Stochastic models

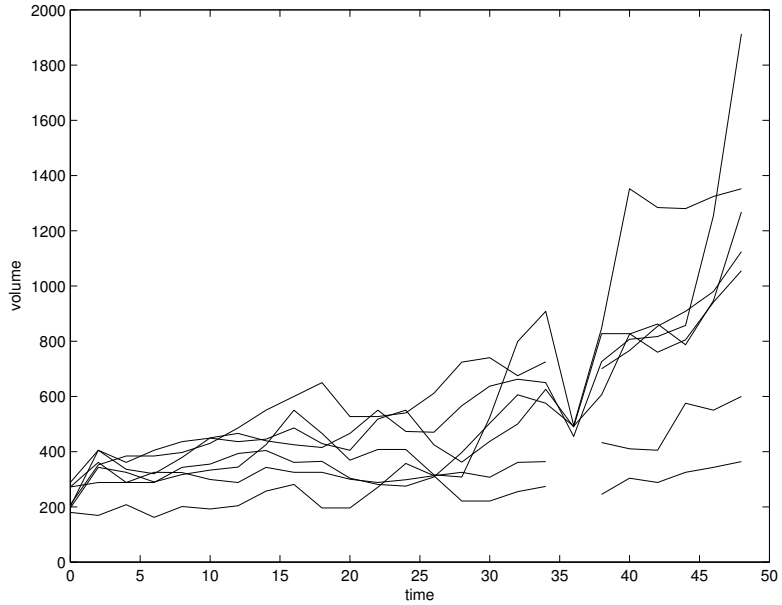
Growth models and some estimation methods

Adeline Leclercq Samson

Laboratoire Jean Kuntzmann, Université Joseph Fourier, Grenoble

Growth data

- Data measured in (pre)-clinical trials
  - Height, weight of subjects
  - Tumor volume, tumor size
  - Circulating biomarkers



- Longitudinal data
  - Several subjects $i = 1, \ldots, n$
  - Reapeated measures at times $t_{ij}, \quad j = 0, \ldots, J$
  - Observations y_{ij} at time t_{ij}
  - Measurement noise

- Mixed-effect models

- The FDA has recommended the use of mixed-effect models to analyze longitudinal data of tumor response to treatment

Mixed-effect model

- Repeated observations:

$$y_{ij} \text{ at time } t_{ij}, \quad i = 1, \ldots, n, \quad j = 0, \ldots, J$$

- Standard regression model

$$\begin{array}{rcl} y_{ij} & = & f(\phi_i, t_{ij}) + g(\phi_i, t_{ij})\varepsilon_{ij} \\ \varepsilon_{ij} & \sim_{iid} & \mathcal{N}(0, \sigma^2) \\ \phi_i & \sim_{iid} & \mathcal{N}(\mu, \Omega) \end{array}$$

  - f: parametric regression function
  - ϕ_i : "biological" or "physiological" random parameters
  - g: error model [homoscedastic $g = 1$ or heteroscedastic $g = f$]

- Parameters to be estimated

$$\theta = (\mu, \Omega, \sigma)$$

Likelihood

- Notations
  - $y_i = y_{i0:J} = (y_{i0}, \ldots, y_{iJ})$: data vector of subject i
  - $y = (y_1, \ldots, y_n)$: global data vector

- Likelihood function

$$\begin{aligned} L(y;\theta) &= \prod_{i=1}^{n} p(y_i;\theta) = \prod_{i=1}^{n} \int p(y_i, \phi_i;\theta) d\phi_i \\ &= \prod_{i=1}^{n} \int p(y_i|\phi_i;\theta) p(\phi_i;\theta) d\phi_i \end{aligned}$$

$\Rightarrow$ If f not linear with respect to ϕ_i, likelihood not explicit

Maximum likelihood estimation

- Approximation of the likelihood
  - Linearization of the likelihood [Pinheiro and Bates, 2000]
- Numerical computation of the likelihood
  - Gaussian quadrature [Davidian and Giltinan, 1995; Guedj et al, 2007; Picchini et al, 2010]
  - Monte Carlo EM algorithm (MCEM) [Wei and Tanner, 1991]
  - Stochastic Approximation EM algorithm (SAEM) [Kuhn and Lavielle, 2005]
- Bayesian approach
  - Prior choice on θ
  - MCMC algorithms [Spiegelhater et al, 1992]
  - Posterior distribution $p(\theta|y)$

Gaussian quadrature methods

[Davidian and Giltinan, 1995; Guedj et al, 2007; Picchini et al, 2010]

- Gauss-Hermite quadrature of order R
  - Individual likelihood

$$L(y_i;\theta) = \int p(y_i|\phi_i;\theta)p(\phi_i;\theta)d\phi_i$$

  - Approximation

$$L^G(y_i;\theta) = \sum_{r=1}^{R} \pi_r p(y_i|\omega\sqrt{2}rz_r/\mu;\theta)$$

    - $z_r, r = 1, \ldots, R$ zeros of the Hermite polynomials $H_R(\cdot)$ of degree R
    - π_r adequate weights

  - Convergence of L^R to the true likelihood when R tends to infinity

- Software: SAS

Stochastic Approximation EM (SAEM) Algorithm

[Dempster, Laird, Rubin, 1977; Delyon, Lavielle and Moulines, 1999; Kuhn, Lavielle, 2005]

- SAEM algorithm
  - *E step*
    - *S step* : simulation of (ϕ_m) under distribution $p(\phi|y;\widehat{\theta}_m)$ with MCMC algorithm
    - *SA step* : approximation of

$$Q_{m+1}(\theta) = \mathbb{E}\left[\log p(y,\phi;\theta) \,|\, y, \widehat{\theta}_m\right]$$

with a stochastic approximation scheme of step size α_m

$$Q_{m+1}(\theta) = (1-\alpha_m)Q_m(\theta) + \alpha_m \log p(y,\phi_m;\theta)$$

  - *M step*: update of $\widehat{\theta}_m$

$$\widehat{\theta}_{m+1} = \arg\max_{\theta} Q_{m+1}(\theta)$$

  - Convergence of $\widehat{\theta}_m$ to the maximum likelihood estimator
- Software: Monolix

