{dt} Ab_i(t) = \varphi_{S_i} e^{-\delta_S t} + \varphi_{L_i} e^{-\delta_L t} - \delta_{Ab_i} Ab_i(t)$$

$$Y_{ij} = \log_{10}(Ab_i(t_j)) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

$$t_j \in \{0, 7, 21, 123, 180, 300\}.$$

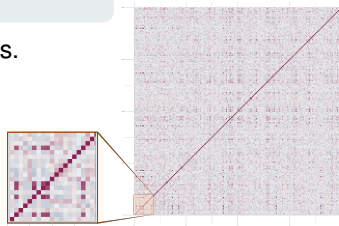
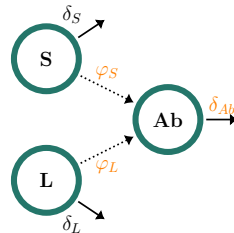
$$\varphi_S = \theta_S \times S(0), \quad \varphi_L = \theta_L \times L(0).$$

- $p = 200$ covariables Gaussiennes corrélées.

Vraies associations :

- $cAGE$ associé à φ_S ,
- G_1 associé à φ_L ,
- G_2 associé à δ_{Ab} .

- Comparaison sur **100 réplicats** indépendants :
step-SAMBA (Prague et Lavielle, 2022), **Lasso-SAMBA** ($\alpha = 10\%$).



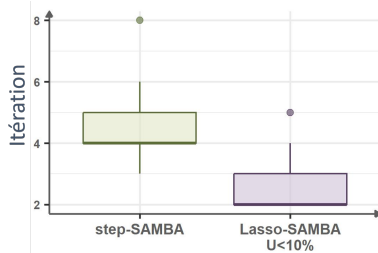
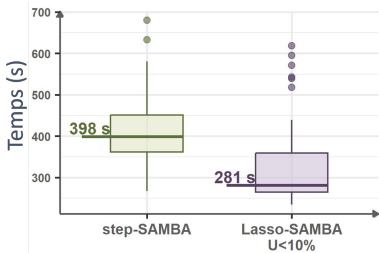
Critères d'évaluation

Lien Paramètre–Covariable	Sélectionné dans le modèle final	Non sélectionné dans le modèle final
Présent dans le vrai modèle	Vrai Positif (TP)	Faux Négatif (FN)
Absent du vrai modèle	Faux Positif (FP)	Vrai Négatif (TN)

- $FDR = \frac{FP}{FP + TP}$: taux de fausses découvertes positives.
- $FNR = \frac{FN}{FN + TP}$: taux de faux négatifs.
- $F_1\text{-score} = \frac{2TP}{2TP + FN + FP}$: compromis sensibilité / spécificité.

Résultats des simulations

Métrique	Méthode	Médiane (IQ 95%)
FDR (%) :	step-SAMBA	72.7 [50.0 ;83.8]
	Lasso-SAMBA	0.0 [0.0 ;25.0]
FNR (%) :	step-SAMBA	0.0 [0.0 ;0.0]
	Lasso-SAMBA	0.0 [0.0 ;0.0]
F_1-score (%) :	step-SAMBA	42.9 [27.9 ;66.7]
	Lasso-SAMBA	100.0 [85.7 ;100.0]

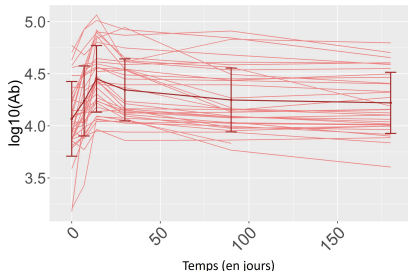


Application sur données réelles

Réponse vaccinale anti-VZV (ZOSTAVAX)

Données issues de l'étude **publique** par **Li et al. (2017)** sur le vaccin **ZOSTAVAX** contre le virus **Varicella-Zona**.

- $N = 35$ volontaires adultes vaccinés avec **ZOSTAVAX** (Arvin, 1996).
- Anticorps mesurés à $t = \{0, 7, 14, 30, 90, 180\}$ jours, et expression génique à $t = \{0, 1, 3, 7\}$ jours.
- Parmi 10 086, on étudie $p = 784$ gènes présélectionnés, annotés pour les voies : interféron, inflammation, cytokines, cycle cellulaire, neutrophiles (Li et al., 2017).



Relier les **signatures précoces d'expression génique (J1)** aux paramètres décrivant la **réponse humorale** du modèle mécaniste.

Modèle mécaniste

Modèle de dynamique humorale :

Modèle de **Andraud et al. (2012); Pasin et al. (2019)**

$$\frac{d}{dt} Ab_i(t) = \varphi_{S_i} e^{-\delta_S t} + \varphi_{L_i} e^{-\delta_L t} - \delta_{Ab_i} Ab_i(t),$$

$$Y_{ij} = \log_{10}(Ab_i(t_j)) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Variabilité inter-individuelle :

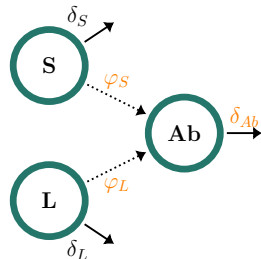
Modèle statistique

$$\log(\varphi_{S_i}) = \log(\varphi_{S_{pop}}) + \mathbf{X}_i \boldsymbol{\beta}^{(S)} + \eta_i^{(S)},$$

$$\log(\varphi_{L_i}) = \log(\varphi_{L_{pop}}) + \mathbf{X}_i \boldsymbol{\beta}^{(L)} + \eta_i^{(L)},$$

$$\log(\delta_{Ab_i}) = \log(\delta_{Ab_{pop}}) + \mathbf{X}_i \boldsymbol{\beta}^{(Ab)} + \eta_i^{(Ab)},$$

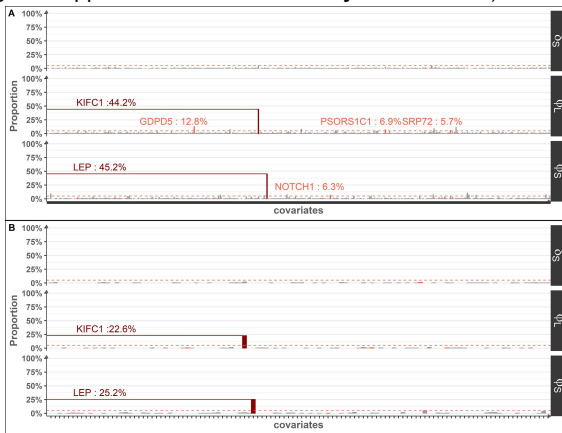
$$\eta_i^{(\cdot)} \sim \mathcal{N}(0, \omega^2), \mathbf{X}_i : \text{gènes candidats (expression à J1)}.$$



Résultats

Résultats de sélection :

- **step-SAMBA** : 15 gènes sélectionnés.
- **Lasso-SAMBA** : modèle parcimonieux avec $LEP \rightarrow \varphi_S$ et $KIFC1 \rightarrow \varphi_L$
(Ces deux gènes appartiennent à la voie du **cycle cellulaire**)



Genes sélectionnés par Lasso- et step-SAMBA

Genes sélectionnés par step-SAMBA seulement

Conclusion & Discussion



Lasso-SAMBA étend l'algorithme SAMBA en intégrant une **sélection Lasso stable** pour la construction du **modèle de covariables** dans les **modèles NLEM basés sur des EDOs**.

