

# *Estimation paramétrique dans des modèles définis par un système d'équations différentielles ordinaires*

Sophie Donnet<sup>1</sup> et Adeline Samson<sup>2</sup>

<sup>1</sup> Université Paris XI, Orsay

<sup>2</sup> INSERM U 738, Hôpital Bichat-Claude Bernard

# Contexte

---

- Utilisation de modèles mixtes en pharmacocinétique-pharmacodynamie (PK/PD), pour l'analyse de données longitudinales, en imagerie médicale, en agronomie, etc...
- Description des phénomènes physiologiques/biologiques à l'aide d'un système d'équations différentielles
  - modèles compartimentaux en PK/PD
  - modèle physiologique en IRMf
  - modèle de décroissance de charge virale chez des patients VIH
- La plupart de ces systèmes n'ont pas de solution analytique

## Exemple de modèle

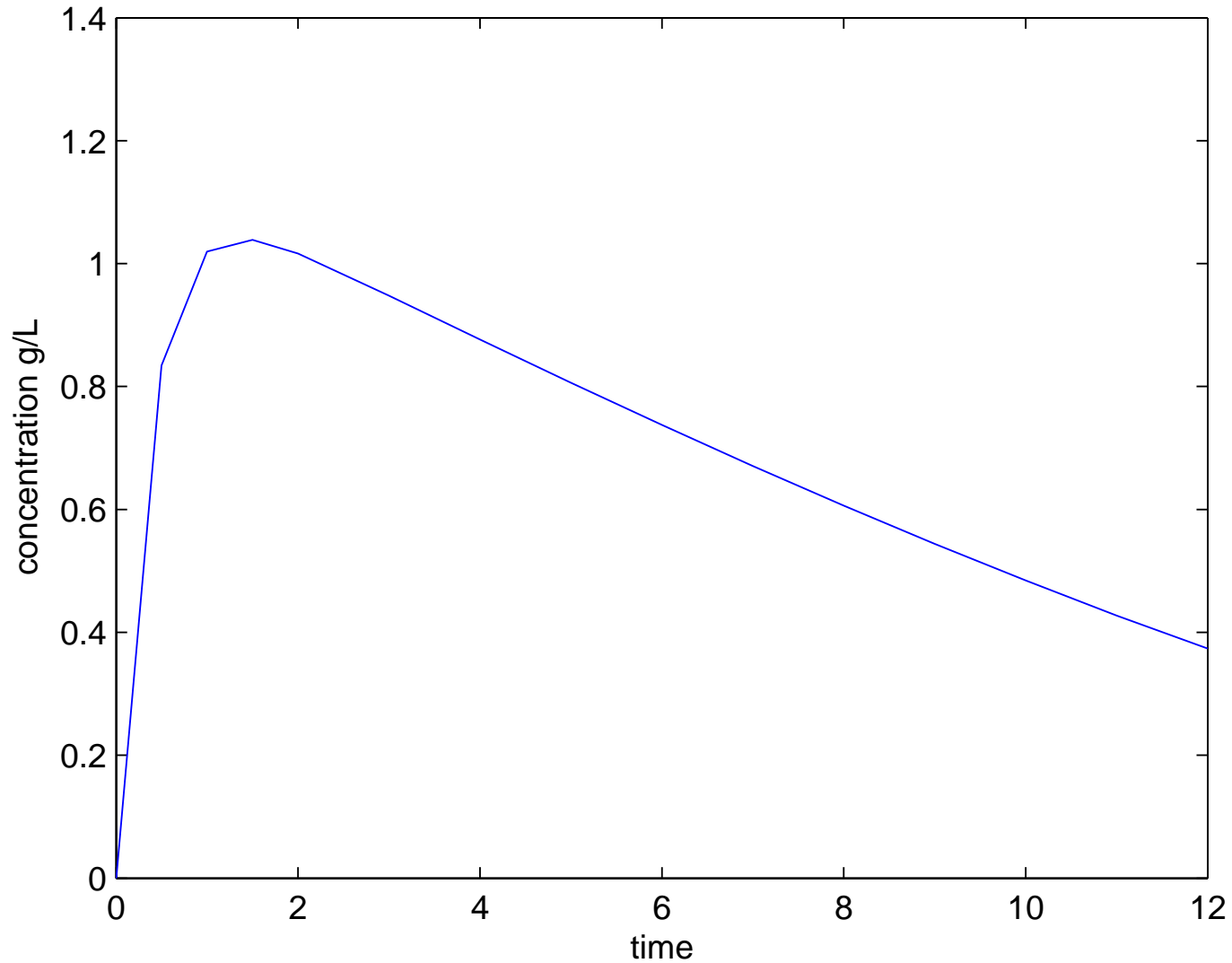
Modèle de PK de Michaelis-Menten à un compartiment avec absorption du premier ordre :

$$\frac{dC}{dt}(t, \phi) = \frac{k_a \cdot Dose}{V} e^{-k_a t} - \frac{V_m \cdot C(t, \phi)}{k_m + C(t, \phi)} \quad (1)$$

- $C(t, \phi)$  (g/L) concentration du médicament
- $Dose$  quantité connue fixée
- $\phi = (k_m, V_m, k_a, V)$
- $k_m$  (g/L) constante d'élimination
- $V_m$  (g/L/h) volume maximal d'élimination
- $k_a$  (h<sup>-1</sup>) constante d'absorption de la molécule
- $V$  (L) volume total de distribution de la molécule dans le corps
  
- ODE sans solution analytique

# Évolution de la concentration

Modèle avec absorption du premier ordre et élimination saturable de Michaelis–Menten



# Modèle mixte avec équation différentielle

- Modèle statistique non linéaire à effets mixtes

$$y_{ij} = a(t_{ij}, \phi_i) + \varepsilon_{ij} \quad 1 \leq j \leq n_i, \quad 1 \leq i \leq N$$

$$\varepsilon_{ij} \sim_{i.i.d} \mathcal{N}(0, \sigma^2)$$

$$\phi_i \sim_{i.i.d.} \pi_\phi(\cdot; \beta)$$

- $\phi_i$  vecteur de paramètres individuels aléatoires non observés
- Estimation de  $\theta = (\beta, \sigma^2)$
- Fonction de régression  $a = H \circ f$  telle que

$$\frac{\partial f(t, \phi)}{\partial t} = F(f(t, \phi), t, \phi)$$

$$f(t_0, \phi) = f_0$$

$$t \in [t_0, T]$$

# Méthodes d'estimation des modèles mixtes

---

- Par maximum de vraisemblance : algorithme SAEM
  - Simulation des données manquantes sous la loi  $p_{\phi|y}(\cdot; \theta)$  par algorithme de Monte Carlo par Chaîne de Markov (MCMC) (Kuhn et Lavielle, 2004)
- Par approche bayésienne
  - Loi *a priori*  $\pi_{\theta}$  sur  $\theta$
  - Estimation de la distribution *a posteriori* de  $\theta$   $p(\theta|y)$  par Gibbs sampling
  - Simulation par algorithme MCMC

## Difficultés

---

- Utilisation d'un algorithme MCMC dans toutes les méthodes d'estimation:
  - ⇒ Besoin de calculer la quantité  $f(\phi, t)$  à chaque itération de l'algorithme MCMC
- Introduction d'une méthode numérique de résolution d'ODE
  - Runge-Kutta, local linearization scheme..
- Travail sur un modèle où on approche la fonction de régression

# Introduction d'un modèle approché

- Modèle  $\mathcal{M}_h$

$$y_{ij} = a_h(t_{ij}, \phi_i) + \varepsilon_{ij} \quad 1 \leq j \leq n_i, \quad 1 \leq i \leq N$$

$$\varepsilon_{ij} \sim_{i.i.d} \mathcal{N}(0, \sigma^2)$$

$$\phi_i \sim_{i.i.d.} \pi_\phi(\cdot; \beta)$$

- Fonction de régression  $a_h = H \circ f_h$
- $f_h$  est l'évaluation de la solution de l'ODE par une méthode numérique d'ordre  $p$  et de pas  $h$ :

$$\sup_t |f_h(t, \phi) - f(t, \phi)| \leq Ch^p$$

- On indice par  $h$  tout ce qui concerne ce modèle
- Quel contrôle de l'estimation des paramètres du modèle ?

# Objectifs

---

1. Étude des différentes méthodes numériques de résolution d'ODE
2. Étude de l'algorithme MCMC de simulation des données manquantes pour les modèles définis par ODE et contrôle de la distance entre les lois conditionnelles  $p_h(\phi|y; \theta)$  et  $p(\phi|y; \theta)$
3. Étude de l'algorithme SAEM et contrôle de la distance entre les vraisemblances
4. Étude de l'algorithme de Gibbs et contrôle de la distance entre les distributions a-posteriori  $p_h(\theta|y)$  et  $p(\theta|y)$
5. Étude sur données simulées
6. Étude d'un jeu de données réelles

## 1.1 Méthodes de résolution d'un système d'ODE

- Système d'équations différentielles considéré :

$$\begin{aligned}\frac{\partial f(t, \phi)}{\partial t} &= F(f(t, \phi), t, \phi) \\ f(t_0, \phi) &= f_0\end{aligned}$$

- $t \in [t_0, T]$
  - $f_0 \in \mathbb{R}^d$  condition initiale du système d'ODE
  - $f : \mathbb{R} \times \mathbb{R}^k \longrightarrow \mathbb{R}^d$ ,  $H : \mathbb{R}^d \longrightarrow \mathbb{R}$ ,  $F : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^k \longrightarrow \mathbb{R}^d$
- Hypothèses supplémentaires
    - H1: la loi  $\pi$  de  $\phi$  est à support compact
    - H2: la fonction  $H$  est lipschitzienne
    - H3: la fonction  $F$  est de classe  $\mathcal{C}^2$

## 1.2 Méthodes de résolution classiques d'ODE

---

- Méthodes à un pas
  - Schéma d'Euler
  - Runge-Kutta avec coefficients à déterminer selon l'ordre
- Méthodes à pas multiples
  - Méthode d'Adams-Bashforth : explicite
  - Méthode d'Adams-Moulton : implicite (pour les systèmes "Stiff")
- Méthode PECE : Prédiction-Évaluation-Correction-Évaluation
- Fonctions dans Matlab6.5
  - ODE45 : Runge-Kutta d'ordre 4 (coef. de Dormand-Prince)
  - ODE23 : Runge-Kutta d'ordre 3 (coef. de Bogacki-Shampine)
  - ODE113 : PECE utilisant Adams-Bashforth et Adams-Moulton
  - ODE23s : Runge-Kutta d'ordre 3 implicite pour systèmes "Stiff"