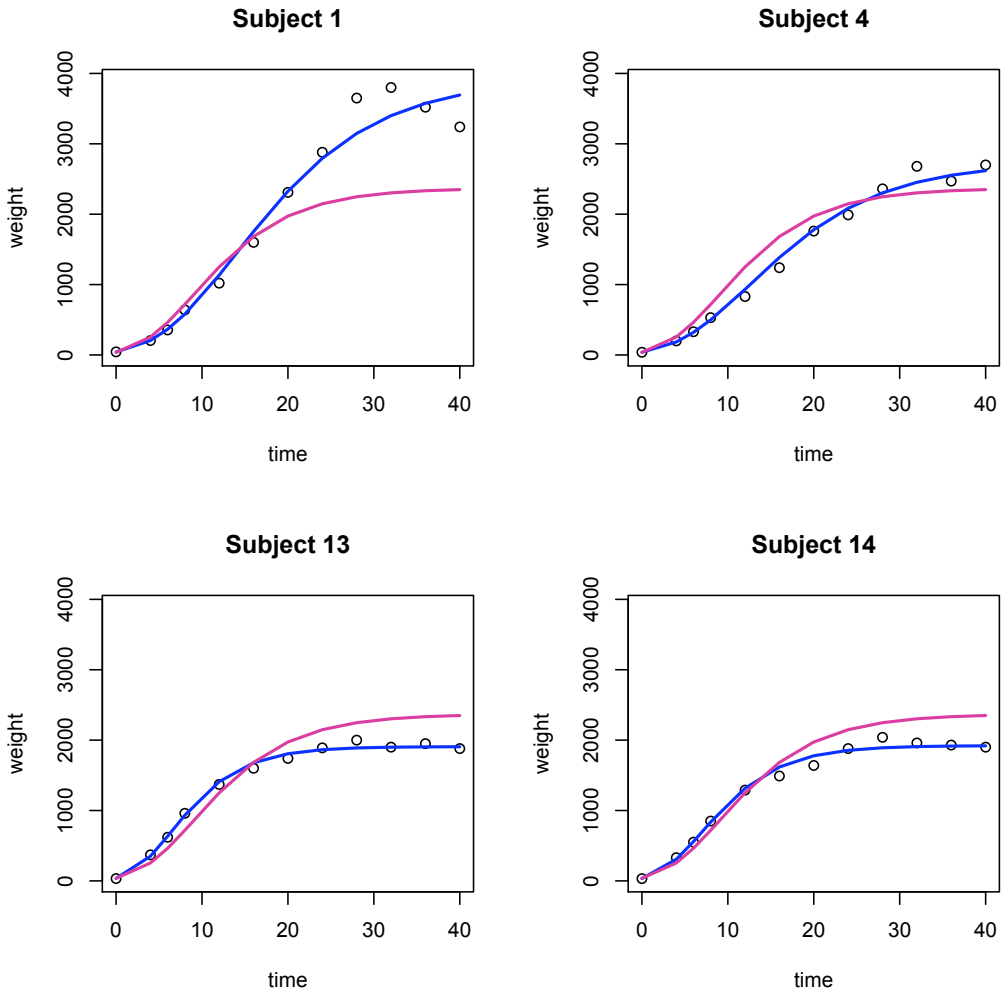
Growth models (regression function f)

- Several classes of models
  - Standard growth functions
    - Logistic, Gompertz, Richards, Weibull [Zimmerman and Nunez-Anton, 2001]
    - Monotone increase
  - Phenomenological models
    - System theory [Wiener, 1948; Bertalanffy, 1960; Bastogne et al; 2009]
    - Holistic representation, black-box models
  - Mechanistic models
    - System of ordinary differential equations [Simeoni et al, 2004; Ribba et al, 2010, 2011, 2012]
    - Partial differential equations [Ribba, Colin, Schnell 2006; Colin et al, 2013; Lagaert PhD]
    - Dynamic of angiogenesis

Growth data

[Donnet, Foulley, Samson, 2010]

- Population
  - 50 pigs
  - 11 weight measures per subject
- Gompertz function
  - $f(\phi, t) = A\, e^{-Be^{-Ct}}$
  - $\phi = (A, B, C)$



Population prediction $f(\hat{\mu}, t)$, Individual prediction $f(\hat{\phi}_i, t)$

Growth inhibition study

[Bastogne, Samson et al, 2010]

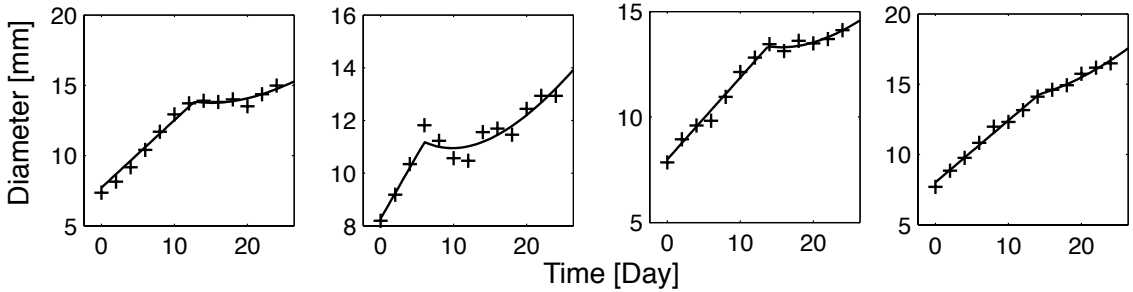
- Data
  - Population
    - 96 Female mice, with tumor implantation
    - Treatments: no treatmtent (NT) or radiotherapy (RT) or concomitant radiochemotherapy (RCT) or photodynamic therapy (PDT)
  - Measurements
    - $v(t)$ tumor volume at time t
    - $y_{ij} =^3 \sqrt{\frac{6v(t)}{\pi}}$