- Taux de fausses découvertes faible sans perte de puissance ;
 - Application sur données réelles : vaccin contre **Varicella-Zona** et **Ebola**.
 - Comparaison sur les simulations proposés dans le papier de la méthode SAEMVS (Naveau et al., 2023) ;
 - Comparaison sur des simulations avec distributions hétérogènes ;
 - Comparaison en plus grande dimension $p = 1\,000$.
- Une publication en cours de révision pour Statistics in Medicine.
- Un paquet  LSAMBA, étendant le paquet Rsm1x, disponible sur le **CRAN**.

Conclusion & Discussion

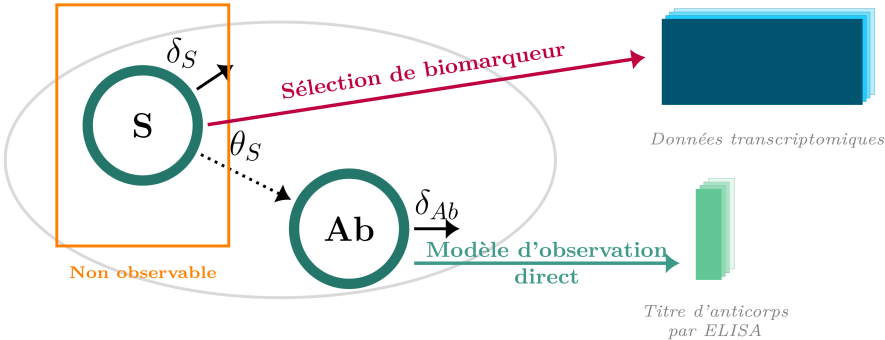
✖ Limites :

- Hypothèse de normalité des effets aléatoires ;
- Coût computationnel élevé pour des covariables de très grande dimension.

💡 Perspectives :

- Pénalités SCAD, MCP ou Lasso adaptatif ;
- Extensions structurées (*group Lasso*, *sparse group Lasso*) pour une sélection par voie ou groupe de gènes ;
- Simulations supplémentaires dans des contextes plus complexes.

Notations et contexte



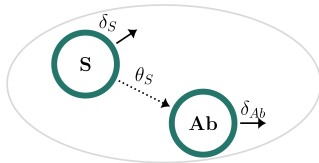
💡 Utiliser les **données d'expression génique** pour **inférer des processus latents** au sein de systèmes mécanistes et ainsi **améliorer l'estimation des paramètres de population**.

Notations et contexte

Modèle structurel

$$\begin{cases} \frac{dS_{pi}}{dt}(t) = F_{Sp}(\mathbf{S}_i(t), R_i(t), \psi_i, t), & p = 1, \dots, P, \\ \frac{dR_i}{dt}(t) = F_R(\mathbf{S}_i(t), R_i(t), \psi_i, t). \end{cases}$$

F_{Sp} et F_R sont des fonctions décrivant les dynamiques.



Modèle statistique

$$h(\psi_i) = h(\psi_0) + \mathbf{X}_i\beta + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \Omega).$$

h est une fonction de lien.

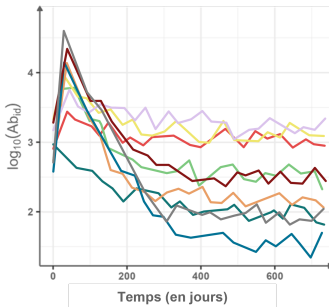
Modèle d'observation

Relie à la fois les compartiments observés $\mathbf{S}$ et le compartiment latent R aux données expérimentales :

$$Y_{pij} = g_p(S_{pi}(t_{pij})) + \epsilon_{pij}, \quad \epsilon_{pij} \sim \mathcal{N}(0, \varsigma_p^2)$$

$$Z_{kij} = \mu_k + \alpha_k s_k(R_i(t_{kij})) + \varepsilon_{kij}, \quad \varepsilon_{kij} \sim \mathcal{N}(0, \sigma_k^2), \quad k \leq K.$$

g_p , s_k sont des fonctions de lien.



Objectif

? Identifier, parmi les K biomarqueurs **longitudinaux** Z_k , ceux informatifs sur la dynamique du compartiment latent R .

💡 Estimation conjointe des paramètres θ et des paramètres pénalisés $\alpha = (\alpha_k)_{k \leq K}$, par maximisation de la **log-vraisemblance pénalisée** :

$$LL_{\text{pen}}(\alpha, \theta) = LL(\alpha, \theta) - \lambda \|\alpha\|_1.$$

$$\theta = (\psi_0, \beta, \Omega, (\sigma_k^2)_{k \leq K}, (s_p^2)_{p \leq P}, (\mu_k)_{k \leq K}).$$

Principe de l'algorithme REMixed

$$\max_{\alpha, \theta} LL_{\text{pen}}(\alpha, \theta) = \max_{\alpha, \theta} LL(\alpha, \theta) - \lambda \|\alpha\|_1.$$

- Utilisation d'un **algorithme cyclique de descente de coordonnées** : alternance entre estimation des paramètres non pénalisés θ et mise à jour des paramètres régularisés α .

REMixed (*Regularized Estimation for non-linear Mixed Effects Models*)

Soit $\lambda > 0$ fixé.

Initialisation : $(\alpha^{(0)}, \theta^{(0)})$.

Itération l (paramètres courants $\alpha^{(l)}, \theta^{(l)}$)

- Mettre à jour $\alpha^{(l+1)}$ par **descente de coordonnées proximale** (mise à jour de type Lasso en fixant $\theta^{(l)}$).
- Mettre à jour $\theta^{(l+1)}$ via **SAEM**, en fixant $\alpha = \alpha^{(l+1)}$.