## 1.3 Schéma de Linéarisation Locale (LL)

---

- Principe
  - Linéarisation locale par rapport au temps de l'ODE
  - Intégration exacte de l'ODE linéaire obtenue
- Convergence (Ramos et Garcia-López, 1997)
  - $f_h$  l'approximation de  $f$  obtenue par LL de pas  $h$
  - Sous les hypothèses H1 et H3
  - Il existe une constante  $C$  indépendante de  $\phi$  telle que

$$\sup_{t \in [t_0, T], \phi} |f(t, \phi) - f_h(t, \phi)| \leq Ch^2$$

## 1.4 Implémentation du schéma LL

---

- Calcul d'exponentielles de matrices
  - Nouvelles méthodes de Padé, Schur, ...
- Ordre moins élevé que ODE45
  - Moins précis sur les systèmes "non Stiff"
- Avantages
  - Grande stabilité dans les systèmes de grande dimension et/ou "Stiff"
  - Généralisation aux SDE avec bruit additif

## 1.5 Comparaison des temps de calculs

- Résolution de l'ODE pour 100 valeurs de  $\phi$  distinctes, choisies aléatoirement, chaque résolution étant réalisée dix fois

méthode	ODE45	ODE23	ODE23s	ODE113	$LL_{h=0.05}$	$LL_{h=0.1}$
temps	0.107	0.096	0.177	0.382	0.327	0.162
écart type	0.100	0.011	0.029	0.736	0.031	0.022

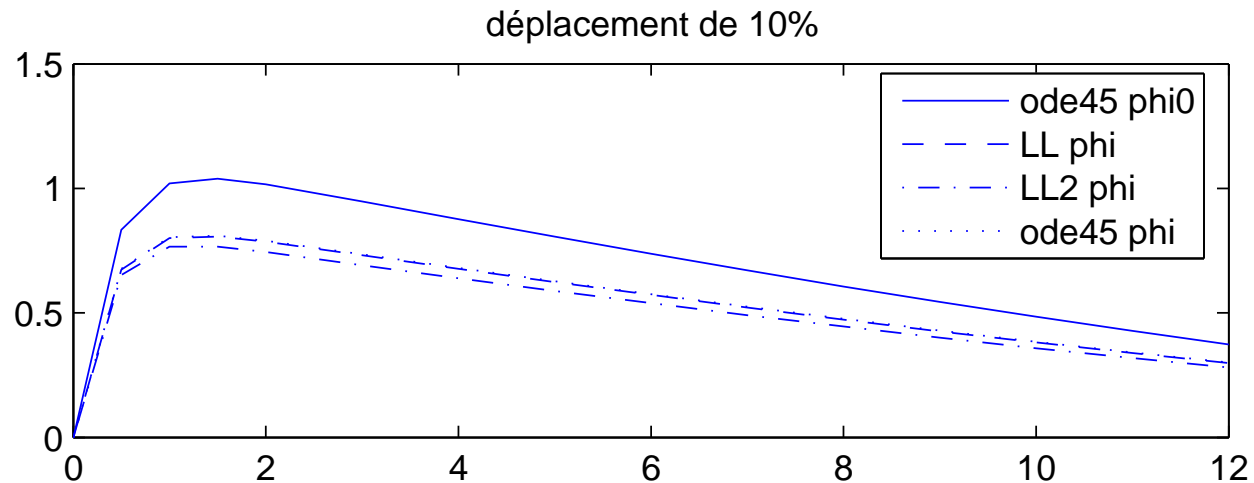
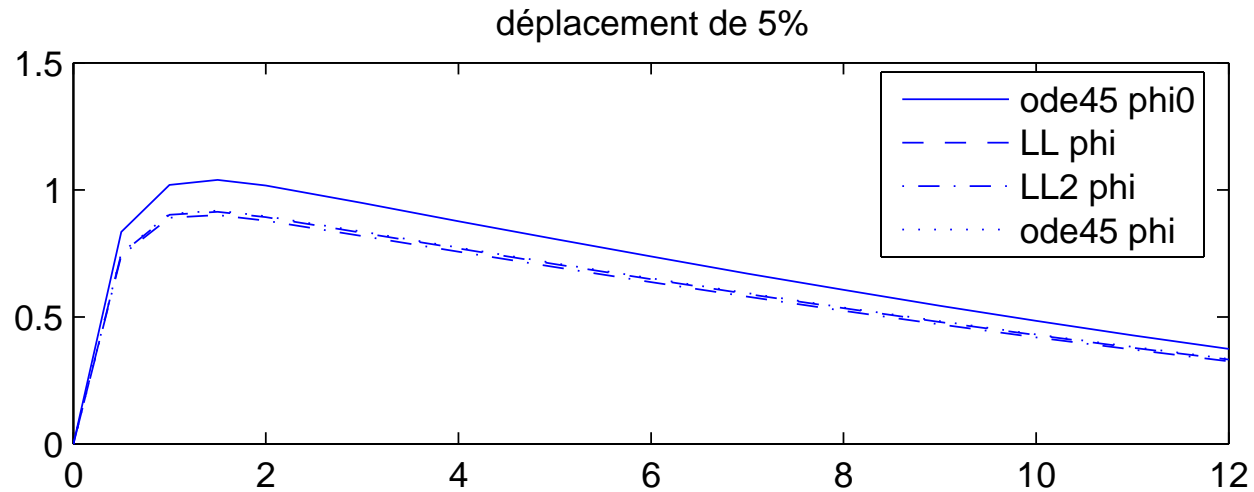
- ODE45 et ODE23 plus rapides

## 1.6 Un deuxième schéma de linéarisation locale : LL2

- But : optimisation du temps de calcul pour le noyau  $K_2$  dans l'algorithme MCMC
- Principe
  - Évaluation de  $f(t, \phi)$  pour  $\phi$  dans un voisinage de  $\phi_0$  et avec  $f(t, \phi_0)$  déjà évalué par LL
  - Développement de Taylor en  $(t, \phi_0)$  de l'ODE
  - Pas de calculs d'exponentielles de matrices
- Convergence
  - Sous H1, H3 et pour  $\phi$  proche de  $\phi_0$ ,
  - on note  $f_{h, \phi_0}$  la solution obtenue par LL2
  - Il existe  $C_1$  et  $C_2$  indépendantes de  $\phi$  telle que

$$\sup_{t \in [t_0, T]} |f(t, \phi) - f_{h, \phi_0}(t, \phi)| \leq \max(C_1 h^2, C_2 \|\phi - \phi_0\|_{\mathbb{R}^k}^2)$$

## 1.7 LL versus LL2



LL2 sept fois plus rapide que LL

## 2.1 Principe de la méthode MCMC

- A chaque itération de l'algorithme d'estimation, à  $\theta$  fixé
- Génère une chaîne de Markov de loi invariante  $p_h(\phi|y; \theta)$ , stationnaire sous certaines conditions
- Utilise un noyau de transition  $q_\theta(\phi'|\phi)$ , simulable, connu analytiquement ou symétrique
- A l'itération  $\ell$ 
  1. Génération d'un candidat  $\phi^c$  sous la loi  $q_\theta(\cdot|\phi^{(\ell)})$
  2. Actualisation

$$\phi^{(\ell+1)} = \left\{ \begin{array}{ll} \phi^c & \text{avec probabilité } \rho(\phi^{(\ell)}, \phi^c) \\ \phi^{(\ell)} & \text{avec probabilité } 1 - \rho(\phi^{(\ell)}, \phi^c) \end{array} \right\}$$

$$\text{où } \rho(\phi^{(\ell)}, \phi^c) = \min \left\{ \frac{p_h(\phi^c|y; \theta) q_\theta(\phi^{(\ell)}|\phi^c)}{p_h(\phi^{(\ell)}|y; \theta) q_\theta(\phi^c|\phi^{(\ell)})}, 1 \right\}$$

## 2.2 Noyau 1 pour la méthode MCMC

---

- Noyau  $K_1$ : la distribution à priori  $q_\theta(\phi'|\phi) = \pi(\phi'; \beta)$

- Probabilité d'acceptation

$$\rho(\phi^{(\ell)}, \phi^c) = \min \left\{ \frac{p_h(y|\phi^c, \theta)}{p_h(y|\phi^{(\ell)}; \theta)}, 1 \right\}$$

- Algorithme de Hastings-Metropolis *indépendant*
- Requiert d'évaluer  $f(\phi^c, t)$

## 2.3 Noyau 2 pour la méthode MCMC

- Noyau  $K_2$ : un petit déplacement :  $\phi^c = \phi^{(\ell)} + \delta$

- On suppose qu'on reste dans un compact (H1)
- Noyau symétrique  $q_\theta(\phi'|\phi) = q_\theta(\phi|\phi')$
- Probabilité d'acceptation

$$\rho(\phi^{(\ell)}, \phi^c) = \min \left\{ \frac{p_h(y|\phi^c; \theta)\pi(\phi^c; \beta)}{p_h(y|\phi^{(\ell)}; \theta)\pi(\phi^{(\ell)}; \beta)}, 1 \right\}$$

- Requier d'évaluer  $f(\phi^c, t)$  connaissant  $f(\phi^{(\ell)}, t)$  et  $\phi^c$  dans un petit voisinage de  $\phi^{(\ell)}$   
⇒ Situation du schéma LL2

## 2.4 Propriétés de l'algorithme MCMC

---

### Théorème 1

- Sous les hypothèses H1, H2 et H3,
  1. L'algorithme MCMC converge sur le modèle  $\mathcal{M}_h$  avec les deux noyaux présentés ci-dessus
  2. Il existe une constante  $C$  telle que pour  $h$  petit

$$D(p_{\phi|y}(\cdot; \theta), p_{h,\phi|y}(\cdot; \theta)) \leq Ch^p$$

## 3.1 L'algorithme EM

- Fonction de vraisemblance appartient à la famille exponentielle

$$p(y, \phi, \theta) = \exp \{ -\psi(\theta) + \langle S(y, \phi), \Phi(\theta) \rangle \}$$