- Linear-Exponential-Linear Model (phenomenological)

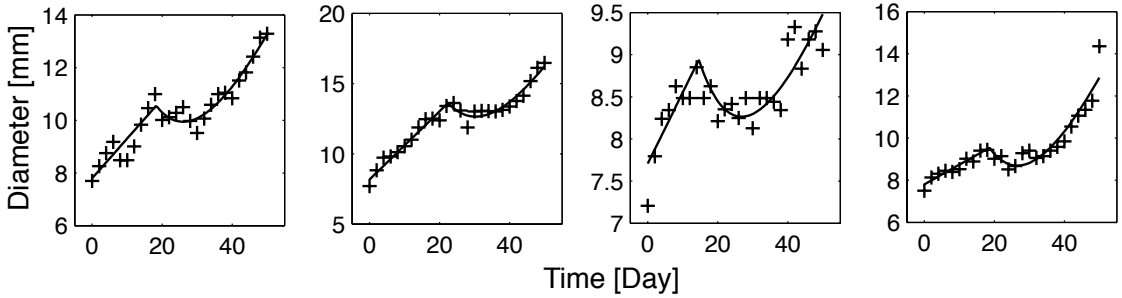
$$f(t,\phi) = s_0 \Big[\underbrace{1 + at}_{\substack{natural \\ growth}} \underbrace{-bt - k_2\, T\,(1 - e^{-(t-\tau)/T})\,\mathbf{1}_{t\geq\tau} - k_3\,(t-\tau)\,\mathbf{1}_{t\geq\tau}}_{treatment\ \ response} \Big] \tag{1}$$

Evaluation of the treatments

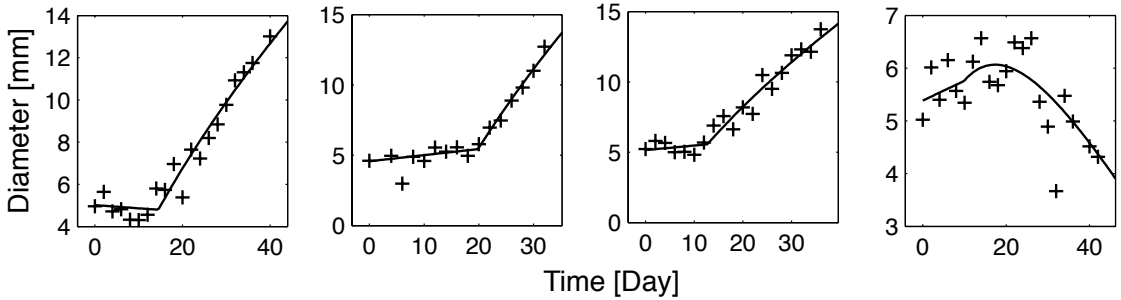
- Treatment effects
  - PDT on transient decrease
  - Duration
  - Dose of radiation



(a) Radiotherapy responses



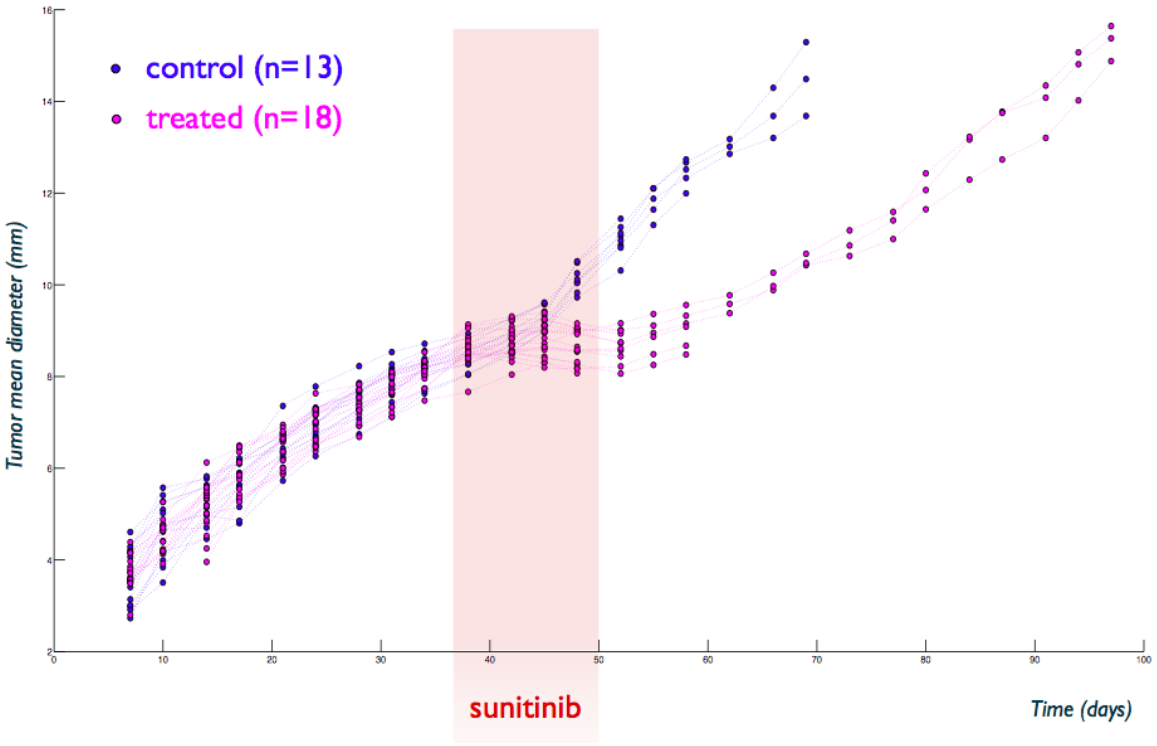
(b) Concomitant radiochemotherapy responses