Critères d'arrêt :

- $|LL_{\text{pen}, \lambda}^{(l+1)} - LL_{\text{pen}, \lambda}^{(l)}| < \tau_1$,
- $\left| \frac{\alpha^{(l+1)} - \alpha^{(l)}}{\alpha^{(l)} + 1} \right| + \left| \frac{\theta^{(l+1)} - \theta^{(l)}}{\theta^{(l)} + 1} \right| < \tau_2$.

Principe de l'algorithme REMixed

$$\max_{\alpha, \theta} LL_{\text{pen}}(\alpha, \theta) = \max_{\alpha, \theta} LL(\alpha, \theta) - \lambda \|\alpha\|_1.$$

- Utilisation d'un **algorithme cyclique de descente de coordonnées** : alternance entre estimation des paramètres non pénalisés θ et mise à jour des paramètres régularisés α .

REMixed (*Regularized Estimation for non-linear Mixed Effects Models*)

Soit $\lambda > 0$ fixé. 🔍

Initialisation : $(\alpha^{(0)}, \theta^{(0)})$. 🔍

Itération l (paramètres courants $\alpha^{(l)}, \theta^{(l)}$)

- Mettre à jour $\alpha^{(l+1)}$ par **descente de coordonnées proximale** (mise à jour de type Lasso en fixant $\theta^{(l)}$). 🔍
- Mettre à jour $\theta^{(l+1)}$ via **SAEM**, en fixant $\alpha = \alpha^{(l+1)}$.

Critères d'arrêt :

- $|LL_{\text{pen}, \lambda}^{(l+1)} - LL_{\text{pen}, \lambda}^{(l)}| < \tau_1,$
- $\left| \frac{\alpha^{(l+1)} - \alpha^{(l)}}{\alpha^{(l)} + 1} \right| + \left| \frac{\theta^{(l+1)} - \theta^{(l)}}{\theta^{(l)} + 1} \right| < \tau_2.$

Mise à jour des paramètres régularisés α

- À cette étape, on fixe $\theta^{(l)}$. On cherche à résoudre :

$$\partial_{\alpha} \left(LL(\alpha, \theta^{(l)}) - \lambda \|\alpha\|_1 \right) = \mathbf{0}.$$

- On approxime $LL(\alpha, \theta^{(l)})$ par un développement de Taylor d'ordre 2 autour de $\alpha^{(l)}$.

$$\begin{aligned} LL(\alpha, \theta^{(l)}) &= LL(\alpha^{(l)}, \theta^{(l)}) + (\alpha - \alpha^{(l)})^T \partial_{\alpha} LL(\alpha, \theta^{(l)})|_{\alpha=\alpha^{(l)}} \\ &\quad - \frac{1}{2} (\alpha - \alpha^{(l)})^T \partial_{\alpha}^2 LL(\alpha, \theta^{(l)})|_{\alpha=\alpha^{(l)}} (\alpha - \alpha^{(l)}). \end{aligned}$$

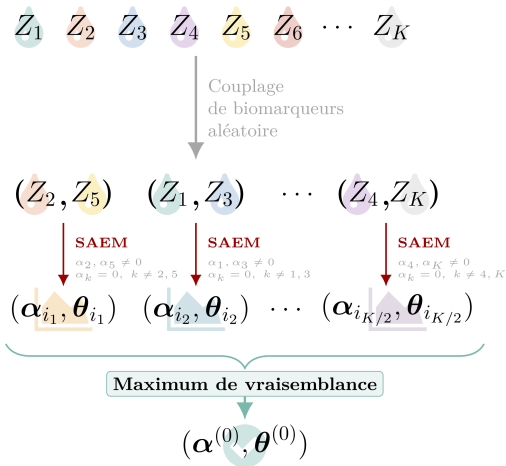
- Mise à jour coordonnée par coordonnée :

$$\alpha_k^{(l+1)} = \begin{cases} \frac{A_k + \lambda}{\partial_{\alpha_k}^2 LL(\alpha, \theta^{(l)})|_{\alpha=\alpha^{(l)}}}, & A_k < -\lambda, \\ \frac{A_k - \lambda}{\partial_{\alpha_k}^2 LL(\alpha, \theta^{(l)})|_{\alpha=\alpha^{(l)}}}, & A_k > \lambda, \\ 0, & \text{sinon.} \end{cases}$$

avec $A_k = \partial_{\alpha_k} LL(\alpha, \theta^{(l)})|_{\alpha=\alpha^{(l)}} + \alpha_k^{(l)} \partial_{\alpha_k}^2 LL(\alpha, \theta^{(l)})|_{\alpha=\alpha^{(l)}}.$

Étape d'initialisation

Objectif : réduire le temps de convergence et garantir une convergence vers une solution plausible et stable.

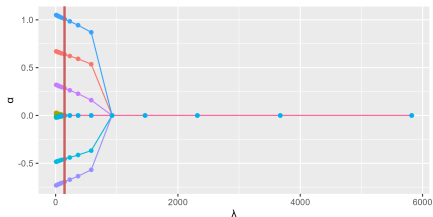


Calibration du paramètre de pénalisation λ

- Application de l'algorithme REMixed pour une grille de $\Lambda = \{\lambda_{\max} \alpha^{l/N_\lambda}\}$ (Park et Hastie, 2007; Friedman *et al.*, 2010).
- Sélection de λ^* minimisant le critère :

$$\lambda^* = \arg \min_{\lambda} BIC_c(\hat{\alpha}_\lambda, \hat{\theta}_\lambda).$$

$$BIC_c(\hat{\alpha}, \hat{\theta}) = -2LL(\hat{\alpha}, \hat{\theta}) + \log(N_{ind}) \dim(\theta_R) + \log(N_{obs}) (\dim(\theta_F) + \#\{k | \alpha_k \neq 0\}).$$



Étape finale

Après sélection de λ^* :

- ré-estimation des paramètres par SAEM sur le sous-ensemble sélectionné pour corriger le biais de pénalisation,
- tests de Wald (Wald, 1943) pour contrôler la proportion de faux positifs.

Simulation de réponse humorale

- Modélisation du processus de production d'anticorps pour $N = 25$ individus.