- A l'itération  $k$ 
  - Étape E : calcul de

$$s_{k+1} = \mathbb{E} [ S(y, \phi) | \theta_k ]$$

- Étape M : réactualisation de  $\theta_k$  par le maximum

$$\theta_{k+1} = \arg \max_{\theta} ( -\Psi(\theta) + \langle s_{k+1}, \Phi(\theta) \rangle )$$

## 3.2 L'algorithme SAEM

### 1. Delyon, Lavielle et Moulines, 1999

- Étape E
  - Étape S : simulation de  $\phi_k$  sous la distribution  $p(\phi|y; \theta_k)$
  - Étape SA : approximation stochastique de  $s_k$  avec  $(\gamma_k)$  suite de pas décroissants vers 0

$$s_{k+1} = s_k + \gamma_k \{S(y, \phi_k) - s_k\}$$

- Étape M : réactualisation de  $\theta_k$  par

$$\theta_{k+1} = \arg \max_{\theta} (-\Psi(\theta) + \langle s_{k+1}, \Phi(\theta) \rangle)$$

### 2. Kuhn et Lavielle, 2004

- Étape S : procédure MCMC pour construire à l'étape  $k$  une chaîne de Markov de probabilité de transition  $\Pi_{\theta_k}$  ayant pour unique loi limite stationnaire la loi  $p(\phi|y; \theta_k)$

## 3.3 Propriétés de l'algorithme SAEM

### Théorème

- Sous les hypothèses H1, H2, H3,
  - Soit  $(\gamma_k)$  tels que  $\sum_k \gamma_k = \infty$  et  $\sum_k \gamma_k^2 < \infty$
1. Supposons que  $(s_k)_{k \geq 0}$  appartient à un compact, alors la suite  $(\theta_k)$  d'estimateurs fournie par l'algorithme SAEM sur le modèle  $\mathcal{M}_h$  converge presque sûrement vers un point  $\theta_\infty$ , maximum (local) de la fonction de vraisemblance  $g_h$
  2. Pour tout  $\sigma_0^2 > 0$ , il existe une constante  $C$  telle que

$$\sup_{\theta=(\beta, \sigma^2) \mid \sigma^2 > \sigma_0^2} |p(y; \theta) - p_h(y; \theta)| \leq Ch^p$$

## 4.1 Algorithme bayésien de Gibbs

---

- Loi *a priori*  $\pi_\theta$  sur  $\theta$
- Estimation de la distribution *a posteriori* de  $\theta$   $p(\theta|y)$  par Gibbs sampling
- A l'iteration  $k$ ,
  1. Simulation de  $\phi^{(k)}$  sous la loi  $p_{\phi|y}(\cdot; \theta^{(k-1)})$  par algorithme MCMC
  2. Simulation de  $\theta^{(k)}$  sous la loi  $p_{\theta|y}(\cdot; \phi^{(k)})$

## 4.2 Propriétés de l'algorithme de Gibbs

---

### Théorème

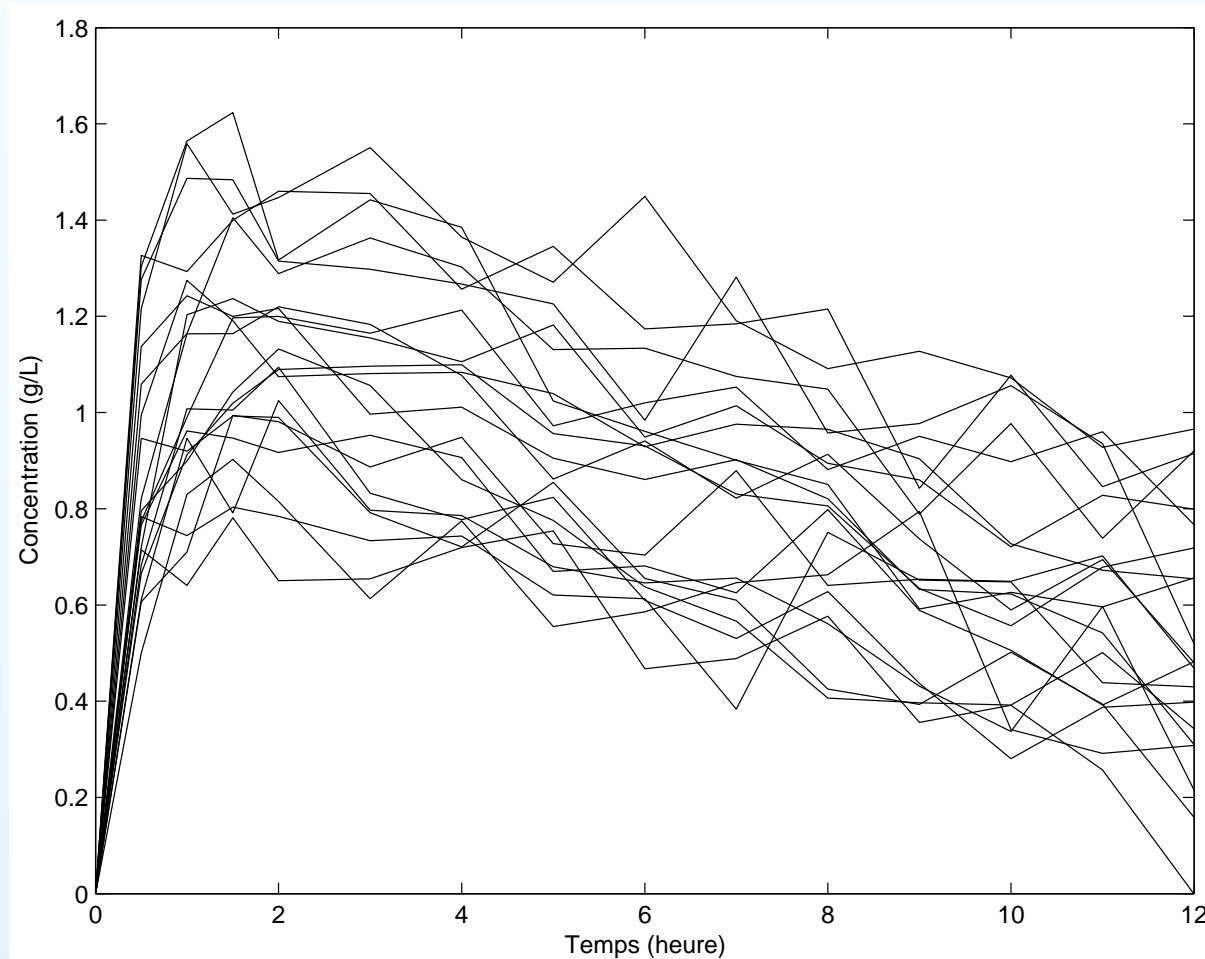
- Sous les hypothèses H1, H2, H3,
  1. L'algorithme de Gibbs converge sur le modèle  $\mathcal{M}_h$
  2. Il existe une constante  $C$  telle que

$$D(p(\theta|y), p_h(\theta|y)) \leq Ch^p$$

## 5.1 Exemple de données simulées

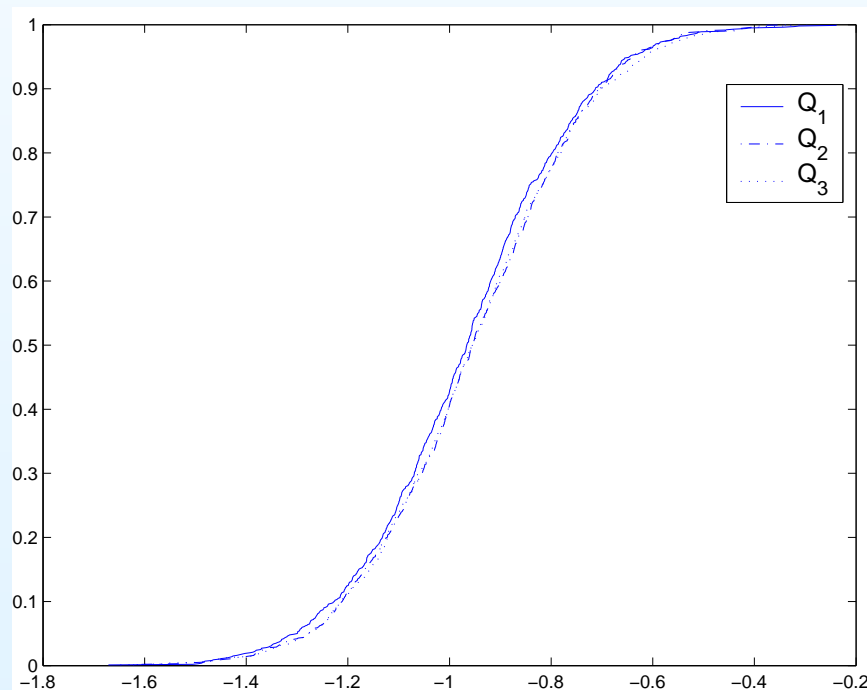
Exemple de pharmacocinétique avec  $N=20$  patients indépendants,