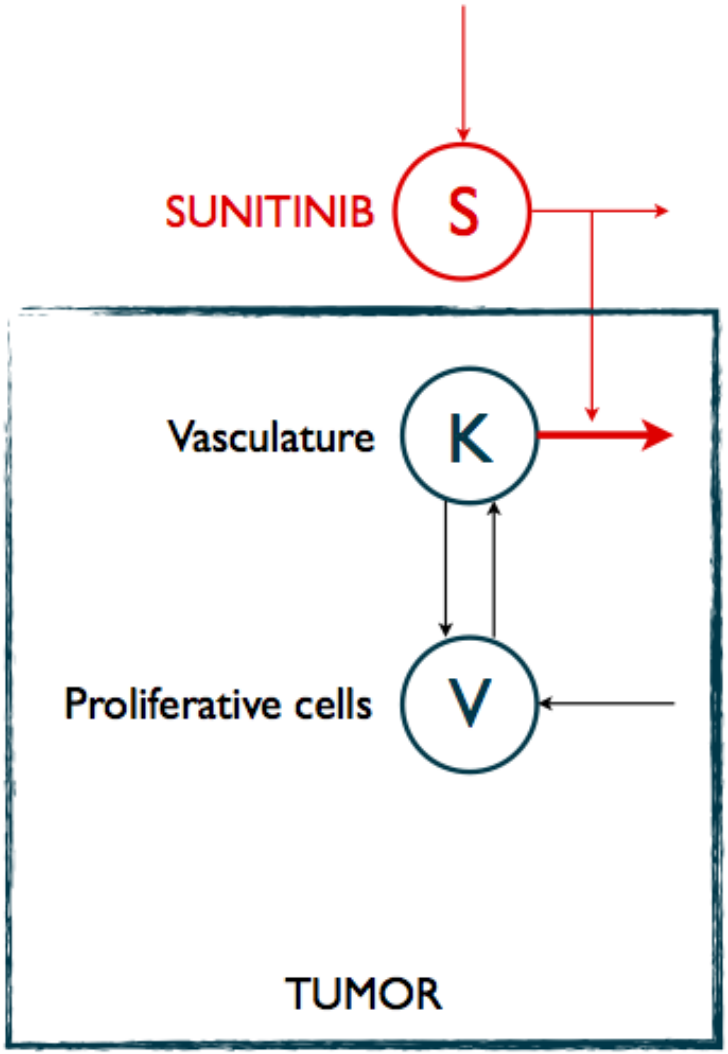
(c) Photodynamic therapy responses

Vasculature normalization [B. Ribba's slides]

- Hypothesis
  - Angiogenesis inhibitors could contribute to normalize vasculature
- Data
  - Sunitinib oral molecule
  - 30 subjects
  - 27 points per subject up to 100 days



Model



$$\frac{dS}{dt} = -k_S S$$

$$\frac{dV}{dt} = \lambda V \left(1 - \frac{V}{K}\right)$$

$$\frac{dK}{dt} = bV^{\gamma} - \beta k_s S K$$

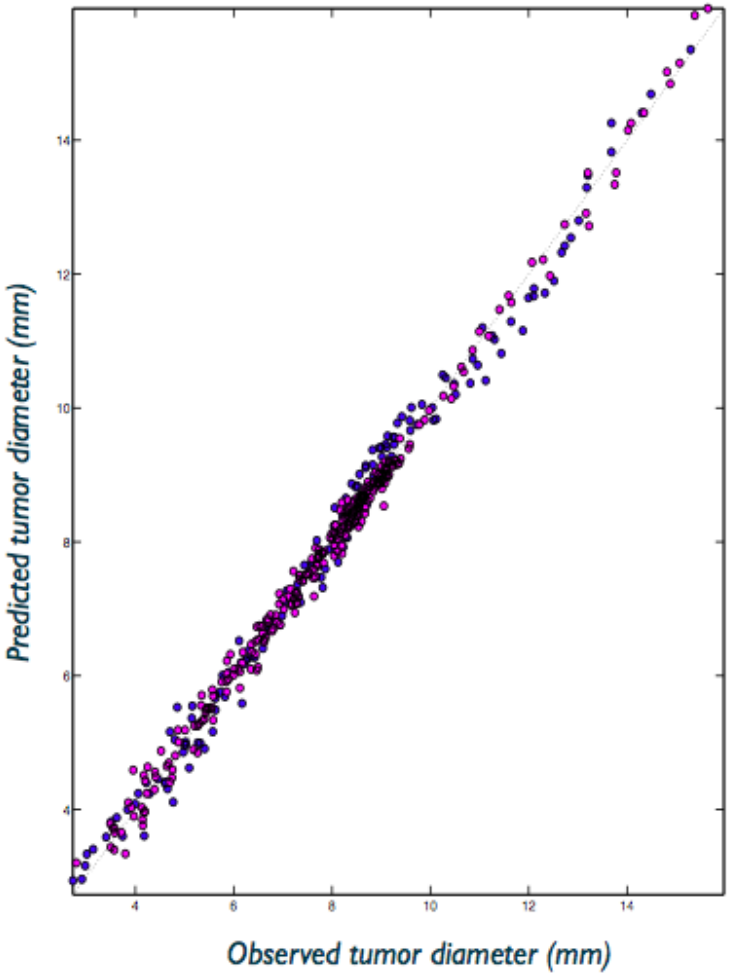
- 7 parameters: $k_S, \lambda, b, \gamma, \beta, V_0, K_0$