Modèle mécaniste

$$\begin{cases} \frac{dS_i}{dt}(t) = -\delta_{S_i} S_i(t), \\ \frac{dAb_i}{dt}(t) = \varphi_{S_i} S_i(t) - \delta_{Ab_i} Ab_i(t), \end{cases} \quad \begin{cases} \log(\varphi_{S_i}) = \log(\varphi_{S_{pop}}) + \eta_i^{\varphi_S}, \\ \log(\delta_{S_i}) = \log(\delta_{S_{pop}}) + \eta_i^{\delta_S}, \\ \log(\delta_{Ab_i}) = \log(\delta_{Ab_{pop}}) + \eta_i^{\delta_{Ab}}, \end{cases}$$

$$\eta_i^{\varphi_S}, \eta_i^{\delta_S}, \eta_i^{\delta_{Ab}} \sim \mathcal{N}(0, \omega^2).$$

Modèle d'observation

$$\begin{cases} Y_{ij} = \log_{10}(Ab_i(t_j)) + \epsilon_{ij}, & \epsilon_{ij} \sim \mathcal{N}(0, \sigma_{Ab}^2), \\ Z_{kij} = \mu_k + \alpha_k S_i(t_{kj}) + \varepsilon_{kij}, & \varepsilon_{kij} \sim \mathcal{N}(0, \sigma_k^2). \end{cases}$$

Y_{ij} : titres d'anticorps (jours 0, 7, 21, 123, 180, 300) ;

Z_{kij} : marqueurs transcriptomiques collectés quotidiennement (jours 1 à 21).

Simulation de réponse humorale

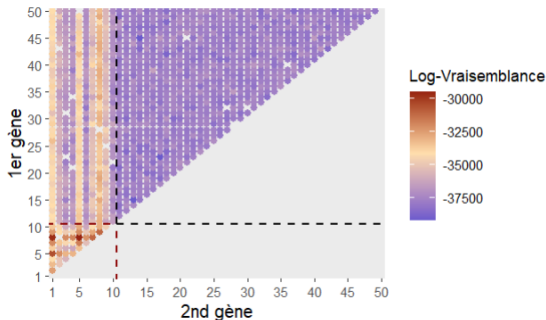
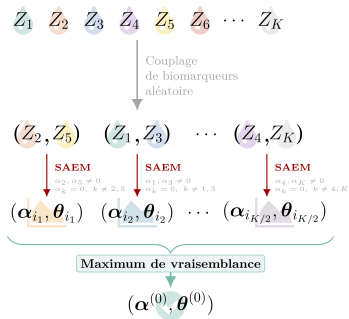
- Deux scénarios étudiés :
 - $K = 20$ biomarqueurs dont $K_1 = 5$ informatifs ;
 - $K = 50$ biomarqueurs dont $K_1 = 10$ informatifs.
- Simulation de **200 réplicats** indépendants pour chaque scénario.

Critères d'évaluation

Lien Paramètre-Covariable	Sélectionné dans le modèle final	Non sélectionné dans le modèle final
Présent dans le vrai modèle	Vrai Positif (TP)	Faux Négatif (FN)
Absent du vrai modèle	Faux Positif (FP)	Vrai Négatif (TN)

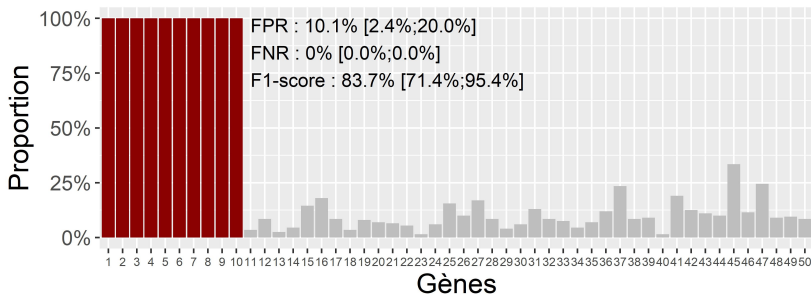
- $FPR = \frac{FP}{FP + TN}$: taux de faux positifs.
- $FNR = \frac{FN}{FN + TP}$: taux de faux négatifs.
- $F_1\text{-score} = \frac{2TP}{2TP + FN + FP}$: compromis sensibilité / spécificité.
- **Évaluation de l'estimation des paramètres :**
 - Biais relatif moyen, écart-type empirique et estimé,
 - Couverture des IC à 95%.

Résultats des simulations



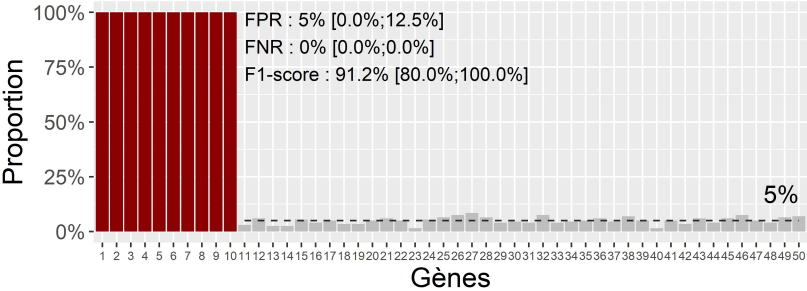
Log-vraisemblance des modèles candidats à l'initialisation.

Résultats des simulations



Résultats de sélection des biomarqueurs avant les tests statistiques.

Résultats des simulations



Résultats après application des tests de Wald (Wald, 1943).

Résultats des simulations

Paramètre	Vraie Valeur	Estimation Moyenne	Biais Relatif (%)	Ecart-Type Empirique	Ecart-Type Estimé	Couverture à 95%
δ_{Ab}	0.025	0.025	0.50	0.0017	0.0015	92.0
δ_S	0.230	0.230	-0.16	0.0071	0.0063	91.5
$\varphi_S \times 10^{-2}$	6.110	6.150	0.71	0.0077	0.0075	91.0
$\omega_{\delta_{Ab}}$	0.300	0.288	-3.90	0.0396	0.0419	93.0
ω_{δ_S}	0.100	0.098	-1.98	0.0184	0.0164	92.5
ω_{φ_S}	0.600	0.573	-4.42	0.1030	0.0903	87.0
α_1	0.650	0.649	-0.15	0.0131	0.0128	93.5
α_2	-0.490	-0.489	-0.22	0.0116	0.0126	96.5
α_3	1.060	1.060	-0.12	0.0149	0.0151	93.5
α_4	-0.720	-0.719	-0.18	0.0140	0.0145	96.0
α_5	0.330	0.329	-0.17	0.0156	0.0148	91.0
μ_1	-0.610	-0.609	-0.14	0.0229	0.0196	91.5
μ_2	0.040	0.041	1.76	0.0189	0.0188	96.0
μ_3	-1.800	-1.800	0.07	0.0253	0.0244	94.5
μ_4	2.400	2.400	-0.04	0.0235	0.0222	95.5
μ_5	1.090	1.090	-0.03	0.0213	0.0214	96.0
σ_{Ab}	0.110	0.110	-0.17	0.0081	0.0077	93.0
σ_1	0.310	0.309	-0.42	0.0087	0.0096	97.0
σ_2	0.310	0.309	-0.17	0.0093	0.0096	96.0
σ_3	0.350	0.348	-0.66	0.0109	0.0109	95.0
σ_4	0.350	0.351	0.16	0.0119	0.0109	94.5
σ_5	0.370	0.368	-0.42	0.0114	0.0114	97.5