$$\pi_{\phi}(\cdot; \beta) = \mathcal{N}(\mu, \Omega)$$



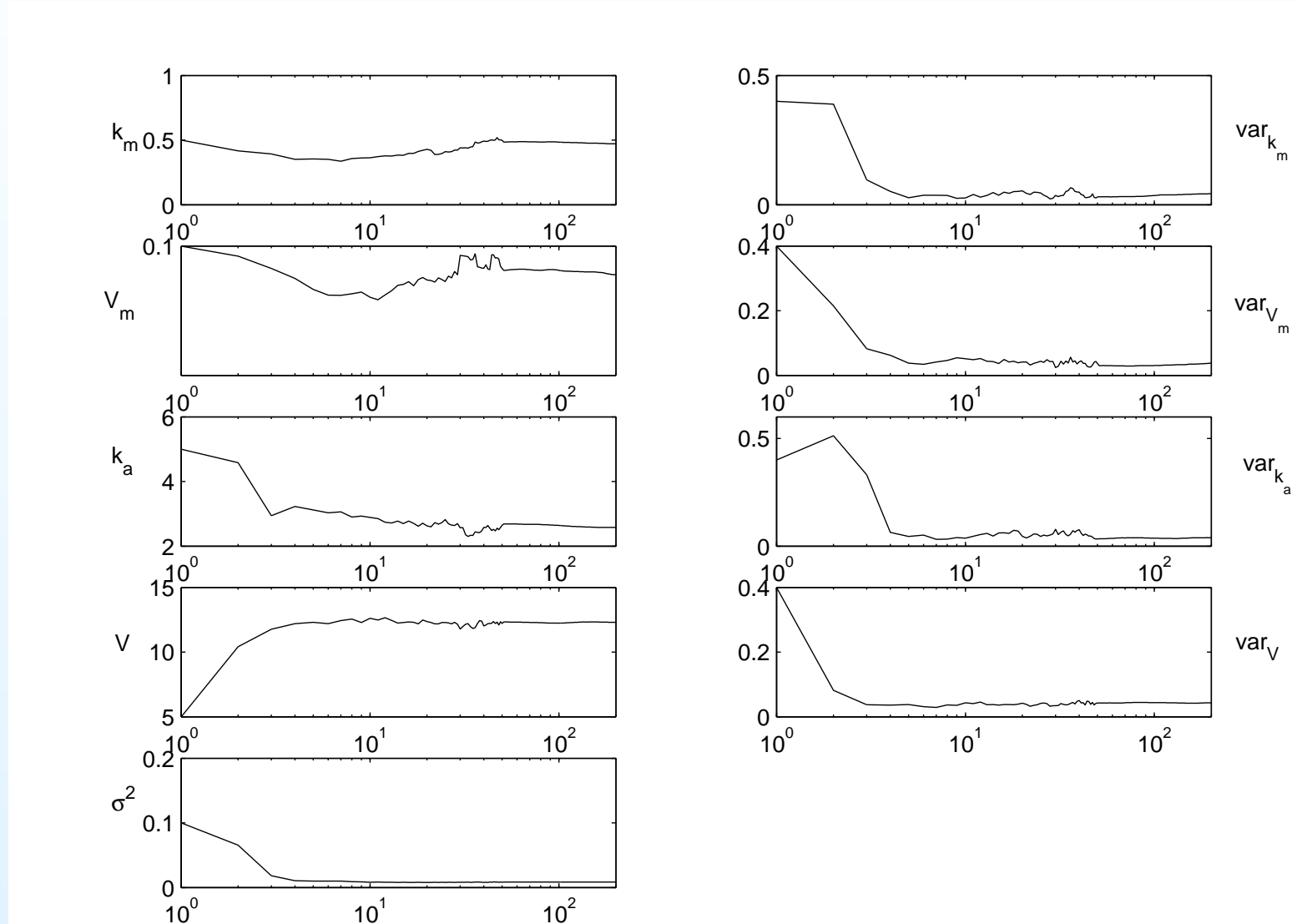
## 5.2 Fonctions de répartition empiriques des distributions simulées par MCMC-ODE

- $Q_1$  : noyau  $K_1$  avec ODE45 + noyau  $K_2$  avec ODE45
- $Q_2$  : noyau  $K_1$  avec ODE45 + noyau  $K_2$  avec LL+LL2
- $Q_3$  : noyau  $K_1$  avec LL + noyau  $K_2$  avec LL+LL2



## 5.3 Convergence de SAEM

200 itérations SAEM, 20 itérations MCMC pour chaque noyau



## 5.4 Résultats d'estimation

- Comparaison de SAEM avec NONMEM
  - NONMEM ne converge pas
  - SAEM estime correctement les paramètres

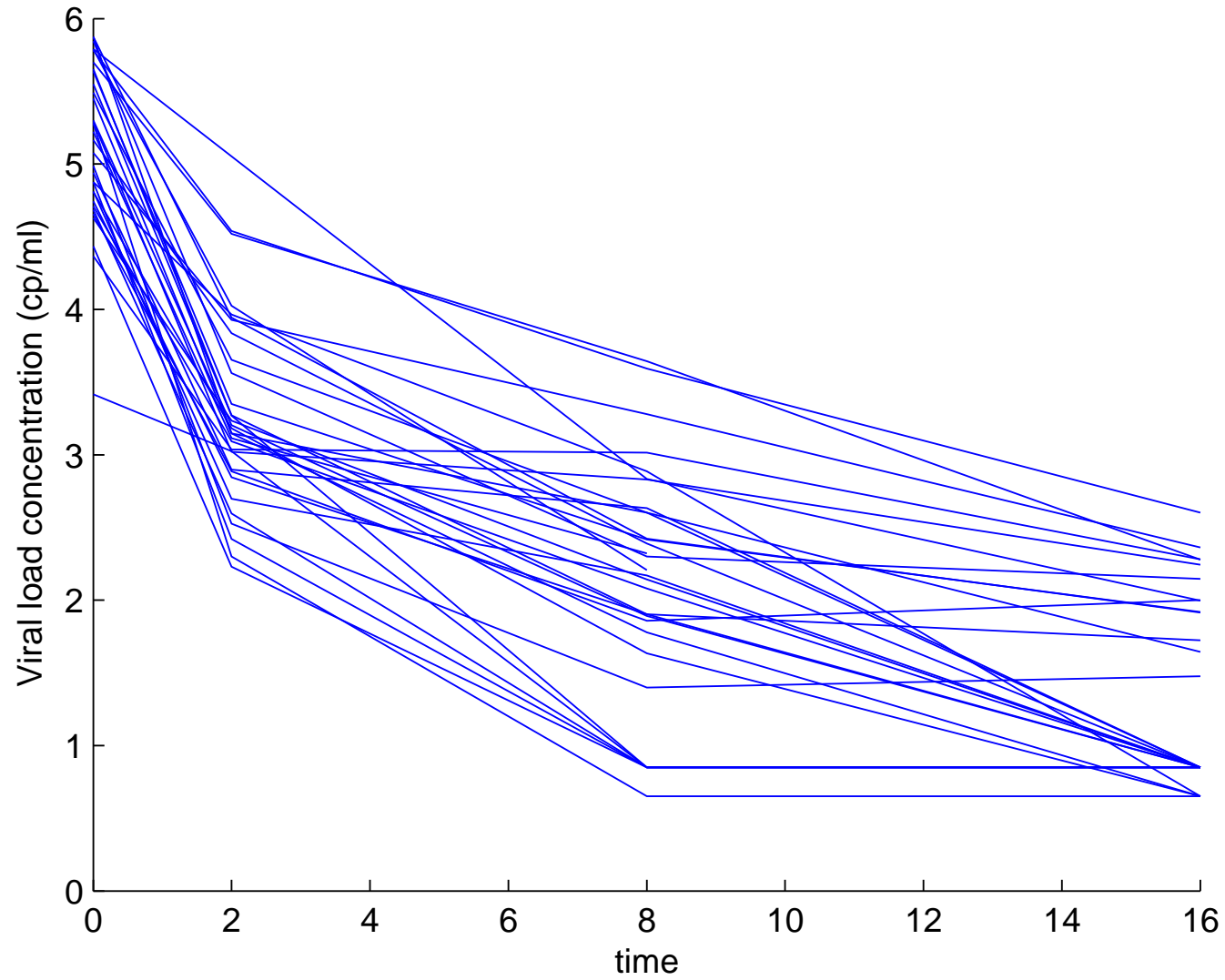
	$k_m$	$V_m$	$k_a$	$V$	$\text{var}_{k_m}$	$\text{var}_{V_m}$	$\text{var}_{k_a}$	$\text{var}_V$	$\sigma^2$
pt init	0.50	0.100	5.00	5.0	0.400	0.400	0.400	0.400	0.1000
valeur	0.37	0.082	2.72	12.2	0.040	0.040	0.040	0.040	0.0100
SAEM	0.47	0.088	2.58	12.3	0.043	0.038	0.039	0.043	0.0084
NONMEM	0.60	0.100	2.57	12.3	$10^{-8}$	0.062	0.068	0.036	0.0088