Results

Treatment effects

	Parameter (s.e)
k_S	2.12 (-)
λ	1.02 (4%)
b	0.00168 (4%)
γ	2 (-)
β	0.0237 (9%)
V_0	1.76 (7%)
K_0	7.43 (1%)

Predictions



Limits of complex growth models

- Complex deterministic models
  - Ordinary differential equations
    - Large number of equations: difficulty to solve the system with discrete numerical scheme. Example: [Lignet et al, 2013] 37 equations, 78 parameters
    - Large number of parameters, some of them have to be fixed
  - Partial Differential Equations
    - few parameters
    - computationally intensive to obtain one realization of the solution

- Alternative: Stochastic models
  - Lot of noise for some biomarkers (different from measurement noise)
  - Reduction of the dimension of the system [Mortensen et al, 2007; Donnet and Samson, 2013]
  - Introduction of a stochastic part to "absorb" all details that are not modeled

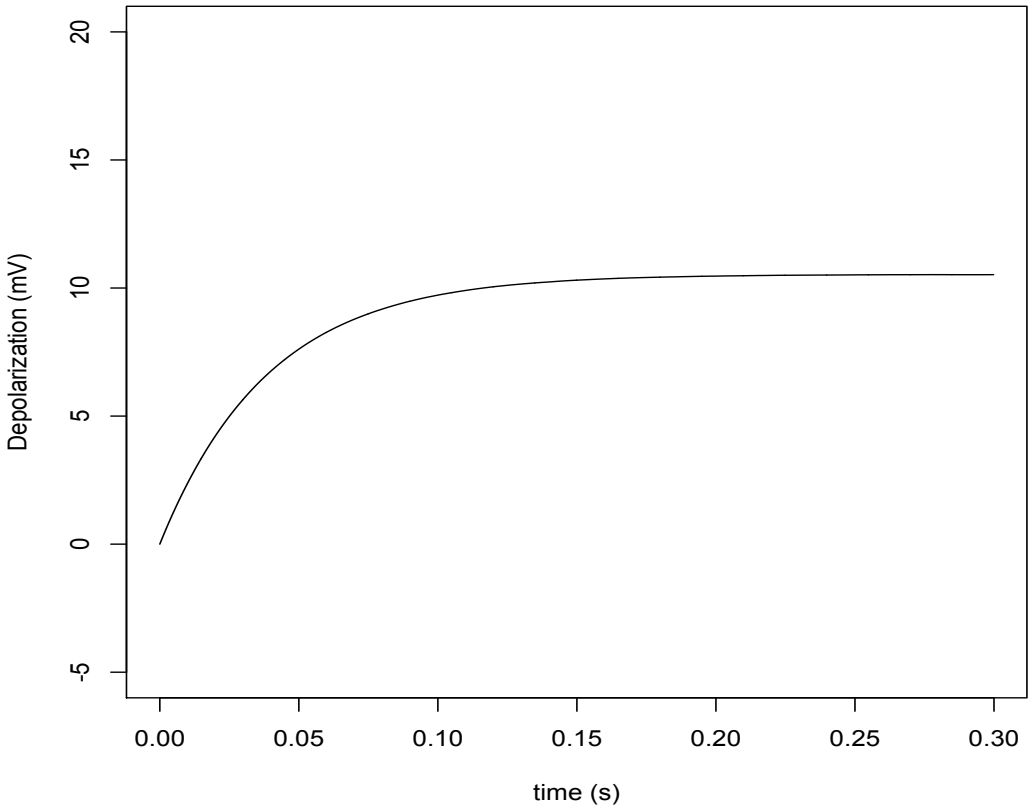
Stochastic Differential Equation

- SDE in biology
  - Pharmacocinetics [Ditlevsen et al, 2005; Ditlevsen, Samson, 2013; Donnet, Samson, 2013]
  - Neurobiology [Hopfner and Broda, 2005; Picchini et al, 2008; Ditlevsen, Samson, 2013]
  - Growth [Donnet et al, 2010]

- New source of variability
  - Variability around the deterministic model: what is not modeled by the deterministic part
  - Within-subject variability: variability in time

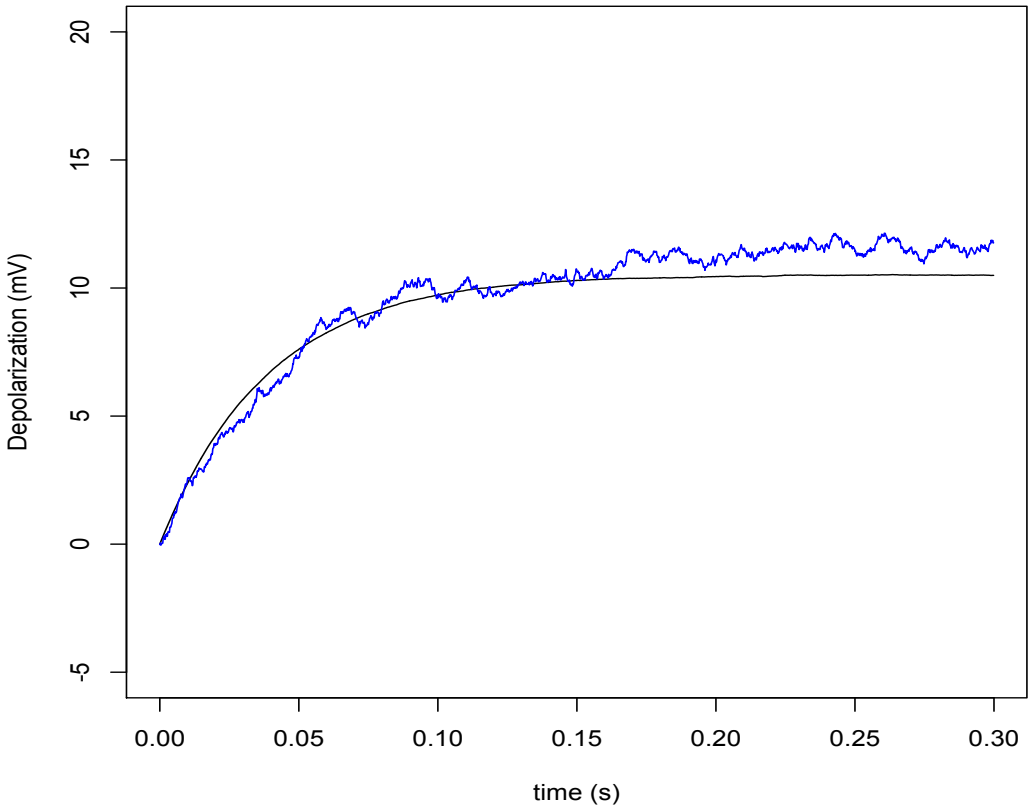
Variability around the deterministic model