Résultats des simulations

Paramètre	Vraie Valeur	Estimation Moyenne	Biais Relatif (%)	Ecart-Type Empirique	Ecart-Type Estimé	Couverture à 95%
δ_{Ab}	0.025	0.025	0.50	0.0017	0.0015	92.0
δ_S	0.230	0.230	-0.16	0.0071	0.0063	91.5
$\varphi_S \times 10^{-2}$	6.110	6.150	0.71	0.0077	0.0075	91.0
$\omega_{\delta_{Ab}}$	0.300	0.288	-3.90	0.0396	0.0419	93.0
ω_{δ_S}	0.100	0.098	-1.98	0.0184	0.0164	92.5
ω_{φ_S}	0.600	0.573	-4.42	0.1030	0.0903	87.0
α_1	0.650	0.649	-0.15	0.0131	0.0128	93.5
α_2	-0.490	-0.489	-0.22	0.0116	0.0126	96.5
α_3	1.060	1.060	-0.12	0.0149	0.0151	93.5
α_4	-0.720	-0.719	-0.18	0.0140	0.0145	96.0
α_5	0.330	0.329	-0.17	0.0156	0.0148	91.0
μ_1	-0.610	-0.609	-0.14	0.0229	0.0196	91.5
μ_2	0.040	0.041	1.76	0.0189	0.0188	96.0
μ_3	-1.800	-1.800	0.07	0.0253	0.0244	94.5
μ_4	2.400	2.400	-0.04	0.0235	0.0222	95.5
μ_5	1.090	1.090	-0.03	0.0213	0.0214	96.0
σ_{Ab}	0.110	0.110	-0.17	0.0081	0.0077	93.0
σ_1	0.310	0.309	-0.42	0.0087	0.0096	97.0
σ_2	0.310	0.309	-0.17	0.0093	0.0096	96.0
σ_3	0.350	0.348	-0.66	0.0109	0.0109	95.0
σ_4	0.350	0.351	0.16	0.0119	0.0109	94.5
σ_5	0.370	0.368	-0.42	0.0114	0.0114	97.5

Résultats des simulations

Paramètre	Vraie Valeur	Estimation Moyenne	Biais Relatif (%)	Ecart-Type Empirique	Ecart-Type Estimé	Couverture à 95%
δ_{Ab}	0.025	0.025	0.50	0.0017	0.0015	92.0
δ_S	0.230	0.230	-0.16	0.0071	0.0063	91.5
$\varphi_S \times 10^{-2}$	6.110	6.150	0.71	0.0077	0.0075	91.0
$\omega_{\delta_{Ab}}$	0.300	0.288	-3.90	0.0396	0.0419	93.0
ω_{δ_S}	0.100	0.098	-1.98	0.0184	0.0164	92.5
ω_{φ_S}	0.600	0.573	-4.42	0.1030	0.0903	87.0
α_1	0.650	0.649	-0.15	0.0131	0.0128	93.5
α_2	-0.490	-0.489	-0.22	0.0116	0.0126	96.5
α_3	1.060	1.060	-0.12	0.0149	0.0151	93.5
α_4	-0.720	-0.719	-0.18	0.0140	0.0145	96.0
α_5	0.330	0.329	-0.17	0.0156	0.0148	91.0
μ_1	-0.610	-0.609	-0.14	0.0229	0.0196	91.5
μ_2	0.040	0.041	1.76	0.0189	0.0188	96.0
μ_3	-1.800	-1.800	0.07	0.0253	0.0244	94.5
μ_4	2.400	2.400	-0.04	0.0235	0.0222	95.5
μ_5	1.090	1.090	-0.03	0.0213	0.0214	96.0
σ_{Ab}	0.110	0.110	-0.17	0.0081	0.0077	93.0
σ_1	0.310	0.309	-0.42	0.0087	0.0096	97.0
σ_2	0.310	0.309	-0.17	0.0093	0.0096	96.0
σ_3	0.350	0.348	-0.66	0.0109	0.0109	95.0
σ_4	0.350	0.351	0.16	0.0119	0.0109	94.5
σ_5	0.370	0.368	-0.42	0.0114	0.0114	97.5

Résultats des simulations

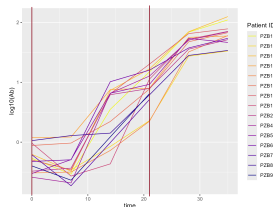
Paramètre	Vraie Valeur	Estimation Moyenne	Biais Relatif (%)	Ecart-Type Empirique	Ecart-Type Estimé	Couverture à 95%
δ_{Ab}	0.025	0.025	0.50	0.0017	0.0015	92.0
δ_S	0.230	0.230	-0.16	0.0071	0.0063	91.5
$\varphi_S \times 10^{-2}$	6.110	6.150	0.71	0.0077	0.0075	91.0
$\omega_{\delta_{Ab}}$	0.300	0.288	-3.90	0.0396	0.0419	93.0
ω_{δ_S}	0.100	0.098	-1.98	0.0184	0.0164	92.5
ω_{φ_S}	0.600	0.573	-4.42	0.1030	0.0903	87.0
α_1	0.650	0.649	-0.15	0.0131	0.0128	93.5
α_2	-0.490	-0.489	-0.22	0.0116	0.0126	96.5
α_3	1.060	1.060	-0.12	0.0149	0.0151	93.5
α_4	-0.720	-0.719	-0.18	0.0140	0.0145	96.0
α_5	0.330	0.329	-0.17	0.0156	0.0148	91.0
μ_1	-0.610	-0.609	-0.14	0.0229	0.0196	91.5
μ_2	0.040	0.041	1.76	0.0189	0.0188	96.0
μ_3	-1.800	-1.800	0.07	0.0253	0.0244	94.5
μ_4	2.400	2.400	-0.04	0.0235	0.0222	95.5
μ_5	1.090	1.090	-0.03	0.0213	0.0214	96.0
σ_{Ab}	0.110	0.110	-0.17	0.0081	0.0077	93.0
σ_1	0.310	0.309	-0.42	0.0087	0.0096	97.0
σ_2	0.310	0.309	-0.17	0.0093	0.0096	96.0
σ_3	0.350	0.348	-0.66	0.0109	0.0109	95.0
σ_4	0.350	0.351	0.16	0.0119	0.0109	94.5
σ_5	0.370	0.368	-0.42	0.0114	0.0114	97.5

Application sur données réelles

Réponse vaccinale anti-SARS-CoV-2 (Pfizer/BioNTech)

Données issues de l'étude **publique** par **Rinchai et al. (2022)** sur le vaccin **Pfizer** contre le virus **SARS-CoV-2**.