## 6.1 Exemple de données réelles

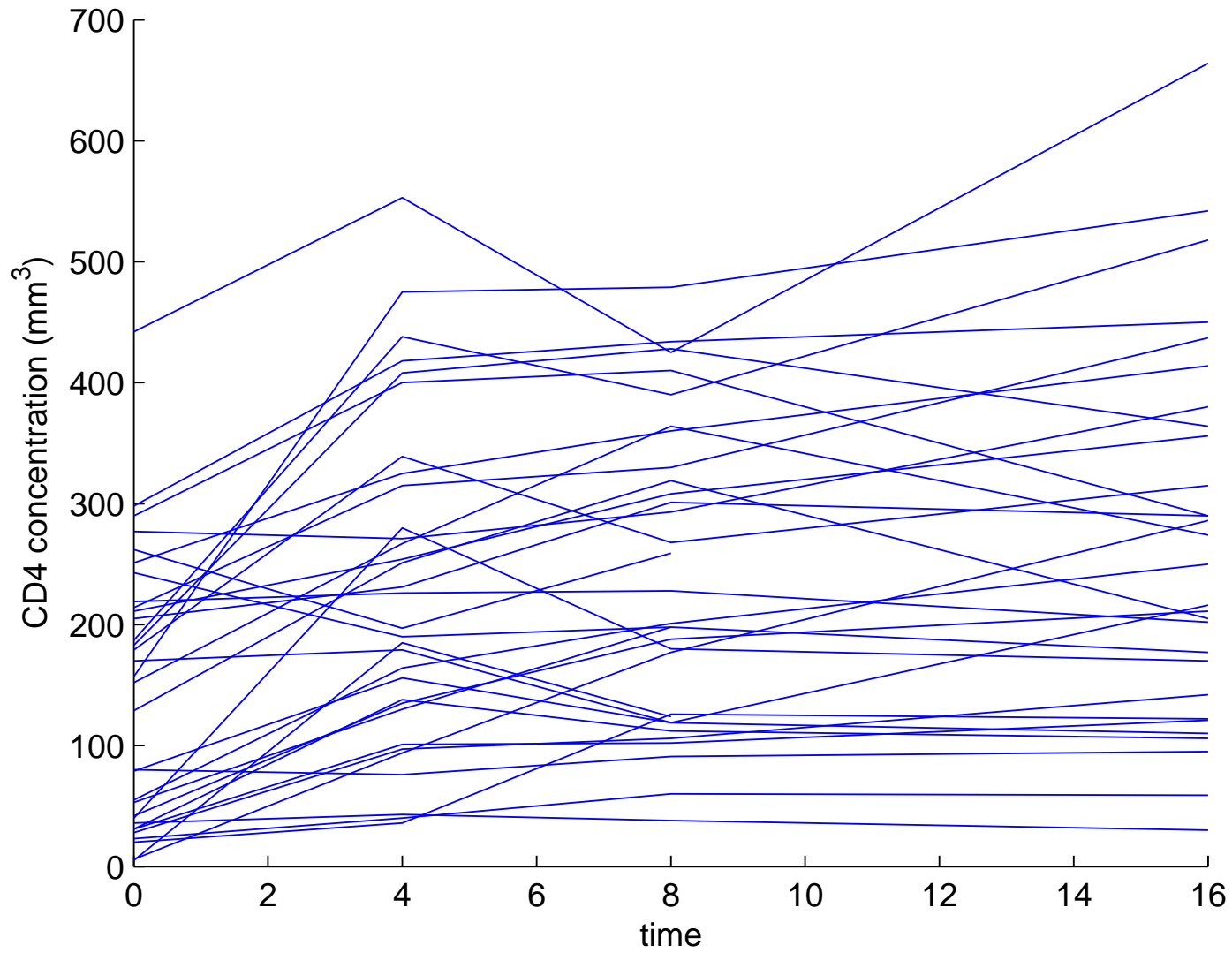
---

- Traitement anti-rétroviral chez des patients atteints du VIH contenant un inhibiteur de protéase (IP)
- Essai Cophar2-ANRS111
- But : adaptation de posologie d'IP
  - Suivi de S-4 à S48
- Données jusqu'à S16 :
  - Modèle utilisé décrit la dynamique initiale du virus et ne tient pas compte de ses mutations
  - Un bras de traitement d'IP analysé (lopinavir)
  - 32 sujets
  - 4 points par sujets
    - Charge Virale (CV) : S0, S2, S8, S16
    - CD4 : S0, S4, S8, S16

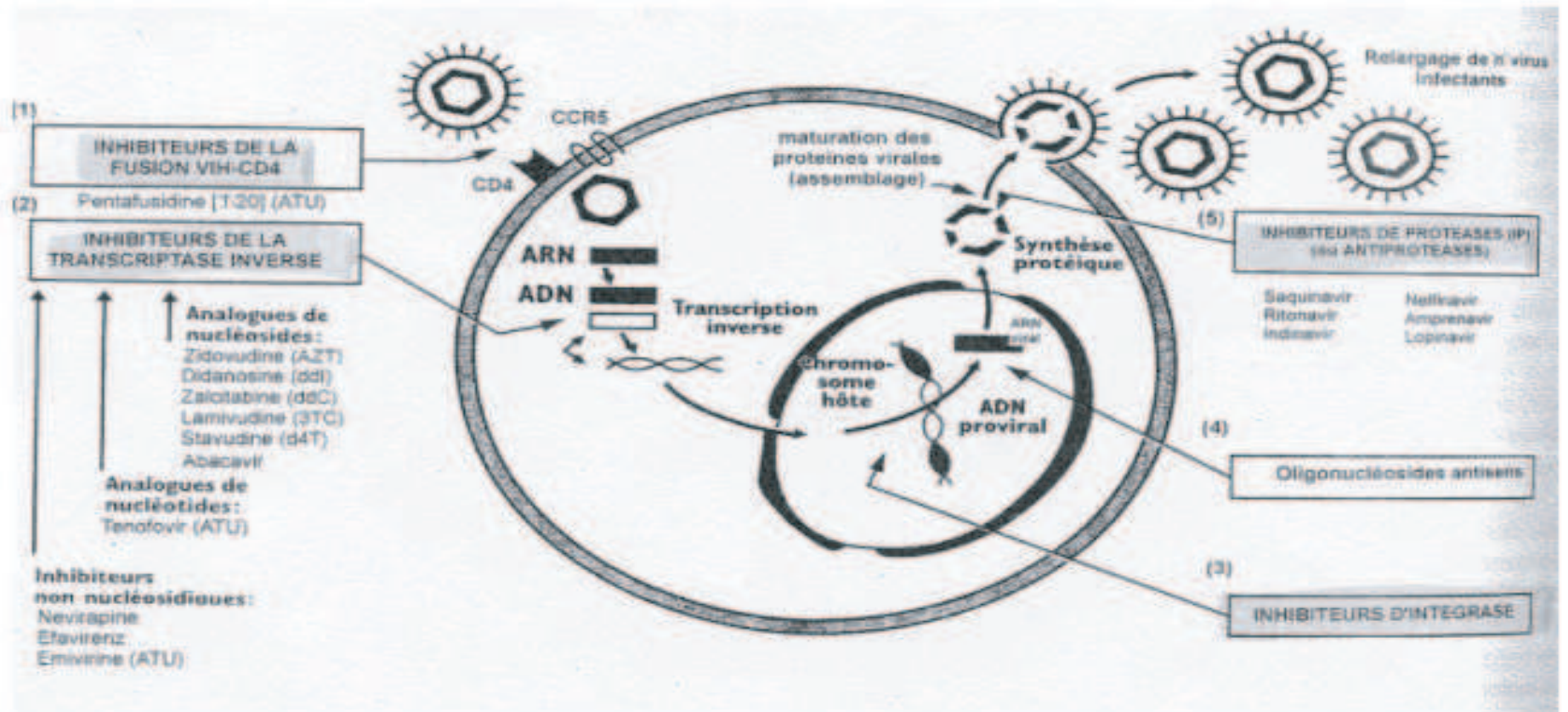
## 6.2 Les données de charge virale



## 6.3 Les données de CD4



## 6.4 Activité du virus VIH et action des médicaments



## 6.5 Modèle avant traitement (Perelson et al, Science, 1996)

---

$$\begin{aligned}\frac{dT_{NI}}{dt} &= \Pi - (\delta_n + \beta V_I)T_{NI} \\ \frac{dT_I}{dt} &= \beta V_I T_{NI} - \delta T_I \\ \frac{dV_I}{dt} &= p_0 T_I - c V_I\end{aligned}$$

- $T_{NI}$  et  $T_I$  : concentration de CD-4 non infectés et infectés
- $V_I$  : concentration de virus infectieux
- $\Pi$  : production thymique de lymphocytes,  $p_0$  : production virale
- $\delta_n$  et  $\delta$  : taux de mortalité cellulaire de  $T_{NI}$  et  $T_I$ ,  $c$  : taux de mortalité des virions
- $\beta$  : taux d'infection cellulaire
- $\eta_0$  : proportion de virus non infectés produits (sans traitement)

## 6.6 Modèle sous traitement

$$\begin{aligned}\frac{dT_{NI}}{dt} &= \Pi - (\delta_n + (1 - \eta_{RTI})\beta V)T_{NI} \\ \frac{dT_I}{dt} &= (1 - \eta_{RTI})\beta V T_{NI} - \delta T_I \\ \frac{dV_I}{dt} &= (1 - \eta_{PI})pT_I - cV_I \\ \frac{dV_{NI}}{dt} &= \eta_{PI}pT_I - cV_{NI}.\end{aligned}\tag{2}$$

- $V_{NI}$  : concentration de virus non infectieux
- $\eta_{RTI}$  : proportion de lymphocytes non infectés produits sous inhibiteur de transcriptase inverse
- $\eta_{IP}$  : proportion de virus non infectants produits sous inhibiteur de protéase
- Modèle non-identifiable  $\Rightarrow$  on estime  $\tilde{\beta} = (1 - \eta_{RTI})\beta$

## 6.7 Conditions initiales

---

- Hypothèse : état d'équilibre avant mise sous traitement

$$T_{NI}(t = 0) = \frac{\delta c}{\beta p(1 - \eta_0)}$$

$$T_I(t = 0) = \frac{c}{p} V(t = 0)$$

$$V_I(t = 0) = V(t = 0)$$

$$V_{NI}(t = 0) = 0$$

- On estime  $V_0 = V(t = 0)$  pour chaque sujet

## 6.8 Modèle final

$$\log CV_{ij} = a_1 \circ f(t_{ij}, \theta_i) + \epsilon_{(1)ij}$$

$$CD4_{ij} = a_2 \circ f(t_{ij}, \theta_i) + \epsilon_{(2)ij}$$

$$\theta_i = \mu_i + b_i$$

- $CV_{ij}$  : mesure de la charge virale chez le patient  $i$  au temps  $t_{ij}$
- $CD4_{ij}$  : mesure du taux de CD4 chez le patient  $i$  au temps  $t_{ij}$
- $\epsilon_{(1)ij} \sim \mathcal{N}(0, \sigma_1^2)$  iid,  $\epsilon_{(2)ij} \sim \mathcal{N}(0, \sigma_2^2)$  iid
- $b_i \sim \mathcal{N}(0, \Omega)$
- $f = (f_1, f_2, f_3, f_4)$  est solution du système précédent
- $a_1 = \log_{10}(f_3 + f_4)$  et  $a_2 = f_1 + f_2$
- $\theta_i = (\beta, \delta, p, \eta, c, V_0, \Pi, \delta_n)$

## 6.9 Estimation

---

- Adaptation de la fonction saemmlx.m de Marc Lavielle
  - Aux ODE
  - A un vecteur d'observation
- Modèle d'équations différentielles stiff
  - Aucune fonction 'ode' de Matlab n'arrive à résoudre le système
  - Implémentation de LL et LL2
- Deux analyses
  1. Données complètes jusque'à S16
  2. Données sous la LOQ censurées
- Résultats présentés avec 200 itérations saem