- Ordinary differential equation
  - $dX_t = \left(-\frac{X_t}{\tau} + \phi\right) dt$
  - Deterministic solution



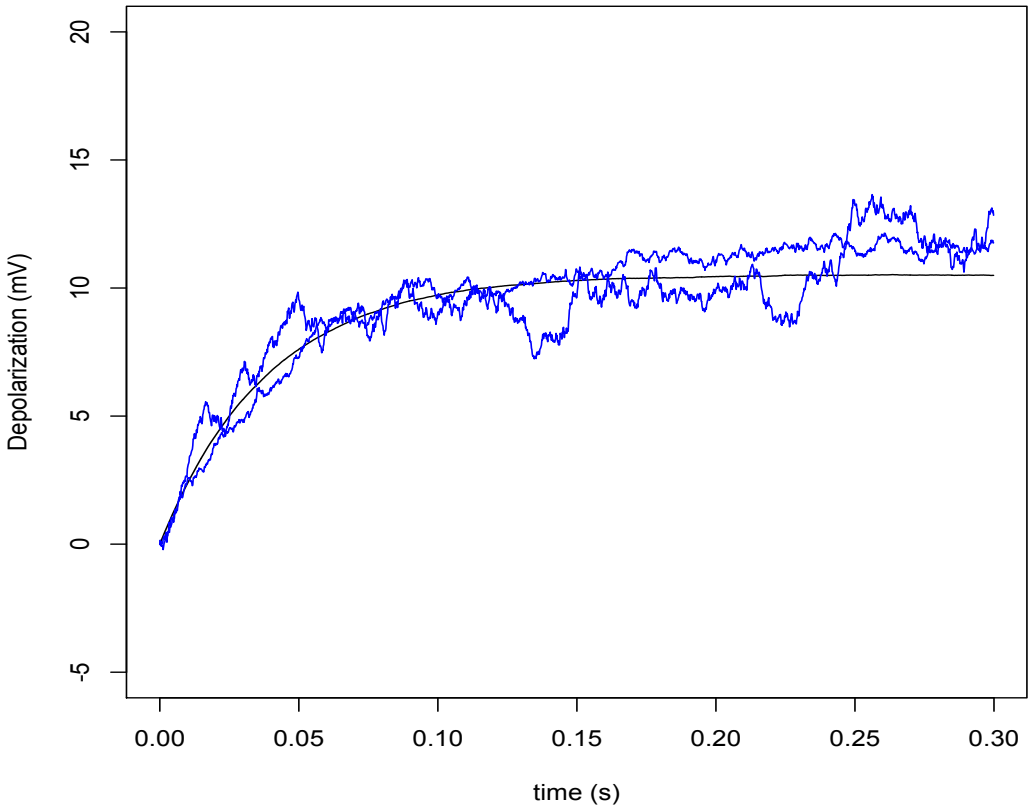
Variability around the deterministic model

- Stochastic differential equation
  - $dX_t = \left(-\frac{X_t}{\tau} + \phi\right) dt + \sigma dB_t$ with B_t a Brownian motion
  - Stochastic solution



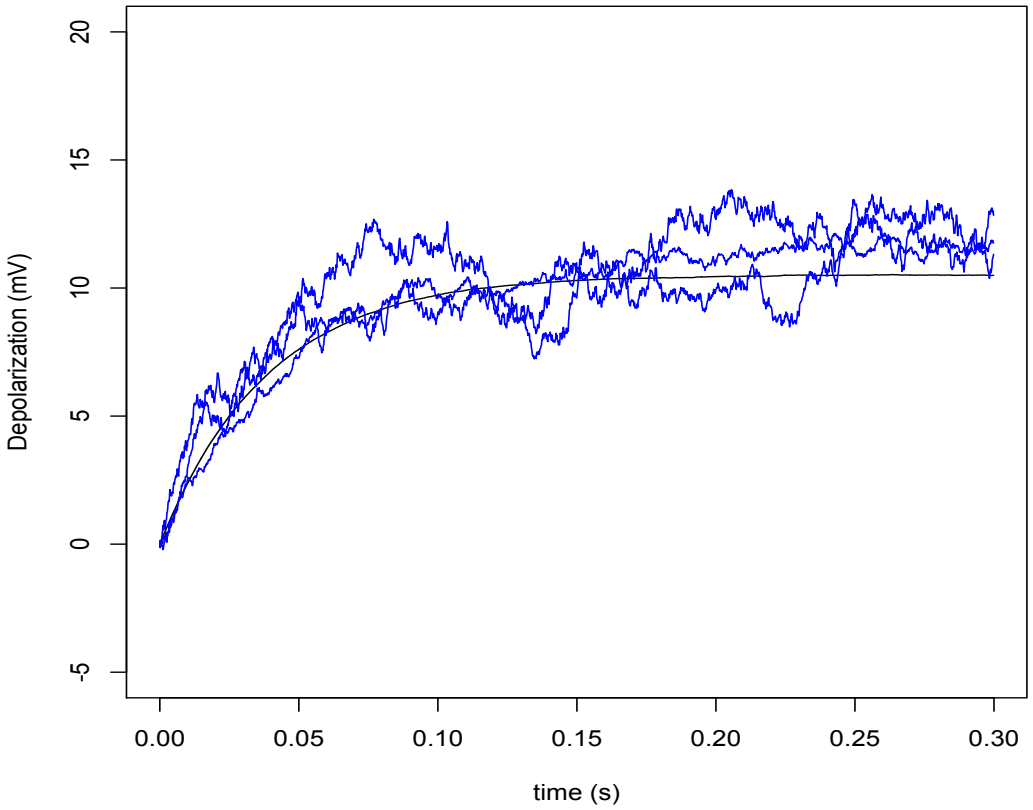
Variability around the deterministic model