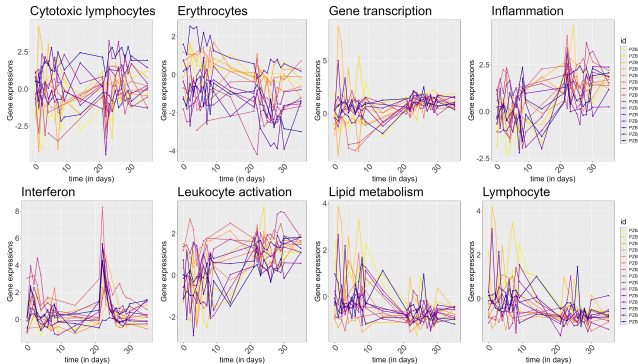
- $N = 15$ volontaires sains, non préalablement infectés.
- Deux doses administrées aux jours 0 et 21.
- Anticorps mesurés à $t = \{0, 7, 14, 21, 28, 35\}$ jours, et expression génique à $t = \{0, \dots, 9, 14, 21, \dots, 30, 35\}$ jours.
- 8 172 gènes regroupés en **34** voies biologiques (**Durinck et al., 2005; Li et al., 2017**). Chaque voie représentée par la moyenne des gènes la composant.



Cinétiques d'anticorps spécifiques SARS-CoV-2 après vaccination (jours 0–35).

Application sur données réelles

Réponse vaccinale anti-SARS-CoV-2 (Pfizer/BioNTech)



Profils d'expression moyens de quelques voies représentatives.

Relier les **profils d'expression génique** à la **production de cellules sécrétices d'anticorps**, afin d'identifier les **voies biologiques informatives** sur la dynamique de la réponse humorale.

Modèle mécaniste

Modèle de *Clairon et al. (2023); Prague et al. (2025)*

$$\begin{cases} \frac{dS_i}{dt} = \rho_i^{(d)} e^{-\delta_V(t-t_d)} - k_m S_i, \\ \frac{dF_1}{dt} = k_m S_i - k_m F_1, \\ \vdots \\ \frac{dF_6}{dt} = k_m F_5 - \delta_S F_6, \\ \frac{dAb_i}{dt} = \theta_i F_6 - \delta_{Ab} Ab_i. \end{cases}$$

$\delta_{Ab} = 0.03$, $\delta_S = 0.1$, $\delta_V = 0.46$, $1/k_m = 4$ jours.

Modèle statistique

$$\begin{aligned} \log(\rho_i^{(2)}) &= \log(\rho_0^{(2)}) + \eta_i^\rho, & \eta_i^{(\cdot)} &\sim \mathcal{N}(0, \omega^2) \\ \log(\theta_i) &= \log(\theta_0) + \eta_i^\theta, \end{aligned}$$

Paramètres individuels : production $\rho_i^{(2)}$ et sécrétion θ_i .

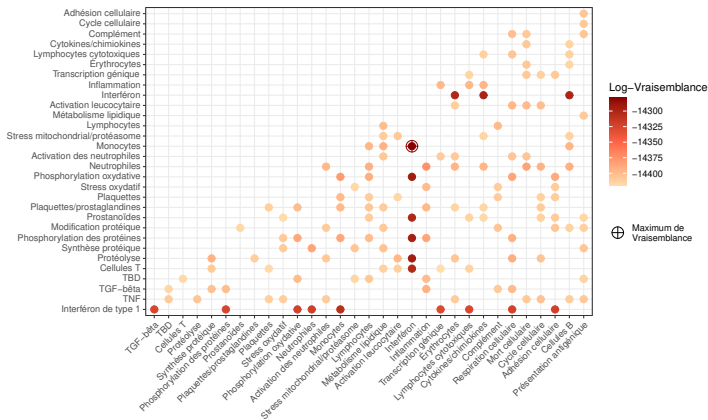
Modèle d'observation

$$\begin{cases} Y_{ij} = \log_{10}(Ab_i(t_{ij})) + \epsilon_{ij}, & \epsilon_{ij} \sim \mathcal{N}(0, \sigma_{Ab}^2), \\ Z_{kij} = \mu_k + \alpha_k \rho^{(d)} e^{-\delta_V(t_{kij} - t_{d-1})} + \varepsilon_{kij}, & \varepsilon_{kij} \sim \mathcal{N}(0, \sigma_k^2). \end{cases}$$

Résultats de l'application

Étape d'initialisation :

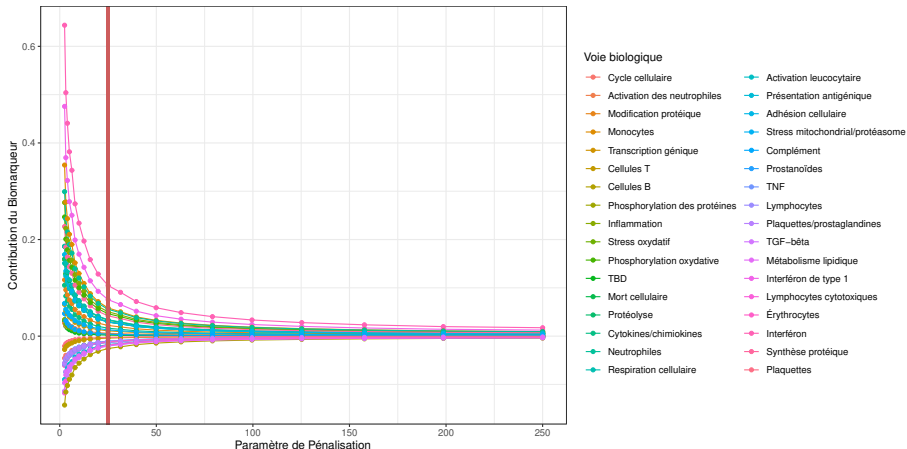
- 17 couples aléatoires de voies biologiques (réalisés sur 10 seeds).
- Meilleur point d'initialisation : **Monocytes–Interféron** avec log-vraisemblance = $-14\,278$.
- Les paires incluant **Interféron** ou **Type I Interféron** ont systématiquement les meilleures vraisemblances.