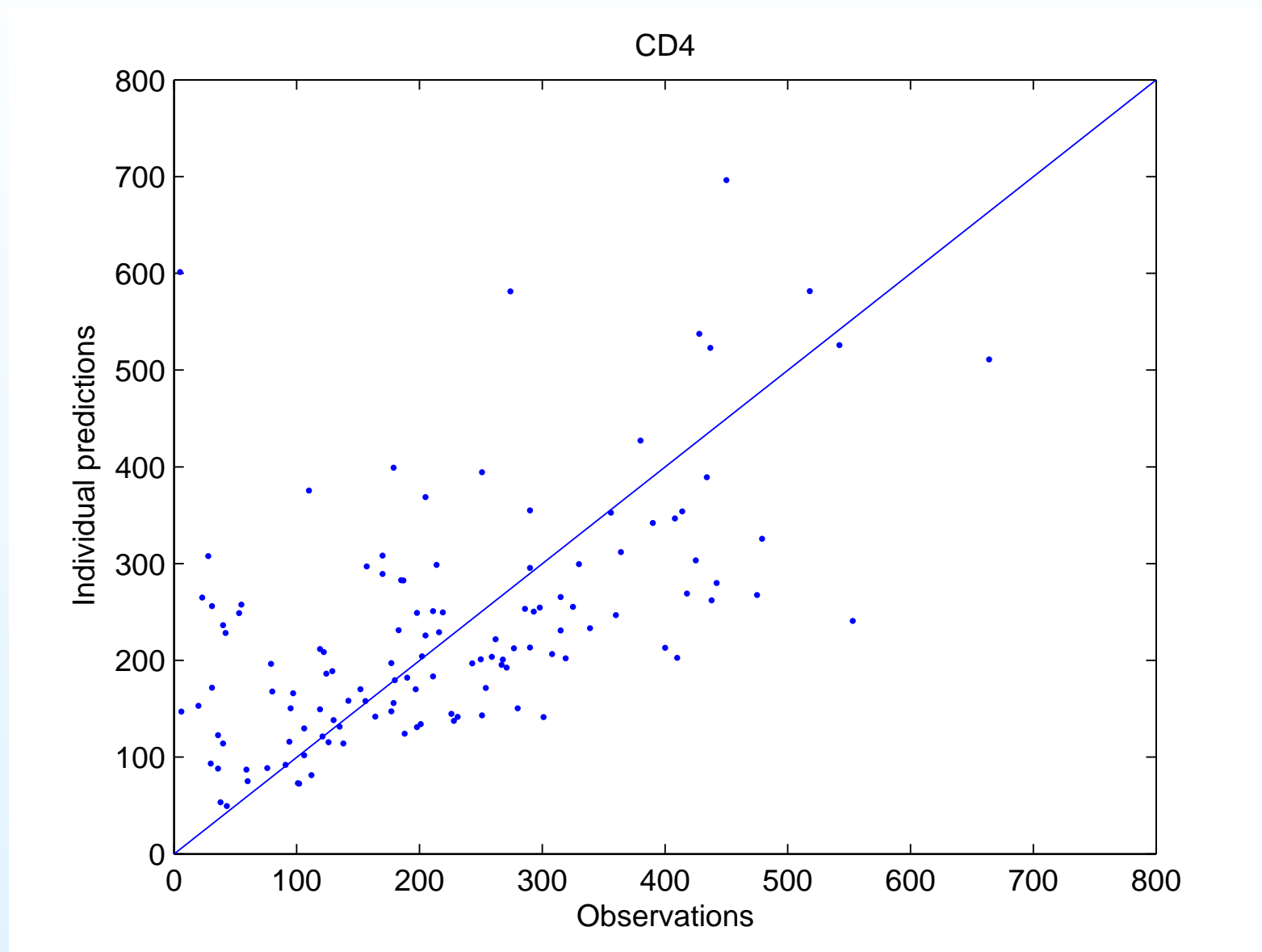
## 6.10 Données complètes (1)

	$\beta$	$\delta$	$p$	$\eta$	$c$	$V_0$	$\Pi$	$\delta_n$
$\mu$	$9.10^{-7}$	0.273	1.19	0.595	155	5.73	19930	0.01
Var	0.005	0.02	1.17	0.0006	0.009	0.0005	0.32	0.003

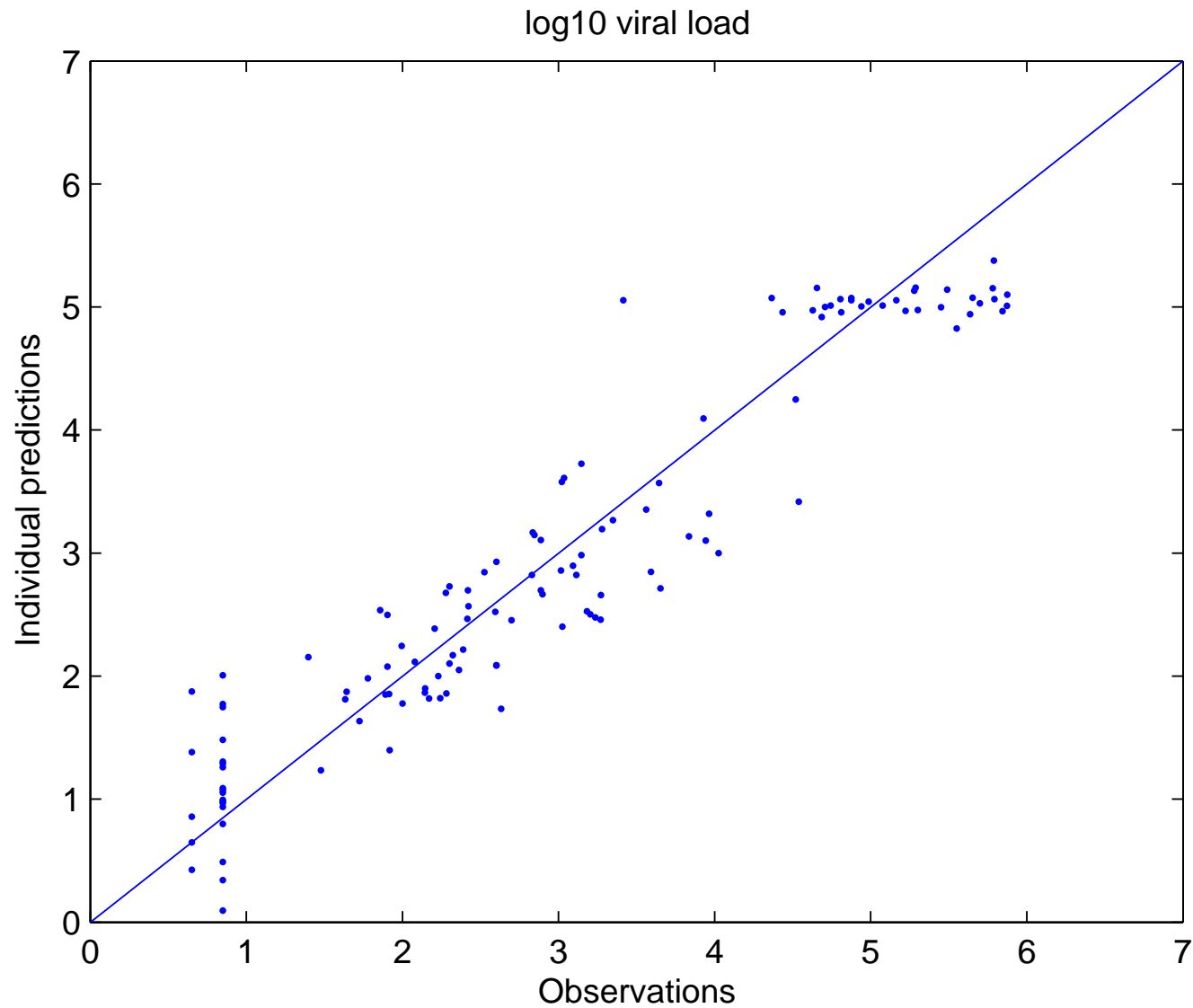
log CV  $\sigma_1^2 = 0.258$

CD4  $\sigma_2^2 = 0.229$

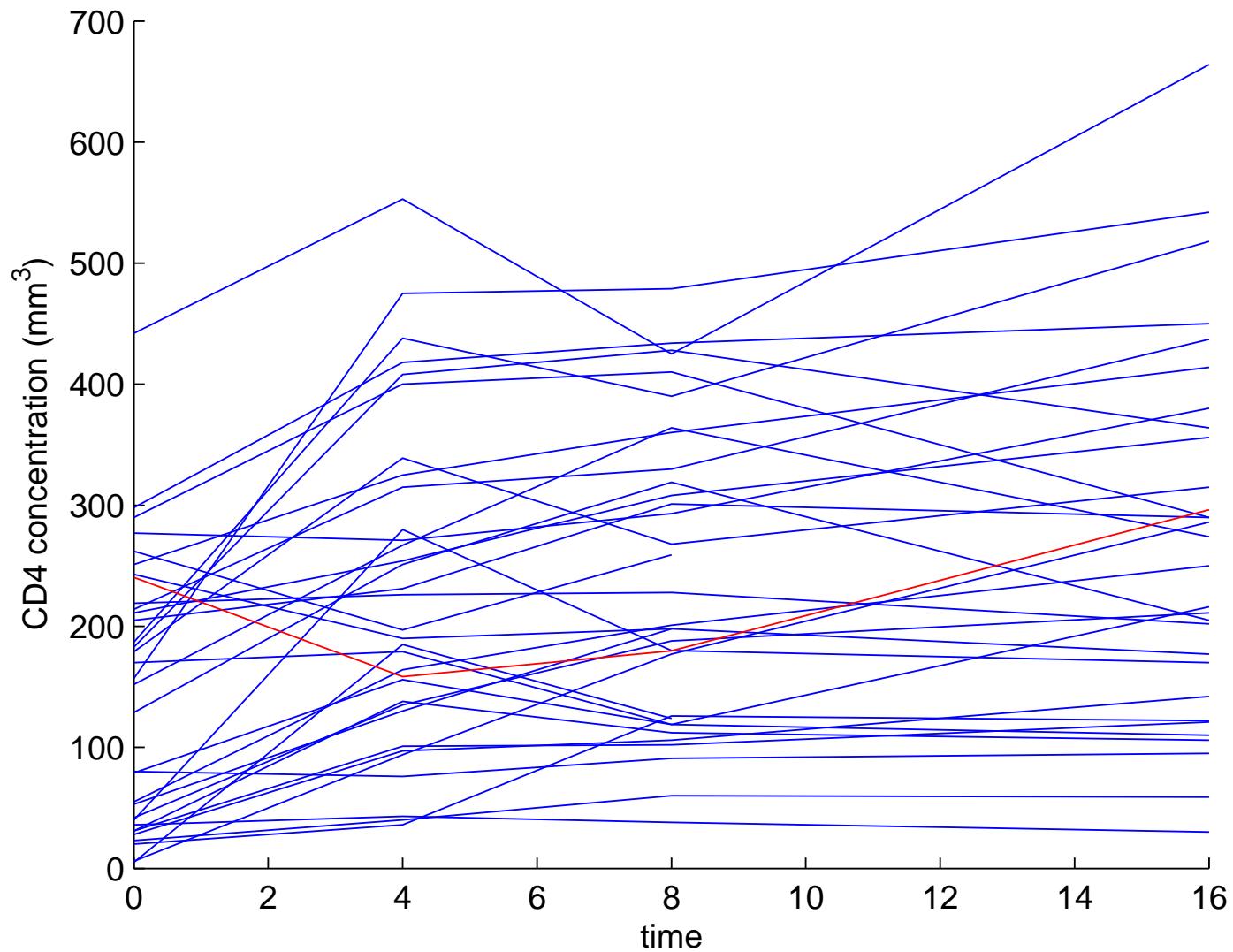
## 6.11 Données complètes (2)