- Stochastic differential equation
  - $dX_t = \left(-\frac{X_t}{\tau} + \phi\right) dt + \sigma dB_t$ with B_t a Brownian motion
  - Stochastic solution



Variability around the deterministic model

- Stochastic differential equation
  - $dX_t = \left(-\frac{X_t}{\tau} + \phi\right) dt + \sigma dB_t$ with B_t a Brownian motion
  - Stochastic solution



Stochastic Differential Equation with random parameters

$$dX_t = a(X_t, \phi)dt + b_\gamma(X_t, \phi)dB_t, \quad X_0 = x_0$$

- SDE with random parameters

$$dX_{it} = a(X_{it}, \phi_i)dt + b_\gamma(X_{it}, \phi_i)dB_{it}, \quad X_{i0} = x_0$$

  with
  - (ϕ_i) random variables
  - (B_{it}) independent brownian motions

- SDE mixed models

$$\begin{aligned} y_{ij} &= X_{it_{ij}} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim_{iid} \mathcal{N}(0, \sigma^2) \\ dX_{it} &= a(X_{it}, \phi_i)dt + b_\gamma(X_{it}, \phi_i)dB_{it}, \quad X_{i0} = x_0 \\ \phi_i &\sim_{iid} \mathcal{N}(\mu, \Omega) \end{aligned}$$

Parameters to be estimated: $\theta = (\mu, \Omega, \gamma, \sigma)$

SDE mixed models

$$\begin{aligned} y_{ij} &= X_{it_{ij}} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim_{iid} \mathcal{N}(0, \sigma^2) \\ dX_{it} &= a(X_{it}, \phi_i)dt + b_\gamma(X_{it}, \phi_i)dB_{it}, \quad X_{i0} = x_0 \\ \phi_i &\sim_{iid} \mathcal{N}(\mu, \Omega) \end{aligned}$$

Notations

- $X_i = X_{i,0:J} = (X_{t_{i0}}, \ldots, X_{t_{iJ}})$: hidden diffusion of subject i
- $X = (X_1, \ldots, X_n)$

Likelihood for subject i

$$\begin{aligned} p(y_i; \theta) &= \int p(y_i, X_i, \phi_i; \theta) dX_i d\phi_i \\ &= \int p(y_i | X_i; \theta) p(X_i | \phi_i; \theta) p(\phi_i; \theta) dX_i d\phi_i \end{aligned}$$

Likelihood

Girsanov formula gives

$$p(X_i;\theta) = \int \exp\left(\int \frac{a(X_i(s),\phi_i)}{b^2(X_i(s),\phi_i)} dX_i(s) - \frac{1}{2}\int \frac{a^2(X_i(s),\phi_i)}{b^2(X_i(s),\phi_i)} ds\right) p(\phi_i;\theta) d\phi_i$$

But explicit only for linear drift and known volatility

Alternative: discretization of the SDE

$$\begin{aligned} p(y_i;\theta) &= \int\int p(y_i|X_i;\theta)p(X_i|\phi_i;\theta)p(\phi_i;\theta)dX_i d\phi_i \\ &= \int\int \prod_{j=0}^{J} p(y_{ij}|X_{t_{ij}};\theta) \\ &\quad \times \prod_{j=1}^{J} p(X_{t_{ij}}|X_{t_{ij-1}},\phi_i;\theta)p(\phi_i;\theta)dX_i d\phi_i \end{aligned}$$

Estimation methods based on approximations

- Approximation of the conditional distribution
  - Extended Kalman filter
  - Stochastic or deterministic maximisation algorithms
  - [Tornoe et al 2005; Overgaard et al 2005; Delattre and Lavielle, 2013]

- Approximation of the likelihood
  - Gaussian quadrature [Picchini et al, 2010]
  - Laplace approximation [Picchini and Ditlevsen, 2011]
  - Hermite expansion of the transition density if needed [Picchini et al, 2010]
  - Simulation of the hidden SDE [Donnet and Samson, 2008; Donnet et al, 2010; Donnet, Samson, 2013]