Résultats de l'application

Calibration du paramètre de pénalisation λ :

- Grille de 30 valeurs de $\lambda \in [2.5, 350]$.
- **BICc minimal atteint pour $\lambda^* = 25$ ($BICc^* = 29\,015$).**



Résultats de l'application

- Après tests de Wald (**Wald, 1943**) : **21 voies biologiques retenues** parmi les 34.
- Principales voies sélectionnées :
 - **Interféron** ($\alpha = 0.105$), **Type I Interféron** ($\alpha = 0.076$), **Monocytes**, **Neutrophiles**, **Cellules B**, **Inflammation**, **TNF**, **Présentation d'antigène**.
- La majorité des voies sélectionnées participent à la **réponse immunitaire humorale précoce** (**Thiébaud et al., 2019; Hagan et al., 2022**).

Conclusion & Discussion



REMixed propose une nouvelle méthodologie pour l'**intégration de données longitudinales** dans des **modèles non-linéaires à effets mixtes basés sur des EDOs**.

- Permet la sélection et l'estimation simultanée des associations gène-compartiment latent et paramètres de population ;
 - Bonne récupération des vrais biomarqueurs, faible biais d'estimation ;
 - Faible taux de faux positifs après test de Wald.
- Une publication en préparation pour être soumis.
- Un paquet  REMixed et une vignette disponible sur le **CRAN**.

Conclusion & Discussion

✖ Limites :

- Coût computationnel élevé ;
- Tendance à sur-sélectionner, amenant le besoin de tests statistiques
→ problème d'inférence post-sélection ;
- Effet moyen μ_k constant dans le temps ;
- Un seul compartiment latent considéré.

💡 Perspectives :

- Modèles multi-compartiments (travail en cours dans l'équipe SISTM) ;
- Corrélation entre biomarqueurs, trajectoires temporelles ;
- Validation croisée pour la calibration du paramètre de pénalisation

Conclusion générale – Bilan & apports

Objectif de la thèse : intégrer des données de grande dimension dans des **modèles NLEM basés sur des EDOs**, pour mieux comprendre les dynamiques biologiques en vaccinologie.



Deux contributions complémentaires :

- **Lasso-SAMBA** ;
- **REMixed**.



Perspectives globales :

- Étudier la convergence et la complexité des algorithmes développés.
- Étudier des contextes de simulations et de modélisations plus complexes, avec des cas mal spécifiés.
- Application dans des contextes plus variés : pharmacogénomique, écologie, etc.

Merci pour votre attention !

Références I

- Mathieu ANDRAUD, Olivier LEJEUNE, Jammbe Z. MUSORO, Benson OGUNJIMI, Philippe BEUTELS et Niel HENS : Living on three time scales : The dynamics of plasma cell and antibody populations illustrated for hepatitis a virus. *PLoS Computational Biology*, 8:e1002418, 3 2012. ISSN 1553-7358.
- A M ARVIN : Varicella-zoster virus. *Clinical Microbiology Reviews*, 9(3):361–381, 1996. URL <https://journals.asm.org/doi/abs/10.1128/cmr.9.3.361>.
- Géraldine AYRAL, Jean-François SI ABDALLAH, Claude MAGNARD et Jonathan CHAUVIN : A novel method based on unbiased correlations tests for covariate selection in nonlinear mixed effects models : The cossac approach. *CPT : Pharmacometrics & Systems Pharmacology*, 10(4):318–329, 2021.
- Barbara BODINIER, Sarah FILIPPI, Therese Haugdahl NØST, Julien CHIQUET et Marc CHADEAU-HYAM : Automated calibration for stability selection in penalised regression and graphical models. *Journal of the Royal Statistical Society Series C : Applied Statistics*, 72:1375–1393, 12 2023. ISSN 0035-9254.

Références II

- Quentin CLAIRON, Mélanie PRAGUE, Delphine PLANAS, Timothée BRUEL, Laurent HOCQUELOUX, Thierry PRAZUCK, Olivier SCHWARTZ, Rodolphe THIÉBAUT et Jérémie GUEDJ : Modeling the kinetics of the neutralizing antibody response against sars-cov-2 variants after several administrations of bnt162b2. *PLOS Computational Biology*, 19:e1011282, 8 2023. ISSN 1553-7358.
- Emmanuelle COMETS, Audrey LAVENU et Marc LAVIELLE : Parameter estimation in nonlinear mixed effect models using saemix , an r implementation of the saem algorithm. *Journal of Statistical Software*, 80, 2017. ISSN 1548-7660.
- Bernard DELYON, Marc LAVIELLE et Eric MOULINES : Convergence of a stochastic approximation version of the em algorithm. *The Annals of Statistics*, 27, 3 1999. ISSN 0090-5364.
- Steffen DURINCK, Yves MOREAU, Arek KASPRZYK, Sean DAVIS, Bart DE MOOR, Alvis BRAZMA et Wolfgang HUBER : Biomart and bioconductor : a powerful link between biological databases and microarray data analysis. *Bioinformatics*, 21:3439–3440, 2005.

Références III

- Jerome FRIEDMAN, Trevor HASTIE et Rob TIBSHIRANI : Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33:1–22, 2010. ISSN 1548-7660. URL <http://www.ncbi.nlm.nih.gov/pubmed/20808728><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2929880>.
- Thomas HAGAN, Bram GERRITSEN, Lewis E. TOMALIN, Slim FOURATI, Matthew P. MULÈ, Daniel G. CHAWLA, Dmitri RYCHKOV, Evan HENRICH, Helen E. R. MILLER, Joann DIRAY-ARCE, Patrick DUNN, Audrey LEE, A. DECKHUT-AUGUSTINE, R. GOTTARDO, E. K. HADDAD, D. A. HAFLE, E. HARRIS, D. FARBER, S. H. KLEINSTEIN, O. LEVY, J. McELRATH, R. R. MONTGOMERY, B. PETERS, B. PULENDRAN, A. RAHMAN, E. F. REED, N. ROUPHAEL, M. M. SARWAL, R. P. SÉKALY, A. FERNANDEZ-SESMA, A. SETTE, K. STUART, A. TOGIAS, J. S. TSANG, Ofer LEVY, Raphael GOTTARDO, Minne M. SARWAL, John S. TSANG, Mayte SUÁREZ-FARIÑAS, Rafick-Pierre SÉKALY, Steven H. KLEINSTEIN et Bali PULENDRAN : Transcriptional atlas of the human immune response to 13 vaccines reveals a common predictor of vaccine-induced antibody responses. *Nature Immunology*, 23:1788–1798, 12 2022. ISSN 1529-2908.