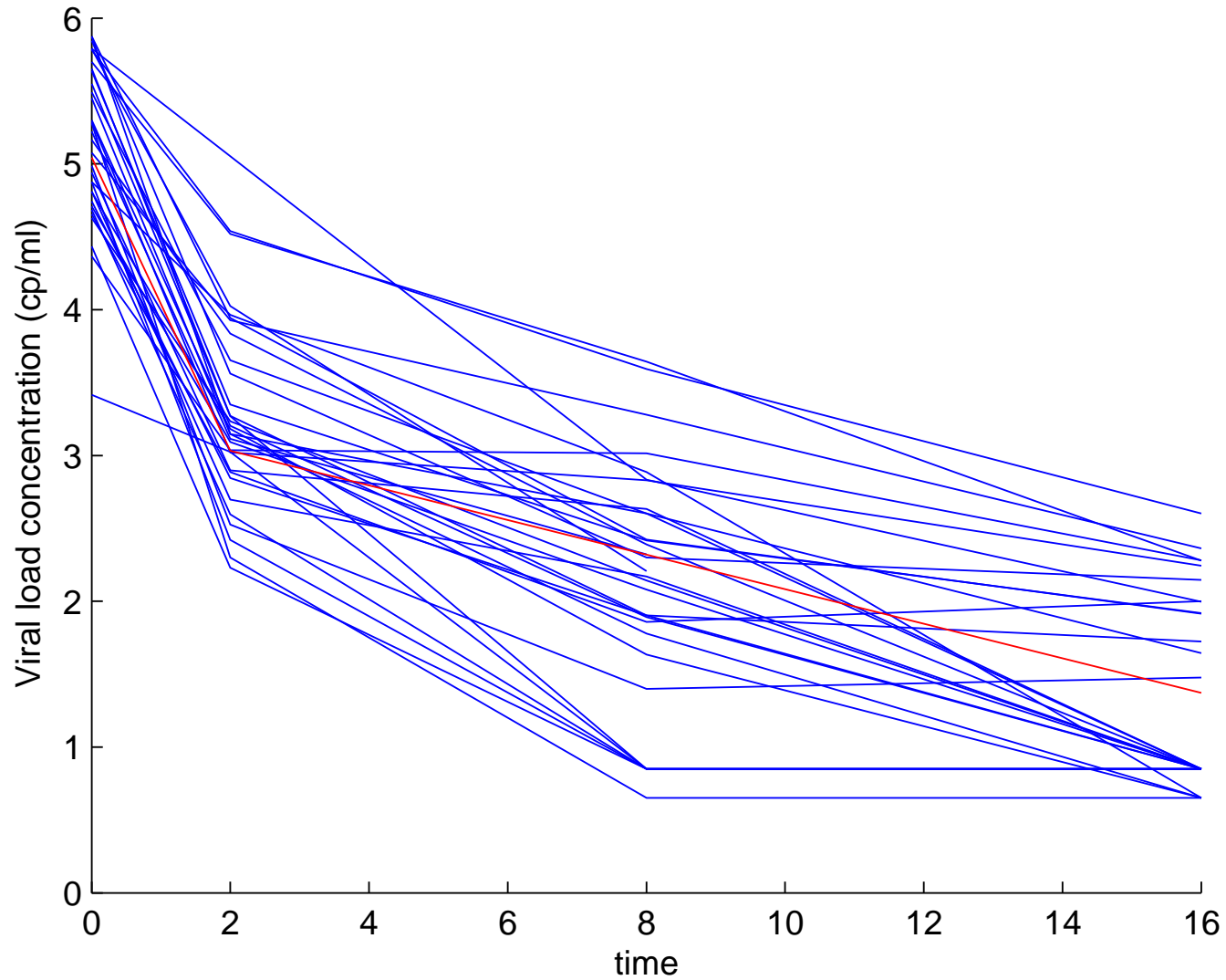
## 6.12 Données complètes (3)



## 6.13 Données complètes (4)



## 6.14 Données complètes (5)



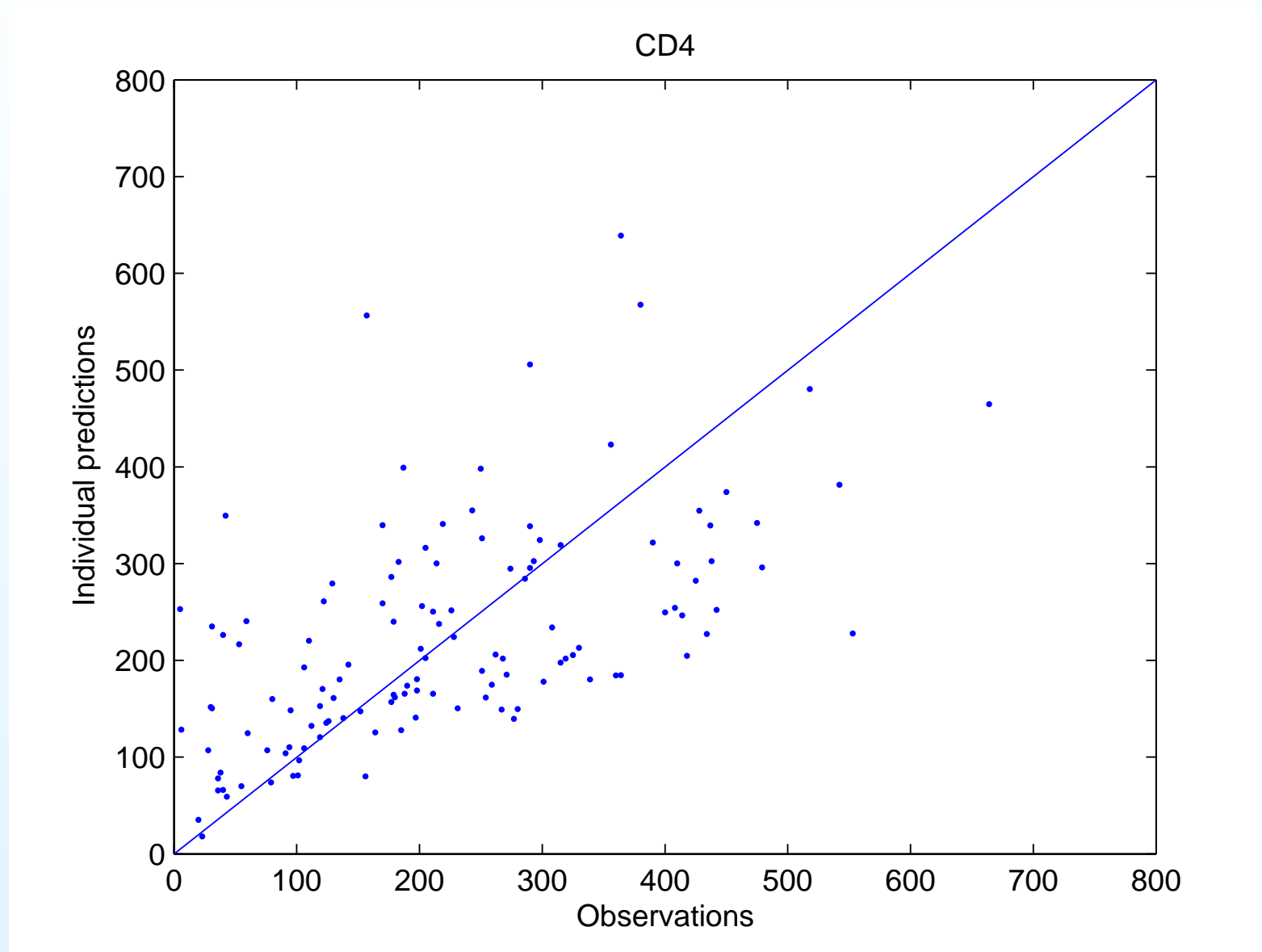
## 6.15 Données censurées (1)

	$\beta$	$\delta$	$p$	$\eta$	$c$	$V_0$	$\Pi$	$\delta_n$
$\mu$	$1.10^{-6}$	0.264	1.25	0.515	152	5.99	18215	0.009
Var	0.002	0.06	0.67	0.01	0.06	0.003	0.132	0.004