Estimation methods based on a simulation step

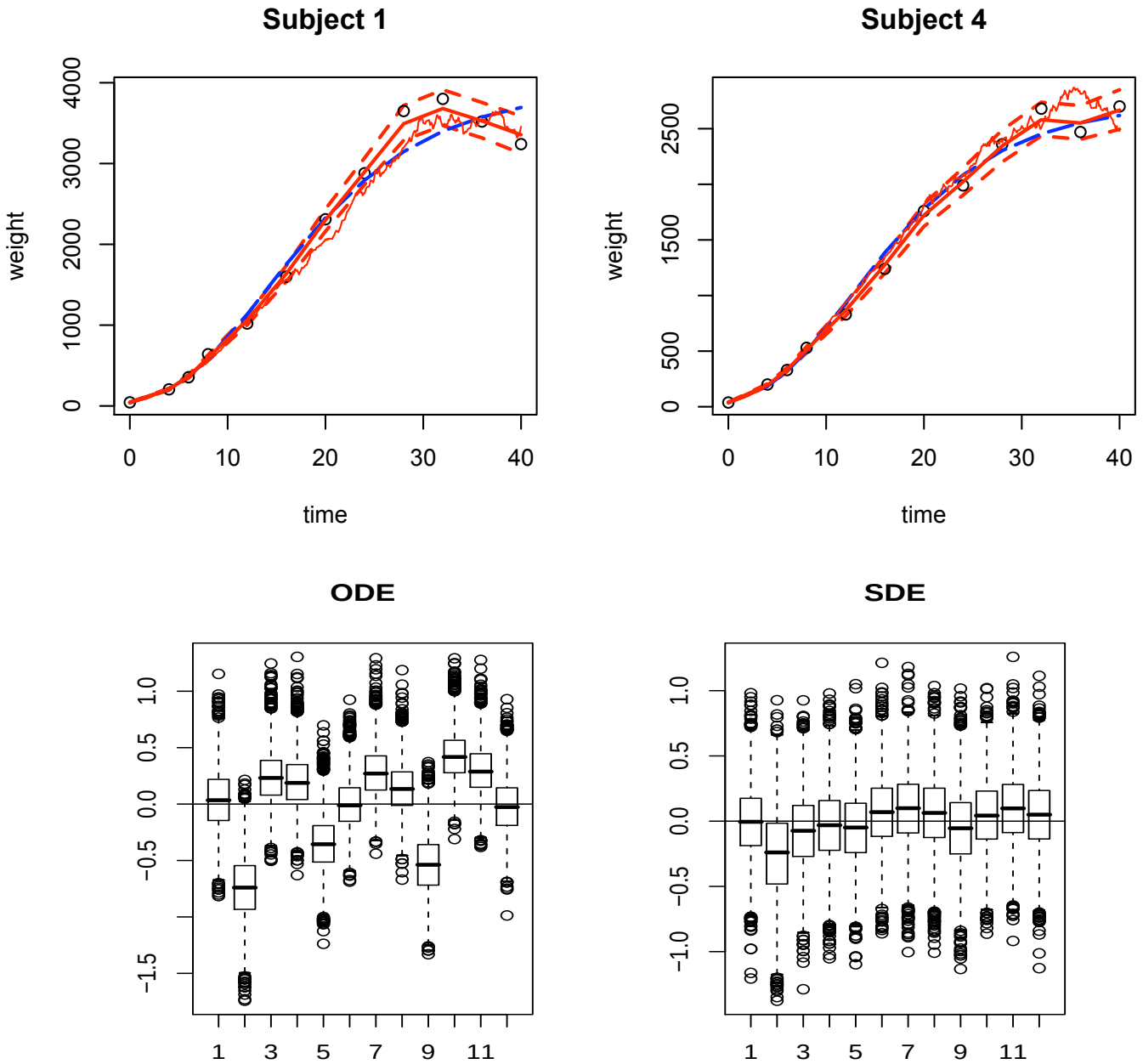
- Estimation algorithms
  - Bayesian [Donnet, Foulley, Samson, 2010]
  - SAEM [Donnet and Samson, 2008; Donnet, Samson, 2013]

- Simulation step
  - For $i = 1, \ldots, n$, simulation of

$$(X_{im}, \phi_{im}) \sim p(X_{i,0:J}, \phi_i | y_{i,0:J}; \widehat{\theta}_m)$$

  - Gibbs algorithm
    - $p(\phi_i | y_{i,0:J}, X_{i,0:J}; \widehat{\theta}_m)$: standard by Metropolis-Hastings
    - $p(X_{i,0:J} | y_{i,0:J}, \phi_i; \widehat{\theta}_m)$: block decomposition and iterative simulation
    - $\Rightarrow$ slow convergence of the chain
  - Particle filter coupled with MCMC (PMCMC)
    - [Del Moral et al, 2001; Doucet et al, 2001; Chopin, 2004; Andrieu et al, 2010]
    - Metropolis Hastings algorithm targeting directly $p(X_{i,0:J}, \phi_i | y_{i,0:J}; \widehat{\theta}_m)$

Improvement of the predictions



Meta-models: another alternative

- Complex mechanistic models
  - Approximation of the solution of the model at each iteration of the estimation method
  - Partial differential equation solution: difficult to obtain

- Meta-model
  - Precomputation on a pre-defined grid
    - Precise evaluation of the solution on the points of the grid
  - Approximation
    - Nearest neighborhood approximation [Barthelemy, Lavielle, submitted]
    - Linear approximation on the grid [Grenier, Louvet, Vigneaux, submitted]
  - Computational cost gain (PDE example) [Grenier, Louvet, Vigneaux, submitted]
    - Exact SAEM: 23 days
    - Interpolation with heterogeneous grid: 26 min

Conclusion

- Mechanistic models
  - Mechanistic description
    - Angiogenesis dynamic
    - Action of molecules
  - Good fitting on real data
  - Need powerful statistical methods because large number of parameters/random effects

- Stochastic models
  - Advantages
    - Less parameters, less equations
    - Allow stochastic individual variations around the deterministic theoretical model
    - Improvement of the predictions
  - Need specific statistical tools to filter the stochastic process
  - Need large number of data $\rightarrow$ pre-clinical data

Perspectives

- Meta-models
  - Nearest neighborhood or linear approximation
    - Estimation on a approximated model
    - Convergence to an approximate MLE
  - More sophisticated meta-models
    - Gaussian process, Reproducing Kernel Hilbert space (RKHS), ...
    - Convergence to the exact MLE joint work with P. Barbillon and C. Barthelemy
- Non parametric estimation
  - Density of the random effects [Comte et al, 2012]
  - Drift function of the stochastic model [Cattiaux et al, submitted]