Références IV

- E. Niclas JONSSON et Mats O. KARLSSON : Automated covariate model building within nonmem. *Pharmaceutical Research*, 15:1463–1468, 1998. ISSN 07248741.
- Estelle KUHN et Marc LAVIELLE : Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM : Probability and Statistics*, 8:115–131, 8 2004. ISSN 1292-8100.
- Shuzhao LI, Nicole L. SULLIVAN, Nadine ROUPHAEL, Tianwei Yu, Sophia BANTON, Mohan S. MADDUR, Megan McCausland, Christopher CHIU, Jennifer CANNIFF, Sheri DUBEY, Ken LIU, ViLinh TRAN, Thomas HAGAN, Sai DURASINGHAM, Andreas WIELAND, Aneesh K. MEHTA, Jennifer A. WHITAKER, Shankar SUBRAMANIAM, Dean P. JONES, Alessandro SETTE, Kalpit VORA, Adriana WEINBERG, Mark J. MULLIGAN, Helder I. NAKAYA, Myron LEVIN, Rafi AHMED et Bali PULENDRAN : Metabolic phenotypes of response to vaccination in humans. *Cell*, 169:862–877.e17, 5 2017. ISSN 00928674.
- LIXOFT : *lixoftConnectors : R connectors for Lixoft Suite (@Lixoft)*. Lixoft SAS, Antony, France, 2023. R package version 2023.1.

Références V

- Nicolai MEINSHAUSEN et Peter BÜHLMANN : Stability selection. *Journal of the Royal Statistical Society Series B : Statistical Methodology*, 72:417–473, 9 2010. ISSN 1369-7412.
- Marion NAVEAU, Guillaume KON KAM KING, Renaud RINCENT, Laure SANSONNET et Maud DELATTRE : Bayesian high-dimensional covariate selection in non-linear mixed-effects models using the saem algorithm. *Statistics and Computing*, 34 (1), décembre 2023. ISSN 1573-1375. URL <http://dx.doi.org/10.1007/s11222-023-10367-4>.
- Mee Young PARK et Trevor HASTIE : L1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society Series B : Statistical Methodology*, 69(4):659–677, 08 2007. ISSN 1369-7412. URL <https://doi.org/10.1111/j.1467-9868.2007.00607.x>.
- Chloé PASIN, Irene BALELLI, Thierry Van EFFELTERRE, Viki BOCKSTAL, Laura SOLFOROSI, Mélanie PRAGUE, Macaya DOUGUIH et Rodolphe THIÉBAUT : Dynamics of the humoral immune response to a prime-boost ebola vaccine : Quantification and sources of variation. *Journal of Virology*, 93, 9 2019. ISSN 0022-538X.

Références VI

- M. PRAGUE, A. GABAUT, K. BA, M. ALEXANDRE, B. HEJBLUM et R. THIÉBAUT : Integrating gene expression from whole blood into mechanistic modeling of the immune system : illustration in a model of the antibody response to covid vaccination. En préparation pour être soumis, 2025.
- Mélanie PRAGUE et Marc LAVIELLE : Samba : A novel method for fast automatic model building in nonlinear mixed-effects models. *CPT : Pharmacometrics & Systems Pharmacology*, 11(2):161–172, 2022.
- Darawan RINCHAI, Sara DEOLA, Gabriele ZOPPOLI, Basirudeen Syed Ahamed KABEER, Sara TALEB, Igor PAVLOVSKI, Selma MAACHA, Giusy GENTILCORE, Mohammed TOUFIQ, Lisa MATHEW, Li LIU, Fazulur Rehaman VEMPALLI, Ghada MUBARAK, Stephan LORENZ, Irene SIVIERI, Gabriella CIRMENA, Chiara DENTONE, Paola CUCCAROLO, Daniele Roberto GIACOBBE, Federico BALDI, Alberto GARBARINO, Benedetta CIGOLINI, Paolo CREMONESI, Michele BEDOGNETTI, Alberto BALLESTRERO, Matteo BASSETTI, Boris P. HEJBLUM, Tracy AUGUSTINE, Nicholas Van PANHUYS, Rodolphe THIEBAUT, Ricardo BRANCO, Tracey CHEW, Maryam SHOJAEI, Kirsty SHORT, Carl G. FENG, SUSU M. ZUGHAIER, Andrea De MARIA, Benjamin TANG, Ali Ait Hssain, Davide BEDOGNETTI, Jean-Charles GRIVEL et Damien CHAUSSABEL : High-temporal

Références VII

resolution profiling reveals distinct immune trajectories following the first and second doses of covid-19 mrna vaccines. *Science Advances*, 8, 11 2022. ISSN 2375-2548.

Rodolphe THIÉBAUT, Edouard LHOMME, Hakim HOCINI, Isabelle PELLEGRIN, Andrea BOIZARD-MORACCHINI, Alexandre DUVIGNAUD, Maud PERPÈRE, Mélanie HUCHON, Mélanie PRAGUE, Christine LACABARATZ, Mathieu SURENAUD, Xavier ANGLARET, Denis MALVY, Boris HEJBLUM, COVERAGE study GROUP et Yves LEVY : Self-collected finger-prick blood for gene expression profiling : Unveiling early immune responses in mild covid-19. *medRxiv*, 2025. URL <https://www.medrxiv.org/content/early/2025/05/13/2025.05.06.25327063>.

Rodolphe THIÉBAUT, Boris P. HEJBLUM, Hakim HOCINI, Henri BONNABAU, Jason SKINNER, Monica MONTES, Christine LACABARATZ, Laura RICHERT, Karolina PALUCKA, Jacques BANCHEREAU et Yves LÉVY : Gene expression signatures associated with immune and virological responses to therapeutic vaccination with dendritic cells in hiv-infected individuals. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00874>.

Références VIII

Robert TIBSHIRANI : Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society : Series B (Methodological)*, 58:267–288, 1 1996. ISSN 00359246. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1996.tb02080.x>.

Abraham WALD : Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54:426–482, 1943. ISSN 0002-9947.