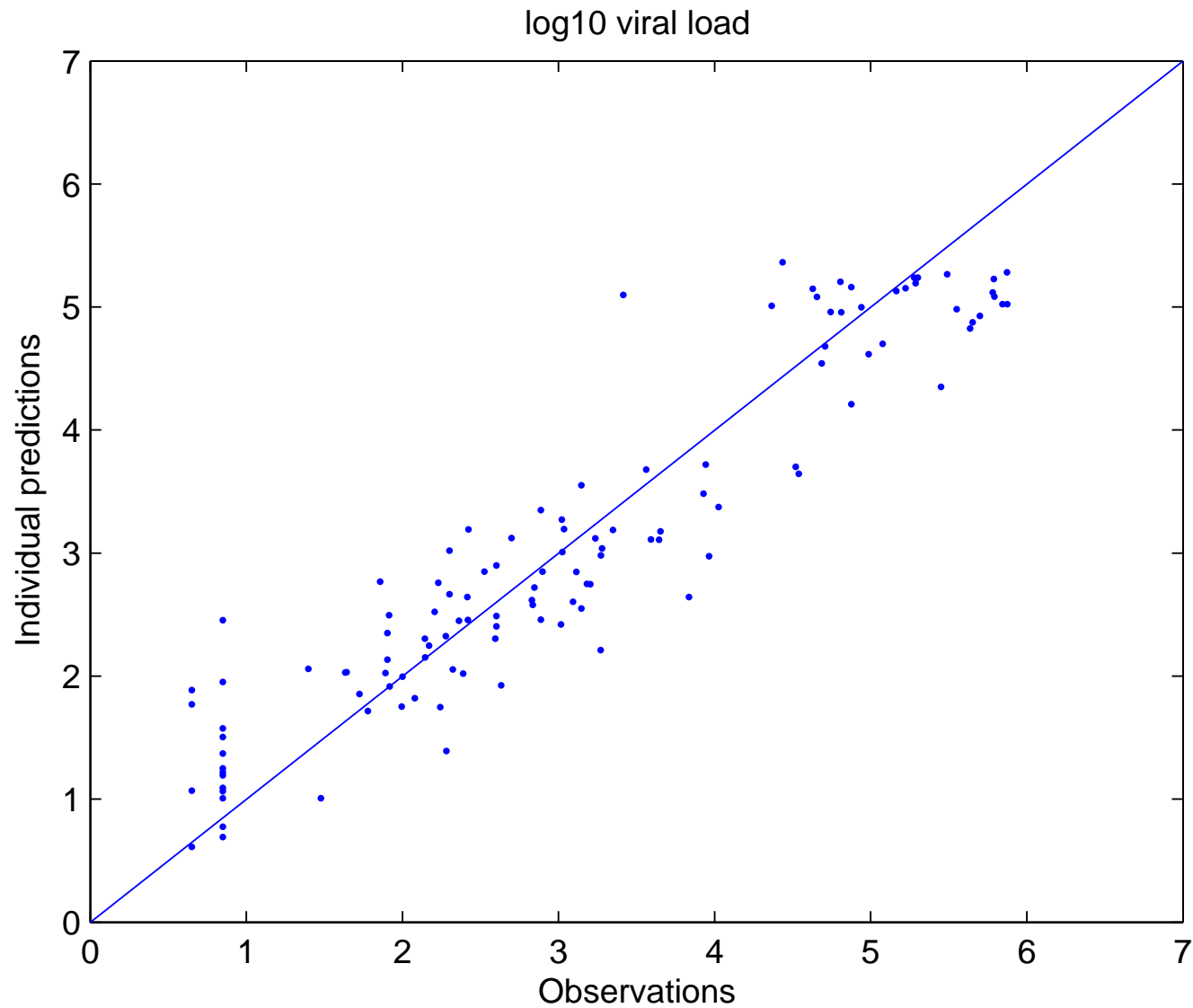
log CV  $\sigma_1^2 = 0.294$

CD4  $\sigma_2^2 = 0.274$

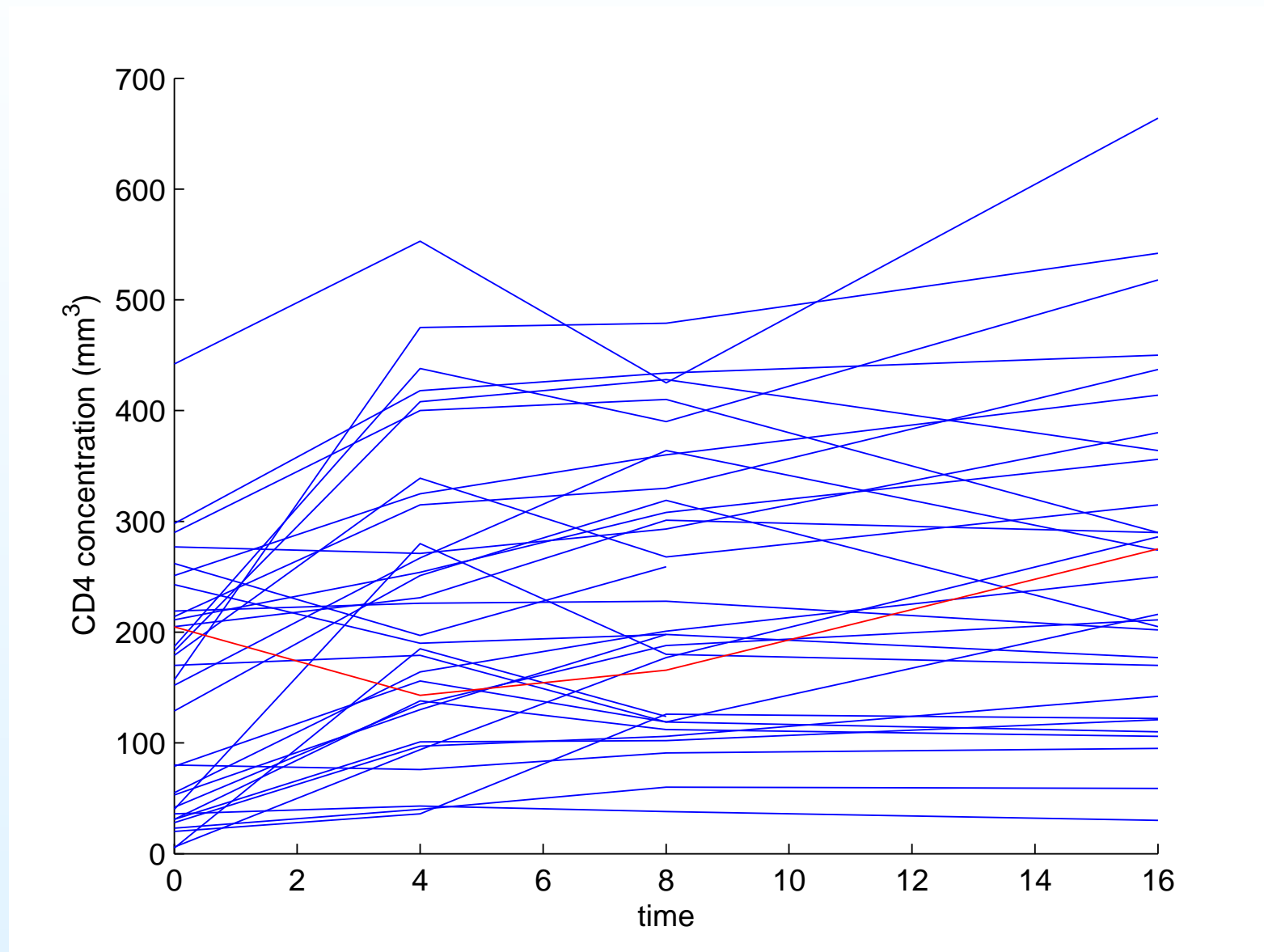
## 6.16 Données censurées (2)



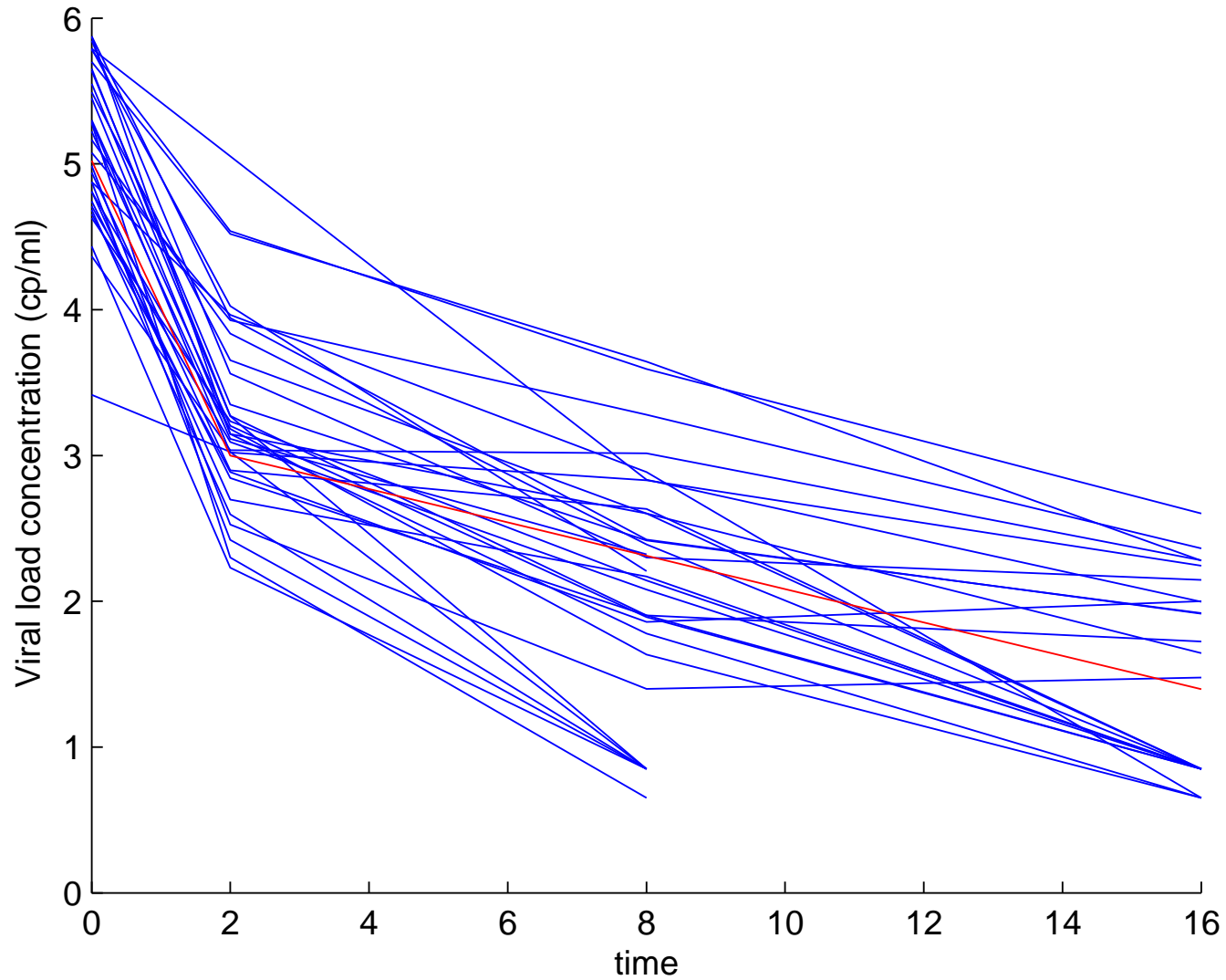
## 6.17 Données censurées (3)



## 6.18 Données censurées (4)



## 6.19 Données censurées (5)



## Conclusion

---

- Résultats théoriques
  - Confirmation par les données simulées
  - Étude du jeu de données réelles à approfondir
- ⇒ Outil puissant pour analyser un modèle mixte défini par équations différentielles
- Inconvénient de l'utilisation du schéma LL : nécessité de calculer le jacobien de la fonction  $F$